ประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบเบส์
โดยใช้การสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ

นายชนะศึก นิชานนท์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2553

THE EFFICIENCY OF BAYESIAN PARAMETER ESTIMATION BASED ON GENERALIZABILITY IN ITEM RESPONSE MODELING

Mr. Chanasuek Nichanong

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Educational Measurement and Evaluation
Department of Educational Research and Psychology
Faculty of Education
Chulalongkorn University
Academic Year 2010

หัวข้อวิทยานิพนธ์	ประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบเบส์โดยใช้การสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ
โดย	นายชนะศึก นิชานนท์
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	Prof. Mark Wilson, Ph.D.

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

....คณบดีคณะครุศาสตร์

(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

คณะกรรมการสอบวิทยานิพนธ์

....ประธานกรรมการ

(รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ)

....อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(Prof. Mark Wilson, Ph.D.)

....กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร. ณัฏฐภรณ์ หลาวทอง)

....กรรมการภายนอกมหาวิทยาลัย

(อาจารย์ ดร. ชูศักดิ์ ขัมภลิขิต)

ขนะศึก นิชานนท์: ประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบเบส์โดยใช้การสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ. (THE EFFICIENCY OF BAYESIAN PARAMETER ESTIMATION BASED ON GENERALIZABILITY IN ITEM RESPONSE MODELING) **อ. ที่ปรึกษาวิทยานิพนธ์หลัก : ศ.ดร.ศิริชัย กาญจนวาสี, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม :** Prof. Mark Wilson, Ph.D., 278 **หน้า.**

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการประมาณของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) 4 รูปแบบ ได้แก่ รูปแบบที่ 1 Original GIRM ซึ่ง พัฒนาโดย Brigg และ Wilson (2007) รูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B และรูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งผู้วิจัยเป็นผู้พัฒนาขึ้น นอกจากนี้ยังศึกษาถึงอิทธิพลของขนาดกลุ่มอย่างและจำนวนข้อสอบ รวมทั้งยังศึกษาความไว (Sensitivity) ของรูปแบบต่าง ๆ ต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบและผู้สอบที่ส่งผลต่อประสิทธิภาพของวิธีการประมาณ ซึ่งวัดได้จากดัชนี 3 ประเภท ได้แก่ ความลำเอียงในการประมาณค่า (Biased estimator) คำนวณจากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (MAD) ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) คำนวณจากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (S.D.) และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณจากการวิเคราะห์ระยะทางยุคลิด (EUC) ข้อมูลที่ใช้ในการศึกษาครั้งนี้ใช้ข้อมูลจำลอง (simulation) จากโปรแกรม R และทำการประมวลผลภายใต้จากการเขียนคำสั่งการประมวลผลด้วยโปรแกรม WinBUGS ด้วย Package R2 WinBUGS

ผลการวิจัยพบว่า

1. การเปรียบเทียบประสิทธิภาพในการประมาณค่าของรูปแบบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดด้วยโมเดลการตอบสนองข้อสอบ(GIRM) พบว่า ความลำเอียงในการประมาณค่า รูปแบบที่ 1 กับ รูปแบบที่ 4 ให้ค่าประสิทธิภาพสูงที่สุด โดยรูปแบบที่ 4 เป็นรูปแบบที่ประมาณค่าพารามิเตอร์ได้เฉพาะลักษณะการแจกแจงเริ่มแรกของผู้สอบและข้อสอบแบบปกติ สำหรับความไม่แน่นอนในการประมาณค่า พบว่า รูปแบบที่ 4 ให้ค่าประสิทธิภาพสูงที่สุด สำหรับลักษณะการแจกแจงเริ่มแรกของผู้สอบและข้อสอบแบบปกติ ส่วนลักษณะการแจกแจกเริ่มแรกของค่าพารามิเตอร์ตัวใดตัวหนึ่งที่ไม่มีลักษณะการแจกแจงเริ่มแรกแบบปกติพบว่า รูปแบบที่ 1 ให้ค่าประสิทธิภาพสูงที่สุด และเมื่อพิจารณาในด้านประสิทธิภาพขององค์ประกอบความแปรปรวนยุคลิด พบว่า รูปแบบที่ 2 ให้ค่าประสิทธิภาพสูงที่สุด

2. การศึกษาอิทธิพลของขนาดกลุ่มตัวอย่างและจำนวนข้อสอบ พบว่า ขนาดกลุ่มตัวอย่าง ส่งผลต่อประสิทธิภาพในด้านความลำเอียงในการประมาณค่าในรูปแบบที่ 1 2 และ 3 ส่วนความไม่แน่นอนในการประมาณค่า พบว่า ขนาดกลุ่มตัวอย่างส่งผลต่อประสิทธิภาพในการประมาณค่าในทุกรูปแบบ และด้านประสิทธิภาพขององค์ประกอบความแปรปรวนยุคลิด พบว่าขนาดกลุ่มตัวอย่างไม่ส่งผลต่อประสิทธิภาพในการประมาณค่า สำหรับความยาวแบบสอบ พบว่า ส่งผลต่อการการวัดประสิทธิภาพในความลำเอียงในการประมาณค่าและการวิเคราะห์ประสิทธิภาพองค์ประกอบความแปรปรวนยุคลิด ในรูปแบบที่ 1 2 และ 3 ส่วนความไม่แน่นอนในการประมาณค่า พบว่า ขนาดกลุ่มตัวอย่างส่งผลต่อประสิทธิภาพในการประมาณค่าในทุกรูปแบบ

3. การศึกษาความไว พบว่า การแจกแจงเริ่มแรกของผู้สอบ ไม่ส่งผลต่อประสิทธิภาพด้านความลำเอียง ส่วนความไม่แน่นอนในการประมาณค่าและประสิทธิภาพขององค์ประกอบความแปรปรวนยุคลิด พบว่าการแจกแจงเริ่มแรกของผู้สอบ ส่งผลต่อการวัดประสิทธิภาพความลำเอียงในทุกรูปแบบ ส่วนการแจกแจงเริ่มแรกของข้อสอบ พบว่า ส่งผลต่อการวัดประสิทธิภาพความลำเอียงในการประมาณค่าและความไม่แน่นอนในการประมาณค่าทุกรูปแบบ และส่งผลการวิเคราะห์ประสิทธิภาพองค์ประกอบความแปรปรวนยุคลิดเฉพาะในกรณีที่การแจกแจงเริ่มแรกของผู้สอบเป็นแบบแกมมาเท่านั้น

ภาควิชา วิจัยและจิตวิทยาการศึกษา ลายมือชื่อนิสิต ขนะศึก นิชานนท์
สาขาวิชา การวัดและประเมินผลการศึกษา ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก
ปีการศึกษา 2553 ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์ร่วม

5084209127: MAJOR IN EDUCATIONAL MEASUREMENT AND EVALUATION

KEYWORDS: QUALITY/THESES QUALITY/ MASTER DEGREE THESES

CHANASUEK NICHANONG: EFFICIENCY OF BAYESIAN PARAMETER ESTIMATION BASED ON GENERALIZABILITY IN ITEM RESPONSE MODELING. THESIS ADVISOR: PROF. SIRICHAI KARNJANAWASEE, Ph.D., THESIS CO-ADVISOR: PROF. MARK WILSON, Ph.D., 278 pp.

This research aimed to compare the efficiency of the 4 forms of GIRM; Form 1 Original GIRM developed by Brigg และ Wilson (2007), Form 2 AGIRM A, Form 3 AGIRM B, and Form 4 Numerical Bayesian GIRM developed by the researcher. The research also studied the impact of the sample size, the test length, and the sensitivity of each form toward the prior distribution setting of parameter of the items and the examiners which affected the efficiency of the estimation which could be measured by 3 types of indicators; Biased estimator calculated from Mean Average Deviation(MAD), Uncertainty estimator calculated from Standard Deviation (S.D.), and efficiency of constituent variance estimation calculated from Euclidean Distance(EUC). The data was simulation from program R and was assessed with program WinBUGS with Package R2 WinBUGS.

The results of research were as follows:

1.From the comparison of GIRM, it was found that when Bias estimated from MAD was considered, Form 1 Original GIRM and Form 4 Numerical Bayesian GIRM had the best efficiency. Form 4 could estimate normal prior distribution of the items and the examiners. For Uncertainty, it was found that Form 4 is the best efficiency for normal prior distribution. For the prior distribution of one parameter without normal prior distribution, Form 1 Original GIRM had the best efficiency. And when the efficiency of constituent variance estimation calculated from EUC was considered, it was found that Form 2 AGIRM A had the best efficiency.

2. From the study of the influence of the sample size and the test length, it was found that the sample size affected the measure of bias efficiency in Form 1 , Form 2 , and Form 3 . For the uncertainty in the estimation, it was found that the sample size affected all forms of efficiency estimation. And when the efficiency of constituent variance EUC was considered, it was found that the sample size did not affect efficiency estimation. The test length affected the measure of bias efficiency and efficiency of constituent variance EUC in Form 1, Form 2, and Form 3. For uncertainty of estimation, it was found that the sample size affected all forms of efficiency estimation.

3. From the study of sensitivity of the test analysis result, it was found that the prior distribution of the examiners did not affect the measure of bias efficiency. For the uncertainty of estimation and analysis of constituent variance EUC, it was found that the prior distribution of the items affected the measure of bias efficiency. It was found that the prior distribution affected the measure of Bias estimator and all forms of uncertainty estimator. It also affected the efficiency of constituent variance EUC only when the prior distribution of the examiners was gamma.

Department: .....Education Research and Psychology........ Student's Signature Chanasuek Nichanong

Field of Study: ..Educational Measurement and Evaluation... Advisor's Signature S. Kajanmasee

Academic Year ..2010 .... Co-Advisor's Signature....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ไม่มีทางสำเร็จได้ด้วยดี ถ้าปราศจากความเมตตาและความกรุณาของศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาวิทยานิพนธ์ที่ได้ให้ความรู้ ช่วยปรับปรุงแก้ไข และให้กำลังใจกระผมตลอดเวลาที่ผมศึกษาต่อในระดับปริญญาเอกทั้งในประเทศและต่างประเทศ พร้อมกันนี้ยังได้มอบประสบการณ์งานวิจัยโดยให้กระผมเป็นผู้ช่วยนักวิจัยในงานวิจัยที่สำคัญหลายโครงการ นอกจากนี้ยังผลักดันกระผมให้ได้รับทุนการศึกษาและมีโอกาสไปศึกษาและวิจัยใน University of California, Berkeley ประเทศสหรัฐอเมริกา กระผมรู้สึกซาบซึ้งและขอกราบขอบพระคุณท่านอาจารย์มา ณ โอกาสนี้ด้วย

ในระหว่างที่กระผมศึกษาที่ University of California, Berkeley USA กระผมได้รับความเมตตาเป็นอย่างสูงจาก Prof. Dr. Mark Wilson อาจารย์ที่ปรึกษาเป็นอย่างดีเยี่ยมสำหรับคำแนะนำในการทำวิทยานิพนธ์และความรู้ต่างๆ เกี่ยวกับทฤษฎีการวัดและประเมินผล นอกจากนี้ยังได้มอบโอกาสในการเป็นผู้ช่วยวิจัยในโครงการวิจัยที่สำคัญๆ หลายโครงการใน BEAR center ซึ่งเป็นศูนย์วิจัยทางด้านการวัดแประเมินผลที่มีชื่อเสียงของประเทศสหรัฐอเมริกา ทำให้ผู้วิจัยได้รับความรู้และประสบการณ์อย่างดียิ่ง นอกจากนี้ผู้วิจัยยังได้รับความเมตตาเกี่ยวกับความรู้และวิธีการเขียนโปรแกรมสำหรับการวิเคราะห์ GIRM จากผู้คิดค้นวิธีการ GIRM โดยตรงนั้นก็คือ Prof.Dr. Derek C. Briggs รวมทั้งผู้วิจัยยังได้รับความช่วยเหลือในการตรวจสอบคำสั่งต่างๆ สำหรับการวิเคราะห์ GIRM จาก Dr. Elizabeth Ayers และ Jinnie Choi เพื่อนร่วมกลุ่มที่สนใจในเรื่องเดียวกันที่ UC Berkeley ผู้วิจัยขอกราบขอบพระคุณมา ณ โอกาสนี้ด้วย

ขอกราบขอบพระคุณ รศ.ดร. ศิริเดช สุชีวะ อ.ดร. ชูศักดิ์ ขัมภลิขิต และผศ.ดร.ณัฏฐภรณ์ หลาวทอง คณะกรรมการวิทยานิพนธ์ที่ช่วยขัดเกลางานวิจัยของกระผมให้มีความชัดเจนและมีความสมบูรณ์เพิ่มมากขึ้น นอกจากนี้กระผมยังขอกราบขอบพระคุณคณาจารย์ภาควิชาวิจัยและจิตวิทยาการศึกษา ได้แก่ ศ.ดร.นงลักษณ์ วิรัชชัย ศ.ดร.สุวิมล ว่องวานิช รศ.ดร.อวยพร เรืองตระกูล และ รศ.ดร.โชติกา ภาษีผล ที่ได้ให้ความรู้กับผมอันเป็นพื้นฐานความรู้ที่สำคัญสำหรับการทำวิทยานิพนธ์ฉบับนี้

ขอขอบคุณเพื่อนๆ พี่ๆ น้องๆ ในภาควิชาวิจัยและจิตวิทยาการศึกษาทุกท่านที่ได้ช่วยเหลือและให้กำลังผู้วิจัยเสมอมาไม่ว่าจะเป็น คุณธเกียรติกมล ทองงอก ที่เป็นคนนำบทความเกี่ยวกับ GIRM มาให้กระผมศึกษาจนเป็นที่มาของวิทยานิพนธ์ฉบับนี้ คุณวราพร เอราวรรณ์ คุณสาธิดา สกุลรัตนกุลชัย คุณศักดิ์สิทธิ์ ฤทธิลัน คุณทัศศิรินทร์ สว่างบุญ และคนอื่นๆที่ไม่สามารถจะเอ่ยนามได้หมด ณ ที่นี้

ผู้วิจัยยังขอขอบพระคุณ สำนักงานกองทุนสนับสนุนการวิจัย(สกว.)ที่ได้ให้ทุนการศึกษาในโครงการปริญญาเอกกาญจนาภิเษก ขอขอบพระคุณ รศ.ดร.ศิโรจน์ ผลพันธิน ท่านอธิการบดีมหาวิทยาลัยราชภัฏสวนดุสิตที่สนับสนุนและส่งเสริมให้กระผมศึกษาต่อในระดับปริญญาเอก นอกจากนี้กระผมยังกราบขอบพระคุณ ดร.สุชาดา โทผล ผู้อำนวยการสถาบันวิจัยและพัฒนาในขณะนั้น ที่คอยสนับสนุนและให้กำลังใจจนกระผมสำเร็จการศึกษาครั้งนี้

ท้ายที่สุดนี้ ถ้าปราศจากบุคคลทั้งสองนี้ กระผมคงไม่สามารถสำเร็จการศึกษาในระดับปริญญาเอกได้นั้นก็คือ บิดา(นายระพินทร์ นิชานนท์) และมารดา(นางจรรยา นิชานนท์)ของกระผม ที่คอยส่งเสริมและสนับสนุนกระผมเสมอมาด้วยความรักและความเอาใจใส่เป็นอย่างดีเยี่ยม

สารบัญ

สารบัญตาราง

สารบัญแผนภาพ

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ปัจจุบันสามารถแบ่งโมเดลการวิเคราะห์ข้อสอบออกได้เป็น 2 ประเภทหลัก ได้แก่ การวิเคราะห์ข้อสอบตามหลักทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) และทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory: MTT) โดยทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) เป็นทฤษฎีลักษณะโมเดลเชิงเส้น (Linear Model) ที่เชื่อว่าคะแนนสอบที่สังเกตได้ (X) เป็นผลรวมของโมเดลเชิงเส้นระหว่างคะแนนจริงหรือความสามารถที่แท้จริงของผู้สอบ (T) กับความคลาดเคลื่อนที่เกิดขึ้นทั้งที่เป็นระบบและไม่เป็นระบบ (ε) โดยความคลาดเคลื่อนที่เกิดขึ้นนี้มีลักษณะเป็นหนึ่งเดียว ไม่สามารถแบ่งแยกได้ (unique error) ซึ่งทำให้เกิดเป็นข้อจำกัดที่สำคัญในการประมาณค่าความเที่ยงของข้อสอบที่ใช้ในการทดสอบเนื่องจากทำให้สามารถวิเคราะห์แหล่งความคลาดเคลื่อนได้ เพียงครั้งละหนึ่งแหล่งเท่านั้น นอกจากนี้จากข้อตกลงเบื้องต้นที่สำคัญของทฤษฎีที่ว่า ค่าพารามิเตอร์ของข้อสอบเป็นค่าเฉพาะของกลุ่มผู้สอบและเป็นอิสระไม่ขึ้นอยู่กับความสามารถของผู้สอบ จึงทำให้ค่าความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement; SEM) เป็นค่าเฉพาะในแต่ละประชากรที่ใช้ในการทดสอบและเป็นค่าเดียวที่คงที่สำหรับทุกคนในประชากรของการทดสอบ ดังนั้นในการเปรียบเทียบคะแนนหรือคุณภาพของข้อสอบ จึงทำได้เฉพาะข้อสอบที่ลักษณะเป็นข้อสอบคู่ขนานเท่านั้น และยังทำให้ค่าพารามิเตอร์ของข้อสอบผันแปรไปตามกลุ่มผู้สอบ คะแนนที่สังเกตที่ได้จากการสอบหรือความสามารถของผู้สอบจึงขึ้นอยู่กับข้อสอบและแบบข้อสอบที่นำมาใช้ในการวิเคราะห์ ดังนั้นทฤษฎีการทดสอบแบบดั้งเดิมจึงมีค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ของผู้สอบไม่เป็นอิสระต่อกัน โดยขึ้นอยู่กับสถานการณ์แต่ละสถานการณ์ในการทดสอบ (ศิริชัย กาญจนวาสี, 2550)

สำหรับทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory; MTT) ในปัจจุบันที่ใช้กันอย่างแพร่หลายมี 2 ทฤษฎี ได้แก่ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory; GT) และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) โดยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เป็นทฤษฎีที่อธิบายแหล่งของความคลาดเคลื่อนที่เกิดจากหลายแหล่ง (Multiple sources of error) โดยใช้วิธีการวิเคราะห์ความแปรปรวน (ANOVA) เข้ามาช่วยในการอธิบายทำให้สามารถแยกส่วนความคลาดเคลื่อนของการวัดได้ออกเป็นสอง

แหล่งหลัก ๆ ประกอบด้วยความคลาดเคลื่อนอย่างเป็นระบบ (systematic error variance) และความคลาดเคลื่อนแบบสุ่ม (random error variance) รวมกันเป็นความคลาดเคลื่อน (ε) ของทฤษฎีการทดสอบแบบดั้งเดิม ดังนั้นคะแนนสอบของผู้สอบที่ได้จากการวิเคราะห์แบบทฤษฎีการสรุปอ้างอิง จึงเป็นคะแนนที่คาดหวังของบุคคล (person 's expected score) ไม่ใช่คะแนนที่สังเกตได้ (person's observed score) เนื่องจากคะแนนตามทฤษฎีนี้เป็นคะแนนที่ได้มาจากการสุ่มตามเงื่อนไขของแหล่งความคลาดเคลื่อนที่น่าจะเกิดขึ้น (randomly sampled facet conditions) ซึ่งสามารถแก้ข้อจำกัดของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ตามข้อตกลงเบื้องต้นที่ว่าความคลาดเคลื่อนมีลักษณะเป็นหนึ่งเดียวที่ไม่สามารถแบ่งแยกได้ (unique error) ซึ่งทำให้สามารถวิเคราะห์ค่าความเที่ยงของการวัดได้เพียงครั้งละแหล่งของความคลาดเคลื่อนเท่านั้น

ส่วนทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่แท้จริงของบุคคลกับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกได้มากน้อยเพียงใด อธิบายโดยใช้โค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันโลจิส (logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (normal ogive function) ซึ่งทำให้ค่าพารามิเตอร์ของข้อสอบที่ได้จากการวิเคราะห์คุณภาพของข้อสอบ ทั้งค่าความยาก (a) ค่าอำนาจจำแนก (b) และค่าโอกาสในการเดาของข้อสอบ (c) แต่ละข้อเป็นคุณลักษณะเฉพาะของข้อสอบที่มีอยู่ประจำ และคงที่ในตัวข้อสอบ จึงทำให้การวิเคราะห์ข้อสอบด้วยทฤษฎีนี้ ค่าพารามิเตอร์ของข้อสอบจึงไม่ผันแปรหรือเปลี่ยนแปลงไปตามกลุ่มผู้สอบ นอกจากนี้ค่าพารามิเตอร์ของผู้สอบหรือค่าความสามารถที่แท้จริงของผู้สอบยังเป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบแต่ละคนที่ไม่เปลี่ยนแปลงไปตามคุณลักษณะของข้อสอบ ผลการวิเคราะห์จึงสามารถแก้ข้อจำกัดที่สำคัญของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ได้

จากทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) ทั้งสองทฤษฎีสามารถทำให้ผลการวิเคราะห์ข้อสอบมีความถูกต้องมากยิ่งขึ้นโดยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เป็นการวิเคราะห์ที่สามารถอธิบายแหล่งความคลาดเคลื่อนของการวัดได้มากขึ้นและสามารถสรุปอ้างอิงไปยังสถานการณ์ต่างๆ ของการวัดในแต่ละแหล่งความคลาดเคลื่อนได้ ซึ่งอาจเรียกทฤษฎีนี้ได้ว่าเป็นโมเดลแบบสุ่ม (sampling model) (Brenan, 2001) แต่ทฤษฎียังไม่สามารถวิเคราะห์ค่าพารามิเตอร์ของข้อสอบเป็นรายข้อและพารามิเตอร์ของผู้สอบเป็นรายบุคคลได้ ในขณะที่ทฤษฎีการตอบสนองข้อสอบ (IRT) สามารถให้ทั้งสารสนเทศที่เป็นค่าพารามิเตอร์ของข้อสอบเป็นรายข้อ (item parameter) และพารามิเตอร์ของผู้สอบเป็นรายบุคคล (person parameter) ได้ ซึ่งอาจเรียกทฤษฎีนี้ได้ว่าเป็นโมเดลการกำหนดมาตรวัด (scaling model) (Brenan, 2001) จาก

จุดเด่นและข้อจำกัดในแต่ละทฤษฎี จึงทำให้นักวัดผลมีความพยายามในการรวมทั้งสองทฤษฎีเข้าไว้ด้วยกันเพื่อรวมจุดเด่นและลดข้อจำกัดในแต่ละทฤษฎี เพื่อการวิเคราะห์ข้อสอบที่มีความถูกต้องมากที่สุด (Kolen และ Harris, 1987; Mislevy, 1993; Verhelst และ Verstralen, 2001; Patz และคณะ, 2002; Bock, Brennan และ Muraki, 2002; Glas และ Van Der Linder, 2003; Briggs และ Wilson, 2007)

การศึกษาวิธีการในการรวมสองทฤษฎีเข้าไว้ด้วยกัน เริ่มต้นขึ้นเมื่อปี 1987 โดย Kolen และ Harris ได้พัฒนา Kolen และ Harris ด้วยการเชื่อมโยงสูตรสำหรับการวิเคราะห์องค์ประกอบความแปรปรวนและค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัดโดย การประมาณได้จากการวิเคราะห์ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบหลายตัวแปร (Multivariate GT) ด้วยกรอบแนวคิดของโมเดลการตอบสนองข้อสอบ (IRT) นอกจากนี้ Mislevy (1993) ซึ่งได้ทำการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบด้วยการใช้วิธีการประมาณค่าแบบเบส์ (Bayesian estimation) โดยการสุ่ม (randomness) เข้าไปในการวิเคราะห์ตามแนวของทฤษฎีการตอบสนองข้อสอบ (IRT) ดังนั้นจะเห็นได้ว่าผลงานวิจัยทั้งสองวิธีนั้นเป็นวิธีในการผ่อนคลายข้อตกลงเบื้องต้นการไม่สรุปอ้างอิงข้อคำถาม (fixed items) ของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อทำให้สามารถเชื่อมโยงกับการสรุปอ้างอิงของการวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ที่สามารถสรุปอ้างอิงข้อคำถาม (Generalizability item) ได้

สำหรับการประยุกต์การเชื่อมโยงวิธีวิเคราะห์ข้อสอบของทฤษฎีทั้งสองในทางปฏิบัติของการออกแบบการวัด เริ่มขึ้นในปี 2001 โดย Verhelst และ Verstralen ได้ศึกษาวิธีการวิเคราะห์ข้อสอบโดยใช้โมเดลการตอบสนองข้อสอบสำหรับการประเมินผลการปฏิบัติงานด้วยผู้ตรวจหลายคน ซึ่งทำให้สารสนเทศที่ได้จากการทฤษฎีการตอบสนองข้อสอบสามารถอธิบายแหล่งความคลาดเคลื่อนของผู้ตรวจได้เหมือนในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) นอกจากนี้ยังมีผลงานวิจัยในการออกแบบที่คล้ายคลึงกันของ Patz และคณะ (2002) และBock, Brennan และ Muraki (2002) โดยการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) ในสถานการณ์ที่มีผู้ประเมินคะแนนการปฏิบัติงานหลายคน ดังนั้นจะเห็นได้ว่า การออกแบบการวัดในการศึกษาของสามเรื่องนี้ได้นำไปสู่ การเชื่อมโยงระหว่าง ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) กับทฤษฎี การตอบสนองข้อสอบ (IRT) และยังเป็นการขยายกรอบแนวคิดที่เด่นชัดเกี่ยวกับโมเดลผลกระทบแบบสุ่มพหุระดับ (Hierarchical random effect models) อีกด้วย และในปี 2003 Glas และ Van Der Linndon จึงได้ริเริ่มมีการประยุกต์ใช้โปรแกรมคอมพิวเตอร์ตามแนวคิดของการ

ทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (Adaptive testing) โดยการใช้คอมพิวเตอร์ช่วยสุ่มข้อสอบมาใช้ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ ซึ่งสามารถทำให้สามารถผ่อนคลายข้อตกลงเบื้องต้นของการไม่สรุปอ้างอิงข้อคำถาม (fixed items) ของทฤษฎีการตอบสนองข้อสอบ (IRT) ได้เช่นกัน

จนกระทั่งมาถึงในปัจจุบันได้มีการพัฒนาวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) (Briggs และ Wilson, 2007) ซึ่งเป็นการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการกำหนดมาตรวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) โดยอาศัยสมมติฐานของการกระจายไปยังแหล่งความคลาดเคลื่อนจากการวัดที่เกี่ยวข้อง การวิเคราะห์ข้อสอบด้วยวิธีการ GIRM นี้จะให้สารสนเทศของผลการวิเคราะห์ทั้งแบบทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และแบบทฤษฎีการตอบสนองข้อสอบ (IRT) ในการวิเคราะห์เพียงครั้งเดียว โดยในส่วนของผลการวิเคราะห์ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) สามารถแยกการประมาณค่าความแปรปรวนขององค์ประกอบความแปรปรวนระหว่างความคลาดเคลื่อนของปฏิสัมพันธ์ของแหล่งความคลาดเคลื่อน (facet interaction effect) และความคลาดเคลื่อน (error effect) ออกจากกันได้ ซึ่งในการวิเคราะห์ข้อสอบด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ไม่สามารถแยกความแปรปรวนดังกล่าวออกจากกันได้ นอกจากนี้การวิเคราะห์ด้วย GIRM ยังไม่ต้องมีการปรับข้อมูลที่ขาดหายไป (missing data) ในเมทริกซ์ของข้อมูล ในขณะที่ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ต้องมีการออกแบบข้อมูลแทนที่ขาดหายไป ทั้งนี้เนื่องจากการประมาณค่าองค์ประกอบความแปรปรวนตั้งอยู่บนพื้นฐานของการคาดคะเน (expected) มากกว่าข้อมูลจริงที่อยู่ในตารางเมทริกซ์ของข้อมูล จึงทำให้ส่งผลต่อการวิเคราะห์ข้อมูล เนื่องจากการลงข้อมูล (specified) ที่ไม่สมดุลสำหรับการออกแบบแบบสุ่ม (unbalanced random effects design) จะไม่สามารถประมาณค่าในขั้นตอนการศึกษาเชิงสรุปอ้างอิง (Generalizability Study; G-Study) และ การศึกษาเชิงตัดสินใจ (Decision Study; D-study) ในโปรแกรม GENOVA ได้

วิธีการประมาณค่าของวิธี GIRM จะประมาณค่าเมทริกซ์ที่คาดหวังโดยใช้เทคนิค MCMC (Markov Chain Monte Carlo) แบบ Metropolis-Hastings algorithm ภายในวิธีการ Gibbs sampler ซึ่งจะสามารถประมาณค่า (Posterior distribution) ของความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ) และค่าพารามิเตอร์ของข้อสอบแต่ละข้อ (β) สามารถแสดงสมการทางคณิตศาสตร์ได้ดังนี้

สมการสำหรับประมาณค่าความสามารถที่แท้จริงของผู้สอบ (θ)

$$P(\theta \mid \beta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\theta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\theta}$$

สมการสำหรับประมาณค่าพารามิเตอร์ของข้อสอบ (β)

$$P(\beta \mid \theta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\beta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\beta}$$

จากสมการ $P(\theta \mid \beta, X)$ และสมการ $P(\beta \mid \theta, X)$ ที่สามารถประมาณได้จากเทคนิค MCMC จะทำให้สามารถหาค่าของ $P(\theta_p, \beta_i)$ หรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบที่มีระดับความสามารถ θ_p ของข้อสอบที่มีระดับความยาก β_i ของแต่ละคนในแต่ละข้อของฟังก์ชันการตอบสนองข้อสอบโมเดลแบบราสซ์ (Rasch model item response function) หลังจากนั้นจึงเอามาสร้างเป็นเมทริกซ์ที่คาดหวังมาประมาณองค์ประกอบความแปรปรวนและค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด โดยประมาณได้จากการวิเคราะห์ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ด้วยกรอบแนวคิดของทฤษฎีการตอบสนองข้อสอบ (IRT) ที่ได้จากการเชื่อมโยงสูตรของ Kolen และ Harris (1987) ดังนี้

$$\mu = \int_{\theta}\int_{\beta} P(\theta_p, \beta_i) f(\beta) g(\theta) d\beta d\theta$$ **(คำอธิบาย μ ใน GT)**

$$\pi(\theta) = \int_{\beta} P(\theta_p, \beta_i) f(\beta) d\beta - \mu$$ **(คำอธิบาย ν_p ใน GT)**

$$\iota(\theta) = \int_{\theta} P(\theta_p, \beta_i) g(\theta) d\theta - \mu$$ **(คำอธิบาย ν_i ใน GT)**

$$\nu(\theta, \beta) = P(\theta_p, \beta_i) - \pi(\theta) - \iota(\beta) - \mu$$ **(คำอธิบาย ν_{pi} ใน GT)**

$$\sigma^2(e) = \int_{\theta}\int_{\beta} P(\theta_p, \beta_i)[1 - P(\theta_P, \beta_i)] f(\beta) g(\theta) d\beta d\theta$$ **(คำอธิบาย ν_e ใน GT)**

ดังนั้นจะเห็นได้ว่าการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) สามารถวิเคราะห์ความแปรปรวนแยกระหว่างองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคล $(\nu(\theta, \beta))$ กับข้อสอบกับความคลาดเคลื่อนแบบสุ่ม

($\sigma^2(e)$) ได้ ในขณะที่การวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ไม่สามารถแยกองค์ประกอบดังกล่าวได้ โดยคำนวณรวมองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบกับความคลาดเคลื่อนแบบสุ่ม ($\sigma^2(pi,e)$) เข้าไว้ด้วยกัน

โดยสรุปในการประมาณค่าในวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เริ่มต้นจากการประมาณค่าการแจกแจงภายหลัง (Posterior distribution) ($P(\theta_p, \beta_i)$) ของค่าพารามิเตอร์ความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ_p) และค่าพารามิเตอร์ของข้อสอบแต่ละข้อ (β_i) หลังจากนั้นจึงนำมาสร้างเป็นเมทริกซ์ที่คาดหวัง (Expected Response Matrix) ระหว่างค่าพารามิเตอร์ความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ_p) และค่าพารามิเตอร์ของข้อสอบแต่ละข้อ (β_i) ในแต่ละเซลล์ของเมทริกซ์ที่คาดหวัง ซึ่งค่าในแต่ละเซลล์นี้ได้มาจากการนำค่าพารามิเตอร์ทั้งสองซึ่งได้แก่ ค่าพารามิเตอร์ความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ_p) และค่าพารามิเตอร์ของข้อสอบแต่ละข้อ (β_i) มาแทนค่าในโมเดลการตอบสนองข้อสอบ (Item Response Model) [$E(X_{pi}) = P(\theta_P, \beta_i)$] หลังจากนั้นจึงเอาค่าในแต่ละเซลล์ของเมทริกซ์ที่คาดหวัง (Expected Response Matrix) มาทำการประมาณค่าองค์ประกอบความแปรปรวน (variance component) ตามสูตรของ Kolen และ Harris ซึ่งจะทำให้ได้ค่า $\pi(\theta)$ $i(\theta)$ และ $v(\theta,\beta)$ ซึ่งในการประมาณค่าแบบนี้จะไม่ทำให้เกิดความลำเอียง (Searle และคณะ, 1992) นอกจากนี้ยังต้องคำนึงถึงสมมติฐานเบื้องต้นของบุคคลและข้อคำถามที่สุ่มว่ามีวิธีการสุ่มที่ถูกต้องหรือไม่ และมีการกระจายก่อนนำมาคำนวณอย่างไร (Prior distribution)

ในทางปฏิบัติค่าพารามิเตอร์ θ_p และ β_i เป็นค่าที่ไม่ทราบและจะต้องถูกประมาณขึ้น โดยสามารถประมาณค่าพารามิเตอร์ดังกล่าว ได้จากสมการการแจกแจงภายหลัง (Posterior distribution) ด้วยวิธีการประมาณค่า MCMC ของโมเดลการตอบสนองข้อสอบ ในแต่ละขั้นของการประมาณค่าด้วยวิธีการ Markov Chain จะทำให้สามารถประมาณค่า θ_p และ β_i ได้ เป็น $\hat{\theta}_p^{(m)}$ และ $\hat{\beta}^{(m)}{}_i$ ซึ่งค่าที่ได้จากการประมาณค่านี้จะเป็นองค์ประกอบในเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) และนำมาใช้ในการคำนวณตามสูตรของ Kolen และ Harris จะทำให้ได้องค์ประกอบของความแปรปรวน $\hat{\sigma}^2(\pi)^{(m)}$ $\hat{\sigma}^2(i)^{(m)}$ $\hat{\sigma}^2(\gamma)^{(m)}$ และ $\hat{\sigma}^2(e)^{(m)}$ ดังนั้นจะเห็นได้ว่าการประมาณค่าตามวิธีการดังกล่าวของวิธีการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) นำไปสู่การฝ่าฝืนข้อตกลงเบื้องต้นแบบสุ่มของการวิเคราะห์ ทั้งนี้สาเหตุน่าจะมาจากการป้อนคำสั่งในกระบวนการการประมาณค่าความแปรปรวนขององค์ประกอบความคลาดเคลื่อนของขั้นตอนการประมาณด้วยวิธี Markov Chain ที่ไม่สามารถทำให้เป็นอิสระจากกันได้ จึงนำไปสู่การฝ่าฝืนข้อตกลงเบื้องต้นซึ่งมีต่อผลกระทบแบบ

สุ่ม (random effects) เช่น การประมาณค่าองค์ประกอบความแปรปรวนของ $\sigma^2(i)$ ตามวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) แบบ p X i design ด้วยวิธีการ Markov Chain เพื่อนำมาสร้างเป็นเมทริกซ์คะแนนที่สังเกตได้ของบุคคล ในการคำนวณดังกล่าวต้องการคำนวณค่า $\sigma^2(i)$ ซึ่งในการประมาณค่าจะต้องมีการเปลี่ยนแปลงเฉพาะคอลัมน์ในเมทริกซ์ โดยค่าในแถวของเมทริกซ์จะต้องคงที่ แต่ในการวิเคราะห์ด้วยโมเดล GIRM นั้นใช้การประมาณค่าด้วยวิธี Markov Chain ซึ่งจะทำให้ทั้งคอลัมน์ และแถวในเมทริกซ์มีค่าที่ผันแปรร่วมกัน จึงทำให้การประมาณพารามิเตอร์ของวิธีการดังกล่าวมีแนวโน้มในการเกิดความไม่แน่นอนในการประมาณค่า (Briggs และ Wilson, 2007)

ดังนั้นในการศึกษาครั้งนี้ ผู้วิจัยจึงสนใจศึกษาวิธีการประมาณค่าของวิธีการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เพื่อไม่ให้นำไปสู่การฝ่าฝืนข้อตกลงเบื้องต้นแบบสุ่ม โดยศึกษารูปแบบการประมาณค่าวิธีการวิเคราะห์ข้อสอบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) 4 รูปแบบ ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM ซึ่งเป็นรูปแบบการประมาณค่าเริ่มแรกของวิธีการ GIRM ที่ Briggs และ Wilson (2007) ได้พัฒนาขึ้นสำหรับรูปแบบที่ 2 Alternative GIRM A รูปแบบที่ 3 Alternative GIRM B ทั้งสองรูปแบบเช่นรูปแบบที่ผู้วิจัยปรับวิธีการประมาณค่าพารามิเตอร์ขึ้นมาใหม่ โดยการนำค่าการแจกแจงภายหลัง (Posterior distribution) ที่ประมาณค่าพารามิเตอร์ (θ_p) และ(β_i) ได้มาจำลองเมทริกซ์ค่าสังเกตใหม่ด้วยการคำนวณคะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) ซึ่งเป็นค่าเฉลี่ยของคะแนนที่สังเกตได้ โดยได้มาจากการวิเคราะห์ได้ใช้เทคนิค MCMC (Markov Chain Monte Carlo) ดังนั้นจะทำให้ได้ Predictive Observed Matrix แทน Expected Response Matrix ที่ได้จากการประมาณค่าด้วยรูปแบบที่ 1 Original GIRM ซึ่งน่าจะช่วยแก้ไขปัญหาการประมาณค่าพารามิเตอร์ที่ไม่เป็นอิสระต่อกันระหว่างแถวและคอลัมน์ในเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) และส่งผลทำให้การประมาณค่าที่ได้จากการวิเคราะห์ในรูปแบบที่ 1 Original GIRM เกิดความไม่แน่นอนในการประมาณค่า โดยรูปแบบที่ 2 Alternative GIRM A และ รูปแบบที่ 3 Alternative GIRM B จะมีลักษณะต่างกันเฉพาะในขั้นตอนการประมาณค่าองค์ประกอบความแปรปรวน (variance components) โดยรูปแบบที่ 2 จะใช้การประมาณด้วยการวิเคราะห์ความแปรปรวน (ANOVA) เหมือนกับการวิเคราะห์ในทฤษฎี GT ส่วนรูปแบบที่ 3 จะประมาณค่าด้วยสูตรของ Kolen และ Harris (1987) เช่นเดียวกับการประมาณค่าในรูปแบบที่ 1 และรูปแบบที่ 4 Numerical Bayesian GIRM เป็นรูปแบบที่ผู้วิจัยประยุกต์ใช้การประมาณค่าแบบเบส์ในการประมาณค่าพารามิเตอร์ของ IRTที่ Swaminathan และ Gifford

(1982, 1985, 1986) พัฒนาขึ้น ซึ่งเป็นการประมาณค่าที่ไม่ได้ใช้เทคนิค MCMC (Markov Chain Monte Carlo) ซึ่งน่าจะช่วยแก้ไขปัญหาการประมาณค่าพารามิเตอร์ที่ไม่เป็นอิสระต่อกันระหว่างแถวและคอลัมน์ได้ โดยในการศึกษาครั้งนี้จะทำการศึกษาประสิทธิภาพของการประมาณค่าใน 3 มิติ ได้แก่ มิติที่1 ความลำเอียงในการประมาณค่า (Biased estimator) คำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) มิติที่ 2 ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) และมิติที่ 3 ประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC)

การศึกษาการประมาณค่าพารามิเตอร์ในครั้งนี้ศึกษาวิธีการประมาณค่าพารามิเตอร์แบบ 1 พารามิเตอร์ของโมเดลการตอบสนองข้อสอบ (1 PL IRT) โดยมีแหล่งความคลาดเคลื่อนที่สนใจศึกษาเพียง 1 แหล่ง (Single facet measurement) ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เนื่องจากการพัฒนาวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) มีการพัฒนาในเบื้องต้นเพียงโมเดลเดียวในปี 2007 โดยโมเดลดังกล่าวยังมีปัญหาความไม่แน่นอนในการประมาณค่า (Uncertainty) เกิดขึ้น ดังนั้นในการศึกษาครั้งนี้จึงศึกษาโมเดลดังกล่าวก่อนเพื่อให้วิธีการ GIRM มีความสมบูรณ์ในโมเดลที่ง่ายต่อการวิเคราะห์ก่อนจึงจะขยายผลการศึกษาไปในรูปแบบการวิเคราะห์ที่ซับซ้อนต่อไป

สำหรับข้อมูลที่ใช้ในการศึกษาครั้งนี้ใช้ข้อมูลจำลองที่ได้จากโปรแกรม R และทำการประมวลผลภายใต้จากการเขียนคำสั่งการประมวลผลด้วยโปรแกรม WinBUGS และกลับมาแสดงผลซึ่งจะแสดงผลในโปรแกรม R โดยใช้โปรแกรมร่วมกันระหว่างโปรแกรม R และโปรแกรม WinBUGS ด้วย Package R2 WinBUGS เนื่องจากการจำลองข้อมูลสามารถศึกษาโอกาสที่จะเกิดขึ้นของข้อมูลได้หลากหลายรูปแบบที่น่าจะมีโอกาสเกิดขึ้นจริง การจำลองข้อมูลจึงเป็นวิธีการที่เป็นประโยชน์สำหรับการศึกษาหรือการพัฒนารูปแบบการประมาณค่าวิธีใหม่ให้มีความชัดเจน โดยงื่อนไขในการจำลองข้อมูลครั้งนี้ ใช้จำนวนกลุ่มตัวอย่างและจำนวนข้อสอบเป็นตัวแปรสำหรับเงื่อนไขในการจำลองข้อมูล เนื่องจากในการศึกษาเอกสารและงานวิจัยที่เกี่ยวกับการวิเคราะห์ข้อสอบ ระบุว่าจำนวนกลุ่มตัวอย่างและจำนวนข้อสอบในการวิเคราะห์ข้อมูลส่งผลต่อการประมาณค่าพารามิเตอร์ในโมเดลการวิเคราะห์ข้อมูลด้วยโมเดลการตอบสนองของข้อสอบ (Swaminathan และ Gifford, 1982; 1985; 1986; Hulin, 1983; Lord, 1986; รัตนา ศรีเหรัญ, 2539; วิชุดา บัวคง, 2532; เกศมณี พยัคฆ์, 2542) พบว่า ถ้าจำนวนกลุ่มตัวอย่างและจำนวนข้อสอบเพิ่มขึ้น จะทำให้มีการประมาณ

ค่าพารามิเตอร์มีความถูกต้องเพิ่มมากขึ้น โดย Rentz และ Bashaw (1975 อ้างถึงใน สุพล นิลกลาง, 2541) กล่าวว่า ผู้เข้าสอบจำนวน 500-1,000 คน จะทำให้การประมาณค่าพารามิเตอร์ของรูปแบบราส์ชมีค่าคงที่เพิ่มขึ้นอย่างรวดเร็ว แต่ถ้าผู้เข้าสอบเพิ่มขึ้นเป็นจำนวน 2,000-4,000 คน การประมาณค่าพารามิเตอร์จะคงที่เพิ่มขึ้นอย่างช้า ดังนั้นผู้วิจัยจึงเลือกเงื่อนไขในการจำลองข้อมูลในครั้งนี้ โดยพิจารณาจาก จำนวนกลุ่มตัวอย่างและจำนวนข้อสอบเป็นเงื่อนไขในการจำลองข้อมูล

ในเรื่องของจำนวนกลุ่มตัวอย่าง Swaminathan และ Gifford (1982, 1985,1986) เสนอแนะว่าควรใช้กลุ่มตัวอย่างในการประมาณ 100 คน ขึ้นไป ดังนั้นในการจำลองข้อมูลขนาดกลุ่มตัวอย่างในการวิเคราะห์ข้อมูล ผู้วิจัยจึงใช้ขนาดกลุ่มตัวอย่างจำนวน 2 เงื่อนไข ได้แก่ 100 คนและ 300 คน สำหรับในเรื่องของจำนวนข้อสอบ Briggs และ Wilson (2007) ได้ทำการจำลองข้อมูลข้อสอบ จำนวน 5 ข้อ สำหรับทดสอบความแกร่งของวิธีการ GIRM เนื่องจากข้อจำกัดของโปรแกรม WinBUGS ที่ใช้หน่วยความจำของคอมพิวเตอร์ในการวิเคราะห์ข้อมูลมาก ดังนั้นในการจำลองข้อมูลขนาดกลุ่มตัวอย่างในการวิเคราะห์ข้อมูล ผู้วิจัยจึงใช้จำนวนข้อสอบรวม 2 เงื่อนไข ได้แก่ 5 ข้อ และ 10 ข้อ

ส่วนเงื่อนไขในการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) จากการศึกษาครั้งนี้ผู้วิจัยได้ศึกษาวิธีการประมาณค่าพารามิเตอร์แบบ 1 พารามิเตอร์ของโมเดลการตอบสนองข้อสอบ (1 PL IRT) โดยมีแหล่งความคลาดเคลื่อนจากการวัด 1 แหล่ง (Single facet measurement) ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) และใช้วิธีการในการประมาณพารามิเตอร์หลักการประมาณค่าแบบเบส์ (Bayesian Estimation) โดยสามารถสรุปเป็นโมเดลการประมาณค่าพารามิเตอร์แบบเบส์ได้ดังนี้

จากโมเดลการประมาณค่าข้างต้นดังกล่าว จะเห็นได้ว่าการวิเคราะห์ประกอบไปด้วย 3 ส่วน ส่วนแรกเกี่ยวกับ เมทริกซ์ของคะแนนที่สังเกตได้ $L(X \mid \theta, \beta)$ ซึ่งประมาณได้จากวิธีการประมาณค่าแม็กซิมั่มไลค์ลิฮูด (Maximum Likelihood Estimation) ส่วนที่สอง เกี่ยวกับลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และส่วนที่สาม เกี่ยวกับลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) ในการศึกษาครั้งนี้ผู้วิจัยมีจุดประสงค์หลักในการพัฒนาวิธีการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ดังนั้นเงื่อนไขของการวิเคราะห์ข้อมูลในการศึกษาครั้งนี้ จึงประกอบไปด้วยตัวแปร จำนวน 3 ตัวแปร ได้แก่ รูปแบบการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β)

การศึกษาความแกร่ง (Robustness) ของผลการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทั้ง 3 รูปแบบ ผู้วิจัยจึงได้ศึกษาความไว (Sensitivity) ของผลการวิเคราะห์ที่มีต่อลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) โดยลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) ผู้วิจัยได้ศึกษา จำนวน 2 ลักษณะ ได้แก่ ลักษณะแรก เป็นลักษณะการแจกแจงเริ่มแรกแบบปกติ (Normal) ซึ่งเป็นลักษณะการแจกแจงตามสมมติฐานของวิธีการ GIRM และลักษณะการแจกแจงแบบแกมมา (Gamma) เนื่องจากการกำหนดลักษณะการแจกแจงแบบแกมมา (Gamma) จะกำหนดเมื่อค่าส่วนใหญ่มีค่าเป็นโค้งและโอกาสในการเกิดค่าพารามิเตอร์มีลักษณะเป็นเส้นโค้ง (Gelman, Carlin, Stern, และ Rubin, 1995) สำหรับลักษณะการแจกแจงเริ่มแรกค่าพารามิเตอร์ของข้อสอบ (β) ผู้วิจัยได้ศึกษา จำนวน 2 ลักษณะ ได้แก่ ลักษณะแรก เป็นลักษณะการแจกแจงเริ่มแรกแบบปกติ (Normal) ซึ่งเป็นลักษณะการแจกแจงตามสมมติฐานของวิธีการ GIRM ลักษณะที่สอง เป็นการแจกแจง แบบเบต้า (Bata) เนื่องจากการกำหนดลักษณะการแจกแจงแบบเบต้า (Beta) กำหนด เมื่อค่าพารามิเตอร์ที่ต้องการประมาณค่ามีค่าอยู่ระหว่าง 0 ถึง 1 (Gelman และคณะ, 1995)

จากผลการศึกษาในครั้งนี้จะทำให้ได้สารสนเทศเกี่ยวกับวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบให้มีประสิทธิภาพ ซึ่งทำให้สามารถวิเคราะห์พารามิเตอร์ของผู้สอบและข้อสอบมีความแม่นยำมากยิ่งขึ้น นอกจากนี้ยังทราบถึงความแกร่งของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบเกี่ยวกับการการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบและการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ เพื่อทำให้ผู้ที่ต้องการวิเคราะห์มีการตรวจสอบข้อมูลว่ามีความเหมาะสมที่จะนำข้อมูลมาวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบหรือไม่ เพื่อให้สามารถประมาณค่าพารามิเตอร์ให้มีความถูกต้องและแม่นยำ โดยไม่ทำให้ผลการวิเคราะห์เกิดความคลาดเคลื่อน รวมทั้งยังทำให้ทราบถึงขนาดกลุ่มตัวอย่างจำนวนข้อสอบที่เหมาะสมสำหรับการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบให้เกิดประสิทธิภาพสูงสุด

คำถามการวิจัย

1. ประสิทธิภาพของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ระหว่างรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 Alternative GIRM A รูปแบบที่ 3 Alternative GIRM B และรูปแบบที่ 4 Numerical Bayesian GIRM รูปแบบใดมีประสิทธิภาพในการประมาณค่าพารามิเตอร์มากที่สุด
2. ขนาดของกลุ่มตัวอย่างและจำนวนข้อสอบที่แตกต่างกันจะส่งผลต่อประสิทธิภาพการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ในรูปแบบการประมาณค่าพารามิเตอร์แบบ Original GIRMX Alternative GIRM A, Alternative GIRM B และ Numerical Bayesian GIRM อย่างไร
3. การกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ (β) จะส่งผลต่อประสิทธิภาพการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ในรูปแบบการประมาณค่าพารามิเตอร์แบบ Original GIRM, Alternative GIRM A, Alternative GIRM B และ Numerical Bayesian GIRM อย่างไร

วัตถุประสงค์การวิจัย

1. เพื่อเปรียบเทียบประสิทธิภาพของวิธีวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) 4 รูปแบบ ระหว่างรูปแบบ Original GIRM, รูปแบบ Alternative GIRM A, รูปแบบ Alternative GIRM B และรูปแบบ Numerical Bayesian GIRM

2. เพื่อศึกษาอิทธิพลของขนาดกลุ่มตัวอย่างและจำนวนข้อสอบที่มีต่อประสิทธิภาพของวิธีวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) 4 รูปแบบ ซึ่งได้แก่ รูปแบบ Original GIRM, รูปแบบ Alternative GIRM A, รูปแบบ Alternative GIRM B และรูปแบบ Numerical Bayesian GIRM

3. เพื่อศึกษาความไว (Sensitivity) ของผลการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) 4 รูปแบบ ซึ่งได้แก่ รูปแบบ Original GIRM, รูปแบบAlternative GIRM A, รูปแบบ Alternative GIRM B และรูปแบบ Numerical Bayesian GIRM ที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ)

สมมติฐานการวิจัย

วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ(GIRM) เป็นวิธีการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด(GT) เข้าไปในโมเดลการกำหนดมาตรวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ(IRT) โดยการประมาณค่าความแปรปรวน (Variance Components) ในแต่ละองค์ประกอบของการวัดไปพร้อมกับการประมาณค่าพารามิเตอร์ในโมเดลการตอบสนองข้อสอบ (IRT) ดังนั้นการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทำให้ได้สารเทศของการวัดทั้งในเชิงของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และทฤษฎีการตอบสนองข้อสอบ (IRT) (Briggs และ Wilson, 2007)

ขั้นตอนการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) สามารถแบ่งการประมาณค่าออกได้ 3 ขั้นตอนใหญ่ๆ ขั้นแรกเป็นการประมาณค่าพารามิเตอร์จากทฤษฎีการตอบสนองข้อสอบ ขั้นตอนที่สองเป็นขั้นตอนของการเตรียมข้อมูลสำหรับการประมาณค่าในขั้นตอนสาม โดยการนำค่าพารามิเตอร์ที่ได้จากขั้นตอนแรกมาวิเคราะห์และสร้างเป็นเมทริกซ์ของข้อมูลเพื่อเตรียมการวิเคราะห์สำหรับในขั้นตอนที่สาม และขั้นตอนสุดท้ายเป็นการประมาณค่าพารามิเตอร์ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

รูปแบบการประมาณค่าของวิธีการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ในการศึกษาครั้งนี้มี 4 รูปแบบ ได้แก่ รูปแบบที่ 1 Original GIRM เป็นรูปแบบการวิเคราะห์แรกเริ่มที่ Briggs และ Wilson (2007) พัฒนาขึ้น รูปแบบที่ 2 Alternative GIRM A เป็นรูปแบบที่ผู้วิจัยมีการปรับเปลี่ยนวิธีการประมาณค่าในวิธีการ GIRM โดยมีการคำนวณค่า Posterior predictive distribution ในการสร้างเป็น Predictive Observed Matrix ก่อนนำมาคำนวณสารสนเทศของ GT ด้วยการวิเคราะห์ความแปรปรวนแบบ ANOVA รูปแบบที่ 3 Alternative GIRM B เป็นรูปแบบที่คล้ายคลึงกับรูปแบบที่ 2 แต่ในรูปแบบนี้มีการวิเคราะห์ความแปรปรวนด้วยสูตร Kolen และ Harris Link (1987) และรูปแบบที่ 4 Numerical Bayesian GIRM เป็นรูปแบบที่ผู้วิจัยนำวิธีการประมาณค่าพารามิเตอร์ของ IRT ที่ Swaminathan และ Gifford (1982, 1985, 1986) พัฒนาขึ้นมาใช้ในการคำนวณสารสนเทศของ IRT ในขั้นตอนที่ 1 ของวิธีการ GIRM

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องที่เกี่ยวกับทฤษฎีการวิเคราะห์ข้อสอบ การประมาณค่าพารามิเตอร์ และการจำลองข้อมูล ทำให้สามารถตั้งสมมติฐานโดยแยกประเด็นตามวัตถุประสงค์ในการวิจัยได้ดังนี้

1. รูปแบบการวิเคราะห์ข้อสอบของวิธีการสรุปอ้างอิงสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ(GIRM)รูปแบบที่ 2,3 Alternative GIRM A,B น่าจะมีประสิทธิภาพในการประมาณค่ามากกว่าวิธีการวิเคราะห์ในรูปแบบที่ 1 Original GIRM และรูปแบบที่ 4 Numerical Bayesian GIRM น่าจะทำการประมาณค่าพารามิเตอร์ได้มีประสิทธิภาพมากกว่ารูปแบบที่ 1 Original GIRM

ทั้งนี้เนื่องจาก รูปแบบที่ 1 Original GIRM เป็นรูปแบบที่มีความไม่แน่นอนในการประมาณค่า ทั้งนี้เป็นเพราะในการประมาณค่าของวิธีนี้เริ่มต้นจากการวิเคราะห์เพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ(IRT) โดยใช้เทคนิค MCMC (Markov Chain Monte Carlo) โดยการคำนวนแบบ Metropolis-Hastings algorithm ภายในวิธีการ Gibbs sampler ด้วยโปรแกรม WinBUGS ซึ่งจะสามารถประมาณค่าการแจกแจงภายหลัง (Posterior distribution) ของค่าพารามิเตอร์ได้ หลังจากนั้นจึงเอาค่าพารามิเตอร์ที่ได้มาคำนวนค่า E(X_{pi}) หรือก็คือค่า $P(\theta_p, \beta_i)$ ด้วยโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ และนำข้อมูลทั้งหมดมาสร้างเป็นเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) จากนั้นจึงทำการวิเคราะห์ขั้นตอนที่สองต่อเพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยการนำข้อมูลจาก เมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) มาเป็นข้อมูลในการวิเคราะห์ด้วยสูตรการคำนวณของ The Kolen และ Harris Link (1987) ซึ่งในขั้นตอนนี้การนำค่าที่ได้ซึ่งมีความไม่เป็นอิสระต่อกันจากการประมาณค่าด้วยวิธี MCMC โดยการประมาณค่าพารามิเตอร์แต่ละค่าระหว่างแถวและคอลัมน์จะสัมพันธ์กันและการนำค่าที่สัมพันธ์กันมาใช้ประมาณต่อในขั้นตอนที่ 3 จึงทำให้ค่าประมาณที่ได้ฝ่าฝืนสมมติฐานของการสุ่ม (Briggs และ Wilson, 2007) ส่วนในรูปแบบที่ 2,3 Alternative GIRM A,B จะใช้การวิเคราะห์เป็นคะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) ซึ่งเป็นคะแนนเฉลี่ยของคะแนนที่สังเกตได้ที่ได้มาจากการทำนาย นำมาสร้างเป็นเมทริกซ์คะแนนสังเกตได้จากการทำนายก่อนนำไปวิเคราะห์ในขั้นตอนที่ 3 ทำให้ข้อมูลที่ได้จากการประมาณค่าขั้นตอนที่ 1 มีลักษณะคล้ายคลึงกับการวิเคราะห์จากคะแนนจริงของการวิเคราะห์ GT ซึ่งเป็นคะแนนที่สังเกตได้ จึงทำให้ไม่ฝ่าฝืนสมมติฐานแบบสุ่มของการวิเคราะห์

ส่วนวิธีการ GIRM ในรูปแบบที่ 4 Numerical Bayesian GIRM น่าจะทำการประมาณค่าพารามิเตอร์ได้มีประสิทธิภาพมากกว่ารูปแบบที่ 1 Original GIRM ทั้งนี้เนื่องจากในรูปแบบที่ 3 Alternative Bayesian GIRM เป็นรูปแบบที่มีวิธีการวิเคราะห์ที่คล้ายคลึงกับการวิเคราะห์ในรูปแบบที่ 1 Original GIRM โดยต่างกันในขั้นตอนที่ 1 การวิเคราะห์ในรูปแบบที่ 3 Alternative GIRM B ใช้การวิเคราะห์ด้วยวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) โดยใช้การหาอนุพันธ์ของ

ฟังก์ชันให้มีค่าเท่ากับศูนย์แล้วจึงหารากของอนุพันธ์นั้น ในขณะที่การวิเคราะห์ในรูปแบบที่ 1 Original GIRM ใช้การวิเคราะห์ด้วยวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation)เช่นกัน แต่วิธีการจำลองข้อมูลแบบห่วงโซ่ (Markov Chain Monte Carlo; MCMC) แทน ซึ่งจากการประมาณค่าที่ไม่ใช่การจำลองข้อมูลแบบห่วงโซ่น่าจะช่วยลดในเรื่องการฝ่าฝืนสมมติฐานของการสุ่มได้ จึงน่าจะทำให้การประมาณค่าพารามิเตอร์จึงไม่น่าจะเกิดความลำเอียงในการประมาณค่า

2. ประสิทธิภาพของการวิเคราะห์ข้อสอบด้วยวิธี GIRM น่าจะให้ประสิทธิภาพเพิ่มมากขึ้นทั้งสามรูปแบบ ถ้ามีจำนวนกลุ่มตัวอย่างและจำนวนข้อสอบที่เพิ่มขึ้น ทั้งนี้เนื่องจากเมื่อพิจารณาเกี่ยวกับเงื่อนไขของการจำลองข้อมูลเพื่อนำมาวิเคราะห์นั้น พบว่า จำนวนกลุ่มตัวอย่างและจำนวนข้อสอบในการวิเคราะห์ข้อมูลส่งผลต่อการประมาณค่าพารามิเตอร์ในโมเดลการวิเคราะห์ข้อมูลด้วยโมเดลการตอบสนองของข้อสอบ (Swaminathan และ Gifford, 1982, 1985, 1986; Hulin, 1983; Lord,1986; รัตนา ศรีเหรัญ, 2539; วิชุดา บัวคง, 2532; เกศมณี พยัคฆ์, 2542) พบว่า ถ้าจำนวนกลุ่มตัวอย่างและจำนวนข้อสอบเพิ่มขึ้นน่าจะทำให้มีการประมาณค่าพารามิเตอร์ที่ถูกต้องเพิ่มมากขึ้น โดย Rentz และ Bashaw (1975 อ้างถึงในสุพล นิลกลาง, 2541) กล่าวว่า ผู้เข้าสอบจำนวน 500-1,000 คนจะทำให้การประมาณค่าพารามิเตอร์ของรูปแบบราส์ชมีค่าคงที่ที่เพิ่มขึ้นอย่างรวดเร็ว แต่ถ้าผู้เข้าสอบเพิ่มขึ้นเป็นจำนวน 2,000-4,000 คนจะเพิ่มขึ้นอย่างช้า ดังนั้นผู้วิจัยจึงเลือกเงื่อนไขในการจำลองข้อมูลในครั้งนี้ โดยพิจารณาจาก จำนวนกลุ่มตัวอย่างและจำนวนข้อสอบเป็นเงื่อนไขในการจำลองข้อมูล

ในเรื่องของจำนวนกลุ่มตัวอย่าง Swaminathan และ Gifford (1982,1985,1986) เสนอแนะว่า ควรใช้กลุ่มตัวอย่างในการประมาณ 100 คนขึ้นไป ส่วน Hulin (1983) และ Lord (1986) เสนอแนะว่าในการประมาณค่าพารามิเตอร์ควรใช้ขนาดกลุ่มตัวอย่างตั้งแต่ 1,000 คนขึ้นไป ดังนั้นในการจำลองข้อมูลขนาดกลุ่มตัวอย่างในการวิเคราะห์ข้อมูล ผู้วิจัยจึงใช้ขนาดกลุ่มตัวอย่างจำนวน 2 เงื่อนไข ได้แก่ 100 คน และ 300 คน

สำหรับในเรื่องของจำนวนข้อสอบ Swaminathan และ Gifford (1982, 1985, 1986) เสนอแนะว่าควรใช้จำนวนข้อสอบสำหรับในการประมาณค่าอย่างน้อย จำนวน 25 ข้อ ส่วน Hulin (1983) เสนอแนะว่า ในการวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ต้องมีแบบสอบอย่างน้อย 30 ข้อ เนื่องจากการวิเคราะห์ด้วยขนาดกลุ่มตัวอย่างดังกล่าวจะทำให้ค่าเฉลี่ยของความแตกต่างระหว่างโค้งคุณลักษณะของข้อสอบของค่าพารามิเตอร์เท่ากับค่าประมาณ RMSE (Root Mean Squared Error) มีค่าต่ำ($RMSE \le 0.05$) สำหรับ Briggs และ Wilson (2007) ได้ทำการจำลองข้อมูลข้อสอบจำนวน 5 ข้อสำหรับทดสอบความแกร่งของการวิเคราะห์ข้อสอบด้วยวิธีการ GIRM เนื่อง

ด้วยข้อจำกัดของโปรแกรม WinBUGS ต้องใช้หน่วยความจำของคอมพิวเตอร์ในการวิเคราะห์มาก แต่ด้วยคุณสมบัติของเครื่องคอมพิวเตอร์ในปัจจุบัน ดังนั้นในการจำลองข้อมูลขนาดกลุ่มตัวอย่างในการวิเคราะห์ข้อมูล ผู้วิจัยจึงใช้จำนวนข้อสอบรวม 2 เงื่อนไข ได้แก่ 5 ข้อ และ 10 ข้อ

จากการศึกษาในเรื่องของจำนวนกลุ่มตัวอย่างและจำนวนข้อสอบที่แตกต่างสำหรับการประมาณค่าพารามิเตอร์ของข้อสอบ พบว่า ถ้าจำนวนกลุ่มตัวอย่างยิ่งมากและจำนวนข้อสอบยิ่งมาก จะทำให้มีการประมาณค่าพารามิเตอร์ที่ถูกต้องเพิ่มมากขึ้น (Swaminathan และ Gifford, 1982, 1985, 1986; Hulin, 1983; Lord, 1986; รัตนา ศรีเหรัญ, 2539; วิชุดา บัวคง,2532; เกศมณี พยัคฆ์, 2542) ดังนั้น จึงตั้งสมมติฐานเกี่ยวกับขนาดกลุ่มตัวอย่างและจำนวนข้อสอบว่า ถ้าจำนวนกลุ่มตัวอย่างที่เพิ่มขึ้นและจำนวนข้อสอบที่เพิ่มขึ้นน่าจะทำให้ประสิทธิภาพของการวิเคราะห์ข้อสอบด้วยวิธีการ GIRM สูงขึ้น

3. การวิเคราะห์ข้อสอบของวิธี GIRM ทั้งสี่รูปแบบ น่าจะมีความไวต่อลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบแกมมา (Gamma) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบเบต้า (Beta) เมื่อเทียบกับผลการวิเคราะห์ที่ได้จากลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบและข้อสอบแบบปกติ (Normal)

ทั้งนี้เนื่องจากในการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ใช้หลักการประมาณค่าแบบเบส์ (Bayesian Estimation) ซึ่งสามารถสรุปเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้

$$P(X,\theta\ ,\beta\ ) = L(X\ \ |\theta\ ,\beta\ \ )f(\theta)g(\beta)$$

จากโมเดลการประมาณค่าดังกล่าว จะเห็นได้ว่าประกอบไปด้วย 3 ส่วน คือ เมทริกซ์ของคะแนนที่สังเกตได้ $[L(X|\theta,\beta)$ ซึ่งประมาณได้จากวิธีการประมาณค่าแม็กซิมั่มไลค์ลิฮู้ด (Maximum Likelihood Estimation) ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ $(\beta\ )$ ในการศึกษาครั้งนี้ผู้วิจัยมีจุดประสงค์หลักในการพัฒนาวิธีการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ดังนั้นในเงื่อนไขของการวิเคราะห์ข้อมูลเกี่ยวกับลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ $(\beta\ )$ จึงเป็นเงื่อนไขที่สำคัญสำหรับในการศึกษาความไวของโมเดลต่อการฝ่าฝืนข้อตกลงเบื้องต้นของโมเดลที่ระบุไว้ว่า โมเดลต้องมีลักษณะการเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และของข้อสอบ $(\beta\ )$ ลักษณะการแจกแจงแบบปกติ (Normal)

ในการศึกษาครั้งนี้ ผู้วิจัยได้ศึกษาลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) ผู้วิจัยได้ศึกษา จำนวน 2 ลักษณะ ลักษณะแรกเป็นการแจกแจงแบบปกติ (Normal) ซึ่งเป็นลักษณะการแจกแจงตามสมมติฐานของวิธีการ GIRM และลักษณะการแจกแจงแบบแกมมา (Gamma) โดยการกำหนดลักษณะการแจกแจงแบบแกมมา (Gamma) จะกำหนดเมื่อค่าส่วนใหญ่มีค่าเป็นโค้งและโอกาสในการเกิดค่าพารามิเตอร์จะมีลักษณะเป็นเส้นโค้ง (Gelman และคณะ, 1995)

ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ $(\beta\ )$ ผู้วิจัยได้ศึกษา จำนวน 2 ลักษณะ ลักษณะแรกเป็นการแจกแจงแบบปกติ (Normal) ซึ่งเป็นลักษณะการแจกแจงตามสมมติฐานของวิธีการ GIRM และลักษณะการแจกแจงแบบเบต้า (Beta) กำหนดเมื่อค่าพารามิเตอร์ที่ต้องการประมาณค่ามีค่าอยู่ระหว่าง 0 ถึง 1 (Gelman และคณะ, 1995)

จากข้อตกลงเบื้องต้นของการวิเคราะห์ข้อมูลด้วยวิธี GIRM ได้ระบุไว้ว่า โมเดลต้องเป็นลักษณะแบบสุ่ม ดังนั้นลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบและค่าพารามิเตอร์ของข้อสอบต้องมีลักษณะการแจกแจงแบบปกติ(Normal) แต่จากการจำลองข้อมูลจากข้อสอบ 5 ข้อ ผู้เข้าสอบ 500 คน ของ Briggs และ Wilson (2007) เพื่อทดสอบความแกร่งของวิธีการวิเคราะห์ พบว่า การวิเคราะห์ข้อสอบด้วยวิธีการ GIRM มีความคงทนต่อลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ทั้งผู้สอบและข้อสอบแบบเอกรูป (Uniform) และมีความไว (Sensitivity) ของผลการวิเคราะห์ต่อลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบแกมมา (Gamma) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบเบต้า (Beta) ดังนั้นในการศึกษาครั้งนี้จึงตั้งสมมติฐานไว้ว่า โมเดลการวิเคราะห์ข้อสอบของวิธี GIRM ทั้งสี่รูปแบบน่าจะมีความไวต่อลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบแกมมา (Gamma) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบเบต้า (Beta) เมื่อเทียบกับผลการวิเคราะห์ที่ได้จากลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบและข้อสอบแบบปกติ (Normal)

ขอบเขตของการวิจัย

1. การประมาณค่าพารามิเตอร์ในครั้งนี้ ศึกษาวิธีการประมาณค่าพารามิเตอร์แบบ 1 พารามิเตอร์ของโมเดลการตอบสนองข้อสอบ (1PL IRT) โดยมีแหล่งความคลาดเคลื่อนที่สนใจศึกษา 1 แหล่ง (single facet measurement) ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เนื่องจากการพัฒนาวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) มีการพัฒนาในเบื้องต้นเพียงโมเดลเดียวในปี 2007 โดยโมเดลดังกล่าวยังมีปัญหาความไม่แน่นอนในการประมาณค่า (Uncertainty) เกิดขึ้น ดังนั้นในการศึกษาครั้งนี้ จึงศึกษาโมเดลดังกล่าวก่อนเพื่อให้วิธีการ GIRM มีความสมบูรณ์ในโมเดลที่ง่ายต่อการวิเคราะห์ก่อน จึงจะขยายผลการศึกษาไปในรูปแบบการวิเคราะห์ที่ซับซ้อนต่อไป

2. ข้อมูลที่ใช้ในการศึกษาครั้งนี้ ใช้ข้อมูลจำลองที่ได้จากโปรแกรม R และทำการประมวลผลภายใต้จากการเขียนคำสั่งการประมวลผลด้วยโปรแกรม WinBUGS และกลับมาแสดงผลจะแสดงผลในโปรแกรม R โดยใช้โปรแกรมร่วมกันระหว่างโปรแกรม R และโปรแกรม WinBUGS ด้วย Package R2 WinBUGS เนื่องจากการจำลองข้อมูลสามารถศึกษาโอกาสที่จะเกิดขึ้นของข้อมูลได้หลากหลายรูปแบบที่น่าจะมีโอกาสเกิดขึ้นจริง ซึ่งการจำลองข้อมูลจึงเป็นวิธีการที่เป็นประโยชน์ สำหรับการศึกษาหรือการพัฒนารูปแบบการประมาณค่าวิธีใหม่ให้มีความชัดเจน โดยมีเงื่อนไขในการจำลองข้อมูลครั้งนี้ การทำกลุ่มข้อมูลซ้ำ 500 ชุดข้อมูล (Data Set) โดยการประมาณค่าซ้ำในแต่ละชุดข้อมูล 10,000 รอบ (number of iteration) หลังจากตัดข้อมูลในส่วนแรกออก 1,000 รอบ (number of burnin)

3. รูปแบบการประมาณค่าวิธีการวิเคราะห์ข้อสอบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ในการศึกษาครั้งนี้มี 4 รูปแบบ ได้แก่

3.1 รูปแบบที่ 1 Original GIRM ซึ่งเป็นรูปแบบการประมาณค่าแบบแรกของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ลักษณะที่ Briggs และ Wilson (2007) พัฒนาขึ้น

3.2 รูปแบบที่ 2 Alternative GIRM A เป็นรูปแบบการประมาณค่าที่ผู้วิจัยปรับวิธีการประมาณค่าพารามิเตอร์ขึ้นมาใหม่จากรูปแบบต้นฉบับที่ Briggs และ Wilson (2007) พัฒนาขึ้น โดยการนำค่าลักษณะการแจกแจงภายหลัง (Posterior distribution) ที่ได้ ซึ่งได้แก่ค่า (θ_p) และ (β_i) มาจำลองเมทริกซ์ค่าสังเกตใหม่ด้วยการคำนวณหาค่าการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) ซึ่งเป็นค่าเฉลี่ยของคะแนนที่สังเกตได้ (Observed score) แทนการนำค่าลักษณะการแจกแจงภายหลัง (Posterior distribution) ที่ได้ (θ_p) และ (β_i) มาคำนวณค่า E (X_{pi})

หรือ $P(\theta_p, \beta_i)$ ด้วยโมเดลการตอบสนองข้อสอบแบบ 1 PL ดังนั้นจะเห็นที่มาข้อมูลก่อนที่นำมาหาค่าองค์ประกอบความแปรปรวนของวิธีการมีลักษณะคล้ายกับข้อมูลจริง (observed data) ซึ่งมีลักษณะเป็นข้อมูล 0, 1 เหมือนกับการวิเคราะห์องค์ประกอบความแปรปรวนในทฤษฎี GT ซึ่งข้อมูลลักษณะนี้จะเป็นอิสระต่อกันระหว่างแถวและคอลัมน์ใน Predictive Observed Matrix หลังจากนั้นจึงประมาณค่าองค์ประกอบความแปรปรวน (variance components) ด้วยการวิเคราะห์ความแปรปรวน (ANOVA)

3.3 รูปแบบที่ 3 Alternative GIRM B รูปแบบนี้มีลักษณะคล้ายคลึงกับวิธีการประมาณค่าในรูปแบบที่ 2 Alternative GIRM A แต่มีความต่างกันเฉพาะในขั้นตอนการประมาณค่าองค์ประกอบความแปรปรวน (variance components) โดยรูปแบบที่ 2 จะใช้การประมาณค่าด้วยการวิเคราะห์ความแปรปรวน (ANOVA) เหมือนกับการวิเคราะห์ในทฤษฎี GT ส่วนรูปแบบที่ 3 จะประมาณค่าด้วยสูตรของ Kolen และ Harris (1987) เช่นเดียวกับการประมาณค่าในรูปแบบที่ 1 ทั้งนี้เพื่อศึกษาว่ารูปแบบการวิเคราะห์องค์ประกอบความแปรปรวนในรูปแบบใดระหว่างการวิเคราะห์ความแปรปรวน (ANOVA) กับการวิเคราะห์ด้วยสูตรของ Kolen และ Harris (1987) จะให้ผลการวิเคราะห์องค์ประกอบความแปรปรวนที่มีประสิทธิภาพมากกว่า

3.4 รูปแบบที่ 4 Numerical Bayesian GIRM เป็นรูปแบบการประมาณที่ผู้วิจัยประยุกต์โดยใช้การประมาณค่าแบบเบส์ในการประมาณค่าพารามิเตอร์ของ IRT ที่ Swaminathan และ Gifford (1982, 1985, 1986) พัฒนาขึ้นมาใช้ในการประมาณค่าพารามิเตอร์ของผู้สอบ และค่าพารามิเตอร์ของข้อสอบก่อนนำมาสร้างเป็นเมทริกซ์ที่คาดหวัง (Expected Response Matrix) และทำการวิเคราะห์ค่าองค์ประกอบความแปรปรวนเหมือนรูปแบบที่ 1 โดยใช้การวิเคราะห์องค์ประกอบความแปรปรวนจากสูตรของ Kolen และ Harris (1987)

4. การจำลองข้อมูลในครั้งนี้ผู้วิจัยได้ใช้เงื่อนไขในการจำลองข้อมูลทั้งสิ้น 4 เงื่อนไขในการศึกษา ซึ่งได้แก่

4.1 จำนวนกลุ่มตัวอย่าง 2 เงื่อนไข ประกอบด้วย

1) กลุ่มตัวอย่างจำนวน 100 คน

2) กลุ่มตัวอย่างจำนวน 300 คน

4.2 จำนวนข้อสอบ 2 เงื่อนไขประกอบด้วย

1) ข้อสอบจำนวน 5 ข้อ

2) ข้อสอบจำนวน 10 ข้อ

5. การประมาณค่าครั้งนี้ประมาณด้วยวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) โดยการประมาณค่าต้องมีการระบุลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ ในการศึกษาครั้งนี้ใช้โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ ดังนั้นจึงต้องมีการระบุเงื่อนไขของลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) โดยมีเงื่อนไขของลักษณะการแจกแจงเริ่มแรกทั้งหมด 4 เงื่อนไข ดังนี้

5.1 ลักษณะการแจกแจงเริ่มแรกของพารามิเตอร์ของผู้สอบ 2 เงื่อนไข ประกอบด้วย

1) ลักษณะการแจกแจงแบบปกติ (Normal distribution)

2) ลักษณะการแจกแจงแบบแกมมา (Gamma distribution)

5.2 ลักษณะการแจกแจงเริ่มแรกของพารามิเตอร์ของข้อสอบ 2 รูปแบบ ประกอบด้วย

1) ลักษณะการแจกแจงแบบปกติ (Normal distribution)

2) ลักษณะการแจกแจงแบบเบต้า (Beta distribution)

6. การวัดประสิทธิภาพของการประมาณค่า ทำการวัดประสิทธิภาพตัวประมาณค่า 3 ดัชนี ได้แก่

6.1 ความลำเอียงในการประมาณค่า (Biased estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD)

6.2 ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.)

6.3 ประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC)

นิยามศัพท์ที่ใช้ในการวิจัย

1. **การประมาณค่าพารามิเตอร์ด้วยวิธีของเบส์** หมายถึง การคำนวณหาค่าพารามิเตอร์ของผู้สอบและค่าพารามิเตอร์ของข้อสอบ โดยใช้วิธีการประมาณค่าของเบส์ (Bayesian Estimation) ซึ่งประกอบด้วย 3 ส่วน ได้แก่ เมทริกซ์ของคะแนนที่สังเกตได้ $L(X \mid \theta, \beta)$ ซึ่งประมาณได้จากวิธีการประมาณค่าแม็กซิมั่มไลค์ลิฮู้ด (Maximum Likelihood Estimation) ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) สามารถสรุปเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้

$$P(X,\theta,\beta) = L(X \mid \theta,\beta) f(\theta) g(\beta)$$

1.1 การแจกแจงเริ่มแรก (Prior distribution) หมายถึง ลักษณะการแจกแจงความน่าจะเป็นที่เก็บรวบรวมสารสนเทศได้จากการศึกษาในอดีตหรือความเชื่อของผู้วิจัยเกี่ยวกับลักษณะการกระจายของค่าพารามิเตอร์ของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ (β)

1.2 การแจกแจงภายหลัง (Posterior distribution) หมายถึง ลักษณะการแจกแจงความน่าจะเป็นของคะแนนสังเกตได้ (P (X)) จากการทดสอบของผู้สอบแต่ละคน (n) สำหรับข้อสอบแต่ละข้อ (i) โดยได้จากการประมาณค่าด้วยวิธีการประมาณค่าของเบส์

1.3 การแจกแจงการทำนายภายหลัง (Posterior predictive distribution) หมายถึง ลักษณะการแจกแจงความน่าจะเป็นของตัวแปรที่ไม่ทราบค่าแต่สามารถสังเกตค่านั้นได้ (observed variable) โดยใช้การอนุมานเชิงทำนาย (Predictive inference) จากความน่าจะเป็นของการแจกแจงภายหลัง (Posterior distribution) มาใช้ในประมาณค่าให้เป็นคะแนนที่สังเกตได้สำหรับการทดสอบหนึ่งๆ หรือสามารถสรุปได้ว่า **การแจกแจงการทำนายภายหลัง (Posterior predictive distribution)** เป็นคะแนนที่สังเกตได้ (Observed score) จากการทดสอบของผู้สอบแต่ละคนในแต่ละข้อ (1=ตอบถูก 0=ตอบผิด) โดยได้มาจากการใช้สถิติอนุมานเชิงทำนาย (Predictive inference) ของความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบแต่ละคนในแต่ละข้อ $(P_i(\theta))$ ที่ได้มาจากการแจกแจงภายหลัง (Posterior distribution) ซึ่งเป็นคะแนนที่คาดหวัง (Expected value)

2. โมเดลการตอบสนองข้อสอบ หมายถึง ระบบที่แสดงความสัมพันธ์ระหว่างโอกาสของการตอบข้อสอบถูก (P_i) กับความสามารถที่มีอยู่ภายในของผู้สอบ (θ) ในรูปโค้งลักษณะข้อสอบ (ICC) ซึ่งมีลักษณะเป็นฟังก์ชันโลจิสติก โดยความสัมพันธ์ดังกล่าวอธิบายด้วยลักษณะของข้อสอบอันประกอบด้วยค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดาข้อสอบถูก (c)

2.1 ค่าความสามารถที่แท้จริงของผู้สอบ (θ) หมายถึง ระดับความสามารถของผู้สอบแต่ละคนที่ประมาณค่าจากการตอบข้อสอบได้ถูกต้องโดยใช้วิธีประมาณค่าของเบส์ (Bayesian estimation) ในโมเดลการตอบสนองข้อสอบ ซึ่งมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0 และ 1 ตามลำดับ ค่าความสามารถที่แท้จริงมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$

2.2 ค่าพารามิเตอร์ของข้อสอบ (β) หมายถึง ค่าที่บ่งบอกถึงลักษณะของข้อสอบซึ่งประกอบด้วยค่าความยากง่าย (b_i) ค่าอำนาจจำแนก (a_i) และค่าการเดา (c_i) ของข้อสอบ

2.2.1 ค่าความยากของข้อสอบ (b_i) หมายถึง ค่าพารามิเตอร์ของข้อสอบที่อยู่ตำแหน่งบนโค้งลักษณะข้อสอบซึ่งอยู่บนมาตรความสามารถซึ่งทำให้โอกาสของการตอบข้อสอบถูกเท่ากับ $\frac{1+c_i}{2}$ โดยข้อสอบที่ยากจะมีค่าความยากของข้อสอบไปทางขวา จุดเปลี่ยนโค้งลักษณะข้อสอบเป็นจุดที่ค่า $\theta = b_i$ สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 และ 2

พารามิเตอร์ ค่าความยากของข้อสอบเป็นระดับความสามารถของผู้สอบที่มีค่าความน่าจะเป็นในการตอบข้อสอบถูกเท่ากับ 0.5 แต่ถ้ามีการเดาแล้ว ค่าความยากของข้อสอบจะเป็นระดับความสามารถของผู้สอบที่มีค่าความน่าจะเป็นในการตอบข้อสอบถูกมากกว่าหรือเท่ากับค่า c แต่มีค่าน้อยกว่าหรือเท่ากับ 1.0 ค่าความยากง่ายที่นิยมใช้โดยมีค่าอยู่ระหว่าง -2.50 ถึง +2.50 แต่ในทางทฤษฎีจะมีค่าอยู่ระหว่าง $(-\infty,+\infty)$

2.2.2 ค่าอำนาจจำแนกของข้อสอบ (a_i) หมายถึง ค่าพารามิเตอร์ของข้อสอบที่เป็นสัดส่วนของความชันของโค้งลักษณะข้อสอบที่ตำแหน่ง b_i ระหว่างผู้ที่มีความสามารถน้อยกว่าหรือเท่ากับค่า θ กับมีค่ามากกว่าค่า θ ณ จุดเปลี่ยนโค้งซึ่งจำแนกค่าความน่าจะเป็นในการตอบข้อสอบถูก ค่าอำนาจจำแนกของข้อสอบที่สูงแสดงถึงความสามารถในการจำแนกผู้สอบที่มีความสามารถสูงออกจากผู้สอบที่มีความสามารถต่ำได้ดี ค่าอำนาจจำแนกของข้อสอบนิยมใช้อยู่ในช่วง 0.50 ถึง +2.50 แต่ในทางทฤษฎีมีค่าอยู่ระหว่าง $(-\infty,+\infty)$

2.2.3 ค่าการเดาของข้อสอบ (c_i) หมายถึง ค่าความน่าจะเป็นของผู้สอบที่ไม่มีความสามารถเลย $(\theta = -\infty)$ ที่จะตอบข้อสอบข้อนั้นได้ถูกต้อง นิยมกำหนดให้ค่าการเดาของข้อสอบมีค่าไม่เกิน 0.30 แต่ในทางทฤษฎีมีค่าอยู่ระหว่างค่า o ถึงค่า 1

2.3 โค้งลักษณะข้อสอบ หมายถึง กราฟของฟังก์ชันการตอบสนองข้อสอบที่แสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับความน่าจะเป็นในการตอบข้อสอบข้อหนึ่งๆ ได้ถูกต้องของผู้สอบ

3. โมเดลตอบสนองข้อสอบแบบหนึ่งพารามิเตอร์ (1PL IRT) หมายถึง รูปแบบการแสดงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบถูกต้องกับระดับความสามารถของผู้สอบซึ่งแสดงในรูปโค้งลักษณะข้อสอบ (ICC) เมื่อค่าความยากของข้อสอบ (b_i) เปลี่ยนแปลงไป แต่ค่าอำนาจจำแนกของข้อสอบ (a_i) คงที่และค่าโอกาสในการเดาของข้อสอบ (c_i) เท่ากับศูนย์

4. การวิเคราะห์ข้อสอบด้วยรูปแบบ Original GIRM หมายถึง วิธีการประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM ที่พัฒนาโดย Briggs และ Wilson (2007) โดยขั้นตอนแรกของการวิเคราะห์เพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ (IRT) ใช้เทคนิค MCMC (Markov Chain Monte Carlo) โดยการคำนวณแบบ Metropolis-Hastings algorithm ภายในวิธีการ Gibbs sampler ด้วยโปรแกรม WinBUGS ซึ่งจะสามารถประมาณค่าการแจกแจงภายหลัง (Posterior distribution) ของค่าพารามิเตอร์ได้ หลังจากนั้นจึงเอาค่าพารามิเตอร์ที่ได้มาคำนวณค่า E(X_{pi}) หรือก็คือค่า $P(\theta_p, \beta_i)$ ด้วยโมเดลการตอบสนองข้อสอบแบบ 1

พารามิเตอร์ และนำข้อมูลทั้งหมดมาสร้างเป็นเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) จากนั้นจึงทำการวิเคราะห์ขั้นตอนที่สามต่อ เพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยการนำข้อมูลจากเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) มาเป็นข้อมูลในการวิเคราะห์หาค่าองค์ประกอบความแปรปรวนด้วยสูตรการคำนวณของ Kolen และ Harris (1987)

5. **การวิเคราะห์ข้อสอบด้วยรูปแบบ** Alternative GIRM A หมายถึง วิธีการประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่นักวิจัยเสนอขึ้น ซึ่งมีวัตถุประสงค์หลักเพื่อแก้ไขปัญหาการประมาณค่าพารามิเตอร์ที่ไม่เป็นอิสระต่อกันระหว่างแถวและคอลัมน์ในเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) โดยใช้คะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) แทนคะแนนความน่าจะเป็นในการตอบข้อสอบถูกซึ่งเป็นคะแนนที่คาดหวัง (Expected value) ก่อนนำไปวิเคราะห์เพื่อให้สารสนเทศของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ด้วยสูตรของการวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA)

6. **การวิเคราะห์ข้อสอบด้วยรูปแบบ** Alternative GIRM B หมายถึง วิธีการประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่นักวิจัยเสนอขึ้น รูปแบบการวิเคราะห์คล้ายคลึงกับรูปแบบการวิเคราะห์ Alternative GIRM A ซึ่งใช้คะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) แทนคะแนนความน่าจะเป็นในการตอบข้อสอบถูกซึ่งเป็นคะแนนที่คาดหวัง (Expected value) ในเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) ก่อนนำไปวิเคราะห์เพื่อให้สารสนเทศของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) แต่การวิเคราะห์องค์ประกอบความแปรปรวนของรูปแบบนี้จะวิเคราะห์โดยใช้สูตรของการคำนวณของ Kolen และ Harris (1987) แทนสูตรของการวิเคราะห์ความแปรปรวน (Analysis of Variance; ANOVA) ในรูปแบบ Alternative GIRM A

7. **การวิเคราะห์ข้อสอบด้วยรูปแบบ** Original Bayesian GIRM หมายถึง วิธีการประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่นักวิจัยเสนอขึ้น ซึ่งมีวัตถุประสงค์หลักเพื่อแก้ไขปัญหา การประมาณค่าพารามิเตอร์ที่ไม่เป็นอิสระต่อกันระหว่างแถวและคอลัมน์ในเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) โดยใช้เทคนิคการประมาณค่าพารามิเตอร์แบบ

เบส์ (Bayesian Estimation) ที่ Swaminathan, Gifford, (1982, 1985, 1986) เป็นผู้เสนอไว้ ซึ่งสามารถประมาณค่าการแจกแจงภายหลัง (Posterior distribution) ของค่าพารามิเตอร์ได้ หลังจากนั้นจึงเอาค่าพารามิเตอร์ที่ได้มาคำนวณค่า E (X_{pi}) หรือก็คือค่า $P(\theta_p, \beta_i)$ ด้วยโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ และนำข้อมูลทั้งหมดมาสร้างเป็นเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) แล้วจึงนำไปวิเคราะห์ตามวิธีการวิเคราะห์เพื่อให้ได้สารสนเทศของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

8. **การจำลองข้อมูล** หมายถึง การจัดสถานการณ์การจำลองข้อมูลตามเงื่อนไขของจำนวนกลุ่มตัวอย่าง และจำนวนข้อสอบเพื่อใช้ในการประมาณค่าพารามิเตอร์ของวิธีการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) โดยการสร้างเลขสุ่มในโปรแกรม R

9. **การกระจายแบบโค้งปกติ** (Normal distribution) หมายถึง ลักษณะการแจกแจงของค่าพารามิเตอร์ความสามารถของผู้สอบและค่าพารามิเตอร์ของค่าความยากของข้อสอบที่มีลักษณะคล้ายรูประฆังคว่ำ ซึ่งมีค่าเฉลี่ยเท่ากับ 0 และมีค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 โดยการแจกแจงลักษณะนี้เป็นการแจกแจงความน่าจะเป็นของตัวแปรที่มีค่าส่วนมากมีค่าใกล้เคียงค่าเฉลี่ย และมีค่าของตัวแปรที่มากกว่าหรือน้อยกว่าค่าเฉลี่ยเป็นส่วนน้อย

10. **การกระจายแบบเอกรูป** (Uniform distribution) หมายถึง ลักษณะการแจกแจงของค่าพารามิเตอร์ความสามารถของผู้สอบและค่าพารามิเตอร์ของค่าความยากของข้อสอบที่มีลักษณะการแจกแจงคล้ายรูปสี่เหลี่ยมผืนผ้า ซึ่งข้อมูลทุกค่ามีโอกาสในการกระจายตัวที่เท่ากัน

11. **การกระจายแบบแกมมา** (Gamma distribution) หมายถึง การแจกแจงของค่าพารามิเตอร์ของผู้สอบ ซึ่งได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบ ซึ่งมีลักษณะการแจกแจงแบบโค้งเบ้ขวา (positive skewed curve) โดยมีคะแนนไม่เกิน 1 และมีลักษณะการกระจายแบบเอ๊กซ์โปแนนเชียล (Exponential function)

12. **การกระจายแบบเบต้า** (Beta distribution) หมายถึง ลักษณะการแจกแจงของค่าพารามิเตอร์ของข้อสอบซึ่งได้แก่ ค่าความยากของข้อสอบ ซึ่งมีลักษณะการแจกแจงแบบโค้งรูปตัวยู (U-shape curve) ค่าพารามิเตอร์ที่ต้องการประมาณค่ามีค่าอยู่ระหว่าง 0 ถึง 1

13. **ประสิทธิภาพของการวิเคราะห์** หมายถึง ดัชนีที่ใช้ในการระบุถึงคุณภาพของวิธีการประมาณค่าด้วยวิธีการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ประกอบด้วย ความลำเอียงในการประมาณค่า (Biased estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD)

14. ประสิทธิภาพการประมาณค่าของสารสนเทศตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด หมายถึง ดัชนีที่ใช้ในการระบุถึงคุณภาพของวิธีการประมาณค่าด้วยวิธีการประมาณค่าของทฤษฎีการตอบสนองข้อสอบ ประกอบด้วย ค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพันธ์และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์

14.1 **องค์ประกอบความแปรปรวน** (Variance components) หมายถึง องค์ประกอบของความผันแปรของคะแนนที่สังเกตได้ (Observed score) ในการศึกษาครั้งนี้ศึกษาแหล่งของความคลาดเคลื่อนจากการวัด 1 แหล่ง ได้แก่ จำนวนข้อสอบ (i) สามารถเขียนรูปแบบได้วิเคราะห์ได้เป็น รูปแบบ p X i ดังนั้นองค์ประกอบของความแปรปรวนของคะแนนที่สังเกตได้ ($\sigma^2(X_{pi})$) ของการออกแบบในการศึกษาครั้งนี้จึงประกอบด้วย ความแปรปรวนของบุคคล($\sigma^2(p)$) ความแปรปรวนของข้อสอบ ($\sigma^2(i)$) ความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบ ($\sigma^2(pi)$) และความแปรปรวนของความคลาดเคลื่อนแบบสุ่ม ($\sigma^2(e)$)

14.2 **ค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพันธ์** หมายถึง ค่าที่บ่งบอกถึงประสิทธิภาพของการวัดตามแนวการวิเคราะห์ข้อสอบด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด มีลักษณะคล้ายคลึงกับค่าความเที่ยงของการทดสอบ โดยค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพันธ์ จะใช้ในสถานการณ์ที่ต้องการเปรียบเทียบคะแนนระหว่างผู้สอบ บางครั้งอาจเรียกว่า Generalizability coefficient มีสูตรการคำนวณแหล่งของความคลาดเคลื่อน 1 แหล่ง ในรูปแบบการทดสอบแบบ p X i ได้ดังนี้

$$E\hat{p}^2 = \frac{\hat{\sigma}^2(p)}{\hat{\sigma}^2(p) + \frac{\hat{\sigma}^2(pi,e)}{n'_i}}$$

14.3 **ค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์** หมายถึง ค่าที่บ่งบอกถึงประสิทธิภาพของการวัดตามแนวการวิเคราะห์ข้อสอบด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด มีลักษณะคล้ายคลึงกับค่าความเที่ยงของการทดสอบ โดยค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ จะใช้ในสถานการณ์ที่ต้องการพิจารณาคะแนนผู้เข้ารับการทดสอบในแต่ละคน หรือ เป็นการแปลความหมายเชิงสัมบูรณ์ (Absolute) บางครั้งอาจเรียกว่า index of dependability มีสูตรการคำนวณแหล่งของความคลาดเคลื่อน 1 แหล่ง ในรูปแบบการทดสอบแบบ p X i ได้ดังนี้

$$\Phi = \frac{\hat{\sigma}^2(p)}{\hat{\sigma}^2(p) + \frac{\hat{\sigma}^2(pi,e)}{n_i'} + \frac{\hat{\sigma}^2(i)}{n'_i}}$$

15. **ประสิทธิภาพของการประมาณค่า** หมายถึง ดัชนีที่ใช้ในการระบุถึงคุณภาพของพารามิเตอร์ที่ได้จากการวิเคราะห์ข้อสอบด้วยวิธีการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ประกอบด้วย ความลำเอียงในการประมาณค่าความไม่แน่นอนในการประมาณค่า และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน

15.1 **ความลำเอียงในการประมาณค่า** (Biased estimator) หมายถึง ดัชนีที่ใช้ในการระบุถึงความมีประสิทธิภาพในการประมาณค่าเกี่ยวกับความลำเอียง โดยวัดได้จากค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง ถ้าตัวแบบใดให้ค่าเฉลี่ยกำลังสองต่ำกว่าแสดงว่าตัวแบบนั้นเป็นตัวประมาณค่าที่มีประสิทธิภาพมากกว่า สามารถแสดงเป็นสูตรทางคณิตศาสตร์ได้ดังนี้

$$MAD(\theta) = \frac{\sum_{i=1}^{r}(\theta_i - \hat{\theta}_i)^2}{r}$$

15.2 **ความไม่แน่นอนในการประมาณค่า** (Uncertainty estimator) หมายถึง ดัชนีที่ใช้ในการระบุถึงความมีประสิทธิภาพในการประมาณค่าเกี่ยวกับความไม่แน่นอนในการประมาณค่า โดยพิจารณาได้จากส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation, S.D.) ของค่าองค์ประกอบความแปรปรวนที่ประมาณค่าได้ในแต่ละองค์ประกอบความแปรปรวนคะแนนของที่สังเกตได้จากการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM)

$$SD = \sqrt{\frac{\Sigma(\theta_i - \bar{x}\bar{\theta})^2}{N}}$$

15.3 **ประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน** หมายถึง ดัชนีที่ใช้เปรียบเทียบวิธีการประมาณค่าองค์ประกอบความแปรปรวนของคะแนนที่สังเกตได้ โดยพิจารณาการเปรียบเทียบขนาดเวกเตอร์ขององค์ประกอบความแปรปรวนที่ศึกษาในแต่ละวิธีการประมาณค่าพารามิเตอร์ของวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ด้วยการคำนวณระยะทางยุคลิด ถ้าวิธีการประมาณค่า

วิธีใดให้ระยะทางยุคลิดต่ำ แสดงว่ามีประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนสูง สามารถแสดงเป็นสูตรทางคณิตศาสตร์ได้ดังนี้

$$\text{EuC} = \frac{\sum_{i=1}^{m}\left\|\hat{\theta}_c - \tilde{\theta}\right\|}{N} = \frac{\sum_{i=1}^{m}\sqrt{(\hat{\sigma}^2_{\alpha c}-\sigma^2_{\alpha})+(\hat{\sigma}^2_{\beta c}-\sigma^2_{\beta})+(\hat{\sigma}^2_{\alpha\beta c}-\sigma^2_{\alpha\beta})+(\hat{\sigma}^2_{ec}-\sigma^2_{e})}}{N}$$

16. **ความไวของผลการวิเคราะห์** (Sensitivity) หมายถึง ความเปลี่ยนแปลงของผลการวิเคราะห์ เมื่อเงื่อนไขในการจำลองข้อมูล ซึ่งได้แก่ จำนวนข้อสอบและจำนวนกลุ่มตัวอย่างและเงื่อนไขของการวิเคราะห์ ซึ่งได้แก่ ลักษณะการแจกแจงเริ่มแรก (Prior distribution) ของค่าพารามิเตอร์ของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ (β) มีลักษณะที่แตกต่างกัน โดยถ้าผลการวิเคราะห์มีการเปลี่ยนแปลงมาก เมื่อเงื่อนไขในการจำลองข้อมูลเปลี่ยนไปก็จะทำให้การวิเคราะห์นั้นมีความไวของผลการวิเคราะห์มาก และถ้าผลการวิเคราะห์มีการเปลี่ยนแปลงน้อย เมื่อเงื่อนไขในการจำลองข้อมูลเปลี่ยนไปก็จะทำให้การวิเคราะห์นั้นมีความไวของผลการวิเคราะห์น้อย

17. **ความแกร่งของวิธีการวิเคราะห์** (Robustness) หมายถึง คุณสมบัติที่แสดงถึงการประมาณค่าพารามิเตอร์ที่ไม่เปลี่ยนแปลงเมื่อมีการเปลี่ยนแปลงเงื่อนไขของผลการประมาณค่าที่ไม่ตรงตามข้อตกลงเบื้องต้นของวิธีการในการประมาณค่าพารามิเตอร์ของทฤษฎีที่ใช้ในการวิเคราะห์ข้อสอบ ดังนั้นจะเห็นได้ว่า ความไวความแกร่งของวิธีการวิเคราะห์มีความสัมพันธ์กับความไวของผลการวิเคราะห์ โดยถ้าการวิเคราะห์ใดมีความแกร่งของวิธีการวิเคราะห์มาก การวิเคราะห์นั้นก็จะมีความไวของผลการวิเคราะห์ต่ำ และถ้าการวิเคราะห์ใดมีความแกร่งของวิธีการวิเคราะห์ต่ำ การวิเคราะห์นั้นจะมีความไวของผลการวิเคราะห์สูง

ประโยชน์ที่คาดว่าจะได้รับ

1. ทราบถึงวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบให้มีประสิทธิภาพ ซึ่งทำให้สามารถวิเคราะห์พารามิเตอร์ของผู้สอบและข้อสอบได้มีแม่นยำมากยิ่งขึ้น

2. ทราบถึงความแกร่งของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบเกี่ยวกับการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบและการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ เพื่อทำให้ผู้ที่ต้องการวิเคราะห์มีการตรวจสอบข้อมูลว่ามีความเหมาะสมที่จะนำข้อมูลมาวิเคราะห์ด้วยวิธีการวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบหรือไม่ เพื่อให้สามารถประมาณค่าพารามิเตอร์ให้มีความถูกต้องและแม่นยำ โดยไม่ทำให้ผลการวิเคราะห์เกิดความคลาดเคลื่อน

3. ทราบถึงขนาดกลุ่มตัวอย่างที่เหมาะสมสำหรับการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบให้เกิดประสิทธิภาพสูงสุด

4. ทราบถึงจำนวนข้อสอบที่เหมาะสมสำหรับการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบให้เกิดประสิทธิภาพสูงสุด

5. ผลจากการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนอง สามารถให้สารสนเทศทั้งจากการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ ซึ่งทั้งทราบพารามิเตอร์ของผู้สอบและพารามิเตอร์ของข้อสอบ ส่วนทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ซึ่งทำให้ทราบแหล่งความคลาดเคลื่อนของผลการวัด ดังนั้นการวิเคราะห์ด้วยวิธีการนี้ในครั้งเดียว จะได้สารสนเทศครอบคลุมทั้งสองทฤษฎี

บทที่ 2

แนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้อง

การนำเสนอสาระสำคัญของรายงานเกี่ยวกับแนวคิด ทฤษฎี และเอกสารรายงานการวิจัยที่เกี่ยวข้องกับวิธีการวิเคราะห์และการประมาณค่าพารามิเตอร์ของการวิเคราะห์ข้อสอบทั้งจากการวิเคราะห์ข้อมูลจริง และข้อมูลจำลอง ผู้วิจัยได้ศึกษาค้นคว้าและสังเคราะห์จากหนังสือ เอกสาร บทความ รายงานการวิจัยทั้งในและต่างประเทศ รวมทั้งฐานข้อมูลการวิจัยที่สำคัญ ซึ่งได้แก่ EBSCOhost, ProQuest, Sage Journals online, Science Direct, Springer Link และWilson web ทำให้ผู้วิจัยสามารถแบ่งการนำเสนอแนวคิด ทฤษฎี เอกสารและงานวิจัยที่เกี่ยวข้องออกเป็น 4 ตอน โดยการนำเสนอในแต่ละตอนผู้วิจัยได้มีการบูรณาการงานวิจัยที่เกี่ยวข้องในตอนนั้น ๆ ไว้เรียบร้อยแล้ว ซึ่งมีรายละเอียดในแต่ละตอนดังต่อไปนี้

ตอนที่ 1 แนวคิด/ทฤษฎีเกี่ยวกับการวิเคราะห์ข้อสอบ

1.1 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

1.2 ทฤษฎีการตอบสนองข้อสอบ

1.3 วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ

1.4 บทสรุปจากการศึกษาแนวคิด/ทฤษฎีเกี่ยวกับการวิเคราะห์ข้อสอบ

ตอนที่ 2 การประมาณค่าพารามิเตอร์ของการวิเคราะห์ข้อสอบ

2.1 วิธีการประมาณค่าพารามิเตอร์แบบฮิวริสติก

2.2 วิธีการประมาณค่าพารามิเตอร์แบบแม็กซิมั่มไลค์ลิฮูด

2.3 วิธีการประมาณค่าพารามิเตอร์แบบเบส์

2.4 บทสรุปจากการศึกษาการประมาณค่าพารามิเตอร์การวิเคราะห์ข้อสอบ

ตอนที่ 3 การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล

3.1 ความเป็นมาของเทคนิคมอนติคาร์โล

3.2 ขั้นตอนการจำลองข้อมูล

3.3 การประยุกต์ใช้เทคนิคมอนติคาร์โลในการวิเคราะห์ข้อสอบ

3.4 ประสิทธิภาพของการจำลองข้อมูล

3.5 ข้อดีและข้อจำกัดของการศึกษาด้วยเทคนิคการจำลองข้อมูล

3.6 บทสรุปจากการศึกษาการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล

ตอนที่ 4 กรอบแนวคิดการวิจัย

ตอนที่ 1 แนวคิด/ทฤษฎีเกี่ยวกับการวิเคราะห์ข้อสอบ

ทฤษฎีการวิเคราะห์ข้อสอบ เป็นทฤษฎีที่มุ่งอธิบายเพื่อให้สามารถสรุปคะแนนจริงหรือความสามารถที่แท้จริงของบุคคลที่ทำการทดสอบได้ โดยคะแนนจริงนั้นมีลักษณะเป็นตัวแปรแฝง (Latent Variable) ซึ่งไม่สามารถวัดหรือสังเกตได้โดยตรง ดังนั้นจึงมีทฤษฎีการวิเคราะห์ข้อสอบเกิดขึ้นเพื่อทำให้สามารถประมาณค่าคะแนนจริงหรือความสามารถที่แท้จริงได้ตามนิยามหรือข้อตกลงเบื้องต้นของแต่ละทฤษฎี ปัจจุบันแนวคิด/ทฤษฎีการวิเคราะห์ข้อสอบที่นิยมใช้และจัดว่าเป็นแนวคิด/ทฤษฎีการวิเคราะห์ข้อสอบแนวใหม่ ประกอบด้วย 3 แนวคิด/ทฤษฎี คือ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ทฤษฎีการตอบสนองข้อสอบ (IRT) และวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) มีรายละเอียดในแต่ละแนวคิด /ทฤษฎีดังต่อไปนี้

1.1 ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory: GT)

ทฤษฎีการทดสอบในปัจจุบันสามารถแบ่งออกได้ใหญ่ๆ เป็น 2 ประเภท ได้แก่ ทฤษฎี การทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) และทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) ทฤษฎีการทดสอบแนวใหม่ในปัจจุบันที่นิยมใช้กันอย่างแพร่หลายมี 2 ทฤษฎี คือ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory: GT) และทฤษฎี การตอบสนองข้อสอบ (Item test Theory: IRT) โดยในส่วนนี้จะได้กล่าวเกี่ยวกับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) มีรายละเอียดดังต่อไปนี้

1.1.1 แนวคิดและวิธีการ

จากทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) คะแนนสอบที่สังเกตได้ (Observed score: X) เป็นผลรวมของโมเดลเชิงเส้นจากผลรวมคะแนนจริง (True score:T) และคะแนนความคลาดเคลื่อน (Error: ε) เมื่อพิจารณาจากโมเดลของทฤษฎีแล้ว พบว่า คะแนนความคลาดเคลื่อนนั้นมีลักษณะเป็นหนึ่งเดียวที่ไม่สามารถแบ่งแยกได้ (unique error) ซึ่งเป็นข้อจำกัดที่สำคัญสำหรับการหาค่าความเที่ยงของแบบวัดที่สามารถวิเคราะห์แหล่งความคลาดเคลื่อนได้เพียงทีละแหล่งเท่านั้น

$$\mathrm{X} = T + \varepsilon$$ (**จาก** Classical Test Theory: CTT)

สำหรับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เป็นที่ทฤษฎีที่อธิบายโดยใช้วิธีการวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) เข้ามาช่วยในการอธิบายแหล่งของความคลาดเคลื่อนที่เกิดขึ้นจากหลายแหล่ง (Multiple sources of error) โดยสามารถแยกส่วนความคลาดเคลื่อนของการวัดได้ออกเป็นสองแหล่งหลัก ๆ ประกอบด้วยความคลาดเคลื่อนอย่างเป็นระบบ (systematic error variance) และความคลาดเคลื่อนแบบสุ่ม (random error variance) รวมกันเป็นความคลาดเคลื่อนที่ได้จากการวัด (ε) ตามทฤษฎีการทดสอบแบบดั้งเดิม โดยในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของการวัดนั้นจะมีการระบุแหล่งของความคลาดเคลื่อนของการวัดที่น่าจะเกิดขึ้นจากการทดสอบซึ่งเรียกว่า ฟาเซต (Facet) โดยฟาเซตในแต่ละฟาเซตของการทดสอบตามเงื่อนไขของทฤษฎีการสรุปอ้างอิงจะต้องเป็นตัวอย่างของแหล่งความคลาดเคลื่อนที่ได้มาจากการสุ่ม (random) ของแหล่งความคลาดเคลื่อนที่น่าจะเกิดขึ้นได้ทั้งหมดของการทดสอบ ดังนั้นคะแนนสอบของผู้สอบที่ได้จากการวิเคราะห์แบบทฤษฎีการสรุปอ้างอิง จึงเป็นคะแนนที่คาดหวังของบุคคล (person's expected score) มากกว่าจะเป็นคะแนนที่สังเกตได้ (person's observed score) เนื่องจากคะแนนตามทฤษฎีนี้เป็นคะแนนที่ได้มาจากการสุ่มตามเงื่อนไขของแหล่งความคลาดเคลื่อนที่น่าจะเกิดขึ้น (randomly sampled facet conditions)

สมมุติว่ามีการทดสอบหนึ่งเป็นการออกแบบการวิเคราะห์แบบ a single-facet measurement โดยกำหนดให้ X_{pi} เป็นคะแนนที่สังเกตได้แบบทวิภาค (dichotomous) ของแต่ละบุคคล p (p = 1, 2, …,P) ในแต่ละข้อ i (i = 1, 2,…,I) โดยสามารถเขียนแบบแผนการวิเคราะห์ได้คือ p x i เมื่อ p คือบุคคลซึ่งเป็นหน่วยของการวัดและ i คือจำนวนข้อสอบซึ่งเป็นฟาเซตของการทดสอบ จากการกำหนดดังกล่าวทำให้สามารถแสดงค่าเฉลี่ยของคะแนนรวมจากบุคคลและข้อคำถาม (The grand mean across persons and items) ได้ดังนี้

$$\mu \equiv E_p E_i X_{pi}$$

คะแนนที่เกิดจากความสามารถของบุคคล (person-specific mean) มีค่าเท่ากับ

$$\mu_p \equiv E_i X_{pi}$$

และคะแนนของข้อคำถาม (item-specific mean) มีค่าเท่ากับ

$$\mu_i \equiv E_p X_{pi}$$

การกำหนดดังกล่าวสามารถเขียนโมเดลเชิงเส้นตรงของคะแนน X_{pi} ในแบบ GT ได้ดังนี้

$$X_{pi} = \mu + \nu_p + \nu_i + \nu_{pi,e} \quad \textbf{(จาก Generalizability Theory: GT)}$$

เมื่อ X_{pi} = คะแนนที่สังเกตได้

μ = ค่าเฉลี่ยรวม (The grand mean)

ν_p = อิทธิพลของบุคคล (The person effect)

$$\nu_p = \mu_p - \mu$$

ν_i = อิทธิพลของข้อสอบ (The item effect)

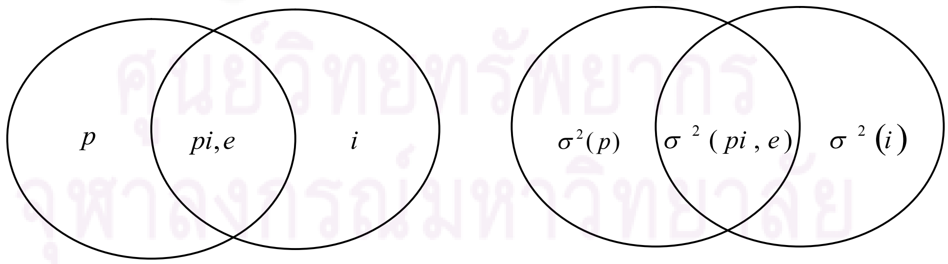
$$\nu_i = \mu_i - \mu$$

$\nu_{pi,e}$ = อิทธิพลของปฏิสัมพันธ์และความคลาดเคลื่อนแบบสุ่ม (The residual/interaction effect)

$$\nu_{pi,e} = X_{pi} - \mu_p - \mu_i + \mu$$

สามารถเขียนสมการในรูปของความแปรปรวนของคะแนนที่สังเกตได้ดังนี้

$$\sigma^2(X_{pi}) = \sigma^2(p) + \sigma^2(i) + \sigma^2(pi,e)$$



แหล่งของความคลาดเคลื่อน **องค์ประกอบความแปรปรวนของ GT**

แผนภาพที่ 1 แผนภาพเวนส์ (Venn Diagram) ของการวิเคราะห์ GT 1 ฟาเซต สำหรับการออกแบบ p x i

การวิเคราะห์ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) มีขั้นตอน การวิเคราะห์ประกอบไปด้วย 2 ขั้นตอนหลักๆ คือ ขั้นตอนแรกเป็นการวิเคราะห์แบบ G–study (Generalizability study) ซึ่งเป็นกระบวนการประมาณค่าความแปรปรวนขององค์ประกอบในแต่ละเงื่อนไขของการออกแบบการวัดและขั้นตอนที่สองเป็นการวิเคราะห์แบบ D-study (Decision study) ซึ่งเป็นกระบวนการประมาณค่าความแปรปรวนเมื่อเงื่อนไขในแต่ละแหล่งความคลาดเคลื่อนเพิ่มขึ้นหรือลดลง เพื่อที่จะตัดสินใจเลือกจำนวนในแต่ละองค์ประกอบของความคลาดเคลื่อนที่เหมาะสม

เป้าหมายหลักที่สำคัญของการคำนวณในขั้น G-Study คือการประมาณองค์ประกอบของความแปรปรวนของตัวแปรโดยใช้สมการประมาณ Expected Mean Square Equation (EMS) ดังนี้

$$EMS(p) = \sigma^2(pi) + n_i\sigma^2(p)$$

$$EMS(i) = \sigma^2(pi) + n_p\sigma^2(i)$$

$$EMS(pi,e) = \sigma^2(pi,e)$$

จากสมการค่า n_i และ n_p แทนจำนวนข้อและคนของกลุ่มตัวอย่างในการศึกษาขั้น G-Study จากสมการในการคำนวณค่า $EMS(.)$ ในขั้นนี้ทำให้เราสามารถประมาณค่าองค์ประกอบของ ความแปรปรวนในการศึกษา G-Study นั้นก็คือสามารถประมาณค่า $\sigma^2(p),\sigma^2(i),\sigma^2(pi,e)$ ซึ่งเป็นแหล่งความแปรปรวนของความคลาดเคลื่อนหลักของการออกแบบการวัดได้

สำหรับในการศึกษาขั้น D-Study นั้นเป็นการศึกษาตามเงื่อนไขของจำนวนตัวอย่างในแต่ละแหล่งความคลาดเคลื่อนที่สนใจศึกษา (facets) เพื่อทำให้การวัดนั้นเกิดประสิทธิภาพสูงสุดตาม การคำนวณค่า Generalizability coefficient เช่น ถ้าการออกแบบการวัดมีลักษณะเป็น p X i design เมื่อ i แสดงถึงจำนวนข้อที่ได้รับการทดสอบ ดังนั้นความแปรปรวนของคะแนนที่สังเกตได้จาก facet ของจำนวนข้อที่ได้รับการทดสอบจะมีค่าเท่ากับ $\frac{\sigma^2(i)}{n'_i}$ และความแปรปรวนของคะแนนที่ได้จากปฏิสัมพันธ์ระหว่างบุคคล ข้อสอบและความคลาดเคลื่อนแบบสุ่มจะมีค่าเท่ากับ $\frac{\sigma^2(pi),e}{n'_i}$

สามารถแสดงเมทริกซ์คะแนนที่สังเกตได้ของผู้สอบตามทฤษฎี GT ได้ดังนี้

ข้อสอบ (item)

		1	…	i	…	I
	1	X_{11}		…		X_{1I}
บุคคล	…					
(person)	p	…		X_{pi}		….
	…					
	P	X_{P1}		…		X_{PI}

แผนภาพที่ 2 เมทริกซ์คะแนนที่สังเกตได้ของผู้สอบตาม GT สำหรับการออกแบบการวัด p x i

1.1.2 ข้อตกลงเบื้องต้นของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เป็นทฤษฎีที่เข้ามาช่วยในการอธิบายแหล่งของความคลาดเคลื่อนที่เกิดจากหลายแหล่งทั้งความคลาดเคลื่อนที่เป็นระบบ (systematic error) และความคลาดเคลื่อนแบบสุ่ม (random error) โดยเรียกรวมกันเป็นความคลาดเคลื่อน (ε) ของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ซึ่งเป็นข้อจำกัดสำคัญในการค่าความเที่ยงของแบบสอบที่สามารถวิเคราะห์ได้ที่ละแหล่งความคลาดเคลื่อนเท่านั้น โดยมีข้อตกลงเบื้องต้นของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ดังนี้

1. คุณลักษณะที่มุ่งวัดต้องเป็นค่าที่อยู่ในสภาวะคงที่ (Steady state)

2. กระบวนการในการวัดจะต้องมีความคลาดเคลื่อนที่เป็นระบบ (Systematic error variance) อย่างน้อย 1 แหล่งที่เป็นอิสระต่อกัน โดยมีข้อตกลงเบื้องต้นว่าวุฒิภาวะ (Maturation) และกระบวนการเรียนรู้ (Learning) ระหว่างการวัดไม่เป็นแหล่งความคลาดเคลื่อนที่สำคัญของการวัด

3. ความแปรปรวนของคะแนนที่สังเกตได้ (Observed score) เกิดจากผลรวมเชิงเส้นของความแปรปรวนของคะแนนจริงหรือความสามารถที่แท้จริงของผู้สอบ ความแปรปรวนของความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่งและความแปรปรวนของคะแนนความคลาดเคลื่อนแบบสุ่ม

1.1.3 การวัดคุณภาพของเครื่องมือตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (G-Theory) มีฟังก์ชันที่บ่งบอกถึงประสิทธิภาพของการวัดอยู่สองประเภทคือ ค่าสัมประสิทธิ์ความน่าเชื่อถือของคะแนน ซึ่งมีลักษณะคล้ายคลึงกับค่าสัมประสิทธิ์ความเที่ยง (Reliability coefficient) ของทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory; CTT) โดยสามารถแบ่งตามการแปลความหมายออกได้เป็น 2 ประเภท ซึ่งได้แก่ Generalizability coefficient และ index of dependability โดย Generalizability coefficient เป็นการแปลความหมายในเชิงเปรียบเทียบ (Relative) (เช่น การตัดสินผลแบบอิงกลุ่ม) สำหรับ index of dependability เป็นการแปลความหมายเชิงสัมบูรณ์ (Absolute) (เช่น การตัดสินผลแบบอิงเกณฑ์) ในกรณี 1 facet pxi design สามารถเขียนแทนเป็นสมการได้ดังนี้

$$E\hat{p}^2 = \frac{\hat{\sigma}^2(p)}{\hat{\sigma}^2(p) + \dfrac{\hat{\sigma}^2(pi,e)}{n_i'}}$$

$$\Phi = \frac{\hat{\sigma}^2(p)}{\hat{\sigma}^2(p) + \dfrac{\hat{\sigma}^2(pi,e)}{n_i'} + \dfrac{\hat{\sigma}^2(i)}{n_i'}}$$

1.1.4 จุดเด่นของการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

1) การวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) สามารถอธิบายแหล่งของความคลาดเคลื่อนที่เกิดขึ้นจากหลายแหล่ง (Multiple sources of error) ในขณะที่ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) อธิบายความเคลื่อนที่ได้จากการวัดรวมเป็นหนึ่งเดียว โดยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) แยกส่วนความคลาดเคลื่อนของการวัดได้ออกเป็นสองแหล่งหลัก ๆ ประกอบด้วยความคลาดเคลื่อนอย่างเป็นระบบ (systematic error variance) และความคลาดเคลื่อนแบบสุ่ม (random error variance) โดยใช้วิธีการวิเคราะห์ทางสถิติที่เรียกว่า การวิเคราะห์ความแปรปรวน (Analysis of Variance: ANOVA) เข้ามาช่วยในการวิเคราะห์ จึงทำให้สามารถแยกส่วนความคลาดเคลื่อนของการวัดได้

2) การวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) มีโปรแกรมคอมพิวเตอร์สำเร็จรูป GENOVA (A GENeralized Analysis Of Variance System) ซึ่งเป็นโปรแกรมเฉพาะสำหรับการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยในการใช้งานนั้นผู้ใช้ระบุเพียงอิทธิพลหลัก (main effect) เท่านั้นโปรแกรมจะระบุ อิทธิพลปฏิสัมพันธ์ (interaction effect) ที่เป็นไปได้ทุกกรณี นอกจากนี้ในการออกแบบของสถานการณ์การทดสอบโปรแกรมนี้สามารถวิเคราะห์ได้ทั้งการออกแบบที่มีลักษณะ Crossed design (pxr) และ Nested design (r:p) และยังวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงสัมพันธ์ (Generalizability coefficient) เป็นการแปลความหมายในเชิงเปรียบเทียบและค่าสัมประสิทธิ์การสรุปอ้างอิงสัมบูรณ์ (index of dependability) เป็นการแปลความหมายเชิงสัมบูรณ์

นอกจากนี้จากการที่มีโปรแกรมเฉพาะสำหรับการวิเคราะห์ ทำให้การวิเคราะห์ข้อสอบตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ด้วยโปรแกรม GENOVA สามารถลดปัญหาความลำเอียงแบบการประมาณค่าต่ำกว่าที่ควรจะเป็น (underestimation) จากการประมาณความแปรปรวนขององค์ประกอบความคลาดเคลื่อน เนื่องจากการประมาณค่าสามารถทำให้เป็นอิสระจากกันได้ จึงไม่นำไปสู่การฝ่าฝืนข้อตกลงเบื้องต้นแบบสุ่มของโมเดล (random effects)

1.1.5 ข้อจำกัดของการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

1) การวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ประกอบไปด้วย แหล่งของความคลาดเคลื่อนที่เกิดจากหลายแหล่งทั้งความคลาดเคลื่อนที่เป็นระบบ (systematic error) และความคลาดเคลื่อนแบบสุ่ม (random error) โดยความคลาดเคลื่อนที่เป็นระบบนั้น จะต้องได้มาจากการสุ่ม (Random) จากเอกภพ (Universe) ของการวัด ดังนั้นในการประมาณค่าองค์ประกอบของความคลาดเคลื่อนของผู้สอบแต่ละคนจะต้องเหมือนกัน (fixed) ซึ่งจากหลักการของทฤษฎีนี้ ทำให้ได้ค่าความเที่ยงมากเกินจริง (Overestimate reliability) และค่าความคลาดเคลื่อนน้อยเกินจริง (underestimate error variance)

จากการสุ่มแหล่งความคลาดเคลื่อนที่เป็นระบบ (systematic error) จากเอกภพ (Universe) ของการวัด ทำให้ทฤษฎีนี้มีความเชื่อว่า การประมาณค่าของคะแนนที่ได้จากการวัด (observed score) สามารถสรุปอ้างอิงไปยังคะแนนจริงของผู้สอบ (true score) ที่ได้จากเอกภพของการวัด (universe score) แต่ในการนำสู่การปฏิบัติ (practical implication) ของนักวิชาการยังไม่เป็นที่ยอมรับ (unacknowledged) ตามหลักการและความเชื่อดังกล่าวของวิธีตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) (Brennan, 2007)

2) เมื่อพิจารณาดัชนีที่บอกประสิทธิภาพของการวัดทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ทั้ง Generalizability coefficient และ index of dependability โดยสมมติสถานการณ์หนึ่งว่าความแปรปรวนที่ได้จากการคะแนนจริงในกลุ่มตัวอย่างมีค่าต่ำ แต่การวัดมีคุณภาพสูง (มีค่าความคลาดเคลื่อนจากการวัดน้อย) จะทำให้การวิเคราะห์ตามทฤษฎีให้ค่า Generalizability coefficient ต่ำ ในทางตรงกันข้าม ถ้าสมมติว่า ความแปรปรวนที่ได้จากคะแนนจริงในกลุ่มตัวอย่างมีค่าสูง แต่การวัดมีคุณภาพต่ำ (มีค่าความคลาดเคลื่อนจากการวัดมาก) ค่า Generalizability coefficient จะมีค่าสูง ดังนั้นจากสถานการณ์ดังกล่าว จะเห็นได้ว่า การแปลความหมายตามดัชนีที่บอกประสิทธิภาพของการวัด ทำให้มีการแปลความหมายเป็นไปในแนวทางที่ผิด (misleading) ไม่สอดคล้องกับความจริงตามทฤษฎี ดังนั้นในการรายงานผลการวิเคราะห์ตามแนว GT ทั้งค่า Generalizability coefficient และ index of dependability จึงต้องมีการรายงานค่าความคลาดเคลื่อนมาตรฐานของการวัดทั้งในเชิงสมบูรณ์และในเชิงสัมพันธ์ ด้วยประกอบในการรายงานผลการวิเคราะห์ตามทฤษฎี (Briggs และ Wilson, 2007)

3) การวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ต้องมีการออกแบบข้อมูลแทนที่ขาดหายไป ทั้งนี้เนื่องจากการประมาณค่าองค์ประกอบความแปรปรวนตั้งอยู่บนพื้นฐานของการคาดคะเน (expected) มากกว่าข้อมูลจริงที่อยู่ในตารางเมทริกซ์ของข้อมูล จึงทำให้ส่งผลต่อการลงข้อมูล (specified) ที่ไม่สมดุลสำหรับการออกแบบแบบสุ่ม (unbalanced random effects design) จะไม่สามารถประมาณค่า G- Study และ D- study

4) การวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ไม่สามารถวิเคราะห์ค่าพารามิเตอร์ของผู้สอบเป็นรายบุคคลและค่าพารามิเตอร์ของการสอบเป็นรายข้อได้ โดยวิเคราะห์ได้แต่ภาพรวมของการทดสอบตามเงื่อนไขหรือสถานการณ์ในการทดสอบที่สนใจศึกษาตามแหล่งความคลาดเคลื่อนที่น่าจะเกิดขึ้นต่อการทดสอบ

1.1.6 สรุปสาระสำคัญของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เป็นทฤษฎีที่อธิบายโดยใช้วิธีการวิเคราะห์ความแปรปรวน (ANOVA) เข้ามาช่วยในการอธิบายแหล่งของความคลาดเคลื่อนที่เกิดขึ้นจากหลายแหล่ง (Multiple sources of error) โดยสามารถแยกส่วนความคลาดเคลื่อนของการวัดได้ออกเป็นสองแหล่งหลัก ๆ ประกอบด้วยความคลาดเคลื่อนอย่างเป็นระบบ (systematic error variance) และความคลาดเคลื่อนแบบสุ่ม (random error variance) รวมกันเป็นความคลาดเคลื่อนที่ได้จากการวัด (ε) ตามทฤษฎีการทดสอบแบบดั้งเดิม ดังนั้นจะเห็นว่าทฤษฎีการสรุปอ้างอิงความ

น่าเชื่อถือของผลการวัด (GT) สามารถอธิบายแหล่งความคลาดเคลื่อนในการวัดในมากกว่าทฤษฎีการทดสอบแบบดั้งเดิมซึ่งส่งผลทำให้ผลการวัดมีความน่าเชื่อถือมากยิ่งขึ้น การวัดคุณภาพของทฤษฎีนี้พิจารณาได้จากค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิง (G-coefficient) โดยสามารถแปลความหมายได้ทั้งในเชิงเปรียบเทียบพิจารณาได้ค่า Generalizability coefficient และแปลความหมายในเชิงสัมบูรณ์พิจารณาได้จากค่า index of dependability

1.2 **ทฤษฎีการตอบสนองข้อสอบ** (Item Response Theory: IRT)

การวิเคราะห์ข้อสอบด้วยทฤษฎีการสรุปอ้างอิง (GT) นั้นสามารถใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) ในการเกื้อหนุนกัน โดยสามารถแปลความหมายในเชิงที่ว่า ผู้สอบมีการตอบสนองข้อสอบอย่างไรเมื่อเจอข้อคำถามในแต่ละชุดเหมือนกับที่พบในทฤษฎีการสรุปอ้างอิง (GT) ซึ่งจะทำให้ทราบสารสนเทศของการทดสอบเพิ่มมากขึ้นในการวิเคราะห์ข้อสอบ (Briggs และ Wilson, 2007)

1.2.1 **แนวคิดและวิธีการ**

เมื่อพิจารณาโมเดลของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) พบว่า คะแนนสอบที่สังเกตได้ (X) เป็นผลมาจาก ผลรวมของโมเดลเชิงเส้นของคะแนนจริง (T) และคะแนนที่มีความคลาดเคลื่อน (ε) โดยตั้งอยู่บนพื้นฐานข้อตกลงเบื้องต้นที่สำคัญ คือ ค่าพารามิเตอร์ของข้อสอบเป็นค่าเฉพาะของกลุ่มผู้สอบและเป็นอิสระไม่ขึ้นอยู่กับความสามารถของผู้สอบ จึงทำให้ค่าความคลาดเคลื่อนมาตรฐานของการวัด (SEM) เป็นค่าเฉพาะในแต่ละประชากรที่ใช้ในการทดสอบและเป็นค่าเดียวคงที่สำหรับทุกคนในประชากร การเปรียบเทียบคะแนนหรือคุณภาพของข้อสอบ จึงทำได้เฉพาะข้อสอบที่นำมาเปรียบเทียบกันมีคุณลักษณะเป็นข้อสอบคู่ขนาน (Parallel forms) ($T_i = T'_i \ \& \ \sigma_E^2 = \sigma_E'^2$) จากข้อตกลงเบื้องต้นดังกล่าวทำให้เป็นข้อจำกัดที่สำคัญของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ซึ่งทำให้ค่าพารามิเตอร์ของข้อสอบผันแปรตามกลุ่มผู้สอบ นอกจากนี้คะแนนที่สังเกตที่ได้จากการสอบหรือความสามารถของผู้สอบขึ้นอยู่กับข้อสอบและแบบสอบที่นำมาใช้ ดังนั้นทฤษฎีการทดสอบแบบดั้งเดิมจึงมีค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ของผู้สอบไม่เป็นอิสระต่อกันโดยขึ้นอยู่กับสถานการณ์แต่ละสถานการณ์ในการสอบ (ศิริชัย กาญจนวาสี, 2550)

ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่แท้จริงของบุคคลกับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกได้มากน้อยเพียงใด โดยใช้โค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันโลจิส (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) (Hambleton and Swaminathan, 1985)

สำหรับโมเดลการวิเคราะห์แบบ p x i ด้วยข้อคำถามแบบสองค่า (dichotomous items) สามารถแสดงสมการแบบ one parameter logistic หรือ Rasch model ได้ดังนี้

$$P(\theta_p, \beta_i) = P(X_{pi} = 1 \mid \theta_p, \beta_i) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}$$

ทฤษฎีการตอบสนองข้อสอบนี้เชื่อว่า พฤติกรรมการตอบสนองข้อสอบของผู้สอบเป็นความสามารถที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้โดยตรง และเชื่อว่าค่าพารามิเตอร์ของข้อสอบที่ได้จากการวิเคราะห์คุณภาพของข้อสอบ ทั้งค่าความยาก (b) ค่าอำนาจจำแนก (a) และ ค่าโอกาสในการเดาของข้อสอบ (c) แต่ละข้อเป็นคุณลักษณะเฉพาะของข้อสอบที่มีอยู่ประจำ และคงที่ในตัวข้อสอบ ฉะนั้นจึงทำให้การวิเคราะห์ข้อสอบด้วยทฤษฎีนี้ ค่าพารามิเตอร์ของข้อสอบจึงไม่ผันแปรหรือเปลี่ยนแปลงไปตามกลุ่มผู้สอบ นอกจากนี้ค่าพารามิเตอร์ของผู้สอบหรือค่าความสามารถของผู้สอบยังเป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบแต่ละคนที่ไม่เปลี่ยนแปลงไปตามคุณลักษณะของข้อสอบหรือค่าพารามิเตอร์ของผู้สอบอีกด้วย ดังนั้นจะเห็นได้ว่าทฤษฎีการตอบสนองข้อสอบนั้นสามารถแก้ข้อจำกัดที่สำคัญของทฤษฎีการทดสอบแบบดั้งเดิมได้ โดยทฤษฎีดังกล่าวสามารถแสดงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบถูกในแต่ละข้อ $P_i(\theta)$ กับระดับความสามารถที่แท้จริงของผู้สอบ (θ) แสดงเป็นกราฟความสัมพันธ์ที่เรียกว่า โค้งคุณลักษณะของข้อสอบ (Item Characteristics Curve: ICC) (ศิริชัย กาญจนวาสี, 2550)

1.2.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ มุ่งอธิบายความสัมพันธ์ระหว่างความสามารถที่แท้จริงของผู้สอบ (θ) กับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกได้มากน้อยเพียงใด $P_i(\theta)$ ด้วยโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ โดยมีข้อตกลงเบื้องต้นของทฤษฎีที่สำคัญดังนี้ (Hambleton และ Swaminathan, 1985)

1) ความเป็นเอกมิติของคุณลักษณะที่ใช้ในการทดสอบ (Unidimensional)

การวัดตามแนวทฤษฎีการตอบสนองข้อสอบในระยะแรกได้ระบุความเป็นเอกมิติของคุณลักษณะในข้อตกลงเบื้องต้นไว้ว่า ข้อคำถามหรือข้อสอบทุกข้อในแบบสอบนั้นต้องมุ่งวัดความสามารถหรือคุณลักษณะเพียงลักษณะเดียวหรือความสามารถเดียว แต่ข้อตกลงนี้สามารถผ่อนปรนได้ในกรณีการทดสอบในสถานการณ์ใดที่มุ่งวัดคุณลักษณะเด่นเพียงคุณลักษณะเดียวก็ถือได้ว่าการวัดในครั้งนั้นมีความเป็นเอกมิติของคุณลักษณะที่ใช้ในการทดสอบเหมือนกันไม่ขัดแย้งกับข้อตกลงเบื้องต้นดังกล่าว

การตรวจสอบความเอกมิติของคุณลักษณะที่ใช้ในการทดสอบมีวิธีการตรวจสอบได้หลายวิธี สามารถสรุปวิธีการที่สำคัญได้ดังนี้

1. การหาค่าความสัมพันธ์ระหว่างค่าน้ำหนักองค์ประกอบรายข้อ (factor loading) ขององค์ประกอบที่หนึ่งกับค่าสหสัมพันธ์แบบไบซีเรียล (biserial correlation coefficient) ของข้อสอบรายข้อกับคะแนนรวม ถ้ามีค่าสัมประสิทธิ์สหสัมพันธ์มากกว่า .80 ทำให้สามารถสรุปได้ว่าข้อสอบหรือแบบสอบฉบับนั้นมีความเป็นเอกมิติของคุณลักษณะที่ใช้ในการทดสอบ

2. การวิเคราะห์องค์ประกอบ (factor analysis) ของข้อสอบทั้งฉบับ พิจารณาได้จากค่าไอเกน (eigen value) โดยผลการวิเคราะห์องค์ประกอบใดมีค่าไอแกนในองค์ประกอบใดองค์ประกอบหนึ่งสูงกว่าค่าอื่นอย่างชัดเจน สามารถสรุปได้ว่าข้อสอบหรือแบบสอบฉบับมีความเป็นเอกมิติของคุณลักษณะในการทดสอบ

3. การใช้โปรแกรม TESTFACT ในการพิจารณาความเป็นมิติของแบบสอบจากการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confermatory Factor Analysis) ซึ่งพัฒนาขึ้นโดย Wilson และ Hoskens (1991) โดยวิเคราะห์ข้อสอบและทดสอบความตรงของโครงสร้าง ด้วย χ^2 สำหรับ Likelihood Ratio (G^2) ในการตรวจสอบมิติของแบบสอบ ดัชนีที่ใช้นี้ทดสอบด้วยการกำหนดจำนวนองค์ประกอบของชุดข้อมูลไว้ล่วงหน้าแล้วทดสอบด้วย χ^2 ที่ประมาณค่าด้วยวิธี G^2 เพื่อทดสอบ

ความเหมาะสมของโมเดล เมื่อค่าG^2 ไม่มีนัยสำคัญแสดงว่าข้อมูลมีจำนวนองค์ประกอบเท่าที่กำหนดในการทดสอบ

4. การวิเคราะห์ การจัดกลุ่มระดับชั้น (Hierarchical Cluster Analysis) เป็นเทคนิคสำหรับการทดสอบความเป็นพหุมิติของชุดแบบสอบ โดยการพิจารณาการแบ่งกลุ่มของตัวแปร โดยกระบวนการแบ่งกลุ่มนี้เป็นการแบ่งกลุ่มจำนวนข้อสอบที่มีลักษณะคล้ายคลึงกันให้อยู่ในกลุ่มเดียวกัน นอกจากนี้ยังใช้วิธีในการหมุนซ้ำ (Iteration) จนสำเร็จหรืออยู่ในระดับที่น่าพอใจ ซึ่งทำให้ได้ความเป็นไปได้ของผลลัพธ์ (Outcome) โดยสามารถช่วยอธิบายให้ข้อมูลมีความถูกต้องมากยิ่งขึ้น ซึ่งการวิเคราะห์ด้วยวิธีการนี้สามารถใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป CCPROX และ HCA ในการวิเคราะห์ได้

5. การใช้โปรแกรม DETECT เป็นการตรวจสอบมิติแฝงเชิงยืนยันแบบ Nonparametric ซึ่งจะใช้ในการประมาณค่าจำนวนของมิติแฝงที่มีคุณลักษณะเด่นในชุดของข้อมูลและสามารถตรวจสอบความเป็นเอกมิติของแบบสอบ โดยระบุคุณลักษณะเด่นของมิติแฝงในแต่ละข้อ ซึ่งผู้ใช้โปรแกรมสามารถระบุจำนวนมิติแฝงสูงสุดที่ต้องการศึกษาได้ เนื่องจากการจัดกลุ่มชุดของข้อสอบแต่กระบวนการดังกล่าวยังมีลักษณะแบบไม่เป็นทางการเท่าใดนัก เนื่องจากการระบุการจัดกลุ่มเพื่อจำแนกความแตกต่างของมิติจะอาศัยกระบวนการในการระบุความเป็นหนึ่งเดียว (Zhang และ Stout, 1999 อ้างถึงในพัชรี จันทร์เพ็ง, 2550)

6. การใช้โปรแกรม DIMTEST เป็นกระบวนการ ตรวจสอบสมมติฐานของ แบบสอบด้วย Nonparametric Statistical โดยมีลักษณะคล้ายคลึงกับการตรวจสอบด้วยโปรแกรม DETECT โดยการตรวจสอบจะตรวจสอบความสัมพันธ์ระหว่างชุดข้อสอบย่อยภายใต้เงื่อนไขความแปรปรวนร่วมของข้อสอบ ซึ่งแตกต่างจากการใช้โปรแกรม DETECT ที่มีลักษณะคล้ายการวิเคราะห์องค์ประกอบเชิงยืนยัน

2) ความเป็นอิสระของแบบสอบ (Local independence)

การออกแบบข้อสอบที่สร้างขึ้นโดยมีจุดหมายหลักสำหรับการใช้วัดคุณลักษณะเพียงอย่างเดียวนั้น ทั้งโมเดลการวัดแบบดั้งเดิม (CTT) หรือโมเดลการตอบสนองข้อสอบ (IRT) ความเป็นอิสระของข้อสอบถือว่าเป็น ข้อตกลงหลักที่สำคัญของทฤษฎี ซึ่งหมายถึงโอกาสของการตอบข้อสอบถูกในแต่ละข้อเป็นอิสระต่อกัน หรือกล่าวได้ว่า การตอบข้อสอบข้อหนึ่งจะต้องไม่ส่งผลกระทบต่อการตอบข้อสอบในอีกข้อหนึ่ง โดยถ้าไม่สามารถทำให้ข้อสอบเกิดความเป็นอิสระตามทฤษฎีนี้จะทำให้การประมาณค่าความสามารถของผู้สอบสารสนเทศของแบบสอบ ค่าความเที่ยงค่าสถิติของข้อสอบ และประสิทธิภาพของผู้สอบเกิดความผิดพลาดได้ (Zennisky และคณะ,2003;

Lee, 2004) ความเป็นอิสระสามารถแยกพิจารณาเป็นความเป็นอิสระระหว่างข้อสอบ และความเป็นอิสระระหว่างผู้สอบ โดยมีรายละเอียดดังนี้

2.1) ความเป็นอิสระระหว่างข้อสอบ

การตอบข้อสอบข้อหนึ่งจะต้องไม่มีผลต่อการตอบข้อสอบอีกข้อสอบหนึ่ง หรือเมื่อพิจารณาจากผลการตอบข้อสอบรายข้อของผู้สอบคนเดียวกัน ต้องไม่มีความเกี่ยวข้องกัน จึงเรียกว่า ข้อสอบแต่ละข้อนั้นเป็นอิสระต่อกัน โดยความน่าจะเป็นของแบบแผนการตอบของข้อสอบ k ข้อของผู้สอบที่มีความสามารถ θ จะเท่ากับผลคูณระหว่างความน่าจะเป็นของผลการตอบในแต่ละข้อ สามารถเขียนแทนด้วยสมการได้ดังนี้

$$\Pr ob[U_1 = u_1, U_2 = u_2 ... U_k = u_k / \theta] = P_1(\theta)^{u_1}.Q_1(\theta)^{1-u_1}.P_2(\theta)^{u_2}.Q_2(\theta)^{1-u_2}$$
$$......P_k(\theta)^{u_n}.Q_k(\theta)^{1-u_n}.$$
$$= \prod_{i=1}^{k} P_i(\theta)^{u_i}.Q_i(\theta)^{1-u_i}$$

2.2) ความเป็นอิสระหว่างผู้สอบ

การตอบข้อสอบข้อเดียวกันของผู้สอบคนหนึ่งจะต้องไม่มีผลต่อการตอบข้อสอบของผู้สอบอีกคนหนึ่ง หรือเมื่อพิจารณาจากผลการตอบข้อสอบของผู้สอบแต่ละคนในข้อเดียวกันต้องไม่มีความเกี่ยวข้องสัมพันธ์ซึ่งกันและกัน จึงเรียกว่า ผู้สอบแต่ละคนนั้นเป็นอิสระต่อกัน โดยความน่าจะเป็นของแบบแผนการตอบข้อสอบของผู้สอบ n คนจะเท่ากับผลคูณระหว่าง ความน่าจะเป็นของผลการตอบข้อนั้นของผู้สอบแต่ละคน สามารถเขียนแทนด้วยสมการได้ดังนี้

$$\Pr ob[U_1 = u_1, U_2 = u_2 ... U_n = u_n / \theta] = P_1(\theta)^{u_1}.Q_1(\theta)^{1-u_1}.P_2(\theta)^{u_2}.Q_2(\theta)^{1-u_2}$$
$$......P_n(\theta)^{u_n}.Q_n(\theta)^{1-u_n}.$$
$$= \prod_{i=1}^{n} P_i(\theta)^{u_i}.Q_i(\theta)^{1-u_i}$$

ในการตรวจสอบความอิสระของข้อสอบสามารถตรวจสอบได้จากการพิจารณาเมทริกซ์ความแปรปรวนและความแปรปรวนร่วม (variance–Covariance matrix) หรือ เมทริกซ์สหสัมพันธ์ (correlation matrix) ของคะแนนคำตอบรายข้อ โดยพิจารณาจากกลุ่มผู้สอบที่มีช่วงความสามารถที่เท่ากัน ซึ่งค่านอกแนวทแยงมุมของเมทริกซ์ต้องมีค่าต่ำหรือมีค่าเข้าใกล้ศูนย์ (ศิริชัย กาญจนวาสี, 2550) หรือสามารถตรวจสอบได้ โดยการตรวจสอบค่าสูงสุดของสหสัมพันธ์ภายในระหว่างข้อสอบ (Magnitude of inter-item correlations) ณ ที่ระดับช่วงคะแนนที่แตกต่างกันของผู้สอบ (Ferrara, Huynh และ Baghi, 1997) หรือสามารถตรวจสอบได้ โดยการสร้างรูปแบบอิทธิพลหลักและอิทธิพลการปฏิสัมพันธ์

ของข้อสอบเพื่ออธิบายความไม่เป็นอิสระของข้อสอบ โดยได้เสนอโมเดล 2 รูปแบบในการตรวจสอบความไม่เป็นอิสระของข้อสอบได้แก่ โมเดลปฏิสัมพันธ์คงที่ (constant interaction model) และโมเดลปฏิสัมพันธ์ของมิติที่ไม่เป็นอิสระ (dimension–dependent interaction model) นอกจากนี้ยังสามารถตรวจสอบความเป็นอิสระด้วยประยุกต์ใช้วิธีการวิเคราะห์พหุระดับ (HLM) พิจารณาจากลักษณะความเป็นเอกพันธ์ของการกระจายค่าความคลาดเคลื่อน (homoscesdasticity) และความเป็นอิสระของตัวแปรกลุ่มตัวอย่างภายในกลุ่มเดียวกันจะมีลักษณะคล้ายกันมากกว่ากลุ่มตัวอย่างระหว่างกลุ่ม ดังนั้นองค์ประกอบของกลุ่มแต่ละกลุ่มจะเป็นอิสระต่อกัน แต่จะมีความสัมพันธ์กันภายในกลุ่ม บางกลุ่มอาจมีความเป็นเอกพันธ์มากกว่ากลุ่มอื่น ๆ ดังนั้นความแปรปรวนขององค์ประกอบระหว่างกลุ่มจะต่างกันจากแนวคิดดังกล่าว จึงสามารถนำมาประยุกต์ใช้ในการพิจารณาแบบสอบและลักษณะที่ร่วมกันระหว่างแบบสอบย่อยได้ (Jiao, Wong และ Kamata, 2005)

2.3) **โค้งลักษณะข้อสอบ** (Item Characteristic Curve: ICC)

ทฤษฎีการตอบสนองข้อสอบ ให้ความสำคัญกับฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบโดยเป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบถูกกับระดับความสามารถของผู้สอบมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันโลจิส (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) ดังนั้น จะเห็นได้ว่าโอกาสที่ผู้สอบตอบข้อสอบถูกจึงขึ้นอยู่กับโค้งลักษณะข้อสอบ ซึ่งเป็นอิสระจาก การกระจายของความสามารถของผู้สอบหรือโอกาสที่ผู้ตอบตอบข้อสอบถูกไม่ขึ้นอยู่กับจำนวนของผู้สอบที่มีความสามารถเหมือนกัน ลักษณะของโค้งลักษณะข้อสอบในแต่ละข้อจะมีคุณสมบัติ ไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ จึงทำให้โอกาสในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยน ซึ่งไม่เหมือนกับการวิเคราะห์ด้วยทฤษฎีการทดสอบแบบดั้งเดิม (CTT)

2.4) **การทดสอบที่ไม่แข่งขันด้านเวลา** (Nonspeed Test Administration)

ทฤษฎีการตอบสนองข้อสอบมุ่งอธิบายความสัมพันธ์ระหว่างความสามารถ ที่แท้จริงของผู้สอบกับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อ ดังนั้นความสามารถที่แท้จริงของผู้สอบจึงเป็นสิ่งสำคัญสำหรับการวิเคราะห์ตามแนวทฤษฎีนี้ ซึ่งความเร็วในการตอบของผู้สอบจะต้องไม่มีอิทธิพลต่อผลการสอบ ดังนั้นการดำเนินการทดสอบต้องไม่อยู่ในสถานการณ์ของ การแข่งขันด้วยเวลาในการทำข้อสอบ โดยผู้สอบจะต้องได้รับเวลาในการทดสอบที่สมเหตุสมผล จึงทำให้สามารถวัดความสามารถที่แท้จริงของผู้สอบได้

การตรวจสอบข้อตกลงเบื้องต้นข้อนี้มีวิธีการตรวจสอบได้หลายวิธี สามารถสรุปวิธีการที่สำคัญได้ดังนี้

1. การพิจารณาจากสัดส่วนหรือร้อยละของจำนวนผู้สอบที่สามารถทำข้อสอบได้ ครบทุกข้อโดยผู้สอบส่วนใหญ่ต้องสามารถตอบข้อสอบได้ครบหรือเกือบครบทุกข้อ

2. การเปรียบเทียบความแปรปรวนของจำนวนข้อที่เว้นกับความแปรปรวนของจำนวนข้อที่ตอบผิด ถ้าอัตราส่วนของความแปรปรวนเข้าใกล้ 0 แสดงว่าการดำเนินการจัดการสอบไม่เน้น การแข่งขันด้านเวลา

1.2.3 โมเดลการตอบสนองข้อสอบ (Item Response Model)

โมเดลการตอบสนองข้อสอบเป็นโมเดลที่มุ่งอธิบายความสัมพันธ์ระหว่างความสามารถที่แท้จริงของผู้สอบกับพฤติกรรมการตอบสนองข้อสอบแต่ละข้อว่ามีโอกาสในการตอบถูกได้มากน้อยเพียงใด โดยใช้โค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ที่เรียกว่าฟังก์ชันโลจิส (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function) โดยฟังก์ชันโลจิส (Logistic function) เป็นฟังก์ชันที่มีความทนทานต่อความคลาดเคลื่อนที่เกิดขึ้นกับผู้สอบที่มีความสามารถสูง จึงทำให้เป็นที่แพร่หลายและนิยมนำไปใช้จริงมากกว่าฟังก์ชันปกติสะสม (Normal ogive function) (Hambleton และ Swaminathan, 1985; Lord, 1980 อ้างถึงใน ศิริชัย กาญจนวาสี, 2550) ในการนำเสนอครั้งนี้จะนำเสนอโมเดลการตอบสนองข้อสอบแบบทวิภาค (Dichotomous Item Response Theory) ด้วยฟังก์ชันโลจิส (Logistic function)

1) โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (1PL Model)

One–Parameter Logistic Model เป็นโมเดลที่รากฐานมาจากโมเดลราส์ชพัฒนาขึ้นโดย จอร์จ ราส์จ (George Rasch) นักคณิตศาสตร์ชาวเดนมาร์ก ในปี1960 ซึ่งมีความเชื่อว่า ในกระบวนการวัดผลเพื่อให้เกิดประสิทธิภาพที่ดีที่สุดนั้น องค์ประกอบที่สำคัญ คือ การใช้เครื่องมือในการวัดผลต้องมีความเป็นอิสระในตัวเองนั้นคือ คุณภาพของการวัดผลต้องไม่ได้ขึ้นอยู่กับสิ่งที่ต้องการวัด ดังนั้นจึงได้เกิดแนวคิดขึ้นว่า ควรจะมีการพัฒนาการวิเคราะห์ข้อสอบที่ให้ผลการวิเคราะห์เป็นอิสระ โดยราส์ชได้พยายามคิดหาค่าความยากของข้อสอบโดยไม่ต้องไปสัมพันธ์กับผู้สอบและหาค่าความสามารถของผู้สอบโดยไม่ต้องไปสัมพันธ์กับระดับความยาก-ง่ายของข้อสอบนั้นก็คือเครื่องมือวัดและสิ่งที่ถูกวัดเป็นอิสระต่อกันและกัน

ฟังก์ชันของโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์นี้เป็นโมเดลที่มีค่าพารามิเตอร์ของข้อสอบเพียงค่าเดียว คือ ค่าความยากของข้อสอบ (β_i) โดยมีข้อตกลงเบื้องต้นว่า ข้อสอบทุกข้อมีค่าอำนาจจำแนกของข้อสอบ (α_i) ทุกข้อเท่ากันหมดและไม่มีโอกาสในการเดาข้อสอบถูก $(c_i = 0)$ สามารถแสดงสูตรทางคณิตศาสตร์ของโมเดลในสถานการณ์การทดสอบที่มีข้อคำถามแบบสองค่า (dichotomous items) ดังนี้

$$P(\theta_p, \beta_i) = P(X_{pi} = 1 | \theta_p, \beta_i) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}$$

เมื่อ $P(\theta_p, \beta_i)$ แทนความน่าจะเป็นซึ่งผู้ตอบที่มีระดับความสามารถเท่ากับθ ตอบข้อที่มีค่าความยากข้อสอบเท่ากับ β ได้ถูกต้อง

X_{pi} แทนคะแนนจากการทดสอบของผู้สอบคนที่ p ข้อสอบข้อที่ i

β_i แทนค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

θ_p แทนค่าพารามิเตอร์ระดับความสามารถของผู้สอบคนที่ p

2) โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (2PL Model)

ฟังก์ชันของโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์นี้เป็นโมเดลที่มีค่าพารามิเตอร์ของข้อสอบสองค่า คือ ค่าความยากของข้อสอบ (β_i) และค่าอำนาจจำแนกของข้อสอบ (α_i) โดยมีข้อตกลงเบื้องต้นว่า ข้อสอบทุกข้อไม่มีโอกาสในการเดาข้อสอบถูก $(c_i = 0)$ สูตรทางคณิตศาสตร์ของโมเดลในสถานการณ์การทดสอบที่มีข้อคำถามแบบสองค่า (dichotomous items) ดังนี้

$$\mathrm{P}(\theta_p, \alpha_i, \beta_i) = \mathrm{P}(X_{pi} = 1 | \theta_p, \alpha_i, \beta_i) = \frac{\exp[D\alpha_i(\theta_p - \beta_i)]}{1 + \exp[(D\alpha_i)(\theta_p - \beta_i)]}$$

เมื่อ $P(\theta_p, \beta_i)$ แทนความน่าจะเป็นซึ่งผู้ตอบที่มีระดับความสามารถเท่ากับθ ตอบข้อที่มีค่าความยากข้อสอบเท่ากับ β ได้ถูกต้อง

X_{pi} แทนคะแนนจากการทดสอบของผู้สอบคนที่ p ข้อสอบข้อที่ i

α_i แทนค่าพารามิเตอร์ค่าอำนาจจำแนกของข้อสอบข้อที่ i

β_i แทนค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

θ_p แทนค่าพารามิเตอร์ระดับความสามารถของผู้สอบคนที่ p

D เป็นค่าคงที่ซึ่งมีค่าเท่ากับ 1.70

3) โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (3PL Model)

ฟังก์ชันของโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์นี้เป็นโมเดลที่มีค่าพารามิเตอร์ของข้อสอบสามค่า คือ ค่าความยากของข้อสอบ (β_i) ค่าอำนาจจำแนกของข้อสอบ (α_i) และค่าโอกาสในการเดาข้อสอบถูก (c_i) หรือจุดต่ำที่สุดของโค้งการตอบข้อสอบ ซึ่งหมายถึงความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้สอบที่มีระดับความสามารถต่ำสุดสูตรทางคณิตศาสตร์ของโมเดลในสถานการณ์การทดสอบที่มีข้อคำถามแบบสองค่า (dichotomous items) ดังนี้

$$P(\theta_p,\alpha_i,\beta_i,c_i) = P(X_{pi}=1\,|\,\theta_p,\alpha_i,\beta_i,c_i) = c_i + \frac{\exp[D\alpha_i(\theta_p-\beta_i)][1-c_i]}{1+\exp[(D\alpha_i)(\theta_p-\beta_i)]}$$

เมื่อ $P(\theta_p,\beta_i)$ แทนความน่าจะเป็นซึ่งผู้ตอบที่มีระดับความสามารถเท่ากับθ ตอบข้อที่มีค่าความยากข้อสอบเท่ากับ β ได้ถูกต้อง

X_{pi} แทนคะแนนจากการทดสอบของผู้สอบคนที่ p ข้อสอบข้อที่ i

α_i แทนค่าพารามิเตอร์ค่าอำนาจจำแนกของข้อสอบข้อที่ i

β_i แทนค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i

c_i แทนค่าการเดาข้อสอบข้อที่ i ได้ถูกต้อง

θ_p แทนค่าพารามิเตอร์ระดับความสามารถของผู้สอบคนที่ p

D เป็นค่าคงที่ซึ่งมีค่าเท่ากับ 1.70

1.2.4 การวัดคุณภาพการทดสอบตามทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีที่มุ่งอธิบายความสัมพันธ์ระหว่างคุณลักษณะของผู้สอบกับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกได้มากน้อยเพียงใด ดังนั้นในการวิเคราะห์ข้อสอบจึงให้ทั้งค่าพารามิเตอร์ของผู้สอบและค่าพารามิเตอร์ข้อสอบในการวัดคุณภาพของการวิเคราะห์แบบสอบจึงมีทั้งดัชนีที่วิเคราะห์คุณภาพของข้อรายข้อและวิเคราะห์คุณภาพของแบบสอบทั้งฉบับ โดยมีดัชนีที่สำคัญสำหรับการวิเคราะห์คุณภาพ ดังนี้

1) **ฟังก์ชันสารสนเทศของข้อสอบ** (Item Information)

ฟังก์ชันสารสนเทศของข้อสอบ เป็นค่าที่บ่งบอกถึงความถูกต้องในการประมาณค่าความสามารถที่แท้จริงของผู้สอบในการตอบข้อสอบในแต่ละข้อ ซึ่งเป็นค่าที่สร้างขึ้นมาจากค่าคุณลักษณะของข้อสอบหลายตัว ประกอบด้วยค่าพารามิเตอร์ความยาก ค่าพารามิเตอร์อำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อ โดยสามารถคำนวณได้จากสูตรดังต่อไปนี้

$$I_i(\theta) = \frac{\left[P_i'(\theta)\right]^2}{P_i(\theta)Q_i(\theta)}$$

เมื่อ $I_i(\theta)$ = ค่าฟังก์ชันสารสนเทศที่ได้รับจากข้อสอบข้อที่ i สำหรับผู้สอบที่มีระดับความสามารถเท่ากับ θ

$P_i'(\theta)$ = ความชันของโค้งลักษณะของข้อสอบข้อที่ i ที่ระดับความสามารถเท่ากับ θ

$P_i(\theta)$ = ความน่าจะเป็นที่ผู้สอบที่มีความสามารถเท่ากับ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$Q_i(\theta)$ = 1- $P_i(\theta)$

เมื่อพิจารณาสูตรการคำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ พบว่า ค่าของฟังก์ชันนี้ขึ้นอยู่กับความชันของโค้งลักษณะข้อสอบ ถ้าโค้งลักษณะข้อสอบมีความชัน ($P_i'(\theta)$) ที่มีค่ามากขึ้นในขณะที่ความแปรปรวนของการตอบข้อสอบ ($P_i(\theta)\,Q_i(\theta)$) มีค่าน้อยลง จะทำให้โค้งสารสนเทศของข้อสอบที่ระดับความสามารถนั้นๆ มีค่ามากขึ้น ความสูงของโค้งสารสนเทศข้อสอบที่สูงที่สุดตรงกับความสามารถในระดับใด แสดงว่า ข้อสอบข้อนั้นจะให้ค่าสารสนเทศสูงสุด ณ ระดับความสามารถ ดังนั้นจึงทำให้สามารถเลือกข้อสอบที่เหมาะสมกับระดับความสามารถของผู้สอบได้อย่างถูกต้อง

โดยทั่วไป ค่าฟังก์ชันสารสนเทศของข้อสอบจะมีค่าสูงก็ต่อเมื่อ 1) ผู้สอบมีค่าระดับความสามารถ (θ) ใกล้เคียงกับค่าพารามิเตอร์ความยากของข้อสอบ (b) 2) ค่าพารามิเตอร์อำนาจจำแนก (a) มีค่าสูงขึ้น และ 3) ค่าพารามิเตอร์โอกาสในการเดาข้อสอบถูก (c) มีค่าเข้าใกล้ศูนย์ (ศิริชัย กาญจนวาสี, 2550)

2) **ฟังก์ชันสารสนเทศของแบบสอบ** (Test Information)

ค่าฟังก์ชันสารสนเทศของแบบสอบ เป็นดัชนีที่แสดงถึงความถูกต้องแม่นยำในการประมาณค่าความสามารถที่แท้จริงของผู้สอบ (θ) ของแบบสอบทั้งฉบับ โดยฟังก์ชันนี้เป็นผลรวมเชิงพีชคณิตของฟังก์ชันสารเทศของข้อสอบแต่ละข้อรวมเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกัน โดยสามารถคำนวณได้จากสูตรดังต่อไปนี้

$$I\ (\theta)=\sum_{i=1}^{k} I_i(\theta)=\frac{\sum\left[P_i^{'}(\theta)\right]^2}{P_i(\theta)Q_i(\theta)}$$

เมื่อ $I\ (\theta)$ = ค่าฟังก์ชันสารสนเทศที่ได้รับจากแบบสอบสำหรับผู้ที่มีระดับความสามารถเท่ากับ θ

$I_i(\theta)$ = ค่าฟังก์ชันสารสนเทศที่ได้รับจากข้อสอบข้อที่ i สำหรับผู้สอบที่มีระดับความสามารถเท่ากับ θ

เมื่อพิจารณาสูตรการคำนวณค่าฟังก์ชันสารสนเทศของแบบสอบ พบว่า ค่าของฟังก์ชันนี้ขึ้นอยู่กับค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อที่เป็นอิสระต่อกัน ดังนั้นถ้าต้องการให้แบบสอบมีค่าสารสนเทศสูงจึงต้องให้ออกแบบข้อสอบในแต่ละข้อมีค่าสารสนเทศที่สูง ไม่เหมือนกับทฤษฎีการทดสอบแบบดั้งเดิมที่ค่าความเที่ยงของแบบสอบทั้งฉบับขึ้นอยู่กับ ค่าความยากและค่าอำนาจจำแนกของข้อสอบในแต่ละข้อที่ไม่มีความเป็นอิสระต่อกัน ดังนั้นค่าความเที่ยงของแบบสอบทั้งฉบับที่ได้จากการวิเคราะห์นี้จึงขึ้นอยู่กับลักษณะสถานการณ์เฉพาะที่นำแบบสอบมาใช้เท่านั้น

3) **ความคลาดเคลื่อนมาตรฐานของการประมาณค่า** (Standard Error of Estimation; SE(θ))

ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเป็นค่าที่แสดงถึงความคลาดเคลื่อนที่เกิดขึ้นจากการวัด ซึ่งเป็นค่าส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็นของการประมาณค่าความสามารถที่แท้จริง(θ) มีค่าเป็นสัดส่วนผกผันกับค่าฟังก์ชันสารสนเทศของข้อสอบ โดยสามารถคำนวณได้จากสูตรดังต่อไปนี้

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

เมื่อ $SE(\theta)$ = ความคลาดเคลื่อนมาตรฐานของการประมาณค่า

สำหรับผู้สอบที่มีระดับความสามารถเท่ากับ θ

$I\ (\theta)$ = ค่าฟังก์ชันสารสนเทศที่ได้รับจากแบบสอบ

สำหรับผู้ที่มีระดับความสามารถเท่ากับ θ

เมื่อพิจารณาสูตรการคำนวณค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า พบว่า ค่าของฟังก์ชันนี้เป็นส่วนกลับของค่าฟังก์ชันสารสนเทศของแบบสอบ โดยถ้าแบบสอบใดมีค่าฟังก์ชันสารสนเทศของแบบสอบสูง แบบสอบนั้นก็จะมีค่า ความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำหรือมีความแม่นยำในการประมาณค่าความสามารถสูง ณ ตำแหน่ง θ นั้น

ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเป็นค่าที่มีความหมายคล้ายคลึงกับค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement; SEM) ในทฤษฎีการทดสอบแบบดั้งเดิมเป็นค่าที่แสดงถึงความคลาดเคลื่อนในการวัดหรือในการประมาณค่าพารามิเตอร์ของการวัด แต่มีความแตกต่างกันที่ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า ($SE(\theta)$) มีความผันแปรไปตามระดับความสามารถที่แท้จริงของผู้สอบ (θ) ในขณะที่ค่าความคลาดเคลื่อนมาตรฐานในการวัด (SEM) เป็นค่าคงที่ของแบบสอบสำหรับผู้สอบทุกคน

4) **ค่าประสิทธิภาพสัมพันธ์** (Relative Efficience; RE(θ))

ค่าประสิทธิภาพสัมพันธ์ เป็นค่าที่แสดงการเปรียบเทียบค่าฟังก์ชันสารสนเทศระหว่างแบบสอบที่วัดคุณลักษณะเดียวกัน ซึ่งคำนวณได้จากอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบสอบต่างฉบับ ณ ตำแหน่งระดับความสามารถของผู้สอบ (θ) เดียวกัน สามารถแสดงได้ด้วยสูตรทางคณิตศาสตร์ดังนี้

$$RE(\theta) = \frac{I_A(\theta)}{I_B(\theta)}$$

เมื่อ $I_A(\theta)$ = ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับ A ณ ตำแหน่งบนสเกล

ความสามารถร่วมกัน ณ ระดับ θ

$I_B(\theta)$ = ค่าฟังก์ชันสารสนเทศของแบบสอบฉบับ B ณ ตำแหน่งบนสเกล

ความสามารถร่วมกัน ณ ระดับ θ

ในการแปลความหมายค่าประสิทธิภาพสัมพันธ์ สามารถพิจารณาจากเกณฑ์ดังนี้

RE(θ) = 1 แสดงว่าแบบสอบทั้งสองฉบับมีประสิทธิภาพเท่ากัน สำหรับผู้ตอบที่มีระดับความสามารถระดับ θ

RE(θ) > 1 แสดงว่าแบบสอบฉบับ A มีประสิทธิภาพสูงกว่าแบบสอบฉบับ B สำหรับผู้ตอบที่มีระดับความสามารถระดับ θ

RE(θ) < 1 แสดงว่าแบบสอบฉบับ A มีประสิทธิภาพต่ำกว่าแบบสอบฉบับ B สำหรับผู้ตอบที่มีระดับความสามารถระดับ θ

นอกจากนี้ในการเปรียบเทียบประสิทธิภาพระหว่างวิธีการประมาณค่าพารามิเตอร์ด้วยแบบสอบฉบับเดียวกัน สามารถคำนวณได้จากการหาอัตราส่วนระหว่างค่าความแปรปรวนของวิธีประมาณค่าพารามิเตอร์ที่หนึ่งกับค่าความแปรปรวนของวิธีประมาณค่าพารามิเตอร์วิธีที่สองนั้น คือ V (M_1) / V (M_2) แต่จากดัชนีกำหนดคุณภาพของแบบทดสอบ พบว่า ความแปรปรวนที่ได้จากการประมาณค่าพารามิเตอร์เป็นสัดส่วนกลับของค่าฟังก์ชันสารสนเทศของแบบสอบ (Lord, 1980) ดังนั้นเราจึงสามารถสรุปสูตรการเปรียบเทียบประสิทธิภาพระหว่างวิธีการประมาณค่า 2 วิธีด้วยแบบสอบฉบับเดียวกันได้ว่า

$$\frac{V(M_1)}{V(M_2)} = \frac{I(\theta, M_2)}{I(\theta, M_1)}$$

จากสมการข้างต้นทำให้สามารถเปรียบเทียบประสิทธิภาพของวิธีประมาณค่าพารามิเตอร์ 2 วิธีใดๆ ด้วยการวิเคราะห์ค่าฟังก์ชันสารสนเทศฉบับเดียวกันจากการวิเคราะห์ด้วยวิธีการประมาณค่าพารามิเตอร์ 2 วิธีแทน นอกจากนี้ยังพบอีกว่า สมการคำนวณวิธีการประมาณค่าพารามิเตอร์ X เทียบกับวิธีประมาณค่าพารามิเตอร์ Y ในแบบทดสอบฉบับหนึ่ง จะมีค่าเท่ากับค่าประสิทธิภาพสัมพันธ์ของแบบสอบฉบับนั้น (RE (Y,X)) ที่เป็นผลมาจากวิธีประมาณค่าพารามิเตอร์ด้วยวิธี Y เมื่อเทียบกับวิธี X ณ ระดับความสามารถใด ๆ โดยสามารถสรุปเป็นสมการทางคณิตศาสตร์ได้ดังนี้

$$EFF(X,Y) = \frac{V(\theta, X)}{V(\theta, Y)} = \frac{I(\theta, Y)}{I(\theta, X)} = RE(Y, X)$$

ในการแปลความหมายสามารถพิจารณาจากเกณฑ์ ดังนี้

$EFF(X,Y) = 1$ แสดงว่าการประมาณค่าด้วยวิธี Y มีประสิทธิภาพเท่ากับการประมาณค่าด้วยวิธี X เนื่องจากการประมาณค่าพารามิเตอร์ด้วยวิธี X มีความแปรปรวนเท่ากับวิธี Y

$EFF(X,Y) > 1$ แสดงว่า การประมาณค่าด้วยวิธี Y มีประสิทธิภาพสูงกว่าการประมาณค่าด้วยวิธี X เนื่องจากการประมาณค่าพารามิเตอร์ด้วยวิธี X มีความแปรปรวนสูงกว่าวิธี Y

$EFF(X,Y) < 1$ แสดงว่า การประมาณค่าด้วยวิธี Y มีประสิทธิภาพต่ำกว่าการประมาณค่าด้วยวิธี X เนื่องจากการ) ประมาณค่าพารามิเตอร์ด้วยวิธี X มีความแปรปรวนต่ำกว่าวิธี Y

5) **ค่าประสิทธิภาพสัมพันธ์เฉลี่ย** (RAI)

ค่าประสิทธิภาพสัมพันธ์เฉลี่ย เป็นค่าที่แสดงการเปรียบเทียบค่าฟังก์ชันสารสนเทศเฉลี่ยระหว่างแบบสอบที่วัดคุณลักษณะเดียวกัน ซึ่งคำนวณได้จากอัตราส่วนเฉลี่ยของแบบสอบต่างฉบับ ณ ทุกตำแหน่งของระดับความสามารถของผู้สอบ (θ) สามารถแสดงได้ด้วยสูตรทางคณิตศาสตร์ดังนี้

$$RAI(\theta; X,Y) = \frac{AI(\theta,X)}{AI(\theta,Y)}$$

เมื่อ $AI(\theta,X)$ = ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบสอบฉบับ X ณ ทุกตำแหน่งระดับความสามารถ (θ)

$AI(\theta,Y)$ = ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบสอบฉบับ Y ณ ทุกตำแหน่งระดับความสามารถ (θ)

ในการแปลความหมายค่าประสิทธิภาพสัมพันธ์ สามารถพิจารณาจากเกณฑ์ดังนี้

$RAI(\theta; X,Y) = 1$ แสดงว่าแบบสอบทั้งสองฉบับมีประสิทธิภาพเฉลี่ยเท่ากันสำหรับผู้ตอบทุกระดับความสามารถระดับ θ

$RAI(\theta; X,Y) > 1$ แสดงว่าแบบสอบฉบับ A มีประสิทธิภาพเฉลี่ยสูงกว่าแบบสอบฉบับ B สำหรับผู้ตอบทุกระดับความสามารถระดับ θ

$RAI(\theta; X,Y) < 1$ แสดงว่าแบบสอบฉบับ A มีประสิทธิภาพเฉลี่ยต่ำกว่าแบบสอบฉบับ B สำหรับผู้ตอบทุกระดับความสามารถระดับ θ

1.2.5) จุดเด่นของการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ

การวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบให้ผลการวิเคราะห์ทั้งค่าพารามิเตอร์ของผู้สอบ ซึ่งได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบซึ่งได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าโอกาสในการเดาคำตอบที่ถูกต้อง (c) การวิเคราะห์ด้วยทฤษฎีนี้มีจุดเด่นหลายประการสามารถสรุปเป็นประเด็นสำคัญได้ดังนี้

1) สารสนเทศของการวิเคราะห์

การวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบสามารถวิเคราะห์ค่าพารามิเตอร์ของผู้สอบเป็นรายบุคคลโดยการประมาณค่าความสามารถที่แท้จริงของผู้สอบ (θ) นอกจากนี้ยังสามารถวิเคราะห์ค่าพารามิเตอร์ของการสอบเป็นรายข้อได้อีกด้วยตามโมเดลการทดสอบ โดยโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (1 PL) จะเป็นโมเดลที่มีค่าพารามิเตอร์ของข้อสอบเพียงค่าเดียว คือ ค่าความยากของข้อสอบ (b_i) ถ้าโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (2 PL) จะเป็นโมเดลที่มีค่าพารามิเตอร์ของข้อสอบสองค่า คือ ค่าความยากของข้อสอบ (b_i) และค่าอำนาจจำแนก (a) และถ้าโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (3 PL) จะเป็นโมเดลที่มีค่าพารามิเตอร์ของข้อสอบสามค่า คือ ค่าความยากของข้อสอบ (b_i) และค่าอำนาจจำแนก (a) และค่าโอกาสในการเดาคำตอบที่ถูกต้อง (c)

2) ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariance)

ผลการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบมีจุดเด่นที่สำคัญ เมื่อเปรียบเทียบกับทฤษฎีการทดสอบแบบดั้งเดิม (CTT) คือ ค่าพารามิเตอร์ของผู้สอบและค่าพารามิเตอร์ของข้อสอบมีความคงที่ไม่แปรเปลี่ยนของค่าพารามิเตอร์ โดยค่าพารามิเตอร์ของข้อสอบจะเป็นอิสระจากกลุ่มผู้สอบที่ใช้ในการประมาณค่านั้นคือ ทั้งค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าโอกาสในการเดาคำตอบที่ถูกต้อง (c) จะมีค่าไม่เปลี่ยนแปลงไปตามกลุ่มผู้สอบไม่ว่าจะนำแบบสอบไปใช้กับผู้สอบกลุ่มใดก็ตาม ค่าพารามิเตอร์ของข้อสอบก็จะไม่เปลี่ยนแปลง สำหรับค่าพารามิเตอร์ของผู้สอบก็เช่นเดียวกันจะเป็นอิสระจากข้อสอบที่ใช้ในการประมาณค่า โดยจะไม่เปลี่ยนแปลงไปตามค่าพารามิเตอร์ของข้อสอบ ดังนั้นจึงทำให้สามารถคำนวณค่าความสามารถที่แท้จริงของผู้สอบได้โดยไม่ขึ้นกับสถานการณ์การทดสอบและกลุ่มของผู้สอบ

3) ความเป็นอิสระของข้อสอบ (Item-Free)

การเปรียบเทียบความสามารถของผู้สอบจะไม่ขึ้นอยู่กับข้อคำถามในแบบสอบ โดยในการประมาณค่าความสามารถของผู้สอบจะใช้ชุดข้อคำถามที่ต่างกันก็สามารถนำค่าความสามารถ มาเปรียบเทียบกันได้ ทั้งนี้ข้อสอบที่นำมาใช้ต้องได้รับการคัดเลือกจากคลังข้อสอบขนาดใหญ่ ที่ข้อสอบแต่ละข้อได้ระบุคุณลักษณะที่วัดเอาไว้ เนื่องจากค่าความสามารถที่แท้จริงที่ประมาณค่าได้เป็นคะแนนโลจิท (Logit) ซึ่งอยู่ในมาตรวัดเดียวกัน

4) การวัดคุณภาพตามแนวทฤษฎีการตอบสนองข้อสอบ

คุณภาพของการวัดตามแนวทฤษฎีการตอบสนองข้อสอบนั้น สามารถรายงานได้ทั้งรายข้อและทั้งฉบับ เนื่องจากสามารถวิเคราะห์คุณภาพของข้อสอบได้ทั้ง ค่าฟังก์ชันสารสนเทศของข้อสอบและค่าฟังก์ชันสารสนเทศของแบบสอบ ซึ่งค่าทั้งสองเป็นดัชนีที่แสดงถึงความถูกต้องแม่นยำในการประมาณค่าความสามารถ โดยสามารถนำมาใช้แปลความหมายได้เหมือนกันค่าความเที่ยงและค่าความคลาดเคลื่อนมาตรฐานตามแนวทฤษฎีการทดสอบแบบดั้งเดิมได้

1.2.6 ข้อจำกัดของการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ

1) ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีที่มุ่งอธิบายความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่แท้จริงของบุคคลกับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกได้มากน้อยเพียงใด ดังนั้นจะเห็นได้ว่าทฤษฎีนี้มุ่งที่จะให้สารสนเทศของข้อสอบเป็นข้อๆ (Fixed) จึงไม่สามารถสรุปอ้างอิงผลการวัดไปใช้ได้ทั้งเอกภพของการทดสอบเหมือนกับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

2) เมื่อพิจารณาคะแนนจริง (true score) ที่ได้จากการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) พบว่า มีเพียงการวิเคราะห์ด้วยโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (3PL) เท่านั้นที่เป็นคะแนนจริงตามที่คาดหวังตามทฤษฎีเท่านั้น ทั้งนี้เนื่องจากการวิเคราะห์ด้วยโมเดล 3 PL เป็นโมเดลที่ให้ ค่าโอกาสในการเดาคำตอบที่ถูกต้อง (c) ซึ่งเป็นค่าที่สามารถบอกได้ว่า ถ้าผู้มีความสามารถต่ำจะมีโอกาสในการตอบข้อสอบข้อนั้น ๆ ถูกเท่าใด ส่วนโมเดลอื่นๆ ของทฤษฎีการตอบสนองข้อสอบไม่ได้กล่าวถึงในประเด็นดังกล่าวจึงไม่สามารถวิเคราะห์คะแนนจริงได้จากการทดสอบ

3) เมื่อพิจารณาค่าความเที่ยงของการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) เปรียบเทียบกับทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) สำหรับในแต่ละคน และแต่ละข้อ สามารถคำนวณค่า marginal reliability ด้วยสูตรดังนี้ (Adams, 2006; Mislevy, Beaton, Kaplan และ Sheehan, 1992)

$$R_P = 1 - \frac{\overline{\sigma_p}^2}{\mathrm{var}(\theta)}$$

$$R_i = 1 - \frac{\overline{\sigma_i}^2}{\mathrm{var}(\beta)}$$

เมื่อพิจารณาค่าความเที่ยงด้วยการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบพบว่า มีข้อจำกัดอยู่ 2 ประการของการพิจารณาค่า marginal reliability ประการแรก การวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) ไม่ได้คำนึงถึงขนาดของกลุ่มตัวอย่าง (sample size) ทำให้ไม่สามารถอธิบายได้ด้วยสูตรของ Spearman-Brown ซึ่งใน GT ยังสามารถพิจารณาค่า marginal reliability เมื่อมีจำนวนคนและจำนวนข้อสอบเปลี่ยนแปลงไป ประการที่สอง เมื่อมีการวัดแบบหลายฟาเซต (multifaceted) จากสูตรการคำนวณไม่สามารถอธิบายได้ว่าค่าความเที่ยงว่ามีการเปลี่ยนแปลงไปในทิศทางอย่างไรในแต่ละฟาเซต

1.2.7 สรุปสาระสำคัญของทฤษฎีการตอบสนองข้อสอบ (IRT)

ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่แท้จริงของบุคคลกับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกได้มากน้อยเพียงใด โดยใช้โค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve : ICC) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันโลจิส (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function)

ผลการวิเคราะห์ของทฤษฎีนี้สามารถประมาณค่าพารามิเตอร์ได้ทั้งค่าพารามิเตอร์ของผู้สอบเป็นรายบุคคลโดยการประมาณค่าความสามารถที่แท้จริง (θ) ของผู้สอบ และยังสามารถวิเคราะห์ค่าพารามิเตอร์ของการสอบเป็นรายข้อได้อีกด้วยตามโมเดลการทดสอบ โดยโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (1 PL) จะเป็นโมเดลที่มีค่าพารามิเตอร์ของข้อสอบเพียงค่าเดียว คือ ค่าความยากของข้อสอบ (b_i) ถ้าโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ (2 PL) จะเป็นโมเดลที่มีค่าพารามิเตอร์ของข้อสอบสองค่า คือ ค่าความยากของข้อสอบ (b_i) และค่าอำนาจจำแนก (a)

และถ้าโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (3 PL) จะเป็นโมเดลที่มีค่าพารามิเตอร์ของข้อสอบสามค่า คือ ค่าความยากของข้อสอบ (b_i) และค่าอำนาจจำแนก (a) และค่าโอกาสในการเดาคำตอบที่ถูกต้อง (c) ดังนั้นจะเห็นได้ว่าทฤษฎีนี้มีข้อดีที่สำคัญ คือ สามารถคำนวณค่าพารามิเตอร์ของข้อสอบได้ทั้งรายข้อโดยจะมีค่าไม่เปลี่ยนแปลงไปตามกลุ่มผู้สอบไม่ว่าจะนำแบบสอบไปใช้กับผู้สอบกลุ่มใดก็ตาม และยังสามารถคำนวณค่าพารามิเตอร์ของผู้สอบโดยจะไม่เปลี่ยนแปลงไปตามค่าพารามิเตอร์ของข้อสอบ เหมือนกับทฤษฎีการทดสอบแบบดั้งเดิม (CTT)

สำหรับการวัดคุณภาพของทฤษฎีนี้สามารถพิจารณาได้จากดัชนี 3 ตัว ได้แก่ ค่าฟังก์ชันสารสนเทศของข้อสอบ (Item Information) ค่าฟังก์ชันสารสนเทศของแบบสอบ (Test Information) และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation; SE(θ)) โดยค่าฟังก์ชันสารสนเทศของแบบสอบมีลักษณะคล้ายคลึงกับค่าความเที่ยง และค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า มีลักษณะคล้ายคลึงกับค่าความคลาดเคลื่อนมาตรฐาน (SEM) ตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT)

1.3 **วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ** (Generalizability in Item Response Modeling:GIRM)

วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (Generalizability in Item Response Modeling: GIRM) เป็นการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการกำหนดมาตรวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) โดยอาศัยสมมติฐานของการกระจายไปยังแหล่งความคลาดเคลื่อนจากการวัดที่เกี่ยวข้อง และทำการประมาณค่า GT ในแต่ละองค์ประกอบของความแปรปรวน (Variance Components) ไปพร้อมกับการประมาณค่าพารามิเตอร์ ในโมเดลการตอบสนองข้อสอบ (IRT)

การวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เป็นการวิเคราะห์ที่ให้สารสนเทศเกี่ยวกับปัจจัยที่ส่งผลกระทบต่อความคลาดเคลื่อนในการวัด ซึ่งจะทำให้สามารถออกแบบเหตุการณ์ได้เหมาะสมกับความต้องการในแต่ละสถานการณ์และทำให้การออกแบบการวัดนั้นเกิดประสิทธิภาพสูงที่สุดในแต่ละสถานการณ์ได้

ส่วนการวิเคราะห์ด้วยโมเดลการตอบสนองข้อสอบ (IRT) เป็นการวิเคราะห์ที่ช่วยเสริมการวิเคราะห์จากทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เนื่องจากการวิเคราะห์นี้สามารถให้ข้อมูลเกี่ยวกับคุณภาพของแหล่งความเคลื่อนที่เราสนใจ (Quality of facet conditions)

ดังนั้นจะเห็นได้ว่าการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นการวิเคราะห์ที่ให้ข้อมูลที่ครบถ้วนและครอบคลุมเนื่องจากเป็นการรวมโมเดลการวัดทั้งสองเข้าไว้ด้วยกันในโมเดลการวัดเพียงโมเดลเดียว

1.3.1 **พัฒนาการวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM)**

วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นวิธีการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการกำหนดมาตรวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) อาศัยสมมติฐานของการกระจายไปยังแหล่งความคลาดเคลื่อนจากการวัดที่เกี่ยวข้อง โดยการประมาณค่า GT ในแต่ละองค์ประกอบของความแปรปรวน (Variance Components) ไปพร้อมกับการประมาณค่าพารามิเตอร์ ในโมเดลการตอบสนองข้อสอบ (IRT)

เมื่อพิจารณาโมเดลของการวัดทั้งสองแล้ว พบว่า ทั้งโมเดลตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และโมเดลตามทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นโมเดลที่มีมโนทัศน์ตามโมเดลสถิติที่เรียกว่าการวิเคราะห์พหุระดับ (Multilevel statistical models) (Goldstein, 1995; Verhelst และ Verstralen, 2001; Patz และคณะ, 2002) ดังนั้นในการรวมทั้งสองโมเดลของการวัดเข้าด้วยกันเบื้องต้นต้องทำการผ่อนคลายข้อตกลงเบื้องต้นการไม่สรุปอ้างอิงข้อคำถาม (fixed items) ของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ให้เป็นการวิเคราะห์ที่สามารถสรุปอ้างอิงข้อคำถาม (random items) ได้จากการวิเคราะห์ข้อสอบ เพื่อให้สอดคล้องกับการวิเคราะห์ตามแนวของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ซึ่งทำให้การวิเคราะห์ของทั้งสองโมเดลกลายเป็นการวิเคราะห์ที่มีมโนทัศน์แบบโมเดลการวิเคราะห์ผลกระทบสุ่มแบบพหุระดับ (Multilevel random effects models) (Briggs และ Wilson, 2007)

การเชื่อมโยงระหว่างทฤษฎีการวัดทั้งสองนั้นเริ่มแรกจากการคิดค้นการเชื่อมโยงสูตรของ Kolen และ Harris (1987) ในการวิเคราะห์องค์ประกอบความแปรปรวนและค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัดโดยการประมาณได้จากการวิเคราะห์ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบหลายตัวแปร (Multivariate GT) ด้วยกรอบแนวคิดของโมเดลการตอบสนองข้อสอบ (IRT) ในโมเดลของการวิเคราะห์ข้อสอบ ซึ่งทำให้การวิเคราะห์ข้อสอบด้วยวิธีการนี้สามารถผ่อนคลายข้อตกลงเบื้องต้นการไม่สรุปอ้างอิงข้อคำถาม (fixed items) ของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) โดยสามารถวิเคราะห์เป็นตัวอย่างของข้อสอบจากเอกภพของข้อสอบ (universe of item) ได้

นอกจากนี้ Mislevy (1993) ยังเสนออีกวิธีการหนึ่งในการผ่อนคลายข้อตกลงเบื้องต้นการไม่สรุปอ้างอิงข้อคำถาม (fixed items) ของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) โดยสามารถวิเคราะห์ข้อสอบตัวอย่างจากข้อสอบจากเอกภพของข้อสอบ (universe of item) ได้ด้วยการใช้วิธีการประมาณค่าของเบส์ (Bayesian) โดยการสุ่ม (randomness) ข้อสอบเข้าไปในการวิเคราะห์ตามแนวของทฤษฎีการตอบสนองข้อสอบ (IRT) ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบจึงทำให้สามารถผ่อนคลายการไม่สรุปอ้างอิงข้อคำถาม (fixed items) ของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT)

ในการประยุกต์ใช้การเชื่อมโยงการวิเคราะห์ข้อสอบของทฤษฎีทั้งสองในทางปฏิบัติตามสถานการณ์การออกแบบการวัด พบว่า ได้มีการออกแบบการวัดโดยการเชื่อมโยงในสถานการณ์การวิเคราะห์ข้อสอบเพียงรูปแบบเดียวจนถึงปัจจุบัน (ปี ค.ศ. 2008) คือการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) โดยมีผู้ประเมินคะแนนการปฏิบัติงานหลายคนมีผลงานวิจัยจำนวน 3 เรื่อง เรื่องแรกเป็นการวิเคราะห์ข้อสอบโดยใช้โมเดลการตอบสนองข้อสอบสำหรับการประเมินผลการปฏิบัติงานด้วยผู้ตรวจหลายคนของ Verhelst และ Verstralen (2001) โดยสามารถเขียนในรูปของการออกแบบการวัดแบบทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ได้เป็น two crossed measurement facets (โดยมีแหล่งความคลาดเคลื่อนทั้งสองแหล่ง ได้แก่ ข้อสอบและผู้ประเมิน) ซึ่งคล้ายคลึงกับงานวิจัยเรื่องที่สองของ Patz และคณะ (2002) ที่ประเมินการปฏิบัติงานด้วยผู้ตรวจหลายคนโดยใช้ Hierarchical rater models (HRM) สำหรับคะแนนที่มีการตรวจให้คะแนนหลายค่า (Polytomous item response data score) ด้วยผู้ประเมินหลายคนและเรื่องที่สามของ Bock, Brennan และ Muraki (2002) ได้ศึกษาวิธีการรวมแหล่งความคลาดเคลื่อนหลายแหล่งเข้าไว้ในการวิเคราะห์ตามแนวทฤษฎีการตอบสนองข้อสอบ (IRT) สำหรับการทดสอบความคงเส้นคงวาในกรณีที่มีผู้ตรวจข้อสอบหลายคนโดยใช้การวิเคราะห์การสรุปอ้างอิง (Generalizability analysis) ในการประมาณค่า ซึ่งจะทำให้ได้ผลการวิเคราะห์ด้วยฟังก์ชันการตอบสนองข้อสอบและค่าความคลาดเคลื่อนมาตรฐานที่ถูกต้องสำหรับผู้ตรวจข้อสอบหลายคนที่มีผู้ตรวจซ้อนอยู่ในคะแนนรายข้อ (raters are nested within items) ดังนั้นจะเห็นได้ว่าการออกแบบการวัดในการศึกษาของ 3 เรื่องนี้ได้นำไปสู่ การเชื่อมโยงทฤษฎีการวัดสองทฤษฎีระหว่าง ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) กับทฤษฎีการตอบสนองข้อสอบ (IRT) และยังเป็นการขยายกรอบแนวคิดที่เด่นชัดเกี่ยวกับโมเดลผลกระทบแบบสุ่มพหุระดับ (Hierarchical random effect models) อีกด้วย

ต่อมาในปี 2003 ได้มีการประยุกต์ใช้โปรแกรมคอมพิวเตอร์ตามแนวคิดของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบเพื่อผ่อนคลายข้อตกลงเบื้องต้นการไม่สรุปอ้างอิงข้อคำถาม (fixed items) ของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ให้เป็นการวิเคราะห์ที่สามารถสรุปอ้างอิงข้อคำถาม (random items) ได้จากการวิเคราะห์ข้อสอบ โดย Glas และ Van Der Linder (2003) ได้ศึกษาการเพิ่มจำนวนของข้อสอบสำหรับการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (Adaptive Testing) และนำข้อมูลดังกล่าวมาใช้ในการวิเคราะห์ Multilevel item response model ด้วยการใช้วิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) ที่ได้จากการศึกษาของMislevy (1993) และวิธีการ The marginal maximum likelyhood โดยการสุ่ม (randomness) เข้าไปในการวิเคราะห์ตามแนวของทฤษฎีการตอบสนองข้อสอบ (IRT) ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ

จากการศึกษาและพัฒนาวิธีการรวมระหว่างทฤษฎีการวิเคราะห์ข้อสอบระหว่างทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) กับทฤษฎีการตอบสนองข้อสอบ (IRT) จนกลายมาเป็นวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่ Briggs และ Wilson ได้พัฒนาโดยมีการเผยแพร่ขึ้นครั้งแรกในปี 2004 โดยได้นำเสนอในงานประชุมประจำปี 2004 ของสมาคมนักวัดทางจิต (Psychometric Society) ที่เมืองแปซิฟิค กรูป (Pacific Grove) มลรัฐแคลิฟอร์เนีย (California) ประเทศสหรัฐอเมริกาแต่การนำเสนอในครั้งนี้ยังเป็นการนำเสนอถึงวิธีการวิเคราะห์ GIRM ในเบื้องต้นและได้พัฒนาวิธีการ GIRM ให้มีความสมบูรณ์มากยิ่งขึ้นในปี 2007 โดยตีพิมพ์เผยแพร่ในวารสารทางวิชาการที่ชื่อว่า Journal of Educational Measurement (JEM) ในบทความที่ชื่อว่า Generalizability in Item Response Modeling (GIRM) จากการศึกษาพัฒนาการการรวมวิธีการวิเคราะห์ระหว่างทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) กับทฤษฎีการตอบสนองข้อสอบ (IRT) จนกลายมาเป็นวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) สามารถสรุปพัฒนาการเกี่ยวกับการรวมวิธีการวิเคราะห์ข้อสอบจากทฤษฎีทั้งสองได้ดังตารางต่อไปนี้

ตารางที่ 1 พัฒนาการของการผสมผสานวิธีการวิเคราะห์ข้อสอบระหว่าง GT กับ IRT

ปี	ผู้พัฒนา	ชื่อผลงาน	สาระสำคัญ
1987	Kolen และ Harris	Kolen และ Harris	การประมาณจากการวิเคราะห์ Multivariate GTด้วยกรอบแนวคิดของ IRT
1993	Mislevy	Formulas for use with Bayesian ability estimated	การใช้วิธีการ Bayesianใน การวิเคราะห์ตามแนวของ IRT สำหรับการประมาณค่าพารามิเตอร์ความสามารถของข้อสอบ
2001	Verhelst และ Verstralen	IRT model for multiple raters	การวิเคราะห์ข้อสอบด้วย IRT สำหรับประเมินผลการปฏิบัติงานโดยใช้ผู้ตรวจหลายคน (two crossed measurement facets)
2002	Patz และคณะ	Hierarchical rater model for rated test items	การประเมินผลการปฏิบัติงานด้วยผู้ตรวจหลายคนโดยวิธี HRM สำหรับ Polytomous item response data score
2002	Bock, Brennan และ Muraki	Information in Multiple Rating	การทดสอบความคงเส้นคงวาในกรณีที่มีผู้ตรวจข้อสอบหลายคน โดยใช้การวิเคราะห์การสรุปอ้างอิง (Generalizability analysis)
2003	Glas และ Van Der Linnden	Computerized adaptive testing with item cloning	การประยุกต์ใช้คอมพิวเตอร์ตามแนวคิดของ Adaptive testing เพื่อผ่อนคลายข้อตกลงเบื้องต้นของข้อที่กำหนดให้เป็นค่าคงที่ (fixed items) ของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT)
2004	Briggs และ Wilson	Running Head: Generalizability in Item Response Modeling	การรวม Sampling Model ; GT เข้าไปใน Scaling Model ; IRT โดยอาศัยสมมติฐานของการกระจายไปยังแหล่งความคลาดเคลื่อนจากการวัดที่เกี่ยวข้อง
2007	Briggs และ Wilson	Generalizability in Item Response Modeling	ขยายผลของ GIRM โดยวิเคราะห์เปรียบเทียบผลการวิเคราะห์จากข้อมูลจริงและข้อมูลจำลอง

1.3.2 แนวคิดและวิธีการ

เมื่อพิจารณาทฤษฎีการตอบสนองข้อสอบ (IRT) พบว่า เป็นทฤษฎีที่มุ่งอธิบายความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่แท้จริงของบุคคลของแต่ละบุคคลกับพฤติกรรมในการตอบสนองข้อสอบในแต่ละข้อ ดังนั้นจะเห็นได้ว่า ทฤษฎีการตอบสนองข้อสอบ (IRT) นี้เป็นมีลักษณะแบบไม่เน้นการสรุปอ้างอิงไปยังผู้สอบคนอื่นและข้อสอบข้ออื่น (fixed effect) ดังนั้นโมเดลตามทฤษฎีการตอบสนองข้อสอบจึงเป็นโมเดลการวัด (scaling model) (Brennan, 2001) อย่างไรก็ตาม เมื่อมีการขยายการสำหรับการทดสอบขนาดใหญ่ โดยการวิเคราะห์ผลการสอบด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) ก็มีการใช้โมเดลการสุ่มของผู้สอบเข้ามาในโมเดลการวิเคราะห์ข้อสอบโดยประยุกต์ใช้วิธีการประมาณค่าวิธีมาร์จินัลแม็กซิมั่มไลค์ลิฮูด (Marginal Maximum Likelihood Estimation) ทั้งนี้เนื่องจากสมมติฐานของการประมาณค่าด้วยวิธีการประมาณค่าวิธีมาร์จินัลแม็กซิมั่มไลค์ลิฮูด (Marginal Maximum Likelihood) เป็นการประมาณค่าตัวแปรแฝง (latent variable) ของแต่ละหน่วยจากลักษณะการกระจายของประชากรที่สนใจ ซึ่งทำให้การวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) สามารถสรุปอ้างอิงไปยังประชากรในการทดสอบได้จากการขยายการสรุปอ้างอิงของผู้สอบไปยังประชากรที่ใช้ในการสอบ ทำให้โมเดลการวัดของทฤษฎีการตอบสนองข้อสอบ (IRT) มีความใกล้เคียงกับโมเดลการวัดของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

ในขั้นต่อไปเพื่อทำให้สามารถรวมทฤษฎีการตอบสนองข้อสอบ (IRT) เข้ากับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ได้ ในการวิเคราะห์หาค่าพารามิเตอร์ของผู้สอบและข้อสอบนั้นจะต้องสุ่มผู้สอบและข้อสอบมาจากประชากรที่สนใจ ซึ่งจะทำให้มโนทัศน์ของการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) จากโมเดลการวัด (scaling model) กลายเป็นโมเดลการวัดและการสุ่ม (scaling and sampling model) อย่างสมบูรณ์โดยสามารถเรียกโมเดลนี้ว่าเป็นโมเดลการตอบสนองข้อสอบแบบผลกระทบสุ่ม (random effect IRT model) ซึ่งโมเดลนี้จะใช้การประมาณค่าด้วยการประมาณค่าด้วยวิธีการประมาณค่าของเบส์ (Bayesian)

นอกจากนี้จากข้อตกลงเบื้องต้นของเงื่อนไขความเป็นอิสระในโมเดลการตอบสนองข้อสอบแบบผลกระทบสุ่มจะต้องมีการระบุการแจกแจงเริ่มแรก (prior distribution) ของค่าพารามิเตอร์ทั้งข้อสอบและผู้สอบ ในกรณี 1 PL model จะต้องระบุการแจกแจงเริ่มแรก (prior distribution) ของ β_i และ θ_p ให้มีลักษณะการกระจายแบบโค้งปกติ (normal distribution) โดยในการประมาณค่าพารามิเตอร์ในโมเดลการตอบสนองข้อสอบแบบผลกระทบสุ่ม (a random effect

IRT) จะใช้ฟังก์ชันไลค์ลิฮูด (likelihood) ในการประมาณค่า ซึ่งสามารถแสดงได้ดังสูตรคณิตศาสตร์สำหรับการวัดที่มีการออกแบบ 1 แหล่งความคลาดเคลื่อน เช่น pxi ได้ดังนี้

$$L(X \mid \theta, \beta) = \prod_{P=1}^{I} \prod_{i=1}^{I} P(\theta_p, \beta_i)^{X_{PI}} \left[1 - P(\theta_P, \beta_i\ )\right]^{1-X_{pi}}$$

(Scaling Model: Maximum Likelihood Function)

ดังนั้นจึงสามารถคำนวณ $P(X,\theta,\beta)$ (ความน่าจะเป็นของผู้ตอบที่มีระดับความสามารถเท่ากับ θ ตอบข้อสอบที่มีค่าความยากข้อสอบเท่ากับ β ได้ถูกต้อง) จากการรวมเมทริกซ์ของคะแนนที่สังเกตได้ $L(X \mid \theta, \beta)$ จากสมการ Scaling Model ในข้างต้นกับ Sampling Model ซึ่งเป็นฟังก์ชันทางคณิตศาสตร์ของพารามิเตอร์ของผู้สอบ $\int(\theta)$ และฟังก์ชันพารามิเตอร์ของข้อสอบ $g(\beta)$ โดยสามารถสรุปเป็นสูตรทางคณิตศาสตร์ของสมการการแจกแจงภายหลัง (posterior distribution) ได้ดังนี้

$$P(X,\theta\ ,\beta\ ) = L(X\ \mid \theta\ ,\beta\ ) f(\theta) g(\beta)$$

ในการประมาณค่าพารามิเตอร์ของ Random effect IRT นั้นสามารถทำได้โดยการประมาณค่าโดยใช้เทคนิค MCMC (Markov Chain Monte Carlo) โดยการคำนวนแบบ Metropolis-Hastings algorithm ภายในวิธีการ Gibbs sampler ซึ่งจะสามารถประมาณค่าการแจกแจงภายหลัง (Posterior distribution) ของ ความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ) ซึ่งสามารถเขียนเป็นสมการสำหรับประมาณค่าความสามารถที่แท้จริงของผู้สอบ (θ) ได้ดังนี้

$$P(\theta \mid \beta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\theta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\theta}$$

และสามารถประมาณค่า Posterior distribution ของค่าพารามิเตอร์ของข้อสอบแต่ละข้อ $(\beta\ )$ ซึ่งสามารถเขียนเป็นสมการสำหรับประมาณค่าพารามิเตอร์ของข้อสอบ $(\beta\ )$ ได้ดังนี้

$$P(\beta \mid \theta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\beta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\beta}$$

จากสมการ $P(\theta \mid \beta, X)$ และสมการ $P(\beta \mid \theta, X)$ ที่สามารถประมาณได้จากเทคนิค MCMC จะทำให้สามารถหาค่าของ $P(\theta_p, \beta_i)$ ของแต่ละคนในแต่ละข้อของฟังก์ชันการตอบสนองข้อสอบแบบราสช์โมเดล (Rasch model item response function)

สำหรับการประมาณองค์ประกอบความแปรปรวนและค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด สามารถประมาณได้จากการวิเคราะห์ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ด้วยกรอบแนวคิดของทฤษฎีการตอบสนองข้อสอบ (IRT) ที่ได้จากการเชื่อมโยงสูตรของ Kolen และ Harris (1987) ดังนี้

$$\mu = \int_\theta \int_\beta P(\theta_p, \beta_i) f(\beta) g(\theta) d\beta d\theta$$ **(คำอธิบาย μ ใน GT)**

$$\pi(\theta) = \int_\beta P(\theta_p, \beta_i) f(\beta) d\beta - \mu$$ **(คำอธิบาย ν_p ใน GT)**

$$\iota(\theta) = \int_\theta P(\theta_p, \beta_i) g(\theta) d\theta - \mu$$ **(คำอธิบาย ν_i ใน GT)**

$$\nu(\theta, \beta) = P(\theta_p, \beta_i) - \pi(\theta) - \iota(\beta) - \mu$$ **(คำอธิบาย ν_{pi} ใน GT)**

จากสมการทั้งสี่สมการดังกล่าวในสามารถประมาณพารามิเตอร์ที่ได้จากการวิเคราะห์ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ด้วยกรอบแนวคิดของทฤษฎีการตอบสนองข้อสอบ (IRT) สามารถเขียนเป็นสมการรวมได้ดังนี้

$$P(\theta_p, \beta_i) = \mu + \pi(\theta) + \iota(\beta) + \nu(\theta, \beta)$$

จาก $P(\theta_P, \beta_i) = E(X_{pi})$ ดังนั้นยังต้องมีส่วนสุดท้ายที่ต้องคำนึงถึงในสมการอีก ก็คือ ความคลาดเคลื่อนแบบสุ่ม (random error term) $e_{pi} = e(\theta_P, \beta_i)$ ดังนั้นจากสมการของ Kolen และ Harris สามารถสรุปเป็นสมการของคะแนนสังเกตได้ดังนี้

$$\mathrm{X}_{pi} = E(X_{pi}) + e_{pi} = \mu + \pi(\theta) + i(\theta) + \nu(\theta, \beta) + e_{pi}$$

เนื่องด้วยภายใต้ข้อตกลงเบื้องต้นของสมการ Scaling Model: Random IRT คะแนนที่สังเกตได้ (X_{pi}) มีลักษณะการกระจายตัวแบบแบร์นูลีย์ (Bernoulli distribution) ดังนั้นความแปรปรวนของความคลาดเคลื่อนแบบสุ่มจะมีค่าเท่ากับ $[P(\theta_P,\beta_i)(1-P(\theta_p,\beta_i)]$ โดยสามารถสรุปเป็นสมการได้

$$\sigma^2(e)=\int_\theta\int_\beta P(\theta_p,\beta_i)[1-P(\theta_P,\beta_i)]f(\beta)g(\theta)d\beta d\theta$$

สามารถเขียนในรูปของความแปรปรวนทั้งหมดของคะแนนที่สังเกตได้ดังนี้

$$\sigma^2(X_{pi})=\sigma^2(\pi)+\sigma^2(i)+\sigma^2(v)+\sigma^2(e)$$

จากสมการข้างต้นสามารถสรุปเปรียบเทียบกรณี p X i design สมการความแปรปรวนของ Kolen และ Harris (1987) กับสมการความแปรปรวนของทฤษฎีการสรุปอ้างอิง ความน่าเชื่อถือของผลการวัด (GT) ได้ดังนี้

ตารางที่ 2 การเปรียบเทียบแหล่งความแปรปรวนที่ได้จากการวิเคราะห์ระหว่างGT กับสูตรของ Kolen และ Harris

แหล่งความแปรปรวน (Source of error)	G-Theory Variance component	Kolen และ Harris Variance component
บุคคล (p)	$\sigma^2(p)$	$\sigma^2(\pi)$
ข้อสอบ (i)	$\sigma^2(i)$	$\sigma^2(i)$
ปฏิสัมพันธ์ระหว่าง บุคคลกับข้อสอบ (pi)	$\sigma^2(pi,e)$	$\sigma^2(v)$
ความคลาดเคลื่อนแบบสุ่ม (e)		$\sigma^2(e)$

จากตารางจะเห็นได้ว่า สมการของ Kolen และ Harris (1987) สามารถแยกความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบและความคลาดเคลื่อนแบบสุ่มได้ ซึ่ง การวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ไม่สามารถแบ่งแยกแหล่งของความแปรปรวนดังกล่าวออกจากกันได้

ดังนั้นสามารถแสดงเมทริกซ์การตอบสนองที่คาดหวัง (an expected response matrix) ของพารามิเตอร์ของโมเดลการตอบสนองข้อสอบ (item response model) ทั้ง θ_p สำหรับทุกคน และ β_i สำหรับทุกข้อ ได้ดังนี้

พารามิเตอร์ของข้อสอบ

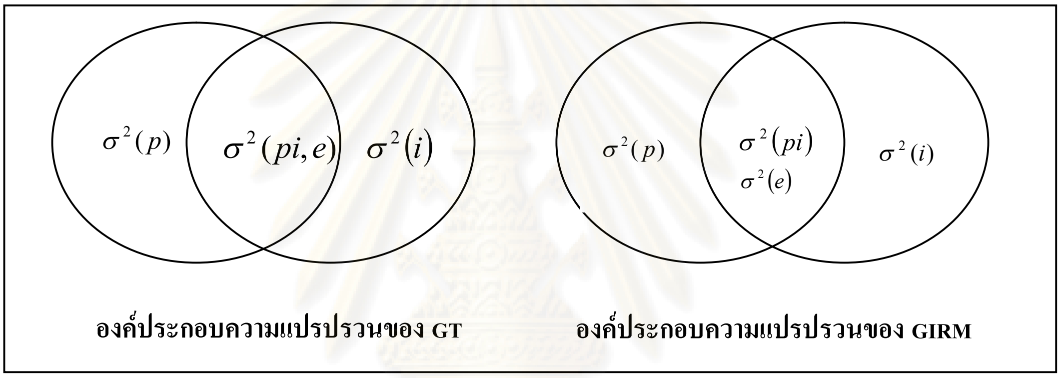
		β_1	$\dots$	β_i	$\dots$	β_I
พารามิเตอร์ของผู้สอบ	θ_1	E (X_{11})		$\dots$		E (X_{1I})
	$\dots$					
	θ_p	$\dots$		E (X_{pi})		$\dots.$
	$\dots$					
	θ_P	E (X_{P1})		$\dots$		E (X_{PI})

แผนภาพที่ 3 เมทริกซ์การตอบสนองที่คาดหวังของผู้สอบตามวิธี GIRM สำหรับการออกแบบการวัด p x i

การประมาณค่าในวิธี GIRM เริ่มต้นจากการประมาณค่าพารามิเตอร์ในสมการ Scaling Model: Random IRT ซึ่งได้ค่าในแต่ละเซลล์ของเมทริกซ์ดังแผนภาพที่ 3 $E(X_{pi}) = P(\theta_P, \beta_i)$ หลังจากนั้นจึงทำการประมาณค่าองค์ประกอบความแปรปรวน (variance component) ตามสูตรของ Kolen และ Harris ซึ่งจะทำให้ได้ค่า $\pi(\theta)$ $i(\theta)$ และ $\nu(\theta, \beta)$ การประมาณค่าแบบนี้จะไม่ทำให้เกิดความลำเอียง (Searle และคณะ, 1992) นอกจากนี้ยังต้องคำนึงถึงสมมติฐานเบื้องต้นของบุคคลและข้อคำถามที่สุ่มว่ามีวิธีการสุ่มที่ถูกต้องหรือไม่ และมีการกระจายก่อนนำมาคำนวณอย่างไร (Prior distribution) แต่ในทางปฏิบัติค่าพารามิเตอร์ θ_p และ β_i เป็นค่าที่ไม่ทราบและจะต้องถูกประมาณขึ้น โดยสามารถประมาณค่าพารามิเตอร์ดังกล่าว ได้จากสมการการแจกแจงภายหลัง (Posterior distribution) ด้วยวิธีการประมาณค่า MCMC ของโมเดลการตอบสนองข้อสอบในแต่ละขั้นของการประมาณค่าด้วยวิธีการ Markov Chain จะทำให้สามารถประมาณค่า θ_p และ β_i ได้ เป็น $\hat{\theta}_p{}^{(m)}$ และ $\hat{\beta}^{(m)}{}_i$ ซึ่งค่าที่ได้จากการประมาณค่านี้จะเป็นองค์ประกอบในเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) และนำมาใช้ในการคำนวณตามสูตรของ Kolen และ Harris จะทำให้ได้องค์ประกอบของความแปรปรวน $\hat{\sigma}^2(\pi)^{(m)}$ $\hat{\sigma}^2(i)^{(m)}$ $\hat{\sigma}^2(\gamma)^{(m)}$ และ $\hat{\sigma}^2(e)^{(m)}$ การประมาณค่าด้วยวิธีการ Markov Chain นี้จะให้ค่าประมาณที่ถูกต้องและเป็นค่าเดียวกับการประมาณด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ก็ต่อเมื่อลักษณะการกระจายของข้อมูลต้องมีลักษณะการกระจายตัวที่มีแนวโน้มเข้าสู่ส่วนกลาง (central tendency) ในการ

ประมาณค่าด้วยวิธี GIRM จึงต้องมีการระบุลักษณะการกระจายตัวของข้อมูลด้วย เนื่องจากการประมาณค่าด้วยวิธี GIRM มีลักษณะของการประมาณค่าที่มีความลำเอียงในการประมาณค่าต่ำกว่าค่าที่ควรจะเป็น (to be biased downward) (Briggs และ Wilson, 2007)

ความแตกต่างที่เห็นได้ชัดเจนจากผลวิเคราะห์ข้อสอบด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) กับวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) คือ ผลการวิเคราะห์ขององค์ประกอบของความแปรปรวนระหว่างทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ซึ่งสามารถสรุปได้ตามแผนภาพดังนี้



แผนภาพที่ 4 ความแตกต่างขององค์ประกอบความแปรปรวนระหว่าง GT และ GIRM

จากแผนภาพจะเห็นได้ว่า การวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) สามารถวิเคราะห์แยกความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบ ($\sigma^2(pi)$) และความคลาดเคลื่อนแบบสุ่ม ($\sigma^2(e)$) ได้ในขณะที่การวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ไม่สามารถแยกองค์ประกอบของความแปรปรวนดังกล่าวได้

1.3.3 **ข้อตกลงเบื้องต้น** (Assumption)

วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าในโมเดลการกำหนดมาตรวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) อาศัยสมมติฐานของการกระจายไปยังแหล่งความคลาดเคลื่อนจากการวัดที่เกี่ยวข้อง โดยการประมาณค่า GT ในแต่ละองค์ประกอบของความแปรปรวน (Variance Components) ไปพร้อมกับการประมาณค่าพารามิเตอร์ในโมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) ดังนั้นข้อตกลงเบื้องต้นของวิธีการจึงเป็น

ข้อตกเบื้องต้นทั้งข้อตกลงเบื้องต้นของการสุ่มตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และข้อตกลงเบื้องต้นของการกระจายตามทฤษฎีการตอบสนองข้อสอบ (IRT) โดยมีรายละเอียดดังนี้

1. ข้อตกลงเบื้องต้นของการสุ่มตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) (Sampling assumption: common GT assumption)

ข้อตกลงเบื้องต้นของการสุ่มนั้นถือว่าเป็นข้อตกลงเบื้องต้นที่สำคัญของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เนื่องจากทฤษฎีนี้ต้องการสรุปอ้างอิงผลการวัดจากกลุ่มตัวอย่างไปยังประชากร ดังนั้นในการสุ่มผู้สอบ (persons) และแหล่งความคลาดเคลื่อนของการวัด (facets) จึงต้องถูกสุ่มอย่างเป็นอิสระจากประชากรผู้ซึ่งเป็นกลุ่มเป้าหมายในการทดสอบและแหล่งความคลาดเคลื่อนที่น่าจะเกิดขึ้นจริงจากการทดสอบ แต่ข้อตกลงเบื้องต้นข้อนี้สามารถผ่อนคลายได้ถ้าสถานการณ์ในการวัดใดไม่เน้นในการสรุปอ้างอิงผลการวัดไปยังประชากร (Briggs และ Wilson, 2007)

นอกจากนี้ในการวิเคราะห์ข้อสอบด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ใช้หลักการวิเคราะห์ความแปรปรวน (Analysis of Variance; ANOVA) ดังนั้นในการวิเคราะห์วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) จึงต้องพิจารณาข้อตกลงเบื้องต้นของการวิเคราะห์ความแปรปรวนด้วย ซึ่งมีรายละเอียดของข้อตกลงเบื้องต้นดังนี้

1) ความเป็นอิสระต่อกัน (idependence) หน่วยตัวอย่างในแต่ละกลุ่มต้องสุ่มมาจากประชากรที่เป็นอิสระต่อกันทั้งภายในกลุ่มและระหว่างกลุ่ม

2) ความเป็นเอกพันธ์ของความแปรปรวน (variance homegeneity) ตัวอย่างแต่ละกลุ่มสุ่มมาจากประชากรที่มีความแปรปรวนเท่ากัน

3) ความเป็นปกติ (normality) ตัวอย่างแต่ละกลุ่มสุ่มมาจากประชากรที่มีการแจกแจงแบบปกติ

2. ข้อตกลงเบื้องต้นของการวัดตามทฤษฎีการตอบสนองข้อสอบ (the standard assumptions of IRT) เนื่องด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นวิธีการที่กระจายแหล่งของความคลาดเคลื่อนเข้าไปในสมการของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) ดังนั้นจึงต้องใช้ข้อตกลงเบื้องต้นของมาตรฐานของทฤษฎีการตอบสนองข้อสอบ ซึ่งได้แก่

2.1 ความเป็นเอกมิติของคุณลักษณะที่ใช้ในการทดสอบ (Unidimensional)

2.2 ความเป็นอิสระของแบบสอบ (Local independence)

2.2.1 ความเป็นอิสระระหว่างข้อสอบ (item local independence)

2.2.2 ความเป็นอิสระหว่างผู้สอบ (person local independence)

2.3 โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) มีลักษณะเป็นฟังก์ชันโลจิส (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม(Normal ogive function)

2.4 การทดสอบที่ไม่แข่งขันด้านเวลา (Nonspeed Test Administration)

3.ข้อตกลงเบื้องต้นของการกระจายแหล่งความคลาดเคลื่อน (Distributional assumption about measurement facet)

การกระจายของแหล่งความคลาดเคลื่อนในแต่ละแหล่งจะต้องมีลักษณะเป็นโค้งปกติ (normal distribution) แต่จากการจำลองข้อมูลเบื้องต้นของ BriggsและWilson (2007) พบว่า วิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GIRM) มีความแกร่ง (robustness) ต่อการกระจายของข้อมูลก่อนนำมาวิเคราะห์ (misspecification of prior distributions) และยังแกร่ง (robust) สำหรับ การวิเคราะห์ข้อมูลด้วยฟังก์ชันการตอบสนองข้อสอบ (the parametric form of the item response function) ที่ใช้ในการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) อีกด้วย

1.3.4 ความแกร่งของการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบ สนองข้อสอบ (Robustness)

เนื่องด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นวิธีการที่รวมโมเดลการสุ่มของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการวัดของโมเดลตามทฤษฎีการตอบสนองข้อสอบ (IRT) โดยอาศัยสมมติฐานของ การกระจายไปยังแหล่งความคลาดเคลื่อนของการวัดที่เกี่ยวข้อง ดังนั้นในข้อตกลงเบื้องต้นของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) จึงต้องมีทั้งข้อตกลงเบื้องต้นทั้งสองทฤษฎีรวมกับข้อตกลงเบื้องต้นของการกระจายแหล่งความคลาดเคลื่อนของการวัดการนำเสนอในส่วนนี้ได้มีการตรวจสอบความแกร่ง (Robustness) เบื้องต้นด้วยวิธีการจำลองข้อมูล (simulation) ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ตามข้อตกลงเบื้องต้นของวิธีการ ดังนี้

1) **ความไม่แน่นอนในการประมาณค่าองค์ประกอบของความแปรปรวน** (The Uncertainty of Variance Component Estimates)

ความไม่แน่นอนในการประมาณค่าขององค์ประกอบความแปรปรวนวิธีการสรุปอ้างอิง ความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่ประมาณค่าการแจกแจงภายหลัง (posterior distributions) ขององค์ประกอบความแปรปรวนและค่าสัมประสิทธิ์การสรุปอ้างอิงด้วยค่าเฉลี่ยของการกระจายที่ประมาณค่าแบบจุดโดยมากมักจะใช้วิธีการคำนวณ ANOVA เหมือนกับการประมาณค่าในวิธีการตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ความไม่แน่นอนของการประมาณค่าการแจกแจงภายหลัง (posterior distributions) โดยมากมักสรุปในรูปของ Posterior standard deviation ซึ่งได้แก่ ความแปรปรวนในการประมาณของแหล่งความคลาดเคลื่อนต่างๆ เช่น การออกแบบการวัด แบบ p x i จะมีแหล่งความแปรปรวนจากบุคคล ($\sigma^2(p)$) เท่ากับ .029 แหล่งความแปรปรวนจากข้อสอบ ($\sigma^2(i)$) เท่ากับ .043 แหล่งความคลาดเคลื่อนที่เกิดปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบและความคลาดเคลื่อนแบบสุ่ม $\sigma^2(pi,e)$ เท่ากับ .177

จากการศึกษาโดยการจำลองข้อมูลของ Brigg และ Wilson (2007) พบว่า การประมาณค่าความแปรปรวนของการวิเคราะห์ด้วยวิธี GIRM มีแนวโน้มในการเกิดความลำเอียงโดยมีค่าที่ประมาณได้ต่ำกว่าค่าที่ควรจะเป็น (to be biased downward) ทั้งนี้สาเหตุน่าจะมาจากการป้อนคำสั่งในกระบวนการการประมาณค่าความแปรปรวนขององค์ประกอบความคลาดเคลื่อนของขั้นตอนการประมาณด้วยวิธี Markov Chain ที่ไม่สามารถทำให้เป็นอิสระจากกันได้ จึงนำไปสู่การฝ่าฝืนข้อตกลงเบื้องต้นต่อผลกระทบแบบสุ่ม (random effects) เช่น การประมาณค่าองค์ประกอบความแปรปรวนของ $\sigma^2(i)$ ตามวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) แบบ p X i design ด้วยวิธีการ Markov Chain เพื่อนำมาสร้างเป็นเมทริกซ์คะแนนที่สังเกตได้ของบุคคลดังแผนภาพที่ 3 ในการคำนวณดังกล่าวต้องการคำนวณค่า $\sigma^2(i)$ ซึ่งในการประมาณค่าจะต้องมีการเปลี่ยนแปลงเฉพาะคอลัมน์ในเมทริกซ์ โดยค่าในแถวของเมทริกซ์จะต้องคงที่ แต่ในการวิเคราะห์ด้วยโมเดล GIRM นั้นใช้การประมาณค่าด้วยวิธี Markov Chain ซึ่งจะทำให้ทั้งคอลัมน์และแถวในเมทริกซ์มีค่าที่ผันแปรร่วมกัน ซึ่งแตกต่างจากการคำนวณในวิธีการตาม ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ที่มีผลกระทบจากเรื่องนี้น้อยมากเพราะการวิเคราะห์นั้นได้ทำการวิเคราะห์โดยใช้โปรแกรม GENOVA ซึ่งเป็นโปรแกรมในการประมาณค่าความคลาดเคลื่อนมาตรฐานโดยมีรูปแบบที่ปิดสำหรับสมมติฐานของการแจกแจงแบบโค้งปกติจึงทำให้ไม่เกิดความลำเอียงในการประมาณค่า

ดังนั้นจะเห็นได้ว่าการประมาณค่าแบบ GIRM เป็นการประมาณค่าที่ประยุกต์การวิเคราะห์แบบ GT ด้วยข้อมูลเมทริกซ์ที่คาดหวังที่ได้จากการประมาณค่าการตอบสนองที่คาดหวังในแต่ละขั้นตอนของ MCMC ซึ่งสามารถเขียนการประมาณค่าการวิเคราะห์ GT ในวิธีการ GIRM ได้ด้วย $E\left[X_{pi} \mid \theta_p^{(m)}, \beta_i^{(m)}\right]$ หรือสามารถอธิบายได้ว่า ในการวิเคราะห์ด้วยวิธีการ GIRM สามารถวิเคราะห์ GT โดยเอาข้อมูลมาจากการคำนวณคะแนนที่คาดหวัง (expected score) ที่ได้มาจากค่าพารามิเตอร์ของโมเดลการตอบสนองข้อสอบ (IRT) ที่ประมาณได้จากการประมาณค่าในแต่ละขั้นตอนของ MCMC ซึ่งในขั้นตอน MCMC จะทำให้การประมาณค่าไม่เป็นอิสระต่อกันระหว่าง $\theta_p^{(m)}$ กับ $\beta_i^{(m)}$ มีผู้เสนอแนะการประมาณค่าอีกวิธีการหนึ่ง ซึ่งน่าจะช่วยแก้ปัญหาความลำเอียงในการประมาณค่า ไว้ในบทความ Generalizability in Item Response Modeling (Brigg และ Willson, 2007) ใน Journal of Educational Measurement ไว้ว่า ควรจะประมาณค่าโดยใช้ค่าพารามิเตอร์ของโมเดลการตอบสนองข้อสอบ (IRT) เข้าไปในแต่ละขั้นของการจำลองข้อมูลเมทริกซ์ของการตอบสนองข้อสอบ หลังจากนั้นจึงนำค่าพารามิเตอร์ที่ได้มาประมาณค่าคะแนนที่สังเกตได้โดยการคำนวณค่าการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) แล้วจึงค่อยประยุกต์ใช้การวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ในการทำนาย สามารถเขียนการประมาณค่าการวิเคราะห์ GT ในวิธีการ GIRM วิธีใหม่นี้ได้ด้วย $X_{pi}^{(m)} \sim Rasch(\theta_p, \beta_i)$ นั้นก็คือการคำนวณค่าพารามิเตอร์จากทฤษฎีการตอบสนองข้อสอบ (IRT) ก่อนแล้วจึงสร้างเลียนแบบข้อมูลที่สังเกตได้ หลังจากนั้นจึงวิเคราะห์ค่าองค์ประกอบของความแปรปรวนตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ซึ่งจะไม่ต้องผลกระทบต่อความไม่เป็นอิสระต่อกันระหว่าง $\theta_p^{(m)}$ กับ $\beta_i^{(m)}$

ความแตกต่างระหว่างการวิเคราะห์ด้วย ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) คือ การวิเคราะห์ด้วย ทฤษฎีแรกเป็นการประมาณค่าจากคะแนนที่สังเกตได้ (X_{pi}) ในขณะที่การประมาณค่าจากวิธีหลังเป็นการประมาณค่าจากคะแนนพยากรณ์ที่ได้หลังจากการจำลองข้อมูล $E(X_{pi})$

2) ความไวของผลจากการวิเคราะห์ด้วยวิธี GIRM ตามสมมติฐานในการกระจาย (Sensitivity of GIRM Results to Distributional Assumptions)

เมื่อพิจารณาค่าความไวของผลการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) แบบฟังก์ชันการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ ที่มีต่อสมมติฐานการกระจาย (Sensitivity of GIRM Results to Distributional Assumptions) โดยค่าพารามิเตอร์ข้อสอบ (β_i) ศึกษาลักษณะการกระจายใน 3 รูปแบบ คือ แบบแรกมีลักษณะการกระจายที่เป็นข้อตกลงเบื้องต้นของวิธีการ GIRM คือมีลักษณะการกระจายแบบ Normal (0, 1) ข้อมูลที่มีลักษณะการแจกแจงของแบบโค้งปกติคล้ายรูประฆังคว่ำ มีค่าเฉลี่ยเท่ากับ 0 และค่าความแปรปรวนเท่ากับ 1 แบบที่สองลักษณะการกระจายแบบ Uniform (-2, 2) ข้อมูลที่มีลักษณะการแจกแจงคล้ายรูปสี่เหลี่ยมผืนผ้านั้นก็คือข้อมูลมีโอกาสในการกระจายตัวที่เท่ากัน (equally probable) ของค่าพารามิเตอร์ที่มีค่าอยู่ระหว่าง -2 ถึง 2 และแบบสุดท้ายลักษณะการกระจายแบบ Beta (.5,.5) ข้อมูลที่มีลักษณะการแจกแจงแบบโค้งรูปตัวยู (U-shapes curve) ของการกระจายร่วมกันในตัวแปร 2 ตัว โดยมีโอกาสในการร่วมกันของแต่ละตัวแปรเท่ากันซึ่งเท่ากับ .5 สำหรับค่า θ_p ศึกษาลักษณะการกระจายตัวใน 3 รูปแบบ โดยศึกษาลักษณะการกระจายตัวเหมือนกันกับค่า β_i ในสองรูปแบบแรก คือ Normal (0, 1) และ Uniform (-2, 2) ส่วนในลักษณะการกระจายตัวในรูปแบบที่สามจะศึกษาลักษณะการกระจายตัวแบบ Gamma (1) ข้อมูลที่มีลักษณะการแจกแจงแบบโค้งเบ้ขวา (positive skewed curve) โดยมาตรวัดของค่าพารามิเตอร์ไม่เกิน 1และมีลักษณะการกระจายแบบ Exponential

ผลการศึกษา พบว่า การวิเคราะห์ด้วยวิธี GIRM ให้ผลการวิเคราะห์ที่มีลักษณะข้อมูลของการกระจายทั้งค่า β_i และค่า θ_p ของการกำหนดการแจกแจงเริ่มแรก (prior distribution) ที่มีลักษณะแบบ Normal (0, 1) และแบบ Uniform (-2,2) ให้ผลไม่แตกต่างกับผลการวิเคราะห์ที่มีลักษณะการกระจายแบบ Normal (0, 1) ของ β_i กับการกระจาย Normal (0, 1) ของ θ_p แต่ถ้ามีลักษณะการกระจายของค่า β_i แบบ Beta (.5,.5) กับการกระจายของค่า θ_p แบบ Gamma (1) จะส่งผลต่อผลการวิเคราะห์ด้วยวิธี GIRM ส่วนการวิเคราะห์ GT ไม่ว่าข้อมูลของการแจกแจงเริ่มแรกจะมีลักษณะอย่างไรให้ผลการวิเคราะห์เหมือนกับข้อมูลที่มีลักษณะการแจกแจงแบบ Normal (0, 1) โดยศึกษาจากการจำลองข้อมูล ผู้ตอบ 500 คน ข้อสอบ 20 ข้อ ซึ่งมีการกำหนดเงื่อนไขที่ผันแปรของการกระจายค่าพารามิเตอร์ θ_p 3 แบบ คือ Uniform (-2, 2) และ Gamma (1) และกำหนดเงื่อนไขที่ผันแปรของการกระจายค่าพารามิเตอร์ของข้อสอบ (β_i) แบบ Normal (0, 1) Uniform (-2, 2) และ Beta (.5,.5) ซึ่งสามารถสรุปความสอดคล้องได้ดังตารางที่ 3

ตารางที่ 3 ความสอดคล้องของผลการวิเคราะห์ที่มีลักษณะการกระจายในรูปแบบต่าง ๆ กับการกระจายตามข้อตกลงเบื้องต้นในการวิเคราะห์

ค่าพารามิเตอร์ของผู้สอบ (θ_p)	**ค่าพารามิเตอร์ของข้อสอบ** (β_i)					
	Normal (0, 1)		Uniform (-2,-2)		Beta (.5,.5)	
	GIRM	GT	GIRM	GT	GIRM	GT
Normal (0, 1)	*	*	√	√	X	√
Uniform (-2,2)	√	√	√	√	X	√
Gamma (1)	X	√	X	√	X	√

* หมายถึง การวิเคราะห์ตามเงื่อนไขของข้อตกลงเบื้องต้น

√ หมายถึง ผลการวิเคราะห์สอดคล้องกับผลการวิเคราะห์ตามเงื่อนไขของข้อตกลงเบื้องต้น

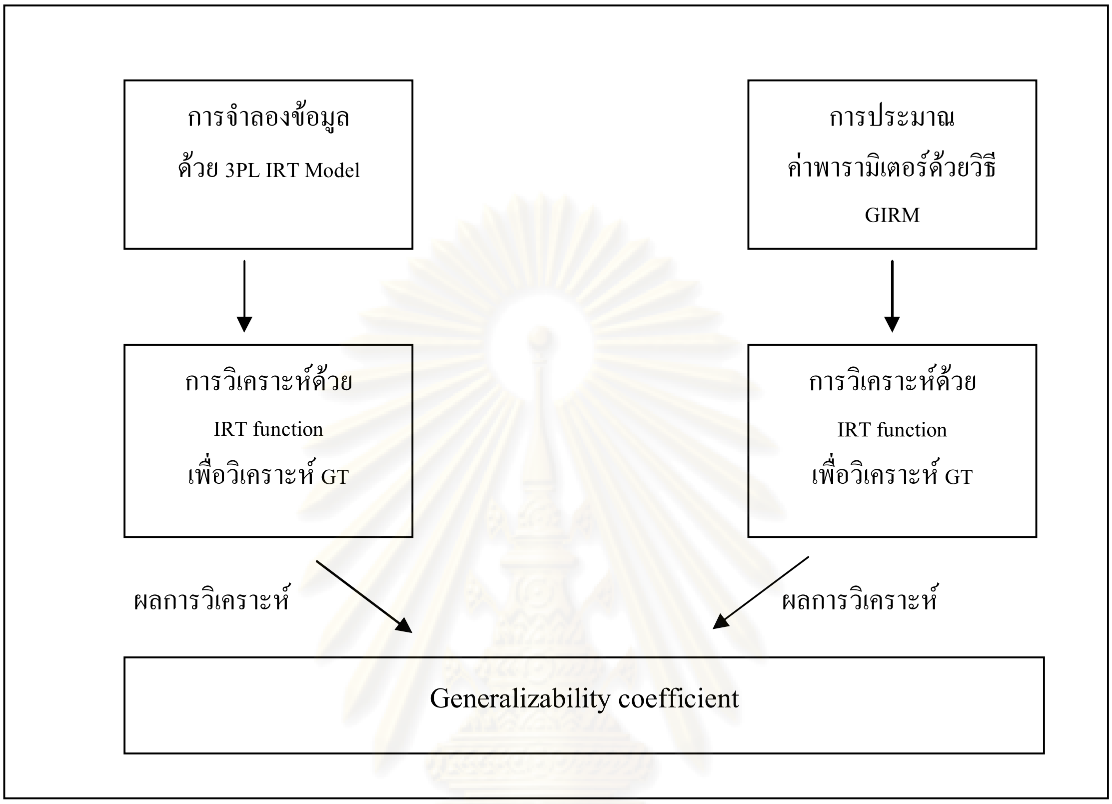
X หมายถึง ผลการวิเคราะห์ไม่สอดคล้องกับผลการวิเคราะห์ตามเงื่อนไขของข้อตกลงเบื้องต้น

ดังนั้นจะเห็นได้ว่าการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) มีความไวตามสมมติฐานในการกระจายมากกว่าการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เมื่อข้อมูลไม่เป็นไปตามเงื่อนไขของสมมติฐาน

3) ความไวของผลจากการวิเคราะห์ข้อมูลด้วยวิธี GIRM ด้วยฟังก์ชันการตอบสนองข้อสอบ (Sensitivity of GIRM Results to Specification of Item Response Function)

จากการจำลองข้อมูลโดยใช้โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (3PL model) ข้อสอบ 20 ข้อ ผู้สอบ 500 คน สามารถให้ข้อค้นพบเบื้องต้นสำหรับการออกแบบการวัดแบบ p X i พบว่า การประมาณค่าทั้งสองวิธีให้ผลการประมาณค่าองค์ประกอบความแปรปรวน (variance component) และสัมประสิทธิ์สหสัมพันธ์การสรุปอ้างอิง (Generalizability coefficient) ที่เหมือนกันทุกประการ ระหว่างวิธีการประมาณค่าด้วยฟังก์ชันการตอบสนองโดยการจำลองข้อมูลจากโมเดลการตอบสนองข้อสอบ (IRT) กับการประมาณค่าด้วยฟังก์ชันการตอบสนองโดยใช้ผลการประมาณค่าพารามิเตอร์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) การประมาณค่าด้วยวิธีGIRM นี้ใช้การจำลองข้อมูลโดยใช้โมเดลราช์ล (Rasch model) ที่มีลักษณะข้อมูลก่อนการจำลอง (Prior distributions) ทั้งพารามิเตอร์ของข้อสอบและผู้สอบมีลักษณะการกระจายแบบโค้งปกติ (Normal) ดังนั้นจึงสามารถสรุปได้ว่า การประมาณค่าด้วยฟังก์ชันการตอบสนองโดยใช้ผลการประมาณค่าพารามิเตอร์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดล

การตอบสนองข้อสอบ (GIRM) ไม่มีความไวต่อการคำนวณที่ผิดพลาดด้วยการใช้ฟังก์ชันการตอบสนองข้อสอบ สามารถสรุปได้ด้วยแผนภาพดังนี้



แผนภาพที่ 5 ความไวของผลจากการวิเคราะห์ข้อมูลด้วยวิธี GIRM ด้วยฟังก์ชันการตอบสนองข้อสอบ

1.3.5 ตัวอย่างงานวิจัยเกี่ยวกับวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ

วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นวิธีการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) อาศัยสมมติฐานของการกระจายไปยังแหล่งความคลาดเคลื่อนจากการวัดที่เกี่ยวข้อง โดยการเชื่อมโยงสูตรของ Kolen และ Harris (1987) ในการวิเคราะห์องค์ประกอบความแปรปรวนและค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัดโดยการประมาณได้จากการวิเคราะห์ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ด้วยกรอบแนวคิดของโมเดลการตอบสนองข้อสอบ (IRT) และการนำกรอบแนวความคิดของ Verhelst และ Verstralen (2001) ในบริบทของการประเมินซ้ำของการประเมิน

การปฏิบัติงาน ซึ่งนำไปสู่การเชื่อมโยงระหว่างทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) กับทฤษฎีการตอบสนองข้อสอบ (IRT) และขยายกรอบแนวคิดที่เด่นชัดเกี่ยวกับโมเดลผลกระทบแบบสุ่มพหุระดับ (Hierarchical random effect models) จึงกลายมาเป็นวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่ Briggs และ Wilson ได้พัฒนา โดยมีการเผยแพร่ขึ้นครั้งแรกในปี 2004 นำเสนองานประชุมประจำปี 2004 ของสมาคมนักวัดทางจิต (Psychometric Society) ที่เมืองแปซิฟิค กรูป (Pacific Grove) มลรัฐแคลิฟอร์เนีย (California) ประเทศสหรัฐอเมริกาแต่การนำเสนอยังเป็นการนำเสนอถึงวิธีการวิเคราะห์ GIRM ในเบื้องต้นและได้พัฒนาวิธีการ GIRM แล้วเสร็จในปี2007 โดยตีพิมพ์เผยแพร่ในวารสารทางวิชาการที่ชื่อว่า Journal of Educational Measurement (JEM) ในบทความที่ชื่อว่า Generalizability in Item Response Modeling

ดังนั้นจะเห็นได้ว่า วิธีการวิเคราะห์ข้อสอบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นวิธีการวิเคราะห์ใหม่ที่เกิดขึ้นในปี 2007 ดังนั้นผลการศึกษาวิจัยด้วยวิธีการวิเคราะห์ GIRM นี้ ในปัจจุบันจึงมีเพียงบทความเดียว นั้นก็คือบทความ Generalizability in Item Response Modeling ซึ่งตีพิมพ์ในวารสาร Journal of Educational Measurement (JEM) ดังนั้นในส่วนนี้จึงนำตัวอย่างงานวิจัยเกี่ยวกับวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบของ Briggs และ Wilson (2007) เพียงบทความเดียว โดยแบ่งแยก การนำเสนอออกเป็น 2 ส่วน คือ การวิเคราะห์ GIRM โดยใช้การจำลองข้อมูลในการวิเคราะห์ (Simulation) และการวิเคราะห์ GIRM โดยการศึกษาจากการเก็บข้อมูลจริง (Real data) ซึ่งมีรายละเอียดดังนี้

1) การศึกษาด้วยวิธีการจำลองข้อมูล (Simulation)

Briggs และ Wilson (2007) ได้ศึกษาเปรียบเทียบผลการวิเคราะห์ข้อมูลด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) กับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ด้วยการจำลองข้อมูลโดยมุ่งเปรียบเทียบการประมาณค่าองค์ประกอบความแปรปรวน (variance components) และค่าสัมประสิทธิ์การสรุปอ้างอิง (generalizability coefficients) ระหว่างการประมาณค่าด้วยวิธีวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) กับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ข้อมูลที่ใช้ในการวิเคราะห์ได้มาจากการจำลองข้อมูลจากการประมาณค่าด้วยวิธีการ MCMC 3 ห่วงโซ่ ด้วยการประมาณค่าซ้ำ (iteration) 10,000 ครั้ง หลังจากที่ได้ตัดข้อมูลในส่วนแรก (burn-in) 1,000 ข้อมูล โดยใช้โปรแกรม WinBUGS จากการประมาณค่าด้วยวิธีการMCMC (Markov Chain Monte Carlo) ผู้ตอบ 500 คนใน 5 ข้อคำถาม โดยทำกลุ่มข้อมูลซ้ำ (data sets) 100 ชุด ตามรูปแบบของ Rasch Model ซึ่ง

คำนวณค่าพารามิเตอร์ θ_p และ β_i แบบอิสระจากการกระจายของข้อมูลแบบโค้งปกติ การประมาณค่า GT ใช้สูตรมาตรฐานในโปรแกรม R

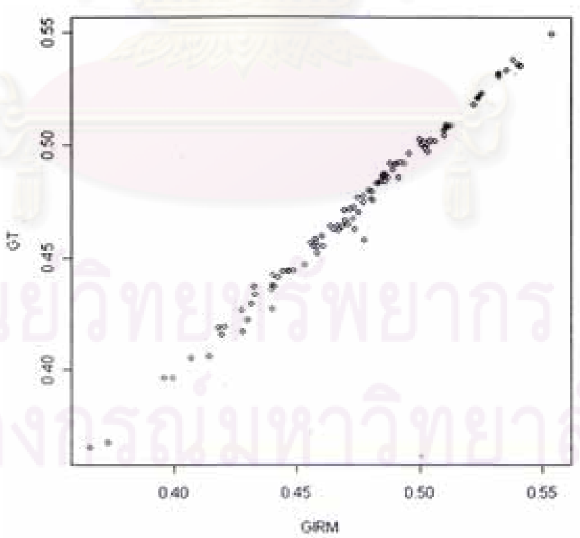
ตารางที่ 4 องค์ประกอบความแปรปรวนด้วยการวิเคราะห์ GT และ GIRM โดยการจำลองข้อมูล

องค์ประกอบความแปรปรวน/ ค่าค่าสัมประสิทธิ์การสรุปอ้างอิง	**ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) ที่ได้จากวิธี GIRM***	**ค่าเฉลี่ย (ส่วนเบี่ยงเบนมาตรฐาน) ที่ได้จากวิธี GT***
$\sigma^2(p)$	.033 (.006)	.033 (.006)
$\sigma^2(i)$	.028 (.016)	.034 (.020)
$\sigma^2(pi)$	.002 (.001)	-
$\sigma^2(e)$	.181 (.014)	-
$\sigma^2(pi,e)$	.002+.181	.183 (.013)
$E\hat{p}^2$	.474 (.037)	.471 (.037)
Φ	.439 (.048)	.430 (.052)

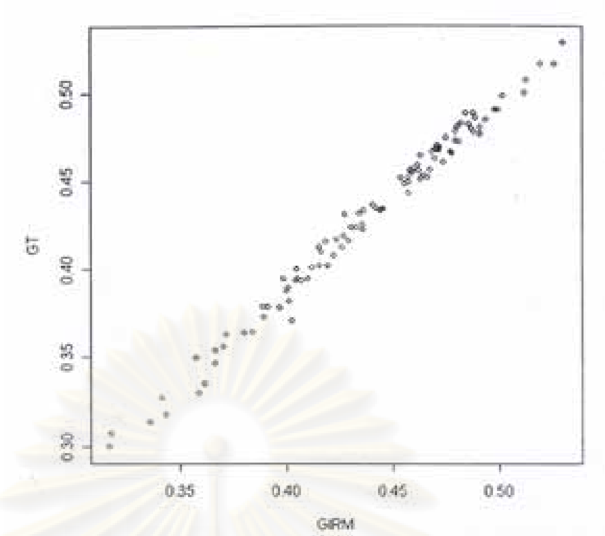
*** จากการจำลองข้อมูลซ้ำ 100 ครั้ง**

ผลการศึกษา พบว่า ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานขององค์ประกอบความแปรปรวนจากการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ให้ผลการวิเคราะห์ที่ใกล้เคียงกันมาก นอกจากนี้การวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) สามารถวิเคราะห์ความแปรปรวนแยกระหว่างองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคล ($\sigma^2(pi)$) กับข้อสอบกับความคลาดเคลื่อนแบบสุ่ม ($\sigma^2(e)$) ได้ ในขณะที่การวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ไม่สามารถแยกองค์ประกอบดังกล่าวได้ โดยคำนวณรวมองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบกับความคลาดเคลื่อนแบบสุ่ม ($\sigma^2(pi,e)$) เข้าไว้ด้วยกัน

เมื่อพิจารณาค่าสัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient) ระหว่างการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) และทฤษฎี การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) พบว่า ผลจากการวิเคราะห์ให้ค่าใกล้เคียงกัน โดยค่าสัมประสิทธิ์การสรุปอ้างอิงแบบสัมพันธ์ โดยได้จากการคำนวณตามทฤษฎีการสรุปอ้างอิง ความน่าเชื่อถือของผลการวัด (GT) มีค่าเท่ากับ .474 และจากการคำนวณตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) มีค่าเท่ากับ .471 สำหรับค่าสัมประสิทธิ์การสรุปอ้างอิงแบบสัมบูรณ์ โดยได้จากการคำนวณด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) มีค่าเท่ากับ .439 และจากการคำนวณตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) มีค่าเท่ากับ .430 และเมื่อนำมาสร้างกราฟแบบจุด (scatter plots) แสดงความสัมพันธ์ของค่าสัมประสิทธิ์การสรุปอ้างอิงแบบสัมพันธ์ และแบบสัมบูรณ์ระหว่างการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ดังแสดงในแผนภาพที่ 5 และ 6 พบว่า ผลที่ได้จากการวิเคราะห์ทั้งสองวิธีมีความสัมพันธ์เชิงเส้นแบบสมบูรณ์ (perfect linear association) โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ 1



แผนภาพที่ 6 ความสัมพันธ์ของการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงแบบสัมพันธ์ ระหว่างการวิเคราะห์ด้วยวิธี GIRM และ GT (Briggs และ Wilson, 2007)



แผนภาพที่ 7 ความสัมพันธ์ของการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงแบบสัมบูรณ์ระหว่างการวิเคราะห์ด้วยวิธี GIRM และ GT (Briggs และ Wilson, 2007)

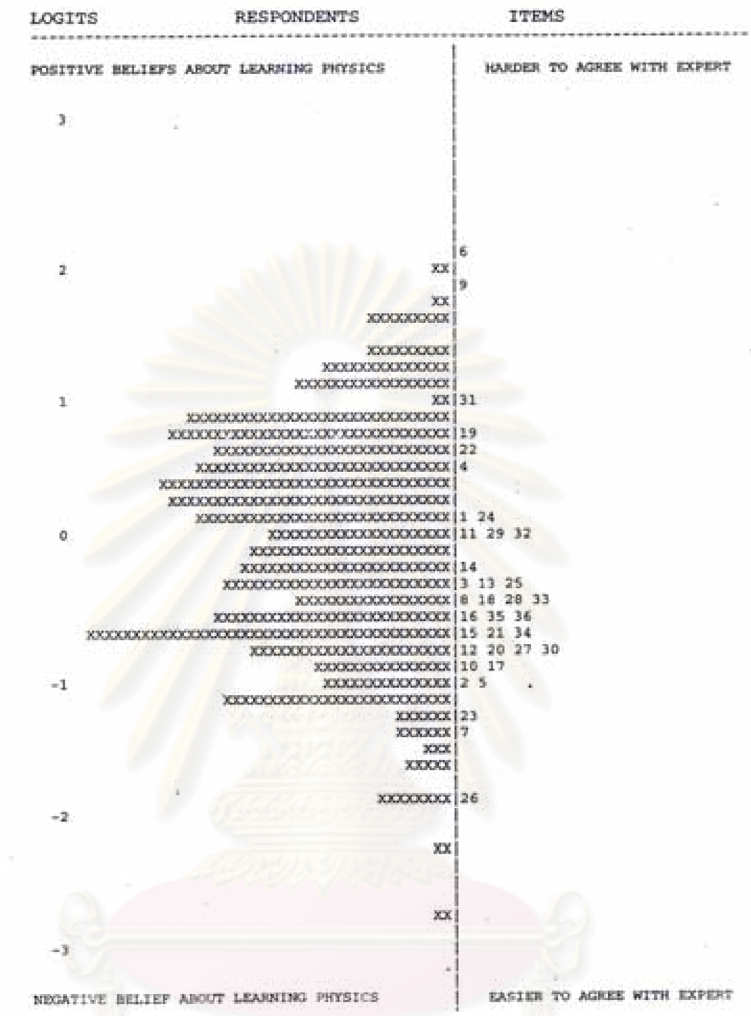
2) การศึกษาจากการเก็บรวบรวมข้อมูลจริง (Real data)

ในการนำเสนอผลการวิเคราะห์จากการเก็บรวบรวมข้อมูลจริง ได้ออกแบบการวิเคราะห์ในลักษณะรูปแบบ pXi โดย Briggs และ Wilson (2007:142-143) ได้ทำการเก็บรวบรวมข้อมูลจริงในโครงการ Colorado Learning Attitude about Science Survey (CLASS) ซึ่งเป็นโครงการเกี่ยวกับการศึกษาความเชื่อของนักศึกษาที่มีต่อการเรียนในวิชาฟิสิกส์เบื้องต้น เก็บรวบรวมข้อมูลจากนักศึกษาปริญญาตรี 349 คนที่เข้าเรียนในวิชาฟิสิกส์ ที่มหาวิทยาลัยโคโรลาโด (University of Colorado) ประเทศสหรัฐอเมริกา แบบสอบถามมีทั้งหมด 36 ข้อ ตัวอย่างข้อคำถามเช่น ท่านมีความคิดเห็นอย่างไรเกี่ยวกับประโยคที่ว่า “**ปัญหาที่สำคัญในการเรียนฟิสิกส์คือ ความสามารถในการจำข้อมูลทั้งหมดที่เราจำเป็นต้องรู้**” เป็นต้น แบบสอบถามมีลักษณะแบบ Likert Scale 5 ระดับ ตั้งแต่ไม่เห็นด้วยอย่างยิ่ง จนถึง เห็นด้วยอย่างยิ่ง ในการแปลผลจะถูกแปลงให้เป็นการให้คะแนนแบบสองค่า (1,0) (dichotomous) ถ้ามีความคิดเห็นในทิศทางเดียวกับนักปฏิบัติที่เชี่ยวชาญด้านฟิสิกส์ให้ 1 คะแนน และถ้าไม่สอดคล้องในทิศทางเดียวกับนักปฏิบัติที่เชี่ยวชาญด้านฟิสิกส์ให้0 คะแนน

การนำเสนอผลการวิเคราะห์สามารถนำเสนอผลการวิเคราะห์ได้ใน 2 รูปแบบ คือ การนำเสนอด้วยแผนภาพที่ 8 ซึ่งเป็นการนำเสนอในลักษณะของสารสนเทศที่ได้จากการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) และการนำเสนอในตารางที่ 5 ซึ่งเป็นการนำเสนอในลักษณะของสารสนเทศที่ได้จากการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยผลการวิเคราะห์ในลักษณะของทฤษฎีการตอบสนองข้อสอบ (IRT) สามารถแสดงได้ด้วยแผนภาพ Wright map ซึ่งเป็นแผนภาพที่นำเสนอในลักษณะของการวิเคราะห์ข้อสอบด้วยโมเดลการตอบสนองข้อสอบ (IRT) ที่ Wilson (2005) ได้นำเสนอไว้ในหนังสือ Constructing measures: An item response modeling approach ลักษณะในการนำเสนอแบบแผนภาพ จะประมาณค่าพารามิเตอร์ของผู้สอบ (θ_p) ของกับความยากของข้อสอบ (β_i) ในแกนของเส้นจำนวนเดียวกันในรูปของคะแนนมาตรฐาน ดังนั้นในการศึกษาครั้งนี้แผนภาพ Wright map จึงได้นำเสนอความสัมพันธ์ระหว่างตำแหน่งของความเชื่อของนักเรียนเกี่ยวกับการเรียนในวิชาฟิสิกส์กับระดับความยากของข้อคำถามตามที่นักเรียนตอบ ดังแสดงไว้ในแผนภาพที่ 7

เมื่อพิจารณาแผนภาพที่ 8 พบว่า ในภาพรวมแบบวัดในโครงการ CLASS ง่าย นั้นคือ นักศึกษาส่วนใหญ่มีความเชื่อต่อการเรียนในวิชาฟิสิกส์ที่สอดคล้องกับนักปฏิบัติที่เชี่ยวชาญด้านฟิสิกส์ โดยมี ข้อคำถามจำนวน 25 ข้อจากข้อคำถามทั้งหมด 36 ข้อ ในข้อคำถามที่มีระดับความยาก 0 ถึง -1 (ข้อ 1 24 11 29 32 14 3 13 25 8 18 38 33 16 35 36 15 21 34 12 20 27 30 10 และ 17) ในทางตรงกันข้ามมีข้อคำถาม 3 ข้อ ที่ยากเกินไป (6 9 และ 31) และมีข้อคำถามอีก 3 ข้อที่ง่ายเกินไป (23 7 และ 26) ผลที่ได้นี้นับว่าสำคัญสำหรับการออกแบบเครื่องมือสำหรับการเก็บรวบรวม

นอกจากนี้ยังพบว่า การกระจายตัวของคะแนนจำนวนมากในข้อที่มีความยาก-ง่ายในระดับ 0 ถึง 2 ซึ่งมีข้อคำถามอยู่ในช่วงนี้เพียงเล็กน้อย ในขณะเดียวกันมีการกระจายตัวของคะแนนจำนวนน้อยในข้อที่มีความยาก-ง่ายในระดับ 0 ถึง -1 ทั้งๆที่มีข้อคำถามอยู่ในช่วงนี้เป็นจำนวนมาก ซึ่งส่งผลต่อคุณภาพและความถูกต้องของผลการวัด



แผนภาพที่ 8 ผลการวิเคราะห์ด้วย Wright map จากข้อมูลโครงการ CLASS (Briggs และ Wilson, 2007)

สำหรับผลการวิเคราะห์ในลักษณะของ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) จะให้ผลการวิเคราะห์ในลักษณะค่าเฉลี่ยขององค์ประกอบความแปรปรวน (variance component) จากแหล่งความแปรปรวนต่างๆ และค่าสัมประสิทธิ์การสรุปอ้างอิง (generalizability coefficient) ทั้งในเชิงสัมบูรณ์และในเชิงสัมพันธ์ โดยการสรุปผลการวิเคราะห์ข้อมูลได้ดังนี้

ตารางที่ 5 องค์ประกอบความแปรปรวนด้วยการวิเคราะห์ GT และ GIRM จากการเก็บรวบรวมข้อมูลโครงการ CLASS (กรณี 1 facet p x i design)

องค์ประกอบความแปรปรวน / ค่าสัมประสิทธิ์การสรุปอ้างอิง	**ค่าเฉลี่ยที่ได้จากวิธี GIRM หลังจากการจำลองข้อมูล**	**การประมาณค่าเฉลี่ยที่ได้จากวิธี GT**
$\sigma^2(p)$	.030	.031
$\sigma^2(i)$	.025	.025
$\sigma^2(pi)$	.001	-
$\sigma^2(e)$	.190	-
$\sigma^2(pi,e)$	.001+.190 = .191	.192
$E\hat{p}^2$	.851	*
Φ	.834	*

* ไม่มีผลการคำนวณจากโปรแกรม urGENOVA แต่สามารถคำนวณได้จากสูตรของ Brennan (2001)

เมื่อพิจารณาตารางที่ 5 พบว่า ความแปรปรวนของคะแนนสอบของผู้สอบแต่ละคนในแต่ข้อ (X_{Pi}) พบว่า มีสัดส่วนความแปรปรวนของความคลาดเคลื่อนสูงมากคือ 77 % และเป็นความแปรปรวนที่มาจากข้อคำถามเพียงแค่ 10 % นอกจากนี้ยังสามารถพิจารณาสัดส่วนความแปรปรวนที่มีปฏิสัมพันธ์ระหว่างบุคคลและข้อสอบด้วย เท่ากับ .4% ซึ่งนับว่าน้อยมากจึงสามารถจะละเลยในเรื่องปฏิสัมพันธ์ระหว่างบุคคลและข้อสอบได้

นอกจากนี้ยังพบอีกว่า การวิเคราะห์ข้อมูลจริงจากเครื่องมือ CLASS จำนวน 36 ข้อ สามารถวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิง (generalizability coefficient) สัมพันธ์และสัมบูรณ์ ซึ่งมีค่าเท่ากับ .85 และ .83 ตามลำดับ ทำให้สามารถสรุปได้ว่าคะแนนที่ได้มีความสามารถในการสรุปอ้างอิงไปยังเอกภพของข้อสอบสูง ในขณะเดียวกันบางครั้งที่เรามีเวลาจำกัดในการสอบ เราสามารถลดจำนวนข้อสอบเหลือเพียง 18 ข้อ ก็จะทำให้ได้ค่าสัมประสิทธิ์ในการสรุปอ้างอิงสัมพันธ์และสัมบูรณ์เหลือเท่ากับ .74 และ .71 ตามลำดับ

1.3.6 จุดเด่นสำหรับการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM)

1. การวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ให้ทั้งผลการวิเคราะห์ทั้งแบบทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และแบบทฤษฎีการตอบสนองข้อสอบ (IRT) โดยการวิเคราะห์แบบทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) –ให้ผลการวิเคราะห์ทั้งองค์ประกอบความแปรปรวน (variance component) และค่าสัมประสิทธิ์การสรุปอ้างอิง (generalizability coefficient) ทำให้สามารถใช้ประโยชน์ได้เชิงการคาดการณ์การออกแบบของจำนวนเงื่อนไขขั้นต่ำในแต่ละฟาเซต ในขณะที่ การวิเคราะห์แบบทฤษฎีการตอบสนองข้อสอบ (IRT) –ให้ผลประมาณค่าพารามิเตอร์ของผู้สอบและค่าพารามิเตอร์ของข้อสอบ โดยผลการวิเคราะห์สามารถช่วยเสริมผลจากการวิเคราะห์ด้วยวิธี ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ซึ่งให้ข้อมูลที่เป็นประโยชน์เกี่ยวกับคุณภาพเงื่อนไขของฟาเซต ดังนั้นจะเห็นได้ว่า การวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) สามารถให้ผลการตีความผลการวิเคราะห์ได้ทั้งในรูปแบบขององค์ประกอบการสุ่ม (the sampling component typically) ซึ่งวิเคราะห์ได้จากวิธีทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และในรูปแบบขององค์ประกอบในการวัด (the scaling component typically) ซึ่งวิเคราะห์ได้จากทฤษฎีการตอบสนองข้อสอบ (IRT)

ยกตัวอย่างเช่น ในการวัดผลสัมฤทธิ์ทางการเรียนในวิชาคณิตศาสตร์โดยใช้แบบสอบจำนวน 30 ข้อ ในการวิเคราะห์ข้อสอบเมื่อวิเคราะห์ตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ทำให้เราสามารถพิจารณาองค์ประกอบของความคลาดเคลื่อน (Facet) ได้จำนวน1 แหล่ง คือ จำนวนข้อสอบ ผลการวิเคราะห์ตามแนว GT จากสถานการณ์ดังกล่าวนี้ จะทำให้เราทราบว่า องค์ประกอบของความแปรปรวนของคะแนนที่สังเกตได้ มาจากความแปรปรวน อันเนื่องมาจากสามารถของบุคคล (p) ข้อสอบ (i) และปฏิสัมพันธ์ระหว่างบุคคล ข้อสอบและความคลาดเคลื่อนแบบสุ่ม (pi,e) แต่ละแหล่งจำนวนเท่าใด นอกจากนี้ยังสามารถพิจารณาผล การวิเคราะห์จาก D- study ซึ่งทำให้ทราบว่า ข้อสอบจำนวนเท่าใดที่จะมีค่าสัมประสิทธิ์สรุปอ้างอิงที่เหมาะสม เพื่อการกำหนดจำนวนข้อสอบที่เหมาะสมสำหรับการทดสอบและให้ผลการทดสอบที่พึงพอใจ ส่วนการวิเคราะห์ข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ (IRT) จะทำให้เราทราบว่าข้อสอบในแต่ละข้อมีคุณภาพมากน้อยเพียงใด โดยพิจารณาได้จากค่าพารามิเตอร์ของข้อสอบในแต่ละข้อ ดังนั้นจะเห็นได้ว่า ถ้าเราทำการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ซึ่งจะให้ผลการวิเคราะห์ของทั้งสองวิธี ซึ่งจะทำให้เราทราบว่าจำนวน

ข้อสอบที่เหมาะสมสำหรับการวัดนั้นมีจำนวนเท่าใด และข้อสอบในแต่ละข้อมีคุณภาพมากน้อยเพียงใด การวิเคราะห์ด้วยวิธี GIRM จึงให้ประโยชน์ทั้งในชิงการคาดการณ์การออกแบบขั้นต่ำของจำนวนเงื่อนไขของฟาเซตและยังสามารถให้ข้อมูลที่เป็นประโยชน์เกี่ยวกับคุณภาพเงื่อนไขของฟาเซตที่ทำการวิเคราะห์ได้อีกด้วย

2. การวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) สามารถแยกการประมาณค่าความแปรปรวนขององค์ประกอบความแปรปรวนระหว่างความคลาดเคลื่อนของปฏิสัมพันธ์ของแหล่งความคลาดเคลื่อน (facet interaction effect) และความคลาดเคลื่อน (error effect) ออกจากกันได้ ซึ่งในการวิเคราะห์ข้อสอบด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ไม่สามารถแยกความแปรปรวนดังกล่าวออกจากกันได้ในการวิเคราะห์แยกองค์ประกอบความแปรปรวนนี้ทำให้ผู้วิเคราะห์ทราบว่าสามารถละเลยการพิจารณาปฏิสัมพันธ์ระหว่างองค์ประกอบของความแปรปรวนได้หรือไม่ ถ้ามีปฏิสัมพันธ์ระหว่างองค์ประกอบก็จะไม่สามารถพิจารณาเฉพาะองค์ประกอบหลักองค์ประกอบใดองค์ประกอบหนึ่งได้ แต่จะต้องทำให้พิจารณาทั้งสององค์ประกอบที่มีปฏิสัมพันธ์ร่วมกันจึงจะทำให้ผลการวัดมีความถูกต้องมากยิ่งขึ้น

จากตัวอย่างในสถานการณ์เดิมในข้อที่ 1 ในการวิเคราะห์ข้อสอบเมื่อวิเคราะห์ตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยพิจารณาองค์ประกอบของความคลาดเคลื่อน (Facet) 1 แหล่ง ได้แก่ จำนวนข้อสอบ ผลการวิเคราะห์ตามแนว GT จากสถานการณ์ดังกล่าวนี้ จะทำให้เราทราบว่า ความแปรปรวนของคะแนนที่สังเกตได้ เป็นผลมาจากความแปรปรวนอันเนื่องมาจากสามารถของบุคคล (p) ข้อสอบ (i) และปฏิสัมพันธ์ระหว่างบุคคล ข้อสอบและความคลาดเคลื่อนแบบสุ่ม (pi,e) แต่ละตัวจำนวนเท่าใด แต่เรายังไม่สามารถแยกผลการวิเคราะห์ความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคล ข้อสอบและความคลาดเคลื่อนแบบสุ่ม (pi,e) ได้ ดังนั้นจึงไม่สามารถทราบได้ว่าแท้ที่จริงแล้วนั้นความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบ (pi) และความแปรปรวนของความคลาดเคลื่อนแบบสุ่ม (e) มีจำนวนเท่าใด แต่สำหรับการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) สามารถวิเคราะห์แยกปฏิสัมพันธ์ระหว่างบุคคล กับ ข้อสอบ (pi) และความแปรปรวนของความคลาดเคลื่อนแบบสุ่ม (e) ได้ ดังนั้นจึงทำให้ทราบว่าแท้ที่จริงแล้วนั้น องค์ประกอบของความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคล กับ ข้อสอบ (pi) มีจำนวนเท่าใด ถ้ามีจำนวนมากพออย่างมีนัยสำคัญก็จะทำให้ไม่สามารถพิจารณาแยกเฉพาะองค์ประกอบหลักองค์ประกอบใดองค์ประกอบหนึ่ง นั่นก็คือ ไม่สามารถแยกพิจารณาองค์ประกอบที่เกิดจากบุคคล (p) และองค์ประกอบที่เกิดจากข้อสอบ (i)

ได้ แต่จะต้องทำให้พิจารณาทั้งสององค์ประกอบที่มีปฏิสัมพันธ์ร่วมกันจึงจะทำให้ผลการวัดมีความถูกต้องมากยิ่งขึ้นตามหลักการวิเคราะห์ความแปรปรวน (Analysis of Variance ; ANOVA)

3. การวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นการวิเคราะห์ของฟังก์ชันของการประมาณค่า (a function of expected) มากกว่าเป็นฟังก์ชันของคะแนนที่สังเกตได้ (observed response) จึงทำให้สามารถออกแบบการวัดครบถ้วน และสมบูรณ์มากกว่าการวิเคราะห์จากคะแนนที่สังเกตได้ซึ่งอาจจะมีข้อมูลที่คาดหายไปได้

เมื่อพิจารณาแผนภาพที่ 2 เมทริกซ์คะแนนที่สังเกตได้ของผู้สอบตามแนว GT ซึ่งเป็นเมทริกซ์ที่ได้จากคะแนนการสอบของบุคคลในแต่ละข้อ ดังนั้นในการวิเคราะห์จึงต้องมีการเก็บรวบรวมข้อมูลจริงซึ่งเป็นคะแนนที่สังเกตได้มาทำการวิเคราะห์ตามวิธีการวิเคราะห์ด้วยทฤษฎี GT แต่สำหรับการวิเคราะห์ด้วยวิธี GIRM เมื่อพิจารณาแผนภาพที่ 3 เมทริกซ์คะแนนที่คาดหวังของผู้สอบด้วยวิธี GIRM พบว่า ข้อมูลที่ได้มานั้นเป็นคะแนนที่คาดหวังสำหรับในการวิเคราะห์ ซึ่งประมาณค่าได้จากวิธีการประมาณค่าแบบ MCMC ดังนั้นจะเห็นได้ว่า คะแนนที่ได้จากการประมาณค่าจึงน่าจะมีความสมบูรณ์มากกว่าคะแนนที่ได้จากกการเก็บรวบรวมข้อมูลจริงซึ่งอาจจะมีข้อมูลที่คาดหายไปได้ (missing-value)

4. การวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ไม่ต้องมีการปรับข้อมูลที่ขาดหายไป (missing data) ในเมทริกซ์ของข้อมูล ในขณะที่ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ต้องมีการออกแบบข้อมูลแทนที่ขาดหายไป ทั้งนี้เนื่องจากการประมาณค่าองค์ประกอบความแปรปรวนตั้งอยู่บนพื้นฐานของการคาดคะเน (expected) มากกว่าข้อมูลจริงที่อยู่ในตารางเมทริกซ์ของมูล จึงทำให้ส่งผลต่อการวิเคราะห์ข้อมูลเนื่องจากการลงข้อมูล (specified) ที่ไม่สมดุลสำหรับการออกแบบแบบสุ่ม (unbalanced random effects design) จะไม่สามารถประมาณค่า G- Study และ D- study ในโปรแกรม GENOVA ได้

1.3.7 ข้อจำกัดสำหรับการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM)

1. การวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ขาดโปรแกรมสำเร็จรูปสำหรับการวิเคราะห์โดยเฉพาะในปัจจุบันสามารถวิเคราะห์ได้โดยประยุกต์ใช้โปรแกรม WinBUGS 1.4 ในการประมาณค่า posterior distributions สำหรับการประมาณค่าพารามิเตอร์ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ซึ่งโปรแกรมนี้สามารถวิเคราะห์ได้ไม่ยากสำหรับการวิเคราะห์แบบ Simple item response function และข้อมูลแบบ

Dichotomous data แต่จะวิเคราะห์ยากมากเมื่อใช้โมเดลการตอบสนองข้อสอบที่ซับซ้อนมาก (More complex item response model) และข้อมูลแบบ polytomous data นอกจากนี้โปรแกรม WinBUGS เป็นโปรแกรมที่ใช้เวลามากสำหรับในการศึกษาครั้งนี้ เช่นในการวิเคราะห์ข้อมูลจำนวนตัวอย่าง 500 คน 20 ข้อคำถามใช้เวลาทั้งสิ้น 4 ชั่วโมงสำหรับคอมพิวเตอร์ทั่วไป (PC Pentium 4, 3 GHz CPU, 1 GB RAM)

2. การวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) โดยใช้โปรแกรม WinBUGS มีแนวโน้มทำให้เกิดความลำเอียงโดยเกิด underestimator spread ในการกระจายเนื่องมาจาก การป้อนคำสั่งในกระบวนการการประมาณค่าองค์ประกอบความคลาดเคลื่อนของขั้นตอนในการประมาณด้วยวิธี Markov Chain ซึ่งการประมาณค่าความแปรปรวนขององค์ประกอบไม่สามารถทำให้เป็นอิสระจากกันได้ จึงทำให้นำไปสู่การฝ่าฝืนข้อตกลงเบื้องต้นต่อผลกระทบแบบสุ่ม (random effects) ดังนั้นในการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) จึงควรมีสมมติฐานที่แข็งเกร่งมากกว่า (a stronger set of assumptions) การวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

1.3.8 สรุปสาระสำคัญของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ

วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นวิธีการวิเคราะห์ข้อสอบที่รวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการกำหนดมาตรวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) โดยอาศัยสมมติฐานของการกระจายไปยังแหล่งความคลาดเคลื่อนจากการวัดที่เกี่ยวข้องโดยการประมาณค่า GT ในแต่ละองค์ประกอบของความแปรปรวน (Variance Components) ไปพร้อมกับการประมาณค่าพารามิเตอร์ ในโมเดลการตอบสนองข้อสอบ (IRT) ด้วยวิธีการประมาณค่า MCMC (Markov Chain Monte Carlo) โดยใช้โปรแกรม WinBUGS ในการวิเคราะห์ ดังนั้นข้อตกลงเบื้องต้นของทฤษฎีนี้จึงต้องมีทั้งข้อตกลงเบื้องต้นของทั้งสองทฤษฎีคือข้อตกลงเบื้องต้นของการสุ่มตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และข้อตกลงเบื้องต้นตามทฤษฎีการตอบสนองข้อสอบ (IRT) รวมกับข้อตกลงเบื้องต้นของการกระจายแหล่งความคลาดเคลื่อนของการวัด

ดังนั้นจะเห็นได้ว่าการวิเคราะห์ด้วยวิธีนี้จะให้ผลการวิเคราะห์ทั้งแบบทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และแบบทฤษฎีการตอบสนองข้อสอบ (IRT) นอกจากนี้ยังสามารถแยกการประมาณค่าความแปรปรวนขององค์ประกอบความแปรปรวนระหว่าง ความคลาดเคลื่อนของปฏิสัมพันธ์ของแหล่งความคลาดเคลื่อน (facet interaction effect) และความคลาดเคลื่อน (error effect) ออกจากกันได้ ในขณะที่วิธีการวิเคราะห์แบบ GT ไม่สามารถแยกความแปรปรวนดังกล่าวออกจากกันได้ นอกจากนี้การวิเคราะห์ด้วย GIRM เป็นการวิเคราะห์ของฟังก์ชันของการประมาณค่า (a function of expected) มากกว่าเป็นฟังก์ชันของคะแนนที่สังเกตได้ (observed response) จึงทำให้สามารถออกแบบการวัดครบถ้วนและสมบูรณ์มากกว่าการวิเคราะห์จากคะแนนที่สังเกตได้ซึ่งอาจจะมีข้อมูลที่คาดหายไปได้ โดยไม่ต้องมีการปรับข้อมูลที่ขาดหายไปในเมทริกซ์ของข้อมูล แต่การวิเคราะห์ด้วยวิธีการนี้ยังขาดโปรแกรมสำเร็จรูปสำหรับการวิเคราะห์โดยเฉพาะในปัจจุบัน สามารถวิเคราะห์ได้ โดยประยุกต์ใช้โปรแกรม WinBUGS 1.4 ในการประมาณค่า posterior distributions สำหรับการประมาณค่าพารามิเตอร์ ซึ่งใช้เวลามากและยังให้ผลการวิเคราะห์ที่มีแนวโน้มทำให้เกิดความลำเอียงโดยเกิด underestimator spread ในการกระจายเนื่องมาจากการป้อนคำสั่งในกระบวนการการประมาณค่าองค์ประกอบความคลาดเคลื่อนของขั้นตอนในการประมาณด้วยวิธี Markov Chain ซึ่งการประมาณค่าความแปรปรวนขององค์ประกอบไม่สามารถทำให้เป็นอิสระจากกันได้

โมเดลการวิเคราะห์ข้อสอบเป็นโมเดลที่มุ่งวัดหรืออธิบายเพื่อให้สามารถสรุปคะแนนจริงหรือความสามารถที่แท้จริง (true score) ของบุคคลที่ทำการทดสอบได้ ธรรมชาติของคะแนนจริงเป็นคุณลักษณะแฝง (latent proficiency) ที่ไม่สามารถวัดหรือสังเกตได้โดยตรง (unobservable) ดังนั้นจึงต้องมีโมเดลการวิเคราะห์เพื่อให้สามารถประมาณค่าของคะแนนจริงได้ตามนิยามหรือสมการทางคณิตศาสตร์ที่แต่ละโมเดลกำหนดขึ้น ในการประมาณค่าของแต่ละโมเดลนั้น นักวัดและประเมินผลพิจารณาผลของการประมาณในแต่ละครั้งว่ามีคุณภาพหรือไม่ จากค่าความเที่ยง (reliability) หรือความคงเส้นคงวาของผลการวัด ซึ่งสามารถแสดงให้เป็นปริมาณได้จาก ค่าสัมประสิทธิ์สหสัมพันธ์ (coefficient) หรือค่าความคลาดเคลื่อนมาตรฐานในการวัด (standard of measurement) ตามแต่ละมโนทัศน์หรือสมการทางคณิตศาสตร์ในแต่ละโมเดลของการวัด (Brennan, 2007)

1.4 บทสรุปจากการศึกษาแนวคิด/ทฤษฎีเกี่ยวกับการวิเคราะห์ข้อสอบ

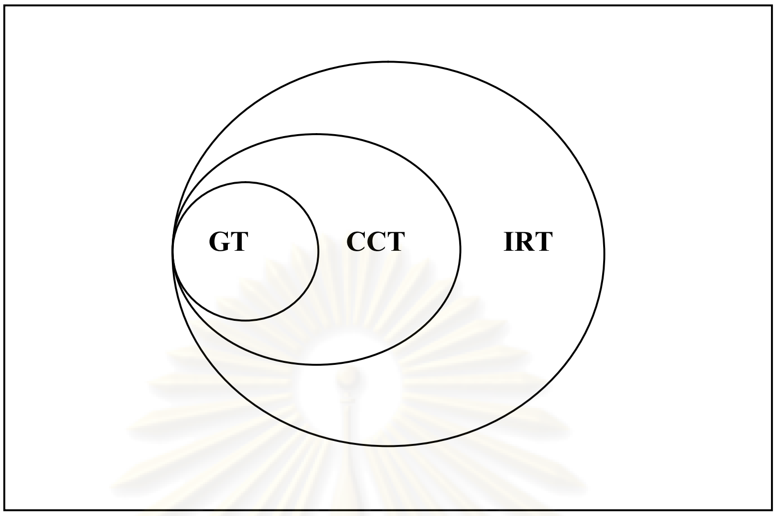
โมเดลหรือทฤษฎีการทดสอบในปัจจุบันสามารถแบ่งออกเป็น 2 ประเภทหลักๆ คือ ประเภทแรก ได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ซึ่งมีแนวคิดว่าคะแนนสอบที่สังเกตได้ (X) นั้นเป็นผลรวมมาจากโมเดลเชิงเส้นของผลรวมคะแนนจริง (T) และความคลาดเคลื่อน (ε) โดยความคลาดเคลื่อนนั้นมีลักษณะเป็นหนึ่งเดียวที่ไม่สามารถแบ่งแยกได้ (unique error) ซึ่งทำให้เป็นข้อจำกัดที่สำคัญสำหรับการค่าความเที่ยงของแบบวัดที่สามารถวิเคราะห์แหล่งความคลาดเคลื่อนได้ที่ละแหล่งเท่านั้น นอกจากนี้ในการทดสอบยังตั้งอยู่บนพื้นฐานข้อตกลงเบื้องต้นที่สำคัญ คือ ค่าพารามิเตอร์ของข้อสอบเป็นค่าเฉพาะของกลุ่มผู้สอบและเป็นอิสระไม่ขึ้นอยู่กับความสามารถของผู้สอบ จึงทำให้ค่าความคลาดเคลื่อนมาตรฐานของการวัด (SEM) เป็นค่าเฉพาะในแต่ละประชากรที่ใช้ในการทดสอบและเป็นค่าเดียวคงที่สำหรับทุกคนในประชากร การเปรียบเทียบคะแนนหรือคุณภาพของข้อสอบ จึงทำได้เฉพาะข้อสอบที่นำมาเปรียบเทียบกันมีคุณลักษณะเป็นข้อสอบคู่ขนานเท่านั้น จากข้อตกลงเบื้องต้นดังกล่าวจึงเป็นข้อจำกัดที่สำคัญของทฤษฎี ซึ่งทำให้ค่าพารามิเตอร์ของข้อสอบผันแปรตามกลุ่มผู้สอบ นอกจากนี้คะแนนที่สังเกตที่ได้จากการสอบหรือความสามารถของผู้สอบขึ้นอยู่กับข้อสอบและแบบสอบที่นำมาใช้ ดังนั้นทฤษฎีการทดสอบแบบดั้งเดิมจึงมีค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ของผู้สอบไม่เป็นอิสระต่อกันโดยขึ้นอยู่กับสถานการณ์แต่ละสถานการณ์ในการสอบ (ศิริชัย กาญจนวาสี, 2550)

ส่วนทฤษฎีประเภทที่สองของทฤษฎีการทดสอบในปัจจุบัน ได้แก่ ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory) ซึ่งทฤษฎีการทดสอบแนวใหม่ในปัจจุบันที่ใช้กันอย่างแพร่หลายมีอยู่ด้วยกัน 2 ทฤษฎี คือ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory: GT) และทฤษฎีการตอบสนองข้อสอบ (Item test Theory: IRT) ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เป็นทฤษฎีที่อธิบายโดยใช้วิธีการวิเคราะห์ความแปรปรวน (ANOVA) เข้ามาช่วยในการอธิบายแหล่งของความคลาดเคลื่อนที่เกิดจากหลายแหล่ง (Multiple sources of error) ทำให้สามารถแยกส่วนความคลาดเคลื่อนของการวัดได้ออกเป็นสองแหล่งหลัก ๆ ประกอบด้วยความคลาดเคลื่อนอย่างเป็นระบบ (systematic error variance) และความคลาดเคลื่อน แบบสุ่ม (random error variance) รวมกันเป็นความคลาดเคลื่อน (ε) ของทฤษฎีการทดสอบแบบดั้งเดิม ดังนั้นคะแนนสอบของผู้สอบที่ได้จากการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิง จึงเป็นคะแนนที่คาดหวังของบุคคล (Person 's expected score) ไม่ใช่คะแนนที่สังเกตได้ (Person 's observed score) เนื่องจากคะแนนตามทฤษฎีนี้เป็นคะแนนที่ได้มาจากการสุ่มตามเงื่อนไขของแหล่งความคลาดเคลื่อนที่น่าจะเกิดขึ้น (Randomly sampled facet conditions) ส่วนทฤษฎีการตอบสนอง

ข้อสอบ (IRT) เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่แท้จริงของบุคคลกับพฤติกรรมการตอบสนองข้อสอบในแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกได้มากน้อยเพียงใด โดยใช้โค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชันโลจิส (Logistic function) หรือใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function)

ดังนั้นจะเห็นได้ว่า การวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) นั้นเป็นการวิเคราะห์ที่สามารถอธิบายแหล่งความคลาดเคลื่อนของการวัดได้มากยิ่งขึ้น และสามารถสรุปอ้างอิงไปยังสถานการณ์ต่างๆ ของการวัดในแต่ละแหล่งความคลาดเคลื่อนได้อีกด้วย ซึ่งอาจเรียกทฤษฎีนี้ได้ว่าเป็นโมเดลแบบสุ่ม (Sampling Model) แต่ทฤษฎีไม่สามารถให้ค่าพารามิเตอร์ของข้อสอบเป็นรายข้อและพารามิเตอร์ของผู้สอบเป็นรายบุคคลได้ ในขณะที่ทฤษฎีการตอบสนองข้อสอบ (IRT) สามารถให้ทั้งสารสนเทศที่เป็นค่าพารามิเตอร์ของข้อสอบเป็นรายข้อ (item parameter) และพารามิเตอร์ของผู้สอบเป็นรายบุคคลได้ ซึ่งอาจเรียกทฤษฎีนี้ได้ว่าเป็นโมเดลการวัด (Scaling Model)

เมื่อพิจารณาเปรียบเทียบทฤษฎีการวิเคราะห์ข้อสอบในปัจจุบันทั้ง 3 ทฤษฎีทั้งในเชิงสมการทางคณิตศาสตร์และมโนทัศน์ของแต่ละทฤษฎี โดยสมการทางคณิตศาสตร์สามารถอธิบายโครงสร้างของทฤษฎีในแต่ละทฤษฎีได้ ในขณะที่เมื่อพิจารณาจากมโนทัศน์ของแต่ละทฤษฎีจะสามารถอธิบายเชิงความหมายและข้อสรุปของในแต่ละทฤษฎีได้ จากการพิจารณาทั้งในเชิงสมการทางคณิตศาสตร์และมโนทัศน์แต่ละทฤษฎีทำให้สามารถสรุปได้ว่า ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เป็นเพียงกรณีเฉพาะกรณีหนึ่งของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) และทฤษฎีการทดสอบแบบดั้งเดิม (CTT) เป็นเพียงกรณีเฉพาะกรณีหนึ่งของทฤษฎี การตอบสนองข้อสอบ (IRT) (Holland และ Hoskens, 2003) โดยสามารถสรุปเป็นความสัมพันธ์ได้ดังแผนภาพต่อไปนี้



แผนภาพที่ 9 ความสัมพันธ์ระหว่าง CCT GT และ IRT

จากจุดเด่นและข้อจำกัดในแต่ละทฤษฎีทำให้นักวัดผลจึงมีความพยายามในการรวมทั้งสองทฤษฎีเข้าไว้ด้วยกันเพื่อรวมจุดเด่นและลดข้อจำกัดในแต่ละทฤษฎี (Kolen และ Harris, 1987; Mislevy, 1993; Verhelst และ Verstralen, 2001; Patz และคณะ, 2002; Bock, Brennan และ Muraki, 2002; Glas และ Van Der Linder, 2003; Brennan, 2006; Briggs และ Wilson, 2007) โดยเริ่มแรก Kolen และ Harris (1987) ได้คิดค้นการเชื่อมโยงสูตรสำหรับการวิเคราะห์องค์ประกอบความแปรปรวนและค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัดโดย การประมาณได้จากการวิเคราะห์ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบหลายตัวแปร (Multivariate GT) ด้วยกรอบแนวคิดของโมเดลการตอบสนองข้อสอบ (IRT) หลังจากนั้น Mislevy (1993) การประมาณค่าพารามิเตอร์ความสามารถของผู้สอบด้วยการใช้วิธีการ Bayesian โดยการสุ่ม (randomness) เข้าไปในการวิเคราะห์ตามแนวของทฤษฎีการตอบสนองข้อสอบ (IRT) ทั้งสองวิธีนั้นเป็นการผ่อนคลายข้อตกลงเบื้องต้นการไม่สรุปอ้างอิงข้อคำถาม (fixed items) ของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT)

ในการประยุกต์ใช้การเชื่อมโยงการวิเคราะห์ข้อสอบของทฤษฎีทั้งสองในทางปฏิบัติของการออกแบบการวัด พบว่า มีงานวิจัยจำนวน 3 เรื่องของ Verhelst และ Verstralen (2001) Patz และคณะ (2002) และBock, Brennan และ Muraki (2002) โดยการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) ในสถานการณ์ที่มีผู้ประเมินคะแนนการปฏิบัติงานหลายคน ดังนั้นจะเห็นได้ว่า การออกแบบการวัดในการศึกษาของ 3 เรื่องนี้ได้นำไปสู่การเชื่อมโยงระหว่าง ทฤษฎีการสรุป

อ้างอิงความน่าเชื่อถือของผลการวัด (GT) กับทฤษฎีการตอบสนองข้อสอบ (IRT) และยังเป็น การขยายกรอบแนวคิดที่เด่นชัดเกี่ยวกับโมเดลผลกระทบแบบสุ่มพหุระดับ (Hierarchical random effect models) อีกด้วย และในปี 2003 ได้มีการประยุกต์ใช้โปรแกรมคอมพิวเตอร์ตามแนวคิดของการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบเพื่อผ่อนคลายข้อตกลงเบื้องต้นการไม่สรุปอ้างอิงข้อคำถาม (fixed items) ของการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT)

ปัจจุบันได้มีวิธีการรวมระหว่างทฤษฎีการวิเคราะห์ข้อสอบระหว่างทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) กับทฤษฎีการตอบสนองข้อสอบ (IRT) ที่เรียกว่า วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (Briggs และ Wilson, 2007) ซึ่งเป็นการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) โดยอาศัยสมมติฐานของการกระจายไปยังแหล่งความคลาดเคลื่อนจากการวัดที่เกี่ยวข้องโดยการประมาณค่า GT ในแต่ละ องค์ประกอบของความแปรปรวน (Variance Components) ไปพร้อมกับการประมาณค่าพารามิเตอร์ในโมเดลการตอบสนองข้อสอบ (IRT) ด้วยวิธีการ MCMC (Markov Chain Monte Carlo) โดยใช้โปรแกรม WinBUGS ดังนั้นข้อตกลงเบื้องต้นของทฤษฎีนี้จึงต้องมีทั้งข้อตกลงเบื้องต้นของทั้งสองทฤษฎีคือข้อตกเบื้องต้นของการสุ่มตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และข้อตกลงเบื้องต้นตามทฤษฎีการตอบสนองข้อสอบ (IRT) รวมกับข้อตกลงเบื้องต้นของการกระจายแหล่งความคลาดเคลื่อนของการวัด

จากการศึกษาด้วยวิธีการจำลองข้อมูล (Simulation) แล้ววิเคราะห์เปรียบเทียบผลการวิเคราะห์ข้อสอบระหว่างวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) กับทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) พบว่า ผลที่ได้จากการวิเคราะห์ทั้งสองวิธีมีความสัมพันธ์เชิงเส้นแบบสมบูรณ์ (Perfect linear association) โดยมีค่าสัมประสิทธิ์สหสัมพันธ์เข้าใกล้ 1 โดยการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) สามารถวิเคราะห์ความแปรปรวนแยกระหว่างองค์ประกอบ ความแปรปรวนของปฏิสัมพันธ์ระหว่างบุคคล ($\sigma^2(pi)$) กับข้อสอบกับความคลาดเคลื่อนแบบสุ่ม ($\sigma^2(e)$) ได้ ในขณะที่การวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ไม่สามารถแยกองค์ประกอบดังกล่าวได้ โดยคำนวณรวมองค์ประกอบความแปรปรวนของ ปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบกับความคลาดเคลื่อนแบบสุ่ม ($\sigma^2(pi,e)$) เข้าไว้ด้วยกัน (Briggs และ Wilson, 2007)

นอกจากนี้เมื่อศึกษาจากการเก็บข้อมูลจริง (Real data) โดยทำการเก็บรวบรวมข้อมูลจริงจากเครื่องมือ Colorado Learning Attitude about Science Survey (CLASS) เพื่อศึกษาความเชื่อของนักศึกษาเกี่ยวกับการเรียนในวิชาฟิสิกส์ของนักศึกษาระดับปริญญาตรีที่มหาวิทยาลัยโคโรลาโด (Colorado university) ประเทศสหรัฐอเมริกา พบว่า ผลการวิเคราะห์นี้สามารถให้ผลได้ทั้งเชิงผลการวิเคราะห์ที่ได้จากทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยทำให้ทราบเกี่ยวกับปัจจัยที่ส่งผลกระทบต่อความคลาดเคลื่อนในการวัด ซึ่งทำให้ผู้วัดสามารถออกแบบสถานการณ์ได้เหมาะสมกับความต้องการในแต่ละเหตุการณ์ (Quantity of facet conditions) และทำให้การออกแบบการวัดนั้นเกิดประสิทธิภาพสูงที่สุดในแต่ละสถานการณ์ นอกจากนี้ยังทำให้ผลการวิเคราะห์ในเชิงทฤษฎีการตอบสนองข้อสอบ (IRT) ที่เป็นตัวเสริมผลการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิง ความน่าเชื่อถือของผลการวัด (GT) โดยทำให้ทราบข้อมูลเกี่ยวกับคุณภาพของแหล่งความเคลื่อนที่วิเคราะห์ได้ (Quality of facet conditions) จากทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ดังนั้นจะเห็นได้ว่าการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดล การตอบสนองข้อสอบ (GIRM) เป็นการวิเคราะห์ที่ให้ข้อมูลที่ครบถ้วนและครอบคลุมเนื่องจากเป็นการรวมโมเดลการวัดทั้งสองเข้าไว้ด้วยกันในโมเดลการวัดเพียงโมเดลเดียว

ข้อดีของการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) นอกจากให้ทั้งผลการวิเคราะห์ทั้งแบบทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) และแบบทฤษฎีการตอบสนองข้อสอบ (IRT) แล้วการวิเคราะห์ด้วยวิธี GIRM ยังสามารถแยกการประมาณค่าความแปรปรวนขององค์ประกอบความแปรปรวนระหว่าง ความคลาดเคลื่อนของปฏิสัมพันธ์ของแหล่งความคลาดเคลื่อน (facet interaction effect) และความคลาดเคลื่อน (error effect) ออกจากกันได้ ในขณะที่วิธีการวิเคราะห์แบบ GT) ไม่สามารถแยกความแปรปรวนดังกล่าวออกจากกันได้ นอกจากนี้การวิเคราะห์ด้วย GIRM เป็นการวิเคราะห์ของฟังก์ชันของการประมาณค่า (a function of expected) มากกว่าเป็นฟังก์ชันของคะแนนที่สังเกตได้ (observed response) จึงทำให้สามารถออกแบบการวัดครบถ้วนและสมบูรณ์มากกว่าการวิเคราะห์จากคะแนนที่สังเกตได้ซึ่งอาจจะมีข้อมูลที่ขาดหายไปได้ โดยไม่ต้องมีการปรับข้อมูลที่ขาดหายไปในเมทริกซ์ของข้อมูล

เมื่อพิจารณาข้อจำกัดของการวิเคราะห์ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) พบว่า การวิเคราะห์ด้วยวิธีการนี้ยังขาดโปรแกรมสำเร็จรูปสำหรับการวิเคราะห์โดยเฉพาะในปัจจุบันสามารถวิเคราะห์ได้โดยประยุกต์ใช้โปรแกรม WinBUGS 1.4 ในการประมาณค่า posterior distributions สำหรับการประมาณค่าพารามิเตอร์ ซึ่งใช้เวลามากสำหรับใน

การศึกษา ในการวิเคราะห์ข้อมูลจำนวนตัวอย่าง 500 คน 20 ข้อคำถามใช้เวลาทั้งสิ้น 4 ชั่วโมง สำหรับคอมพิวเตอร์ทั่วไป (PC Pentium 4, 3 GHz CPU, 1 GB RAM) นอกจากนี้การวิเคราะห์ด้วยโปรแกรม WinBUGS ยังมีแนวโน้มทำให้เกิดความลำเอียงโดยเกิด underestimator spread ในการกระจายเนื่องมาจาก การป้อนคำสั่งในกระบวนการการประมาณค่าองค์ประกอบความคลาดเคลื่อนของขั้นตอนในการประมาณด้วยวิธี Markov Chain ซึ่งในการประมาณค่าความแปรปรวนขององค์ประกอบไม่สามารถทำให้เป็นอิสระจากกันได้

ตอนที่ 2 การประมาณค่าพารามิเตอร์

การวิเคราะห์ข้อสอบมีจุดมุ่งหมายหลักเพื่อวิเคราะห์หาสารสนเทศของการทดสอบ ซึ่งได้แก่สารสนเทศของข้อสอบและสารสนเทศของผู้สอบ การหาสารสนเทศดังกล่าวจะต้องมี การประมาณค่าพารามิเตอร์ผู้สอบ ซึ่งได้แก่ระดับความสามารถของผู้สอบและประมาณค่าพารามิเตอร์ของข้อสอบซึ่งได้ ค่าความยาก ค่าอำนาจจำแนก และค่าโอกาสในการเดาคำตอบของข้อสอบถูก ดั้งนั้นการประมาณค่าพารามิเตอร์สำหรับการวิเคราะห์ข้อสอบจึงมีความสำคัญและมีความจำเป็น สำหรับการวิเคราะห์ข้อสอบ โดยในตอนนี้จะเป็นการนำเสนอวิธีการประมาณค่าพารามิเตอร์ที่ใช้ในการวิเคราะห์ข้อสอบที่ได้รับความนิยมซึ่งมีอยู่ด้วยกัน 3 วิธี ได้แก่ วิธีฮิวริสติก (Heuristic) แม็กซิมั่มไลค์ลิฮูด (Maximum Likelihood) และวิธีของเบส์ (Bayesian) (Hambleton และ Cook, 1977; Lord, F.M., 1980; รัตนา ศรีเหรัญ, 2539; ศิริชัย กาญจนวาสี, 2550; สุนทร เทียนงาม, 2551) ซึ่งแต่ละวิธีมีรายละเอียด ดังนี้

2.1 วิธีการประมาณค่าแบบฮิวริสติก (Heuristic)

การประมาณค่าพารามิเตอร์วิธีการนี้เป็นวิธีการที่ยูริเป็นผู้พัฒนาขึ้น (Urry, 1974) เพื่อใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ ซึ่งได้แก่ ค่าอำนาจจำแนก และค่าความยาก ของโมเดลการตอบสนองข้อสอบโดยค่าอำนาจจำแนกนั้นสามารถคำนวณได้จากค่าสัมประสิทธิ์สหสัมพันธ์แบบไบซีเรียล หรือค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยท์ไบซีเรียล สำหรับค่าความยากสามารถคำควณได้จากสัดส่วนของผู้ตอบข้อสอบข้อนั้นถูกต่อผู้เข้าสอบทั้งหมดซึ่งมีพื้นฐานมาจากการวิเคราะห์ข้อสอบด้วยทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory; CTT) และเพื่อใช้เป็นค่าเริ่มต้นในการประมาณค่าพารามิเตอร์ของโมเดลโลจิสติกด้วยวิธีแมกซิมัมไลค์ลิฮู้ด ต่อมา Schmidt (1977) ได้ศึกษาเกี่ยวกับการประมาณค่าพารามิเตอร์ด้วยวิธีฮิวริสติก (Heuristic) พบว่าการประมาณค่าพารามิเตอร์ด้วยวิธีฮิวริสติก (Heuristic) มีความคลาดเคลื่อนอย่างเป็นระบบ คือการประมาณค่าอำนาจจำแนกได้ค่าประมาณต่ำกว่าค่าจริงเสมอ และการประมาณค่าความยากได้

ค่าประมาณสูงกว่าค่าจริงเสมอ Schmidt (1977) จึงได้เสนอแนะให้ใช้ค่าความเที่ยงของแบบสอบ (KR-20) ในการปรับสมการของวิธีฮิวริสติก (Heuristic) ที่ใช้ในการประมาณค่า เพื่อทำให้การหาค่าประมาณได้ค่าที่มีความถูกต้องมากขึ้น

กระบวนการในประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีฮิวริสติก (Heuristic) เป็นกระบวนการที่ง่ายไม่มีการทำซ้ำ สามารถสรุปสมการ ที่ใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบดังนี้

$$a_j = \left[R_j\left(P_jQ_j\right)^{1/2}\right] / \left[(KR-20)Y_j^{\,2} - R_j^{\,2}\left(P_jQ_j\right)\right]^{1/2}$$

$$b_j = \left[Y_jZ_j(KR-20)^{1/2}\right] / \left[R_j\left(P_jQ_j\right)^{1/2}\right]$$

c_j = สัดส่วนของผู้ตอบถูกที่อยู่ปลายล่างของโค้งที่เกิดจากการสร้างกราฟจำนวนคนตอบถูกสะสม ในแต่ละระดับคะแนนรวม หลังจากหักคะแนนของข้อที่กำลังหาค่าพารามิเตอร์ออกจากคะแนนรวมทั้งฉบับก่อน

เมื่อ

a_j = ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อ j

b_j = ค่าพารามิเตอร์ความยากของข้อสอบข้อ j

R_j = ค่าสัมพันธ์พอยท์ไบซีเรียลของข้อสอบข้อที่ j

P_j = ค่าสัดส่วนการตอบถูกของข้อสอบข้อที่ j

Q_j = 1 - P_j

$KR-20$ = ค่าความเที่ยงของแบบสอบ

Z_j = ค่า Z ของการแจกแจงปกติมาตรฐานที่พื้นที่ใต้โค้งปกติมาตรฐานด้านขวามือมีค่าเท่ากับ P_j

Y_j = ค่าความสูงของโค้งปกติมาตรฐานที่ตรงจุด Z_j

สำหรับการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบในวิธีฮิวริสติก (Heuristic) จะไม่มีโดยตรง แต่จะนำข้อมูลที่ได้จากการประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีฮิวริสติก (Heuristic) ไปประยุกต์ใช้ต่อด้วยวิธีแม็กซิมั่มไลค์ลิฮูดแบบมีเงื่อนไข (Conditional Maximum Likelihood Estimation of Ability) ซึ่งเป็นการประมาณค่าความสามารถภายหลังจากทราบค่าพารามิเตอร์ของข้อสอบแต่ละข้อแล้ว โดยใช้กระบวนการนิวตัน ราฟสัน และกำหนดค่าความสามารถเริ่มต้น สำหรับผู้สอบคนที่ a เป็น

$$\theta_{0a} = \ln[r_a/(n-r_a)] = \theta_0$$

เมื่อ

r_a = จำนวนข้อสอบผู้เข้าสอบคนที่ a ตอบถูก

n = จำนวนข้อสอบที่ผู้เข้าสอบคนที่ a ทำทั้งหมด

จะทำให้ได้ค่า θ ของการทำซ้ำครั้งที่ m+1 ดังสมการ

$$\theta_a^{(m+1)} = \theta_a^{\ m} - ha^{(m)}$$

โดยที่

$$ha^{(m)} = \frac{d\ln L/d\theta_a}{d^2 \ln L/d\theta_a^{\ 2}}$$

$$\frac{d\ln L}{d\theta_a} = D\sum_{j=1}^{n} a_j(U_{ja} - P_{ja})(P_{ja} - C_j)/P_{ja}(1-C_j)$$

$$\frac{d\ln L}{d\theta_a^2} = D^2\sum_{j=1}^{n} a^2(P_{ja} - C_j)(U_{ja}C_j - P_{jA^2})Q_{ja}/P_{ja^2}(1-C_j)^2$$

เมื่อ

U_{ja} = ค่าผลการตอบ (ตอบถูกเป็น 1 ตอบผิดเป็น 0) ของผู้เข้าสอบคนที่ a

P_{ja} = โอกาสที่ผู้เข้าสอบคนที่ a ทำข้อสอบข้อ j ถูกต้อง

= $C_j + (1-c_j)/(1+\exp(-Da_j(\theta_a - b_j)))$

D = ค่าคงที่ขององค์ประกอบในการวัดมีค่าเท่ากับ 1.7

θ_a = ค่าความสามารถของผู้เข้าสอบคนที่ a

a_j = ค่าอำนาจจำแนกของข้อสอบข้อที่ j ที่ประมาณด้วยวิธีฮิวริสติก

b_j = ค่าความยากของข้อสอบข้อที่ j ที่ประมาณด้วยวิธีฮิวริสติก

c_j = ค่าการเดาของข้อสอบข้อที่ j ที่ประมาณด้วยวิธีฮิวริสติก

การประมาณค่าความสามารถของผู้สอบด้วยวิธีนี้จะสิ้นสุดกระบวนการก็ต่อเมื่อ $ha^{(m)}$ มีค่าน้อยกว่าหรือเท่ากับ 0.001 นอกจากนี้ในการประมาณค่าความสามารถของผู้สอบในโมเดลการตอบสนองข้อสอบ 2 พารามิเตอร์ต้องกำหนดให้ค่า c_j = 0 และการประมาณค่าความสามารถของผู้สอบในโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ต้องกำหนดให้ค่า c_j และค่า a_j เท่ากับ 0

โดยสรุปการประมาณค่าพารามิเตอร์วิธีฮิวริสติกเป็นวิธีการประมาณค่าที่ง่ายต่อการคำนวณค่า เนื่องจากสามารถใช้สารสนเทศจากทฤษฎีการทดสอบแบบดั้งเดิม (CTT) มาใช้ในการประมาณค่าพารามิเตอร์ แต่การประมาณค่าด้วยวิธีการนี้ไม่สามารถประมาณค่าพารามิเตอร์ความสามารถที่แท้จริงของผู้สอบได้โดยตรง ซึ่งสามารถประมาณค่าพารามิเตอร์ความสามารถที่แท้จริงของผู้สอบได้ก็ต่อเมื่อต้องใช้การประมาณค่าร่วมกับวิธีอื่นๆ แต่การประมาณค่าวิธีฮิวริสติกนี้สามารถประมาณค่าพารามิเตอร์ของผู้สอบได้ทั้ง ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และค่าโอกาสในการเดาข้อสอบถูกได้

2.2 วิธีการประมาณแบบแม็กซิมั่มไลค์ลิฮูด (Maximum Likelihood: ML)

การประมาณค่าพารามิเตอร์ด้วยวิธีแม็กซิมั่มไลค์ลิฮูด เป็นการประมาณค่าพารามิเตอร์โดยอาศัยผลที่ได้จากตัวอย่างที่สุ่มเลือกมาจากการแจกแจงที่ทราบรูปแบบของฟังก์ชันความหนาแน่นแต่ไม่ทราบค่าพารามิเตอร์ ดังนั้นจึงใช้หลักของความน่าจะเป็นในการเลือกตัวอย่าง และวัดค่าได้จากกลุ่มตัวอย่างที่ถูกเลือก ($U_1 = U_1,\ U_2 = U_2, \ldots\ldots, U_n = U_n$) มาพิจารณาค่าประมาณของค่าพารามิเตอร์ที่ต้องการ

จากการนิยามความน่าจะเป็นของผู้เข้าสอบในการตอบข้อสอบข้อที่ i เมื่อ$U_{ij} = 1$ สำหรับการตอบถูก และ $U_{ij} = 0$ สำหรับการตอบผิด ซึ่งสามารถแสดงสมการได้ดังนี้

$$
\begin{aligned}
P\,(U_{ij}|\theta,b,a,c) &= P(U_{ij}=1|\theta,b,a,c)\ \ P(U_{ij}=0|\theta,b,a,c) \\
&= P_{ij}^{u_{ij}}(1-P_{ij1})^{1-u_{ij}} \\
&= P_{ij}^{u_{ij}}\ Q_{ij}^{1-u_{ij}} Q_{ij} = 1-P_{ij}
\end{aligned}
$$

ถ้าผู้เข้าสอบตอบข้อสอบ n ข้อ และแบบทดสอบมีลักษณะการวัดเพียงมิติเดียว (คือ มีความเป็นอิสระเฉพาะที่) แล้วความน่าจะเป็นของการตอบแสดงได้ด้วยฟังก์ชันความหนาแน่นร่วม (Joint Density Function) ซึ่งสามารถแสดงสมการได้ดังนี้

$$
P(U:\theta,a,b) = \prod_{i=1}^{N}\prod_{j=1}^{n} P_{ij}^{U_{ij}} Q_{ij}^{1-U_{ij}}
$$

เมื่อ

$P(U:\theta,a,b)$ = ความน่าจะเป็นในการตอบข้อสอบ n ข้อของผู้เข้าสอบ N คน

U = เวคเตอร์ (Vector) ของตัวแปรที่แสดงผลการตอบข้อสอบ n ข้อ ของผู้เข้าสอบ N คน

θ = เวคเตอร์ (Vector) ความสามารถของผู้เข้าสอบ N คน

a = เวคเตอร์ (Vector) ค่าอำนาจจำแนกของข้อสอบ n ข้อ

b = เวคเตอร์ (Vector) ค่าความยากของข้อสอบ n ข้อ

สมการดังกล่าวเป็นความน่าจะเป็นของการตอบข้อสอบ n ข้อ ที่สามารถวัดหรือสังเกตได้โดยที่ $U_1,U_2,...,U_n$ เป็นตัวแปรสุ่มที่มีค่าเฉพาะเป็น $U_1,U_2,...,U_n$ เมื่อ U_1 มีค่าเท่ากับ 1 หรือ 0 และเนื่องจากสมการนี้เป็นฟังก์ชันทางคณิตศาสตร์ของค่า θ,b,a,c ที่จะบอกว่าตัวแปรสุ่มนี้มีโอกาสเกิดขึ้นเมื่อใด จึงเรียกฟังก์ชันนี้ว่าฟังก์ชันความน่าจะเป็นหรือ ฟังก์ชันไลค์ลิฮูด (Likelihood Function) ซึ่งแสดงได้ดังสมการ

$$L(U_1,U_2,...,U_n|\theta,b,a,c)=P_j^{u_j}\,Q_j^{1-u_j}$$

เมื่อU_j = 1 ค่าของ Q_j ก็หมดไป เพราะ $Q_j^{\,1-u_j}$ = 1 และเมื่อ U_j = 0 ค่าของ P_j จะหมดไป เพราะ $P_j^{\,u_j}$ = 1 และฟังก์ชันไลค์ลิฮูดที่มีผู้สอบ N คน ตอบข้อสอบ n ข้อ มีสมการดังนี้

$$L(U|\theta,b,a,c) = L(U_1,U_2,...,U_n|\theta_1,\theta_2,...,\theta_n, b_1,b_2,...,b_n,a_1,a_2,...a_n, c_1,c_2,...,c_n)$$

$$L(U|\theta,b,a,c) = \underset{i=1}{\overset{N}{\pi}}\ \underset{j=1}{\overset{n}{\pi}}\ L(U_{ij}|\theta,b,a,c)$$

$$L(U|\theta,b,a,c) = \underset{i=1}{\overset{N}{\pi}}\ \underset{j=1}{\overset{n}{\pi}}\ P_{ij}^{u_{ij}}\ Q_{ij}^{1-u_{ij}}$$

เมื่อ U = เวคเตอร์ผลการตอบข้อสอบ n ข้อ ของผู้เข้าสอบ N คน

U_i = เวคเตอร์ผลการตอบข้อสอบ n ข้อ ของผู้เข้าสอบคนที่ i

P_{ij} = $P_{ij}(\theta_i,b_1,a_1,c_1)$ = ฟังก์ชันไลค์ลิฮูดของผลการสอบ

การประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัม ไลค์ลิฮูด คือ การหาค่าประมาณของค่าพารามิเตอร์ที่จะทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด ซึ่งโดยปกติจะทำให้การหาค่าพารามิเตอร์ที่ทำให้ลอการิทึม (logorithm) ของฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด ทั้งนี้เพราะค่าประมาณของพารามิเตอร์ที่ทำให้ฟังก์ชันทั้งสองฟังก์ชันมีค่าสูงสุดเป็นค่าเดียวกัน แต่การหาค่าประมาณที่ทำให้ลอการิทึมของฟังก์ชันไลค์ลิฮูดมีค่าสูงสุดนั้นทำได้ง่ายกว่า ซึ่งลอการิทึมของฟังก์ชันไลค์ลิฮูดของตัวแปร θ, b, a, c คือ

$$\ln L(U|\theta, b, a, c) = \sum_{i=1}^{N} \sum_{j=1}^{n} (U_{ij}\, In\, P_{ij} + (1-U_{ij})\, In\, Q_{ij})$$

และการหาค่าพารามิเตอร์ θ, b, a, c ที่ทำให้ $\ln L(U|\theta, b, a, c)$ มีค่าสูงสุดทำได้โดยกำหนดให้อนุพันธ์ของ $\ln L(U|\theta, b, a, c)$ มีค่าเท่ากับศูนย์ แล้วแก้สมการหาค่ารากของอนุพันธ์ดังนี้

$$d \ln L(U|\theta, b, a, c) / d\,\theta_i \quad = 0 \quad \text{เมื่อ} \quad i = 1,2,...,N$$
$$d \ln L(U|\theta, b, a, c) / d\,b_i \quad = 0 \quad \text{เมื่อ} \quad i = 1,2,...,N$$
$$d \ln L(U|\theta, b, a, c) / d\,a_i \quad = 0 \quad \text{เมื่อ} \quad i = 1,2,...,N$$
$$d \ln L(U|\theta, b, a, c) / d\,c_i \quad = 0 \quad \text{เมื่อ} \quad i = 1,2,...,N$$

โดยทั่วไปสมการไลค์ลิฮูดนี้ไม่เป็นสมการเส้นตรง ดังนั้น การหาค่า θ_i, a_j, b_j, c_j ที่ทำให้ $InL(U|\theta, b, a, c)$ มีค่าสูงสุดจึงไม่สามารถหาได้ด้วยการใช้วิธีการอย่างง่าย แต่หาได้โดยใช้วิธีของนิวตัน ราฟสัน (Newton-Raphson Procedure) (Lord, 1980) ซึ่งเป็นการหาค่าประมาณโดยการประมาณค่าซ้ำ (Iterative) จนค่าที่ได้มีค่าคงที่ ซึ่งมีขั้นตอนดังต่อไปนี้

ขั้นที่ 1 กำหนดค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i และค่าพารามิเตอร์ a_j, b_j และ c_j

1. ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i คือ

$$\theta_i^{(0)} = In(x_i / (n - x_i))$$

เมื่อ In = Natural Logarithm

x_i = คะแนนสอบของคนที่ i

n = จำนวนข้อสอบ

2. ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ a_j, b_j และ c_j อาจใช้ค่าประมาณจากวิธีฮิวริสติค หรือค่าอื่นๆ ที่เหมาะสม

ขั้นที่ 2 ประมาณค่าความสามารถของผู้เข้าสอบแต่ละคน θ_i โดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่ ดังนี้

$$\theta_i^{(m+1)} = \theta_i^{(m)} - g(\theta_i^{(m)}) / h(\theta_i^{(m)})$$

เมื่อ $\theta_i^{(m)}$, $\theta_i^{(m+1)}$ = ค่าประมาณความสามารถของผู้เข้าสอบคนที่ i ครั้งที่ m และครั้งที่ m+i

$g(\theta_i^{(m)}) = d\, InL(U|\theta,b,a,c) / d\,\theta_i$; ประมาณที่ $\theta_i^{(m)}$

$h(\theta_i^{(m)}) = d^2\, InL(U|\theta,b,a,c) / d\,\theta_i$; ประมาณที่ $\theta_i^{(m)}$

การประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณความสามารถ (θ_i) จะมีค่าเข้าใกล้ค่าใดค่าหนึ่ง (convergence) คือ ค่าประมาณครั้งที่ m+1 และครั้งที่ m มีค่าแตกต่างกันน้อยกว่าค่าที่กำหนดไว้ เช่น 0.001

ขั้นที่ 3 ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ a_j ,b_j และ c_j โดยกำหนดให้ค่าความสามารถของผู้เข้าสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่

$$x_j^{(m+1)} = x_j^{(m)} - g(x_j^{(m)}) / h(x_j^{(m)})$$

เมื่อ x_j = เวคเตอร์ของค่า a , b และ c ของข้อสอบที่ j

$x_j^{(m)}, x_j^{(m+1)}$ = ค่าประมาณพารามิเตอร์ของข้อสอบที่ j ครั้งที่ m และครั้งที่ m+1

$g(x_j^{(m)})$ = $d\, InL(U|\theta,b,a,c) / d\,\theta_i$ ของค่า $x_j^{(m)}$

$h(x_j^{(m)})$ = $d^2\, InL(U|\theta,b,a,c) / d\,\theta_i$ ของค่า $x_j^{(m)}$

ในกระบวนการประมาณค่า จะมีการประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณจะลู่เข้าใกล้ค่าใดค่าหนึ่ง (convergence)

<u>ขั้นที่ 4</u> ประมาณค่าซ้ำในขั้นที่ 2 และขั้นที่ 3 จนกว่าค่าประมาณ θ_i, a_j, b_j และ c_j จะมีค่าที่คงที่ และมีความถูกต้องเพียงพอหรือทำให้ $\ln L(\theta, b, a, c)$ มีค่าสูงที่สุด

ค่าอนุพันธ์อันดับที่ 1 และค่าอนุพันธ์อันดับที่ 2 ของค่าพารามิเตอร์ θ, a, b, c แสดงค่าสุดท้ายที่ได้จากการประมาณจะเรียกว่า ค่าที่ได้จากการประมาณโดยวิธีแมกซิมัม ไลค์ลิฮูด และเพื่อให้ได้ค่าคงที่เร็วขึ้นในการประมาณคำซ้ำแต่ละครั้ง ค่าพารามิเตอร์ความสามารถของข้อสอบจะถูกปรับให้มีค่าเฉลี่ยเท่ากับ 0 และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ซึ่งจะเป็นผลให้ค่า a, b และ c ต้องถูกปรับตามค่าพารามิเตอร์ความสามารถของผู้สอบไปด้วย (Hambleton และ Swaminathan, 1985)

ตารางที่ 6 อนุพันธ์อันดับที่ 1 และอนุพันธ์อันดับที่ 2 สำหรับฟังก์ชันไลค์ลูฮูดของค่าพารามิเตอร์ข้อสอบ (a, b, c) และความสามารถของผู้เข้าสอบ (θ) ในแบบจำลองโลจิสติค 3 พารามิเตอร์

อนุพันธ์ (Derivative)	**สัญลักษณ์แสดง** (Expression)
$d\,InL/d\theta_i$	$D\sum_{j=1}^{N}\frac{a_j(P_{ij}-c_j)(u_{ij}-P_{ij})}{(1-c_j)P_{ij}}$
$d\,InL/d\,a_j$	$\frac{D}{(1-c_j)}\sum_{i=1}^{N}\frac{(\theta_i-b_j)(P_{ij}-c_j)(u_{ij}-P_{ij})}{P_{ij}}$
$d\,InL/d\,b_j$	$\frac{-Da_j}{(1-c_j)}\sum_{i=1}^{N}\frac{(P_{ij}-c_j)(u_{ij}-P_{ij})}{P_{ij}}$
$d\,InL/d\,c_j$	$\frac{1}{(1-c_j)}\sum_{i=1}^{N}\frac{(u_{ij}-P_{ij})}{P_{ij}}$
$d^2\,InL/d\theta_i$	$D^2\sum_{j=1}^{n}\frac{a_j(P_{ij}-c_j)Q_{ij}}{(1-c_j)^2P_{ij}}\left[\frac{u_{ij}\,c_j}{P}-P_{ij}\right]$
$d^2\,InL/d\,a_j$	$\frac{D^2}{(1-c_j)}\sum_{i=1}^{N}\left[\frac{(\theta_i-b_j)^2(P_{ij}-c_j)Q_{ij}}{P_{ij}}\left\{\frac{u_{ij}}{P_{ij}}c_j-P_{ij}\right\}\right]$

ตารางที่ 6 (ต่อ)

อนุพันธ์ (Derivative)	สัญลักษณ์แสดง (Expression)
$d^2\ InL/d\,b_j$	$\frac{D^2 a_j}{(1-c_j)^2}\sum_{i=1}^{N}(P_{ij}-c_j)\frac{Q_{ij}}{P_{ij}}\left[\frac{u_{ij}\,c_j}{P_{ij}}-P_{ij}\right]$
$d^2\ InL/d\,c_j$	$\frac{1}{(1-c_j)^2}\sum_{i=1}^{N}\left[\frac{u_{ij}-1}{P_{ij}}\right]-\left[\frac{u_{ij}\ {}_{ij}Q}{P_{ij}}\right]$

เมื่อ $P_{ij}=c_j+(2-c_j)/(1+\exp^{-Da_j(\theta_j-b_j)})$

u_{ij} = ผลการตอบข้อที่ j ของคนที่ i

จากวิธีการประมาณค่าพารามิเตอร์ข้างต้นแป็นการประมาณค่าตามโมเดลการตอบสนองข้อสอบโลจิสติค 3 พารามิเตอร์ (3 PL) ถ้าผู้วิเคราะห์ข้อสอบต้องการประมาณค่าแบบโมเดลการตอบสนองข้อสอบโลจิสติค 2 พารามิเตอร์ (2 PL) ให้กำหนดค่า c_j เป็น 0 และถ้าต้องการประมาณค่าแบบโมเดลการตอบสนองข้อสอบโลจิสติค 3 พารามิเตอร์ (3 PL) ให้กำหนดค่า c_j เป็น 0 และ a_j เป็น 1

ตารางที่ 7 ค่าฟังก์ชันสารสนเทศสำหรับค่าพารามิเตอร์ข้อสอบและค่าความสามารถของผู้เข้าสอบในแบบจำลองโลจิสติค 3 พารามิเตอร์

พารามิเตอร์ (Parameter)	**ฟังก์ชันสารสนเทศ** (Information Function)
θ_i	$D^2\sum_{j=1}^{n}\frac{a_j^2(P_{ij}-c_j)Q_{ij}(c_j-P_{ij})}{P_{ij}(1-c_j)^2}$
a_j	$\frac{D^2}{(1-c_j)^2}\sum_{i=1}^{N}(\theta_i-b_j)^2(P_{ij}-c_j)^2Q_{ij}$
b_j	$\frac{D^2a_j^2}{(1-c_j)^2}\sum_{i=1}^{N}\frac{(\theta_i-b_j)(P_{ij}-c_j)^2Q_{ij}}{P_{ij}}$
c_j	$\frac{D}{(1-c_j)^2}\sum_{i=1}^{N}\frac{(\theta_i-b_j)(P_{ij}-c_j)Q_{ij}}{P_{ij}}$

วิธีมาร์จินัลแม็กซิมั่มไลค์ลิฮูด (Marginal Maximum Likelihood Procedure)

การประมาณค่าพารามิเตอร์ด้วยวิธีมาร์จินัลแม็กซิมั่มไลค์ลิฮูดนี้เป็นวิธีการประมาณค่าโดยสามารถประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบได้พร้อมกัน ซึ่งการประมาณค่าด้วยวิธีการนี้จะไม่ได้อิงกับค่าพารามิเตอร์ของผู้สอบ นอกจากนี้ยังมีการแจกแจงความสามารถของผู้สอบ เพื่อให้ได้ฟังก์ชันความน่าจะป็นของพารามิเตอร์ของข้อสอบ หลังจากนั้นจึงนำค่าที่ได้ไปใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ การประมาณค่าด้วยวิธีนี้สามารถใช้ได้กับแบบทดสอบที่มีทั้งจำนวนข้อมากและน้อยได้ วิธีการประมาณค่าเริ่มจากรูปแบบการตอบข้อสอบของผู้สอบโดยสามารถแทนได้ด้วยสมการดังนี้

$$X = (x_1, x_2, ..., x_n)$$

และสำหรับผู้ที่มีความสามารถ θ ค่าความเป็นไปได้สามารถแสดงได้ดังนี้

$$P(X|\theta) \quad = \quad \mathop{\pi}_{j=1}^{n} \; P_{ij}(\theta)^{x_j}\left[1-P(\theta)\right]^{1-x_j}$$

ค่าความน่าจะเป็นที่ได้เป็นความน่าจะเป็นที่มีเงื่อนไขว่าทราบค่าความสามารถของผู้สอบ (θ) ซึ่งต่างจากความน่าจะเป็นของรูปแบบ X จากผู้ไม่ทราบค่าความสามารถของผู้สอบ (θ) ซึ่งสุ่มมาจากประชากรที่มีการกระจายของค่าระดับความสามารถ (θ) อยู่ในรูปฟังก์ชันความหนาแน่นต่อเนื่อง (Continuous Density g (θ)) ที่อยู่ในรูปของความน่าจะเป็นที่ไม่มีเงื่อนไขและสามารถแสดงได้ดังนี้

$$P(X) = \int_{-\infty}^{+\infty} P(X|\theta)g(\theta)de$$

ค่านี้เรียกว่าความน่าจะเป็นโดยปราศจากเงื่อนไขของ X และเนื่องจากค่าความสามารถ θ ถูกอินทริเกรต (Integrate) ออกไป ค่านี้จึงเป็นฟังก์ชันของพารามิเตอร์ของข้อสอบเท่านั้น ในวิธีมาร์จินัลแมกซิมัมไลค์ลิฮูด ค่าของพารามิเตอร์ของข้อสอบจะพิจารณาใช้ค่าที่ทำให้สมการข้างล่าง มีค่าสูงสุด

$$\log L_M = \sum_{l=1}^{s} r_l \log_\theta \overline{P}(X_l)$$

เมื่อ P (X_l) ประมาณได้จากสูตร Gaussian Guadrature

r_l คือ ความถี่ของรูปแบบการตอบข้อสอบรูปแบบ X_l จากจำนวนผู้สอบทั้งหมด (N) และ

S คือ จำนวนรูปแบบที่เป็นไปได้ทั้งหมดจากข้อมูลการตอบข้อสอบ

ในแบบจำลองโลจิสติค 3 พารามิเตอร์ ค่าสูงสุดของสมการดังกล่าวสำหรับ ข้อ j หาได้จากสมการไลค์ลิฮูด ดังนี้

$$\sum_{k=1}^{a} \frac{\bar{r}_{jk} - \bar{N}_k P_j(X_k)}{P_j(X_k)\left[1 - P_j(X_k)\right]} \frac{d\,P_j(X_k)}{d\begin{bmatrix} c_j \\ a_j \\ b_j \end{bmatrix}} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

เมื่อ $\bar{r}_{jk} = \sum_{l}^{s} r_l X_{jl} P(X_l / X_k) A(X_k) / \bar{P}(X_l)$

$$\bar{N}_k = \sum_{l}^{s} r_l P(X_l / X_k) A(X_k) / \bar{P}(X_l)$$

โดยสมการดังกล่าวจะสามารถคำนวณค่าความคาดหวังภายหลังของจำนวนข้อสอบที่ถูก และจำนวนข้อสอบที่ทำที่จุด X_k (X_{jl})คือคะแนน 0, 1 สำหรับข้อสอบ j ในรูปแบบการตอบ i นอกจากนี้การประมาณค่าด้วยวิธีการวิธีมาร์จินัลแม็กซิมั่มไลค์ลิฮูด สามารถสรุปเป็นสมการทางคณิตศาสตร์ได้ดังนี้

Marginal Maximum Likelihood of Item Parameters(β):

Maximize $L(a,b,c) = \prod_{a=1}^{N} \int_{-\infty}^{\infty} g(a) L(\theta_a ; a,b,c) d\theta_a$

ปัจจุบันในการคำนวณเพื่อประมาณค่าพารามิเตอร์ด้วยวิธีมาร์จินัลแม็กซิมั่มไลค์ลิฮูด สามารถทำได้โดยสะดวก เนื่องจากมิสเลวี่ และบ็อค (Mislevy และ Bock, 1990) ได้พัฒนาโปรแกรมคำสั่งชื่อ BILOG แต่ในการใช้โปรแกรมดังกล่าวต้องใช้กลุ่มผู้สอบที่มีขนาดใหญ่ เพื่อทำให้สามารถประมาณค่าพารามิเตอร์ของข้อสอบได้อย่างคงที่ นอกจากนี้จำนวนข้อสอบสำหรับ การประมาณค่ายังต้องมีจำนวนมาก สำหรับการประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้โมเดลการตอบสนองข้อสอบแบบ 3 พาราเตอร์ (3 PL) ต้องมีผู้สอบที่มีระดับความสามารถต่ำที่เพียงพอด้วย

จึงจะสามารถทำการประมาณค่าพารามิเตอร์ค่าโอกาสในการเดาข้อสอบถูก (c) ได้อย่างมีความน่าเชื่อถือ นอกจากมีผลกระทบต่อค่าพารามิเตอร์โอกาสในการเดาข้อสอบถูก (c) แล้วยังอาจจะส่งผลกระทบถึงค่าพารามิเตอร์ความยากของข้อสอบ (b) และค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) อีกด้วย แต่ปัญหาดังกล่าวสามารถแก้ไขได้ โดยการประมาณค่าพารามิเตอร์ด้วยวิธีการของเบย์ (Bayesian) ซึ่งจะมีการกำหนดการแจกแจงเบื้องต้นของค่าพารามิเตอร์ค่าโอกาสในการเดาข้อสอบถูก (c) ก่อนในโปรแกรม BILOG (ศิริชัย กาญจนวาสี, 2550) นอกจากนี้ยังมีโปรแกรมคอมพิวเตอร์บางโปรแกรม เช่น โประแกรม LOGIST สามารถทำการค้นหาค่าความสามารถและค่าพารามิเตอร์ของข้อสอบ (a,b,c) ที่จะทำให้ฟังก์ชันประมาณค่าที่เป็นไปได้สูงสุด

วิธีการประมาณค่าแบบแม็กซิมั่มไลค์ลิฮูดร่วมกัน (Joint Maximum Likelihood Estimation Procedure)

สำหรับการประมาณค่าด้วยวิธีประมาณค่าที่เป็นไปได้สูงสุดวิธีการประมาณค่าแบบแม็กซิมั่มไลค์ลิฮูดร่วมกัน (Joint Maximum Likelihood Estimation Procedure) ในการประมาณค่าจะใช้ N แทนจำนวนผู้สอบ, $P_i(\theta_a)$ แทนความน่าจะเป็นในการตอบข้อสอบถูกในข้อ i ณ ผู้สอบระดับความสามารถ θ_a ,$Q_i(\theta_a) = 1 - P_i(\theta_a)$, u_{ia} แทนคำตอบข้อที่ i (0, 1) และ $g(\theta_a)$แทนการแจกแจงพารามิเตอร์โดยมีสูตรดังนี้ (Lord, 1984)

$$\text{Maximize } L(\theta; a,b,c) = \prod_{a=1}^{N}\prod_{i=1}^{n}\left[P_i(\theta_a)\right]^{u_{ia}}\left[Q_i(\theta_a)\right]^{1-u_{ia}}$$

หรือ
$$\log L(\theta; a,b,c) = \sum_{a=1}^{N}\sum_{i=1}^{n}\left[u_{ia}\log P_i(\theta_a) + (1+u_{ia})\log Q_i(\theta_a)\right]$$

สำหรับขั้นตอนแรกของการประมาณค่าด้วยวิธีประมาณแบบแม็กซิมั่มไลค์ลิฮูดร่วมกันมีการกำหนดค่าเริ่มต้นของ θ_a โดยใช้ค่า log ของอัตราส่วนจำนวนข้อที่ตอบถูกต่อจำนวนข้อที่ตอบผิดสำหรับผู้ตอบแต่ละคนแปลงเป็นคะแนนมาตรฐานเพื่อใช้เป็นค่าเริ่มต้นของ θ_a จากนั้นทำประมาณค่าพารามิเตอร์ของข้อสอบจากค่าเริ่มต้นที่กำหนดขึ้นของ θ_a จากขั้นตอนแรก สำหรับขั้นตอนที่สอง จากค่าพารามิเตอร์ของข้อสอบที่ได้จากขั้นตอนแรก ทำเสมือนทราบค่าพารามิเตอร์ของข้อสอบ เพื่อประมาณค่าความสามารถของผู้สอบ วิธีนี้จะกระทำซ้ำตามขั้นตอนทั้งสองจนกระทั่งได้ค่าประมาณ 2 ครั้งหลังที่ไม่เปลี่ยนแปลงโปรแกรมคอมพิวเตอร์สำหรับการประมาณค่าด้วยวิธีการนี้ ได้แก่ โปรแกรม LOGIST สำหรับการวิเคราะห์ข้อสอบตามโมเดลการตอบสนอง 1 2 3 พารามิเตอร์และโปรแกรม BICAL และโปรแกรม BIGSCALE สำหรับการวิเคราะห์ข้อสอบตามโมเดลการตอบสนอง 1 พารามิเตอร์ (ศิริชัย กาญจนวาสี, 2551)

ข้อพึงระวังสำหรับการประมาณค่าด้วยวิธีประมาณแบบแม็กซิมั่มไลค์ลิฮูดร่วมกันมี คือ การประมาณค่าด้วยวิธีการนี้ไม่สามารถประมาณค่าความสามารถได้ในกรณีมีผู้สอบได้คะแนนเต็ม หรือได้ศูนย์คะแนน นอกจากนี้ยังไม่สามารถประมาณค่าพารามิเตอร์ข้อสอบได้ ในกรณีข้อสอบที่มีผู้ตอบถูกหรือผิดหมดทุกคน และในการวิเคราะห์โมเดล 2 และ 3 พารามิเตอร์ ค่าพารามิเตอร์ของโมเดลการตอบสนองข้อสอบจะมีความคงเส้นคงวาได้ ก็ต่อเมื่อมีจำนวนข้อสอบมากและกลุ่มผู้สอบมีขนาดใหญ่ (Hambleton, Swaminathan และ Rogers, 1991 อ้างถึงใน ศิริชัย กาญจนวาสี, 2545)

เมื่อเปรียบเทียบวิธีแม็กซิมั่มไลค์ลิฮูดทั้ง 2 วิธี พบว่าวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบวิธีมาร์จินัลแม็กซิมั่มไลค์ลิฮูด มีข้อดีกว่าวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบร่วมกันที่สำคัญ คือ สามารถประมาณค่าผู้สอบโดยไม่ต้องประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ รวมทั้งความแม่นยำของการวิเคราะห์ เมื่อมีจำนวนผู้สอบ 1000 ถึง 2000 คน ทำข้อสอบ 40 ข้อ จะให้ค่าประมาณที่แตกต่างกันเล็กน้อย แต่ในกรณีที่ผู้สอบทำข้อสอบ 10 ถึง 15 ข้อ วิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบร่วมกัน จะให้ค่าประมาณความสามารถที่ลำเอียง โดยเฉพาะในผู้สอบที่มีความสามารถต่ำ ดังนั้นจึงทำให้การประมาณค่าพารามิเตอร์มีความคลาดเคลื่อนแม้จะมีผู้สอบจำนวนมาก (Hambleton และ Swaminathan, 1985)

ข้อดีของการประมาณค่าด้วยวิธีแม็กซิมั่มไลค์ลิฮูด คือ การประมาณค่าพารามิเตอร์ด้วยวิธีการนี้จะให้สารสนเทศของค่าพารามิเตอร์ที่ต้องการได้ทั้งหมด ไม่ว่าจะเป็นพารามิเตอร์ของผู้สอบ ซึ่งได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบ หรือพารามิเตอร์ของข้อสอบ ซึ่งได้แก่ ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบและค่าโอกาสในการเดาข้อสองได้ถูกต้องการประมาณค่าพารามิเตอร์ของวิธีนี้ขึ้นอยู่กับจำนวนของกลุ่มผู้เข้าสอบและข้อสอบถ้ามีจำนวนเพิ่มขึ้น การประมาณค่าก็จะมีความคงที่ไปสู่ค่าพารามิเตอร์เพิ่มมากขึ้น

ส่วนข้อจำกัดของการประมาณค่าด้วยวิธีแม็กซิมั่มไลค์ลิฮูด คือ ประเด็นแรกในการประมาณค่าพารามิเตอร์ในขั้นที่ 2 และ 3 โดยใช้ค่าอนุพันธ์อันดับที่ 2 ในกระบวนการนิวตัน ราฟสันนั้นมีโอกาสที่ค่าประมาณที่ได้จะไม่ลู่เข้าสู่ค่าคงที่ ประเด็นถัดมาสำหรับการประมาณค่าในสมการไลค์ลิฮูดไม่ใช่สมการเชิงเส้นตรง จะทำให้การหารากของสมการที่ทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุดได้หลายค่า แต่ค่าเหล่านี้ไม่สามารถนำไปใช้หรือประกันได้ว่าเป็นค่าพารามิเตอร์ที่แท้จริงได้ ประเด็นที่สาม ในบางครั้งค่าพารามิเตอร์หรือค่าที่ได้จากการประมาณไม่ตกอยู่ในขอบเขตของค่าพารามิเตอร์ กล่าวคือ อาจมีค่าใดค่าหนึ่งอยู่ภายนอกขอบเขตที่ยอมรับได้ ในกรณีเช่นนี้ต้องมีการกำหนดขอบเขตจำกัดของค่าประมาณไว้ เพื่อให้ค่าประมาณที่ได้ไม่สูงหรือต่ำเกินไปนักแต่ การกระทำเช่นนี้เป็นจุดอ่อนของ

กระบวนการประมาณค่าด้วยวิธีแม็กซิมั่มไลค์ลิฮูด โดยเฉพาะในแบบจำลอง 2 และ 3 พารามิเตอร์ จึงทำให้เกิดปัญหาตามมาเกี่ยวกับความตรง (Validity) ของค่าที่ประมาณได้ และประเด็นสุดท้ายเนื่องจากการประมาณค่าพารามิเตอร์ในแบบจำลอง 3 พารามิเตอร์มีสมการหลายสมการที่ต้องหารากที่ทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด ด้วยวิธีของ นิวตัน ราฟสัน จึงจำเป็นต้องใช้เครื่องคอมพิวเตอร์ขนาดใหญ่ในการทำงาน

2.3 การประมาณค่าพารามิเตอร์แบบเบส์ (Bayesian Estimation)

เนื่องจากการประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัม ไลค์ลิฮูด มีข้อจำกัดและปัญหาเมื่อทำการประมาณค่าพารามิเตอร์ของข้อสอบ และค่าความสามารถของผู้เข้าสอบไปพร้อมๆ กันดังกล่าวแล้ว วิธีของเบส์จึงอาจเป็นวิธีที่เหมาะสมกว่า ทั้งนี้เพราะวิธีของเบส์มีแนวคิดบางประการที่ต่างออกไปจากแนวคิดของวิธีแม็กซิมั่มไลค์ลิฮูด และได้รับการพัฒนาให้สามารถประมาณค่าพารามิเตอร์ของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ (a, b, c) ร่วมกันได้อย่างมีประสิทธิภาพ (Swaminathan และ Gifford 1982) กล่าวคือ การประมาณค่าพารามิเตอร์ด้วยวิธีของเบส์มีแนวคิดว่า ค่าความสามารถของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ ซึ่งได้แก่ ค่าอำนาจจำแนกของข้อสอบ (a) ค่าความยากของข้อสอบ (b) และค่าโอกาสในการเดาข้อสอบ (c) เป็นตัวแปรสุ่ม (Random Variable) จากการแจกแจงที่แสดงได้ด้วยฟังก์ชันความหนาแน่นร่วม (Joint Density Function) f(θ,b,a,c)โดยเรียกฟังก์ชัน f(θ,b,a,c) นี้ว่าการแจกแจงเริ่มแรก (Prior Distribution) ของค่า θ,b,a และ c ซึ่งทำให้การใช้ฟังก์ชันไลค์ลิฮู้ด $L(U|\theta,b,a,c)$ เพียงอย่างเดียวในการประมาณค่า θ,b,a และ c ถูกพิจารณาว่าเป็นการใช้ข้อมูลที่มีอยู่อย่างไม่ครบถ้วน เพราะยังมีการแจกจงเริ่มต้นร่วมกับ f(θ,b,a,c)ที่ควรนำมาใช้ในการประมาณค่าพารามิเตอร์ด้วย

จากการนิยามความน่าจะเป็นของผู้เข้าสอบคนที่ i ในการตอบข้อสอบข้อที่ j เมื่อ $U_{ij}=1$ สำหรับการตอบถูก และ $U_{ij}=0$ สำหรับการตอบผิดสามารถแสดงสมการโมเดลแบบ 3 พารามิเตอร์ของโมเดลการตอบสนองข้อสอบ (3 PL IRT) ได้ดังนี้

$$P(U_j|\theta,b,a,c) \quad = \quad P(U_j=1|\theta,b,a,c)\,P(U_j=1|\theta,b,a,c)$$

$$= \quad P_j^{u_j}\ Q_j^{1-u_j}\ ;\ Q_j=1-P_j$$

ถ้าผู้เข้าสอบตอบข้อสอบจำนวน n ข้อและข้อสอบแต่ละข้อเป็นอิสะต่อกัน ความน่าจะเป็นของการตอบ สามารถแสดงได้ด้วยฟังก์ชันความหนาแน่นร่วม ตามสมการดังต่อไปนี้

$$P(u_1,u_2,...,u_n|\theta,b,a,c) = \underset{j=1}{\overset{n}{\pi}}\ P_j^{u_j}\,Q_j^{1-u_j}$$

จากสมการดังกล่าวข้างต้น เป็นความน่าจะเป็นของการตอบข้อสอบ n ข้อ ที่สามารถวัดหรือสังเกตได้ โดยที่ $U_1, U_2, ..., U_n$ เป็นตัวแปรสุ่มที่มีค่าเฉพาะเป็น $u_1, u_2, ..., u_n$ เมื่อ u_{ij} มีค่าเท่ากับ 1 หรือ 0 และเนื่องจากสมการนี้เป็นฟังก์ชันทางคณิตศาสตร์ของค่า θ, b, a, c ที่จะบอกว่าตัวแปรสุ่มนี้มีโอกาสขึ้นเพียงใด จึงเรียกฟังก์ชันนี้ว่า ฟังก์ชันความน่าจะเป็น หรือ ฟังก์ชันไลค์ลิฮูด (Likelihood Function) ซึ่งแสดงได้ดังนี้

$$L(u_1, u_2, ..., u_n | \theta, b, a, c) = \prod_{j=1}^{n} P_j^{u_j} Q_j^{1-u_j}$$

เมื่อ $u_{ij} = 1$ ค่าของ Q_1 จะหมดไป และเมื่อ $u_{ij} = 0$ ค่าของ P_1 ก็จะหมดไปและฟังก์ชันไลค์ลิฮูด ที่มีผู้สอบ N คน ตอบข้อสอบ n ข้อ มีสมการดังนี้

$$L(U | \theta, b, a, c) = \prod_{i=1}^{N} \prod_{j=1}^{n} P_{ij}^{u_{ij}} Q_{ij}^{1-u_{ij}}$$

เมื่อ u = เวคเตอร์ผลการตอบข้อสอบ n ข้อ ของผู้เข้าสอบ N คน

$$P_{ij} = P_1 (\theta_i, b_j, a_j, c_j)$$

ดังนั้นถ้าพิจารณาความน่าจะเป็นร่วมของการตอบข้อสอบ $P(U | \theta, b, a, c)$ แล้วจะเห็นว่าการแจกแจงของตัวแปร U ขึ้นอยู่กับค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b และ c ถ้าค่า θ, a, b และ c เปลี่ยนไป โอกาสที่ตัวแปร U มีค่าเท่ากับ u ก็จะเปลี่ยนไปด้วย ดังนั้นการทราบผลการตอบข้อสอบ u จึงน่าจะช่วยทำให้ทราบค่า θ, a, b และ c ได้ดียิ่งขึ้น ซึ่งอาจแสดงได้ด้วยการแจกแจงอย่างมีเงื่อนไขของค่า θ, a, b และ c เมื่อทราบผลการตอบข้อสอบ $f(\theta, b, a, c | u)$ และเรียกฟังก์ชันนี้ว่า การแจกแจงภายหลัง (Posterior Distribution)

การแจกแจงภายหลังร่วมกันของค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b และ c เมื่อทราบผลการตอบข้อสอบ u ก็คือ

$$f(\theta, b, a, c | u) = L(U | \theta, b, a, c)\, f(\theta, b, a, c) \,/\, f(u)$$

เมื่อ f (u) = การแจกแจงมาร์จินัล (Marginal) ของผลการตอบข้อสอบ

f (θ,b,a,c) = การแจกแจงเริ่มแรกของค่า θ,a,b และ c

f$(U|\theta,b,a,c)$ = ฟังก์ชันไลค์ลิฮูด

f$(\theta,b,a,c)|u)$ = การแจกแจงภายหลังของค่า θ,a,b และ c เมื่อทราบผลการตอบข้อสอบ u

ซึ่งการแจกแจงภายหลังนี้เป็นฟังก์ชันที่ใช้ในการประมาณค่า θ,a,b และ c ด้วยวิธีของเบส์ โดยมีส่วนแตกต่างจากการประมาณด้วยวิธีแม็กซิมั่มไลค์ลิฮูด คือ การแจกแจงเริ่มแรก f(θ,b,a,c) และการแจกแจงมาร์จินัล f (u) โดยการแจกแจงมาร์จินัลนี้เป็นการแจกแจงที่ไม่ขึ้นอยู่กับค่า θ,a,b และ c จึงถือว่าเป็นค่าคงที่ในการประมาณค่า θ,a,b และ c

จากแนวคิดของการประมาณค่าด้วยวิธีของเบส์ดังกล่าว ทำให้สามารถจำแนกกระบวนการดำเนินการตามแนวคิดของเบส์ออกเป็น 2 กระบวนการ ดังนี้

1. กระบวนการกำหนดลักษณะของการแจกแจงเริ่มแรก (Prior Distribution) มีอยู่ 2 ขั้น คือ (Swaminathan และ Gifford, 1986)

1.1 กำหนดให้การแจกแจงเริ่มแรกของค่าความสามารถ θ ค่าอำนาจจำแนก a ค่าความยาก b และค่าการเดา c เป็นอิสระต่อกัน ดังนี้

$$f(\theta,b,a,c) = f(\theta)\, f(b)\, f(a)\, f(c)$$

โดยกำหนดลักษณะการแจกแจงของ f (θ) f (b) f (a) และ f (c) ไว้ดังนี้

1.1.1 การแจกแจงเริ่มแรกของค่าความสามารถ f (θ) มีข้อตกลงว่า ข้อสารสนเทศที่มีมาก่อนของค่าความสามารถของผู้เข้าสอบแต่ละคนไม่แตกต่างกันสามารถใช้แทนกันได้ (exchangeability) และค่าความสามารถเป็นตัวแปรสุ่มที่มีการแจกแจงเป็นปกติ (normal distribution)

$$f(\theta_i|\mu_\theta,\sigma_\theta^2) = N(\mu_\theta,\sigma_\theta^2)$$

เมื่อ $N(\mu_\theta,\sigma_\theta^2)$ = การแจกแจงปกติ (normal distribution) ที่มีค่าเฉลี่ยเท่ากับ μ_θ และค่าความแปรปรวนเท่ากับ σ_θ^2

1.1.2 การแจกแจงเริ่มแรกของค่าความยากของข้อสอบ f (b) อาจใช้กระบวนการเดียวกับการกำหนดการแจกแจงเริ่มแรกของค่าความสามารถ คือ มีข้อตกลงว่า f (b) มีการแจกแจงเป็นปกติ หรืออาจจะไม่กำหนดการแจกแจงเริ่มแรกไว้ก็ได้ (Swaminathan และ Gifford, 1985)

1.1.3 การแจกแจงเริ่มแรกของค่าอำนาจจำแนกของข้อสอบ f (a) ควรเป็นการแจกแจงแบบไคว์ (Chi-square distribution) เนื่องจากค่าอำนาจจำแนกของข้อสอบโดยทั่วไปจะต้องเป็นค่าบวก และเป็นความชันของเส้นโค้งลักษณะของข้อสอบ ณ จุดเปลี่ยนโค้ง

$$f\ (a_j | v_j\, w_j)\ \alpha \quad a_j^{\,v_j - 1} \exp\left[-a_j^{\,2} / 2w_j\right]$$

เมื่อ v_j = Degree of Freedom

w_j = Scale Parameter

1.1.4 การแจกแจงเริ่มแรกของค่าการเดาของข้อสอบ f (c) ควรมีการแจกแจงแบบเบต้า (Beta Distribution) เนื่องจากค่าพารามิเตอร์ c_j มีขอบเขตอยู่ตั้งแต่ 0-1

$$f\ (c_j | s_j\, t_j)\ \alpha \quad c_j^{\,a_j} (1 - c_j)^{t_j}$$

เมื่อ $s_j\, t_j$ = Scale Parameter

1.2 กำหนดค่าที่เป็นตัวเลขของพารามิเตอร์สำหรับการแจกแจงเริ่มแรก

1.2.1 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความสามารถ θ ได้แก่ μ_θ และ $\sigma_\theta^{\,2}$ อาจกำหนดให้ $\mu_\theta = 0$ และ $\sigma_\theta^{\,2} = 1$ ซึ่งจะทำให้การประมาณค่าความสามารถ θ มีความสะดวก และประมาณค่าได้รวดเร็วขึ้น (Swaminathan และ Gifford. 1985: 351-355)

1.2.2 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความยาก b หากมีการกำหนดลักษณะการแจกแจงไว้ ก็ใช้ค่าเดียวกับพารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความสามารถ θ ดังกล่าวแล้วข้างต้น

1.2.3 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าอำนาจจำแนก a ได้แก่ v_j และ w_j การกำหนดค่าของ v_j และ w_j ที่เหมาะสม อาจจะหาได้จากการกำหนดพิสัย (Range) ของค่า a คือ ถ้าให้ H เป็นขีดจำกัดบนของพิสัย และให้ L เป็นขีดจำกัดล่างของพิสัย จะหาค่า v_j และ w_j ได้จากสูตร

$$v_i = ½\,(1 + Z_{(½)\alpha}\,((H + L)/H - L)^2)$$

$$w_i = ½\,((H - L)/Z_{½\alpha})^2$$

เมื่อ $Z_{(½)\alpha}$ = ค่า Z ของการแจกแจงปกติมาตรฐาน Standard Normal Distribution) ที่ระดับนัยสำคัญ α

v_i = Degree of Freedom

นอกจากวิธีดังกล่าว อาจจะกำหนดให้ v_j และ w_j ของการแจกแจงเริ่มแรก f (a) ของข้อสอบทุกข้อเท่ากัน คือ v_j = 10 และ w_j = 0.1 ซึ่งจะทำให้การประมาณค่าพารามิเตอร์มีความสะดวกยิ่งขึ้น (Swaminathan และ Gifford, 1985) และการกำหนดเช่นนี้ก็ยังคงทำให้ค่าประมาณของค่าพารามิเตอร์ a_j ตกอยู่ในช่วงที่ยอมรับได้ คือ 0.40 < a_j < 1.55 ด้วยความเชื่อมั่น 99 % (Swaminathan และ Gifford, 1985)

1.2.4 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าการเดา c ได้แก่ s_j และ t_j ที่เหมาะสมอาจหาได้จากการสังเกตสัดส่วนการตอบถูกของกลุ่มผู้เข้าสอบที่มีความสามารถในระดับต่ำมาก กล่าวคือ ถ้าให้ m แทนจำนวนผู้เข้าสอบที่มีระดับความสามารถต่ำมาก และให้ M แทนสัดส่วนการตอบถูกของผู้เข้าสอบกลุ่ม m แล้ว สามารถหาค่า s_j และ t_j ได้จากสูตร

s_j = m M

t_j = m (1-M) -2

นอกจากวิธีนี้อาจจะกำหนดให้ s_j และ t_j ของการแจกแจงเริ่มแรก f (c) ของข้อสอบทุกข้อเท่ากัน คือ s_j = 2 และ t_j = 12 ซึ่งจะทำให้การประมาณค่าพารามิเตอร์มีความสะดวกยิ่งขึ้น (Swaminathan และ Gifford, 1986) และการกำหนดเช่นนี้ ก็ยังคงทำให้ค่าประมาณของค่าพารามิเตอร์ c_j ตกอยู่ในช่วงที่ยอมรับได้ คือ 0.026 < c_j < 0.317 ด้วยความเชื่อมั่น 99 % (Novick และ Jackson, 1974 อ้างถึงในวิชุดา บัวคง. 2532)

2. กระบวนการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้เข้าสอบ

การประมาณค่าพารามิเตอร์ด้วยวิธีของเบส์ คือ การหาค่าประมาณ $\theta_i(j=1,2,...,N)$ และ $b_j, a_1, c_1\,(i=1, 2,..., n)$ ที่ทำให้ฟังก์ชันการแจกแจงภายหลัง f$(\theta,b,a,c|u)$ มีค่าสูงสุด ถ้า ln f$(\theta,b,a,c|u)$ เป็นฟังก์ชันที่หาอนุพันธ์ได้ การหาค่า θ_i, b_j, a_j และ c_j จะหาได้จากการอนุพันธ์ของ ln f$(\theta,b,a,c|u)$ และกำหนดให้อนุพันธ์ของ ln f$(\theta,b,a,c|u)$ มีค่าเท่ากับศูนย์ แล้วจึงหาค่ารากของอนุพันธ์ของ ln f$(\theta,b,a,c|u)$ ซึ่งแสดงได้ดังสมการต่อไปนี้

$$f(\theta,b,a,c|u) = L(u|\theta,b,a,c)\ f(\theta)\ f(b)\ (a)\ f(c) / f(u)$$

$$\ln f(\theta, b, a, c|u) = \ln L(u|\theta, b, a, c) + \ln f(\theta) + \ln f(b)$$
$$+ \ln f(a) + \ln f(c) + \text{Constant}$$

$$d \ln f(\theta, b, a, c|u) / d\theta_i = 0$$

$$d \ln f(\theta, b, a, c|u) / d b_j = 0$$

$$d \ln f(\theta, b, a, c|u) / d a_j = 0$$

$$d \ln f(\theta, b, a, c|u) / d c_j = 0$$

สมการอนุพันธ์ของ ln $f(\theta, b, a, c|u) = 0$ นี้ เรียกว่า สมการโมดัล (Modal Equation) ค่ารากของสมการโมดัล คือ ค่าความสามารถ θ และค่าพารามิเตอร์ของข้อสอบ a, b, c ที่ทำให้ฟังก์ชันการแจกแจงภายหลังมีค่าสูงสุด อาจทำได้โดยใช้เทคนิคนิวตัน ราฟสัน (Newton-Raphson) ซึ่งเป็นการหาค่าประมาณโดยการหาค่าซ้ำ (Iterative) มีขั้นตอนดังต่อไปนี้

ขั้นที่ 1 กำหนดค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i และค่าพารามิเตอร์ ของข้อสอบ a_j b_j และ c_j

1.1 ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i

$$\theta_i(0) = \ln(X_i / (n - X_i))$$

เมื่อ ln = Natural Logarithm

X_i = คะแนนสอบของผู้เข้าสอบคนที่ i

n = จำนวนข้อสอบ

1.2 ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ a_j, b และ c_j

$$a_j^{(0)} = R_j / \left(1 - R_j^2\right)^{1/2}$$
$$b_j^{(0)} = Z_j / R_j$$
$$c_j = 1 / m_j$$

เมื่อ R_j = Point – Biserial Correlation

Z_j = ค่า Z ของการแจกแจงปกติมาตรฐานที่พื้นที่ใต้โค้งปกติมาตรฐานด้านขวามือมีค่าเท่ากับ P_j

P_j = U_j / N

N = จำนวนผู้เข้าสอบทั้งหมด

m_j = จำนวนตัวเลือกในข้อสอบข้อที่ j

<u>ขั้นที่ 2</u> ประมาณค่าความสามารถของผู้เข้าสอบแต่ละคน θ_i โดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าได้ในครั้งก่อนเป็นค่าคงที่

$$\theta_i^{m+1} = \theta_i^{(m)} - g\left(\theta_i^{(m)}\right) / h\left(\theta_i^{m}\right)$$

เมื่อ $\theta_i^{m+i}, \theta_i^{(m)}$ = ค่าประมาณความสามารถของคนที่ i ครั้งที่ m + 1และ m

$$g\left(\theta_i^{(m)}\right) = d \ln f(\theta, b, a, c / u) d\theta_i$$

$$g\left(\theta_i^{(m)}\right) = d^2 \ln f(\theta, b, a, c / u) d\theta_i^2$$

การประมาณค่าซ้ำ (Iterative) จะกระทำจนกว่าค่าประมาณความสามารถ θ_i จะลู่เข้าสู่ค่าคงที่ค่าใดค่าหนึ่ง (Convergence) คือ ค่าประมาณครั้งที่ m + 1 และครั้งที่ m มีค่าแตกต่างกันน้อยกว่าค่าคงที่ ที่กำหนดไว้ เช่น 0.001

<u>ขั้นที่ 3</u> ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ a_j, b_j และ c_j โดยกำหนดให้ค่าความสามารถของผู้เข้าสอบที่ประมาณค่าได้ในครั้งก่อนเป็นค่าคงที่

$$a_j^{m+1} = a_j^{(m)} - g\left(a_j^{(m)}\right) / h\left(a_j^{(m)}\right)$$

$$b_j^{m+1} = b_j^{(m)} - g\left(b_j^{(m)}\right) / h\left(b_j^{(m)}\right)$$

$$c_j^{m+1} = c_j^{(m)} - g\left(c_j^{(m)}\right) / h\left(c_j^{(m)}\right)$$

เมื่อ $a_j^m, a_j^{(m+1)}$ = ค่าประมาณอำนาจจำแนกของข้อสอบข้อที่ j ครั้งที่ m และ m + 1

$b_j^m, b_j^{(m+1)}$ = ค่าประมาณความยากของข้อสอบข้อที่ j ครั้งที่ m และ m + 1

$c_j^m, c_j^{(m+1)}$ = ค่าประมาณการเดาของข้อสอบข้อที่ j ครั้งที่ m และ m + 1

$$g(a_j^{(m)}) = d \ln f(\theta, b, a, c / u) / d\ a_j$$

$$h(a_j^{(m)}) = d^2 \ln f(\theta, b, a, c / u) / d\ a_j^2$$

$$g(b_j^{(m)}) = d \ln f(\theta, b, a, c / u) / d\ b_j$$

$$h(b_j^{(m)}) = d^2 \ln f(\theta, b, a, c / u) / d\ b_j^2$$

$$g(c_j^{(m)}) = d \ln f(\theta, b, a, c / u) / d\ c_j$$

$$h(c_j^{(m)}) = d^2 \ln f(\theta, b, a, c / u) / d\ c_j^2$$

การประมาณค่าซ้ำ (Iterative) จะทำจนกว่าค่าประมาณ จะลู่เข้าสู่ค่าคงที่ค่าใดค่าหนึ่ง (Convergence) สำหรับค่าของอนุพันธ์อันดับ 1 g (.) และอนุพันธ์อันดับ 2 h (.) มีส่วนประกอบสองส่วน คือ ส่วนที่ได้จากฟังก์ชันไลค์ลิฮู้ด และส่วนที่ได้จากการแจกแจงเริ่มแรก ซึ่งส่วนประกอบทั้งสองส่วนนี้แสดงไว้ในตารางที่ 8

<u>ขั้นที่ 4</u> ประมาณค่าซ้ำ ชั้นที่ 2 และชั้นที่ 3 จนกว่าค่าประมาณ θ_i, a_j, b และ c_j จะมีค่าคงที่ และมีความถูกต้องเพียงพอ หรือ ทำให้ $f(\theta, b, a, c/u)$ มีค่าสูงสุด

ตารางที่ 8 อนุพันธ์อันดับที่ 1 และอนุพันธ์อันดับที่ 2 ของ $\ln f(\theta, b, a, c/u)$

พารามิเตอร์	ฟังก์ชัน	อนุพันธ์อันดับที่ 1	อนุพันธ์อันดับที่ 2
θ_i	Likelihood	$D\sum_{j=1}^{n} a_j\left(P_{ij}-c_j\right)\left(U_{ij}-P_{ij}\right) / P_{ij}(1-c_{ij})$	$D^2\sum_{j=1}^{n} a_j^2\left(P_{ij}-c_j\right)\left(U_{ij}-P_{ij}^2\right) Q_{ij} / P_{ij}^2(1-c_j)^2$
	Prior	$-\theta_i$	-1
a_j	Likelihood	$D\sum_{i=1}^{N}\left(\theta_i-b_j\right)\left(P_{ij}-c_j\right) \left(U_{ij}-P_{ij}\right)/P_{ij}(1-c_j)$	$D^2\sum_{i=1}^{N}\left(\theta_i-b_j\right)^2\left(P_{ij}-c_j\right) \left(U_{ij}c_j-P_{ij}^2\right)Q_{ij}/P_{ij}^2(1-c_j)^2$
	Prior	$+\left(v_j-1\right)/(a_j-1)/w_j$	$-\left(v_j-1\right)/(a_j^2-1)/w_j$
b_j	Likelihood	$-D\sum_{i=1}^{N} a_j\left(P_{ij}-c_j\right)\left(U_{ij}-P_{ij}\right) / P_{ij}(1-c_j)$	$D^2\sum_{i=1}^{N} a_j^2\left(P_{ij}-c_j\right)\left(U_{ij}c_j-P_{ij}^2\right) Q_{ij}/P_{ij}^2(1-c_j)^2$
	Prior	-	-
c_j	Likelihood	$\sum_{i=1}^{N}\left(U_{ij}-P_{ij}\right)/P_{ij}(1-c_j)$	$\sum_{i=1}^{N}\left[U_{ij}(2P_{ij}-1)-P_{ij}^2\right] / P_{ij}^2(1-c_j)^2$
	Prior	$+s_j/c_j-t_j/(1-c_j)$	$-s_j/c_j^2-t_j/(1-c_j)^2$

$P_{ij}=c_j+(1-c_j)/(1+\exp(-Da_j(\theta_i-b_j)))$

โดยสรุปการประมาณค่าพารามิเตอร์ด้วยวิธีการของเบส์ เป็นการประมาณค่าที่ประกอบไปสองส่วนที่สำคัญ คือ ส่วนที่ใช้ข้อมูลในปัจจุบันมาใช้ในการประมาณค่า ซึ่งก็คือ ส่วนการประมาณค่าด้วยวิธีแม็กซิมั่มไลค์ลิฮูด และส่วนที่เป็นข้อมูลในอดีต ซึ่งก็คือ ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ที่ต้องการประมาณค่า

บทสรุปจากการศึกษาการประมาณค่าพารามิเตอร์การวิเคราะห์ข้อสอบ

จากการศึกษาวิธีการประมาณค่าพารามิเตอร์ที่ใช้ในการวิเคราะห์ข้อสอบที่ได้รับความนิยมซึ่งมีอยู่ด้วยกัน 3 วิธี ได้แก่ วิธีฮิวริสติก (Heuristic) แม็กซิมั่มไลค์ลิฮูด (Maximum Likelihood) และวิธีของเบส์ (Bayesian) (Hambleton และ Cook, 1977; Lord, F.M., 1980; รัตนา ศรีเหรัญ, 2539 ; ศิริชัย กาญจนวาสี, 2550; สุนทร เทียนงาม, 2551) สามารถสรุปได้ว่า การประมาณค่าด้วยวิธีของเบส์นั้น สามารถประมาณค่าพารามิเตอร์ของข้อสอบที่ผู้เข้าสอบทุกคนตอบถูกหรือตอบผิดได้ และประมาณค่าความสามารถของผู้เข้าสอบที่ทำข้อสอบถูกหรือผิดทุกข้อได้ด้วย โดยเฉพาะในกลุ่มตัวอย่างขนาดเล็กประมาณ 200-300 คน (Swaminathan และ Gifford) นอกจากนี้จากการศึกษาของ Swaminathan และ Gifford (1982, 1985, 1986) พบว่า วิธีการประมาณค่าของเบส์มีความเหนือกว่าวิธีการประมาณค่าด้วยวิธีแม็กซิมั่มไลค์ลิฮูด ในเกือบทุกสถานการณ์ พิจารณาได้จากค่า MSD (Mean Squared Difference between estimates and true value) นอกจากนี้วิธีการประมาณค่าด้วยวิธีแม็กซิมั่มไลค์ลิฮูดยังไม่สามารถประมาณค่าได้คงที่ (No convergence) หรืออาจจะประมาณได้ค่าคงที่ แต่ค่าคงที่ที่ได้ไม่ได้ตกอยู่ในขอบเขตของค่าพารามิเตอร์ที่จะยอมรับได้ ส่วนวิธีการประมาณค่าแบบฮิวริสติกเป็นวิธีการประมาณค่าที่คำนวณได้ง่ายเนื่องจากทำการประมาณค่าพารามิเตอร์ของข้อสอบได้จากค่าสถิติที่ได้จากทฤษฎีการทดสอบแบบดั้งเดิม (CTT) แต่ผลที่ได้จากการประมาณค่ามีความคลาดเคลื่อนมากและในการประมาณค่าต้องมีกลุ่ม ผู้เข้าสอบขนาดใหญ่และแบบสอบต้องมีความเที่ยงสูง (Schmidt, 1977) ดังนั้นเมื่อเปรียบเทียบวิธีการประมาณค่าทั้งสามวิธี พบว่า วิธีการประมาณค่าของเบส์เป็นวิธีการที่ดีที่สุดสำหรับการประมาณพารามิเตอร์ของผู้สอบและพารามิเตอร์ของข้อสอบของโมเดลการตอบสนองข้อสอบ

เมื่อเปรียบเทียบคุณลักษณะที่สำคัญเกี่ยวกับวิธีการประมาณค่าทั้ง 3 วิธี ซึ่งได้แก่ การประมาณค่าด้วยวิธีวิธีแม็กซิมั่มไลค์ลิฮูด (Maximum Likelihood) วิธีฮิวริสติค (Heuristic) และวิธีของเบส์ (Bayesian) สามารถสรุปความเหมือนและความแตกต่างของวิธีการประมาณค่าทั้งสามได้ดังตารางที่ 9

ตารางที่ 9 การสรุปความเหมือนและความแตกต่างของกระบวนการประมาณค่าพารามิเตอร์ด้วยวิธีแม็กซิมั่มไลค์ลิฮูด วิธีฮิวริสติค และวิธีของเบส์

การประมาณค่า	วิธีฮิวริสติค	วิธีแมกซิมัมไลค์ลิฮูด	วิธีของเบส์
1. ใช้วิธีการประมาณค่าพารามิเตอร์ในแบบจำลองโลจิสติค	X	X	X
2. ใช้แนวคิดฟังก์ชันสารสนเทศของแบบทดสอบได้	X	X	X
3.ใช้การประมาณค่าแบบซ้ำๆ จนกว่าจะได้ค่าคงที่		X	X
4. ใช้ฟังก์ชันไลค์ลิฮูดในสมการประมาณค่า		X	X
5. ใช้ข้อมูลผลการตอบเพียงอย่างเดียวเพื่อประมาณค่าพารามิเตอร์		X	
6. ประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้เข้าสอบไปพร้อมๆกัน		X	X
7. ประมาณค่าความสามารถของบุคคลด้วยวิธีแม็กซิมั่มไลค์ลิฮูด	X	X	
8. ประมาณค่า θ ด้วยวิธีแม็กซิมั่มไลค์ลิฮูดอย่างมีเงื่อนไข	X		
9. ใช้แนวคิดทางสถิติแบบนอนเบส์เซียน	X	X	
10. ใช้แนวคิดทางสถิติแบบเบส์เซียน			X
11. มีการกำหนดขอบเขตของค่าพารามิเตอร์	X	X	

ตอนที่ 3 การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล

การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลนับเป็นสาขาหนึ่งของวิชาคณิตศาสตร์เชิงทดลอง ซึ่งเป็นเทคนิคที่ใช้สำหรับการแก้ปัญหาในการคำนวณทางคณิตศาสตร์สำหรับสถานการณ์ที่มีกระบวนการที่ซับซ้อน ซึ่งการวิเคราะห์ทั่วไปไม่สามารถดำเนินการได้อย่างครบถ้วนและสมบูรณ์ โดยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลสามารถให้ผลสรุปจากสภาพการณ์ที่สร้างขึ้นใน การทดลองที่แสดงเหตุการณ์เกิดขึ้นทั้งทางบวกและทางลบ เนื่องจากเทคนิคนี้สามารถจำลองข้อมูลได้หลายๆ ครั้งจากข้อมูลลักษณะเดียวกัน ผลการศึกษาจึงสามารถสะท้อนให้เห็นถึงเหตุการณ์ที่น่าจะเกิดขึ้นได้จริง โดยหลักการในการจำลองจะสร้างตัวแบบ (model) โดยการใช้เลขสุ่ม (Random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา โดยการนำเสนอในตอนนี้จะนำเสนอแบ่งออก เป็น 6 ส่วน ได้แก่ ส่วนแรกเป็นการนำเสนอความเป็นมาของเทคนิคมอนติคาร์โล ส่วนที่สอง

เกี่ยวกับขั้นตอนในการจำลองข้อมูล ส่วนที่สามเกี่ยวกับการประยุกต์ใช้เทคนิคมอนติคาร์โลในการวิเคราะห์ข้อสอบ ส่วนที่สี่ประสิทธิภาพของการจำลองข้อมูล ส่วนที่ห้าข้อดีและข้อจำกัดของการศึกษาด้วยเทคนิคการจำลองข้อมูล และส่วนสุดท้าย บทสรุปจากการศึกษาการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล มีรายละเอียดในแต่ละส่วนดังนี้

3.1 ความเป็นมาของการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล

การจำลองข้อมูลเริ่มต้นขึ้นเมื่อปี ค.ศ. 1777 โดยบุฟฟอง เริ่มแรกเขาได้ทำการทดลองเพื่อสร้างสมการจากความสัมพันธ์ระหว่างความยาวของเข็มกับระยะทางระหว่างเส้นขนานทั้งสอง หลังจากนั้นจึงได้มีความพยายามในการพิสูจน์สูตรคณิตศาสตร์ที่เขาได้คิดค้นขึ้นมา จึงทำให้มีการจำลองข้อมูลเกิดขึ้น ส่วนการจำลองข้อมูลที่เรียกว่าการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลนั้นเริ่มปรากฏขึ้นในปี 1940 โดยนักวิทยาศาสตร์ในชั้นเรียนทางคณิตศาตร์ที่ทำงานการพัฒนาอาวุธนิวเครียร์ในลอส อาลามอส

การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลเป็นเทคนิคที่สร้างขึ้นมาจากทฤษฎีความน่าจะเป็น (Probability theory) และทฤษฎีการสุ่ม (Theory of random) โดยทั้งสองทฤษฎีดังกล่าวเป็นรากฐานที่สำคัญของการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ซึ่งพัฒนาการของการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลสามารถแบ่งออกได้เป็น 3 ช่วงหลัก ๆ ได้แก่ ช่วงแรกเป็นช่วงก่อนการใช้เทคนิคมอนติคาร์โล ช่วงนี้เป็นช่วงที่ใช้ในการศึกษาปัญหาต่างๆ ทางสถิติ เช่น การสุ่มตัวอย่างทางสถิติ เกี่ยวกับการแจกแจงของค่าสัมประสิทธิ์สหสัมพันธ์ การแจกแจงสถิติที ช่วงที่ 2 เป็นช่วงที่มีศัพท์คำว่า เทคนิคมอนติคาร์โล เกิดขึ้น โดย Metropolis และ Ulam เป็นผู้ใช้ศัพท์คำว่า “Monte Carlo” เป็นครั้งแรก โดยใช้เทคนิคนี้เป็นเครื่องมือวิจัยเกี่ยวกับระเบิดปรมณูในยุคสงครามโลกครั้งที่ 2 ส่วนช่วงที่ 3 เริ่มต้นประมาณปี ค.ศ.1970 เป็นช่วงนี้ที่มีการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล โดยการนำคอมพิวเตอร์ความเร็วสูงมาใช้ ในการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล จึงทำให้การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลเริ่มเป็นที่นิยมในการศึกษาเพื่อแก้ไขปัญหาทางสถิติ

3.2 ขั้นตอนการจำลองข้อมูล

การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล เป็นวิธีการจำลองข้อมูลด้วยการสร้างตัวแบบ (model) โดยการใช้เลขสุ่ม (Random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา สามารถสรุปขั้นตอนที่สำคัญในการจำลองข้อมูลออกเป็น 3 ขั้นตอนใหญ่ๆ ดังนี้

3.2.1 การสร้างตัวเลขสุ่ม ขั้นตอนนี้ถือเป็นขั้นตอนที่สำคัญมากสำหรับเทคนิคมอนติคาร์โล เนื่องจาก การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลนั้นจะใช้เลขสุ่มมาช่วยในการหาคำตอบให้กับปัญหาโดยลักษณะของตัวเลขสุ่มที่นำมาใช้ จะมีการแจกแจงแบบสม่ำเสมอ (uniform distribution) ในช่วง (0, 1) สำหรับวิธีการสร้างตัวเลขสุ่มที่ดีนั้น ลักษณะของเลขสุ่มที่ถูกสร้างขึ้นจะต้องมีการแจกแจงแบบสม่ำเสมอในช่วง (0, 1) ตัวเลขสุ่มแต่ละตัวเป็นอิสระกันและมีช่วงยาวก่อนจะเกิดเลขสุ่มซ้ำ (มีวัฏจักรยาว)

เนื่องด้วยการสร้างตัวแปรสุ่ม (Random Variable) นับว่าเป็นขั้นตอนที่สำคัญของการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ดังนั้นในขั้นตอนนี้จึงมีการนำเสนอเกี่ยวกับตัวแปรสุ่มเพื่อให้เข้าใจเกี่ยวกับลักษณะของตัวแปรสุ่มจึงแบ่งการนำเสนอตัวแปรสุ่มออกเป็น 3 ส่วน ส่วนแรกเป็นการนำเสนอความหมายของตัวแปรสุ่ม ส่วนที่สองเป็นการนำเสนอประเภทของตัวแปรสุ่ม และส่วนสุดท้ายเป็นการนำเสนอการแจกแจงความน่าจะเป็นของตัวแปรสุ่ม โดยมีรายละเอียดในแต่ละส่วนของตัวแปรสุ่มดังนี้

1) ความหมายของตัวแปรสุ่ม

ตัวแปรสุ่ม เป็นฟังก์ชันที่มีค่าเป็นจำนวนจริง (Real-valued function) ซึ่งกำหนดโดยแต่ละสมาชิกในปริภูมิตัวอย่างจากการลองสุ่ม (random trial) ซึ่งหมายถึงการกระทำใดๆ ที่ผู้กระทำไม่สามารถทราบผลลัพธ์ล่วงหน้าจนกว่าจะได้ทำเสร็จสิ้นไปแล้ว จึงจะทราบผลลัพธ์ที่ถูกต้อง (คณาจารย์ ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2551)

2) ประเภทของตัวแปรสุ่ม

ตัวแปรสุ่ม สามารถแบ่งออกได้เป็น 2 ชนิด ได้แก่

2.1) ตัวแปรสุ่มไม่ต่อเนื่อง (discrete random variable) คือ ตัวแปรสุ่มที่มีค่าเป็นจำนวนที่นับได้ครบ หรือ เป็นตัวแปรที่มีค่าเป็นจำนวนที่จับคู่ 1-1 กับจำนวนเต็มได้ทั้งหมด ปริภูมิตัวอย่างของตัวแปรสุ่มประเภทนี้จะเรียกว่า ปริภูมิตัวอย่างที่ไม่ต่อเนื่อง (discrete sample space)

2.2) ตัวแปรสุ่มต่อเนื่อง (continuous random variable) คือ ตัวแปรสุ่มที่มีค่าต่อเนื่องกันได้หลายค่านับไม่ถ้วน ปริภูมิตัวอย่างที่มีจุดตัวอย่างเป็นตัวแปรสุ่มต่อเนื่อง เรียกว่า ปริภูมิตัวอย่างที่ต่อเนื่อง (continuous sample space) ดังนั้นจะเห็นได้ว่าคะแนนมาตรฐาน (Z-score) ที่ได้จากการทดสอบ หรือค่าพารามิเตอร์ที่ได้จากการวิเคราะห์ข้อสอบ เป็นคะแนนสุ่มแบบต่อเนื่อง

3) การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม

การแจกแจงความน่าจะเป็นของตัวแปรสุ่ม มีหลายประเภท แต่ในการนำเสนอครั้งนี้ จะเป็นการนำเสนอที่พบบ่อยสำหรับการวิเคราะห์ข้อสอบ ซึ่งได้แก่ การแจกแจงแบบยูนิฟอร์ม (Uniform distribution) การแจกแจงแบบปกติ (Normal distribution) การแจกแจงแบบปกติหาค่าโดยประมาณของความน่าจะเป็นของตัวแปรสุ่มที่มีการแจกแจงทวินาม การแจกแจงแกมมา (Gamma distribution) การแจกแจงแบบเบต้า (Beta distribution) การแจกแจงแบบไคสแควร์ (Chi-square distribution) และการแจกแจงแบบที (t-distribution) มีรายละเอียดเกี่ยวกับการแจกแจงแต่ละประเภทดังนี้

3.1) การแจกแจงแบบยูนิฟอร์ม (Uniform distribution)

การแจกแจงแบบยูนิฟอร์มมีลักษณะการแจกแจงที่มีค่าได้ทุกค่าจริงในช่วง [a,b], $-\infty < a,b < \infty$ ด้วยความน่าจะเป็นเท่าๆ กัน โดยสามารถเขียนแทนด้วยสัญลักษณ์ X ~U (a,b) โดยมีฟังก์ชันความน่าจะเป็นดังนี้

$$\mathrm{f}(x) = \begin{cases} \dfrac{1}{b-a} & \text{เมื่อ } a \langle x \langle b \\ & \text{เมื่อ X มีค่าอื่นๆ} \end{cases}$$

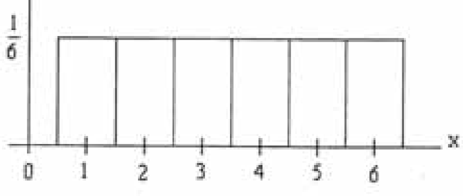
การแจกแจงแบบยูนิฟอร์มจะมีค่าเฉลี่ยของ X เท่ากับ

$$\mu = \frac{a+b}{2}$$

การแจกแจงแบบยูนิฟอร์มจะค่าความแปรปรวนของ X เท่ากับ

$$\sigma^2 = \frac{(b-a)^2}{12}$$

การแจกแจงแบบยูนิฟอร์ม สามารถแสดงตัวอย่างกราฟของฟังก์ชันการแจกแจงแบบยูนิฟอร์ม สำหรับค่าพารามิเตอร์บางค่าได้ดังนี้



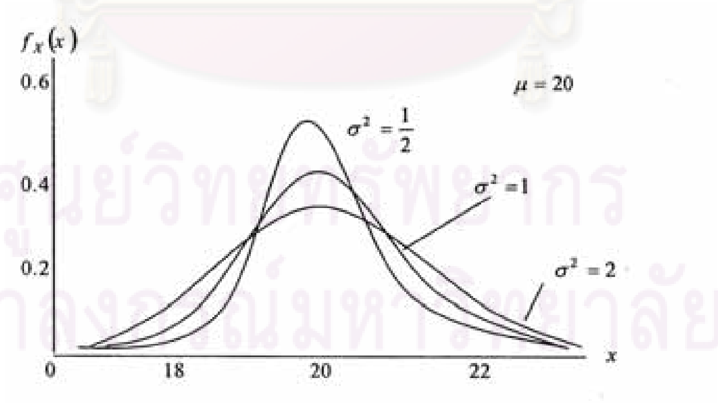
แผนภาพที่ 10 การแจกแจงแบบยูนิฟอร์ม

3.2) **การแจกแจงแบบปกติ** (Normal distribution)

การแจกแจงแบบปกติ นับเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มต่อเนื่องที่สำคัญที่สุด (คณาจารย์ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย, 2551; มานพ วราภักดิ์, 2548) เนื่องจากเป็นการแจกแจงความน่าจะเป็นของตัวแปรสุ่มที่ส่วนมากจะมีค่าใกล้เคียงค่าเฉลี่ยของตัวแปรเหล่านั้น จะมีค่าของตัวแปรที่มากกว่าหรือน้อยกว่าค่าเฉลี่ยเป็นส่วนน้อย การแจกแจงแบบปกตินี้ บางครั้งอาจจะเรียกว่า การแจกแจงแบบเกาส์ (Gaussian distribution) ตัวแปรสุ่มที่มีการแจกแจงแบบปกติ เรียกว่า ตัวแปรสุ่มปกติ (normal randon variable) โดยสามารถเขียนแทนด้วยสัญลักษณ์ $X \sim N(\mu,\sigma^2)$ โดยมีฟังก์ชันความน่าจะเป็นดังนี้

$$f(x)=\frac{2}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{เมื่อ } -\infty \langle x \langle \infty$$

การแจกแจงแบบปกติมจะมีค่าเฉลี่ยของ X เท่ากับ μ และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ σ สามารถแสดงตัวอย่างกราฟของฟังก์ชันการแจกแจงปกติ สำหรับค่าพารามิเตอร์บางค่าได้ดังนี้



แผนภาพที่ 11 การแจกแจงแบบปกติ

3.3) การแจกแจงแบบปกติหาค่าโดยประมาณของความน่าจะเป็นของตัวแปรสุ่มที่มีการแจกแจงทวินาม

การแจกแจงของตัวแปรสุ่ม (x) ที่มีการแจกแจงแบบทวินาม b (x ; n,p) เมื่อมีค่า n น้อยจะหาค่า b (x ; n,p) ได้จากการแจกแจงปัวซอง (Poisson distribution) โดยการประมาณของความน่าจะเป็นที่มีการแจกแจงทวินามเมื่อ n มีค่ามาก และ p มีค่าใกล้ 0 หรื 1 แต่ในครั้งนี้จะใช้การแจกแจงแบบปกติหาค่าโดยประมาณของความน่าจะเป็นของตัวแปรสุ่มที่มีการแจกแจงทวินามหาค่าโดยประมาณของ b (x ; n,p) เมื่อ n มีค่ามากพอ โดยสามารถเขียนแทนด้วยสัญลักษณ์ Z ~ $N(0,1)$ โดยมีฟังก์ชันความน่าจะเป็นดังนี้

$$Z = \frac{X - np}{\sqrt{npq}}$$

การแจกแจงจะมีค่าเฉลี่ยของ X เท่ากับ $\mu = np$ และมีค่าความแปรปรวนของ X เท่ากับ $\sigma^2 = npq$

3.4) **การแจกแจงแกมมา** (Gamma distribution)

การแจกแจงแบบแกมมามีลักษณะการแจกแจงในลักษณะเบ้ขวา (positive skewness) โดยตัวแปรสุ่มที่มีการแจกแจงแบบแกมมา ค่าพารามิเตอร์ $\alpha > 0$ และ $\beta > 0$ โดยสามารถเขียนแทนด้วยสัญลักษณ์ X ~G(α, β) โดยมีฟังก์ชันความน่าจะเป็นดังนี้

$$f(x) = \begin{cases} \dfrac{1}{\beta^{\propto}\gamma(\propto)} x^{\propto-1} e^{\frac{-x}{\mathrm{B}}} & ; \mathrm{X} > 0 \\ 0 & ; \mathrm{X} = \text{อื่นๆ} \end{cases}$$

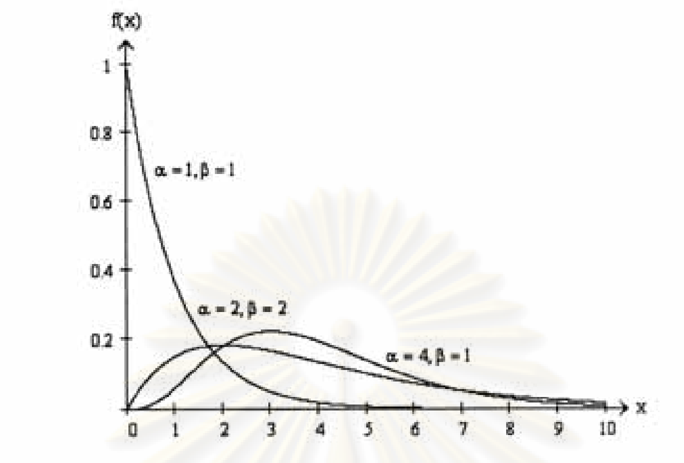
การแจกแจงแบบแกมมาจะมีค่าเฉลี่ยของ X เท่ากับ

$$\mu = \frac{\propto}{\lambda} = \frac{\beta\gamma(\propto + 1)}{\gamma(\propto)} = \propto \beta$$

การแจกแจงแบบแกมมาจะมีค่าความแปรปรวนของ X เท่ากับ

$$\sigma^2 = \frac{\propto}{\lambda^2} = \propto \beta^2$$

การแจกแจงแบบแกมมา สามารถแสดงตัวอย่างกราฟของฟังก์ชันการแจกแจงแกมมา สำหรับค่าพารามิเตอร์บางค่าได้ดังนี้



แผนภาพที่ 12 การแจกแจงแบบแกมมา

3.6) **การแจกแจงแบบเบต้า** (Beta distribution)

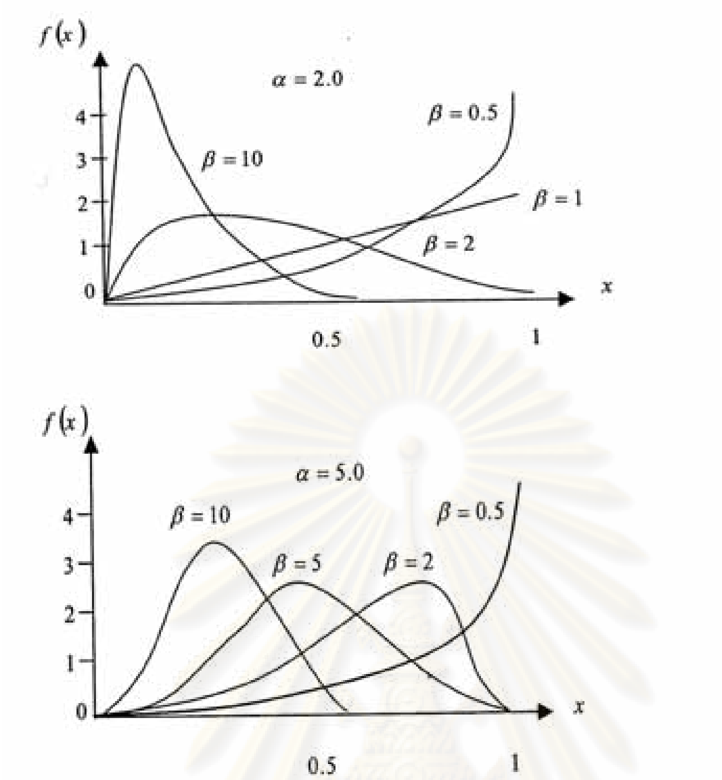
การแจกแจงแบบแกมมามีลักษณะการแจกแจงในลักษณะเบ้ซ้าย (negative skewness) โดยตัวแปรสุ่มที่มีการแจกแจงแบบเบต้า ค่าพารามิเตอร์ $\alpha > 0$ และ $\beta > 0$ โดยสามารถเขียนแทนด้วยสัญลักษณ์ X ~Be (α, β) โดยมีฟังก์ชันความน่าจะเป็นดังนี้

$$\mathrm{B}(\propto, \beta) = \int_0^1 x^{\propto-1}(1-x)^{\beta-1}dx \text{ เมื่อ } \propto\rangle 0, \beta\rangle 0$$

$$f(x)\begin{cases} \dfrac{x^{\propto-1}(1-x)^{\beta-1}}{\mathrm{B}(\propto,\beta)} & ;\ 0 \le x \le 1 \\ 0 & ;\ \mathrm{x} = \text{อื่นๆ} \end{cases}$$

การแจกแจงแบบเบต้าจะมีค่าเฉลี่ยของ X เท่ากับ $\mu = \dfrac{\propto}{\propto + \beta}$ และมีค่าความแปรปรวนของ X เท่ากับ $\sigma^2 = \dfrac{\propto \beta}{(\propto + \beta + 1)(\propto + \beta)^2}$

การแจกแจงแบบเบต้า สามารถแสดงตัวอย่างกราฟของฟังก์ชันการแจกแจงแกมมา สำหรับค่าพารามิเตอร์บางค่าได้ดังนี้



แผนภาพที่ 13 การแจกแจงแบบเบต้า

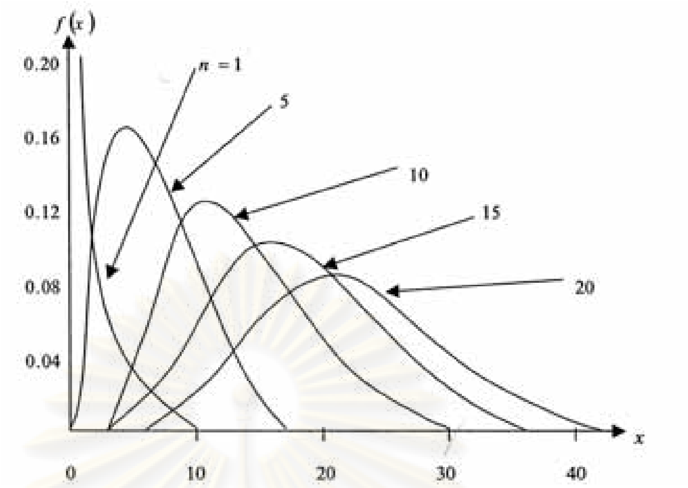
3.7) **การแจกแจงแบบไคสแควร์** (Chi-square distribution)

การแจกแจงแบบไคสแควร์ จัดเป็นการแจกแจงกรณีหนึ่งของการแจกแจงแบบแกมมา G(α,λ) ในกรณีที่ λ = 0.5 และ α = n/2 โดย n เป็นค่าพารามิเตอร์ที่มีจำนวนเต็มบวก และเรียก n ว่าเป็นระดับขั้นความเสรี หรือ จำนวนองศาความเสรี (Number of degree of freedom; df) การแจกแจงแบบนี้จะให้ประโยชน์มากสำหรับการอนุมานเชิงสถิติ โดยสามารถเขียนแทนด้วยสัญลักษณ์ X ~ $\chi^2(n)$ โดยมีฟังก์ชันความน่าจะเป็นดังนี้

$$f(\chi) = \begin{cases} \dfrac{1}{2^{\frac{v}{2}}\gamma(\frac{v}{2})} x^{\frac{v}{2}-1} e^{\frac{-x}{2}} \\ 0 \end{cases}$$

การแจกแจงแบบไคสแควร์จะมีค่าเฉลี่ยของ X เท่ากับ $\mu = v$ และมีค่าความแปรปรวนของ X เท่ากับ $\sigma^2 = 2v$

การแจกแจงแบบไคสแควร์ สามารถแสดงตัวอย่างกราฟของฟังก์ชันการแจกแจงไคสแควร์สำหรับค่าพารามิเตอร์บางค่าได้ดังนี้



แผนภาพที่ 14 การแจกแจงแบบไคสแควร์

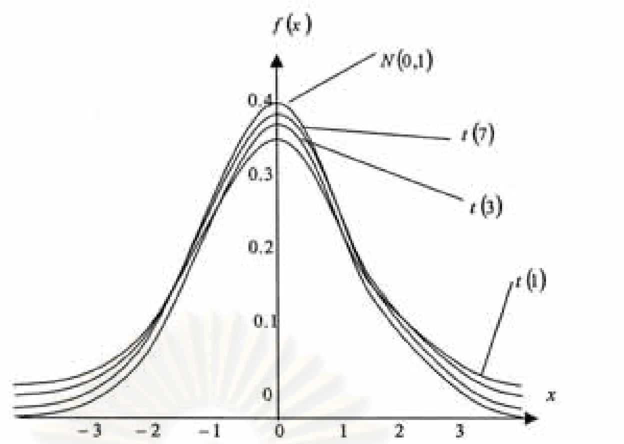
3.8) **การแจกแจงแบบที** (t- distribution)

การแจกแจงแบบที บางครั้งอาจเรียกว่าการแจกแจงทีของสติวเดนท์ (Student's t-distribution) ตัวแปรสุ่ม x ที่มีการแจกแจงความน่าจะเป็นที ด้วยระดับขั้นความเสรี n สามารถเขียนแทนด้วยสัญลักษณ์ X ~ t (n) โดยมีฟังก์ชันความน่าจะเป็นดังนี้

$$h(t) = \frac{\gamma(\frac{\gamma+1}{2})}{\gamma\left(\frac{v}{2}\right)\sqrt{\pi v}}\left(1+\frac{t^2}{v}\right)^{\frac{-v+1}{2}} \quad \text{เมื่อ } -\infty < t < \infty$$

การแจกแจงแบบทีจะมีค่าเฉลี่ยของ X เท่ากับ $\mu = 0$ และมีค่าความแปรปรวนของ X เท่ากับ $\sigma^2 = \frac{v}{v-2}$ เมื่อ $v \geq 3$

การแจกแจงแบบแบบที สามารถแสดงตัวอย่างกราฟของฟังก์ชันการแจกแจงแบบทีสำหรับค่าพารามิเตอร์บางค่าได้ดังนี้



แผนภาพที่ 15 การแจกแจงแบบที

3.2.2 การประยุกต์ปัญหาที่ต้องการศึกษามาใช้กับตัวเลขสุ่ม ขั้นตอนนี้ขึ้นอยู่กับลักษณะของปัญหาที่ศึกษา บางปัญหาอาจจะไม่ใช้ตัวเลขสุ่มโดยตรง แต่อาจจะมีขั้นตอนอื่นอีกหลายๆ ขั้นตอน ซึ่งขั้นตอนเหล่านี้มีบางขั้นตอนที่ต้องใช้ตัวเลขสุ่ม

3.2.3 การทดลองกระทำการสุ่ม เป็นการทดลองโดยใช้กระบวนการสุ่ม (random process) มากระทำในลักษณะซ้ำๆกัน (replication) เพื่อประมาณค่าที่แท้จริง โดยจะใช้ความคลาดเคลื่อนมาตรฐาน (MSE) หรือ รากที่สองของความคลาดเคลื่อนมาตรฐาน (RMSD) ในการประเมินประสิทธิภาพของการประมาณค่า

3.3 การประยุกต์ใช้เทคนิคมอนติคาร์โลในการวิเคราะห์ข้อสอบ

การวิเคราะห์ข้อสอบในปัจจุบัน โดยเฉพาะการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ มีการประยุกต์ใช้เทคนิคมอนติคาร์โลในการวิเคราะห์เป็นจำนวนมาก โดยเฉพาะในการประเด็นเกี่ยวกับการประเมินกระบวนการในการประมาณค่าหรือความครอบคลุมของค่าพารามิเตอร์ นอกจากนี้ยังมีการประยุกต์ใช้เทคนิคมอนติคาร์โลในการประเมินคุณสมบัติทางสถิติของโมเดลในการวิเคราะห์ข้อสอบ รวมทั้งยังมีการประยุกต์ใช้เทคนิคมอนติคาร์โลในการเปรียบเทียบวิธีการต่างๆ ในโมเดลการวิเคราะห์ข้อสอบ ไม่ว่าจะเป็นการทำหน้าที่ต่างกันของข้อสอบหรือการประเมินผลที่ได้จากการวัดหลายมิติ เป็นต้น (Harwell และคณะ, 1996) ดังนั้นจะเห็นได้ว่าการประยุกต์ใช้เทคนิคมอนติคาร์โลในการวิเคราะห์ข้อสอบจึงเป็นเรื่องที่สำคัญและนักวิจัยให้ความนิยมกันเป็นอย่างมาก โดยในครั้งนี้จะมีการนำเสนอแบ่งออกเป็น 3 ส่วนคือ ส่วนแรกเกี่ยวกับประเด็นที่พบบ่อยสำหรับการประยุกต์ใช้เทคนิคมอนติคาร์โลในการวิเคราะห์ข้อสอบ ส่วนที่สองเกี่ยวกับขั้นตอนในการวิเคราะห์ข้อสอบด้วยเทคนิคมอนติคาร์โล และส่วนสุดท้ายเกี่ยวกับข้อพึงระวังสำหรับการวิเคราะห์ข้อสอบด้วยการใช้เทคนิคมอนติคาร์โล

3.3.1 ประเด็นที่พบบ่อยในการประยุกต์ใช้เทคนิคมอนติคาร์โลในการวิเคราะห์ข้อสอบ

การวิเคราะห์ข้อสอบในปัจจุบันมีการประยุกต์ใช้เทคนิคมอนติคาร์โลมาก โดยเฉพาะการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) โดยส่วนใหญ่จะมีการประยุกต์ใช้เทคนิคเทคนิคมอนติคาร์โลสำหรับการวิเคราะห์ข้อสอบ ซึ่งสามารถสรุปการประยุกต์สำหรับการวิเคราะห์ข้อสอบได้ 3 ประเด็น ดังนี้

1) การประเมินประสิทธิภาพของวิธีในการประมาณค่าหรือความครอบคลุมของค่าพารามิเตอร์ที่ได้จากการประมาณด้วยโมเดลการวิเคราะห์ข้อสอบ

2) การประเมินคุณสมบัติของสถิติที่ใช้ในการวิเคราะห์ของโมเดลในการวิเคราะห์ข้อสอบว่ามีความเหมาะสมหรือไม่

3) การเปรียบเทียบวิธีการต่างๆ ในโมเดลการวิเคราะห์ข้อสอบ ไม่ว่าจะเป็นการทำหน้าที่ต่างกันของข้อสอบ หรือการประเมินผลที่ได้จากการวัดหลายมิติ เป็นต้น

ทั้งนี้เนื่องจากการวิเคราะห์การประยุกต์ใช้เทคนิคมอนติคาร์โลในการวิเคราะห์ข้อสอบสามารถวิเคราะห์เพื่อตอบคำถามตามวัตถุประสงค์ที่ซับซ้อน โดยสามารถสรุปลำดับขั้นตอนการวิเคราะห์ของเทคนิคมอนติคาร์โลได้ 8 ขั้นตอนตามลำดับดังนี้

1) การตอบคำถามวิจัยโดยสามารถแบ่งคำถามวิจัยออกเป็นวัตถุประสงค์ที่เฉพาะเจาะจงได้ เช่น เพื่อศึกษาผลของจำนวนข้อสอบและผู้สอบที่แตกต่างกัน เพื่อลักษณะการแจกแจงค่าพารามิเตอร์ข้อสอบที่ส่งผลต่อการประมาณค่าพารามิเตอร์ของโมเดลการวิเคราะห์ข้อสอบ เป็นต้น

2) การกำหนดเงื่อนไขประเภทต่างๆ ของตัวแปรต้นที่ส่งผลต่อตัวแปรตามได้ เช่น จำนวนข้อสอบ ส่งผลต่อค่าพารามิเตอร์ของผู้สอบหรือไม่

3) การออกแบบการทดลองให้มีความสอดคล้องและเหมาะสมกับวัตถุประสงค์ของการวิจัย

4) การจำลองข้อมูลให้สอดคล้องกับเงื่อนไขของโมเดลการวิเคราะห์ข้อสอบ

5) การประมาณค่าพารามิเตอร์โดยใช้ข้อมูลจากการจำลอง

6) การเปรียบเทียบผลที่ได้จากการประมาณค่า โดยใช้ค่าสถิติต่าง ๆได้ เช่น ค่ามัธยฐาน ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า เป็นต้น

7) การกำหนดจำนวนรอบ (Replicated R time)

8) การคำนวนค่าสถิติของผลการจำลองข้อมูลจำนวน R รอบทั้งสถิติเชิงบรรยาย เพื่อพิจารณาการแจกแจงและสถิติเชิงอ้างอิงสำหรับการทดสอบความแตกต่าง

3.3.2 ขั้นตอนในการวิเคราะห์ข้อสอบด้วยเทคนิคมอนติคาร์โล

การวิเคราะห์ข้อสอบด้วยเทคนิคมอนติคาร์โลมีขั้นตอนที่สำคัญสามารถสรุปได้ 4 ขั้นตอนสำหรับการนำไปสู่การปฏิบัติการวิเคราะห์ ขั้นตอนแรก ได้แก่ การกำหนดปัญหาใน การศึกษาวิจัย ขั้นตอนต่อมา เป็นขั้นตอนของการออกแบบการทดลอง ซึ่งรวมถึงการระบุตัวแปรอิสระและตัวแปรตาม การออกแบบการวิจัยเชิงทดลอง จำนวนรอบในการคำนวณ และการเลือกโมเดลการวิเคราะห์ข้อสอบ ขั้นตอนที่สามเป็นการเขียนและระบุโปรแกรมในการจำลองข้อมูลและการประมาณค่าพารามิเตอร์ และขั้นตอนสุดท้ายเป็นขั้นตอนเกี่ยวกับการวิเคราะห์ผลจากการจำลองข้อมูล โดยในแต่ละขั้นตอนมีรายละเอียดดังนี้ (Naylor และคณะ, 1968 อ้างถึงใน Harwell และคณะ, 1996)

1) **การกำหนดปัญหา** ขั้นตอนนี้นับเป็นขั้นตอนที่สำคัญของกระบวนการวิจัย การวิเคราะห์ข้อสอบด้วยเทคนิคมอนติคาร์โลก็เช่นเดียวกัน เริ่มแรกนักวิจัยจะต้องมีการกำหนดปัญหาและข้อคำถามการวิจัยก่อน หลังจากนั้นจึงตั้งสมมติฐานการทดสอบ และมีการวัดผลกระทบจากเงื่อนไขต่างๆ จากการจำลองข้อมูล โดยทั่วไปการกำหนดปัญหาการวิจัยมักจะมาจากการทบทวนเอกสารและรายงานการวิจัยที่เกี่ยวข้องกับเรื่องที่เราสนใจศึกษา

2) **การออกแบบการศึกษาด้วยด้วยเทคนิคมอนติคาร์โล** การออกแบบการศึกษานั้นจะต้องออกแบบให้สามารถตอบข้อคำถามของการวิจัยและสมมติฐานการวิจัยได้ โดยการออกแบบจะต้องมีการออกแบบทั้งตัวแปรต้นหรือตัวเป็นที่เป็นสาเหตุและตัวแปรตามซึ่งเป็นผลกระทบที่เกิดขึ้นจากตัวแปรต้น นอกจากนี้ยังต้องมีการประเมินผลทั้งในเรื่องของความตรงภายในและความตรงภายนอกซึ่งการออกแบบการศึกษาด้วยด้วยเทคนิคมอนติคาร์โลมีลักษณะคล้ายคลึงกับการศึกษาจากข้อมูลเชิงประจักษ์ ดังนี้

2.1) **การกำหนดและระบุค่าของตัวแปรต้น** โดยมีการกำหนดและระบุค่าของตัวแปรรวมทั้งเงื่อนไขในการจำลองข้อมูลให้สอดคล้องกับปัญหาและคำถามของการวิจัย โดยค่าของตัวแปรซึ่งคือว่าเป็นค่าคงที่ ซึ่งได้แก่ ขนาดกลุ่มตัวอย่าง (n) ความยาวของข้อสอบ (L) ความแปรปรวนของตัวแปรอิสระ (S^2) ซึ่งค่าของตัวแปรเหล่านี้ควรจะเกิดจากคำถามวิจัย นอกจากนี้ค่าของพารามิเตอร์ในโมเดลยังควรเป็นตัวแทนของตัวแปรอิสระ เนื่องจากในการศึกษาเทคนิคมอนติคาร์โล ต้องมีการพิจารณาว่าค่าพารามิเตอร์เหล่านี้เป็นค่าของตัวแปรคงที่ (fixed effect) หรือเป็นค่าของตัวแปรสุ่ม (random effect) เนื่องจากถ้ามีการสุ่มค่าอำนาจจำแนกและค่าความยากง่ายมาใช้ในการศึกษา จะทำให้โมเดลในการศึกษาเป็นโมเดลแบบสุ่ม ซึ่งสามารถการอ้างอิงไปยังประชากร แต่ถ้าโมเดลไม่ได้มีการสุ่มขึ้นมาใช้ในการศึกษา โมเดลนั้นจะไม่สามารถอ้างอิงกลับไปยังประชากรได้

2.2) การเลือกแบบการทดลอง โดยทั่วไปตัวแปรอิสระมักจะเป็นตัวกำหนดแบบการทดลองที่เหมาะสม เช่น ถ้าจำนวนตัวแปรอิสระและระดับค่าของตัวแปรน้อย การใช้แบบการทดลองแบบแฟกตอเรียลจะมีความเหมาะสมกว่าแบบการทดลองอื่น ในการศึกษาด้วยเทคนิคมอนติคาร์โล มักจะใช้แบบการทดลองโดยยึดเป้าหมายและวัตถุประสงค์ในการศึกษาเป็นหลัก ซึ่งการเลือกแบบทดลองอย่างระมัดระวังจะช่วยในการวางแผนการวิเคราะห์ผลลัพธ์ได้อย่างถูกต้อง (Lewis และ Ovar, 1989 อ้างถึงใน Harwell และคณะ, 1996) ในงานวิจัยของ Harwell และ Jamoskey (1991) มีตัวแปรจัดกระทำหรือตัวแปรต้นเป็น ขนาดกลุ่มตัวอย่าง ความยาวของแบบสอบ และความแปรปรวนของการแจกแจงค่าอำนาจจำแนก ใช้การออกแบบการทดลองแบบแฟกตอเรียลระหว่างกลุ่มตัวอย่างแบบสมบูรณ์ (completely between-subjects factorial design) นอกจากนี้ยังมีงานวิจัยอีกชิ้นหนึ่งซึ่งถือว่าเป็นตัวอย่างที่ดีเกี่ยวกับการออกแบบการทดลอง คือ งานวิจัยของ Yen (1987 อ้างถึงใน Harwel และคณะ, 1996) ซึ่งได้เปรียบเทียบการศึกษาจากการใช้โปรแกรมการวิเคราะห์ข้อสอบระหว่างโปแกรม BILOG และโปรแกรม LOGIST โดยมีเงื่อนไขด้านความยาวของข้อสอบและลักษณะการกระจายของค่าพารามิเตอร์ของผู้สอบ ออกแบบการทดลองแบบแฟกตอเรียลโดยเรื่องของความยาวของข้อสอบและลักษณะการกระจายของค่าพารามิเตอร์ของผู้สอบ ได้ออกแบบการทดลองแฟกตอเรียลแบบ between-subjects factor และส่วนในการเปรียบเทียบระหว่างโปรแกรมคอมพิวเตอร์ทั้งสองได้ออกแบบการทดลองแฟกตอเรียลแบบ within-subjects factor

2.3) การเลือกตัวแปรตาม การเลือกตัวแปรตามนั้นไม่เพียงแต่จะต้องสอดคล้องกับคำถามในการวิจัยแต่จะต้องเลือกตัวแปรตามที่มีความไวต่อตัวแปรจัดกระทำหรือตัวแปรต้น เนื่องจากการเลือกตัวแปรที่มีความไวและควรใช้ประโยชน์ได้ถ้ามีการแปลงข้อมูลเป็นรูปแบบอื่น เช่น การหาค่า RMSD สามารถแปลงค่าเพื่อให้มีการแจกแจงแบบปกติ ทำให้สามารถนำไปสรุปอ้างอิงได้ทั่วไป นอกจากนี้ถ้าเป็นเรื่องเกี่ยวกับการศึกษาเปรียบเทียบวิธีการในการศึกษา IRT นักวิจัยก็สามารถใช้คุณลักษณะของแบบสอบ เช่น ความเป็นเอกมิติ การทำหน้าที่ต่างกันของข้อสอบหรือผู้สอบ เป็นตัวแปรตามในการศึกษาถึงผลกระทบของตัวแปรอิสระได้อีกด้วย สำหรับค่าความสัมพันธ์ของค่าจริงกับค่าประมาณ ก็สามารถใช้ให้เป็นตัวแปรตามในการใช้ เทคนิคมอนติคาร์โล เนื่องจากค่าสัมพันธ์นั้น สามารถใช้เมตริกที่ต่างกันหาความสัมพันธ์ของตัวแปรอิสระและตัวแปรตามได้ เช่น ค่าความสัมพันธ์ระหว่างค่าพารามิเตอร์ที่ประมาณกับความคลาดเคลื่อนมาตรฐาน ส่วนข้อเสียก็คือความสัมพันธ์เหล่านี้สะท้อนความสัมพันธ์เฉพาะอันดับของตัวแปรและแสดงอิทธิพลของตัวแปรอิสระเท่านั้น เช่น ค่าความสัมพันธ์ระหว่างค่าอำนาจจำแนกที่แท้จริงกับค่าที่ประมาณ มีค่าเท่ากับ 0.9 ซึ่งหมายความว่า โดยค่าเฉลี่ยของค่าอำนาจจำแนกที่แท้จริงนั้น อาจจะสูงกว่าค่าเฉลี่ยของ

ค่าอำนาจจำแนกที่ประมาณได้ แต่ไม่รับรองว่าค่าอำนาจจำแนกที่แท้จริงนั้นกับค่าอำนาจจำแนกที่ประมาณจะใกล้เคียงกันหรือดีกว่ามากน้อยเพียงใด เช่น 0.8 กับ 0.9

2.4) **การกำหนดจำนวนรอบ** การกำหนดจำนวนรอบในการศึกษาด้วยเทคนิคมอนติคาร์โลเปรียบเทียบได้กับการกำหนดขนาดกลุ่มตัวอย่าง โดยมีเกณฑ์ที่ใช้ในการกำหนดประยุกต์ใช้มาจากการกำหนดขนาดกลุ่มตัวอย่างสำหรับในการศึกษาจากข้อมูลเชิงประจักษ์ ในการศึกษาการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ จำนวนรอบจะขึ้นอยู่กับวัตถุประสงค์ในการศึกษาด้วยเทคนิคมอนติคาร์โล โดยพิจารณาได้จากความต้องการในลดค่าความแปรปรวนของการสุ่มตัวอย่างในการประมาณค่าพารามิเตอร์ และความต้องของการทดสอบสถิติของผลการจำลองข้อมูลว่าอำนาจในการตรวจสอบผลกระทบที่สนใจเพียงพอหรือไม่

จำนวนรอบมีอิทธิพลโดยตรงกับความแม่นยำในการประมาณค่าพารามิเตอร์ ถ้ากลุ่มตัวอย่างขนาดใหญ่ (มีจำนวนรอบมาก) จะให้การประมาณค่าพารามิเตอร์ด้วยความแปรปรวนของการสุ่มตัวอย่างน้อย ดังนั้นถ้านักวิจัยไม่กำหนดจำนวนรอบหรือกำหนดจำนวนรอบน้อยจะทำให้ความแปรปรวนของการสุ่มมีมากเพียงพอที่จะให้ทำให้การประมาณค่าพารามิเตอร์มีความลำเอียงมาก ซึ่งจะส่งผลความเที่ยงและความน่าเชื่อถือของผลการวิจัยที่ได้ต่ำมาก

เทคนิคในการลดความแปรปรวนในการประมาณค่า คือ การเพิ่มจำนวนรอบซึ่งจะทำให้ได้ค่าที่คงที่และน่าเชื่อถือได้มากกว่า เนื่องจากสามารถเปรียบเทียบค่าพารามิเตอร์ระหว่างรอบได้ ข้อดีอีกประการหนึ่ง คือ จำนวนรอบสะท้อนความเบี่ยงเบนของค่าประมาณ ซึ่งถ้าเป็นความเบี่ยงเบนระหว่างเงื่อนไขมีค่าน้อย แสดงว่า ตัวแปรอิสระส่งผลต่อตัวแปรตามน้อย โดยมีสมการในการคำนวณความเบี่ยงเบนของค่าพารามิเตอร์ที่แท้จริงและค่าประมาณได้จาก ค่า RMSD (Root Mean Square Deviation) ซึ่งเป็นค่าการถอดรากของค่าเฉลี่ยกำลังสองของ ความแตกต่างระหว่างค่าที่ประมาณได้กับค่าพารามิเตอร์ที่แท้จริงสามารถเขียนเป็นสูตรทางคณิตศาสตร์ได้ดังนี้

$$RMSD = \left[\frac{\sum_{i=1}^{n} (\hat{a}_i - a_i)^2}{n} \right]^{1/2}$$

เมื่อ

$\hat{a}_i$ = ค่าพารามิเตอร์ที่ได้จากการประมาณค่า

a_i = ค่าพารามิเตอร์แท้จริง

n = จำนวนพารามิเตอร์

นอกจากนี้เมื่อพิจารณาความแปรปรวนของการประมาณค่าจากจำนวนรอบ (the variance of the estimates across replications) และฟังก์ชันความแปรปรวนความคลาดเคลื่อนของข้อมูลเชิงประจักษ์ (an empirical error variance) โดยสามารถสรุปเป็นสมการได้ดังนี้

$$\frac{\sum_{r=1}^{R}(\hat{a}_{ir}-a_i)^2}{R}=\left(\bar{\hat{a}}_{ir}-a_i\right)+\frac{\sum_{r=1}^{R}\left(\hat{a}_{ir}-\bar{\hat{a}}_i\right)^2}{R}$$

จากสมการข้างบน ถ้าค่าดังกล่าวมีค่าน้อย แสดงให้ว่า การประมาณค่ามีความคงที่นั้นก็ค่าการประมาณค่ามีความเที่ยง ในทางตรงกันข้ามถ้าค่าดังกล่าวมีค่ามาก นักวิจัยต้องระมัดระวังในการนำค่านั้นไปใช้ในการวิเคราะห์เนื่องจากค่าดังกล่าวไม่คงที่หรือการประมาณค่านั้นไม่มีความเที่ยง

เมื่อพิจารณาในเรื่องของจำนวนรอบกับอำนาจการทดสอบ (power) พบว่า จำนวนรอบนั้นมีความสำคัญกับอำนาจการทดสอบ โดยนักวิจัยต้องเลือกจำนวนรอบให้มีขนาดมากพอกับอำนาจที่ต้องการสำหรับการใช้สถิติตัวนั้น ๆ ในการทดสอบจากการศึกษาของ Stone (1993 อ้างถึงใน Harwell และคณะ, 1996) โดยได้ศึกษาความสัมพันธ์ระหว่างจำนวนรอบกับอำนาจการทดสอบเป็น 2 ขั้นตอน โดยขั้นตอนแรกใช้เทคนิคมอนติคาร์โล ในการศึกษาผลของจำนวนผู้สอบ N (N=250, 500, 1000) , จำนวนข้อสอบ (L=10, 20, 30) และการแจกแจงของความสามารถ (θ) (D= Normal, Skewed, Platykurtic) ทดลองโดยใช้แบบแฟคตอเรียล โดยมีค่า RMSD เป็นตัวแปรตาม และใช้โมเดล 2PL ในการระบุค่าอำนาจจำแนกและค่าความยากในแต่ละข้อ กำหนดให้ R=10 (10 ชุด) ในแต่ละเงื่อนไข แสดง RMSD ด้วยกราฟ และทดสอบด้วย ANOVA เพื่อดูอิทธิพลของตัวแปรอิสระ ค่าอิทธิพลสูงสุดแสดงด้วยค่า η^2 (correlation ratio=η^2) จะเหมาะสมเนื่องจากสอดคล้องกับข้อตกลงเบื้องต้นด้านการแจกแจงเป็นโค้งปกติ

ในขั้นตอนที่ 2 η_s^2 ได้จากการศึกษาเทคนิคมอนติคาร์โล จะใช้ประมาณค่าอำนาจการทดสอบของ ANOVA เพื่อตรวจสอบอิทธิพลอันเกิดจากจำนวนรอบที่แตกต่างกัน (R=10, 25, 50, 100) โดยใช้ η_s^2 เป็นตัวประมาณขนาดอิทธิพล โปรแกรม STAT-POWER จะใช้ประมาณค่าอำนาจการทดสอบที่ระดับ .05 และองศาอิสระเท่ากับ 27 และ η^2 กระบวนการนี้จะทำให้เกิดการเปลี่ยนแปลงอำนาจการทดสอบของ ANOVA F-Test ความสัมพันธ์ระหว่างจำนวนรอบกับอำนาจการทดสอบ โดยอำนาจการทดสอบจะเปลี่ยนแปลง เมื่อจำนวนรอบเปลี่ยนแปลงไป กล่าวคือ จำนวนรอบยิ่งมากอำนาจการทดสอบก็จะมีค่าสูงขึ้น

การเพิ่มจำนวนรอบจนถึง 500 อาจไม่จำเป็น เมื่อจำนวนผู้สอบ (N) และจำนวนข้อสอบ (L) มีขนาดใหญ่พอ อาจน้อยกว่า 100 รอบก็ได้ แต่จากการตรวจสอบจำนวนรอบ (R) อาจจะจำเป็นเพื่อความน่าเชื่อถือของการศึกษาในประเด็นดังต่อไปนี้ เช่น การตรวจสอบการแจกแจงการสุ่ม การวิเคราะห์ข้อสอบที่มีความแปรปรวนขนาดใหญ่ การศึกษาอิทธิพลของเงื่อนไขที่มีความสลับซับซ้อน และการศึกษาค่าพารามิเตอร์ที่เป็นค่าสุดโต่ง

ดังนั้นจะเห็นได้ว่าจำนวนรอบมีความจำเป็นต่อความเที่ยงในการตรวจสอบผลกระทบของผลการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลมากในลักษณะงานวิจัยใน 4 ประเภท ได้แก่ ประเภทแรก งานวิจัยที่ให้ความสำคัญกับลักษณะการกระจายของการสุ่มตัวอย่างจากข้อมูลเชิงประจักษ์ เช่น งานวิจัยที่มีวัตถุประสงค์เพื่อตรวจสอบคุณสมบัติของค่าสถิติหรือการทดสอบนัยสำคัญทางสถิติ ประการที่สอง งานวิจัยที่สนใจศึกษาค่ากลางของการวิจัยระดับข้อสอบที่มีค่าความแปรปรวนของกลุ่มตัวอย่างมาก ประการที่สามงานวิจัยที่มีเป้าหมายในการศึกษาผลกระทบที่เพิ่มขึ้นของบริบทที่ซับซ้อน เช่น ผลกระทบที่เกิดขึ้นจากปฏิสัมพันธ์กับอิทธิพลหลัก ประการสุดท้าย โมเดลที่มีค่าพารามิเตอร์ที่มีค่าสุดโต้งมาก ทั้งนี้จำนวนรอบขั้นต่ำสำหรับการศึกษา การวิเคราะห์ด้วยทฤษฎีการตอบสนองควรในการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลควรมีค่าไม่ต่ำกว่า 25 รอบ ถือเป็นเงื่อนไขที่อาจนำไปใช้ศึกษาในทฤษฎีการตอบสนองข้อสอบด้วยเทคนิคมอนติคาร์โลได้

3) **การเขียนและการระบุโปรแกรมในการจำลองข้อมูล** การเลือกโปรแกรมคอมพิวเตอร์นักวิจัยอาจใช้หลายโปรแกรมในการจำลองข้อมูล รวมทั้งการวิเคราะห์ผลลัพธ์ ในแต่ละขั้นตอนจะต้องมีการประเมินความถูกต้องแม่นยำ ใน 3 ประเด็นดังนี้ ประการแรก ค่าของข้อมูลจากการจำลองมีคุณภาพเพียงใดเมื่อเทียบกับข้อมูลที่มีการศึกษาในอดีต กล่าวคือ ควรมีการตรวจสอบความตรงของค่าที่ได้จากโมเดลหรือกระบวนการจำลองข้อมูล ประการที่สอง โมเดลการจำลองข้อมูลสามารถทำนายปรากฏการณ์ในอนาคตได้ถูกต้องแม่นยำเพียงใด ประเด็นสุดท้าย คือ ผลจาก

การจำลองข้อมูลนั้นเมื่อนำไปศึกษากับข้อมูลเชิงประจักษ์จะสามารถทำนายได้เมื่อเวลาผ่านไปหรือไม่ ซึ่งมีรายละเอียดในการเขียนและระบุโปรแกรมในการจำลองดังนี้

3.1) **การจำลองคำตอบ** การจำลองคำตอบเริ่มด้วยการกำหนดค่าเริ่มต้น (Seed) ให้กับตัวเลขสุ่ม ซึ่งจะแปลงเป็นค่าความน่าจะเป็นผู้ตอบในการตอบคำถามได้ถูกต้องและต่อจากนั้น จึงแปลเป็นคำตอบของผู้สอบในการตอบข้อสอบแบบไม่ต่อเนื่อง (0, 1) โดยทั่วไปในขั้นตอนนี้จะเป็นหน้าที่ของคอมพิวเตอร์เป็นผู้ดำเนินการ

3.2) **ตัวเลือกค่าเริ่มต้น** ซึ่งผู้จำลองข้อมูลจะเป็นผู้กำหนดค่าเริ่มต้นเอง โดยใช้เติมในช่องว่าง (Prompted) ประโยชน์ก็คือ ง่ายต่อการจำลองคำตอบในข้อต่อ ๆ มา ประเด็นที่สำคัญ คือ เป็นค่าที่สัมพันธ์กับความคลาดเคลื่อนในการสุ่ม ซึ่งหลีกเลี่ยงได้ยากในการจำลองข้อมูล เทคนิคอย่างหนึ่งที่จะลดความแปรผัน คือ ใช้พารามิเตอร์ข้อสอบและค่าเริ่มต้นร่วมกันทุกครั้ง เมื่อมีการจำลองข้อมูล เช่น การจำลองข้อสอบ 20 ข้อและ 30 ข้อ ค่าเริ่มต้นที่ใช้ในการจำลองข้อมูล 20 ข้อ ควรจะเป็นค่าเริ่มต้นเดียวกันกับเมื่อจำลองข้อมูล 30 ข้อ เทคนิคอีกประการหนึ่ง คือ การจำลองประชากรข้อมูลคำตอบจำนวนมาก แล้วสุ่มคำตอบมาจากประชากรที่จำลองขึ้นนั้นหลาย ๆ รอบมากกว่าการใช้ค่าเริ่มต้นหลายตัว เพื่อจะจำลองชุดข้อมูลให้ได้ตามต้องการ การใช้โมเดลพารามิเตอร์ที่ต่างกันในการจำลองข้อสอบ 20 ข้อ 30 ข้อ รวมทั้งค่าเริ่มต้นที่ต่างกันจะทำให้คำตอบมีความเป็นอิสระแก่กันมากขึ้น แต่อาจจะเกิดความคลาดเคลื่อนมากกว่าเมื่อเทียบกับการใช้โมเดลพารามิเตอร์และค่าเริ่มต้นแบบเดียวกัน วิธีที่ควรใช้คือ การจำลองข้อมูลทุกชุดควรใช้โมเดลพารามิเตอร์ต่างกันแต่ใช้ค่าเริ่มต้นร่วมกัน

3.3) **การจำลองตัวเลขสุ่ม** จะใช้ตัวเลขสุ่มที่มีการแจกแจงแบบยูนิฟอร์ม (Uniform distribution) เป็นส่วนใหญ่และใช้วิธีการจำลองข้อมูลแบบ Congruential generators (Cewis และ Ovaw, อ้างถึงใน Harwell, M.และคณะ, 1996) วิธี Congruential generators จะใช้โมเดลพีชคณิตใน การจำลองตัวเลขที่สุ่มขึ้นมากับตัวเลขสุ่มที่ผ่านมา ตัวเลขสุ่มจะเริ่มจำนวนจาก 0....m ซึ่งจะสุ่มโดยโปรแกรม โดยจะวิ่งเป็นวงจรที่เรียกว่า “ความยาวรอบ” (period) เมื่อครบรอบก็จะวนกลับมาใช้เลขเดิมอีก จากการศึกษาในอดีต พบว่า การใช้วิธี Congruential generators มีช่วงความยาวรอบเพียงพอกว่าจะเกิดข้อมูลซ้ำ การแจกแจงปกติมาตรฐานจะใช้มากในการใช้เทคนิคมอนติคาร์โล โดยการแปลงข้อมูลจาก Uniform (0, 1) เป็น Normal ~ N (0, 1) วิธีที่ใช้แปลงคือ วิธีของ Box – Muller และ Marsaglia

หลักในการเลือกโปรแกรมเพื่อใช้ในการจำลองข้อมูล ประกอบด้วย 4 อย่าง คือ วิธีการควรง่ายต่อการทำความเข้าใจและเขียนโปรแกรม โปรแกรมควรมีความกะทัดรัด รหัสปลายทางควรมีความสมเหตุสมผล (คำตอบ 0, 1) และขั้นตอนการคำนวณ ควรใช้ตัวเลขสุ่มเทียม (Pseudo-random numbers) (Ripley, 1987)

ความถูกต้องของตัวเลขสุ่มขึ้นอยู่กับตัวจำลองเลขสุ่มและเครื่องคอมพิวเตอร์นักวิจัยควรตรวจสอบตัวจำลองและเอกสารด้านสมรรถนะ แม้บางตัวอาจเป็นที่เชื่อถือได้แต่ก็ควรตรวจสอบว่าสามารถนำไปใช้ตรงกับปัญหาเฉพาะหรือไม่ อาจเป็นการทดสอบและตรวจเอกสารซึ่งจะเกี่ยวข้องกับความเชื่อถือได้ของโปรแกรม การตรวจสอบความพอเพียงของตัวจำลองข้อมูลด้วยกระบวนการทดสอบความเหมาะสมด้วยสถิติ Kolmogorov Smirnov and Coraner – van – Mises goodness - of – fit การทดสอบลำดับขึ้นลง การทดสอบช่วงว่างและค่าสูงสุดและการทดสอบ การสุ่ม

3.4) **การแปลงตัวเลขสุ่มเป็นคำตอบ** ในเบื้องต้น เวกเตอร์ของ N (จำนวนผู้สอบ) , θ (ความสามารถ) ถูกสุ่มจากการแจกแจงที่กำหนดจากกลุ่มผู้สอบ ซึ่งค่าความสามารถมักจะได้จากการสุ่มกลุ่มตัวอย่างที่มีการแจกแจงแบบปกติ และเป็นโมเดล IRT แบบเอกมิติหรือจากการแจกแจงปกติหลายตัวแปรที่มีค่าความสัมพันธ์ระหว่างตัวแปรในกรณีนี้ศึกษาการวัดแบบหลายมิติ เช่น การสุ่มความน่าจะเป็นของคำตอบที่ตอบแบบ (0, 1) จะได้จากสมการ ดังนี้

$$P_i(\theta) = c_i + \frac{1 - c_i}{1 + e^{-Da_i(\theta - b_i)}}$$

$$Q_i = 1 - P_i$$

ค่าความน่าจะเป็นของคำตอบ (P_i) จะถูกแปลงไปเป็นคำตอบ 0, 1 โดยการเปรียบเทียบค่าความน่าจะเป็นจากการสุ่ม การแจกแจงแบบ Uniform ถ้าความน่าจะเป็นในการตอบถูกของผู้สอบสูงกว่าหรือเท่ากับตัวเลขสุ่ม คำตอบในข้อนั้นจะเป็น 1 แต่ถ้าต่ำกว่าตัวเลขสุ่มจะเป็น 0 ในกรณีเป็นคำตอบหลายค่า ถ้าตัวเลขสุ่มตกอยู่ในช่วงใดก็จะเป็นคำตอบ K และ K+1 ในกระบวนการทั้งหมดจะกระทำซ้ำด้วยตัวเลขสุ่มที่แตกต่างกันในแต่ละข้อและผู้สอบทั้งหมด

3.5) **การประมาณค่าพารามิเตอร์ในโมเดล** ผู้วิจัยอาจใช้โปรแกรมสำเร็จรูป เช่น BILOG หรือ MULTILOG หรือการสร้างโปรแกรมด้วยตนเองก็ได้ ประเด็นสำคัญ 2 ประเด็นในการเขียนและการเลือกโปรแกรมวิเคราะห์ คือ การกำหนดค่าเริ่มต้น (Starting Value) และ การแก้ปัญหาเมตริกไม่คอนเวิร์จ (non-convergent solutions) ในกระบวนการทั้งหมดของการประมาณค่าจะเกี่ยวกับขั้นตอนการคำนวณทวนซ้ำและการกำหนดค่าเริ่มต้น โดยปกติคอมพิวเตอร์จะกำหนดให้แต่นักวิจัยส่วนใหญ่จะเป็นผู้กำหนดเอง

หากนักวิจัยพบปัญหาการไม่คอนเวิร์จ ในการศึกษาแบบมอนติคาร์โลนักวิจัยสามารถดำเนินการแยกได้เป็น 3 กรณี ได้แก่ ประการแรก นักวิจัยไม่ต้องให้สนใจความไม่คอนเวิร์จนั้น แต่ใช้ค่าประมาณจากจำนวนครั้งของการคำนวณทวนซ้ำที่มากที่สุด ประการที่สองนักวิจัยควรแยกการประมาณค่าของค่าสถิติสรุปรวม ได้แก่ RMSDs และประการที่สามนักวิจัยควรใช้วิธีการคำนวณอื่น ๆ เช่น วิธีของเบย์ (Bayesian) เพื่อบังคับค่าพารามิเตอร์ โดยควรคำนึงถึงค่าพารามิเตอร์ที่ประมาณค่า กับค่าพารามิเตอร์ที่แท้จริงให้มีค่าใกล้เคียงกันด้วย มิฉะนั้นแล้วค่าที่ได้จะเป็นค่าที่ลำเอียง

ปัญหาสำคัญของการกำหนดค่าเริ่มต้นและการแก้ปัญหาการไม่คอนเวิร์จ มักพบในข้อมูลจำนวนน้อย (N=200, 40 ข้อ) และในโมเดล IRT ที่ซับซ้อนมากขึ้น การแก้ปัญหาคือ การพัฒนาโปรแกรมเพื่อใช้ประมาณค่า วิธีการของเบย์ (Bayesian) สามารถช่วยได้ โดยการพิจารณาอย่างพิถีพิถันกับการแจกแจงของข้อมูลและการแจกแจงของค่าพารามิเตอร์

4) **การวิเคราะห์ผล** การวิเคราะห์ผลจะตั้งอยู่บนพื้นฐานของคำถามวิจัย การออกแบบการทดสอบสมมติฐานทางสถิติ กระบวนการวิเคราะห์ โดยปกติการวิเคราะห์ผลจากการใช้เทคนิคมอนติคาร์โล ประกอบด้วย การใช้ตารางสรุปผลรวม สถิติเชิงบรรยายเบื้องต้นหรือการนำเสนอด้วยกราฟแผนภูมิ การดูผลกระทบจากตัวแปรอิสระ อาจต้องใช้สถิติเชิงอ้างอิงปัญหาของการวิเคราะห์คือการมีค่าต่าง ๆ กันจำนวนมาก เมื่อต้องการรายงานผล Harwell (1991) เสนอว่า ควรจะมีการใช้ทั้งเชิงบรรยายและเชิงอ้างอิง เพื่อเป็นการเพิ่มโอกาสในการตรวจสอบข้อมูล ซึ่งจะทำให้มีเชื่อมั่นมากขึ้น

การวิเคราะห์ผลด้วยการใช้สถิติเชิงสรุปอ้างอิงสามารถใช้ได้หลายวิธี แต่มักเป็นการวิเคราะห์ถดถอยพหุคูณและการวิเคราะห์ความแปรปรวน ถ้าตัวแปรต้นเป็นตัวแปรระดับนามบัญญัติ การวิเคราะห์ความแปรปรวนจะดีกว่า แต่ถ้าเป็นตัวแปรระดับช่วงชั้น การใช้การวิเคราะห์ถดถอยพหุคูณจะดีกว่า แต่ส่วนใหญ่แล้วในทฤษฎีการตอบสนองข้อสอบจะใช้ การวิเคราะห์ถดถอยพหุคูณ เนื่องจากตัวแปรส่วนใหญ่จะเป็นระดับช่วงชั้น หรืออัตราส่วน ในโมเดลการวิเคราะห์ถดถอยพหุคูณประชากรสามารถเขียนแทนด้วย

$$\tau = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_T X_T$$

และ

$$\hat{\tau}_s = \tau_s + \varepsilon_s$$

τ มักจะเป็นผลลัพธ์ เช่น ค่าของ RMSD ซึ่งจะได้รับผลจากตัวแปรทำนาย Xs (s=0,..., T) , β_0 เป็นค่าเฉลี่ย, β_T เป็นสัมประสิทธิ์การถดถอยพหุคูณของประชากร ε_s เป็นความคลาดเคลื่อน และ $\hat{\tau}_s$ เป็นค่าผลลัพธ์สังเกตได้ โมเดลประมาณสำหรับกลุ่มตัวอย่าง คือ

$$\tau_s = b_0 + b_1 x_1 + b_2 x_2 + ... + b_T x_T$$

ข้อตกลงของสมการถดถอย กล่าวว่า ความคลาดเคลื่อนจะต้องเป็นอิสระจากกัน และในการสร้างข้อมูลก็จะต้องมีการตรวจสอบความเป็นอิสระดังกล่าว ความเป็นอิสระของ τ_s อาจยอมรับได้จากความถูกต้องของตัวจำลองข้อมูล และการตรวจสอบการแจกแจงแบบโค้งปกติ หรืออาจใช้วิธีการตรวจสอบด้วยความแข็งแกร่งของการแจกแจงปกติ จากการทบทวนเอกสาร เช่น การแปลงค่า RMSD เป็น Log (RMSD) ถ้านำมาใช้ในการคำนวณต้องมีการแจกแจงเป็นโค้งปกติด้วย แต่ถ้าไม่สามารถแปลงค่าได้ให้มีการแจกแจงเป็นโค้งปกติ อาจต้องนำสถิติ non-parametric มาใช้แทน

Harwell และ Janosky (1991) ได้ศึกษาการวิเคราะห์การถดถอยพหุคูณและการวิเคราะห์ ANOVA เพื่อวิเคราะห์ผลลัพธ์จากการใช้เทคนิค MC โดยการจำลองข้อมูลแบบ 0, 1 ด้วยการวิเคราะห์โมเดล 2 PL ประมาณค่าพารามิเตอร์โดยใช้โปรแกรม BILOG เปรียบเทียบค่าประมาณค่า a และ b ในแต่ละข้อกับค่าแท้จริงโดยใช้ RMSD ผลจากการศึกษาสรุปได้ว่า ขนาดกลุ่มตัวอย่าง (N) ของข้อสอบจำนวน 15 และ 25 ข้อ จะมีผลต่อความถูกต้องในการประมาณค่าพารามิเตอร์ และพบว่าการใช้ log (RMSD) สามารถทำให้ตัวแปรตามสอดคล้องกับข้อตกลงเบื้องต้นทางสถิติด้านการแจกแจงแบบปกติ โดยถ้าตัวแปรต้นเป็นตัวแปรเมทริกซ์ จะมีความเหมาะสมที่จะวิเคราะห์ด้วยการถดถอยพหุคูณ นอกจากนี้ยังทดสอบแฟกตอเรียล ANOVA แบบสุ่มสมบูรณ์ซึ่งให้ผลคล้ายกับการวิเคราะห์ด้วย การถดถอยพหุคูณ

3.3.4 ข้อพึงระวังในการประยุกต์ใช้เทคนิคมอนติคาร์โลสำหรับการวิเคราะห์ข้อสอบ

ปัจจุบันมีการประยุกต์ใช้ทคนิคมอนติคาร์โลสำหรับการวิเคราะห์ข้อสอบกันอย่างแพร่หลาย โดยตั้งแต่ในปี 1981 ถึง ปี 1991 มีงานวิจัยที่ลงตีพิมพ์ในวารสาร Psychometrika และ JEM (Journal of measurement) ซึ่งเป็นวารสารที่มีชื่อเสียงสำหรับการวิจัยทางด้านวัดและประเมินผล ทั้งสินจำนวน 26 เรื่อง โดยในบางเรื่อง มีการประยุกต์ใช้เทคนิค MC ในการศึกษาเป็นบางส่วน ในขณะที่บางเรื่องจะใช้เทคนิค MC ทั้งหมด นอกจากนี้งานวิจัยนั้นมักขาดข้อมูลในการสนับสนุนอย่างเพียงพอ อาทิเช่น มีงานวิจัยจำนวน 16 เรื่อง (คิดเป็นร้อยละ 61.5) ไม่มีการกำหนดจำนวนรอบในการศึกษา ซึ่งจะทำให้ผลการวิเคราะห์เกิดความคลาดเคลื่อน และมีงานวิจัยจำนวน 20 เรื่อง (คิดเป็นร้อยละ 76.9) ขาดหลักฐานในการแจงแจงความสามารถและค่าพารามิเตอร์ของข้อสอบที่ตรงกับสภาพความเป็นจริง ซึ่งส่งผลให้งานวิจัยขาดความตรงภายนอก ดังนั้น Harwell, และ

คณะ (1996) จึงให้ข้อควรระวัง สำหรับงานวิจัยที่จะใช้การวิเคราะห์ข้อสอบด้วยการใช้เทคนิคมอนติคาร์โลที่สำคัญก่อนตัดสินใจใช้เทคนิคมอนติคาร์โลสำหรับการวิเคราะห์ข้อสอบ สามารถสรุปได้ดังนี้ (Harwell, และคณะ, 1996)

1. ปัญหาในการวิจัยสามารถแก้ได้ด้วยการวิเคราะห์ด้วยเทคนิคมอนติคาร์โลจริงหรือไม่
2. การศึกษาในประเด็นนั้นสามารถขยายองค์ความรู้เดิมได้หรือไม่
3. ในการออกแบบการทดลองและการวิเคราะห์เหมาะสมที่จะใช้เทคนิคมอนติคาร์โลมาศึกษาหรือไม่
4. การวิเคราะห์ข้อมูลควรจะนำโปรแกรมที่มีอยู่เดิมมาใช้หรือปรับปรุงโปรแกรมใหม่ เพื่อให้เหมาะสมกับสภาพปัญหาการวิจัยของเราหรือไม่
5. ผลลัพธ์ในการวิเคราะห์ข้อมูลในงานวิจัยขึ้นอยู่กับค่าเริ่มต้นสำหรับวิธีการประมาณค่าพารามิเตอร์หรือไม่
6. ข้อตกลงเบื้องต้นของการแจกแจงลักษณะตัวแปรอิสระและค่าต่างๆ ที่กำหนดในการจำลองข้อมูลตรงกับสภาพความเป็นจริงหรือไม่

3.4 ประสิทธิภาพของการจำลองข้อมูล

การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลเป็นการจำลองข้อมูลโดยการใช้ตัวเลขสุ่ม (Random Numbers) โดยการพิจารณาว่าวิธีการจำลองข้อมูลนั้นดีหรือไม่ สามารถพิจารณาได้ 3 ประการ ได้แก่ ประการแรก วิธีการสุ่มนั้นจะต้องมีลักษณะของตัวเลขสุ่มที่ถูกสร้างขึ้นมามีการแจกแจงอย่างสม่ำเสมอ (Uniformity) ในช่วง 0 ถึง 1 ประการที่สอง ตัวเลขสุ่มในแต่ละตัวต้องเป็นอิสระซึ่งกันและกัน (Independence) และประการสุดท้าย ตัวเลขสุ่มที่ได้มานั้นต้องมีช่วงยาวก่อนจะเกิดการสุ่มซ้ำใหม่ (มีวัฏจักรยาว) ดังนั้น วิธีการจำลองข้อมูลที่ดี มีประสิทธิภาพ จึงสามารถพิจารณาได้จากตัวเลขสุ่มที่ได้มานั้นต้องมีลักษณะ 3 ประการดังกล่าวข้างต้น จึงจะถือว่าข้อมูลที่ได้ ได้มาจากวิธีการจำลองข้อมูลที่ดี (ศิริรัตน์ วงศ์ประกรณ์กุล, 2539)

3.5 ข้อดีและข้อจำกัดของการศึกษาด้วยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล

การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลเป็นเทคนิคที่ใช้สำหรับการแก้ปัญหาในการคำนวณทางคณิตศาสตร์สำหรับสถานการณ์ที่มีกระบวนการที่ซับซ้อนหรือไม่สามารถแก้ไขได้ในเชิงเหตุผล ซึ่งการวิเคราะห์ทั่วไปไม่สามารถดำเนินการได้อย่างครบถ้วนและสมบูรณ์ โดยการจำลองข้อมูลที่แสดงเหตุการณ์เกิดขึ้นทั้งทางบวกและทางลบ โดยการสร้างเลขสุ่ม (random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา ดังนั้นสถานการณ์ส่วนใหญ่ของงานวิจัยที่เหมาะสมกับการใช้ข้อมูลจำลอง จึงเป็น สถานการณ์ที่ต้องการตรวจสอบการแจกแจงทางสถิติ การเปรียบเทียบตัวประมาณค่าที่อยู่ในสถานการณ์ที่เก็บข้อมูลได้ยาก การศึกษาความแกร่งของสถิติ การเปรียบเทียบขั้นตอนการคำนวณของฟังก์ชัน หรือการประเมินขั้นตอนในการคำนวณอย่างใดอย่างหนึ่ง (Harwell, และคณะ, 1996) สามารถสรุปข้อดีของการศึกษาด้วยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ได้ดังนี้

1) ในงานวิจัยบางเรื่องที่ไม่สามารถเก็บข้อมูลได้โดยตรง การวิจัยประเภทนี้สามารถให้กระบวนการทางคณิตศาตร์ในการจำลองข้อมูลขึ้นมาสำหรับการวิเคราะห์แทนการใช้ข้อมูลจริงที่ไม่สามารถเก็บรวบรวมข้อมูลในการทดลอง

2) การจำลองข้อมูลสามารถสร้างตัวแบบที่มีความซับซ้อนมากได้ ซึ่งบางครั้งในการเก็บรวมข้อมูลจริงอาจเก็บได้ยากและมีจำนวนข้อมูลน้อย ซึ่งไม่สามารถนำมาวิเคราะห์สถิติในบางประเภทได้

3) การจำลองข้อมูลสามารถกำหนดและจัดกระทำค่าพารามิเตอร์ ซึ่งนำไปใช้ในการศึกษาปัจจัยที่ส่งผลต่อองค์ประกอบต่างๆ ได้ รวมทั้งค่าใช้จ่ายจากการจำลองข้อมูลยังน้อยกว่าการเก็บรวบรวมข้อมูลจริงอีกด้วย

4) การศึกษาวิจัยด้วยการจำลองข้อมูลสามารถกำหนดระยะเวลาที่แน่นอนในการดำเนินการทดลองได้ ซึ่งบางงานวิจัยต้องการใช้ผลงานวิจัยที่เร่งด่วน

5) การจำลองข้อมูลสามารถทำการสุ่มและสร้างตัวแปรเทียมสำหรับใช้ในตัวแบบได้โดยตรง ซึ่งในการเก็บรวบรวมด้วยข้อมูลจริงบางครั้ง บางสถานการณ์ไม่สามารถกระทำได้โดยตรง

สำหรับเรื่องข้อจำกัดของการศึกษาด้วยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล สามารถสรุปสาระสำคัญได้ดังนี้

1) ในการศึกษาบางครั้งผลที่ได้จากการศึกษาด้วยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ไม่สามารถครอบคลุมได้ในทุกกรณี เหมือนกับสภาพการณ์ที่เกิดขึ้นจริงได้

2) สำหรับในงานวิจัยหรือการศึกษาสภาพปัญหาในบางประเภทมีกระบวนการในการสร้างตัวแบบที่สามารถกระทำได้ยาก

3) การนำผลการวิจัยไปใช้จากการจำลองข้อมูลต้องพิจารณาเงื่อนไขที่ศึกษาว่า เงื่อนไขเหล่านั้นสอดคล้องหรือตรงกับสภาพความเป็นจริงมากน้อยเพียงใด

4) ผลการจำลองข้อมูลนั้นขึ้นอยู่กับจำนวนรอบและความถูกต้องของตัวเลขที่จำลองได้ ดังนั้นคุณธรรมของนักวิจัยที่จำลองข้อมูลจึงเป็นเรื่องที่สำคัญ ซึ่งนับว่าเป็นเรื่องยากต่อการประเมิน

5) การเลือกใช้วิธีการที่เหมาะสมสำหรับการจำลองข้อมูล โดยถ้านักวิจัยเลือกวิธีการที่ไม่เหมาะสม อาจจะทำให้ผลสรุปในงานวิจัยผิดพลาดและไม่สามารถนำไปใช้ประโยชน์ได้จริง

ดังนั้นในการศึกษาด้วยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล จึงต้องพิจารณาด้วยงานวิจัยที่ศึกษามีลักษณะหรือสภาพการณ์อย่างไร เหมาะสมกับการศึกษาด้วยการจำลองข้อมูลเทคนิคมอนติคาร์โลหรือควรจะเก็บจากข้อมูลจริง การวิจัยที่ใช้วิธีการจำลองข้อมูลควรเป็นงานวิจัยที่มีความสลับซับซ้อนไม่สามารถได้จากการเก็บรวบรวมข้อมูลจริง คำถามวิจัยมักเป็นคำถามที่ต้องหาคำตอบได้ด้วยวิธีการทางคณิตศาสตร์ หรือวิธีการทางสถิติ จึงจะทำให้ผลการวิจัยครั้งนั้นได้ผลที่ถูกต้องตามวัตถุประสงค์ของงานวิจัย (Stone, 1993 อ้างถึงใน Harwell, M. และคณะ, 1996)

3.6 บทสรุปจากการศึกษาการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล

การจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลเป็นเทคนิคที่ใช้สำหรับการแก้ปัญหาในการคำนวณทางคณิตศาสตร์สำหรับสถานการณ์ที่มีกระบวนการที่ซับซ้อนหรือไม่สามารถแก้ไขได้ในเชิงเหตุผล ซึ่งการวิเคราะห์ทั่วไปไม่สามารถดำเนินการได้อย่างครบถ้วนและสมบูรณ์ โดยการจำลองข้อมูลที่แสดงเหตุการณ์เกิดขึ้นทั้งทางบวกและทางลบ โดยการสร้างเลขสุ่ม (random number) มาช่วยในการหาคำตอบของปัญหาที่ต้องการศึกษา

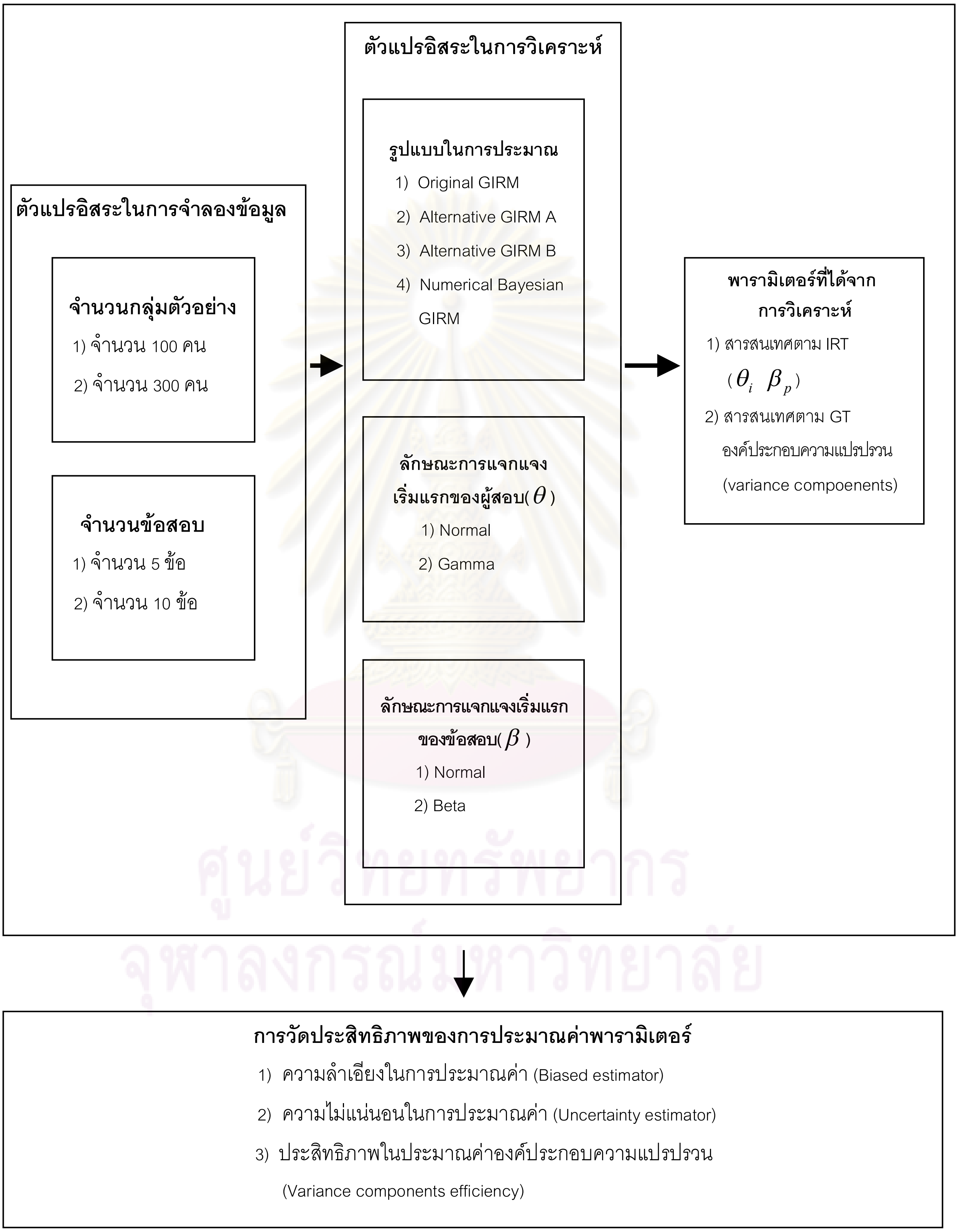
การวิเคราะห์ข้อสอบในปัจจุบัน โดยเฉพาะการวิเคราะห์ข้อสอบด้วยทฤษฎีการตอบสนองข้อสอบ มีการประยุกต์ใช้เทคนิคมอนติคาร์โลในการวิเคราะห์เป็นจำนวนมาก โดยเฉพาะในประเด็นเกี่ยวกับการประเมินกระบวนการในการประมาณค่าหรือความครอบคลุมของค่าพารามิเตอร์ นอกจากนี้ยังมีการประยุกต์ใช้เทคนิคมอนติคาร์โลในการประเมินคุณสมบัติทางสถิติของโมเดลในการวิเคราะห์ข้อสอบ รวมทั้งยังมีการประยุกต์ใช้เทคนิคมอนติคาร์โลในการเปรียบเทียบวิธีการต่างๆ ในโมเดลการวิเคราะห์ข้อสอบ

การวิเคราะห์ข้อสอบด้วยเทคนิคมอนติคาร์โลมีขั้นตอนที่สำคัญสามารถสรุปได้ 4 ขั้นตอน สำหรับการนำไปสู่การปฏิบัติการวิเคราะห์ ขั้นตอนแรก ได้แก่ การกำหนดปัญหาในการศึกษาวิจัย ขั้นตอนต่อมา เป็นขั้นตอนของการการออกแบบการทดลอง ซึ่งรวมถึงการระบุตัวแปรอิสระและตัวแปรตาม การออกแบบการวิจัยเชิงทดลอง จำนวนรอบในการคำนวณ และการเลือกโมเดลการวิเคราะห์ข้อสอบ ขั้นตอนที่สามเป็นการเขียนและระบุโปรแกรมในการจำลองข้อมูลและการประมาณค่าพารามิเตอร์ และขั้นตอนสุดท้ายเป็นขั้นตอนเกี่ยวกับการวิเคราะห์ผลจากการจำลองข้อมูล

ตอนที่ 4 กรอบแนวคิดในการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเกี่ยวกับแนวคิด/ทฤษฎีเกี่ยวกับการวิเคราะห์ข้อสอบ ซึ่งได้แก่ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ทฤษฎีการตอบสนองข้อสอบ และวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ การประมาณค่าพารามิเตอร์ ซึ่งได้แก่ วิธีการประมาณค่าพารามิเตอร์แบบฮิวริสติก แบบแม็กซิมั่มไลค์ลิฮูด และแบบเบส์ และการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โลทำให้สามารถสรุปกรอบแนวคิดในการวิจัยได้ดังนี้

กรอบแนวคิดในการวิจัย



แผนภาพที่ 16 กรอบแนวคิดในการวิจัย

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาในครั้งนี้มีวัตถุประสงค์หลัก เพื่อเปรียบเทียบประสิทธิภาพในการประมาณค่าของวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ ระหว่าง 4 รูปแบบ ได้แก่ Original GIRM, Alternative GIRM A , Alternative GIRM B และ Numerical Bayesian GIRM ที่มีขนาดกลุ่มตัวอย่างและจำนวนข้อในแบบสอบที่แตกต่างกัน นอกจากนี้ยังมีการศึกษาความไว (Sensitivity) ของผลการวิเคราะห์ข้อสอบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ของรูปแบบทั้งสามที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ (β) ข้อมูลที่ใช้ในการศึกษาครั้งนี้เป็นข้อมูลที่จำลองจากโปรแกรม R และทำการประมวลผลภายใต้จากการเขียนคำสั่งการประมวลผลด้วยโปรแกรม WinBUGS และกลับมาแสดงผลจะแสดงผลในโปรแกรม R โดยใช้โปรแกรมร่วมกันระหว่างโปรแกรม R และโปรแกรม WinBUGS ด้วย Package R2 WinBUGS มีรายละเอียดในการดำเนินการวิจัยดังนี้

1. ขั้นตอนการวิจัย

การศึกษาในครั้งนี้มีขั้นตอนการวิจัย ดังต่อไปนี้

1.1 การศึกษาเอกสารและงานวิจัยทั้งในและต่างประเทศที่เกี่ยวกับแนวคิด วิธีการและจุดเด่น ข้อจำกัดในแต่ละทฤษฏีการวิเคราะห์ข้อสอบ ซึ่งได้แก่ ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ทฤษฎีการตอบสนองข้อสอบ (IRT) และวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM)

1.2 การศึกษาแนวคิด หลักการ วิธีการใช้งานรวมทั้งวิธีการเขียนคำสั่งสำหรับการวิเคราะห์ด้วยโปรแกรม R โปรแกรม WinBUGS และ Package R2 WinBUGS

1.3 การจำลองข้อมูลด้วยโปรแกรม R จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า ตัวแปรที่ส่งผลต่อการประมาณค่าพารามิเตอร์ของข้อสอบ ได้แก่ จำนวนกลุ่มตัวอย่างในการวิเคราะห์ ข้อมูลและจำนวนข้อสอบที่ใช้ในการวิเคราะห์ข้อมูล ดังนั้นในการจำลองข้อมูล จึงมีเงื่อนไขในการจำลองข้อมูล 2 ตัวแปร ได้แก่ จำนวนกลุ่มตัวอย่างในการวิเคราะห์ 2 เงื่อนไข ได้แก่ 100 คนและ 300 คน จำนวนข้อสอบที่ใช้ในการวิเคราะห์ข้อมูล 2 เงื่อนไข ได้แก่ 5 ข้อและ10 ข้อ รวมเงื่อนไขในการจำลองข้อมูลทั้งสิ้น 4 เงื่อนไข

1.4 การสร้างโมเดลการวิเคราะห์ข้อมูล โดยในการศึกษาครั้งนี้ ตัวแปรที่เกี่ยวข้องกับการวิเคราะห์ มีจำนวน 3 ตัวแปร ได้แก่ ตัวแปรรูปแบบการวิเคราะห์ข้อสอบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) 4 รูปแบบ ประกอบด้วย รูปแบบที่ 1 Original GIRM ซึ่งเป็นรูปแบบต้นแบบที่ Briggs และ Willson (2007) พัฒนาขึ้น รูปแบบที่ 2 Alternative GIRM A และรูปแบบที่ 3 Alternative GIRM B เป็นรูปแบบที่ผู้วิจัยปรับวิธีการประมาณค่าพารามิเตอร์ขึ้นมาใหม่ และรูปแบบที่ 4 Original Bayesian GIRM เป็นรูปแบบที่ผู้วิจัยประยุกต์ใช้การประมาณค่าแบบเบส์ในการประมาณค่าพารามิเตอร์ของ IRT ที่ Swaminathan และ Gifford (1982, 1985, 1986) พัฒนาขึ้น ตัวแปรลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) 2 ลักษณะ ประกอบด้วย ลักษณะการแจกแจงแบบปกติ (Normal) และลักษณะการแจกแจงแบบแกมมา (Gamma) และตัวแปรลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) 2 ลักษณะ ประกอบด้วย ลักษณะการแจกแจงแบบปกติ (Normal) และลักษณะการแจกแจงแบบเบต้า (Beta) รวมโมเดลการวิเคราะห์ที่ผู้วิจัยต้องตรวจสอบทั้งสิ้น 16 เงื่อนไข (4 รูปแบบการวิเคราะห์ X 2 ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) X 2 ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β))

1.5 การตรวจสอบข้อมูลที่จำลองขึ้นกับโมเดลการวิเคราะห์ข้อมูลด้วยโปรแกรม WinBUGS ภายใต้การเขียนคำสั่ง การประมวลผล และกลับมาแสดงผลในโปรแกรม R โดยใช้โปรแกรมร่วมกันระหว่างโปรแกรม R และโปรแกรม WinBUGS ด้วยโปรแกรม Package R2 WinBUGS ดังนั้นจะมีเงื่อนไขที่ผู้วิจัยต้องตรวจสอบทั้งสิ้น 64 เงื่อนไข (ข้อมูลจำลอง 4 เงื่อนไข X โมเดลการวิเคราะห์ 16 เงื่อนไข) การวิเคราะห์สำหรับการวิจัยในครั้งนี้ มีการวิเคราะห์ข้อมูลที่ได้จากการสุ่มโดยการทำกลุ่มข้อมูลซ้ำ 500 ชุดข้อมูล (data set) โดยการประมาณค่าซ้ำในแต่ละชุดข้อมูล 10,000 รอบ (number of iteration) หลังจากตัดข้อมูลในส่วนแรกออก 1000 รอบ (number of burnin) เนื่องจากการประมาณค่าในรอบแรก ๆ ค่าพารามิเตอร์ที่ต้องการการประมาณค่ายังมีการแกว่งตัวไม่ลู่เข้าค่าใดค่าหนึ่ง

1.6 การวัดประสิทธิภาพของการประมาณค่า ทำการวัดประสิทธิภาพตัวประมาณค่า 3 ดัชนี ได้แก่ ดัชนีแรก ความลำเอียงในการประมาณค่า (Biased estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ดัชนีที่สอง ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) และดัชนีสุดท้าย ประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC)

1.7 สรุปผลการวิเคราะห์ อภิปรายผลการวิเคราะห์และเขียนรายงานผลการวิจัย

1.8 ตีพิมพ์และเผยแพร่ผลการวิจัย

2. เงื่อนไขการจำลองข้อมูล

จากการศึกษาการวิเคราะห์ข้อสอบ พบว่า จำนวนกลุ่มตัวอย่างและจำนวนข้อสอบในการวิเคราะห์ข้อมูลส่งผลต่อการประมาณค่าพารามิเตอร์ในโมเดลการวิเคราะห์ข้อมูลด้วยโมเดลการตอบสนองของข้อสอบ (Swaminathan และ Gifford, 1982, 1985, 1986; Hulin, 1983; Lord, 1986; รัตนา ศรีเหรัญ, 2539; วิชุดา บัวคง, 2532; เกศมณี พยัคฆ์, 2542) พบว่า ถ้าจำนวนกลุ่มตัวอย่างและจำนวนข้อสอบเพิ่มขึ้นจะทำให้มีการประมาณค่าพารามิเตอร์ที่มีความถูกต้องสูงขึ้น โดย Rentz และ Bashaw (1975 อ้างถึงใน สุพล นิลกลาง, 2541) กล่าวว่า ผู้เข้าสอบจำนวน 500-1,000 คนจะทำให้การประมาณค่าพารามิเตอร์ของรูปแบบราส์ชมีค่าคงที่ที่เพิ่มขึ้นอย่างรวดเร็ว แต่ถ้าผู้เข้าสอบเพิ่มขึ้นเป็นจำนวน 2,000-4,000 คนจะเพิ่มขึ้นอย่างช้า ดังนั้นผู้วิจัยจึงเลือกเงื่อนไขในการจำลองข้อมูลในครั้งนี้ โดยพิจารณาจาก **จำนวนกลุ่มตัวอย่างและจำนวนข้อสอบเป็นเงื่อนไขในการจำลองข้อมูล**

2.1 จำนวนกลุ่มตัวอย่าง

ในเรื่องของจำนวนกลุ่มตัวอย่าง Swaminathan และ Gifford (1982, 1985, 1986) เสนอแนะว่าควรใช้กลุ่มตัวอย่างในการประมาณ 100 คนขึ้นไป ดังนั้นในการจำลองข้อมูลขนาดกลุ่มตัวอย่างในการวิเคราะห์ข้อมูล **ผู้วิจัยจึงใช้ขนาดกลุ่มตัวอย่างจำนวน 2 เงื่อนไข ได้แก่ 100 คน และ 300 คน**

2.2 จำนวนข้อสอบ

สำหรับในเรื่องของจำนวนข้อสอบ Briggs และ Wilson (2007) ได้ทำการจำลองข้อมูลข้อสอบ จำนวน 5 ข้อสำหรับทดสอบความแกร่งของวิธีการ GIRM เนื่องจากข้อจำกัดของโปรแกรม WinBUGS ที่ใช้หน่วยความจำของคอมพิวเตอร์ในการวิเคราะห์ข้อมูลมาก (Briggs และ Wilson, 2007) ดังนั้นในการจำลองข้อมูลขนาดกลุ่มตัวอย่างในการวิเคราะห์ข้อมูล **ผู้วิจัยจึงใช้จำนวนข้อสอบรวม 2 เงื่อนไข ได้แก่ 5 ข้อและ 10 ข้อ**

ดังนั้นเงื่อนไขในการจำลองข้อมูลในการศึกษาครั้งนี้ได้แก่

จำนวนกลุ่มตัวอย่าง ประกอบด้วย

1) กลุ่มตัวอย่างจำนวน 100 คน

2) กลุ่มตัวอย่างจำนวน 300 คน

จำนวนข้อสอบ ประกอบด้วย

1) ข้อสอบจำนวน 5 ข้อ

2) ข้อสอบจำนวน 10 ข้อ

3. เงื่อนไขในการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ผู้วิจัยได้ศึกษาวิธีการประมาณค่าพารามิเตอร์แบบ 1 พารามิเตอร์ของโมเดลการตอบสนองข้อสอบ (1 PL IRT) โดยมีแหล่งความคลาดเคลื่อนของการวัดที่สนใจศึกษา 1 แหล่ง (single facet measurement) ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เนื่องจากการพัฒนาวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) มีการพัฒนาในเบื้องต้นเพียงโมเดลเดียวในปี 2007 โดยโมเดลดังกล่าวยังมีปัญหาความไม่แน่นอนในการประมาณค่า (Uncertainty) เกิดขึ้น ดังนั้นในการศึกษาครั้งนี้จึงศึกษาโมเดลดังกล่าวก่อนเพื่อให้วิธีการ GIRM มีความสมบูรณ์ในโมเดลที่ง่ายต่อการวิเคราะห์ก่อนจึงจะขยายผลการศึกษาไปในรูปแบบการวิเคราะห์ที่ซับซ้อนต่อไป โดยโมเดลดังกล่าวใช้หลักการประมาณค่าแบบเบส์ (Bayesian Estimation) ซึ่งสามารถสรุปเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้

$$P(\theta\ ,\beta\ |X) = L(X\ \ |\theta\ ,\beta\ \ )f(\theta)g(\beta)$$

จากโมเดลการประมาณค่าดังกล่าว จะเห็นได้ว่าประกอบไปด้วย 3 ส่วน คือ เมทริกซ์ของคะแนนที่สังเกตได้ $L(X\ \ |\theta\ ,\beta\ \ )$ ซึ่งประมาณได้จากวิธีการประมาณค่าไลค์ลิฮู้ด (Likelihood Estimation) ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) ในการศึกษาครั้งนี้ผู้วิจัยมีจุดประสงค์หลักในการพัฒนาวิธีการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ดังนั้นในเงื่อนไขของการวิเคราะห์ข้อมูลในการศึกษาครั้งนี้จึงประกอบไปด้วยตัวแปรจำนวน 3 ตัวแปร ได้แก่รูปแบบการวิเคราะห์ข้อสอบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β)

3.1 รูปแบบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM)

วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) เป็นวิธีการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการกำหนดมาตรวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) อาศัยสมมติฐานของการกระจายไปยังแหล่งความคลาดเคลื่อนจากการวัดที่เกี่ยวข้อง โดยการประมาณค่า ความแปรปรวน (Variance Components) ในแต่ละองค์ประกอบของการวัดไปพร้อม

กับการประมาณค่าพารามิเตอร์ในโมเดลการตอบสนองข้อสอบ (IRT) โดยในการประมาณค่าพารามิเตอร์จะแบ่งการประมาณค่าออกเป็น 3 ขั้นตอนใหญ่ๆ ขั้นแรกจะเป็นการประมาณค่าพารามิเตอร์จากทฤษฎีการตอบสนองข้อสอบ ขั้นตอนที่สองเป็นขั้นตอนของการเตรียมข้อมูลสำหรับการประมาณค่า ในขั้นตอนที่สามโดยการนำค่าพารามิเตอร์ที่ได้จากขั้นตอนแรกมาวิเคราะห์และสร้างเป็นเมทริกซ์ของข้อมูลเพื่อเตรียมการวิเคราะห์ สำหรับในขั้นตอนที่สามและขั้นตอนสุดท้ายเป็นการประมาณค่าพารามิเตอร์ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) การศึกษาครั้งนี้จะศึกษาการวิเคราะห์ข้อสอบสำหรับการออกแบบการวัดที่มีแหล่งความคลาดเคลื่อน 1 แหล่ง (single Facet Measurement) สถานการณ์ p X i (p X i design) ด้วยโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (1 PL item response model)

รูปแบบการประมาณค่าวิธีการวิเคราะห์ข้อสอบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ในการศึกษาครั้งนี้จะมี 3 รูปแบบ ได้แก่ รูปแบบที่ 1 Original GIRM เป็นวิธีการวิเคราะห์ที่ Briggs และ Wilson (2007) พัฒนาขึ้น รูปแบบที่ 2 Alternative GIRM A รูปแบบที่ 3 Alternative GIRM B และรูปแบบที่ 4 Numerical Bayesian GIRM เป็นวิธีการวิเคราะห์ที่ผู้วิจัยได้พัฒนาขึ้นสำหรับการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) มีรายละเอียดในแต่ละรูปแบบของการวิเคราะห์ดังนี้

รูปแบบที่ 1 Original GIRM

การวิเคราะห์ในรูปแบบ Original GIRM ที่พัฒนาโดย Briggs และ Wilson (2007) ขั้นตอนแรกของการวิเคราะห์เพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ (IRT) ได้ใช้เทคนิค MCMC (Markov Chain Monte Carlo) โดยการคำนวนแบบ Metropolis-Hastings algorithm ภายในวิธีการ Gibbs sampler ด้วยโปรแกรม WinBUGS ซึ่งจะสามารถประมาณค่าการแจกแจงภายหลัง (posterior distribution) ของค่าพารามิเตอร์ได้หลังจากนั้นจึงเอาค่าพารามิเตอร์ที่ได้มาคำนวณค่าที่คาดหวังของคะแนนที่สังเกตได้ (E (X_{pi})) หรือค่าความน่าจะเป็นของผู้สอบที่ตอบข้อสอบถูก $P(\theta_p, \beta_i)$ ด้วยโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ และนำข้อมูลทั้งหมดมาสร้างเป็นเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) จากนั้นจึงทำการวิเคราะห์ขั้นตอนที่สามต่อเพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยการนำข้อมูลจากเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) มาเป็นข้อมูลในการวิเคราะห์ด้วยสูตรการคำนวณของ Kolen และ Harris (1987) สามารถสรุปขั้นตอนการประมาณค่าพารามิเตอร์ของรูปแบบที่ 1 Original GIRM ได้ด้วยแผนภาพ ดังนี้

ขั้นตอนที่ 1

การวิเคราะห์ IRT

ประมาณค่าด้วย MCMC (วิธีGibbs sampler) ด้วยโปรแกรม WinBUGS

Posterior distribution
- พารามิเตอร์ผู้สอบ (θ_p)
- พารามิเตอร์ข้อสอบ (β_i)

ขั้นตอนที่ 2

นำค่า Posterior distribution ที่ได้ $(\theta_p^{(m)})$ และ $(\beta_i^{(m)})$ มาคำนวณค่า $E(X_{Pi}^{(m)})$ หรือ $P(\theta_P, \beta_i)$ ด้วยโมเดลการตอบสนองข้อสอบแบบ 1 PL

	$\beta_1^{(m)}$	…	$\beta_i^{(m)}$	…	$\beta_I^{(m)}$
$\theta_1^{(m)}$	$E(\mathrm{X}_{11}^{(m)})$		…		$E(\mathrm{X}_{1I}^{(m)})$
…					
$\theta_p^{(m)}$	…		$E(\mathrm{X}_{pi}^{(m)})$		…
…					
$\theta_P^{(m)}$	E($\mathrm{X}_{P1}^{(m)}$)		…		E($\mathrm{X}_{1P}^{(m)}$)

Expected Response Matrix

ขั้นตอนที่ 3

โมเดลการวิเคราะห์	**วิธีการประมาณค่า**	**สารสนเทศที่ได้**
การวิเคราะห์ GT	ประมาณค่าด้วย MCMC (วิธีGibbs sampler) ด้วยสูตรของ Kolen และ Harris (1987)	Posterior distribution - Variance component $\sigma^2(\pi)$ $\sigma^2(i)$ $\sigma^2(v)\sigma^2(e)$

แผนภาพที่ 17 ขั้นตอนการวิเคราะห์ข้อสอบด้วยวิธี Original GIRM

โดยมีรายละเอียดและสูตรในการคำนวณดังนี้

ขั้นตอนที่ 1 การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT)

จุดมุ่งหมายของการวิเคราะห์ในขั้นตอนนี้เพื่อให้ได้สารสนเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อนำมาใช้ในการวิเคราะห์ในขั้นตอนต่อไป

การวิเคราะห์ในขั้นตอนนี้ประมาณค่าพารามิเตอร์ของผู้สอบ ได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบเป็นรายบุคคล (θ) และค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยากของข้อสอบ (β) โดยการประมาณค่าสามารถทำได้โดยใช้เทคนิค MCMC (Markov Chain Monte Carlo) โดยการคำนวนแบบ Metropolis-Hastings algorithm ภายในวิธีการ Gibbs sampler ด้วยโปรแกรม WinBUGS ซึ่งจะสามารถประมาณค่า Posterior distribution ของ ความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ) ซึ่งสามารถเขียนเป็นสมการสำหรับประมาณค่าความสามารถที่แท้จริงของผู้สอบ (θ) ได้ดังนี้

$$P(\theta \mid \beta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\theta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\theta}$$

และสามารถประมาณค่า Posterior distribution ของค่าพารามิเตอร์ของข้อสอบแต่ละข้อ (β) ซึ่งสามารถเขียนเป็นสมการสำหรับประมาณค่าพารามิเตอร์ของข้อสอบ (β) ได้ดังนี้

$$P(\beta \mid \theta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\beta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\beta}$$

จากสมการ $P(\theta \mid \beta, X)$ และสมการ $P(\beta \mid \theta, X)$ ที่สามารถประมาณได้จากเทคนิค MCMC จะทำให้ค่าสารสนเทศจากการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (1PL) ซึ่งได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ_p) และ ค่าความยากของข้อสอบแต่ละข้อ (β_i)

ขั้นตอนที่ 2 ขั้นการวิเคราะห์ค่าโอกาสในการทำข้อสอบในแต่ละข้อของแต่ละคนเพื่อใช้สำหรับการวิเคราะห์ในขั้นตอนต่อไป

นำค่าการแจกแจงภายหลัง (Posterior distribution) ที่ได้ในขั้นตอนที่หนึ่งซึ่งได้แก่ค่า (θ_p) และ (β_i) มาคำนวนค่า E (X_{pi}) หรือ $P(\theta_p, \beta_i)$ สำหรับโมเดลการวิเคราะห์แบบ p x i ด้วยข้อคำถามแบบสองค่า (dichotomous items) สามารถแสดงสมการแบบ one parameter logistic ได้ดังนี้

$$P(\theta_p, \beta_i) = P(X_{pi} = 1 \mid \theta_p, \beta_i) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)}$$

จากการประมาณค่าดังกล่าวทำให้สามารถสร้างExpected response matrix ได้ดังนี้

พารามิเตอร์ของข้อสอบ

		β_1	…	β_i	…	β_I
พารามิเตอร์ของผู้สอบ	θ_1	E (X_{11})		…		E (X_{1I})
	…					
	θ_p	…		E (X_{pi})		….
	…					
	θ_P	E (X_{P1})		…		E (X_{PI})

ซึ่งเราจะใช้ค่า E (X_{pi}) หรือค่า $P(\theta_p, \beta_i)$ ที่ได้จากการวิเคราะห์มาใช้ในการคำนวณในขั้นตอนต่อไป

ขั้นตอนที่ 3 การวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

จุดมุ่งหมายของการวิเคราะห์ในขั้นตอนนี้เพื่อให้ได้สารเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) นั้นคือการประมาณองค์ประกอบความแปรปรวน (Variance component) และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability coefficient) สามารถประมาณได้ ด้วยกรอบแนวคิดของทฤษฎีการตอบสนองข้อสอบ (IRT) ที่ได้จากการเชื่อมโยงสูตรของ Kolen และ Harris (1987) ดังนี้

ค่าเฉลี่ยของคะแนนแบบสอบของผู้เข้าสอบทั้งหมด

$$\mu = \int_\theta \int_\beta P(\theta_p, \beta_i) f(\beta) g(\theta) d\beta d\theta$$ (เทียบได้กับ The GT grand mean)

ค่าเฉลี่ยของคะแนนแบบสอบในแต่ละบุคคล

$$\pi(\theta) = \int_\beta P(\theta_p, \beta_i) f(\beta) d\beta - \mu$$ (เทียบได้กับ The GT person effect)

ค่าเฉลี่ยของคะแนนผู้เข้าสอบทั้งหมดในข้อสอบแต่ละข้อ

$$\iota(\theta) = \int_\theta P(\theta_p, \beta_i) g(\theta) d\theta - \mu$$ (เทียบได้กับ The GT item effect)

คะแนนปฏิสัมพันธ์ระหว่างผู้เข้าสอบและข้อสอบ

$$\nu(\theta,\beta) = P(\theta_p,\beta_i) - \pi(\theta) - \iota(\beta) - \mu$$

(ไม่สามารถเทียบได้กับ The GT interaction effect เนื่องจาก ใน Kolen และ Harris ปฏิสัมพันธ์ไม่ได้รวมความคลาดเคลื่อนส่วนที่เหลือ เหมือนในการวิเคราะห์ด้วย GT)

ในทางปฏิบัตินักวัดผลคาดหวังให้ค่า $\pi(\theta)$ $\iota(\theta$ และ ค่า $\nu(\theta,\beta)$ มีค่า Covariance เท่ากับ 0 ดังนั้นจึงสามารถสรุปเป็นสมการได้ดังนี้คือ

$$\int_\theta \pi(\theta)g(\theta)d(\theta) = \int_\beta i(\theta)f(\beta)d(\beta) = \int_\theta \nu(\theta)g(\theta)d(\theta)$$

$$= \int_\beta \nu(\theta,\beta)f(\beta)d(\beta) = \int_\theta\int_\beta \nu(\theta,\beta)f(\beta)g(\theta)d\beta d\theta = 0$$

ดังนั้น

$$\sigma^2(\pi) = \int_\theta \pi^2(\theta)g(\theta)d(\theta) \qquad \textbf{(คำอธิบาย } \nu_e \textbf{ ใน GT)}$$

$$\sigma^2(i) = \int_\beta i^2(\beta)g(\beta)d(\beta) \qquad \textbf{(คำอธิบาย } \nu_i \textbf{ ใน GT)}$$

$$\sigma^2(\nu) = \int_\theta\int_\beta \nu^2(\theta,\beta)f(\beta)g(\theta)d\beta d\theta$$

สามารถสรุป การคำนวณ GT ด้วยการใช้สูตรการคำนวณของ IRT ตามสมการของ Kolen และ Harris ได้ดังนี้ คือ

$$P(\theta_p,\beta_i) = \mu + \pi(\theta) + \iota(\theta) + \nu(\theta,\beta) + \ldots$$

และเนื่องด้วย $P(\theta_p,\beta_i) = E(X_{pi})$ ดังนั้น จึงต้องมีการระบุความคลาดเคลื่อนที่เกิดขึ้น $e_{pi} = e(\theta_p,\beta_i)$ ซึ่งมีค่าเท่ากับ

$$X_{pi} = E(X_{pi}) + e_{pi} = \mu + \pi(\theta) + \iota(\theta) + \nu(\theta,\beta) + e_{pi}$$

สืบเนื่องจากคะแนนที่สังเกตได้ X_{pi} มีลักษณะการแจกแจงแบบแบล์นูลีย์ (Bernoulli distribution) ดังนั้นความแปรปรวนของความคลาดเคลื่อน (e_{pi}) คือ [$P(\theta_p,\beta_i)$] [1- $P(\theta_p,\beta_i)$] โดยสามารถสรุปเป็นสมการได้ดังนี้

$$\sigma^2(e) = \int_\theta\int_\beta P(\theta_p,\beta_i)[1 - P(\theta_P,\beta_i)]f(\beta)g(\theta)d\beta d\theta$$

สามารถเขียนในรูปของความแปรปรวนทั้งหมดของคะแนนที่สังเกตได้ดังนี้

$$\sigma^2(X_{pi}) = \sigma^2(\pi) + \sigma^2(i) + \sigma^2(\nu) + \sigma^2(e)$$

จากสูตรดังกล่าวจะทำให้สามารถประมาณองค์ประกอบความแปรปรวน (Variance components) และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability coefficient)ได้ ซึ่งจะให้สารสนเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

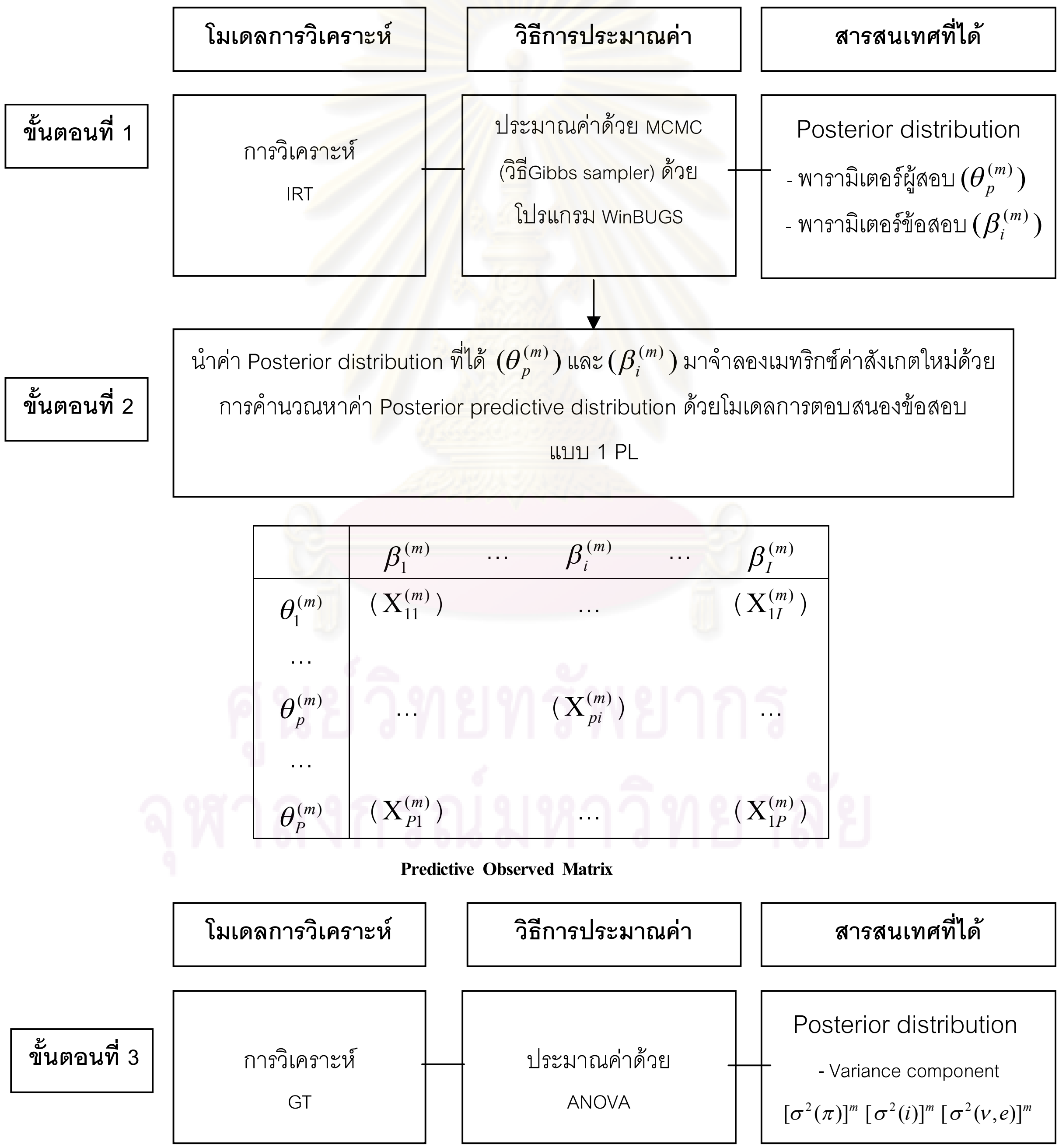
การวิเคราะห์จะประมวลผลภายใต้จากการเขียนคำสั่งการประมวลผลด้วยโปรแกรม WinBUGS (Spiegelhalter และคณะ, 2004) โดยการจำลองข้อมูลด้วยโปรแกรม R และกลับมาแสดงผลจะแสดงผลในโปรแกรม R โดยใช้โปรแกรมร่วมกันระหว่างโปรแกรม R และ WinBUGS ด้วย โปรแกรม Package R2 WinBUGS

รูปแบบที่ 2 Alternative GIRM A

การวิเคราะห์ในรูปแบบ Alternative GIRM A เป็นการวิเคราะห์ที่พัฒนาขึ้นเพื่อแก้ไขปัญหาการประมาณค่าพารามิเตอร์ที่ไม่เป็นอิสระต่อกันระหว่างแถวและคอลัมน์ในเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) และส่งผลทำให้การประมาณค่าที่ได้จากการวิเคราะห์ในรูปแบบ Original GIRM เกิดความไม่แน่นอนในการประมาณค่า ดังนั้นจึงมีผู้เสนอว่าต้องทำให้คะแนนที่ได้จากการวิเคราะห์ในขั้นตอนก่อนนำมาวิเคราะห์ในขั้นตอนที่ 3 ต้องไม่เป็นค่าที่คาดหวัง (Expected value) แต่จะต้องเป็นคะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) แทนเพื่อให้ได้มาก่อนนำไปวิเคราะห์ในขั้นตอนที่ 3 (Briggs และ Wilson, 2007) จึงเป็นที่มาของวิธีการในรูปแบบที่ 2

การวิเคราะห์ในรูปแบบที่ 2 Alternative GIRM A ประกอบด้วยขั้นตอนการวิเคราะห์ 3 ขั้นตอน ได้แก่ ขั้นตอนแรกของการวิเคราะห์เพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ (IRT) ได้ใช้เทคนิค MCMC (Markov Chain Monte Carlo) โดยการคำนวณแบบ Metropolis-Hastings algorithm ภายในวิธีการ Gibbs sampler ด้วยโปรแกรม WinBUGS ซึ่งจะสามารถประมาณค่าการแจกแจงภายหลัง (Posterior distribution) ของค่าพารามิเตอร์ได้ ขั้นตอนที่สอง เป็นการวิเคราะห์โดยการนำค่าการแจกแจงภายหลัง (Posterior distribution) ที่ประมาณค่าพารามิเตอร์ (θ_p) และ (β_i) ได้มาจำลองเมทริกซ์ค่าสังเกตใหม่ด้วยการคำนวณคะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) ซึ่งเป็นค่าเฉลี่ยของคะแนนที่สังเกตได้ โดยได้มาจากการวิเคราะห์ได้ใช้ เทคนิคMCMC (Markov Chain Monte Carlo) ดังนั้น

ในขั้นตอนที่ 2 นี้จะทำให้ได้ Predictive Observed Matrix แทน Expected Response Matrix ที่ได้จากการประมาณค่าด้วยรูปแบบที่ 1 และขั้นตอนที่ 3 จึงทำการวิเคราะห์เพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยการนำข้อมูลจากคะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) มาเป็นข้อมูลในการวิเคราะห์ด้วยสูตรการวิเคราะห์ความแปรปรวน (ANOVA) ประมาณค่าองค์ประกอบความแปรปรวนสามารถสรุปขั้นตอนการวิเคราะห์เป็นแผนภาพได้ ดังนี้



แผนภาพที่ 18 ขั้นตอนการวิเคราะห์ข้อสอบด้วยวิธี Alternative GIRM A

โดยมีรายละเอียดและสูตรในการคำนวณดังนี้

ขั้นตอนที่ 1 การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT)

จุดมุ่งหมายของการวิเคราะห์ในขั้นตอนนี้เพื่อให้ได้สารสนเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อนำมาใช้ในการวิเคราะห์ในขั้นตอนต่อไป

การวิเคราะห์ในขั้นตอนนี้จะประมาณค่าพารามิเตอร์ของผู้สอบ ได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบเป็นรายบุคคล (θ) และค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยากของข้อสอบ (β) โดยการประมาณค่าสามารถทำได้โดยใช้เทคนิค MCMC (Markov Chain Monte Carlo) โดยการคำนวนแบบ Metropolis-Hastings algorithm ภายในวิธีการ Gibbs sampler ด้วยโปรแกรม WinBUGS ซึ่งจะสามารถประมาณค่าลักษณะการแจกแจงภายหลัง (Posterior distribution) ของความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ) ซึ่งสามารถเขียนเป็นสมการสำหรับประมาณค่าความสามารถที่แท้จริงของผู้สอบ (θ) ได้ดังนี้

$$P(\theta \mid \beta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\theta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\theta}$$

และสามารถประมาณค่าลักษณะการแจกแจงภายหลัง (Posterior distribution) ของค่าพารามิเตอร์ของข้อสอบแต่ละข้อ (β) ซึ่งสามารถเขียนเป็นสมการสำหรับประมาณค่าพารามิเตอร์ของข้อสอบ (β) ได้ดังนี้

$$P(\beta \mid \theta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\beta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\beta}$$

จากสมการ $P(\theta \mid \beta, X)$ และสมการ $P(\beta \mid \theta, X)$ ที่สามารถประมาณได้จากเทคนิค MCMC จะทำให้ค่าสารสนเทศจากการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (1PL) ซึ่งได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ_p) และ ค่าความยากของข้อสอบแต่ละข้อ (β_i)

ขั้นตอนที่ 2 ขั้นการวิเคราะห์ค่าโอกาสในการทำข้อสอบในแต่ละข้อของแต่ละคนเพื่อใช้สำหรับการวิเคราะห์ในขั้นตอนต่อไป

ขั้นตอนนี้เป็นการวิเคราะห์คะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) ซึ่งเป็นค่าที่เราต้องการจะสรุปอ้างอิงค่าตัวแปรนั้นที่ไม่ทราบค่าแต่สามารถสังเกตค่า

นั้นได้ ในขั้นตอนนี้เราต้องการจะหาค่าเฉลี่ยของคะแนนของผู้สอบที่มีความสามารถเท่ากับ (θ_p) ทำข้อสอบข้อที่มีระดับความยาก (β_i) ดังนั้นสามารถหาค่าการแจกแจงการทำนายภายหลังได้จากสูตรดังต่อไปนี้

$$p(\tilde{y}) = \int p(\tilde{y},\theta)d\theta = \int p(\theta)p(\tilde{y} \mid \theta)d\theta$$

$$\begin{aligned} p(\tilde{y} \mid y) &= \int p(\tilde{y},\theta \mid y)d\theta \\ &= \int p(\tilde{y} \mid \theta, y)p(\theta \mid y)d\theta \\ &= \int p(\tilde{y} \mid \theta)p(\theta \mid y)d\theta \end{aligned}$$

จากสมการจะเห็นว่าการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) ของค่าทำนายคือ อินทิกรัลของผลคูณระหว่างฟังก์ชันภาวะความควรจะเป็น $p(\tilde{y} \mid \theta)$ และการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์เมื่อกำหนดค่าสังเกต $p(\theta \mid y)$

กรณีในการศึกษาครั้งนี้สามารถนำค่าการแจกแจงภายหลัง (Posterior distribution) ที่ได้ในขั้นตอนที่หนึ่งซึ่งได้แก่ค่า (θ_p) และ (β_i) มาคำนวนคะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) สำหรับโมเดลการวิเคราะห์แบบp x i ด้วยข้อคำถามแบบสองค่า (dichotomous items) สามารถแสดงสมการแบบ one parameter logistic ได้ดังนี้

$$\begin{aligned} p(\hat{x}_p \mid x_p) &= \int_\beta \int_\theta P(\hat{x}_p, \theta_p, \beta_i \mid x_p)d\theta d\beta \\ &= \int_\beta \int_\theta P(\hat{x}_p \mid \theta_p, \beta_i, x_p)P(\theta_p, \beta_i \mid x_p)d\theta d\beta \\ &= \int_\beta \int_\theta P(\hat{x}_p \mid \theta_p, \beta_i)P(\theta_p, \beta_i \mid x_p)d\theta d\beta \end{aligned}$$

จากการประมาณค่าดังกล่าวทำให้สามารถสร้าง Predictive observed matrix ได้ดังนี้

พารามิเตอร์ของข้อสอบ

		β_1	…	β_i	…	β_I
	θ_1	$X_{11}^{(m)}$		…		$X_{1I}^{(m)}$
พารามิเตอร์ของผู้สอบ	…					
	θ_p	…		$X_{pi}^{(m)}$		….
	…					
	θ_P	$X_{P1}^{(m)}$		…		$X_{pi}^{(m)}$

ซึ่งเราจะใช้ค่า $\mathrm{X}_{pi}^{(m)}$ ที่ได้จากการวิเคราะห์มาใช้ในการคำนวณในขั้นตอนต่อไป

ขั้นตอนที่ 3 การวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

จุดมุ่งหมายของการวิเคราะห์ในขั้นตอนนี้เพื่อให้ได้สารเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) นั้นคือการประมาณองค์ประกอบความแปรปรวน (Variance component) และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability coefficient) สามารถประมาณได้ด้วยการวิเคราะห์ความแปรปรวน (ANOVA) ดังนี้

คะแนนรวมจากบุคคลและข้อคำถาม (The grand mean across persons and items)

$$\mu \equiv E_p E_i X_{pi}$$

คะแนนที่เกิดจากความสามารถของบุคคล (person-specific mean)

$$\mu_p \equiv E_i X_{pi}$$

คะแนนของข้อคำถาม (item-specific mean)

$$\mu_i \equiv E_p X_{pi}$$

โดยสามารถคำนวณองค์ประกอบความแปรปรวนได้ดังนี้

อิทธิพลของบุคคล (The person effect)

$$\nu_p = \mu_p - \mu$$

อิทธิพลของข้อสอบ (The item effect)

$$\nu_i = \mu_i - \mu$$

อิทธิพลของปฏิสัมพันธ์และความคลาดเคลื่อนแบบสุ่ม (The residual/interaction effect)

$$\nu_{pi,e} = X_{pi} - \mu_p - \mu_i + \mu$$

สามารถเขียนสมการในรูปของความแปรปรวนของคะแนนที่สังเกตได้ดังนี้

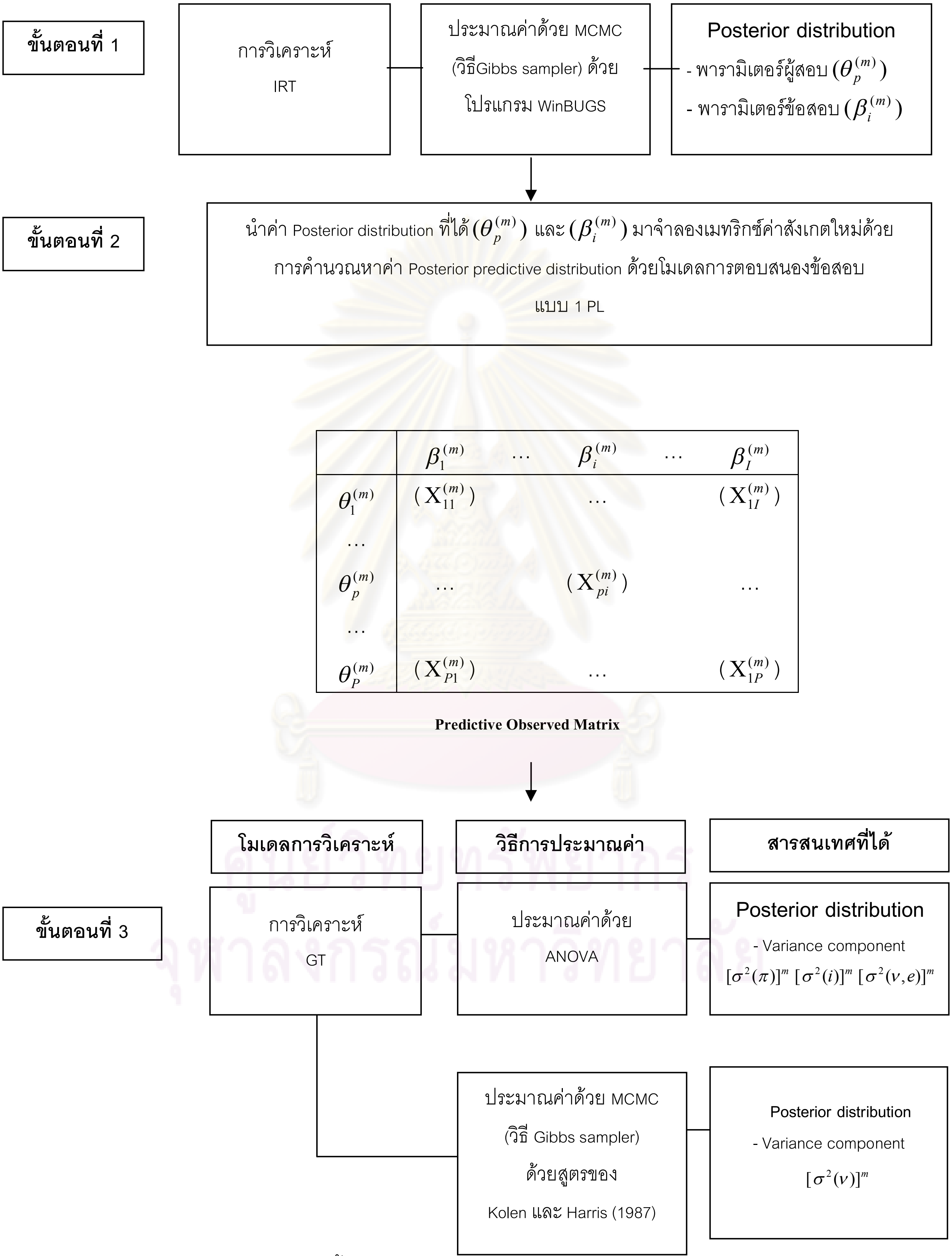
$$\sigma^2(\mathrm{X}_{pi}) = \sigma^2(p) + \sigma^2(i) + \sigma^2(pi,e)$$

จากสูตรดังกล่าวจะทำให้สามารถประมาณองค์ประกอบความแปรปรวน (Variance component) และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability coefficient) ได้ซึ่งจะให้สารสนเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

การวิเคราะห์จะประมวลผลภายใต้จากการเขียนคำสั่งการประมวลผลด้วย โปรแกรม R และกลับมาแสดงผลจะแสดงผลในโปรแกรม R โดยใช้โปรแกรมร่วมกันระหว่างโปรแกรม R และ WinBUGS ด้วย Package R2 WinBUGS

รูปแบบที่ 3 Alternative GIRM B

การประมาณค่าพารามิเตอร์ของวิธีการ GIRM ในรูปแบบที่ 3 มีลักษณะคล้ายคลึงกับการประมาณค่าในรูปแบบที่ 2 แตกต่างกันที่การวิเคราะห์ในรูปแบบที่ 2 สำหรับการประมาณองค์ประกอบความแปรปรวน (Variance component) ใช้การประมาณด้วยการวิเคราะห์ความแปรปรวน (ANOVA) ซึ่งไม่สามารถแยกผลการวิเคราะห์ของอิทธิพลของปฏิสัมพันธ์และความคลาดเคลื่อนแบบสุ่ม ($v_{pi,e}$) ดังนั้นการวิเคราะห์ในรูปแบบที่ 3 นี้จึงใช้สูตรการประมาณองค์ประกอบความแปรปรวน (Variance component) และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability coefficient) ด้วยสูตรของ Kolen และ Harris (1987) สามารถสรุปขั้นตอนการวิเคราะห์เป็นแผนภาพได้ดังนี้



แผนภาพที่ 19 ขั้นตอนการวิเคราะห์ข้อสอบด้วยวิธี Alternative GIRM B

โดยมีรายละเอียดและสูตรในการคำนวณดังนี้

ขั้นตอนที่ 1 การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT)

จุดมุ่งหมายของการวิเคราะห์ในขั้นตอนนี้เพื่อให้ได้สารสนเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อนำมาใช้ในการวิเคราะห์ในขั้นตอนต่อไป

การวิเคราะห์ในขั้นตอนนี้จะประมาณค่าพารามิเตอร์ของผู้สอบ ได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบเป็นรายบุคคล (θ) และค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยากของข้อสอบ (β) โดยการประมาณค่าสามารถทำได้โดยใช้เทคนิค MCMC (Markov Chain Monte Carlo) โดยการคำนวนแบบ Metropolis-Hastings algorithm ภายในวิธีการ Gibbs sampler ด้วยโปรแกรม WinBUGS ซึ่งจะสามารถประมาณค่าลักษณะการแจกแจงภายหลัง (Posterior distribution) ของความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ) ซึ่งสามารถเขียนเป็นสมการสำหรับประมาณค่าความสามารถที่แท้จริงของผู้สอบ (θ) ได้ดังนี้

$$P(\theta \mid \beta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\theta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\theta}$$

และสามารถประมาณค่าลักษณะการแจกแจงภายหลัง (Posterior distribution) ของค่าพารามิเตอร์ของข้อสอบแต่ละข้อ ซึ่งสามารถเขียนเป็นสมการสำหรับประมาณค่าพารามิเตอร์ของข้อสอบ (β) ได้ดังนี้

$$P(\beta \mid \theta, X) = \frac{L(X \mid \theta, \beta) f(\theta) g(\beta)}{\int_{\beta} L(X \mid \theta, \beta) f(\theta) g(\beta) d\beta}$$

จากสมการ $P(\theta \mid \beta, X)$ และสมการ $P(\beta \mid \theta, X)$ ที่สามารถประมาณได้จากเทคนิค MCMC จะทำให้ค่าสารสนเทศจากการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (1PL) ซึ่งได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ_p) และค่าความยากของข้อสอบแต่ละข้อ (β_i)

ขั้นตอนที่ 2 ขั้นการวิเคราะห์ค่าโอกาสในการทำข้อสอบในแต่ละข้อของแต่ละคน เพื่อใช้สำหรับการวิเคราะห์ในขั้นตอนต่อไป

ขั้นตอนนี้เป็นการวิเคราะห์คะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) ซึ่งเป็นค่าที่เราต้องการจะสรุปอ้างอิงค่าตัวแปรนั้นที่ไม่ทราบค่าแต่สามารถสังเกตค่านั้นได้ ในขั้นตอนนี้เราต้องการจะหาค่าเฉลี่ยของคะแนนที่ผู้สอบที่มีความสามารถเท่ากับ (θ_p) ทำข้อสอบข้อที่มีระดับความยาก (β_i) ดังนั้นสามารถหาค่าการแจกแจงการทำนายภายหลังได้จากสูตรดังต่อไปนี้

$$p(\tilde{y}) = \int p(\tilde{y},\theta)d\theta = \int p(\theta)p(\tilde{y} \mid \theta)d\theta$$

$$\begin{aligned} p(\tilde{y} \mid y) &= \int p(\tilde{y},\theta \mid y)d\theta \\ &= \int p(\tilde{y} \mid \theta, y)p(\theta \mid y)d\theta \\ &= \int p(\tilde{y} \mid \theta)p(\theta \mid y)d\theta \end{aligned}$$

จากสมการจะเห็นว่าการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) ของค่าทำนายคือ อินทิกรัลของผลคูณระหว่างฟังก์ชันภาวะความควรจะเป็น $p(\tilde{y} \mid \theta)$ และการแจกแจงความน่าจะเป็นภายหลังของพารามิเตอร์เมื่อกำหนดค่าสังเกต $p(\theta \mid y)$

กรณีในการศึกษาครั้งนี้สามารถนำค่าการแจกแจงภายหลัง (Posterior distribution) ที่ได้ในขั้นตอนที่หนึ่งซึ่งได้แก่ค่า (θ_p) และ (β_i) มาคำนวนคะแนนการแจกแจงการทำนายภายหลัง (Posterior predictive distribution) สำหรับโมเดลการวิเคราะห์แบบ p x i ด้วยข้อคำถามแบบสองค่า (dichotomous items) สามารถแสดงสมการแบบ one parameter logistic ได้ดังนี้

$$\begin{aligned} p(\hat{x}_p \mid x_p) &= \int_\beta \int_\theta P(\hat{x}_p, \theta_p, \beta_i \mid x_p) d\theta d\beta \\ &= \int_\beta \int_\theta P(\hat{x}_p \mid \theta_p, \beta_i, x_p) P(\theta_p, \beta_i \mid x_p) d\theta d\beta \\ &= \int_\beta \int_\theta P(\hat{x}_p \mid \theta_p, \beta_i) P(\theta_p, \beta_i \mid x_p) d\theta d\beta \end{aligned}$$

จากการประมาณค่าดังกล่าวทำให้สามารถสร้าง Predictive observed matrix ได้ดังนี้

พารามิเตอร์ของข้อสอบ

		β_1	$\cdots$	β_i	$\cdots$	β_I
	θ_1	$X_{11}^{(m)}$		$\cdots$		$X_{1I}^{(m)}$
พารามิเตอร์	$\cdots$					
ของผู้สอบ	θ_p	$\cdots$		$X_{pi}^{(m)}$		$\cdots$
	$\cdots$					
	θ_P	$X_{P1}^{(m)}$		$\cdots$		$X_{pi}^{(m)}$

ซึ่งเราจะใช้ค่า $X_{pi}^{(m)}$ ที่ได้จากการวิเคราะห์มาใช้ในการคำนวณในขั้นตอนต่อไป

ขั้นตอนที่ 3 การวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

จุดมุ่งหมายของการวิเคราะห์ในขั้นตอนนี้เพื่อให้ได้สารเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) นั้นคือการประมาณองค์ประกอบความแปรปรวน (Variance component) และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability coefficient) สามารถประมาณได้ ด้วยกรอบแนวคิดของทฤษฎีการตอบสนองข้อสอบ (IRT) ที่ได้จากการเชื่อมโยงสูตรของ Kolen และ Harris(1987) ดังนี้

$$\mu = \int_\theta \int_\beta P(\theta_p, \beta_i) f(\beta) g(\theta) d\beta d\theta$$ **(คำอธิบาย μ ใน GT)**

$$\pi(\theta) = \int_\beta P(\theta_p, \beta_i) f(\beta) d\beta - \mu$$ **(คำอธิบาย ν_p ใน GT)**

$$\iota(\theta) = \int_\theta P(\theta_p, \beta_i) g(\theta) d\theta - \mu$$ **(คำอธิบาย ν_i ใน GT)**

$$\nu(\theta, \beta) = P(\theta_p, \beta_i) - \pi(\theta) - \iota(\beta) - \mu$$ **(คำอธิบาย ν_{pi} ใน GT)**

$$\sigma^2(e) = \int_\theta \int_\beta P(\theta_p, \beta_i)[1 - P(\theta_P, \beta_i)] f(\beta) g(\theta) d\beta d\theta$$ **(คำอธิบาย ν_e ในGT)**

สามารถเขียนในรูปของความแปรปรวนทั้งหมดของคะแนนที่สังเกตได้ดังนี้

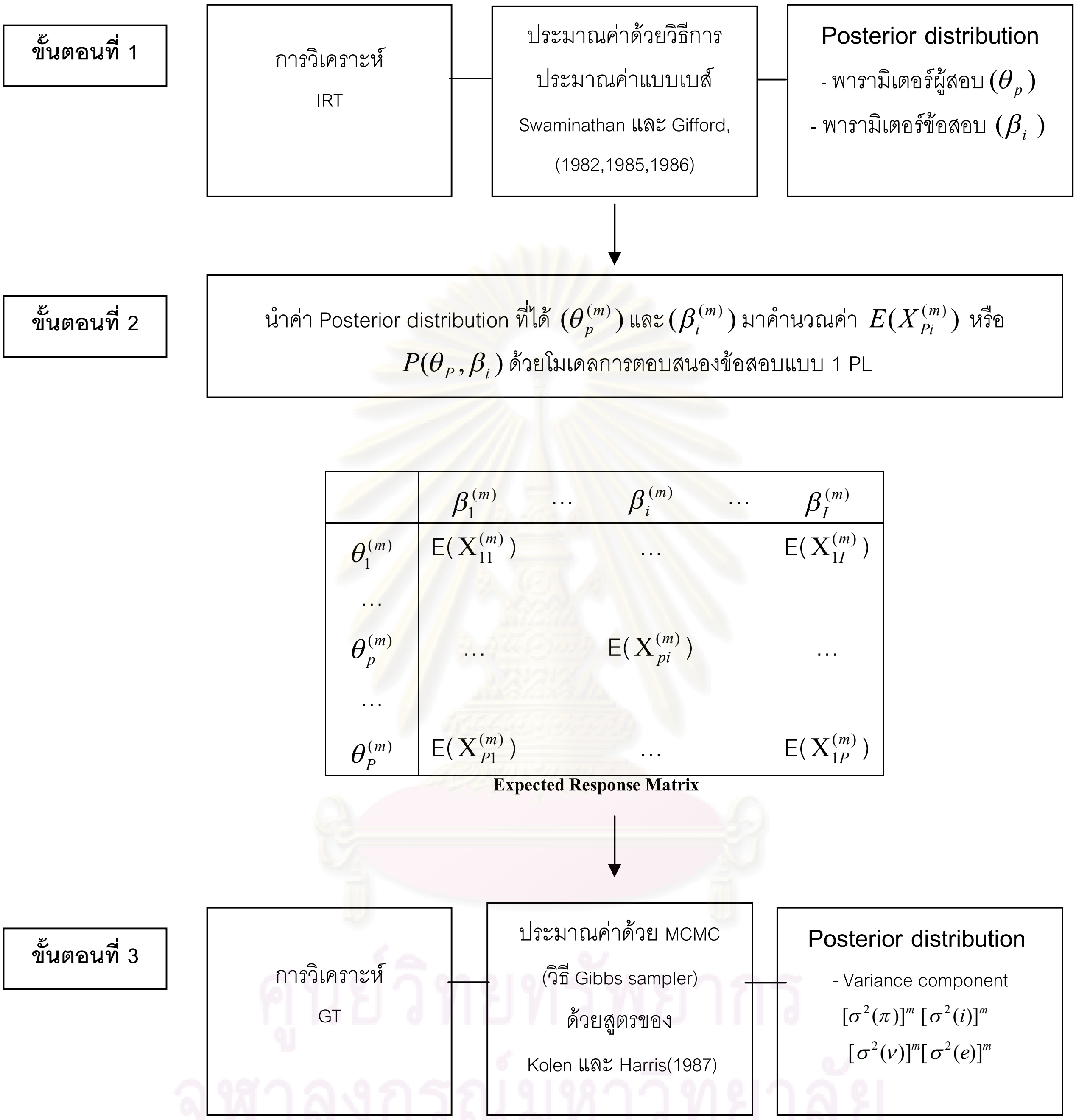
$$\sigma^2(X_{pi}) = \sigma^2(\pi) + \sigma^2(i) + \sigma^2(\nu) + \sigma^2(e)$$

จากสูตรดังกล่าวจะทำให้สามารถประมาณองค์ประกอบความแปรปรวน (Variance component) และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability coefficient) ได้ซึ่งจะให้สารสนเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

การวิเคราะห์จะประมวลผลภายใต้จากการเขียนคำสั่งการประมวลผลด้วยโปรแกรม WinBUGS (Spiegelhalter และคณะ, 2004) โดยการจำลองข้อมูลด้วยโปรแกรม R และกลับมาแสดงผลในโปรแกรม R โดยใช้โปรแกรมร่วมกันระหว่างโปรแกรม R และ WinBUGS ด้วยโปรแกรม Package R2 WinBUGS

รูปแบบที่ 4 Numerical Bayesian GIRM

การวิเคราะห์ในรูปแบบ Original Bayesian GIRM ขั้นตอนแรกของการวิเคราะห์เพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ (IRT) ได้ใช้เทคนิคการประมาณค่าพารามิเตอร์แบบเบส์ (Bayesian) ที่ Swaminathan และGifford (1982, 1985, 1986) เป็นผู้เสนอไว้ ซึ่งจะสามารถประมาณค่าลักษณะการแจกแจงภายหลัง (Posterior distribution) ของค่าพารามิเตอร์ได้ หลังจากนั้นจึงเอาค่าพารามิเตอร์ที่ได้มาคำนวณค่า E($\mathbf{X}_{pi}$) หรือก็คือค่า $P(\theta_p, \beta_i)$ ด้วยโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์(1PL) และนำข้อมูลทั้งหมดมาสร้างเป็นเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) จากนั้นจึงทำการวิเคราะห์ขั้นตอนที่สามต่อเพื่อให้ได้สารสนเทศจากการประมาณค่าพารามิเตอร์ในทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยการนำข้อมูลจากเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) มาเป็นข้อมูลในการวิเคราะห์ด้วยสูตรการคำนวณของ Kolen และ Harris(1987) สามารถสรุปขั้นตอนการวิเคราะห์เป็นแผนภาพได้ดังนี้



แผนภาพที่ 20 ขั้นตอนการวิเคราะห์ข้อสอบด้วยวิธี Numerical Bayesian GIRM

โดยมีรายละเอียดและสูตรในการคำนวณดังนี้

ขั้นตอนที่ 1 การวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ (IRT)

จุดมุ่งหมายของการวิเคราะห์ในขั้นตอนนี้เพื่อให้ได้สารสนเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อนำมาใช้ในการวิเคราะห์ในขั้นตอนต่อไป

การวิเคราะห์ในขั้นตอนนี้เป็นการประมาณค่าพารามิเตอร์ของผู้สอบ ซึ่งได้แก่ค่าความสามารถที่แท้จริงของผู้สอบเป็นรายบุคคล (θ) และค่าพารามิเตอร์ของข้อสอบ ซึ่งได้แก่ค่าความยากของข้อสอบ (β) การประมาณค่าทำได้โดยใช้วิธีการประมาณค่าของ Swaminathan และ Gifford (1982, 1985, 1986) ซึ่งใช้วิธีการประมาณค่าพารามิเตอร์ของเบส์ (Bayesian Estimation) โดยมีวิธีการในการประมาณค่าโดยสามารถจำแนกกระบวนการดำเนินการตามแนวคิดของเบส์ออกเป็น 2 กระบวนการหลักๆ คือ

1. กระบวนการกำหนดลักษณะของการแจกแจงเริ่มแรก (Prior distribution) ประกอบด้วย 2 ขั้นตอน ได้แก่

1.1 กำหนดให้การแจกแจงเริ่มแรกของค่าพารามิเตอร์ความสามารถที่แท้จริงของผู้สอบ (θ) และค่าพารามิเตอร์ความยากของข้อสอบ (b) ให้เป็นอิสระต่อกัน ดังนี้

$$f(\theta,b) = f(\theta) \,.\, f(b)$$

โดยกำหนดลักษณะการแจกแจงของ $f(\theta)$ และ f (b) ไว้ดังนี้

1.1.1 การแจกแจงเริ่มแรกของค่าความสามารถของผู้สอบ ($f(\theta)$) มีข้อตกลงว่าข้อสารสนเทศที่มีมาก่อนของค่าความสามารถของผู้เข้าสอบแต่ละคนไม่แตกต่างกันสามารถใช้แทนกันได้ (exchangeability) และกำหนดให้ค่าความสามารถของผู้สอบเป็นตัวแปรสุ่มที่มีลักษณะการแจกแจงแบบปกติ (normal distribution)

$$f(\theta_i|\mu_\theta,\sigma_\theta^2) = N(\mu_\theta,\sigma_\theta^2)$$

เมื่อ $N(\mu_\theta,\sigma_\theta^2)$ = การแจกแจงปกติ (normal distribution) ที่มีค่าเฉลี่ยเท่ากับ μ_θ และค่าความแปรปรวนเท่ากับ σ_θ^2

1.1.2 การแจกแจงเริ่มแรกของค่าความยากของข้อสอบ f (b) อาจใช้กระบวนการเดียวกับการกำหนดการแจกแจงเริ่มแรกของค่าความสามารถของผู้สอบ (θ) คือ มีข้อตกลงว่า f (b) มีการแจกแจงแบบปกติ หรืออาจจะไม่กำหนดการแจกแจงเริ่มแรกไว้ก็ได้

1.2 กำหนดค่าที่เป็นตัวเลขของพารามิเตอร์สำหรับการแจกแจงเริ่มแรก

1.2.1 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความสามารถ θ ได้แก่ μ_θ และ σ_θ^2 อาจกำหนดให้ $\mu_\theta = 0$ และ $\sigma_\theta^2 = 1$ ซึ่งจะทำให้การประมาณค่าความสามารถ θ มีความสะดวก และประมาณค่าได้รวดเร็วขึ้น

1.2.2 พารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความยาก (b) หากมีการกำหนดลักษณะการแจกแจงไว้ ก็ใช้ค่าเดียวกับพารามิเตอร์ของการแจกแจงเริ่มแรกของค่าความสามารถ θ ดังกล่าวแล้วข้างต้น

2. กระบวนการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้เข้าสอบ

การประมาณค่าพารามิเตอร์ด้วยวิธีของเบส์ คือ การหาค่าประมาณ $\theta_i (j=1,2,...,N)$ และ b_j $(i=1, 2,...,n)$ ที่ทำให้ฟังก์ชันการแจกแจงภายหลัง f$(\theta,b|u)$ มีค่าสูงสุด ถ้า ln f$(\theta,b|u)$ เป็นฟังก์ชันที่หาอนุพันธ์ได้ การหาค่า θ_i, b_j จะหาได้จากการอนุพันธ์ของ ln f$(\theta,b|u)$ และกำหนดให้อนุพันธ์ของ ln f$(\theta,b|u)$ มีค่าเท่ากับศูนย์ แล้วจึงหาค่ารากของอนุพันธ์ของ ln f$(\theta,b|u)$ ซึ่งแสดงได้ดังสมการต่อไปนี้

$$\mathrm{f}(\theta,b|u) = L(u|\theta,b)\ \mathrm{f}(\theta)\ \mathrm{f(b)} / \mathrm{f(u)}$$

$$\ln \mathrm{f}(\theta,b|u) = \ln L(u|\theta,b) + \ln \mathrm{f}(\theta) + \ln \mathrm{f(b)} + \text{Constant}$$

$$d \ln \mathrm{f}(\theta,b|u) / d\,\theta_i = 0$$

$$d \ln \mathrm{f}(\theta,b|u) / d\,b_j = 0$$

สมการอนุพันธ์ของ ln [f$(\theta,b|u) = 0$] นี้ เรียกว่า สมการโมดัล (Modal Equation) ค่ารากของสมการโมดัล คือ ค่าความสามารถที่แท้จริงของผู้สอบ (θ) และค่าพารามิเตอร์ของข้อสอบ (b) ที่ทำให้ฟังก์ชันการแจกแจงภายหลังมีค่าสูงสุด อาจทำได้โดยใช้เทคนิคนิวตันราฟสัน (Newton-Raphson) ซึ่งเป็นการหาค่าประมาณโดยการหาค่าซ้ำ (Iterative) มีขั้นตอนดังต่อไปนี้

ขั้นที่ 1 กำหนดค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถθ_i และค่าพารามิเตอร์ ของข้อสอบ b_j

1.1 ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ θ_i

$$\theta_i(0) = \ln(X_i / (n - X_i))$$

เมื่อ ln = Natural Logarithm

X_i = คะแนนรวมของผู้เข้าสอบคนที่ i

n = จำนวนข้อสอบ

1.2 ค่าเริ่มต้นสำหรับใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ b_j

$$b_j^{(0)} = Z_j / R_j$$

เมื่อ R_j = สัมประสิทธิ์สหสัมพันธ์แบบ Point – Biserial ระหว่างคะแนนรายข้อ (1 = ตอบถูก 0 = ตอบผิด กับคะแนนรวมของแบบสอบทั้งฉบับ)

Z_j = ค่า Z-Score ของการแจกแจงปกติมาตรฐานที่พื้นที่ใต้โค้งปกติมาตรฐานด้านขวามือมีค่าเท่ากับ P_j

P_j = U_j / N

N = จำนวนผู้เข้าสอบทั้งหมด

<u>ขั้นที่ 2</u> ประมาณค่าความสามารถของผู้เข้าสอบแต่ละคน θ_i โดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าได้ในครั้งก่อนเป็นค่าคงที่

$$\theta_i^{m+1} = \theta_i^{(m)} - g\left(\theta_i^{(m)}\right) / h\left(\theta_i^{m}\right)$$

เมื่อ $\theta_i^{m+i}, \theta_i^{(m)}$ = ค่าประมาณความสามารถของคนที่ i ครั้งที่ m + 1และ m

$$g\left(\theta_i^{(m)}\right) = d \ln f(\theta, b/u) d\theta_i$$

$$h\left(\theta_i^{(m)}\right) = d^2 \ln f(\theta, b/u) d\theta_i^2$$

การประมาณค่าซ้ำ (Iterative) จะกระทำจนกว่าค่าประมาณความสามารถ θ_i จะลู่เข้าสู่ค่าคงที่ค่าใดค่าหนึ่ง (Convergence) คือ ค่าประมาณครั้งที่ m + 1 และครั้งที่ m มีค่าแตกต่างกันน้อยกว่าค่าคงที่ กรณีการวิเคราะห์ครั้งนี้กำหนดไว้ที่ 0.000001

<u>ขั้นที่ 3</u> ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ b_j โดยกำหนดให้ค่าความสามารถของผู้เข้าสอบที่ประมาณค่าได้ในครั้งก่อนเป็นค่าคงที่

$$b_j^{m+1} = b_j^{(m)} - g\left(b_j^{(m)}\right) / h\left(b_j^{(m)}\right)$$

เมื่อ $b_j^m, b_j^{(m+1)}$ = ค่าประมาณความยากของข้อสอบข้อที่ j ครั้งที่ m และ m + 1

$g(b_j^{(m)})$ = $d \ln f(\theta, b/u) / d\ b_j$

$h(b_j^{(m)})$ = $d^2 \ln f(\theta, b/u) / d\ b_j^2$

การประมาณค่าพารามิเตอร์จะดำเนินการตามขั้นที่หนึ่งและขั้นที่สอง ซ้ำๆ (Iterative) จนกว่าค่าประมาณ จะลู่เข้าสู่ค่าคงที่ค่าใดค่าหนึ่ง (Convergence) สำหรับค่าของอนุพันธ์อันดับ

1 g (.) และอนุพันธ์อันดับ 2 h (.) มีส่วนประกอบสองส่วน คือ ส่วนที่ได้จากฟังก์ชันไลค์ลิฮู้ด และส่วนที่ได้จากการแจกแจงเริ่มแรก ซึ่งส่วนประกอบทั้งสองส่วนนี้แสดงไว้ในตารางที่ 10

ตารางที่ 10 อนุพันธ์อันดับที่ 1 และ อนุพันธ์อันดับที่ 2 ของ $\ln f(\theta, b/u)$ กรณี 1 PL IRT

พารามิเตอร์	ฟังก์ชัน	อนุพันธ์อันดับที่ 1	อนุพันธ์อันดับที่ 2
θ_i	Likelihood	$\sum_{j=1}^{n}(U_{ij} - P_{ij})$	$-\sum_{j=1}^{n} (P_{ij})(1 - P_{ij})$
	Prior		
	Normal (0, 1)	$-\theta_i$	-1
	Gamma (1, 1)	$(a-1)\theta_i - (1/b)$	$(a-1)$
β_j	Likelihood	$-\sum_{i=1}^{N}(U_{ij} - P_{ij})$	$-\sum_{i=1}^{N} (P_{ij})(1 - P_{ij})$
	Prior		
	Normal (0, 1)	$-(b_j - b.)^2 / \sigma_b^2$	$-[\sigma_b^2(1-(1/n)) - \{(2(b_j - b.)^2)/(\nu_b + n - 1)$
	Beta (0.5,0.5)	$(a-1)(1/\beta_j) - (b-1)(1/(1-\beta_j)$	$-(a-1)(\beta^2) + (b-1)(1/(1-\beta)^2$

$$P_{ij} = \exp(\theta_p - \beta_j)/1 + \exp(\theta_j - \beta_j)$$

$$b. = \sum b_j / n_j$$

$$\sigma_b^2 = [\lambda_b + \sum_{j=1}^{n}(b_j - b.)^2]/(\nu_b + n - 1)$$

$\lambda_b = 10$, $\nu_b = 10$

N = จำนวนผู้เข้าสอบ

n = จำนวนข้อสอบ

การประมาณในขั้นตอนที่ 1 จะทำให้ค่าสารสนเทศจากการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (1PL) ซึ่งได้แก่ ค่าความสามารถที่แท้จริงของผู้สอบแต่ละคน (θ_p) และ ค่าความยากของข้อสอบแต่ละข้อ (β_i)

ขั้นตอนที่ 2 ขั้นการวิเคราะห์ค่าโอกาสในการทำข้อสอบในแต่ละข้อของแต่ละคน เพื่อใช้สำหรับการวิเคราะห์ในขั้นตอนต่อไป

นำค่าการแจกแจงภายหลัง (Posterior distribution) ที่ได้ในขั้นตอนที่หนึ่งซึ่งได้แก่ ค่า (θ_p) และ (β_i) มาคำนวนค่า E (X_{pi}) หรือ $P(\theta_p, \beta_i)$ สำหรับโมเดลการวิเคราะห์แบบp x i ด้วยข้อคำถามแบบสองค่า (dichotomous items) สามารถแสดงสมการแบบ one parameter logistic ได้ดังนี้

$$P(\theta_p, \beta_i) = P(X_{pi} = 1 \mid \theta_p, \beta_i) = \frac{\exp(\theta_p - \beta_i)}{1 + \exp(\theta_p - \beta_i)} \quad \text{(Item Response Theory : IRT)}$$

จากการประมาณค่าดังกล่าวทำให้สามารถสร้างExpected response matrix ได้ดังนี้

พารามิเตอร์ของข้อสอบ

		β_1	…	β_i	…	β_I
	θ_1	E (X_{11})		…		E (X_{1I})
พารามิเตอร์ของผู้สอบ	…					
	θ_p	…		E (X_{pi})		….
	…					
	θ_P	E (X_{P1})		…		E (X_{PI})

ซึ่งเราจะใช้ค่า E (X_{pi}) หรือก็คือค่า $P(\theta_p, \beta_i)$ ที่ได้จากการวิเคราะห์มาใช้ในการคำนวณในขั้นตอนต่อไป

ขั้นตอนที่ 3 การวิเคราะห์ตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

จุดมุ่งหมายของการวิเคราะห์ในขั้นตอนนี้เพื่อให้ได้สารเทศตามแนวการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) นั้นคือการประมาณองค์ประกอบความแปรปรวน (Variance components) และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability coefficient) สามารถประมาณได้ ด้วยกรอบแนวคิดของทฤษฎีการตอบสนองข้อสอบ (IRT) ที่ได้จากการเชื่อมโยงสูตรของ Kolen และ Harris (1987) ดังนี้

$$\mu = \int_{\theta}\int_{\beta} P(\theta_p, \beta_i) f(\beta) g(\theta) d\beta d\theta$$ **(คำอธิบาย μ ใน GT)**

$$\pi(\theta) = \int_{\beta} P(\theta_p, \beta_i) f(\beta) d\beta - \mu$$ **(คำอธิบาย ν_p ใน GT)**

$$\iota(\theta) = \int_{\theta} P(\theta_p, \beta_i) g(\theta) d\theta - \mu$$ **(คำอธิบาย ν_i ใน GT)**

$$\nu(\theta, \beta) = P(\theta_p, \beta_i) - \pi(\theta) - \iota(\beta) - \mu$$ **(คำอธิบาย ν_{pi} ใน GT)**

$$\sigma^2(e) = \int_{\theta}\int_{\beta} P(\theta_p, \beta_i)[1 - P(\theta_P, \beta_i)] f(\beta) g(\theta) d\beta d\theta$$ **(คำอธิบาย ν_e ในGT)**

สามารถเขียนในรูปของความแปรปรวนทั้งหมดของคะแนนที่สังเกตได้ดังนี้

$$\sigma^2(\mathrm{X}_{pi}) = \sigma^2(\pi) + \sigma^2(i) + \sigma^2(\nu) + \sigma^2(e)$$

จากสูตรดังกล่าวจะทำให้สามารถประมาณองค์ประกอบความแปรปรวน (Variance component) และค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของการวัด (Generalizability coefficient) ได้ ซึ่งจะให้สารสนเทศตามแนวการวิเคราะห์ด้วยทฤษฏีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT)

การเปรียบเทียบวิธีการวิเคราะห์ 4 รูปแบบของวิธีการ GIRM

เมื่อพิจารณาวิธีการวิเคราะห์ของวิธีการ GIRM ทั้งสามรูปแบบ ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM เป็นรูปแบบต้นฉบับในการวิเคราะห์ด้วยวิธีการ GIRM ที่ Briggs และ Wilson พัฒนาขึ้นในปี 2007 รูปแบบที่ 2 Alternative GIRM A เป็นรูปแบบที่ผู้วิจัยพัฒนาขึ้น ซึ่งมีวิธีการวิเคราะห์ที่คล้ายคลึงกับการวิเคราะห์ในรูปแบบที่ 1 Original GIRM โดยต่างกันในขั้นตอนที่ 2 การวิเคราะห์ในรูปแบบ Alternative GIRM A จะนำค่าสารสนเทศที่ได้จากขั้นตอนที่ 1 มาวิเคราะห์เป็นคะแนน การแจกแจงการทำนายภายหลัง (Posterior predictive distribution) ซึ่งเป็นคะแนนเฉลี่ยของคะแนนที่สังเกตได้ที่ได้มาจากการทำนาย นำมาสร้างเป็น Predictive Observed Matrix จากการทำนายก่อนนำไปวิเคราะห์ในขั้นตอนที่ 3 ในขณะที่การวิเคราะห์ในรูปแบบที่ 1

Original GIRM จะนำค่าการแจกแจงภายหลัง (Posterior distribution) ที่ได้ในขั้นตอนที่หนึ่งซึ่งได้แก่ค่า (θ_p) และ (β_i) มาคำนวนค่า E (X_{pi}) หรือ $P(\theta_p, \beta_i)$ แล้วนำมาสร้างเป็นเมทริกซ์การตอบสนองที่คาดหวัง (Expected response matrix) แล้วจึงนำไปวิเคราะห์ในขั้นตอนที่ 3 ต่อไป ในขั้นตอนนี้สามารถสรุปเป็นสัญลักษณ์ทางคณิตศาสตร์ของความแตกต่างในการวิเคราะห์ระหว่างรูปแบบวิธีที่ 1 กับวิธีที่ 2 ได้ดังนี้ การวิเคราะห์ในรูปแบบที่ 1 จะคำนวณสารสนเทศของ GT ด้วย $E\left[X_{pi} \mid \theta_p^{(m)}, \beta_i^{(m)}\right]$ ในขณะที่การวิเคราะห์ในรูปแบบที่ 2 จะคำนวณสารสนเทศของ GT ด้วย $X_{pi}^{(m)} \sim Rasch(\theta_p, \beta_i)$ ส่วนการวิเคราะห์ในรูปแบบที่ 3 Alternative GIRM B มีลักษณะคล้ายคลึงกับวิธีการวิเคราะห์ในรูปแบบที่ 2 โดยแตกต่างกันในขั้นตอนที่ 3 โดยการวิเคราะห์ Alternative GIRM A จะใช้สูตรการประมาณค่าองค์ประกอบความแปรปรวนของ Kolen และ Harris (1987) ส่วนวิธีการวิเคราะห์ในรูปแบบที่ 3 จะใช้การวิเคราะห์ความแปรปรวน (ANOVA) ประมาณค่าองค์ประกอบความแปรปรวน

รูปแบบสุดท้ายในการศึกษาครั้งนี้ คือ รูปแบบที่ 4 Numerical Bayesian GIRM เป็นรูปแบบที่มีวิธีการวิเคราะห์ที่คล้ายคลึงกับการวิเคราะห์ในรูปแบบที่ 1 Original GIRM โดยต่างกันในขั้นตอนที่ 1 โดยการวิเคราะห์ในรูปแบบที่ 4 Numerical Bayesian GIRM ใช้การวิเคราะห์ด้วยวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) ด้วยการหาอนุพันธ์ของฟังก์ชันและให้มีค่าเท่ากับศูนย์หลังจากนั้นจึงหารากของอนุพันธ์นั้น ในขณะที่การวิเคราะห์ในรูปแบบที่ 1 Original GIRM ใช้การวิเคราะห์ด้วยวิธีการประมาณค่าแบบเบส์ (Bayesian Estimation) แต่ใช้วิธีการจำลองข้อมูลแบบห่วงโซ่ (Markov Chain Monte Carlo: MCMC) แทนการหาค่าอนุพันธ์ ส่วนการวิเคราะห์ในขั้นตอนที่ 2 และขั้นตอนที่ 3 จะมีรูปแบบ การวิเคราะห์ที่เหมือนกันทั้งในรูปแบบที่ 1 Original GIRM และรูปแบบที่ 4 Numerical Bayesian GIRM การเปรียบเทียบทั้งสามวิธีและเปรียบเทียบกับทฤษฎีการวิเคราะห์การตอบสนองข้อสอบ (IRT) และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) สามารถเปรียบเทียบได้ดังแสดงในตารางที่ 11

ตารางที่ 11 การเปรียบเทียบลักษณะของรูปแบบในการวิเคราะห์ข้อสอบ

ลักษณะ	รูปแบบของวิธี GIRM			Traditional IRT	Traditional GT
	original GIRM	Alternative GIRM	Original Bayesian GIRM		
1. **ขั้นตอนที่** 1 สารสนเทศ IRT	การวิเคราะห์ MCMC ด้วยการหาค่า Posterior disribution	การวิเคราะห์ MCMC ด้วยการหาค่า Posterior disribution	การวิเคราะห์ด้วยวิธีการประมาณค่าแบบเบส์ กำหนดให้อนุพันธ์เท่ากับ 0	ใช้สมการ Logistic function	*
2. **ขั้นตอนที่** 2 2.1) การเตรียมข้อมูลในการวิเคราะห์ GT	ผลจากการวิเคราะห์ด้วยการใช้ฟังก์ชัน 1 PL IRT	ผลจากการวิเคราะห์ MCMC ด้วยการหาค่า Posterior predictive distribution	ผลจากการวิเคราะห์ด้วย การใช้ฟังก์ชัน 1 PL IRT	*	นำคะแนนดิบซึ่งเป็นตัวแปรสังเกตได้มาใช้ในการคำนวณโดยตรง
2.2) ลักษณะของข้อมูลก่อนการวิเคราะห์ GT	$E(X_{pi})$	$X^{(m)}_{pi}$	$E(X_{pi})$	$E(X_{pi})$	X_{pi}
3. **ขั้นตอนที่** 3 สารสนเทศ GT	สูตร Kolen และ Harris	AGIRM A การวิเคราะห์ ANOVA AGIRM B สูตรKolen และ Harris	สูตร Kolen และ Harris	*	ANOVA
4. การประมาณค่าพารามิเตอร์	Bayesian Estimation	Bayesian Estimation	Bayesian Estimation	Maximum Likehood Estimation**	*
5. การวิเคราะห์ตามแนว MCMC	ใช่	ใช่	ไม่ใช่	ใช่/ไม่ใช่***	ใช่/ไม่ใช่***

ตารางที่ 11 (ต่อ)

ลักษณะ	รูปแบบของวิธี GIRM			Traditional IRT	Traditional GT
	Classical GIRM	Alternative GIRM	Original Bayesian GIRM		
6. โปรแกรมการวิเคราะห์	Package R2 WinBUGS	Package R2 WinBUGS	R Program	IRT (BAY) BILOG TESTFACT เป็นต้น	GENOVA

* ไม่เกี่ยวข้องกับการเปรียบเทียบในส่วนนี้

** การประมาณค่าพารามิเตอร์ส่วนใหญ่จะพบได้ใน 3 วิธี วิธีฮิวริสติก วิธีแมกซิมัมไลค์ลิฮู้ด และวิธีการของเบส์ แต่ที่นิยมใช้กันมากที่สุด คือ วิธีแมกซิมัมไลค์ลิฮู้ด

*** ใช่หรือไม่ใช่ก็ได้ขึ้นอยู่กับการเลือกวิธีการประมาณค่าของผู้วิเคราะห์

3.2 ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ(θ)

ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) ผู้วิจัยได้ศึกษา จำนวน 2 ลักษณะ ได้แก่ ลักษณะที่ 1 การแจกแจงแบบปกติ (Normal) ซึ่งเป็นลักษณะการแจกแจงตามสมมติฐานของวิธีการ GIRM และลักษณะที่ 2 การแจกแจงแบบแกมม่า (Gamma) เนื่องจากการกำหนดลักษณะการแจกแจงแบบแกมมา (Gamma) จะกำหนดเมื่อค่าส่วนใหญ่มีค่าเป็นโค้งและโอกาสในการเกิดค่าพารามิเตอร์จะมีลักษณะเป็นเส้นโค้ง (Gelman,A., Carlin,J.B., Stern,H.S. และ Rubin, D.B., 1995)

3.3 ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ(β)

ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) ผู้วิจัยได้ศึกษา จำนวน 2 ลักษณะ ได้แก่ ลักษณะที่ 1 การแจกแจงแบบปกติ (Normal) ซึ่งเป็นลักษณะการแจกแจงตามสมมติฐานของวิธีการ GIRM และลักษณะที่ 2 การแจกแจงแบบเบต้า (Beta) เนื่องจากการกำหนดลักษณะการแจกแจงแบบเบต้า (Beta) จะกำหนดเมื่อค่าพารามิเตอร์ที่ต้องการประมาณค่ามีค่าอยู่ระหว่าง 0 ถึง 1 (Gelman, Carlin, Stern และ Rubin, 1995)

ดังนั้นในเงื่อนไขของการวิเคราะห์ข้อมูลในการศึกษาครั้งนี้จึงประกอบไปด้วยตัวแปร จำนวน 3 ตัวแปร ได้แก่

รูปแบบการวิเคราะห์ข้อสอบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) 4 รูปแบบ ประกอบด้วย

รูปแบบที่ 1 Original GIRM

รูปแบบที่ 2 Alternative GIRM A

รูปแบบที่ 3 Alternative GIRM B

รูปแบบที่ 4 Numerical Bayesian GIRM

ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) 2 ลักษณะ ประกอบด้วย

ลักษณะที่ 1 การแจกแจงแบบปกติ (Normal)

ลักษณะที่ 2 การแจกแจงแบบแกมมา (Gamma)

ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) 2 ลักษณะ ประกอบด้วย

ลักษณะที่ 1 การแจกแจงแบบปกติ (Normal)

ลักษณะที่ 2 การแจกแจงแบบเบต้า (Beta)

4. การวัดประสิทธิภาพของการประมาณค่า

การวิจัยครั้งนี้ศึกษาวิธีการประมาณค่าพารามิเตอร์ของข้อสอบโดยมีดัชนี้ชี้วัดดังต่อไปนี้

4.1 ความลำเอียง (Biased estimator) ในการประมาณค่าพารามิเตอร์วัดได้จากค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean Square Error; MSE)

$$MSE(\theta) = \frac{\sum_{i=1}^{r} (\theta_i - \hat{\theta}_i)^2}{r}$$

4.2 ความไม่แน่นอนในการประมาณค่า (uncertainty estimator) ของค่าพารามิเตอร์ วัดได้จากส่วนเบี่ยงเบนมาตรฐานขององค์ประกอบความแปรปรวนในแต่ละองค์ประกอบของคะแนนสังเกตได้จากการวิเคราะห์ด้วยวิธีสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

4.3 ประสิทธิภาพการในประมาณค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean distance) ซึ่งเป็นดัชนีที่ใช้เปรียบเทียบวิธีการประมาณค่าองค์ประกอบความแปรปรวนโดยพิจารณาการเปรียบเทียบขนาดเวกเตอร์ขององค์ประกอบความแปรปรวนที่ศึกษา

กำหนดให้

$\widetilde{\theta}$ แทนเวกเตอร์ค่าพารามิเตอร์ขององค์ประกอบความแปรปรวน

$\hat{\theta}_c$ แทนเวกเตอร์ค่าพารามิเตอร์ขององค์ประกอบความแปรปรวนรูปแบบที่ 1

$\hat{\theta}_a$ แทนเวกเตอร์ค่าพารามิเตอร์ขององค์ประกอบความแปรปรวนรูปแบบที่ 2

$\hat{\theta}_b$ แทนเวกเตอร์ค่าพารามิเตอร์ขององค์ประกอบความแปรปรวนรูปแบบที่ 3

$\hat{\theta}_o$ แทนเวกเตอร์ค่าพารามิเตอร์ขององค์ประกอบความแปรปรวนรูปแบบที่ 4

ซึ่ง $\widetilde{\theta} = \begin{pmatrix} \sigma^2_\alpha \\ \sigma^2_\beta \\ \sigma^2_{\alpha\beta} \\ \sigma^2_e \end{pmatrix}$ $\hat{\theta}_c = \begin{pmatrix} \hat{\sigma}^2_{\alpha c} \\ \hat{\sigma}^2_{\beta c} \\ \hat{\sigma}^2_{\alpha\beta c} \\ \hat{\sigma}^2_{ec} \end{pmatrix}$ $\hat{\theta}_a = \begin{pmatrix} \hat{\sigma}^2_{\alpha a} \\ \hat{\sigma}^2_{\beta a} \\ \hat{\sigma}^2_{\alpha\beta a} \\ \hat{\sigma}^2_{ea} \end{pmatrix}$ $\hat{\theta}_a = \begin{pmatrix} \hat{\sigma}^2_{\alpha b} \\ \hat{\sigma}^2_{\beta b} \\ \hat{\sigma}^2_{\alpha\beta b} \\ \hat{\sigma}^2_{eb} \end{pmatrix}$ $\hat{\theta}_o = \begin{pmatrix} \hat{\sigma}^2_{\alpha o} \\ \hat{\sigma}^2_{\beta o} \\ \hat{\sigma}^2_{\alpha\beta o} \\ \hat{\sigma}^2_{eo} \end{pmatrix}$

รูปแบบที่ 1 Origianal GIRM

$$\text{EuC} = \frac{\sum_{i=1}^{m}\left\|\hat{\theta}_c - \widetilde{\theta}\right\|}{N} = \frac{\sum_{i=1}^{m}\sqrt{(\hat{\sigma}^2_{\alpha c} - \sigma^2_\alpha) + (\hat{\sigma}^2_{\beta c} - \sigma^2_\beta) + (\hat{\sigma}^2_{\alpha\beta c} - \sigma^2_{\alpha\beta}) + (\hat{\sigma}^2_{ec} - \sigma^2_e)}}{N}$$

รูปแบบที่ 2 Alternative GIRM A

$$\text{EuA} = \frac{\sum_{i=1}^{m}\left\|\hat{\theta}_a - \widetilde{\theta}\right\|}{N} = \frac{\sum_{i=1}^{m}\sqrt{(\hat{\sigma}^2_{\alpha a} - \sigma^2_\alpha) + (\hat{\sigma}^2_{\beta a} - \sigma^2_\beta) + (\hat{\sigma}_{\alpha\beta a} - \sigma^2_{\alpha\beta}) + (\hat{\sigma}^2_{ea} - \sigma^2_e)}}{N}$$

รูปแบบที่ 3 Alternative GIRM B

$$\text{EuB} = \frac{\sum_{i=1}^{m}\left\|\hat{\theta}_b - \widetilde{\theta}\right\|}{N} = \frac{\sum_{i=1}^{m}\sqrt{(\hat{\sigma}^2_{\alpha b} - \sigma^2_\alpha) + (\hat{\sigma}^2_{\beta b} - \sigma^2_\beta) + (\hat{\sigma}_{\alpha\beta b} - \sigma^2_{\alpha\beta}) + (\hat{\sigma}^2_{eb} - \sigma^2_e)}}{N}$$

รูปแบบที่ 4 Numerical Bayesian GIRM

$$\text{EuO} = \frac{\sum_{i=1}^{m}\left\|\hat{\theta}_O - \widetilde{\theta}\right\|}{N} = \frac{\sum_{i=1}^{m}\sqrt{(\hat{\sigma}^2_{\alpha o} - \sigma^2_\alpha) + (\hat{\sigma}^2_{\beta o} - \sigma^2_\beta) + (\hat{\sigma}_{\alpha\beta o} - \sigma^2_{\alpha\beta}) + (\hat{\sigma}^2_{eo} - \sigma^2_e)}}{N}$$

N = จำนวนรอบที่ทำให้ค่าประมาณองค์ประกอบความแปรปรวนทั้ง 3 แบบ
มีค่าลู่เข้าแบบคงที่

จากขั้นตอนการจำลองเงื่อนไขของข้อมูล เงื่อนไขของการวิเคราะห์ข้อมูล และการวัดประสิทธิภาพของการประมาณค่า สามารถสรุปเป็นแผนภาพได้ดังนี้

1. การจำลองข้อมูลด้วยโปรแกรม R

1.1 จำนวนกลุ่มตัวอย่างในการจำลอง 2 เงื่อนไข (100 300 คน)

1.2 จำนวนข้อสอบในการจำลอง 2 เงื่อนไข (5 10ข้อ)

รวมเงื่อนไขข้อมูลทั้งสิน 4 เงื่อนไข

↓

2. เงื่อนไขการวิเคราะห์ข้อมูลด้วยโปรแกรม WinBUGS

2.1 รูปแบบในการประมาณค่าพารามิเตอร์ 4 รูปแบบ (original GIRM Alternative GIRM A Alternative GIRM B และ Numerical Bayesian GIRM)

2.2 ลักษณะการแจกแจงเริ่มแรกของพารามิเตอร์ของผู้สอบ 2 รูปแบบ (Normal และUniform)

2.3 ลักษณะการแจกแจงเริ่มแรกของพารามิเตอร์ของข้อสอบ 2 รูปแบบ (Normal และBeta)

รวมเงื่อนไขที่ต้องวิเคราะห์ทั้งสิ้น 16 **เงื่อนไข**

↓

3. การวิเคราะห์ข้อมูล

จากเงื่อนไขของข้อมูล 4 เงื่อนไข และ เงื่อนไขการวิเคราะห์ 16 เงื่อนไข

ดังนั้นจะต้องทำการวิเคราะห์ข้อมูลทั้งสิ้น 64 ชุด

↓

4. การวัดประสิทธิภาพของการประมาณค่า

1) ความลำเอียงในการประมาณค่า (bias estimator)

2) ความไม่แน่นอนในการประมาณค่า (uncertainty estimator)

3) ประสิทธิภาพในประมาณค่าองค์ประกอบความแปรปรวน

แผนภาพที่ 21 ขั้นตอนการจำลองเงื่อนไขของข้อมูล การวิเคราะห์ข้อมูล และการวัดประสิทธิภาพของการประมาณค่า

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มุ่งเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทั้ง 4 วิธี ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM ซึ่ง พัฒนาโดย Brigg และ Wilson (2007) กับรูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B รูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งผู้วิจัยเป็นผู้พัฒนาขึ้น นอกจากนี้ยังศึกษาถึงอิทธิพลของขนาดกลุ่มอย่างและจำนวนข้อสอบที่มีต่อประสิทธิภาพของวิธีวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) รวมทั้งยังศึกษาถึงความไว (Sensitivity) ของผลการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) ดังนั้นผู้วิจัยจึงได้นำเสนอในส่วนนี้แบ่งออก เป็น 6 ตอน ซึ่งมีรายละเอียดในแต่ละตอนดังต่อไปนี้

ตอนที่ 1 การวิเคราะห์ค่าพารามิเตอร์ของผู้สอบและข้อสอบตามแนวทฤษฎีการตอบสนองของข้อสอบ

ตอนที่ 2 การวิเคราะห์องค์ประกอบความแปรปรวนตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ตอนที่ 3 การวิเคราะห์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าในแต่ละรูปแบบการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM)

ตอนที่ 4 การวิเคราะห์ความแตกต่างของขนาดกลุ่มตัวอย่างและความยาวของแบบสอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์

ตอนที่ 5 การวิเคราะห์ความแตกต่างของการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบและค่าพารามิเตอร์ของข้อสอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์

ตอนที่ 1 ผลการวิเคราะห์ค่าพารามิเตอร์ของผู้สอบและข้อสอบตามแนวทฤษฎีการตอบสนองของข้อสอบ

วิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (Generalizability in Item Response Modeling: GIRM) เป็นการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการกำหนดมาตรวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) ดังนั้นสารสนเทศที่ได้จากการวิเคราะห์จึงให้สารสนเทศทั้งสารสนเทศที่ได้จากการวิเคราะห์จากทฤษฎีการตอบสนองข้อสอบ (IRT) และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยในตอนนี้นำเสนอสารสนเทศเกี่ยวกับค่าพารามิเตอร์ของผู้สอบและข้อสอบตามแนวทฤษฎีการตอบสนองของข้อสอบ (IRT)

นอกจากนี้ในการนำเสนอตอนนี้ยังได้มีการวิเคราะห์ความแตกต่างของค่าพารามิเตอร์ของผู้สอบและข้อสอบระหว่างวิธีการประมาณค่าพารามิเตอร์จากวิธีการ MCMC ซึ่งในวิเคราะห์ค่าพารามิเตอร์ของผู้สอบและข้อสอบในวิธีการประมาณค่าพารามิเตอร์ผู้สอบและข้อสอบในรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B โดยเปรียบเทียบกับวิธีการประมาณค่าพารามิเตอร์จากวิธีการ Numerical Method ซึ่งใช้ในการวิเคราะห์ค่าพารามิเตอร์ของผู้สอบและข้อสอบในรูปแบบที่ 4 Numerical Bayesian GIRM โดยมีรายละเอียดของผลการวิเคราะห์ในตอนนี้ ดังนี้

1.1 การวิเคราะห์ค่าพารามิเตอร์ของผู้สอบตามแนวทฤษฎีการตอบสนองของข้อสอบ

การวิเคราะห์ค่าพารามิเตอร์ของผู้สอบตามแนวทฤษฎีการตอบสนองข้อสอบทำการวิเคราะห์เปรียบเทียบระหว่างวิธีการประมาณค่าพารามิเตอร์แบบ MCMC ซึ่งใช้ในวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของของโมเดลการตอบสนองข้อสอบ (GIRM) ในรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และ รูปแบบที่ 3 AGIRM B เปรียบเทียบกับ วิธีการประมาณค่าพารามิเตอร์แบบ Numerical Method ซึ่งใช้ในวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ในรูปแบบที่ 4 Numerical Bayesian GIRM โดยวิเคราะห์เปรียบเทียบในแต่ละกรณีศึกษาทั้ง 16 กรณีศึกษา มีผลการศึกษาดังนี้

ตารางที่ 12 ผลการวิเคราะห์ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐานและการวิเคราะห์ความแตกต่างด้วยสถิติการทดสอบที ของค่าพารามิเตอร์ข้อสอบระหว่างวิธีการ GIRM ที่ประมาณค่าของวิธี MCMC และวิธี Numerical จำแนกตามเงื่อนไขของการวิเคราะห์

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	GIRM (MCMC)		GIRM (Numerical)		t-test	p-value
				Mean	S.D.	Mean	S.D.		
Normal	Normal	300	5	-0.0358	0.0103	-0.0344	0.0102	-0.2169	0.8337
		300	10	0.0059	0.0084	0.0018	0.0081	1.1033	0.2844
		100	5	-0.0222	0.0113	-0.0255	0.0103	0.4797	0.6444
		100	10	-0.0107	0.0113	-0.0074	0.0108	-0.0074	0.7663
	Beta	300	5	0.1651	0.0048	NA	NA	NA	NA
		300	10	0.1786	0.0062	NA	NA	NA	NA
		100	5	0.2343	0.0081	NA	NA	NA	NA
		100	10	0.2743	0.0053	NA	NA	NA	NA
Gamma	Normal	300	5	0.8454	0.0105	NA	NA	NA	NA
		300	10	0.8901	0.0085	NA	NA	NA	NA
		100	5	0.8232	0.0109	NA	NA	NA	NA
		100	10	0.8265	0.0109	NA	NA	NA	NA
	Beta	300	5	0.7529	0.0098	NA	NA	NA	NA
		300	10	0.7954	0.0059	NA	NA	NA	NA
		100	5	0.7121	0.0029	NA	NA	NA	NA
		100	10	0.7073	0.0088	NA	NA	NA	NA

NA = ไม่สามารถทำการวิเคราะห์ได้

จากผลการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ พบว่า การวิเคราะห์ค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ MCMC ในรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B ให้ค่าเฉลี่ยของพารามิเตอร์ของข้อสอบอยู่ระหว่าง -0.0358 ถึง 0.8901 และให้ค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0029 ถึง 0.0113 ส่วนการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ Numerical Method ในรูปแบบที่ 4 Numerical Bayesian GIRM สามารถวิเคราะห์ได้เฉพาะกรณีลักษณะการแจกแจงเริ่มแรกผู้สอบและลักษณะการแจกแจงเริ่มแรกของข้อสอบมีลักษณะการกระจายแบบโค้งปกติเท่านั้น โดยให้

ค่าเฉลี่ยของพารามิเตอร์ของข้อสอบอยู่ระหว่าง -0.0344 ถึง 0.0018 และให้ค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0081 ถึง 0.0108

เมื่อพิจารณาเปรียบเทียบเป็นรายกรณีของค่าส่วนเบี่ยงเบนมาตรฐาน (S.D) ระหว่างการประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ MCMC กับ การประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ Numerical Method พบว่า การประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ Numerical Method ให้ค่าต่ำกว่าทุกกรณี แสดงว่า การประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ Numerical Method มีความไม่แน่นอนในการประมาณค่าที่ต่ำกว่า การประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ MCMC

สำหรับการทดสอบความแตกต่างของค่าเฉลี่ยในการประมาณค่าพารามิเตอร์ของข้อสอบระหว่างการประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ MCMC กับ การประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ Numerical พบว่า ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 แสดงว่า การประมาณค่าพารามิเตอร์ของข้อสอบทั้งสองวิธีให้ผลที่ไม่แตกต่างกัน

1.2 การวิเคราะห์ค่าพารามิเตอร์ของข้อสอบตามแนวทฤษฎีการตอบสนองของข้อสอบ

การวิเคราะห์ค่าพารามิเตอร์ของข้อสอบตามแนวทฤษฎีการตอบสนองข้อสอบ ทำการวิเคราะห์เปรียบเทียบระหว่างวิธีการประมาณค่าพารามิเตอร์แบบ MCMC ซึ่งใช้ในวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของของโมเดลการตอบสนองข้อสอบ (GIRM) ในรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และ รูปแบบที่ 3 AGIRM B เปรียบเทียบกับ วิธีการประมาณค่าพารามิเตอร์แบบ Numerical ซึ่งใช้ในวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของของโมเดลการตอบสนองข้อสอบ (GIRM) ในรูปแบบที่ 4 Numerical Bayesian GIRM โดยวิเคราะห์เปรียบเทียบในแต่ละกรณีศึกษาทั้ง 16 กรณีศึกษา มีผลการศึกษา ดังนี้

ตารางที่ 13 ผลการวิเคราะห์ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐานและการวิเคราะห์ความแตกต่างด้วยสถิติการทดสอบที่ ของค่าพารามิเตอร์ของผู้สอบระหว่างวิธีการ GIRM ที่ประมาณค่าของวิธี MCMC และ Numerical จำแนกตามเงื่อนไขของการวิเคราะห์

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	GIRM (MCMC)		GIRM (Numerical)		t-test	p-value
				Mean	S.D.	Mean	S.D.		
Normal	Normal	300	5	0.0002	0.0266	-0.0004	0.0227	0.3128	0.7546
		300	10	0.0003	0.0329	-0.0019	0.0297	0.8488	0.3963
		100	5	-0.0019	0.0215	-0.0037	0.0194	0.6162	0.5385
		100	10	-0.0019	0.0352	-0.0039	0.0307	0.6178	0.4776
	Beta	300	5	0.0913	0.0225	NA	NA	NA	NA
		300	10	0.1076	0.0302	NA	NA	NA	NA
		100	5	0.1136	0.0216	NA	NA	NA	NA
		100	10	0.1770	0.0324	NA	NA	NA	NA
Gamma	Normal	300	5	0.9574	0.0258	NA	NA	NA	NA
		300	10	0.9491	0.0317	NA	NA	NA	NA
		100	5	0.9441	0.0225	NA	NA	NA	NA
		100	10	0.9232	0.0347	NA	NA	NA	NA
	Beta	300	5	0.9219	0.0246	NA	NA	NA	NA
		300	10	0.8994	0.0307	NA	NA	NA	NA
		100	5	0.9003	0.0214	NA	NA	NA	NA
		100	10	0.8597	0.0331	NA	NA	NA	NA

NA = ไม่สามารถทำการวิเคราะห์ได้

จากผลการวิเคราะห์ค่าพารามิเตอร์ของผู้สอบตามแนวทฤษฎีการตอบสนองข้อสอบ พบว่า การวิเคราะห์ค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ MCMC ในรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B ให้ค่าเฉลี่ยของพารามิเตอร์ของผู้สอบอยู่ระหว่าง -0.0019 ถึง 0.9574และให้ค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0214 ถึง 0.0352 ส่วนการวิเคราะห์ค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ Numerical Method ในรูปแบบที่ 4 Numerical Bayesian GIRM สามารถวิเคราะห์ได้เฉพาะกรณีการแจกแจงเริ่มแรกผู้สอบและการแจกแจงเริ่มแรกของข้อสอบมีลักษณะการกระจายแบบโค้งปกติเท่านั้น โดยให้ค่าเฉลี่ยของพารามิเตอร์ของ

ผู้สอบอยู่ระหว่าง -0.0004 ถึง -0.0039 และให้ค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0194 ถึง 0.0307

เมื่อพิจารณาเปรียบเทียบเป็นรายกรณีของค่าส่วนเบี่ยงเบนมาตรฐานระหว่างการประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ MCMC กับ การประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ Numerical พบว่า การประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ Numerical ให้ค่าต่ำกว่าทุกกรณี แสดงว่า การประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ Numerical มีความไม่แน่นอนในการประมาณค่าที่ต่ำกว่า การประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ MCMC

สำหรับการทดสอบความแตกต่างของค่าเฉลี่ยในการประมาณค่าพารามิเตอร์ของผู้สอบระหว่างการประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ MCMC กับ การประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ Numerical พบว่า ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 แสดงว่า การประมาณค่าพารามิเตอร์ของผู้สอบทั้งสองวิธีให้ผลที่ไม่แตกต่างกัน

ตอนที่ 2 ผลการวิเคราะห์องค์ประกอบความแปรปรวนตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

เนื่องจากวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (Generalizability in Item Response Modeling; GIRM) เป็นการรวมโมเดลการสุ่ม (Sampling Model) ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) เข้าไปในโมเดลการกำหนดมาตรวัด (Scaling Model) ของทฤษฎีการตอบสนองข้อสอบ (IRT) ดังนั้นสารสนเทศที่ได้จากการวิเคราะห์จึงให้สารสนเทศทั้งสารสนเทศที่ได้จากการวิเคราะห์จากทฤษฎีการตอบสนองข้อสอบ (IRT) และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) โดยในตอนนี้นำเสนอสารสนเทศเกี่ยวกับผลการวิเคราะห์องค์ประกอบความแปรปรวนและค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) จำแนกตามรูปแบบในการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบทั้ง 4 รูปแบบ และจำแนกตามเงื่อนไขของการวิเคราะห์ทั้ง 16 กรณี มีรายละเอียดของผลการวิเคราะห์ ดังนี้

ตารางที่ 14 ผลการวิเคราะห์องค์ประกอบความแปรปรวนและค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

องค์ประกอบแปรปรวน	Original GIRM		AGIRM A		AGIRM B		Numerical GIRM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
กรณีที่ 1 Normal-Normal-300-5								
$\sigma^2(p)$	0.0117	0.0120	0.0120	0.0134	0.0120	0.0134	0.0198	0.0103
$\sigma^2(i)$	0.0009	0.0007	0.0012	0.0014	0.0012	0.0014	0.0009	0.0007
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	0.0000	0.0000
$\sigma^2(e)$	0.2309	0.0140	NA	NA	0.2307	0.0152	0.2223	0.0301
$\sigma^2(pi,e)$	0.2309	0.0140	0.2307	0.0151	0.2307	0.0151	0.2223	0.0101
$\hat{E}_p^2$	0.1776	0.1491	0.1756	0.1708	0.1756	0.1708	0.2525	0.1364
Φ	0.1771	0.1488	0.1751	0.1703	0.1751	0.1703	0.2519	0.1311
กรณีที่ 2 Normal-Normal-300-10								
$\sigma^2(p)$	0.0189	0.0109	0.0191	0.0110	0.0191	0.0110	0.0262	0.0092
$\sigma^2(i)$	0.0015	0.0007	0.0016	0.0011	0.0016	0.0011	0.0016	0.0011
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	0.0000	0.0001
$\sigma^2(e)$	0.2261	0.0113	NA	NA	0.2260	0.0115	0.2193	0.0127
$\sigma^2(pi,e)$	0.2261	0.0113	0.2260	0.0115	0.2260	0.0115	0.2193	0.0127
$\hat{E}_p^2$	0.4207	0.1532	0.4229	0.1568	0.4229	0.1568	0.4059	0.1831
Φ	0.4192	0.1530	0.4214	0.1565	0.4214	0.1565	0.4046	0.1826
กรณีที่ 3 Normal-Normal-100-5								
$\sigma^2(p)$	0.0118	0.0132	0.0116	0.0163	0.0116	0.0163	0.0202	0.0120
$\sigma^2(i)$	0.0027	0.0019	0.0034	0.0039	0.0034	0.0039	0.0027	0.0028
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	0.0000	0.0001
$\sigma^2(e)$	0.2284	0.0153	NA	NA	0.2285	0.0185	0.2180	0.0125
$\sigma^2(pi,e)$	0.2284	0.0153	0.2285	0.0184	0.2285	0.0184	0.2181	0.0125
$\hat{E}_p^2$	0.1765	0.1622	0.1540	0.2235	0.1540	0.2235	0.2602	0.1421
Φ	0.1751	0.1613	0.1528	0.2214	0.1528	0.2214	0.2583	0.1398

ตารางที่ 14 (ต่อ)

องค์ประกอบแปรปรวน	Original GIRM		AGIRM A		AGIRM B		Numerical GIRM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
กรณีที่ 4 Normal-Normal-100-10								
$\sigma^2(p)$	0.0184	0.0114	0.0186	0.0123	0.0186	0.0123	0.0254	0.0091
$\sigma^2(i)$	0.0048	0.0022	0.0053	0.0034	0.0053	0.0034	0.0047	0.0024
$\sigma^2(pi)$	0.0001	0.0001	NA	NA	0.0001	0.0001	0.0001	0.0001
$\sigma^2(e)$	0.2232	0.0120	NA	NA	0.2230	0.0130	0.2178	0.0156
$\sigma^2(pi,e)$	0.2232	0.0119	0.2231	0.0129	0.2231	0.0129	0.2179	0.0154
$\hat{E}_p^2$	0.4112	0.1667	0.4043	0.1945	0.4043	0.1945	0.4049	0.1404
Φ	0.4067	0.1660	0.3996	0.1929	0.3996	0.1929	0.4011	0.1377
กรณีที่ 5 Normal-Beta-300-5								
$\sigma^2(p)$	0.0131	0.0123	0.0132	0.0133	0.0132	0.0133	NA	NA
$\sigma^2(i)$	0.0003	0.0005	0.0005	0.0011	0.0005	0.0011	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2329	0.0138	NA	NA	0.2328	0.0146	NA	NA
$\sigma^2(pi,e)$	0.2329	0.0138	0.2328	0.0146	0.2328	0.0146	NA	NA
$\hat{E}_p^2$	0.1959	0.1490	0.1923	0.1643	0.1923	0.1643	NA	NA
Φ	0.1958	0.1489	0.1921	0.1641	0.1921	0.1641	NA	NA
กรณีที่ 6 Normal-Beta-300-10								
$\sigma^2(p)$	0.0199	0.0110	0.0200	0.0115	0.0200	0.0115	NA	NA
$\sigma^2(i)$	0.0007	0.0006	0.0008	0.0009	0.0008	0.0009	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2269	0.0114	NA	NA	0.2268	0.0119	NA	NA
$\sigma^2(pi,e)$	0.2269	0.0114	0.2268	0.0119	0.2268	0.0119	NA	NA
$\hat{E}_p^2$	0.4344	0.1483	0.4316	0.1597	0.4316	0.1597	NA	NA
Φ	0.4338	0.1483	0.4309	0.1595	0.4309	0.1595	NA	NA

ตารางที่ 14 (ต่อ)

องค์ประกอบแปรปรวน	Original GIRM		AGIRM A		AGIRM B		Numerical GIRM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
กรณีที่ 7 Normal-Beta-100-5								
$\sigma^2(p)$	0.0134	0.0136	0.0134	0.0162	0.0134	0.0162	NA	NA
$\sigma^2(i)$	0.0009	0.0010	0.0013	0.0026	0.0013	0.0026	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2315	0.0149	NA	NA	0.2315	0.0172	NA	NA
$\sigma^2(pi,e)$	0.2315	0.0149	0.2315	0.0172	0.2315	0.0172	NA	NA
$\hat{E}_p^2$	0.1945	0.1644	0.1784	0.2172	0.1784	0.2172	NA	NA
Φ	0.1940	0.1641	0.1778	0.2166	0.1778	0.2166	NA	NA
กรณีที่ 8 Normal-Beta-100-10								
$\sigma^2(p)$	0.0202	0.0114	0.0205	0.0126	0.0205	0.0126	NA	NA
$\sigma^2(i)$	0.0019	0.0010	0.0020	0.0021	0.0020	0.0021	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2253	0.0119	NA	NA	0.2252	0.0130	NA	NA
$\sigma^2(pi,e)$	0.2253	0.0118	0.2252	0.0130	0.2252	0.0130	NA	NA
$\hat{E}_p^2$	0.4362	0.1570	0.4318	0.1824	0.4318	0.1824	NA	NA
Φ	0.4343	0.1567	0.4299	0.1821	0.4299	0.1821	NA	NA
กรณีที่ 9 Gamma-Normal-300-5								
$\sigma^2(p)$	0.0132	0.0041	0.0135	0.0065	0.0135	0.0065	NA	NA
$\sigma^2(i)$	0.0009	0.0007	0.0012	0.0015	0.0012	0.0015	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2296	0.0083	NA	NA	0.2292	0.0100	NA	NA
$\sigma^2(pi,e)$	0.2296	0.0083	0.2292	0.0100	0.2292	0.0100	NA	NA
$\hat{E}_p^2$	0.2202	0.0526	0.2196	0.0895	0.2196	0.0895	NA	NA
Φ	0.2195	0.0526	0.2188	0.0893	0.2188	0.0893	NA	NA

ตารางที่ 14 (ต่อ)

องค์ประกอบแปรปรวน	Original GIRM		AGIRM A		AGIRM B		Numerical GIRM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
กรณีที่ 10 Gamma-Normal-300-10								
$\sigma^2(p)$	0.0181	0.0056	0.0183	0.0065	0.0183	0.0065	NA	NA
$\sigma^2(i)$	0.0015	0.0007	0.0017	0.0012	0.0017	0.0012	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2270	0.0064	NA	NA	0.2268	0.0074	NA	NA
$\sigma^2(pi,e)$	0.2270	0.0064	0.2269	0.0074	0.2269	0.0074	NA	NA
$\hat{E}_p^2$	0.4346	0.0779	0.4337	0.0923	0.4337	0.0923	NA	NA
Φ	0.4331	0.0779	0.4319	0.0922	0.4319	0.0922	NA	NA
กรณีที่ 11 Gamma-Normal-100-5								
$\sigma^2(p)$	0.0126	0.0043	0.0129	0.0103	0.0129	0.0103	NA	NA
$\sigma^2(i)$	0.0024	0.0017	0.0035	0.0043	0.0035	0.0043	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2283	0.0087	NA	NA	0.2275	0.0127	NA	NA
$\sigma^2(pi,e)$	0.2284	0.0086	0.2276	0.0127	0.2276	0.0127	NA	NA
$\hat{E}_p^2$	0.2121	0.0575	0.1998	0.1497	0.1998	0.1497	NA	NA
Φ	0.2104	0.0573	0.1978	0.1479	0.1978	0.1479	NA	NA
กรณีที่ 12 Gamma-Normal-100-10								
$\sigma^2(p)$	0.0167	0.0057	0.0168	0.0076	0.0168	0.0076	NA	NA
$\sigma^2(i)$	0.0044	0.0020	0.0048	0.0032	0.0048	0.0032	NA	NA
$\sigma^2(pi)$	0.0001	0.0001	NA	NA	0.0001	0.0001	NA	NA
$\sigma^2(e)$	0.2254	0.0069	NA	NA	0.2256	0.0091	NA	NA
$\sigma^2(pi,e)$	0.2255	0.0068	0.2257	0.0090	0.2257	0.0090	NA	NA
$\hat{E}_p^2$	0.4152	0.0830	0.4057	0.1214	0.4057	0.1214	NA	NA
Φ	0.4107	0.0831	0.4010	0.1208	0.4010	0.1208	NA	NA

ตารางที่ 14 (ต่อ)

องค์ประกอบแปรปรวน	Original GIRM		AGIRM A		AGIRM B		Numerical GIRM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
กรณีที่ 13 Gamma-Beta-300-5								
$\sigma^2(p)$	0.0122	0.0036	0.0118	0.0063	0.0118	0.0063	NA	NA
$\sigma^2(i)$	0.0005	0.0006	0.0007	0.0013	0.0007	0.0013	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2331	0.0060	NA	NA	0.2336	0.0083	NA	NA
$\sigma^2(pi,e)$	0.2332	0.0060	0.2336	0.0083	0.2336	0.0083	NA	NA
$\hat{E}_p^2$	0.2048	0.0485	0.1935	0.0900	0.1935	0.0900	NA	NA
Φ	0.2045	0.0485	0.1930	0.0897	0.1930	0.0897	NA	NA
กรณีที่ 14 Gamma-Beta-300-10								
$\sigma^2(p)$	0.0167	0.0049	0.0169	0.0058	0.0169	0.0058	NA	NA
$\sigma^2(i)$	0.0008	0.0007	0.0009	0.0010	0.0009	0.0010	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2302	0.0055	NA	NA	0.2300	0.0062	NA	NA
$\sigma^2(pi,e)$	0.2302	0.0055	0.2300	0.0062	0.2300	0.0062	NA	NA
$\hat{E}_p^2$	0.4120	0.0750	0.4121	0.0895	0.4121	0.0895	NA	NA
Φ	0.4111	0.0750	0.4112	0.0895	0.4112	0.0895	NA	NA
กรณีที่ 15 Gamma-Beta-100-5								
$\sigma^2(p)$	0.0114	0.0036	0.0111	0.0093	0.0111	0.0093	NA	NA
$\sigma^2(i)$	0.0011	0.0011	0.0012	0.0026	0.0012	0.0026	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2334	0.0067	NA	NA	0.2338	0.0114	NA	NA
$\sigma^2(pi,e)$	0.2334	0.0067	0.2338	0.0114	0.2338	0.0114	NA	NA
$\hat{E}_p^2$	0.1937	0.0501	0.1743	0.1401	0.1743	0.1401	NA	NA
Φ	0.1930	0.0500	0.1736	0.1395	0.1736	0.1395	NA	NA

ตารางที่ 14 (ต่อ)

องค์ประกอบแปรปรวน	Original GIRM		AGIRM A		AGIRM B		Numerical GIRM	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
กรณีที่ 16 Gamma-Beta-100-10								
$\sigma^2(p)$	0.0150	0.0047	0.0154	0.0071	0.0154	0.0071	NA	NA
$\sigma^2(i)$	0.0021	0.0012	0.0023	0.0023	0.0023	0.0023	NA	NA
$\sigma^2(pi)$	0.0000	0.0000	NA	NA	0.0000	0.0000	NA	NA
$\sigma^2(e)$	0.2305	0.0058	NA	NA	0.2302	0.0083	NA	NA
$\sigma^2(pi,e)$	0.2305	0.0058	0.2302	0.0083	0.2302	0.0083	NA	NA
$\hat{E}_p^2$	0.3861	0.0779	0.3812	0.1216	0.3812	0.1216	NA	NA
Φ	0.3841	0.0779	0.3790	0.1213	0.3790	0.1213	NA	NA

NA = ไม่สามารถทำการวิเคราะห์ได้

จากผลการวิเคราะห์ค่าองค์ประกอบความแปรปรวนตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด พบว่า การวิเคราะห์ค่าองค์ประกอบความแปรปรวนในแต่ละแหล่งความแปรปรวนที่ได้จากรูปแบบการวิเคราะห์ทั้ง 4 รูปแบบในการประมาณค่าให้ผลที่ใกล้เคียงกัน โดยรูปแบบการประมาณค่าในรูปแบบที่ 4 Numerical Bayesian GIRM สามารถวิเคราะห์ได้เฉพาะกรณีการแจกแจงเริ่มแรกผู้สอบและการแจกแจงเริ่มแรกของข้อสอบมีลักษณะการกระจายแบบโค้งปกติเท่านั้น และรูปแบบการประมาณค่าในรูปแบบที่ 2 AGIRM A ไม่สามารถวิเคราะห์แยกองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ ($\sigma^2(pi)$) และองค์ประกอบของความคลาดเคลื่อนในส่วนที่เหลือได้ ($\sigma^2(e)$)

โดยองค์ประกอบความแปรปรวนในการประมาณค่าของผู้สอบ ($\sigma^2(p)$) มีค่าเฉลี่ยอยู่ระหว่าง 0.0262 ถึง 0.0111 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0292 ถึง 0.0036 องค์ประกอบความแปรปรวนของข้อสอบ ($\sigma^2(i)$) มีค่าเฉลี่ยอยู่ระหว่าง 0.0053 ถึง 0.0005 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0064 ถึง 0.0005 องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ ($\sigma^2(pi)$) มีค่าเฉลี่ยอยู่ระหว่าง 0.0001 ถึง 0.0000 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0001 ถึง 0.0000 องค์ประกอบความแปรปรวนของความคลาดเคลื่อนในส่วนที่เหลือ ($\sigma^2(e)$) มีค่าเฉลี่ยอยู่ระหว่าง 0.2338 ถึง 0 .2309 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0356 ถึง 0.0055 องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบรวมกับความคลาดเคลื่อนในส่วนที่เหลือ ($\sigma^2(pi,e)$) มีค่าเฉลี่ยอยู่ระหว่าง

0.2338 ถึง 0.2179 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0354 ถึง 0.0055 ค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) มีค่าเฉลี่ยอยู่ระหว่าง 0.4362 ถึง 0.1756 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.2831 ถึง 0.0485 และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) มีค่าเฉลี่ยอยู่ระหว่าง 0.4343 ถึง 0.1528 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.2828 ถึง 0.0485

เมื่อพิจารณาในกรณีที่มีลักษณะการแจกแจงเริ่มแรกรูปแบบเดียวกัน พบว่า ถ้ามีจำนวนข้อเพิ่มมากขึ้น จะส่งผลทำให้มีค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) เพิ่มมากขึ้นในทุกกรณีการแจกแจงเริ่มแรก เช่น กรณี Normal-Normal 300-5 มีค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) เท่ากับ 0.1776 และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) เท่ากับ 0.1771 เทียบกับกรณี Normal-Normal 300-10 มีค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) เท่ากับ 0.4207 และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) เท่ากับ 0.4192

สำหรับจำนวนผู้เข้าสอบก็เช่นเดียวกับจำนวนข้อสอบ โดยในกรณีที่มีลักษณะการแจกแจงเริ่มแรกรูปแบบเดียวกัน พบว่า ถ้ามีจำนวนผู้เข้าสอบมากขึ้น จะส่งผลทำให้มีค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) เพิ่มมากขึ้นในทุกกรณีการแจกแจงเริ่มแรก เช่น กรณี Normal-Normal 300-5 มีค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) เท่ากับ 0.1776 และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) เท่ากับ 0.1771 เทียบกับกรณี Normal-Normal 100-10 มีค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) เท่ากับ 0.1765 และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) เท่ากับ 0.1751

ตอนที่ 3 การวิเคราะห์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าในแต่ละรูปแบบการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM)

การวิเคราะห์ในตอนนี้มุ่งเปรียบเทียบประสิทธิภาพของการประมาณค่าในแต่ละรูปแบบการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทั้ง 4 รูปแบบ ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM ซึ่ง พัฒนาโดย Brigg และ Wilson (2007) และรูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B รูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งพัฒนาวิธีการประมาณค่าโดยผู้วิจัย ในการศึกษาครั้งนี้จะวัดประสิทธิภาพของตัวประมาณค่าจากดัชนี 3 ประเภท ซึ่งได้แก่ ความลำเอียงในการประมาณค่า (Biased estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ดังนั้นจะเห็นได้ว่า การวิเคราะห์ในตอนนี้จะมุ่งตอบคำถามการวิจัยและวัตถุประสงค์ในการวิจัยข้อที่ 1 โดยรายละเอียดจะแบ่งการนำเสนอในแต่ละประเภทของดัชนีดังนี้

3.1 การวิเคราะห์ความลำเอียงในการประมาณค่าด้วยการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD)

ความลำเอียงในการประมาณค่าในส่วนนี้จะแยกการนำเสนอตามรูปแบบการประมาณค่าทั้ง 4 รูปแบบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GIRM) ทั้ง 4 รูปแบบ และจำแนกตามเงื่อนไขของการวิเคราะห์ทั้ง 16 กรณี มีรายละเอียดของผลการวิเคราะห์ดังนี้

ตารางที่ 15 ผลการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ขององค์ประกอบความแปรปรวนและสัมประสิทธิ์สหสัมพันธ์ความน่าเชื่อถือของผลการวัดจำแนกตามรูปแบบการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบทั้ง 16 รายกรณี

องค์ประกอบแปรปรวน/ค่าสัมประสิทธิ์	Original GIRM	AGIRM A	AGIRM B	Numerical GIRM
กรณีที่ 1 Normal – Normal 300 - 5				
$\sigma^2(p)$	0.7132	0.7140	0.7140	0.6763
$\sigma^2(i)$	0.8416	0.9042	0.9042	0.8529
$\sigma^2(pi)$	0.9776	NA	0.9776	0.9654
$\sigma^2(e)$	0.2042	NA	0.2037	0.2046
$\sigma^2(pi,e)$	0.2010	0.2006	0.2006	0.2022
$\hat{E}_p^2$	0.6508	0.6577	0.6577	0.5694
Φ	0.6465	0.6541	0.6541	0.5694
กรณีที่ 2 Normal – Normal 300 - 10				
$\sigma^2(p)$	0.5160	0.5111	0.5111	0.7221
$\sigma^2(i)$	0.7678	0.7668	0.7668	0.8143
$\sigma^2(pi)$	0.9552	NA	0.9552	0.9426
$\sigma^2(e)$	0.1784	NA	0.1780	0.2023
$\sigma^2(pi,e)$	0.1750	0.1746	0.1746	0.1997
$\hat{E}_p^2$	0.3676	0.3641	0.3641	0.4739
Φ	0.3615	0.3582	0.3582	0.4739
กรณีที่ 3 Normal – Normal 100 - 5				
$\sigma^2(p)$	0.7156	0.7433	0.7433	0.6631
$\sigma^2(i)$	0.7255	0.9946	0.9946	0.9160
$\sigma^2(pi)$	0.9279	NA	0.9279	0.9620
$\sigma^2(e)$	0.1948	NA	0.1966	0.1961
$\sigma^2(pi,e)$	0.1919	0.1937	0.1937	0.1938
$\hat{E}_p^2$	0.6536	0.7100	0.7100	0.5586
Φ	0.6515	0.7083	0.7083	0.5593

ตารางที่ 15 (ต่อ)

องค์ประกอบแปรปรวน/ค่าสัมประสิทธิ์	Original GIRM	AGIRM A	AGIRM B	Numerical GIRM
กรณีที่ 4 Normal – Normal 100 - 10				
$\sigma^2(p)$	0.5370	0.5323	0.5323	0.7327
$\sigma^2(i)$	0.5191	0.5873	0.5873	0.8721
$\sigma^2(pi)$	0.8614	NA	0.8614	0.9600
$\sigma^2(e)$	0.1667	NA	0.1661	0.2034
$\sigma^2(pi,e)$	0.1636	0.1629	0.1629	0.2009
$\hat{E}_p^2$	0.3833	0.3939	0.3939	0.4712
Φ	0.3821	0.3929	0.3929	0.4734
กรณีที่ 5 Normal – Beta 300 - 5				
$\sigma^2(p)$	0.6817	0.6868	0.6868	NA
$\sigma^2(i)$	0.9284	0.9421	0.9421	NA
$\sigma^2(pi)$	0.9910	NA	0.9910	NA
$\sigma^2(e)$	0.2137	NA	0.2136	NA
$\sigma^2(pi,e)$	0.2104	0.2104	0.2104	NA
$\hat{E}_p^2$	0.6159	0.6251	0.6251	NA
Φ	0.6106	0.6205	0.6205	NA
กรณีที่ 6 Normal – Beta 300 - 10				
$\sigma^2(p)$	0.4927	0.4950	0.4950	NA
$\sigma^2(i)$	0.8959	0.8832	0.8832	NA
$\sigma^2(pi)$	0.9779	NA	0.9779	NA
$\sigma^2(e)$	0.1793	NA	0.1789	NA
$\sigma^2(pi,e)$	0.1760	0.1756	0.1756	NA
$\hat{E}_p^2$	0.3474	0.3527	0.3527	NA
Φ	0.3399	0.3457	0.3457	NA

ตารางที่ 15 (ต่อ)

องค์ประกอบแปรปรวน/ค่าสัมประสิทธิ์	Original GIRM	AGIRM A	AGIRM B	Numerical GIRM
กรณีที่ 7 Normal – Beta 100 - 5				
$\sigma^2(p)$	0.6795	0.6928	0.6928	NA
$\sigma^2(i)$	0.8587	1.0063	1.0063	NA
$\sigma^2(pi)$	0.9693	NA	0.9693	NA
$\sigma^2(e)$	0.2110	NA	0.2117	NA
$\sigma^2(pi,e)$	0.2080	0.2086	0.2086	NA
$\hat{E}_p^2$	0.6191	0.6558	0.6558	NA
Φ	0.6149	0.6537	0.6537	NA
กรณีที่ 8 Normal – Beta 100 - 10				
$\sigma^2(p)$	0.6464	0.6377	0.6377	NA
$\sigma^2(i)$	0.8028	0.8432	0.8432	NA
$\sigma^2(pi)$	0.9428	NA	0.9428	NA
$\sigma^2(e)$	0.1925	NA	0.1906	NA
$\sigma^2(pi,e)$	0.1896	0.1877	0.1877	NA
$\hat{E}_p^2$	0.5524	0.5533	0.5533	NA
Φ	0.5450	0.5463	0.5463	NA
กรณีที่ 9 Gamma – Normal 300 - 5				
$\sigma^2(p)$	0.5190	0.5152	0.5152	NA
$\sigma^2(i)$	0.7833	0.7601	0.7601	NA
$\sigma^2(pi)$	0.9405	NA	0.9405	NA
$\sigma^2(e)$	0.1800	NA	0.1794	NA
$\sigma^2(pi,e)$	0.1768	0.1762	0.1762	NA
$\hat{E}_p^2$	0.3426	0.3441	0.3441	NA
Φ	0.3355	0.3375	0.3375	NA
กรณีที่ 10 Gamma – Normal 300 - 10				
$\sigma^2(p)$	0.6605	0.6541	0.6541	NA
$\sigma^2(i)$	0.7314	0.9686	0.9686	NA
$\sigma^2(pi)$	0.8865	NA	0.8865	NA

ตารางที่ 15 (ต่อ)

องค์ประกอบแปรปรวน/ค่าสัมประสิทธิ์	Original GIRM	AGIRM A	AGIRM B	Numerical GIRM
$\sigma^2(e)$	0.1837	NA	0.1807	NA
$\sigma^2(pi,e)$	0.1811	0.1781	0.1781	NA
$\hat{E}_p^2$	0.5662	0.5913	0.5913	NA
Φ	0.5617	0.5886	0.5886	NA
กรณีที่ 11 Gamma – Normal 100 - 5				
$\sigma^2(p)$	0.5490	0.5499	0.5499	NA
$\sigma^2(i)$	0.5469	0.6077	0.6077	NA
$\sigma^2(pi)$	0.8180	NA	0.8180	NA
$\sigma^2(e)$	0.1704	NA	0.1718	NA
$\sigma^2(pi,e)$	0.1674	0.1688	0.1688	NA
$\hat{E}_p^2$	0.3681	0.3831	0.3831	NA
Φ	0.3653	0.3810	0.3810	NA
กรณีที่ 12 Gamma – Normal 100 - 10				
$\sigma^2(p)$	0.6753	0.6874	0.6874	NA
$\sigma^2(i)$	0.9149	0.9089	0.9089	NA
$\sigma^2(pi)$	0.9675	NA	0.9675	NA
$\sigma^2(e)$	0.2102	NA	0.2125	NA
$\sigma^2(pi,e)$	0.2070	0.2093	0.2093	NA
$\hat{E}_p^2$	0.5849	0.6081	0.6081	NA
Φ	0.5771	0.6010	0.6010	NA
กรณีที่ 13 Gamma – Beta 300 - 5				
$\sigma^2(p)$	0.6753	0.6874	0.6874	NA
$\sigma^2(i)$	0.9149	0.9089	0.9089	NA
$\sigma^2(pi)$	0.9675	NA	0.9675	NA
$\sigma^2(e)$	0.2102	NA	0.2125	NA
$\sigma^2(pi,e)$	0.2070	0.2093	0.2093	NA
$\hat{E}_p^2$	0.5849	0.6081	0.6081	NA
Φ	0.5771	0.6010	0.6010	NA

ตารางที่ 15 (ต่อ)

องค์ประกอบแปรปรวน/ ค่าสัมประสิทธิ์	Original GIRM	AGIRM A	AGIRM B	Numerical GIRM
กรณีที่ 14 Gamma – Beta 300 - 10				
$\sigma^2(p)$	0.5566	0.5514	0.5514	NA
$\sigma^2(i)$	0.8848	0.8720	0.8720	NA
$\sigma^2(pi)$	0.9548	NA	0.9548	NA
$\sigma^2(e)$	0.1971	NA	0.1962	NA
$\sigma^2(pi,e)$	0.1935	0.1926	0.1926	NA
$\hat{E}_p^2$	0.3764	0.3763	0.3763	NA
Φ	0.3682	0.3683	0.3683	NA
กรณีที่ 15 Gamma – Beta 100 - 5				
$\sigma^2(p)$	0.6927	0.7026	0.7026	NA
$\sigma^2(i)$	0.8216	0.9714	0.9714	NA
$\sigma^2(pi)$	0.9321	NA	0.9321	NA
$\sigma^2(e)$	0.2137	NA	0.2160	NA
$\sigma^2(pi,e)$	0.2106	0.2130	0.2130	NA
$\hat{E}_p^2$	0.6049	0.6451	0.6451	NA
Φ	0.5985	0.6397	0.6397	NA
กรณีที่ 16 Gamma – Beta 100 - 10				
$\sigma^2(p)$	0.5879	0.5792	0.5792	NA
$\sigma^2(i)$	0.7122	0.7483	0.7483	NA
$\sigma^2(pi)$	0.8903	NA	0.8903	NA
$\sigma^2(e)$	0.1984	NA	0.1969	NA
$\sigma^2(pi,e)$	0.1952	0.1936	0.1936	NA
$\hat{E}_p^2$	0.4094	0.4170	0.4170	NA
Φ	0.4034	0.4115	0.4115	NA

NA = ไม่สามารถทำการวิเคราะห์ได้

จากการประมาณค่าความลำเอียงในการประมาณค่า (Biased estimator) ขององค์ประกอบความแปรปรวนตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ด้วยวิธีการประมาณค่าทั้ง 4 วิธีซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) พบว่า การประมาณค่าด้วยรูปแบบที่ 1 Original GIRM มีค่า MAD อยู่ในช่วงระหว่าง 0.1704 ถึง 0.9910 การประมาณค่าด้วยรูปแบบที่ 2 AGIRM Aมีค่า MAD อยู่ในช่วงระหว่าง 0.1629 ถึง 1.0063 การประมาณค่าด้วยรูปแบบที่ 3 AGIRM B มีค่า MAD อยู่ในช่วงระหว่าง 1.0063 ถึง 0.1746 และการประมาณค่าด้วยรูปแบบที่ 4 Numerical Bayesian GIRM มีค่า MAD อยู่ในช่วงระหว่าง 0.1938 ถึง 0.9654 โดยรูปแบบการประมาณค่าในรูปแบบที่ 4 Numerical Bayesian GIRM สามารถวิเคราะห์ได้เฉพาะกรณีการแจกแจงเริ่มแรกผู้สอบและการแจกแจงเริ่มแรกของข้อสอบมีลักษณะการกระจายแบบโค้งปกติเท่านั้น และรูปแบบการประมาณค่าในรูปแบบที่ 2 AGIRM A ไม่สามารถวิเคราะห์แยกองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ ($\sigma^2(pi)$) และองค์ประกอบของความคลาดเคลื่อนในส่วนที่เหลือ ($\sigma^2(e)$) ออกจากกันได้

เมื่อทำการเปรียบเทียบผลการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างรูปแบบการประมาณค่าทั้ง 4 วิธี โดยใน 4 กรณีแรกที่มีลักษณะการแจกแจงเริ่มแรกของข้อสอบและผู้สอบแบบปกติ (normal distribution) ที่รูปแบบที่ 4 สามารถประมาณค่าพารามิเตอร์ได้ พบว่า รูปแบบที่ 4 Numerical Bayesian GIRM มีวิเคราะห์ให้ค่า MAD น้อยที่สุดใน 2 กรณี ได้แก่ กรณีที่ 2 Normal-Normal 300-5 และ กรณีที่ 4 Normal-Normal 100-5 ส่วน 2 กรณีที่เหลือ ได้แก่ กรณีที่ 1 Normal-Normal 300-10 และ กรณีที่ 3 Normal-Normal 100-10 รูปแบบการประมาณค่าที่ 1 Original GIRM มีค่า MAD น้อยที่สุด

สำหรับกรณีอื่นๆ พบว่า การประมาณค่าด้วยรูปแบบที่ 1 Original GIRM จะให้ค่า MAD ต่ำกว่ารูปแบบการประมาณค่าวิธีอื่นๆ รองลงมา รูปแบบที่ 2 AGIRM A และ รูปแบบที่ 3 AGIRM B ให้ค่า MAD เท่ากัน สำหรับรูปแบบที่ 4 Numerical Bayesian GIRM ให้ค่า MAD สูงที่สุด

3.2 การวิเคราะห์ความไม่แน่นอนในการประมาณค่าด้วยการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน

ความไม่แน่นอนในการประมาณค่าในส่วนนี้จะแยกการนำเสนอตามรูปแบบการประมาณค่าทั้ง 4 รูปแบบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GIRM) ทั้ง 4 รูปแบบและจำแนกตามเงื่อนไขของการวิเคราะห์ทั้ง 16 กรณี มีรายละเอียดของผลการวิเคราะห์ ดังนี้

ตารางที่ 16 ผลการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐานรวมในการประมาณค่าองค์ประกอบความแปรปรวนจำแนกตามรูปแบบการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบทั้ง 16 รายกรณี

การแจกแจงเริ่มแรกผู้สอบ	**การแจกแจงเริ่มแรกข้อสอบ**	**ขนาดตัวอย่าง**	**จำนวนข้อสอบ**	Original GIRM	AGIRM A	AGIRM B	Numerical GIRM
Normal	Normal	300	5	0.0092	0.0117	0.0101	0.0087
		300	10	0.0078	0.0092	0.0079	0.0065
		100	5	0.0101	0.0143	0.0124	0.0050
		100	10	0.0083	0.0105	0.0091	0.0074
	Beta	300	5	0.0092	0.0114	0.0098	NA
		300	10	0.0079	0.0095	0.0082	NA
		100	5	0.0100	0.0137	0.0118	NA
		100	10	0.0082	0.0105	0.0091	NA
Gamma	Normal	300	5	0.0046	0.0069	0.0060	NA
		300	10	0.0042	0.0057	0.0049	NA
		100	5	0.0049	0.0097	0.0084	NA
		100	10	0.0045	0.0070	0.0061	NA
	Beta	300	5	0.0035	0.0060	0.0052	NA
		300	10	0.0037	0.0049	0.0042	NA
		100	5	0.0038	0.0086	0.0074	NA
		100	10	0.0037	0.0064	0.0055	NA

NA = ไม่สามารถทำการวิเคราะห์ได้

จากผลการวิเคราะห์ความไม่แน่นอนในการประมาณค่า ด้วยการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐานรวมในการประมาณค่าองค์ประกอบความแปรปรวน พบว่า มีค่าส่วนเบี่ยงเบนมาตรฐานรวมอยู่ระหว่าง 0.0037 ถึง 0.0143 โดยการวิเคราะห์ด้วยรูปแบบที่ 4 Numerical Bayesian GIRM ให้ค่าค่าส่วนเบี่ยงเบนมาตรฐานรวมขององค์ประกอบความแปรปรวนตามทฤษฎีการอ้างอิงความน่าเชื่อของผลการวัดต่ำสุดในทุกกรณีทั้ง 4 กรณีที่รูปแบบที่ 4 สามารถประมาณค่าองค์ประกอบความแปรปรวนได้ รองลงมารูปแบบที่ 1 Original GIRM รูปแบบที่ 3 AGIRM A และรูปแบบที่ 2 AGIRM B ตามลำดับ

3.3 **การวิเคราะห์ค่าประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนด้วยการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด** (Euclidean Distance: EUC)

ค่าประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนในส่วนนี้จะแยกการนำเสนอตามรูปแบบการประมาณค่าทั้ง 4 รูปแบบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GIRM) ทั้ง 4 รูปแบบและจำแนกตามเงื่อนไขของการวิเคราะห์ทั้ง 16 กรณี มีรายละเอียดของผลการวิเคราะห์ดังนี้

ตารางที่ 17 ผลการวิเคราะห์ค่าประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนจำแนกตามรูปแบบการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบทั้ง 16 รายกรณี

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Original GIRM	AGIRM A	AGIRM B	Numerical GIRM
Normal	Normal	300	5	0.0489	0.0485	0.0488	0.0639
		300	10	0.0406	0.0399	0.0404	0.0638
		100	5	0.0495	0.0485	0.0488	0.0614
		100	10	0.0375	0.0372	0.0376	0.0644
	Beta	300	5	0.0497	0.0494	0.0498	NA
		300	10	0.0407	0.0403	0.0407	NA
		100	5	0.0499	0.0490	0.0494	NA
		100	10	0.0395	0.0392	0.0397	NA
Gamma	Normal	300	5	0.0461	0.0453	0.0457	NA
		300	10	0.0406	0.0400	0.0405	NA
		100	5	0.0488	0.0442	0.0446	NA
		100	10	0.0400	0.0399	0.0403	NA
	Beta	300	5	0.0534	0.0496	0.0500	NA
		300	10	0.0441	0.0435	0.0439	NA
		100	5	0.0564	0.0501	0.0505	NA
		100	10	0.0447	0.0439	0.0444	NA

จากผลการวิเคราะห์การวิเคราะห์ค่าประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนด้วยการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) พบว่า มีค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด อยู่ระหว่าง 0.0375 ถึง 0.0644 โดยการวิเคราะห์ด้วยรูปแบบที่ 2 AGIRM A ให้ค่าค่าองค์ประกอบความแปรปรวนระยะทางยุคลิดต่ำสุดในทุกกรณี รองลงมารูปแบบที่ 3 AGIRM A รูปแบบที่ 1 Original GIRM และ รูปแบบที่ 4 Numerical Bayesian GIRM ตามลำดับ

ตอนที่ 4 การวิเคราะห์ความแตกต่างของขนาดกลุ่มตัวอย่างและจำนวนข้อสอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์

การวิเคราะห์ในตอนนี้มุ่งศึกษาอิทธิพลของขนาดกลุ่มอย่างและจำนวนข้อสอบที่มีต่อประสิทธิภาพของวิธีวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทั้ง 4 รูปแบบการประมาณค่า ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM ซึ่ง พัฒนาโดย Brigg และ Wilson (2007) และรูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B รูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งพัฒนาวิธีการประมาณค่าพารามิเตอร์โดยผู้วิจัย โดยเปรียบเทียบรูปแบบการประมาณค่าในแต่ละรูปแบบทั้ง 3 ดัชนีวัดประสิทธิภาพของการประมาณค่าพารามิเตอร์ ซึ่งได้แก่ ความลำเอียงในการประมาณค่า (Biased estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ค่าความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ดังนั้นจะเห็นได้ว่า การวิเคราะห์ในตอนนี้จะมุ่งตอบคำถามการวิจัยและวัตถุประสงค์ในการวิจัยข้อที่ 2 เกี่ยวกับอิทธิพลของขนาดกลุ่มตัวอย่างและความยาวของแบบสอบที่ส่งผลต่อประสิทธิภาพในการประมาณค่าของวิธีการ GIRM

4.1 การวิเคราะห์ความแตกต่างของขนาดกลุ่มตัวอย่างที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์

การวิเคราะห์ในตอนนี้มุ่งศึกษาอิทธิพลของขนาดกลุ่มอย่างที่มีต่อประสิทธิภาพของวิธีวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทั้ง 4 รูปแบบการประมาณ โดยเปรียบเทียบรูปแบบการประมาณค่าในแต่ละรูปแบบทั้ง 3 ดัชนีวัดประสิทธิภาพของการประมาณค่าพารามิเตอร์ มีรายละเอียดการนำเสนอดังนี้

ตารางที่ 18 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีขนาดกลุ่มตัวอย่างแตกต่างกัน

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
รูปแบบที่ 1 Original GIRM								
Normal	Normal	300	5	500	0.6767	0.1067	0.8023	0.4227
		100	5	500	0.6653	0.2995		
Normal	Normal	300	10	500	0.6058	0.0922	7.8313	0.0000**
		100	10	500	0.5333	0.1855		
Normal	Beta	300	5	500	0.7045	0.0880	3.5095	0.0004**
		100	5	500	0.6809	0.1221		
Normal	Beta	300	10	500	0.6363	0.0862	8.5082	0.0000**
		100	10	500	0.5840	0.1068		
Gamma	Normal	300	5	500	0.6772	0.5857	2.5385	0.0114*
		100	5	500	0.6086	0.1483		
Gamma	Normal	300	10	500	0.6042	0.0687	15.1935	0.0000**
		100	10	500	0.5163	0.1094		
Gamma	Beta	300	5	500	0.6863	0.0680	1.7220	0.0855
		100	5	500	0.6704	0.1949		
Gamma	Beta	300	10	500	0.6444	0.0656	9.1058	0.0000**
		100	10	500	0.5987	0.0912		
รูปแบบที่ 2 AGIRM A								
Normal	Normal	300	5	500	0.5873	0.1896	-3.0344	0.0025**
		100	5	500	0.6719	0.5931		
Normal	Normal	300	10	500	0.4833	0.1198	2.3418	0.0194*
		100	10	500	0.4563	0.2563		
Normal	Beta	300	5	500	0.6142	0.1503	-1.4127	0.1582
		100	5	500	0.6428	0.4265		
Normal	Beta	300	10	500	0.5164	0.1199	5.2987	0.0000**
		100	10	500	0.4729	0.1389		
Gamma	Normal	300	5	500	0.6219	1.3627	0.1317	0.8952
		100	5	500	0.6133	0.5229		

ตารางที่ 18 (ต่อ)

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
Gamma	Normal	300	10	500	0.4867	0.1005	5.4784	0.0000**
		100	10	500	0.4380	0.1714		
Gamma	Beta	300	5	500	0.6044	0.1386	-1.508	0.1319
		100	5	500	0.6469	0.6148		
Gamma	Beta	300	10	500	0.5349	0.0847	4.4893	0.0000**
		100	10	500	0.5034	0.1315		
รูปแบบที่ 3 AGIRM B								
Normal	Normal	300	5	500	0.6853	0.1460	-2.604	0.0094**
		100	5	500	0.7402	0.4481		
Normal	Normal	300	10	500	0.6026	0.0984	4.5894	0.0000**
		100	10	500	0.5533	0.2042		
Normal	Beta	300	5	500	0.7092	0.1131	-1.124	0.2613
		100	5	500	0.7261	0.3169		
Normal	Beta	300	10	500	0.6326	0.0955	6.4426	0.0000**
		100	10	500	0.5893	0.1159		
Gamma	Normal	300	5	500	0.7096	1.1448	0.5527	0.5807
		100	5	500	0.6798	0.3810		
Gamma	Normal	300	10	500	0.6000	0.0832	9.2787	0.0000**
		100	10	500	0.5345	0.1340		
Gamma	Beta	300	5	500	0.6940	0.1055	-1.041	0.2981
		100	5	500	0.7167	0.4755		
Gamma	Beta	300	10	500	0.6402	0.0746	6.4463	0.0000**
		100	10	500	0.6008	0.1141		
รูปแบบที่ 4 Numerical Bayesian GIRM								
Normal	Normal	300	5	500	0.6988	0.4921	-0.715	0.4746
		100	5	500	0.7263	0.7053		
Normal	Normal	300	10	500	0.6686	0.1491	-0.617	0.5268
		100	10	500	0.6810	0.4223		

NA = ไม่สามารถทำการวิเคราะห์ได้

จากการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีขนาดกลุ่มตัวอย่างแตกต่างกัน ทั้ง 26 คู่กรณี พบว่า มีคู่กรณีที่มีความแตกต่างกันของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ยอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิ้น 16 กรณี คิดเป็นร้อยละ 61.53 โดยแบ่งเป็น รูปแบบที่ 1 ทั้งสิ้น 6 คู่กรณี รูปแบบที่ 2 ทั้งสิ้น 5 คู่กรณี และรูปแบบที่ 3 ทั้งสิ้น 5 คู่กรณี จากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ของกรณีศึกษาจำนวนขนาดของกลุ่มตัวอย่างส่งผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์เมื่อพิจารณาจากค่า ความคลาดเคลื่อนเฉลี่ย (MAD) ใน 3 รูปแบบการประมาณค่า ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 โดยขนาดกลุ่มตัวอย่างมากขึ้นจะยิ่งช่วยลดความลำเอียงในการประมาณค่าพารามิเตอร์ ส่วนรูปแบบที่ 4 Numerical Bayesian GIRM ขนาดตัวอย่างไม่มีผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์

ตารางที่ 19 ผลการวิเคราะห์ความแตกต่างของส่วนเบี่ยงเบนมาตรฐานในการประมาณค่าองค์ประกอบความแปรปรวน (Standard Deviation: S.D.) ระหว่างกรณีศึกษาที่มีขนาดตัวอย่างแตกต่างกัน

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
รูปแบบที่ 1 Original GIRM									
Normal	Normal	300	5	0.0120	0.0007	0.0000	0.0140	0.0140	0.0092
		100	5	0.0132	0.0019	0.0000	0.0153	0.0153	0.0101
Normal	Normal	300	10	0.0109	0.0007	0.0000	0.0113	0.0113	0.0078
		100	10	0.0114	0.0022	0.0001	0.0120	0.0119	0.0083
Normal	Beta	300	5	0.0123	0.0005	0.0000	0.0138	0.0138	0.0092
		100	5	0.0136	0.0010	0.0000	0.0149	0.0149	0.0100
Normal	Beta	300	10	0.0110	0.0006	0.0000	0.0114	0.0114	0.0079
		100	10	0.0114	0.0010	0.0000	0.0118	0.0118	0.0082
Gamma	Normal	300	5	0.0041	0.0007	0.0000	0.0083	0.0083	0.0046
		100	5	0.0043	0.0017	0.0010	0.0087	0.0086	0.0049
Gamma	Normal	300	10	0.0056	0.0007	0.0000	0.0064	0.0064	0.0042
		100	10	0.0057	0.0020	0.0000	0.0069	0.0068	0.0045

ตารางที่ 19 (ต่อ)

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
Gamma	Beta	300	5	0.0036	0.0006	0.0000	0.0060	0.0060	0.0037
		100	5	0.0036	0.0011	0.0000	0.0067	0.0067	0.0038
Gamma	Beta	300	10	0.0049	0.0007	0.0000	0.0055	0.0055	0.0035
		100	10	0.0047	0.0012	0.0000	0.0058	0.0058	0.0037
รูปแบบที่ 2 AGIRM A									
Normal	Normal	300	5	0.0134	0.0014	NA	NA	0.0000	0.0117
		100	5	0.0163	0.0039	NA	NA	0.0184	0.1430
Normal	Normal	300	10	0.0110	0.0011	NA	NA	0.0115	0.0092
		100	10	0.0123	0.0034	NA	NA	0.0129	0.0105
Normal	Beta	300	5	0.0133	0.0011	NA	NA	0.0146	0.0114
		100	5	0.0162	0.0026	NA	NA	0.0172	0.0137
Normal	Beta	300	10	0.0115	0.0009	NA	NA	0.0119	0.0095
		100	10	0.0126	0.0021	NA	NA	0.0130	0.0105
Gamma	Normal	300	5	0.0065	0.0015	NA	NA	0.0100	0.0069
		100	5	0.0103	0.0043	NA	NA	0.0127	0.0097
Gamma	Normal	300	10	0.0065	0.0012	NA	NA	0.0074	0.0057
		100	10	0.0076	0.0032	NA	NA	0.0090	0.0070
Gamma	Beta	300	5	0.0063	0.0013	NA	NA	0.0083	0.0060
		100	5	0.0093	0.0020	NA	NA	0.0114	0.0086
Gamma	Beta	300	10	0.0058	0.0010	NA	NA	0.0062	0.0049
		100	10	0.0071	0.0023	NA	NA	0.0083	0.0064
รูปแบบที่ 3 AGIRM B									
Normal	Normal	300	5	0.0222	0.0013	0.0000	0.0301	0.0130	0.0101
		100	5	0.0220	0.0038	0.0001	0.0325	0.0325	0.0124
Normal	Normal	300	10	0.0292	0.0022	0.0001	0.0327	0.0327	0.0079
		100	10	0.0123	0.0034	0.0001	0.0130	0.0129	0.0091
Normal	Beta	300	5	0.0133	0.0011	0.0000	0.0146	0.0146	0.0098
		100	5	0.0162	0.0026	0.0000	0.0172	0.0172	0.0118

ตารางที่ 19 (ต่อ)

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
Normal	Beta	300	10	0.0125	0.0009	0.0000	0.0119	0.0119	0.0082
		100	10	0.0126	0.0021	0.0000	0.0130	0.0130	0.0091
Gamma	Normal	300	5	0.0065	0.0015	0.0000	0.0100	0.0100	0.0060
		100	5	0.0103	0.0043	0.0000	0.0127	0.0127	0.0084
Gamma	Normal	300	10	0.0065	0.0012	0.0000	0.0074	0.0074	0.0049
		100	10	0.0076	0.0032	0.0001	0.0091	0.0091	0.0061
Gamma	Beta	300	5	0.0063	0.0013	0.0000	0.0083	0.0083	0.0052
		100	5	0.0093	0.0026	0.0000	0.0114	0.0114	0.0074
Gamma	Beta	300	10	0.0058	0.0010	0.0000	0.0062	0.0062	0.0042
		100	10	0.0071	0.0023	0.0000	0.0083	0.0083	0.0055
รูปแบบที่ 4 Numerical Bayesian GIRM									
Normal	Normal	300	5	0.0103	0.0007	0.0000	0.0301	0.0301	0.0087
		100	5	0.0120	0.0028	0.0001	0.0125	0.0125	0.0097
Normal	Normal	300	10	0.0092	0.0011	0.0001	0.0127	0.0127	0.0065
		100	10	0.0091	0.0024	0.0001	0.0156	0.0154	0.0074

NA = ไม่สามารถทำการวิเคราะห์ได้

จากการวิเคราะห์ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ที่คำนวณได้จาก การวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) พบว่า เมื่อเปรียบเทียบค่าส่วนเบี่ยงเบนมาตรฐานระหว่างกรณีที่จำนวนขนาดตัวอย่างเพิ่มขึ้น ในขณะที่จำนวนข้อสอบเท่ากัน จะให้ค่าส่วนเบี่ยงเบนมาตรฐานที่ต่ำกว่าในทุกกรณี และทุกรูปแบบการประมาณค่าทั้ง 4 รูปแบบ แสดงให้เห็นว่า จำนวนขนาดกลุ่มตัวอย่าง ส่งผลต่อความไม่แน่นอนในประมาณค่าองค์ประกอบความแปรปรวน โดยขนาดกลุ่มตัวอย่างเพิ่มขึ้นจะส่งผลให้การประมาณค่าองค์ประกอบความแปรปรวนมีความแน่นอนในประมาณค่าเพิ่มมากขึ้น

ตารางที่ 20 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีขนาดตัวอย่างแตกต่างกัน

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
รูปแบบที่ 1 Original GIRM								
Normal	Normal	300	5	500	0.0489	0.0152	1.3280	0.1845
		100	5	500	0.0495	0.0167		
Normal	Normal	300	10	500	0.0406	0.0135	3.3768	0.0007**
		100	10	500	0.0376	0.0148		
Normal	Beta	300	5	500	0.0497	0.0155	0.8103	0.4180
		100	5	500	0.0499	0.0164		
Normal	Beta	300	10	500	0.0408	0.0140	1.4024	0.1611
		100	10	500	0.0395	0.0148		
Gamma	Normal	300	5	500	0.0462	0.0091	2.1483	0.0319*
		100	5	500	0.0448	0.0106		
Gamma	Normal	300	10	500	0.0407	0.0097	1.1440	0.2529
		100	10	500	0.0400	0.0097		
Gamma	Beta	300	5	500	0.0534	0.0081	-0.629	0.5294
		100	5	500	0.0564	0.0095		
Gamma	Beta	300	10	500	0.0442	0.0086	-0.962	0.3359
		100	10	500	0.0447	0.0100		
รูปแบบที่ 2 AGIRM A								
Normal	Normal	300	5	500	0.0485	0.0163	0.0027	0.9979
		100	5	500	0.0485	0.0194		
Normal	Normal	300	10	500	0.0400	0.0136	2.9183	0.0035**
		100	10	500	0.0372	0.0158		
Normal	Beta	300	5	500	0.0494	0.0163	0.3635	0.7163
		100	5	500	0.0490	0.0192		
Normal	Beta	300	10	500	0.0403	0.0145	1.0669	0.2863
		100	10	500	0.0393	0.0154		

ตารางที่ 20 (ต่อ)

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
Gamma	Normal	300	5	500	0.0454	0.0112	1.2600	0.2080
		100	5	500	0.0443	0.0151		
Gamma	Normal	300	10	500	0.0401	0.0102	0.2484	0.8039
		100	10	500	0.0399	0.0120		
Gamma	Beta	300	5	500	0.0496	0.0107	-0.618	0.5361
		100	5	500	0.0501	0.0149		
Gamma	Beta	300	10	500	0.0435	0.0092	-0.661	0.5088
		300	10	500	0.0440	0.0093		
รูปแบบที่ 3 AGIRM B								
Normal	Normal	300	5	500	0.0489	0.0164	0.0165	0.9868
		100	5	500	0.0489	0.0195		
Normal	Normal	300	10	500	0.0404	0.0137	2.9459	0.0032**
		100	10	500	0.0377	0.0159		
Normal	Beta	300	5	500	0.0498	0.0164	0.3666	0.7140
		100	5	500	0.0494	0.0193		
Normal	Beta	300	10	500	0.0408	0.0145	1.0721	0.2839
		100	10	500	0.0397	0.0155		
Gamma	Normal	300	5	500	0.0457	0.0113	1.3041	0.1925
		100	5	500	0.0446	0.0152		
Gamma	Normal	300	10	500	0.0405	0.0103	0.3052	0.7603
		100	10	500	0.0403	0.0120		
Gamma	Beta	300	5	500	0.0500	0.0108	-0.596	0.5510
		100	5	500	0.0505	0.0150		
Gamma	Beta	300	10	500	0.0440	0.0093	-0.643	0.5204
		100	10	500	0.0444	0.0127		
รูปแบบที่ 4 Numerical Bayesian GIRM								
Normal	Normal	300	5	500	0.0634	0.0220	0.7486	0.4543
		100	5	500	0.0623	0.0240		
Normal	Normal	300	10	500	0.0635	0.0256	-0.437	0.6617
		100	10	500	0.0643	0.0285		

จากการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีขนาดกลุ่มตัวอย่างแตกต่างกัน ทั้ง 26 คู่กรณี พบว่า มีคู่กรณีที่มีความแตกต่างกันของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิ้น 4 กรณี คิดเป็นร้อยละ 15.38 โดยแบ่งเป็น รูปแบบที่ 1 ทั้งสิ้น 2 คู่กรณี รูปแบบที่ 2 ทั้งสิ้น 1 คู่กรณี และรูปแบบที่ 3 ทั้งสิ้น 1 คู่กรณี จากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ของกรณีศึกษาจำนวนขนาดของกลุ่มตัวอย่างไม่ส่งผลต่อประสิทธิภาพในประมาณค่าองค์ประกอบความแปรปรวนเมื่อพิจารณาจากค่า องค์ประกอบความแปรปรวนระยะทางยุคลิด (EUC) อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

4.2 การวิเคราะห์ความแตกต่างของความยาวของแบบสอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์

การวิเคราะห์ในตอนนี้มุ่งศึกษาอิทธิพลของความยาวของแบบสอบที่มีต่อประสิทธิภาพของวิธีวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทั้ง 4 รูปแบบการประมาณค่า โดยเปรียบเทียบรูปแบบการประมาณค่าในแต่ละรูปแบบทั้ง 3 ดัชนีวัดประสิทธิภาพของการประมาณค่าพารามิเตอร์ ซึ่งได้แก่ ความลำเอียงในการประมาณค่า (Biased estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ค่าความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) โดยมีรายละเอียดการนำเสนอดังนี้

ตารางที่ 21 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีจำนวนข้อแตกต่างกัน

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
รูปแบบที่ 1 Original GIRM								
Normal	Normal	300	5	500	0.6768	0.1068	11.2355	0.0000**
		300	10	500	0.6059	0.0922		
Normal	Normal	100	5	500	0.6654	0.2995	8.3806	0.0000**
		100	10	500	0.5333	0.1855		

ตารางที่ 21 (ต่อ)

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
Normal	Beta	300	5	500	0.7046	0.0880	12.3851	0.0000**
		300	10	500	0.6363	0.0863		
Normal	Beta	100	5	500	0.6810	0.1221	13.3543	0.0000**
		100	10	500	0.5840	0.1069		
Gamma	Normal	300	5	500	0.6772	0.5858	2.7675	0.0058**
		300	10	500	0.6042	0.0687		
Gamma	Normal	100	5	500	0.6086	0.1483	11.1875	0.0000**
		100	10	500	0.5164	0.1095		
Gamma	Beta	300	5	500	0.6863	0.0681	9.8885	0.0000**
		300	10	500	0.6445	0.0656		
Gamma	Beta	100	5	500	0.6704	0.1949	7.4472	0.0000**
		100	10	500	0.5987	0.0912		
รูปแบบที่ 2 AGIRM A								
Normal	Normal	300	5	500	0.5874	0.1897	10.3714	0.0000**
		300	10	500	0.4833	0.1198		
Normal	Normal	100	5	500	0.6719	0.5931	7.5516	0.0000**
		100	10	500	0.4537	0.2563		
Normal	Beta	300	5	500	0.6143	0.1503	11.3725	0.0000**
		300	10	500	0.5165	0.1200		
Normal	Beta	100	5	500	0.6428	0.4265	8.4688	0.0000**
		100	10	500	0.4730	0.1389		
Gamma	Normal	300	5	500	0.6219	1.3628	2.2121	0.0274*
		300	10	500	0.4867	0.1005		
Gamma	Normal	100	5	500	0.6133	0.5230	7.1210	0.0000**
		100	10	500	0.4380	0.1714		
Gamma	Beta	300	5	500	0.6044	0.1387	9.5633	0.0000**
		300	10	500	0.5349	0.0847		
Gamma	Beta	100	5	500	0.6469	0.6149	5.1010	0.0000**
		100	10	500	0.5035	0.1315		

ตารางที่ 21 (ต่อ)

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
รูปแบบที่ 3 AGIRM B								
Normal	Normal	300	5	500	0.6854	0.1460	10.4975	0.0000**
		300	10	500	0.6027	0.0985		
Normal	Normal	100	5	500	0.7402	0.4481	8.4842	0.0000**
		100	10	500	0.5534	0.2043		
Normal	Beta	300	5	500	0.7092	0.1132	11.5606	0.0000**
		300	10	500	0.6326	0.0955		
Normal	Beta	100	5	500	0.7261	0.3169	9.0622	0.0000**
		100	10	500	0.5894	0.1159		
Gamma	Normal	300	5	500	0.7096	1.1449	2.1344	0.0332*
		300	10	500	0.6001	0.0832		
Gamma	Normal	100	5	500	0.6798	0.3810	8.0394	0.0000**
		100	10	500	0.5346	0.1341		
Gamma	Beta	300	5	500	0.6940	0.1056	9.2983	0.0000**
		300	10	500	0.6403	0.0706		
Gamma	Beta	100	5	500	0.7167	0.4755	5.2981	0.0000**
		100	10	500	0.6008	0.1141		
รูปแบบที่ 4 Numerical Bayesian GIRM								
Normal	Normal	300	5	500	0.6988	0.4921	1.313	0.1895
		300	10	500	0.6686	0.1491		
Normal	Normal	100	5	500	0.7263	0.7053	1.233	0.2179
		100	10	500	0.6810	0.4223		

NA = ไม่สามารถทำการวิเคราะห์ได้

จากการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีจำนวนข้อสอบที่แตกต่างกัน ทั้ง 26 คู่กรณี พบว่า มีคู่กรณีที่มีความแตกต่างกันของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ยอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิ้น 24 กรณี คิดเป็นร้อยละ 92.31 โดยแบ่งเป็น รูปแบบที่ 1 รูปแบบที่ 2

และรูปแบบที่ 3 ทุกกรณี (รูปแบบละ 8 คู่กรณี) จากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ของกรณีศึกษาจำนวนข้อสอบส่งผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์เมื่อพิจารณาจากค่าความคลาดเคลื่อนเฉลี่ย (MAD) ใน 3 รูปแบบการประมาณค่า ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 โดยจำนวนข้อสอบมากขึ้นจะยิ่งช่วยลดความลำเอียงในการประมาณค่าพารามิเตอร์ ส่วนรูปแบบที่ 4 Numerical Bayesian GIRM จำนวนข้อสอบไม่มีผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์

ตารางที่ 22 ผลการวิเคราะห์ความแตกต่างของส่วนเบี่ยงเบนมาตรฐานในการประมาณค่าองค์ประกอบความแปรปรวน (Standard Deviation: S.D.) ระหว่างกรณีศึกษาที่มีจำนวนข้อแตกต่างกัน

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
รูปแบบที่ 1 Original GIRM									
Normal	Normal	300	5	0.0120	0.0007	0.0000	0.0140	0.0140	0.0092
		300	10	0.0109	0.0007	0.0000	0.0113	0.0113	0.0078
Normal	Normal	100	5	0.0132	0.0019	0.0000	0.0153	0.0153	0.0101
		100	10	0.0114	0.0022	0.0001	0.0120	0.0119	0.0083
Normal	Beta	300	5	0.0123	0.0005	0.0000	0.0138	0.0138	0.0092
		300	10	0.0110	0.0006	0.0000	0.0114	0.0114	0.0079
Normal	Beta	100	5	0.0136	0.0010	0.0000	0.0149	0.0149	0.0100
		100	10	0.0114	0.0010	0.0000	0.0119	0.0118	0.0082
Gamma	Normal	300	5	0.0041	0.0007	0.0000	0.0083	0.0083	0.0046
		300	10	0.0056	0.0007	0.0000	0.0064	0.0064	0.0042
Gamma	Normal	100	5	0.0043	0.0017	0.0000	0.0086	0.0086	0.0049
		100	10	0.0057	0.0020	0.0001	0.0067	0.0068	0.0045
Gamma	Beta	300	5	0.0036	0.0006	0.0000	0.0060	0.0060	0.0037
		300	10	0.0049	0.0007	0.0000	0.0055	0.0055	0.0035
Gamma	Beta	100	5	0.0036	0.0011	0.0000	0.0067	0.0067	0.0038
		100	10	0.0047	0.0012	0.0000	0.0058	0.0058	0.0037

ตารางที่ 22 (ต่อ)

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
รูปแบบที่ 2 AGIRM A									
Normal	Normal	300	5	0.0134	0.0014	NA	NA	0.0151	0.0117
		300	10	0.0110	0.0011	NA	NA	0.0115	0.0092
Normal	Normal	100	5	0.0163	0.0039	NA	NA	0.0184	0.0143
		100	10	0.0123	0.0034	NA	NA	0.0129	0.0105
Normal	Beta	300	5	0.0133	0.0011	NA	NA	0.0146	0.0114
		300	10	0.0115	0.0009	NA	NA	0.0119	0.0095
Normal	Beta	100	5	0.0162	0.0026	NA	NA	0.0172	0.0137
		100	10	0.0126	0.0021	NA	NA	0.0130	0.0105
Gamma	Normal	300	5	0.0065	0.0015	NA	NA	0.0100	0.0069
		300	10	0.0065	0.0012	NA	NA	0.0074	0.0057
Gamma	Normal	100	5	0.0103	0.0043	NA	NA	0.0127	0.0097
		100	10	0.0076	0.0032	NA	NA	0.0090	0.0070
Gamma	Beta	300	5	0.0063	0.0013	NA	NA	0.0083	0.0060
		300	10	0.0058	0.0010	NA	NA	0.0062	0.0049
Gamma	Beta	100	5	0.0093	0.0026	NA	NA	0.0114	0.0086
		100	10	0.0071	0.0023	NA	NA	0.0083	0.0064
รูปแบบที่ 3 AGIRM B									
Normal	Normal	300	5	0.0134	0.0014	0.0000	0.0151	0.0151	0.0101
		300	10	0.0110	0.0011	0.0000	0.0115	0.0115	0.0079
Normal	Normal	100	5	0.0163	0.0039	0.0000	0.0184	0.0184	0.0124
		100	10	0.0123	0.0034	0.0001	0.0129	0.0129	0.0091
Normal	Beta	300	5	0.0133	0.0011	0.0000	0.0146	0.0146	0.0098
		300	10	0.0115	0.0009	0.0000	0.0119	0.0119	0.0082
Normal	Beta	100	5	0.0162	0.0026	0.0000	0.0172	0.0172	0.0118
		100	10	0.0126	0.0021	0.0000	0.0130	0.0130	0.0091
Gamma	Normal	300	5	0.0065	0.0015	0.0000	0.0100	0.0100	0.0060
		300	10	0.0065	0.0012	0.0000	0.0074	0.0074	0.0049

ตารางที่ 22 (ต่อ)

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
Gamma	Normal	100	5	0.0103	0.0043	0.0000	0.0127	0.0127	0.0084
		100	10	0.0076	0.0032	0.0001	0.0091	0.0090	0.0061
Gamma	Beta	300	5	0.0063	0.0013	0.0000	0.0083	0.0083	0.0052
		300	10	0.0058	0.0010	0.0000	0.0062	0.0062	0.0042
Gamma	Beta	100	5	0.0093	0.0026	0.0000	0.0114	0.0114	0.0074
		100	10	0.0071	0.0023	0.0000	0.0083	0.0083	0.0055
รูปแบบที่ 4 Numerical Bayesian GIRM									
Normal	Normal	300	5	0.0103	0.0007	0.0000	0.0301	0.0301	0.0087
		300	10	0.0092	0.0011	0.0001	0.0127	0.0127	0.0065
Normal	Normal	100	5	0.0120	0.0028	0.0001	0.0125	0.0125	0.0097
		100	10	0.0091	0.0024	0.0001	0.0156	0.0154	0.0074

NA = ไม่สามารถทำการวิเคราะห์ได้

จากการวิเคราะห์ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ที่คำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) พบว่า เมื่อเปรียบเทียบค่าส่วนเบี่ยงเบนมาตรฐานระหว่างกรณีที่จำนวนข้อสอบเพิ่มขึ้น ในขณะที่จำนวนขนาดกลุ่มตัวอย่างเท่ากัน จะให้ค่าส่วนเบี่ยงเบนมาตรฐานที่ต่ำกว่าในทุกกรณี และทุกรูปแบบการประมาณค่าทั้ง 4 รูปแบบ แสดงให้เห็นว่า จำนวนข้อสอบ ส่งผลต่อความไม่แน่นอนในประมาณค่าองค์ประกอบความแปรปรวน โดยจำนวนข้อสอบเพิ่มขึ้น จะส่งผลให้การประมาณค่าองค์ประกอบความแปรปรวนมีความแน่นอนในประมาณค่าเพิ่มมากขึ้น

ตารางที่ 23 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีจำนวนข้อแตกต่างกัน

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
รูปแบบที่ 1 Original GIRM								
Normal	Normal	300	5	500	0.0489	0.0152	9.1593	0.0000**
		300	10	500	0.0406	0.0135		
Normal	Normal	100	5	500	0.0495	0.0167	10.016	0.0000**
		100	10	500	0.0376	0.0148		
Normal	Beta	300	5	500	0.0497	0.0155	9.5918	0.0000**
		300	10	500	0.0408	0.0140		
Normal	Beta	100	5	500	0.0499	0.0164	9.5210	0.0000**
		100	10	500	0.0395	0.0148		
Gamma	Normal	300	5	500	0.0462	0.0091	9.5148	0.0000**
		300	10	500	0.0407	0.0091		
Gamma	Normal	100	5	500	0.0448	0.0106	7.4723	0.0000**
		100	10	500	0.0400	0.0097		
Gamma	Beta	300	5	500	0.0534	0.0081	9.9437	0.0000**
		300	10	500	0.0442	0.0086		
Gamma	Beta	100	5	500	0.0564	0.0095	8.1220	0.0000**
		100	10	500	0.0447	0.0100		
รูปแบบที่ 2 AGIRM A								
Normal	Normal	300	5	500	0.0485	0.0163	8.9763	0.0000**
		300	10	500	0.0400	0.0136		
Normal	Normal	100	5	500	0.0485	0.0194	10.061	0.0000**
		100	10	500	0.0372	0.0158		
Normal	Beta	300	5	500	0.0494	0.0163	9.3547	0.0000**
		300	10	500	0.0403	0.0145		
Normal	Beta	100	5	500	0.0490	0.0192	8.8541	0.0000**
		100	10	500	0.0393	0.0154		

ตารางที่ 23 (ต่อ)

การแจกแจงเริ่มแรกผู้สอบ	การแจกแจงเริ่มแรกข้อสอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
Gamma	Normal	300	5	500	0.0454	0.0112	7.7580	0.0000**
		300	10	500	0.0401	0.0102		
Gamma	Normal	100	5	500	0.0443	0.0151	5.0785	0.0000**
		100	10	500	0.0399	0.0120		
Gamma	Beta	300	5	500	0.0496	0.0107	9.7144	0.0000**
		300	10	500	0.0435	0.0092		
Gamma	Beta	100	5	500	0.0501	0.0149	7.0796	0.0000**
		100	10	500	0.0440	0.0125		
รูปแบบที่ 3 AGIRM B								
Normal	Normal	300	5	500	0.0489	0.0164	8.8667	0.0000**
		300	10	500	0.0404	0.0137		
Normal	Normal	100	5	500	0.0489	0.0195	9.9728	0.0000**
		100	10	500	0.0377	0.0159		
Normal	Beta	300	5	500	0.0498	0.0164	9.2475	0.0000**
		300	10	500	0.0408	0.0145		
Normal	Beta	100	5	500	0.0494	0.0193	8.7599	0.0000**
		100	10	500	0.0397	0.0155		
Gamma	Normal	300	5	500	0.0457	0.0113	7.5973	0.0000**
		300	10	500	0.0405	0.0103		
Gamma	Normal	100	5	500	0.0446	0.0152	4.9738	0.0000**
		100	10	500	0.0403	0.0120		
Gamma	Beta	300	5	500	0.0500	0.0108	9.5441	0.0000**
		300	10	500	0.0440	0.0093		
Gamma	Beta	100	5	500	0.0505	0.0150	6.9705	0.0000**
		100	10	500	0.0444	0.0127		
รูปแบบที่ 4 Numerical Bayesian GIRM								
Normal	Normal	300	5	500	0.0634	0.0220	-0.078	0.9371
		300	10	500	0.0635	0.0256		
Normal	Normal	100	5	500	0.0623	0.0240	-1.176	0.2397
		100	10	500	0.0643	0.0285		

จากการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีขนาดกลุ่มตัวอย่างแตกต่างกัน ทั้งสิ้น 24 กรณี คิดเป็นร้อยละ 92.31 โดยแบ่งเป็น รูปแบบที่ 1 รูปแบบที่ 2 และรูปแบบที่ 3 ทุกกรณี (รูปแบบละ 8 คู่กรณี) จากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ของกรณีศึกษาจำนวนข้อสอบส่งผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์เมื่อพิจารณาจากค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (EUC) ใน 3 รูปแบบการประมาณค่า ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 โดยจำนวนข้อสอบมากขึ้นจะยิ่งช่วยเพิ่มประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ส่วนรูปแบบที่ 4 Numerical Bayesian GIRM จำนวนข้อสอบไม่มีผลต่อประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน

ตอนที่ 5 การวิเคราะห์ความแตกต่างของการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบและค่าพารามิเตอร์ของข้อสอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์

การวิเคราะห์ในตอนนี้มุ่งศึกษาความไว (Sensitivity) ของผลการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทั้ง 4 รูปแบบการประมาณค่า ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM ซึ่ง พัฒนาโดย Brigg และ Wilson (2007) และรูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B รูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งพัฒนาวิธีการประมาณค่าของวิธีการโดยผู้วิจัย ที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) โดยจะเปรียบเทียบรูปแบบการประมาณค่าในแต่ละรูปแบบทั้ง 3 ดัชนีวัดประสิทธิภาพของการประมาณค่าพารามิเตอร์ ซึ่งได้แก่ ความลำเอียงในการประมาณค่า (Biased estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ค่าความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ดังนั้นจะเห็นได้ว่า การวิเคราะห์ในตอนนี้จะมุ่งตอบคำถามการวิจัยและวัตถุประสงค์ในการวิจัยข้อที่ 3 โดยมีรายละเอียดการนำเสนอดังนี้

6.1 ความไว (Sensitivity) ของผลการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์

ในการนำเสนอผลการวัดประสิทธิภาพในการประมาณค่าพารามิเตอร์ของการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ จะนำเสนอแยกออกเป็น 3 ตารางในแต่ละดัชนีวัดประสิทธิภาพของการประมาณค่าพารามิเตอร์ ซึ่งได้แก่ ความลำเอียงในการประมาณค่า (Biased estimator) คำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) จำแนกตามเงื่อนไขของการวิเคราะห์ทั้ง 16 กรณี มีรายละเอียดการนำเสนอ ดังนี้

ตารางที่ 24 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแตกต่างกัน

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
รูปแบบที่ 1 Original GIRM								
Normal	Normal	300	5	500	0.6768	0.1068	-0.0160	0.987
	Gamma	300	5	500	0.6772	0.5858		
Normal	Normal	300	10	500	0.6059	0.0922	0.3230	0.7468
	Gamma	300	10	500	0.6042	0.0687		
Normal	Normal	100	5	500	0.6654	0.2995	3.7966	0.0001**
	Gamma	100	5	500	0.6086	0.1483		
Normal	Normal	100	10	500	0.5333	0.1855	1.7577	0.0791
	Gamma	100	10	500	0.5164	0.1095		
Beta	Normal	300	5	500	0.7046	0.0880	3.6725	0.0002**
	Gamma	300	5	500	0.6863	0.0681		
Beta	Normal	300	10	500	0.6363	0.0863	-1.6870	0.0918
	Gamma	300	10	500	0.6445	0.0656		

ตารางที่ 24 (ต่อ)

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
Beta	Normal	100	5	500	0.6810	0.1221	1.0258	0.3053
	Gamma	100	5	500	0.6704	0.1949		
Beta	Normal	100	10	500	0.5840	0.1069	-2.3370	0.0196 *
	Gamma	100	10	500	0.5987	0.0912		
รูปแบบที่ 2 AGIRM A								
Normal	Normal	300	5	500	0.5874	0.1897	-0.5600	0.5752
	Gamma	300	5	500	0.6219	1.3628		
Normal	Normal	300	10	500	0.4833	0.1198	-0.4850	0.6267
	Gamma	300	10	500	0.4867	0.1005		
Normal	Normal	100	5	500	0.6719	0.5931	1.6571	0.0978
	Gamma	100	5	500	0.6133	0.5230		
Normal	Normal	100	10	500	0.4537	0.2563	1.1358	0.2563
	Gamma	100	10	500	0.4380	0.1714		
Beta	Normal	300	5	500	0.6143	0.1503	1.0784	0.2811
	Gamma	300	5	500	0.6044	0.1387		
Beta	Normal	300	10	500	0.5165	0.1200	-2.810	0.0050**
	Gamma	300	10	500	0.5349	0.0847		
Beta	Normal	100	5	500	0.6428	0.4265	-0.122	0.9026
	Gamma	100	5	500	0.6469	0.6149		
Beta	Normal	100	10	500	0.4730	0.1389	-3.569	0.0003**
	Gamma	100	10	500	0.5035	0.1315		
รูปแบบที่ 3 AGIRM B								
Normal	Normal	300	5	500	0.6854	0.1460	-0.470	0.6383
	Gamma	300	5	500	0.7096	1.1449		
Normal	Normal	300	10	500	0.6027	0.0985	0.4521	0.6513
	Gamma	300	10	500	0.6001	0.0832		
Normal	Normal	100	5	500	0.7402	0.4481	2.2976	0.0218
	Gamma	100	5	500	0.6798	0.3810		
Normal	Normal	100	10	500	0.5534	0.2043	1.7215	0.0855
	Gamma	100	10	500	0.5346	0.1341		

ตารางที่ 24 (ต่อ)

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
Beta	Normal	300	5	500	0.7092	0.1132	2.1932	0.0285
	Gamma	300	5	500	0.6940	0.1056		
Beta	Normal	300	10	500	0.6326	0.0955	-1.408	0.1593
	Gamma	300	10	500	0.6403	0.0746		
Beta	Normal	100	5	500	0.7261	0.3169	0.3683	0.7127
	Gamma	100	5	500	0.7167	0.4755		
Beta	Normal	100	10	500	0.5894	0.1159	-1.579	0.1145
	Gamma	100	10	500	0.6008	0.1141		

จากการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบปกติ (Normal) กับแบบแกมมา (Gamma) ทั้ง 24 คู่กรณี พบว่า มีคู่กรณีศึกษาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิ้น 5 คู่กรณี คิดเป็นร้อยละ 20.83 โดยแบ่งเป็นรูปแบบที่ 1 Original GIRM 3 คู่กรณี คิดเป็นร้อยละ 12.50 และรูปแบบที่ 2 AGIRM 2 คู่กรณี คิดเป็นร้อยละ 8.33 ดังนั้นจากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ของกรณีศึกษาการกำหนดลักษณะการแจกแจงเริ่มแรกของผู้สอบระหว่าง การแจกแจงแบบปกติ (Normal) และการแจกแจงแบบแกมมา (Gamma) ไม่ส่งผลต่อผลการประมาณค่าพารามิเตอร์ของวิธีการ GIRM ทั้ง 4 รูปแบบของวิธีการประมาณค่า

ตารางที่ 25 ผลการวิเคราะห์ความแตกต่างของส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation; S.D) ในการประมาณค่าองค์ประกอบความแปรปรวนระหว่างกรณีศึกษาที่มีการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแตกต่างกัน

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
รูปแบบที่ 1 Original GIRM									
Normal	Normal	300	5	0.0120	0.0007	0.0000	0.0140	0.0140	0.0092
	Gamma	300	5	0.0136	0.0010	0.0000	0.0149	0.0149	0.0100
Normal	Normal	300	10	0.0123	0.0005	0.0000	0.0138	0.0138	0.0092
	Gamma	300	10	0.0056	0.0007	0.0000	0.0064	0.0064	0.0042

ตารางที่ 25 (ต่อ)

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
Normal	Normal	100	5	0.0132	0.0019	0.0000	0.0153	0.0153	0.0101
	Gamma	100	5	0.0043	0.0017	0.0000	0.0086	0.0086	0.0049
Normal	Normal	100	10	0.0114	0.0022	0.0001	0.0120	0.0119	0.0083
	Gamma	100	10	0.0057	0.0020	0.0001	0.0067	0.0068	0.0045
Beta	Normal	300	5	0.0123	0.0005	0.0000	0.0138	0.0138	0.0092
	Gamma	300	5	0.0036	0.0006	0.0000	0.0060	0.0060	0.0035
Beta	Normal	300	10	0.0110	0.0006	0.0000	0.0114	0.0114	0.0079
	Gamma	300	10	0.0049	0.0007	0.0000	0.0055	0.0055	0.0037
Beta	Normal	100	5	0.0136	0.0010	0.0000	0.0149	0.0149	0.0100
	Gamma	100	5	0.0036	0.0011	0.0000	0.0067	0.0067	0.0038
Beta	Normal	100	10	0.0114	0.0010	0.0000	0.0119	0.0118	0.0082
	Gamma	100	10	0.0047	0.0012	0.0000	0.0058	0.0058	0.0037
รูปแบบที่ 2 AGIRM A									
Normal	Normal	300	5	0.0134	0.0014	NA	NA	0.0151	0.0117
	Gamma	300	5	0.0162	0.0026	NA	NA	0.0172	0.0137
Normal	Normal	300	10	0.0110	0.0011	NA	NA	0.0115	0.0092
	Gamma	300	10	0.0065	0.0012	NA	NA	0.0074	0.0057
Normal	Normal	100	5	0.0163	0.0039	NA	NA	0.0184	0.0143
	Gamma	100	5	0.0103	0.0043	NA	NA	0.0127	0.0097
Normal	Normal	100	10	0.0123	0.0034	NA	NA	0.0129	0.0105
	Gamma	100	10	0.0076	0.0032	NA	NA	0.0090	0.0070
Beta	Normal	300	5	0.0133	0.0011	NA	NA	0.0146	0.0114
	Gamma	300	5	0.0063	0.0013	NA	NA	0.0083	0.0060
Beta	Normal	300	10	0.0115	0.0009	NA	NA	0.0119	0.0095
	Gamma	300	10	0.0058	0.0010	NA	NA	0.0062	0.0049
Beta	Normal	100	5	0.0162	0.0026	NA	NA	0.0172	0.0137
	Gamma	100	5	0.0093	0.0026	NA	NA	0.0114	0.0086
Beta	Normal	100	10	0.0126	0.0021	NA	NA	0.0130	0.0105
	Gamma	100	10	0.0071	0.0023	NA	NA	0.0083	0.0064

ตารางที่ 25 (ต่อ)

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
รูปแบบที่ 3 AGIRM B									
Normal	Normal	300	5	0.0134	0.0014	0.0000	0.0151	0.0151	0.0101
	Gamma	300	5	0.0162	0.0026	0.0000	0.0172	0.0172	0.0118
Normal	Normal	300	10	0.0110	0.0011	0.0000	0.0115	0.0115	0.0079
	Gamma	300	10	0.0065	0.0012	0.0000	0.0074	0.0074	0.0049
Normal	Normal	100	5	0.0163	0.0039	0.0000	0.0184	0.0184	0.0124
	Gamma	100	5	0.0103	0.0043	0.0000	0.0127	0.0127	0.0084
Normal	Normal	100	10	0.0123	0.0034	0.0001	0.0129	0.0129	0.0091
	Gamma	100	10	0.0076	0.0032	0.0001	0.0091	0.0090	0.0061
Beta	Normal	300	5	0.0133	0.0011	0.0000	0.0146	0.0146	0.0098
	Gamma	300	5	0.0063	0.0013	0.0000	0.0083	0.0083	0.0052
Beta	Normal	300	10	0.0115	0.0009	0.0000	0.0119	0.0119	0.0082
	Gamma	300	10	0.0058	0.0010	0.0000	0.0062	0.0062	0.0042
Beta	Normal	100	5	0.0162	0.0026	0.0000	0.0172	0.0172	0.0118
	Gamma	100	5	0.0093	0.0026	0.0000	0.0114	0.0114	0.0074
Beta	Normal	100	10	0.0126	0.0021	0.0000	0.0130	0.0130	0.0091
	Gamma	100	10	0.0071	0.0023	0.0000	0.0083	0.0083	0.0055

NA = ไม่สามารถทำการวิเคราะห์ได้

จากการวิเคราะห์ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ที่คำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) พบว่า เมื่อเปรียบเทียบค่าส่วนเบี่ยงเบนมาตรฐานระหว่างกรณีที่การแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ พบว่า ในจำนวนคู่กรณีทั้งหมด 24 คู่กรณี ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบแกมมา ให้ผลการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) ต่ำกว่า 21 คู่กรณี คิดเป็นร้อยละ 87.50 โดยมีคู่กรณีจำนวน 3 คู่กรณี คิดเป็นร้อยละ 12.50 ที่ลักษณะการแจกแจงเริ่มแรกของของค่าพารามิเตอร์ของผู้สอบแบบปกติ ให้ผลการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) สูงกว่า จากผลการวิเคราะห์ สามารถสรุปได้ว่า โดยส่วนใหญ่ของกรณีศึกษาลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบ

แกมมาส่งผลต่อความไม่แน่นอนในประมาณค่าองค์ประกอบความแปรปรวนต่ำกว่าลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบปกติ

ตารางที่ 26 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (prior distribution) แตกต่างกัน

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
รูปแบบที่ 1 Original GIRM								
Normal	Normal	300	5	500	0.0489	0.0152	3.4841	0.0005
	Gamma	300	5	500	0.0462	0.0091		
Normal	Normal	300	10	500	0.0406	0.0135	-0.132	0.8948
	Gamma	300	10	500	0.0407	0.0091		
Normal	Normal	100	5	500	0.0495	0.0167	3.1151	0.0018**
	Gamma	100	5	500	0.0448	0.0106		
Normal	Normal	100	10	500	0.0376	0.0148	-3.079	0.0021**
	Gamma	100	10	500	0.0400	0.0097		
Beta	Normal	300	5	500	0.0497	0.0155	0.4301	0.6672
	Gamma	300	5	500	0.0534	0.0081		
Beta	Normal	300	10	500	0.0408	0.0140	-4.621	0.0000**
	Gamma	300	10	500	0.0442	0.0086		
Beta	Normal	100	5	500	0.0499	0.0095	-0.981	0.3266
	Gamma	100	5	500	0.0564	0.0095		
Beta	Normal	100	10	500	0.0395	0.0148	-6.547	0.0000**
	Gamma	100	10	500	0.0447	0.0100		
รูปแบบที่ 2 AGIRM A								
Normal	Normal	300	5	500	0.0485	0.0163	3.555	0.0003**
	Gamma	300	5	500	0.0454	0.0112		
Normal	Normal	300	10	500	0.0400	0.0136	-0.168	0.8861
	Gamma	300	10	500	0.0401	0.0102		
Normal	Normal	100	5	500	0.0485	0.0194	3.823	0.0001**
	Gamma	100	5	500	0.0443	0.0151		

ตารางที่ 26 (ต่อ)

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
Normal	Normal	100	10	500	0.0372	0.0158	-3.023	0.0025**
	Gamma	100	10	500	0.0399	0.0120		
Beta	Normal	300	5	500	0.0494	0.0163	-0.217	0.8300
	Gamma	300	5	500	0.0496	0.0107		
Beta	Normal	300	10	500	0.0403	0.0145	-4.181	0.0000**
	Gamma	300	10	500	0.0435	0.0092		
Beta	Normal	100	5	500	0.0490	0.0192	-1.017	0.3094
	Gamma	100	5	500	0.0501	0.0149		
Beta	Normal	100	10	500	0.0393	0.0154	-5.262	0.0000**
	Gamma	100	10	500	0.0440	0.0093		
รูปแบบที่ 3 AGIRM B								
Normal	Normal	300	5	500	0.0489	0.0164	3.5362	0.0004**
	Gamma	300	5	500	0.0457	0.0113		
Normal	Normal	300	10	500	0.0404	0.0137	-0.156	0.8755
	Gamma	300	10	500	0.0405	0.0103		
Normal	Normal	100	5	500	0.0489	0.0195	3.8370	0.0001**
	Gamma	100	5	500	0.0446	0.0152		
Normal	Normal	100	10	500	0.0377	0.0159	-2.992	0.0028**
	Gamma	100	10	500	0.0403	0.0120		
Beta	Normal	300	5	500	0.0498	0.0164	-0.208	0.8350
	Gamma	300	5	500	0.0500	0.0108		
Beta	Normal	300	10	500	0.0408	0.0145	-4.133	0.0000**
	Gamma	300	10	500	0.0440	0.0093		
Beta	Normal	100	5	500	0.0494	0.0193	-0.997	0.3177
	Gamma	100	5	500	0.0505	0.0150		
Beta	Normal	100	10	500	0.0397	0.0155	-5.209	0.0000**
	Gamma	100	10	500	0.0444	0.0127		

จากการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบปกติ (Normal) และแบบแกมมา (Gamma) ทั้งสิ้น 24 คู่กรณี พบว่ามีคู่กรณีศึกษาที่มีค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 คิดเป็นร้อยละ 58.33 โดยแบ่งเป็น รูปแบบที่ 1 Original GIRM 4 คู่กรณี รูปแบบที่ 2 AGIRM A 5 คู่กรณี และรูปแบบที่ 3 AGIRM B 5 คู่กรณี คิดเป็น ร้อยละ 16.67 20.83 และ 20.83 ตามลำดับ จากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ของกรณีศึกษาลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบส่งผลต่อประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนเมื่อพิจารณาจากค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (EUC)

6.2 ความไว (Sensitivity) ของผลการประมาณค่าของวิธีการสรุปอ้างอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์

ในการนำเสนอผลการวัดประสิทธิภาพในการประมาณค่าพารามิเตอร์ของการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ จะนำเสนอแยกออกเป็น 3 ตารางในแต่ละดัชนีวัดประสิทธิภาพของการประมาณค่าพารามิเตอร์ ซึ่งได้แก่ ความลำเอียงในการประมาณค่า (Biased estimator) คำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) จำแนกตามเงื่อนไขของการวิเคราะห์ทั้ง 16 กรณี มีรายละเอียดการนำเสนอดังนี้

ตารางที่ 27 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างข้อสอบที่มีการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแตกต่างกัน

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
รูปแบบที่ 1 Original GIRM								
Normal	Normal	300	5	500	0.6768	0.1068	-4.490	0.0000**
	Beta	300	5	500	0.7046	0.0880		
Normal	Normal	300	10	500	0.6059	0.0922	-5.386	0.0000**
	Beta	300	10	500	0.6363	0.0863		
Normal	Normal	100	5	500	0.6654	0.2995	-1.076	0.2821
	Beta	100	5	500	0.6810	0.1221		
Normal	Normal	100	10	500	0.5333	0.1855	-5.296	0.0000**
	Beta	100	10	500	0.5840	0.1069		
Gamma	Normal	300	5	500	0.6772	0.5858	-0.344	0.7308
	Beta	300	5	500	0.6863	0.0681		
Gamma	Normal	300	10	500	0.6042	0.0687	-9.471	0.0000**
	Beta	300	10	500	0.6445	0.0656		
Gamma	Normal	100	5	500	0.6886	0.1483	-5.639	0.0000**
	Beta	100	5	500	0.6704	0.1949		
Gamma	Normal	100	10	500	0.5164	0.1095	-12.92	0.0000**
	Beta	100	10	500	0.5035	0.1315		
รูปแบบที่ 2 AGIRM A								
Normal	Normal	300	5	500	0.5874	0.1897	-2.282	0.0132*
	Beta	300	5	500	0.6143	0.1503		
Normal	Normal	300	10	500	0.4833	0.1198	-4.368	0.0000**
	Beta	300	10	500	0.5165	0.1200		
Normal	Normal	100	5	500	0.6719	0.5931	0.8895	0.3704
	Beta	100	5	500	0.6428	0.4265		

ตารางที่ 27 (ต่อ)

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
Normal	Normal	100	10	500	0.4537	0.2563	-1.477	0.1401
	Beta	100	10	500	0.4730	0.1389		
Gamma	Normal	300	5	500	0.6219	1.3628	0.2856	0.7753
	Beta	300	5	500	0.6044	0.1387		
Gamma	Normal	300	10	500	0.4867	0.1005	-8.194	0.0000**
	Beta	300	10	500	0.5349	0.0847		
Gamma	Normal	100	5	500	0.6133	0.5230	-0.931	0.3517
	Beta	100	5	500	0.6469	0.6149		
Gamma	Normal	100	10	500	0.4380	0.1714	-6.775	0.0000**
	Beta	100	10	500	0.5035	0.1315		
รูปแบบที่ 3 AGIRM B								
Normal	Normal	300	5	500	0.6554	0.1460	-2.868	0.0039**
	Beta	300	5	500	0.7092	0.1132		
Normal	Normal	300	10	500	0.6027	0.0985	-4.885	0.0000**
	Beta	300	10	500	0.6326	0.0955		
Normal	Normal	100	5	500	0.7402	0.4481	0.5752	0.5653
	Beta	100	5	500	0.7261	0.3169		
Normal	Normal	100	10	500	0.5534	0.2043	-3.424	0.0006**
	Beta	100	10	500	0.5894	0.1159		
Gamma	Normal	300	5	500	0.7096	1.1449	0.3034	0.7617
	Beta	300	5	500	0.6940	0.1056		
Gamma	Normal	300	10	500	0.6001	0.0832	-8.045	0.0000**
	Beta	300	10	500	0.6403	0.0746		
Gamma	Normal	100	5	500	0.6798	0.3810	-1.354	0.1759
	Beta	100	5	500	0.7167	0.4755		
Gamma	Normal	100	10	500	0.5346	0.1341	-8.418	0.0000**
	Beta	100	10	500	0.6008	0.1141		

จากการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบปกติ (Normal) กับแบบเบต้า (Beta) ทั้ง 24 คู่กรณี พบว่า มีคู่กรณีศึกษาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิ้น 15 คู่กรณี คิดเป็นร้อยละ 62.5 โดยแบ่งเป็นรูปแบบที่ 1 Original GIRM 6 คู่กรณี คิดเป็นร้อยละ 25.00 รูปแบบที่ 2 AGIRM A 4 คู่กรณี คิดเป็นร้อยละ 16.67 และรูปแบบที่ 3 AGIRM B 5 คู่กรณี คิดเป็นร้อยละ 20.83 จากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ของกรณีศึกษาการแจกแจงแบบปกติ จะให้ค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (MAD) ต่ำกว่า การแจกแจงแบบเบต้า ยกเว้นกรณี การกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบแกมมา ที่ให้ค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (MAD) การกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบเบต้าต่ำกว่าการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบปกติ

ตารางที่ 28 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) ระหว่างข้อสอบที่มีการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแตกต่างกัน

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
รูปแบบที่ 1 Original GIRM									
Normal	Normal	300	5	0.0120	0.0007	0.0000	0.0140	0.0140	0.0092
	Beta	300	5	0.0123	0.0005	0.0000	0.0138	0.0138	0.0092
Normal	Normal	300	10	0.0123	0.0005	0.0000	0.0138	0.0138	0.0092
	Beta	300	10	0.0110	0.0006	0.0000	0.0114	0.0114	0.0079
Normal	Normal	100	5	0.0132	0.0019	0.0000	0.0153	0.0153	0.0101
	Beta	100	5	0.0136	0.0010	0.0000	0.0149	0.0149	0.0100
Normal	Normal	100	10	0.0114	0.0022	0.0001	0.0120	0.0119	0.0083
	Beta	100	10	0.0114	0.0010	0.0000	0.0119	0.0118	0.0082
Gamma	Normal	300	5	0.0136	0.0010	0.0000	0.0149	0.0149	0.0100
	Beta	300	5	0.0036	0.0006	0.0000	0.0060	0.0060	0.0035
Gamma	Normal	300	10	0.0114	0.0010	0.0000	0.0119	0.0118	0.0082
	Beta	300	10	0.0049	0.0007	0.0000	0.0055	0.0055	0.0037
Gamma	Normal	100	5	0.0043	0.0017	0.0000	0.0086	0.0086	0.0049
	Beta	100	5	0.0036	0.0011	0.0000	0.0067	0.0067	0.0038

ตารางที่ 28 (ต่อ)

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
Gamma	Normal	100	10	0.0057	0.0020	0.0001	0.0067	0.0068	0.0045
	Beta	100	10	0.0047	0.0012	0.0000	0.0058	0.0058	0.0037
รูปแบบที่ 2 AGIRM A									
Normal	Normal	300	5	0.0134	0.0014	NA	NA	0.0151	0.0117
	Beta	300	5	0.0133	0.0011	NA	NA	0.0146	0.0114
Normal	Normal	300	10	0.0110	0.0011	NA	NA	0.0115	0.0092
	Beta	300	10	0.0115	0.0009	NA	NA	0.0119	0.0095
Normal	Normal	100	5	0.0163	0.0039	NA	NA	0.0184	0.0143
	Beta	100	5	0.0162	0.0026	NA	NA	0.0172	0.0137
Normal	Normal	100	10	0.0123	0.0034	NA	NA	0.0129	0.0105
	Beta	100	10	0.0126	0.0021	NA	NA	0.0130	0.0105
Gamma	Normal	300	5	0.0162	0.0026	NA	NA	0.0172	0.0137
	Beta	300	5	0.0063	0.0013	NA	NA	0.0083	0.0060
Gamma	Normal	300	10	0.0126	0.0021	NA	NA	0.0130	0.0105
	Beta	300	10	0.0058	0.0010	NA	NA	0.0062	0.0049
Gamma	Normal	100	5	0.0103	0.0043	NA	NA	0.0127	0.0097
	Beta	100	5	0.0093	0.0026	NA	NA	0.0114	0.0086
Gamma	Normal	100	10	0.0076	0.0032	NA	NA	0.0090	0.0070
	Beta	100	10	0.0071	0.0023	NA	NA	0.0083	0.0064
รูปแบบที่ 3 AGIRM B									
Normal	Normal	300	5	0.0134	0.0014	0.0000	0.0151	0.0151	0.0101
	Beta	300	5	0.0133	0.0011	0.0000	0.0146	0.0146	0.0098
Normal	Normal	300	10	0.0110	0.0011	0.0000	0.0115	0.0115	0.0079
	Beta	300	10	0.0115	0.0009	0.0000	0.0119	0.0119	0.0082
Normal	Normal	100	5	0.0163	0.0039	0.0000	0.0184	0.0184	0.0124
	Beta	100	5	0.0162	0.0026	0.0000	0.0172	0.0172	0.0118
Normal	Normal	100	10	0.0123	0.0034	0.0001	0.0129	0.0129	0.0091
	Beta	100	10	0.0126	0.0021	0.0000	0.0130	0.0130	0.0091

ตารางที่ 28 (ต่อ)

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	Standard Deviation					
				$\sigma^2(p)$	$\sigma^2(i)$	$\sigma^2(pi)$	$\sigma^2(e)$	$\sigma^2(pi,e)$	รวม
Gamma	Normal	300	5	0.0162	0.0026	0.0000	0.0172	0.0172	0.0118
	Beta	300	5	0.0063	0.0013	0.0000	0.0083	0.0083	0.0052
Gamma	Normal	300	10	0.0126	0.0021	0.0000	0.0130	0.0130	0.0091
	Beta	300	10	0.0058	0.0010	0.0000	0.0062	0.0062	0.0042
Gamma	Normal	100	5	0.0103	0.0043	0.0000	0.0127	0.0127	0.0084
	Beta	100	5	0.0093	0.0026	0.0000	0.0114	0.0114	0.0074
Gamma	Normal	100	10	0.0076	0.0032	0.0001	0.0091	0.0090	0.0061
	Beta	100	10	0.0071	0.0023	0.0000	0.0083	0.0083	0.0055

NA = ไม่สามารถทำการวิเคราะห์ได้

จากการวิเคราะห์ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ที่คำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) โดยเมื่อเปรียบเทียบค่าส่วนเบี่ยงเบนมาตรฐานระหว่างกรณีที่การแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบในจำนวนคู่กรณีทั้งหมด 24 คู่กรณี พบว่า ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบเบต้า ให้ผลการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) ต่ำกว่าที่มีลักษณะการแจกแจงเริ่มแรกของของค่าพารามิเตอร์ของผู้สอบแบบปกติเป็นส่วนใหญ่ จำนวน 18 คู่กรณี คิดเป็นร้อยละ 75 จากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ของกรณีศึกษาลักษณะการแจกแจงเริ่มแรกของของค่าพารามิเตอร์ของข้อสอบแบบเบต้า ส่งผลต่อความไม่แน่นอนในประมาณค่าองค์ประกอบความแปรปรวนต่ำกว่าลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบปกติ

ตารางที่ 29 ผลการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแตกต่างกัน

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
รูปแบบที่ 1 Original GIRM								
Normal	Normal	300	5	500	0.0489	0.0152	-0.846	0.3974
	Beta	300	5	500	0.0497	0.0155		
Normal	Normal	300	10	500	0.0406	0.0135	-0.205	0.8376
	Beta	300	10	500	0.0408	0.0140		
Normal	Normal	100	5	500	0.0495	0.0167	-1.281	0.2002
	Beta	100	5	500	0.0499	0.0164		
Normal	Normal	100	10	500	0.0376	0.0148	-2.052	0.0403
	Beta	100	10	500	0.0395	0.0148		
Gamma	Normal	300	5	500	0.0462	0.0091	-5.953	0.0000**
	Beta	300	5	500	0.0534	0.0081		
Gamma	Normal	300	10	500	0.0407	0.0091	-6.237	0.0000**
	Beta	300	10	500	0.0442	0.0086		
Gamma	Normal	100	5	500	0.0448	0.0106	-7.771	0.0000**
	Beta	100	5	500	0.0564	0.0095		
Gamma	Normal	100	10	500	0.0400	0.0097	-7.550	0.0000**
	Beta	100	10	500	0.0447	0.0100		
รูปแบบที่ 2 AGIRM A								
Normal	Normal	300	5	500	0.0485	0.0163	-0.804	0.3713
	Beta	300	5	500	0.0494	0.0163		
Normal	Normal	300	10	500	0.0400	0.0136	-0.380	0.7037
	Beta	300	10	500	0.0403	0.0145		
Normal	Normal	100	5	500	0.0485	0.0194	-0.424	0.6716
	Beta	100	5	500	0.0490	0.0192		

ตารางที่ 29 (ต่อ)

การแจกแจงเริ่มแรกข้อสอบ	การแจกแจงเริ่มแรกผู้สอบ	ขนาดตัวอย่าง	จำนวนข้อสอบ	N	Mean	S.D.	t-test	p-value
Normal	Normal	100	10	500	0.0372	0.0158	-2.086	0.0371
	Beta	100	10	500	0.0393	0.0154		
Gamma	Normal	300	5	500	0.0454	0.0112	-6.152	0.0000**
	Beta	300	5	500	0.0496	0.0107		
Gamma	Normal	300	10	500	0.0401	0.0102	-5.561	0.0000**
	Beta	300	10	500	0.0435	0.0092		
Gamma	Normal	100	5	500	0.0443	0.0151	-6.148	0.0000**
	Beta	100	5	500	0.0501	0.0149		
Gamma	Normal	100	10	500	0.0399	0.0120	-5.219	0.0000**
	Beta	100	10	500	0.0440	0.0093		
รูปแบบที่ 3 AGIRM B								
Normal	Normal	300	5	500	0.0489	0.0164	-0.904	0.3662
	Beta	300	5	500	0.0498	0.0164		
Normal	Normal	300	10	500	0.0404	0.0137	-0.386	0.6989
	Beta	300	10	500	0.0408	0.0145		
Normal	Normal	100	5	500	0.0489	0.0195	-0.442	0.6580
	Beta	100	5	500	0.0494	0.0193		
Normal	Normal	100	10	500	0.0377	0.0159	-2.109	0.0351
	Beta	100	10	500	0.0397	0.0155		
Gamma	Normal	300	5	500	0.0457	0.0113	-6.111	0.0000**
	Beta	300	5	500	0.0500	0.0108		
Gamma	Normal	300	10	500	0.0405	0.0103	-5.514	0.0000**
	Beta	300	10	500	0.0440	0.0093		
Gamma	Normal	100	5	500	0.0446	0.0152	-6.163	0.0000**
	Beta	100	5	500	0.0505	0.0150		
Gamma	Normal	100	10	500	0.0403	0.0120	-5.221	0.0000**
	Beta	100	10	500	0.0444	0.0127		

NA = ไม่สามารถทำการวิเคราะห์ได้

จากการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบปกติ (Normal) และแบบเบต้า (Beta) ทั้งสิ้น 24 คู่กรณี พบว่า มีคู่กรณีศึกษาที่มีค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 คิดเป็นร้อยละ 50.00 โดยแบ่งเป็น รูปแบบที่ 1 Original GIRM 4 คู่กรณี รูปแบบที่ 2 AGIRM A 4 คู่กรณี และรูปแบบที่ 3 AGIRM B 4 คู่กรณี โดยเป็นที่น่าสังเกตว่า คู่กรณีที่มีความแตกต่างกันอย่างมีนัยสำคัญจะมีลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบแกมมา (Gamma) ทั้งหมดทั้ง 9 รูปแบบการวิเคราะห์ ดังนั้นจึงสามารถสรุปได้ว่า ถ้าลักษณะการแจกแจงเริ่มแรกของผู้สอบมีลักษณะการแจกแจงแบบแกมมา (Gamma) การแจกแจงเริ่มแรกของข้อสอบส่งผลต่อการประมาณค่าพารามิเตอร์ โดยลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบปกติ (Normal) ให้ค่าประสิทธิภาพของการประมาณค่าขององค์ประกอบความแปรปรวนที่สูงกว่าลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบเบต้า (Beta)

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มุ่งเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทั้ง 4 วิธี ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM ซึ่ง พัฒนาโดย Brigg และ Wilson (2007) กับรูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B และรูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งผู้วิจัยเป็นผู้พัฒนาขึ้น นอกจากนี้ยังศึกษาถึงอิทธิพลของขนาดกลุ่มตัวอย่างและจำนวนข้อสอบที่มีต่อประสิทธิภาพของวิธีวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) รวมทั้งยังศึกษาถึงความไว (Sensitivity) ของผลการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบ (β) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) ข้อมูลที่ใช้ในการศึกษาครั้งนี้ใช้ข้อมูลจำลองที่ได้จากโปรแกรม R และทำการประมวลผลภายใต้จากการเขียนคำสั่งการประมวลผลด้วยโปรแกรม WinBUGS และกลับมาแสดงผลจะแสดงผลในโปรแกรม R โดยใช้โปรแกรมร่วมกันระหว่างโปรแกรม R และโปรแกรม WinBUGS ด้วย Package R2 WinBUGS

การวัดประสิทธิภาพของรูปแบบวิธีการประมาณวัดจากดัชนีวัดประสิทธิภาพของการประมาณค่าพารามิเตอร์ 3 ดัชนี ได้แก่ ความลำเอียงในการประมาณค่า (Biased estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ค่าความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ในบทนี้ได้นำเสนอข้อสรุป อภิปรายผลการวิจัย และข้อเสนอแนะ ดังต่อไปนี้

สรุปผลการวิจัย

1. ผลการวิเคราะห์ค่าพารามิเตอร์ตามแนวทฤษฎีการตอบสนองข้อสอบ พบว่า ค่าพารามิเตอร์ของข้อสอบด้วยวิธีการ MCMC ในรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B ให้ค่าเฉลี่ยของพารามิเตอร์ของข้อสอบอยู่ระหว่าง -0.0358 ถึง 0.8901 และให้ค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0029 ถึง 0.0113 ส่วนการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบด้วยวิธีการ Numerical Method ในรูปแบบที่ 4 Numerical Bayesian GIRM สามารถวิเคราะห์ได้เฉพาะกรณีการแจกแจงเริ่มแรกผู้สอบและการแจกแจงเริ่มแรกของข้อสอบมีลักษณะการกระจายแบบโค้งปกติเท่านั้น โดยให้ค่าเฉลี่ยของพารามิเตอร์ของข้อสอบอยู่ระหว่าง -0.0344 ถึง 0.0018 และให้ค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0081 ถึง 0.0108 นอกจากนี้เมื่อพิจารณาเปรียบเทียบความแตกต่างของค่าเฉลี่ยในการประมาณค่าพารามิเตอร์ของข้อสอบระหว่างการประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ MCMC กับการประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้วิธีการ Numerical Method พบว่า ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 แสดงว่า การประมาณค่าพารามิเตอร์ของข้อสอบทั้งสองวิธีให้ผลที่ไม่แตกต่างกัน

สำหรับค่าพารามิเตอร์ของผู้สอบตามแนวทฤษฎีการตอบสนองข้อสอบ พบว่า การวิเคราะห์ค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ MCMC ในรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B ให้ค่าเฉลี่ยของพารามิเตอร์ของผู้สอบอยู่ระหว่าง -0.0019 ถึง 0.9574 และให้ค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0214 ถึง 0.0352 ส่วนการวิเคราะห์ค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ Numerical Method ในรูปแบบที่ 4 Numerical Bayesian GIRM สามารถวิเคราะห์ได้เฉพาะกรณีการแจกแจงเริ่มแรกผู้สอบและการแจกแจงเริ่มแรกของผู้สอบมีลักษณะการกระจายแบบโค้งปกติเท่านั้น โดยให้ค่าเฉลี่ยของพารามิเตอร์ของผู้สอบอยู่ระหว่าง -0.0039 ถึง -0.0004 และให้ค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0194 ถึง 0.0307 นอกจากนี้เมื่อพิจารณาเปรียบเทียบความแตกต่างของค่าเฉลี่ยในการประมาณค่าพารามิเตอร์ของผู้สอบระหว่างการประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ MCMC กับการประมาณค่าพารามิเตอร์ของผู้สอบโดยใช้วิธีการ Numerical Method พบว่า ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 แสดงว่า การประมาณค่าพารามิเตอร์ของผู้สอบทั้งสองวิธีให้ผลที่ไม่แตกต่างกัน

2. ผลการวิเคราะห์ค่าองค์ประกอบความแปรปรวนตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด พบว่า องค์ประกอบความแปรปรวนในการประมาณค่าของผู้สอบ ($\sigma^2(p)$) มีค่าเฉลี่ยอยู่ระหว่าง 0.0262 ถึง 0.0111 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0292 ถึง 0.0036 องค์ประกอบความแปรปรวนของข้อสอบ ($\sigma^2(i)$) มีค่าเฉลี่ยอยู่ระหว่าง 0.0053 ถึง 0.0005 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0064 ถึง 0.0005 องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ ($\sigma^2(pi)$) มีค่าเฉลี่ยอยู่ระหว่าง 0.0001 ถึง 0.0000 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0001 ถึง 0.0000 องค์ประกอบความแปรปรวนของความคลาดเคลื่อนในส่วนที่เหลือ ($\sigma^2(e)$) มีค่าเฉลี่ยอยู่ระหว่าง 0.2338 ถึง 0.2309 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0356 ถึง 0.0055 องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบรวมกับความคลาดเคลื่อนในส่วนที่เหลือ ($\sigma^2(pi,e)$) มีค่าเฉลี่ยอยู่ระหว่าง 0.2338 ถึง 0.2179 และ ค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.0354 ถึง 0.0055 ค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) มีค่าเฉลี่ยอยู่ระหว่าง 0.4362 ถึง 0.1756 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.2831 ถึง 0.0485 และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) มีค่าเฉลี่ยอยู่ระหว่าง 0.4343 ถึง 0.1528 และค่าส่วนเบี่ยงเบนมาตรฐานอยู่ระหว่าง 0.2828 ถึง 0.0485

เมื่อพิจารณาในกรณีที่มีลักษณะการแจกแจงเริ่มแรกรูปแบบเดียวกัน พบว่า ถ้ามีจำนวนข้อเพิ่มมากขึ้น จะส่งผลทำให้มีค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) เพิ่มมากขึ้นในทุกกรณีการแจกแจงเริ่มแรก สำหรับจำนวนผู้เข้าสอบก็เช่นเดียวกับจำนวนข้อสอบ โดยในกรณีที่มีลักษณะการแจกแจงเริ่มแรกรูปแบบเดียวกัน พบว่า ถ้ามีจำนวนผู้เข้าสอบมากขึ้น จะส่งผลทำให้มีค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมพัทธ์ ($\hat{E}_p^2$) และค่าสัมประสิทธิ์สหสัมพันธ์ของการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดแบบสัมบูรณ์ (Φ) เพิ่มมากขึ้นในทุกกรณีการแจกแจงเริ่มแรก

3. ผลการวิเคราะห์เพื่อเปรียบเทียบประสิทธิภาพของตัวประมาณค่าในแต่ละรูปแบบการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) 4 รูปแบบ เมื่อพิจารณาเรื่องของความลำเอียงในการประมาณค่า (biased estimator) ของวิธีการประมาณค่าทั้ง 4 วิธีซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) พบว่า โดยใน 4 กรณีแรกที่มีลักษณะการแจกแจงเริ่มแรกของข้อสอบ

และผู้สอบแบบปกติ (normal distribution) ที่รูปแบบที่ 4 สามารถประมาณค่าพารามิเตอร์ได้ พบว่า รูปแบบที่ 4 Numerical Bayesian GIRM มีวิเคราะห์ให้ค่า MAD น้อยที่สุดใน 2 กรณี ได้แก่ กรณีที่ 2 Normal-Normal 300-5 และ กรณีที่ 4 Normal-Normal 100-5 ส่วน 2 กรณีที่เหลือ ได้แก่ กรณีที่ 1 Normal-Normal 300-10 และ กรณีที่ 3 Normal-Normal 100-10 รูปแบบการประมาณค่าที่ 1 Original GIRM มีค่า MAD น้อยที่สุด สำหรับกรณีอื่นๆ พบว่า การประมาณค่าด้วยรูปแบบที่ 1 Original GIRM จะให้ค่า MAD ต่ำกว่ารูปแบบการประมาณค่าวิธีอื่นๆ รองลงมา รูปแบบที่ 2 AGIRM A และ รูปแบบที่ 3 AGIRM B ให้ค่า MAD เท่ากัน สำหรับรูปแบบที่ 4 Numerical Bayesian GIRM ให้ค่า MAD สูงที่สุด

สำหรับความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) พบว่า การประมาณค่าในรูปแบบที่ 4 Numerical Bayesian GIRM ให้ค่าส่วนเบี่ยงเบนมาตรฐานรวมขององค์ประกอบความแปรปรวนตามทฤษฎีการอ้างอิงความน่าเชื่อของผลการวัดต่ำสุดในทุกกรณี รองลงมารูปแบบที่ 1 Original GIRM รูปแบบที่ 3 AGIRM A และรูปแบบที่ 2 AGIRM B ตามลำดับ

ส่วนประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) พบว่า มีค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด อยู่ระหว่าง 0.0375 ถึง 0.0644 โดยการวิเคราะห์ด้วยรูปแบบที่ 2 AGIRM A ให้ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิดต่ำสุดในทุกกรณี รองลงมารูปแบบที่ 3 AGIRM A รูปแบบที่ 1 Original GIRM และรูปแบบที่ 4 Numerical Bayesian GIRM ตามลำดับ

จากการวัดประสิทธิภาพในการประมาณค่าทั้ง 3 ดัชนีชี้วัด ทำให้สามารถเปรียบเทียบการวัดประสิทธิภาพในการประมาณค่าพารามิเตอร์ โดยจำแนกตามรูปแบบการประมาณค่าทั้ง 4 รูปแบบ และกรณีศึกษาทั้ง 16 กรณีศึกษาได้ดังตารางที่ 30

ตารางที่ 30 ผลการสรุปประสิทธิภาพของการประมาณค่าในแต่ละดัชนีชี้วัดจำแนกตามรูปแบบการประมาณค่าพารามิเตอร์ และกรณีศึกษาทั้ง 16 กรณี

กรณีศึกษา	การวัดประสิทธิภาพของการประมาณค่า											
	MAD				SD				EUC			
	M1	M2	M3	M4	M1	M2	M3	M4	M1	M2	M3	M4
1. Nor-Nor-300-5				*				*		*		
2. Nor-Nor-300-10	*							*		*		
3. Nor-Nor-100-5				*				*		*		
4. Nor-Nor-100-10	*							*		*		
5. Nor-Beta-300-5	*			NA	*			NA		*		NA
6. Nor-Beta-300-10	*			NA	*			NA		*		NA
7. Nor-Beta-100-5	*			NA	*			NA		*		NA
8. Nor-Beta-100-10		*	*	NA	*			NA		*		NA
9. Gam-Nor-300-5		*	*	NA	*			NA		*		NA
10. Gam-Nor-300-10		*	*	NA	*			NA		*		NA
11. Gam-Nor-100-5	*			NA	*			NA		*		NA
12. Gam-Nor-100-10	*			NA	*			NA		*		NA
13. Gam-Beta-300-5		*	*	NA	*			NA		*		NA
14. Gam-Beta-300-10		*	*	NA	*			NA		*		NA
15. Gam-Beta-100-5	*			NA	*			NA		*		NA
16. Gam-Beta-100-10		*	*	NA	*			NA		*		NA
รวม	8 (50%)	6 (37.5%)	6 (37.5%)	2 (50%)	12 (75%)	0 (0%)	0 (0%)	4 (100%)	0 (0%)	16 (100%)	0 (0%)	0 (0%)

* เป็นวิธีการที่ให้ประสิทธิภาพดีสุดในแต่ละเงื่อนไข

NA ไม่สามารถวิเคราะห์ได้

ดังนั้นจะเห็นได้ว่า เมื่อพิจารณาความลำเอียง (Bias) ในการประมาณค่าโดยพิจารณาจากค่า MAD รูปแบบที่ 1 Original GIRM กับ รูปแบบที่ 4 Numerical Bayesian GIRM มีประสิทธิภาพสูงที่สุด แต่รูปแบบที่ 4 เป็นรูปแบบที่ประมาณค่าพารามิเตอร์ได้เฉพาะลักษณะการแจกแจงเริ่มแรกแบบปกติ (Normal) เท่านั้น โดยเมื่อพิจารณาในแต่ละกรณีพบว่า รูปแบบที่ 1

สามารถประมาณค่าได้มีประสิทธิภาพในกรณีที่ลักษณะการแจกแจงเริ่มแรกของผู้สอบแบบปกติ (Normal) และลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบเบต้า (Beta) ส่วนรูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B จะประมาณค่าได้มีประสิทธิภาพในกรณีที่ลักษณะการแจกแจงเริ่มแรกของผู้สอบแบบแกมมา (Gamma) และลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบเบต้า (Beta) ส่วนรูปแบบที่ 4 Numerical Bayesian GIRM จะประมาณค่าได้มีประสิทธิภาพในกรณีที่ลักษณะการแจกแจงเริ่มแรกของผู้สอบและข้อสอบแบบปกติ(Normal)

เมื่อพิจารณาประสิทธิภาพในการประมาณค่าด้านความไม่แน่นอนในการประมาณค่า (Uncertainty) พบว่า รูปแบบที่ 4 Numerical Bayesian GIRM มีประสิทธิภาพสูงที่สุด สำหรับลักษณะการแจกแจงเริ่มแรกแบบปกติ (Normal) ส่วนถ้าลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ตัวหนึ่งตัวใดไม่ได้มีลักษณะการแจกแจงเริ่มแรกแบบปกติ รูปแบบที่ 1 Original GIRM มีประสิทธิภาพสูงที่สุด และเมื่อพิจารณาประสิทธิภาพในการประมาณค่าด้านประสิทธิภาพขององค์ประกอบความแปรปรวนยุคลิด พบว่า รูปแบบที่ 2 AGIRM A มีประสิทธิภาพสูงที่สุด

4. ผลการวิเคราะห์ความแตกต่างของขนาดกลุ่มตัวอย่างที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ เมื่อพิจารณาความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีขนาดกลุ่มตัวอย่างแตกต่างกัน ทั้ง 26 คู่กรณี พบว่า มีคู่กรณีที่มีความแตกต่างกันของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ยอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิ้น 16 กรณี คิดเป็นร้อยละ 61.53 โดยแบ่งเป็น รูปแบบที่ 1 ทั้งสิ้น 6 คู่กรณี รูปแบบที่ 2 ทั้งสิ้น 5 คู่กรณี และรูปแบบที่ 3 ทั้งสิ้น 5 คู่กรณี จากผลการวิเคราะห์แสดงให้เห็นว่า จำนวนขนาดของกลุ่มตัวอย่างส่งผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์เมื่อพิจารณาจากค่า ความคลาดเคลื่อนเฉลี่ย (MAD) ใน 3 รูปแบบการประมาณค่าซึ่งได้แก่ รูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 โดยขนาดกลุ่มตัวอย่างมากขึ้นจะยิ่งช่วยลดความลำเอียงในการประมาณค่าพารามิเตอร์ ส่วนรูปแบบที่ 4 Numerical Bayesian GIRM ขนาดตัวอย่างไม่มีผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์

สำหรับความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ที่คำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) พบว่า เมื่อเปรียบเทียบค่าส่วนเบี่ยงเบนมาตรฐานระหว่างกรณีที่จำนวนขนาดตัวอย่างเพิ่มขึ้น ในขณะที่จำนวนข้อสอบเท่ากัน จะให้ค่าส่วนเบี่ยงเบนมาตรฐานที่ต่ำกว่าในทุกกรณี และทุกรูปแบบการประมาณค่าทั้ง 4 รูปแบบ แสดงให้เห็นว่า จำนวนขนาดกลุ่มตัวอย่าง ส่งผลต่อความไม่แน่นอนในประมาณค่า

องค์ประกอบความแปรปรวน โดยขนาดกลุ่มตัวอย่างเพิ่มขึ้นจะส่งผลให้การประมาณค่าองค์ประกอบความแปรปรวนมีความแน่นอนในประมาณค่าเพิ่มมากขึ้น

ส่วนการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีขนาดกลุ่มตัวอย่างแตกต่างกัน ทั้ง 26 คู่กรณี พบว่า มีคู่กรณีที่มีความแตกต่างกันของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิดอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิ้น 4 กรณี คิดเป็นร้อยละ 15.38 โดยแบ่งเป็น รูปแบบที่ 1 ทั้งสิ้น 2 คู่กรณี รูปแบบที่ 2 ทั้งสิ้น 1 คู่กรณี และรูปแบบที่ 3 ทั้งสิ้น 1 คู่กรณี จากผลการวิเคราะห์แสดงให้เห็นว่า จำนวนขนาดของกลุ่มตัวอย่างไม่ส่งผลต่อประสิทธิภาพในประมาณค่าองค์ประกอบความแปรปรวนเมื่อพิจารณาจากค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (EUC) อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

5. ผลการวิเคราะห์ความแตกต่างของความยาวของแบบสอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ เมื่อพิจารณาความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีจำนวนข้อสอบที่แตกต่างกัน ทั้ง 26 คู่กรณี พบว่า มีคู่กรณีที่มีความแตกต่างกันของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ยอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิ้น 24 กรณี คิดเป็นร้อยละ 92.31 โดยแบ่งเป็น รูปแบบที่ 1 รูปแบบที่ 2 และรูปแบบที่ 3 ทุกกรณี (รูปแบบละ 8 คู่กรณี) จากผลการวิเคราะห์แสดงให้เห็นว่า จำนวนข้อสอบส่งผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์เมื่อพิจารณาจากค่าความคลาดเคลื่อนเฉลี่ย (MAD) ใน 3 รูปแบบการประมาณค่า ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 โดยจำนวนข้อสอบมากขึ้นจะยิ่งช่วยลดความลำเอียงในการประมาณค่าพารามิเตอร์ ส่วนรูปแบบที่ 4 Numerical Bayesian GIRM จำนวนข้อสอบไม่มีผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์

สำหรับความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ที่คำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) พบว่า เมื่อเปรียบเทียบค่าส่วนเบี่ยงเบนมาตรฐานระหว่างกรณีที่จำนวนข้อสอบเพิ่มขึ้น ในขณะที่จำนวนขนาดกลุ่มตัวอย่างเท่ากัน จะให้ค่าส่วนเบี่ยงเบนมาตรฐานที่ต่ำกว่าในทุกกรณี และทุกรูปแบบการประมาณค่าทั้ง 4 รูปแบบ แสดงให้เห็นว่า จำนวนข้อสอบ ส่งผลต่อความไม่แน่นอนในประมาณค่าองค์ประกอบความแปรปรวน โดยจำนวนข้อสอบเพิ่มขึ้น จะส่งผลให้การประมาณค่าองค์ประกอบความแปรปรวนมีความแน่นอนในประมาณค่าเพิ่มมากขึ้น

ส่วนการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีขนาดกลุ่มตัวอย่างแตกต่างกัน ทั้งสิ้น 24 กรณี คิดเป็นร้อยละ 92.31 โดยแบ่งเป็น รูปแบบที่ 1 รูปแบบที่ 2 และรูปแบบที่ 3 ทุกกรณี (รูปแบบละ 8 คู่กรณี) จากผลการวิเคราะห์แสดงให้เห็นว่า จำนวนข้อสอบส่งผลต่อความลำเอียงในการประมาณค่าพารามิเตอร์เมื่อพิจารณาจากค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (EUC) ใน 3 รูปแบบการประมาณค่า ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 โดยจำนวนข้อสอบมากขึ้นจะยิ่งช่วยเพิ่มประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ส่วนรูปแบบที่ 4 Numerical Bayesian GIRM จำนวนข้อสอบไม่มีผลต่อประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน

6. ผลการวิเคราะห์ความไว (Sensitivity) ของผลการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ เมื่อพิจารณาการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบปกติ (Normal) กับแบบแกมมา (Gamma) ทั้ง 24 คู่กรณี พบว่า มีคู่กรณีศึกษาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิน 5 คู่กรณี คิดเป็นร้อยละ 20.83 โดยแบ่งเป็นรูปแบบที่ 1 Original GIRM 3 คู่กรณี คิดเป็นร้อยละ 12.50 และรูปแบบที่ 2 AGIRM 2 คู่กรณี คิดเป็นร้อยละ 8.33 ดังนั้นจากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ การกำหนดลักษณะการแจกแจงเริ่มแรกของผู้สอบระหว่าง การแจกแจงแบบปกติ (Normal) และการแจกแจงแบบแกมมา (Gamma) ไม่ส่งผลต่อผลการประมาณค่าพารามิเตอร์ของวิธีการ GIRM ทั้ง 4 รูปแบบของวิธีการประมาณค่า

สำหรับการวิเคราะห์ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ที่คำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) พบว่า เมื่อเปรียบเทียบค่าส่วนเบี่ยงเบนมาตรฐานระหว่างกรณีที่การแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ พบว่า ในจำนวนคู่กรณีทั้งหมด 24 คู่กรณี ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบแกมมา ให้ผลการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) ต่ำกว่า 21 คู่กรณี คิดเป็นร้อยละ 87.50 โดยมีคู่กรณีจำนวน 3 คู่กรณี คิดเป็นร้อยละ 12.50 ที่ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบปกติ ให้ผลการ

วิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) สูงกว่า จากผลการวิเคราะห์สามารถสรุปได้ว่า โดยส่วนใหญ่ ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบแกมมา ส่งผลต่อความไม่แน่นอนในประมาณค่าองค์ประกอบความแปรปรวนต่ำกว่าลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบปกติ

ส่วนการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบปกติ (Normal) และแบบแกมมา (Gamma) ทั้งสิ้น 24 คู่กรณี พบว่า มีคู่กรณีศึกษาที่มีค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 คิดเป็นร้อยละ 58.33 โดยแบ่งเป็น รูปแบบที่ 1 Original GIRM 4 คู่กรณี รูปแบบที่ 2 AGIRM A 5 คู่กรณี และรูปแบบที่ 3 AGIRM B 5 คู่กรณี คิดเป็นร้อยละ 16.67 20.83 และ 20.83 ตามลำดับ จากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่ลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบส่งผลต่อประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน เมื่อพิจารณาจากค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (EUC)

7. ผลการวิเคราะห์ความไว (Sensitivity) ของผลการประมาณค่าของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบที่มีต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ เมื่อพิจารณาการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation: MAD) ระหว่างกรณีศึกษาที่มีการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบปกติ (Normal) กับแบบเบต้า (Beta) ทั้ง 24 คู่กรณี พบว่า มีคู่กรณีศึกษาที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ทั้งสิ้น 15 คู่กรณี คิดเป็นร้อยละ 62.5 โดยแบ่งเป็นรูปแบบที่ 1 Original GIRM 6 คู่กรณี คิดเป็นร้อยละ 25.00 รูปแบบที่ 2 AGIRM A 4 คู่กรณี คิดเป็นร้อยละ 16.67 และรูปแบบที่ 3 AGIRM B 5 คู่กรณี คิดเป็นร้อยละ 20.83 จากผลการวิเคราะห์แสดงให้เห็นว่า โดยส่วนใหญ่การแจกแจงแบบปกติ จะให้ค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (MAD) ต่ำกว่า การแจกแจงแบบเบต้า ยกเว้นกรณี การกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบแกมมา ที่ให้ค่าเฉลี่ยของความคลาดเคลื่อนเฉลี่ย (MAD) การกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบเบต้าต่ำกว่าการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบปกติ

สำหรับการวิเคราะห์ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ที่คำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) โดยเมื่อเปรียบเทียบค่าส่วนเบี่ยงเบนมาตรฐานระหว่างกรณีที่การแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบในจำนวนคู่กรณีทั้งหมด 24 คู่กรณี พบว่าลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบเบต้า ให้ผลการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation: S.D.) ต่ำกว่าที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบแบบปกติเป็นส่วนใหญ่จำนวน 18 คู่กรณี คิดเป็นร้อยละ 75 จากผลการวิเคราะห์แสดงให้เห็นว่า ลักษณะการแจกแจงเริ่มแรกของของค่าพารามิเตอร์ของข้อสอบแบบเบต้า ส่งผลต่อความไม่แน่นอนในประมาณค่าองค์ประกอบความแปรปรวนต่ำกว่าลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบปกติ

ส่วนการวิเคราะห์ความแตกต่างของค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance: EUC) ระหว่างกรณีศึกษาที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของข้อสอบแบบปกติ (Normal) และแบบเบต้า (Beta) ทั้งสิ้น 24 คู่กรณี พบว่ามีคู่กรณีศึกษาที่มีค่าเฉลี่ยของค่าองค์ประกอบความแปรปรวนระยะทางยุคลิดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 คิดเป็นร้อยละ 50.00 โดยแบ่งเป็น รูปแบบที่ 1 Original GIRM 4 คู่กรณี รูปแบบที่ 2 AGIRM A 4 คู่กรณี และรูปแบบที่ 3 AGIRM B 4 คู่กรณี โดยเป็นที่น่าสังเกตว่าคู่กรณีที่มีความแตกต่างกันอย่างมีนัยสำคัญจะมีลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบแกมมา (Gamma) ทั้งหมดทั้ง 9 รูปแบบการวิเคราะห์ ดังนั้นจึงสามารถสรุปได้ว่า ถ้าลักษณะการแจกแจงเริ่มแรกของผู้สอบมีลักษณะการแจกแจงแบบแกมมา (Gamma) การแจกแจงเริ่มแรกของข้อสอบจะส่งผลต่อการประมาณค่าพารามิเตอร์ โดยลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบปกติ (Normal) จะให้ค่าประสิทธิภาพของการประมาณค่าขององค์ประกอบความแปรปรวนที่สูงกว่าลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบเบต้า (Beta)

จากการศึกษาเรื่องความแกร่ง (Robustness) ของวิธีการ โดยศึกษาจากขนาดของกลุ่มตัวอย่าง ความยาวของแบบสอบ ลักษณะการแจกแจงเริ่มแรกของผู้สอบและข้อสอบ สามารถสรุปได้ดังตารางที่ 31

ตารางที่ 31 ผลการศึกษาอิทธิพลของขนาดตัวอย่าง ความยาวของแบบสอบ การแจกแจงเริ่มแรกของผู้สอบ และการแจกแจงเริ่มแรกของข้อสอบจำแนกตามตัววัดประสิทธิภาพและรูปแบบการประมาณค่า

ตัววัดประสิทธิภาพ	รูปแบบการประมาณค่า	ขนาดตัวอย่าง	ความยาวของแบบสอบ	การแจกแจงเริ่มแรกของผู้สอบ	การแจกแจงเริ่มแรกของข้อสอบ
MAD	Original GIRM	✓	✓	X	✓
	AGIRM A	✓	✓	X	✓
	AGIRM A	✓	✓	X	✓
	Numerical GIRM	X	X	X	✓
SD	Original GIRM	✓	✓	✓	✓
	AGIRM A	✓	✓	✓	✓
	AGIRM A	✓	✓	✓	✓
	Numerical GIRM	✓	✓	✓	✓
EUC	Original GIRM	X	✓	✓	*
	AGIRM A	X	✓	✓	*
	AGIRM A	X	✓	✓	*
	Numerical GIRM	X	X	NA	NA

✓ ส่งผลต่อการประมาณค่าประสิทธิภาพ

X ไม่ส่งผลต่อการประมาณค่าประสิทธิภาพ

NA ไม่สามารถทำการวิเคราะห์ได้

* เฉพาะการแจกแจงเริ่มแรกของผู้สอบแบบแกมมา การระบุการแจกแจงเริ่มแรกของข้อสอบจึงจะส่งผลต่อการประมาณค่า

จากตารางที่ 31 แสดงให้เห็นว่า ขนาดของกลุ่มตัวอย่างส่งผลต่อการวัดประสิทธิภาพในด้านความลำเอียงในการประมาณค่าเกือบทุกรูปแบบ ยกเว้นรูปแบบที่ 4 Numerical GIRM ส่วนความไม่แน่นอนในการประมาณค่า พบว่า ขนาดกลุ่มตัวอย่างส่งผลต่อประสิทธิภาพในการประมาณค่าในทุกรูปแบบ และเมื่อมองในเรื่องของการวิเคราะห์ประสิทธิภาพขององค์ประกอบความแปรปรวนยุคลิด พบว่า ขนาดของกลุ่มตัวอย่างไม่ส่งผลประสิทธิภาพในการประมาณค่า

เมื่อพิจารณาในเรื่องของ ความยาวของแบบสอบ พบว่า ส่งผลต่อการการวัดประสิทธิภาพในด้านความลำเอียงในการประมาณค่าและการวิเคราะห์ประสิทธิภาพขององค์ประกอบความแปรปรวนยุคลิด เกือบทุกรูปแบบ ยกเว้นรูปแบบที่ 4 Numerical GIRM ส่วนความไม่แน่นอนในการประมาณค่า พบว่า ขนาดกลุ่มตัวอย่างส่งผลต่อประสิทธิภาพในการประมาณค่าในทุกรูปแบบ

สำหรับในเรื่องการแจกแจงเริ่มแรกของผู้สอบ พบว่า ไม่ส่งผลต่อการวัดประสิทธิภาพในด้านความลำเอียง ส่วนในเรื่องของส่วนความไม่แน่นอนในการประมาณค่าและการวิเคราะห์ประสิทธิภาพขององค์ประกอบความแปรปรวนยุคลิด พบว่าการแจกแจงเริ่มแรกของผู้สอบ ส่งผลต่อการวัดประสิทธิภาพในด้านความลำเอียงในรูปแบบ

ส่วนในเรื่องการแจกแจงเริ่มแรกของข้อสอบ พบว่า ส่งผลต่อการวัดประสิทธิภาพในด้านความลำเอียงในการประมาณค่าและความไม่แน่นอนในการประมาณค่าในทุกรูปแบบ และส่งผลการวิเคราะห์ประสิทธิภาพขององค์ประกอบความแปรปรวนยุคลิดเฉพาะในกรณีที่การแจกแจงเริ่มแรกของผู้สอบเป็นแบบแกมมาเท่านั้น

อภิปรายผลการวิจัย

สืบเนื่องจากความพยามในการรวมโมเดลของการวัดผลที่สำคัญ 2 ทฤษฎี ระหว่างทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory: GT) ซึ่งเป็นโมเดลแบบสุ่ม (Sampling model) กับทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ซึ่งเป็นโมเดลการกำหนดมาตรวัด (Scaling model) (Brenan, 2001) จนกลายมาเป็นวิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดด้วยโมเดลการตอบสนองข้อสอบ (Generalizability in Item Response Modeling: GIRM) ที่ Briggs และ Wilson ได้พัฒนาขึ้นในปี 2007 โดยในเริ่มแรกได้พัฒนาเฉพาะการประมาณค่าพารามิเตอร์ในบริบทของการออกแบบการวัดผลอย่างง่ายซึ่งได้แก่ 1PL-IRT แบบ P X I และเนื่องจากเป็นการพัฒนาในเบื้องต้น ผลที่ได้จากการประมาณค่ายังไม่คงที่ในการประมาณค่าพารามิเตอร์ (Uncertainty) กอปรกับยังไม่ได้ศึกษาในเรื่อง อิทธิพลของจำนวนข้อสอบและขนาดของผู้สอบว่ามีผลกระทบต่อการประมาณค่าพารามิเตอร์หรือไม่ และยังไม่ได้มีการศึกษาในเรื่องของความไวในการประมาณค่าต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ จึงเป็นที่มาของงานวิจัยในครั้งนี้เกิดขึ้น ดังนั้นในการวิจัยครั้งนี้จึงมุ่งศึกษาประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ทั้ง 4 รูปแบบการประมาณค่า ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM ซึ่งพัฒนาโดย Brigg และ Wilson (2007) กับรูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B รูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งผู้วิจัยเป็นผู้พัฒนาขึ้น นอกจากนี้ยังศึกษาถึงอิทธิพลของขนาดกลุ่มอย่างและจำนวนข้อสอบที่มีต่อประสิทธิภาพของวิธีวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) รวมทั้งยังศึกษาถึงความไว (Sensitivity) ของผลการวิเคราะห์ข้อสอบด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) ที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์

ของข้อสอบ (β) และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ของผู้สอบ (θ) ดังนั้น ในอภิปรายผลการวิจัยนี้จะนำเสนอโดยยึดตามวัตถุประสงค์และผลที่ได้รับจากการวิจัย มีรายละเอียดตามประเด็นดังต่อไปนี้

1. การประมาณค่าพารามิเตอร์ตามแนวการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ (IRT) ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดด้วยโมเดลการตอบสนองข้อสอบ (GIRM)

จากการวิเคราะห์ค่าพารามิเตอร์ตามแนวทฤษฎีการตอบสนองข้อสอบขอรูปแบบการประมาณของวิธีการ GIRM พบว่า ค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ของผู้สอบด้วยวิธีการ MCMC ซึ่งใช้ในการประมาณค่าของวิธีการ GIRM ในรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B ให้ค่าเฉลี่ยของพารามิเตอร์ของข้อสอบและพารามิเตอร์ของผู้สอบไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 กับการประมาณค่าพารามิเตอร์ด้วยวิธีการ Numerical bayesian method ในรูปแบบที่ 4 Numerical Bayesian GIRM แสดงว่าการประมาณค่าพารามิเตอร์ของผู้สอบทั้งสองวิธีให้ผลที่ไม่แตกต่างกัน แต่เมื่อพิจารณค่าส่วนเบี่ยงเบนมาตรฐาน (S.D.) พบว่า ค่าส่วนเบี่ยงเบนมาตรฐานที่ได้จากการประมาณค่าด้วยวิธีการ Numerical bayesian method ในรูปแบบที่ 4 ให้ค่าที่ต่ำกว่าค่าส่วนเบี่ยงเบนมาตรฐานที่ได้จากการประมาณค่าวิธีการ MCMC ในรูปแบบที่ 1 รูปแบบที่ 2 และ รูปแบบที่ 3 แสดงให้เห็นว่าค่าเฉลี่ยของพารามิเตอร์ที่ได้จากการประมาณค่าตามแนวแนวทฤษฎีการตอบสนองข้อสอบไม่แตกต่างกันระหว่างการประมาณค่าด้วยวิธี MCMC กับวิธี Numerical แต่การประมาณค่าด้วยวิธีการ Numerical มีความแน่นอนในการประมาณค่ามากกว่าโดยพิจารณาจาก ค่าส่วนเบี่ยงเบนมาตรฐาน ทั้งนี้สาเหตุอาจจะเป็นเพราะในการประมาณค่าด้วยวิธีการประมาณค่าด้วยวิธีการ MCMC ของการประมาณค่าแบบเบส์นั้นการประมาณค่าพารามิเตอร์ของผู้สอบจะขึ้นอยู่กับการประมาณค่าของข้อสอบ และการประมาณค่าของข้อสอบจะขึ้นอยู่กับการประมาณค่าผู้สอบ จึงทำให้ค่าพารามิเตอร์ทั้งสองไม่เป็นอิสระต่อกัน ซึ่งฝ่าฝืนข้อตกลงเบื้องต้นในการประมาณค่าของค่าพารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบ

นอกจากนี้เมื่อพิจารณาการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบตามแนวทฤษฎีการตอบสนองของข้อสอบ (IRT) พบว่า สามารถประมาณค่าพารามิเตอร์ได้เฉพาะกรณีการแจกแจงเริ่มแรกผู้สอบและการแจกแจงเริ่มแรกของข้อสอบมีลักษณะการกระจายแบบโค้งปกติเท่านั้น ทั้งนี้เนื่องมาจากการประมาณค่าพารามิเตอร์ของรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B ใช้การประมาณค่าด้วยวิธีการ MCMC ซึ่งเป็นการประมาณ

ค่าพารามิเตอร์โดยการสุ่มค่าพารามิเตอร์ขึ้นมาทำซ้ำในหลายๆ ครั้ง ซึ่งต่างจากการประมาณค่าพารามิเตอร์ของรูปแบบที่ 4 Numerical bayesian method ซึ่งใช้การประมาณค่าด้วยวิธีการ Numerical method ซึ่งใช้วิธีการทำให้ฟังก์ชันการแจกแจงภายหลังมีค่าสูงสุด จากการหาอนุพันธ์ของฟังก์ชันและกำหนดให้ค่าของอนุพันธ์มีค่าเท่ากับศูนย์ ซึ่งการประมาณค่าในลักษณะนี้ ถ้าลักษณะการแจกแจงไม่ใช่รูปแบบการแจกแจงที่เข้ากันได้ เช่น ลักษณะการแจกแจงแบบแกมมา (Gamma) กับ ลักษณะการแจกแจงแบบเบต้า (Beta) จะทำให้ไม่สามารถมีการดำเนินการทางคณิตศาสตร์ (Operation) ได้ (Ntzoufras, 2009) จึงเป็นเหตุผลที่รูปแบบการประมาณค่าในรูปแบบที่ 4 Numerical bayesian method จึงสามารถประมาณค่าพารามิเตอร์ได้เฉพาะลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบปกติกับลักษณะการแจกแจงเริ่มแรกของผู้สอบแบบปกติเท่านั้น

ดังนั้นจะเห็นได้ว่าในเรื่องการประมาณค่าพารามิเตอร์ตามแนวทฤษฎีการตอบสนองข้อสอบ ทั้งการประมาณค่าจากรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B ซึ่งใช้การประมาณค่าด้วยวิธี MCMC กับการประมาณค่าจากรูปแบบที่ 4 Numerical bayesian method ซึ่งใช้การประมาณค่าด้วยวิธี Numerical method ให้ผลที่ไม่แตกต่างกัน แต่จากการประมาณค่าจากวิธีการที่ 4 จะให้ค่าประมาณที่คงที่มากกว่า และสามารถทำการประมาณค่าได้เฉพาะกรณีที่มีลักษณะการแจกแจงเริ่มแรกของข้อสอบและผู้สอบแบบปกติเท่านั้น

2. การประมาณค่าพารามิเตอร์ตามแนวการวิเคราะห์ด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (GT) ด้วยวิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดด้วยโมเดลการตอบสนองข้อสอบ(GIRM)

จากการวิเคราะห์ค่าพารามิเตอร์ตามแนวทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด(GT) ของรูปแบบการประมาณค่าของวิธีการ GIRM พบว่า การประมาณค่าองค์ประกอบความแปรปรวนในรูปแบบที่1 Original GIRM และรูปแบบที่ 4 Numerical bayesian GIRM ซึ่งใช้การประมาณค่าด้วยสูตร Kolen และ Harris กับรูปแบบที่ 2 AGIRM A ซึ่งใช้การประมาณค่าด้วยสูตร ANOVA และรูปแบบที่ 3 AGIRM B ซึ่งใช้การประมาณค่าด้วยสูตร Kolen และ Harris และ ANOVA ให้ผลการประมาณค่าองค์ประกอบความแปรปรวนไม่แตกต่างกัน ซึ่งสอดคล้องกับงานวิจัยของ Briggs และ Wilson (2007) ที่เปรียบเทียบค่าองค์ประกอบความแปรปรวนของค่าสารสนเทศที่ได้จากวิธีการวิเคราะห์ด้วยวิธี GIRM กับการวิเคราะห์ด้วยวิธี Classical GT พบว่า ค่าองค์ประกอบความแปรปรวนที่ได้จากสองวิธีการวิเคราะห์ให้ผลที่ไม่แตกต่างกัน

เมื่อพิจารณาค่าส่วนเบี่ยงเบนมาตรฐาน (S.D.) ของค่าองค์ประกอบความแปรปรวนในแต่ละชุดข้อมูล (Data set) พบว่า การประมาณค่าองค์ประกอบความแปรปรวนจากรูปแบบที่ 4 Numerical bayesian GIRM ให้ค่าส่วนเบี่ยงเบนมาตรฐานต่ำที่สุด แสดงว่ามีความคงที่ในการประมาณค่าองค์ประกอบความแปรปรวนมากที่สุด ทั้งนี้สาเหตุน่าจะเนื่องมาจากขั้นตอนการวิเคราะห์องค์ประกอบความแปรปรวนเป็นขั้นตอนที่ 3 ของวิธีการ GIRM ดังนั้น ข้อมูลก่อนนำมาวิเคราะห์เป็นข้อมูลที่ได้จากผลการวิเคราะห์ในขั้นตอนที่ 1 ซึ่งในรูปแบบที่ 4 นี้ใช้การประมาณค่าแบบ Numerical Bayesian ซึ่งไม่ฝ่าฝืนข้อตกลงเบื้องต้นของความไม่เป็นอิสระต่อกัน เหมือนการประมาณค่าด้วยวิธีการ MCMC ดังที่ได้ให้เหตุผลไว้ในอภิปรายผลการวิจัยในข้อที่ 1

นอกจากนี้เมื่อพิจารณาผลการวิเคราะห์องค์ประกอบความแปรปรวนของรูปแบบที่ 2 AGIRM B พบว่า ในการประมาณค่าองค์ประกอบความแปรปรวนไม่สามารถวิเคราะห์แยกองค์ประกอบความแปรปรวนของปฏิสัมพันธ์ (interaction) กับความคลาดเคลื่อนส่วนที่เหลือ (residual) ได้ ทั้งนี้เนื่องจากการวิเคราะห์ในรูปแบบที่ 2 AGIRM B ใช้การวิเคราะห์องค์ประกอบความแปรปรวนด้วย การวิเคราะห์ความแปรปรวน (ANOVA) ซึ่งการวิเคราะห์ด้วย ANOVA ไม่สามารถวิเคราะห์แยกแหล่งความแปรปรวนที่มาจากปฏิสัมพันธ์ (interaction) กับความคลาดเคลื่อนส่วนที่เหลือ (residual) ได้

3.การเปรียบเทียบประสิทธิภาพในการประมาณค่าของรูปแบบการประมาณค่าทั้ง 4 รูปแบบของวิธีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดด้วยโมเดลการตอบสนองข้อสอบ (GIRM)

การเปรียบเทียบประสิทธิภาพของการประมาณค่าในแต่ละรูปแบบการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ(GIRM) ทั้ง 4 รูปแบบ ซึ่งได้แก่ รูปแบบที่ 1 Original GIRM ซึ่ง พัฒนาโดย Brigg และ Wilson (2007) และรูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B รูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งพัฒนาวิธีการประมาณค่าโดยผู้วิจัย โดยในการวัดประสิทธิภาพของตัวประมาณค่าจะวัดจากดัชนี 3 ประเภท ซึ่งได้แก่ ความลำเอียงในการประมาณค่า (Biased estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าความคลาดเคลื่อนเฉลี่ย (Mean Average Deviation; MAD) ความไม่แน่นอนในการประมาณค่า (Uncertainty estimator) ซึ่งคำนวณได้จากการวิเคราะห์ค่าส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation; S.D.) และประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวน ซึ่งคำนวณได้จากการวิเคราะห์ค่าองค์ประกอบความแปรปรวนระยะทางยุคลิด (Euclidean Distance; EUC)

ผลการศึกษา พบว่า เมื่อพิจารณาความลำเอียง (Bias) ในการประมาณค่าโดยพิจารณาจากค่า MAD รูปแบบที่ 1 Original GIRM กับ รูปแบบที่ 4 Numerical Bayesian GIRM มีประสิทธิภาพสูงที่สุด แต่รูปแบบที่ 4 เป็นรูปแบบที่ประมาณค่าพารามิเตอร์ได้เฉพาะลักษณะการแจกแจงเริ่มแรกแบบปกติ (Normal) เท่านั้น โดยเมื่อพิจารณาในแต่ละกรณีพบว่า รูปแบบที่ 1 จะประมาณค่าได้มีประสิทธิภาพในกรณีที่ลักษณะการแจกแจงเริ่มแรกของผู้สอบแบบปกติ (Normal) และลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบเบต้า (Beta) ส่วนรูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B จะประมาณค่าได้มีประสิทธิภาพในกรณีที่ลักษณะการแจกแจงเริ่มแรกของผู้สอบแบบแกมมา (Gamma) และลักษณะการแจกแจงเริ่มแรกของข้อสอบแบบเบต้า(Beta) ส่วนรูปแบบที่ 4 Numerical Bayesian GIRM จะประมาณค่าได้มีประสิทธิภาพในกรณีที่ลักษณะการแจกแจงเริ่มแรกของผู้สอบและข้อสอบแบบปกติ(Normal) ดังนั้นจะเห็นได้ว่า ถ้าลักษณะการแจกแจงเริ่มแรกที่ไม่ปกติ เช่น ลักษณะการแจกแจงเริ่มแรกแบบแกมมา (Gamma) กับลักษณะการแจกแจงเริ่มแรกแบบเบต้า (Beta) รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B จะให้การประมาณค่าที่มีประสิทธิภาพดีกว่า ทั้งนี้สาเหตุมาจากในขั้นตอนที่ 2 ของการประมาณใน รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B มีการคำนวณหา Predictive Observed Matrix มีลักษณะคล้ายคลึงกับข้อมูลจริงแบบสองค่า (dichotomous) ซึ่งถ้าผู้ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน ในขณะที่รูปแบบที่ 1 Original GIRM แบะรูปแบบที่ 4 Numerical Bayesian GIRM มีการคำนวณหา Expected Response Matrix ซึ่งข้อมูลมีลักษณะเป็นความน่าจะเป็น ซึ่งได้มาจากการคำนวณจากสูตร Logistic function ดังนั้นค่าที่ได้จึงไม่เป็นอิสระจากกัน จึงทำให้การคำนวณในขั้นตอนที่ 3 จึงไม่เป็นอิสระต่อกัน นอกจากนี้ในการคำนวณ Predictive Observed Matrix ยังเป็นการปรับลักษณะการกระจายจึงทำให้สามารถประมาณค่า ลดความลำเอียงให้มากที่สุดเมื่อลักษณะการแจกแจงไม่ใช่ลักษณะแบบปกติ(Normal distribution)

เมื่อพิจารณาประสิทธิภาพในการประมาณค่าด้านความไม่แน่นอนในการประมาณค่า (Uncertainty) พบว่า รูปแบบที่ 4 Numerical Bayesian GIRM มีประสิทธิภาพสูงที่สุด สำหรับลักษณะการแจกแจงเริ่มแรกแบบปกติ (Normal) ถ้าลักษณะการแจกแจกเริ่มแรกของค่าพารามิเตอร์ตัวหนึ่งตัวใดไม่ได้มีลักษณะการแจกแจงเริ่มแรกแบบปกติ รูปแบบที่ 1 Original GIRM มีประสิทธิภาพสูงที่สุด ทั้งนี้น่าจะมีสาเหตุมาจากการประมาณค่าพารามิเตอร์ในรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B ประมาณค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ของผู้สอบด้วยวิธีการ MCMC ในขณะที่รูปแบบที่ 4 Numerical Bayesian GIRM ประมาณค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ของผู้สอบด้วยวิธีการ

Numerical bayesian method ทั้งนี้สาเหตุอาจจะเป็นเพราะในการประมาณค่าด้วยวิธีการประมาณค่าด้วยวิธีการ MCMC ของการประมาณค่าแบบเบส์นั้นการประมาณค่าพารามิเตอร์ของผู้สอบจะขึ้นอยู่กับการประมาณค่าของข้อสอบ และการประมาณค่าของข้อสอบจะขึ้นอยู่กับการประมาณค่าผู้สอบ จึงทำให้ค่าพารามิเตอร์ทั้งสองไม่เป็นอิสระต่อกัน ซึ่งฝ่าฝืนข้อตกลงเบื้องต้นในการประมาณค่าของค่าพารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบ จึงทำให้การประมาณค่าด้วยรูปแบบที่ 1 Original GIRM รูปแบบที่ 2 AGIRM A และรูปแบบที่ 3 AGIRM B มีความไม่แน่นอนในการประมาณค่า (Uncertainty) มากกว่า การประมาณค่าด้วยรูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งสอดคล้องกับข้อเสนอแนะของบทความ GIRM ของ Briggs และ Wilson (2007)

ข้อเสนอแนะ

1.ข้อเสนอแนะสำหรับการนำผลการวิจัยไปใช้

การวิจัยครั้งนี้เป็นการศึกษาประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ(GIRM) ซึ่งวิธีการนี้เป็นวิธีการวิเคราะห์ข้อสอบที่ให้ค่าพารามิเตอร์ทั้งในเชิงการวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ(IRT) และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด(GT) ดังนั้น จึงเป็นการวิเคราะห์ครั้งเดียวแต่ให้ค่าพารามิเตอร์ทั้งสองวิธี การวิจัยในครั้งนี้ได้เปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าใน 4 รูปแบบ ได้แก่ รูปแบบที่ 1 Original GIRM ซึ่ง พัฒนาโดย Brigg และ Wilson (2007) กับรูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B และรูปแบบที่ 4 Numerical Bayesian GIRM ซึ่งผู้วิจัยเป็นผู้พัฒนาขึ้น ให้ผลการวิจัยที่จะนำไปเป็นข้อมูลในการเลือกใช้รูปแบบการประมาณค่าของวิธีการ GIRM สามารถสรุปได้ดังนี้

1.การประมาณค่ารูปแบบที่ 1 Original GIRM เหมาะสมกับสถานการณ์ที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ตัวใดตัวหนึ่งไม่ใช่การแจกแจงแบบปกติ และลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์อีกตัวเป็นแบบปกติ เช่นลักษณะการแจกแจงเริ่มแรกของข้อสอบมีลักษณะแบบปกติ และลักษณะการแจกแจงของผู้สอบแบบเบต้า เป็นต้น เนื่องจากจะให้ค่าพารามิเตอร์ที่ได้มีความลำเอียงต่ำและมีความคงที่ในการประมาณค่าของค่าพารามิเตอร์

นอกจากนี้สิ่งที่ต้องคำนึงถึงเมื่อใช้รูปแบบการประมาณค่ารูปแบบนี้คือ ขนาดกลุ่มตัวอย่างและความยาวของแบบสอบ โดยถ้าใช้ขนาดกลุ่มตัวอย่างและจำนวนข้อสอบจำนวนมากจะยิ่งทำให้ประสิทธิภาพของการประมาณค่ายิ่งเพิ่มสูงขึ้น และเมื่อพิจารณาถึงความไวที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ พบว่า รูปแบบนี้มีความทนทาน (Robustness) ต่อการกำหนด

ลักษณะการแจกแจงเริ่มแรกของผู้สอบ แต่มีความไว (Sensitivity) ต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของข้อสอบ

2. การประมาณค่ารูปแบบที่ 2 AGIRM A รูปแบบที่ 3 AGIRM B เหมาะสมกับสถานการณ์ที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ไม่ใช่การแจกแจงแบบปกติ เช่นลักษณะการแจกแจงเริ่มแรกของข้อสอบมีลักษณะแบบแกมมา และลักษณะการแจกแจงของผู้สอบแบบเบต้า เป็นต้น เนื่องจากจะให้ค่าพารามิเตอร์ที่ได้มีความคงที่ในการประมาณค่าของค่าพารามิเตอร์และมีประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนของยุคลิดสูง

นอกจากนี้สิ่งที่ต้องคำนึงถึงเมื่อใช้รูปแบบการประมาณค่ารูปแบบนี้คือ ขนาดกลุ่มตัวอย่างและความยาวของแบบสอบ โดยถ้าใช้ขนาดกลุ่มตัวอย่างและจำนวนข้อสอบจำนวนมากจะยิ่งทำให้ประสิทธิภาพของการประมาณค่ายิ่งเพิ่มสูงขึ้น และเมื่อพิจารณาถึงความไวที่มีต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์ พบว่า รูปแบบนี้มีความไว (Sensitivity) ทั้งต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของข้อสอบและผู้สอบ

3. การประมาณค่ารูปแบบที่ 4 Numerical Bayesian GIRM สามารถประมาณค่าพารามิเตอร์เฉพาะสถานการณ์ที่มีลักษณะการแจกแจงเริ่มแรกของค่าพารามิเตอร์แบบปกติเท่านั้น เช่นลักษณะการแจกแจงเริ่มแรกของข้อสอบมีลักษณะแบบปกติ และลักษณะการแจกแจงของผู้สอบแบบปกติ แต่จะให้ค่าพารามิเตอร์ที่ประมาณได้มีประสิทธิภาพสูงมาก เนื่องจากจะให้ค่าพารามิเตอร์ที่ได้มีความลำเอียงต่ำ มีความคงที่ในการประมาณค่าของค่าพารามิเตอร์และมีประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนของยุคลิดสูง

รูปแบบที่ 4 เป็นรูปแบบที่ไม่ผันแปรต่อขนาดตัวอย่างและความยาวของแบบสอบ ดังนั้น การวิเคราะห์ด้วยรูปแบบนี้จึงสามารถวิเคราะห์จำนวนของข้อสอบหรือผู้สอบน้อยได้มีประสิทธิภาพ นอกจากนี้ รูปแบบนี้มีความทนทาน (Robustness) ต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของผู้สอบ แต่มีความไว (Sensitivity) ต่อการกำหนดลักษณะการแจกแจงเริ่มแรกของข้อสอบ

2. ข้อเสนอแนะสำหรับการทำวิจัยครั้งต่อไป

1. การศึกษาครั้งนี้เป็นการศึกษาวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบในรูปแบบการวัด 1-PL IRT แบบ P X I ซึ่งพิจารณาแหล่งความคลาดเคลื่อนเพียงแค่ 2 แหล่ง ดังนั้นในการวิจัยครั้งต่อควรมีขยายการศึกษาในรูปแบบการวัด 2-PL IRT หรือ 3-PL IRT นอกจากนี้ยังอาจจะมีการศึกษาแหล่งความคลาดเคลื่อนมากกว่า 2 แหล่งความคลาดเคลื่อนเพื่อให้สอดคล้องกับลักษณะการออกแบบการวัดผลในสภาพจริง

2. การประมาณค่าองค์ประกอบความแปรปรวนของแหล่งความคลาดเคลื่อนในครั้งนี้ใช้วิธีการประมาณค่าโดยใช้สูตร Kolen และ Harris ในรูปแบบที่ 1 Original GIRM รูปแบบที่ 3 AGIRM B และรูปแบบที่ 4 Numerical Bayesian GIRM และใช้การประมาณค่าโดยใช้สูตร ANOVA ในวิธีการที่ 2 AGIRM A แต่จากงานวิจัยของ Wiley (2001) พบว่า การประมาณค่าองค์ประกอบความแปรปรวนด้วยวิธีการ Bootstrap ทำให้สามารถลดความลำเอียงในการประมาณได้ดี ดังนั้นในการทำวิจัยครั้งต่อไป จึงควรมีการศึกษารูปแบบการประมาณค่าของวิธีการ GIRM โดยใช้การประมาณค่าแบบ Bootstrap มาใช้ในการประมาณค่าองค์ประกอบความแปรปรวน

3. การวิจัยครั้งนี้ยังไม่มีการศึกษาเกี่ยวกับ อิทธิพลของจำนวนข้อมูลขาดหายไป (missing value) ในระดับต่างๆ ที่ส่งผลต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ในรูปแบบต่างๆ ของวิธีการสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ เช่น ถ้ามีข้อมูลขาดหายไปร้อยละ5 ร้อยละ 10 และร้อยละ 20 จะส่งผลอย่างไรต่อประสิทธิภาพในการประมาณค่าในแต่ละรูปแบบการประมาณค่าของวิธีการ GIRM

4. การวิจัยครั้งนี้ รูปแบบที่ 4 Numerical Bayesian GIRM สามารถทำการวิเคราะห์ได้เฉพาะลักษณะที่มีการแจกแจงเริ่มแรกแบบปกติ (Normal DIstribution)เท่านั้น เนื่องจากลักษณะการแจกแจงเริ่มแรกแบบอื่นๆไม่สามารถดำเนินการทางคณิตศาสตร์ที่เข้ากันได้ ดังนั้น ในการวิจัยครั้งต่อไป ถ้าต้องมีการวิเคราะห์ในรูปแบบที่ 4 Numerical Bayesian GIRM สำหรับลักษณะการแจกแจงเริ่มแรกแบบอื่นที่ไม่ใช่แบบปกติให้แปลงคะแนนดิบผลการสอบ (raw data) เป็นรูปแบบคะแนนมาตรฐาน (standardized score) ก่อนทำการวิเคราะห์ ซึ่งน่าจะทำให้สามารถวิเคราะห์ผลการวัดจากรูปแบบที่ 4 ได้

5. การวิจัยครั้งเป็นการศึกษาจากข้อมูลจำลอง (Simulation) เพียงอย่างเดียวสำหรับการศึกษาประสิทธิภาพของรูปแบบการประมาณค่าของวิธีการ GIRM เพียงอย่างเดียว ดังนั้นในการศึกษาครั้งต่อไป ควรจะมีการศึกษาเปรียบเทียบกับข้อมูลจริงด้วย ว่าให้ผลการศึกษาที่สอดคล้องหรือใกล้เคียงกันกับข้อมูลจำลองหรือไม่

รายการอ้างอิง

ภาษาไทย

เกศมณี พยัคฆ์. (2543). *การเปรียบเทียบความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ของแบบทดสอบความเข้าใจในการอ่านที่มีจำนวนข้อและจำนวนกลุ่มตัวอย่างต่างกันที่วิเคราะห์ด้วยราส์ชโมเดล.* วิทยานิพนธ์ปริญามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

คณาจารย์ภาควิชาคณิตศาสตร์ คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย. (2551). *ความน่าจะเป็นและสถิติ.* พิมพ์ครั้งที่ 11. กรุงเทพมหานคร: โรงพิมพ์ห้างหุ้นส่วนจำกัดพิทักษ์การพิมพ์.

ชยุตม์ ภิรมย์สมบัติ. (2547). *คุณสมบัติของตัวประมาณค่าความเข้มอิทธิพล: การเปรียบเทียบระหว่างทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ.* วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ธีระพร วีระถาวร. (2539). *ความน่าจะเป็นกับการประยุกต์.* พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: บริษัทวิทยพัฒน์ จำกัด.

พัชรี จันทร์เพ็ง. (2550). *การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีตอบสนองข้อสอบแบบพหุมิติภายใต้การหมุนแกนโครงสร้างเชิงมิติและระดับความสัมพันธ์ที่แตกต่างกัน.* วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

มานพ วราศักดิ์. (2548). *ทฤษฎีความน่าจะเป็น.* กรุงเทพมหานคร: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

รัตนา ศรีเหรัญ. (2539). *การเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์ของแบบจำลองโลจิสติค 3 พารามิเตอร์ ระหว่างวิธีแมกซิมัม ไลค์ลิฮูด วิธีฮิวริสติคและวิธีของเบส์ เมื่อขนาดของกลุ่มตัวอย่างและจำนวนข้อสอบต่างกันด้วยเทคนิคมอนติคาร์โล.* วิทยานิพนธ์ปริญาดุษฎีบัณฑิต, สาขาวิชาการทดสอบและวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

วิชุดา บัวคง. (2532). *การเปรียบเทียบประสิทธิผลของวิธีประมาณค่าพารามิเตอร์ของแบบจำลองโลจิสติก 3 พารามิเตอร์ระหว่างวิธีแมกซิมัสไลค์ลิฮด วิธีฮิวริสติก และวิธีของเบย์ในแบบสอบวัดผลสัมฤทธิ์*. วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วินัย วงศ์ฤทัยวัฒนา. (2532). *การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ของโมเดลโลจิสติกแบบสองพารามิเตอร์ระหว่างวิธีของเบส์กับวิธีแมกซิมไมลค์ลีฮูด*. วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2548). *ทฤษฎีการทดสอบแบบดั้งเดิม*. พิมพ์ครั้งที่ 5. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่*. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริรัตน์ วงศ์ประกรณ์กุล. (2539). *การจำลองโดยใช้เทคนิคมอนติคาร์โล*. วารสารวิทยาศาสตร์มหาวิทยาลัยขอนแก่น. 24 (4) : 240 – 246.

สุกัญญรัตน์ คงงาม. (2539). *การเปรียบเทียบคุณสมบัติของตัวประมาณค่าพารามิเตอร์ที่ได้จากกลุ่มตัวอย่างสุ่มแบบหลายขั้นตอน ระหว่างวิธีสุ่มแบบง่ายกับแบบมีระบบ*. วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

สมกิต วงศ์มาก. (2549). *การศึกษาอำนาจการทดสอบของการวิเคราะห์ความแปรปรวนทางเดียวภายหลังการแปลงข้อมูลจากกลุ่มตัวอย่างขนาดเล็ก โดยใช้เทคนิคการจำลองแบบมอนติคาร์โลและการใช้ข้อมูลจริง*. วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาการวิจัยและสถิติการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

สุนทร เทียนงาม. (2551). *ผลของความไม่เป็นอิสระของข้อสอบที่มีต่อการประมาณค่าพารามิเตอร์ของข้อสอบและค่าความสามารถของผู้สอบเมื่อมีเงื่อนไขการทดสอบที่แตกต่างกัน.* วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

สุพรรณี อึ้งปัญสัตวงศ์. (2541). *แนวคิดเกี่ยวกับการเปรียบเทียบสถิติทดสอบโดยใช้การจำลองแบบปัญหาด้วยเทคนิคมอนติคาร์โล*. วารสารวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น. 26 (1) : 7 – 11.

สุพล นิลกลาง. (2541). *การเปรียบเทียบคุณลักษณะของแบบทดสอบเลือกตอบระหว่างกลุ่มตัวอย่างที่มีจำนวนต่างกันที่วิเคราะห์ด้วยราส์ชโมเดล.* วิทยานิพนธ์ปริญญามหาบัณฑิต, สาขาวิชาการวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

ภาษาอังกฤษ

Adams, R. (2006). *Reliability as a measurement design effect.* Paper presented at the 2006 bi-annual meeting of the Institute for Objective Measurement Workshop, Berkeley, C. A.

Betebenner, D. M. (1998) . *Improved Confidence Interval Estimation for Variance Components and Error Variances in Generalizability Theory*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego. USA.

Bechger, T. M., Maris, G. , Verstralen, H. H. F. M. & Beguin, A. A. (2003). Using Classical Test Theory in Combination With Item Response Theory. *Applied Psychological Measurement*, 27 (5), p. 319-334.

Bock, Brennan & Muraki. (2002). *The Information in Multiple Rating.* Applied Psychological Measurement . 26 (4), p. 364-375.

Brennan, R. (1992). The Context of Context Effects. *Applied Measurement in Education* 5 (3) , p. 225-264.

Brennan, R. (2001). An Essay on the History and Future of Reliability from the Perspective of Replications. *Journal of Educational Measurement,38 (4)* , p. 295-317.

Brennan, R. (2004). *Some Perspectives on Inconsistencies Among Measurement Model.* CASMA Rearch Report. University of Iowa. Available at www. Uiowa. edu/casma.

Brennan. L. R. (2007). Integration of Model. *Handbook of Statistics.* Elsevier B. V.

Briggs, C. D. & Wilson, M. (2004). *Running Head : Generalizability in Item Response Modeling*. Paper presented at the 2004 Annual Meeting of the Psychometric Society, Pacific Grove, California.

Briggs, C. D. & Wilson, M. (2007). Generalizability in Item Response Modeling. *Journal of Educational Measurement,*44 (2) ,p. 131-155.

Engelhard, G. (1997). History of Modern Psychometrics. *Educational Measurement*, 16,p. 5-52

Ferara, S., Huynh H. & Baghi H. (1997). Contexual Characteristics of locally Dependent Open – Ended Item Cluster in a Large - Scale Performance. Assessment. *Applied Measurement in Education.* 10 (2) , p. 123-144.

Gelman, A., Carlin, J. B. , Stern, H. S. & Rubin, D. B. (1995). *Bayesian Data Analysis.* London: Champman and Hall.

Glass, G. V. & Hopkin, K. D. (1995). *Statisitcal methods in education and psychology*. Neddham Heights,MA: AllynandBacon.

Glas, C. A. W. , & Van der Linnden, W. J. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement,* 27 (4) ,p. 247-261.

Goldstein, H. (1995). *Multilevel statistical models. (2nd ed.).* London: Edward Arnold.

Hambleton, R. K. & Cook, L. L. (1977). Latent Triat Model and Their Use in The Analysis of Educatonal Test Data. *Journal of Educational Measuremen.* 14 (2) ,p. 75-96.

Hambleton, R. K. & Swaminathan. H. (1985). *Item Response Theory: Principles and Application.* Boston: Kluner-Nijhoff Publishing.

Harwell, M., et. al. (1996). Monte Carlo Studies in Item Response Theory. *Applied Psychological Measurement.* Vol. 20, No. 2, June, p. 101-125.

Holland, P. W. & Hoskens. M. (2003). Classical test theory as a first-order item response theory: Application to true-score prediction from a possibly nonparallel test. *Psychometrika*, 68,p. 123-149.

Jiao, H. & Kamata, A. (2005). Modeling Local Item Dependence with the Hierarchical Generalized Linear Model. *Journal of Applied Measurement* ,6 (3) , p. 311-321.

Kolen, M. & Harris, D. (1987). *A multivariate test theory model based on item response theory and generalizability.* Paper presented at the American Educational Research Association, Washington DC. , USA.

Lee. G. (2000). A Comparison of Methods of Estimating Conditional Standard Errors of Measurement for Testlet-Based Test Scores Using Simulation Techniques. *Journal of Educational Measurement,*36 (2) ,p. 91-112.

Lee, Y. W. (2004). *Examining passage-related local item dependence (LID) and measurement construct using Q_3 statistics in an EFL reading comprehension test*. Center for Validity Research/Division of Research and Development, Educational Testing service, Rosedale Road, Princeton,p. 74-100.

Lord, F. M. (1980). *Application of item response theory to practical testing problems.* Hillsdale,N. J. : Erlbaum.

Lord, F. M. (1986). Maximun Likelihood and Bayesian Parameter Estimationin Item Response Theory. *Journal of Educational Measurement.* 23 (2) ,p. 157-162.

Mandenhall, W. & Beaver, R. J. (1994). *Introduction to orobability and Statistic.* Belmont, CA: Wadsworth Publishing.

Mariano, L. T. & Junker, B. W. (2007). Covariates of the Rating Process in Hierarchical Models for Multiple Ratings of Test Items. *Journal of Educational and Behavioral statistics. 32 (3),* p. 287-314.

Mislevy, R., Beaton, A., Kaplan, B., & Sheehan, K. (1992). Estimating population characteristics from sparse matrix samples of item responses. *Journal of Educational Measurement , 29,* p. 133-161.

Mislevy, R. J. (1993). Some formuls for use with Bayesian ability estiamates. *Educational and Psychology Measurement*,52,p. 315-328.

Nugent, R. W. & Hankins, A. J. (1992). A Comparison of Classical, Item Response and Generalizability Theories of Measurement. *Journal of Social Service Research.* 16 (1) . p. 11-39.

Patz, R. J. , Junker, B. W. , Johnson, M. S. & Mariano, L. T. (2002). The Hierarchical Rater Model for Rated Test Items and its Application to Large-Scale Educational Assessment Data. *Journal of Educational and Behavioral Statistics*. 27 (4) ,p. 341-384.

Pelton, W. T. (2002). *The accuracy of unidimensional measurement models in the presence of deviation from the underlying assumotion.* Doctoral dissertation, Department of Instructional Psychology and Technology Faculty of Philosophy, Brigham Young University[Online]. Available from: http: //web. uvic. ca~tpelton/timdissertation. pdf [2009,July 26].

Searle, S. R., Casella, G. & McCulloch, C. E. (1992). *Variance Components*. New York: John Wiley.

Spiegelhalter, D., Thomas, A., Best, N., & Lunn, D. (2004). WinBUGS 1. 4 [Online]. Available from: http: //www. mrc-bsu. cam. ac. uk/bugs [2009,July 25].

Swaminathan, H. & Gifford, J. A. (1982). Bayesina Estimation in the Rasch Model. *Journal of Edicational Statistics.* 7,p. 175-192.

Swaminathan, H. & Gifford, J. A. (1985). Bayesina Estimation in the Two-Parameter Logistic Model. *Psychometrika.* 50 (September) ,p. 349-364.

Swaminathan, H. & Gifford, J. A. (1986). Bayesina Estimation in the Three-Parameter Logistic Model. *Psychometrika.* 51 (December) p. 599-601.

Verhelst, N. & Verstralen, H. (2001). An IRT model for multiple raters. In A. Boomsma, M. van Duijn, and T. Snijders (Eds.) *Essays on item response theory* (pp. 88-108). New York: Springer-Verlag.

Wendy, M. Y. (1984). Effect of local item dependence on the fit and equation performance of the three-parameter logistic model. *Applied Psychological measurement*, 8 (2) , p. 125-145.

Wilson, M. & Hoskens, M. (2001). The rater bundle models. *Journal of Educational and Behavioral Statistics*, 26 (3) ,p. 283-306.

Yamane, T. (1967). *Elementary Sampling Theory.* New York: Prentice-Hall.

Zenisky, A. L. , Hambleton, R. K. & Sireci, S. G. (2002). Identification and evaluation of local item dependences in the Medical College Admissions Test. *Journal of Educational Measurement*, 39, p. 1-16.

ภาคผนวก

ตัวอย่างคำสั่งที่ใช้ประมาณค่าพารามิเตอร์ในรูปแบบที่ 1 Original GIRM รูปแบบที่2 AGIRM A และรูปแบบที่ 3 AGIRM B ด้วยโปรแกรม R และใช้ Package R2WinBUGS สำหรับกรณี Normal-Normal-300-10

```
######################################## ALL COMMAND ############################################
# Simulate data set(s) according to Rasch Model and then run GIRM from WinBUGS
  # Variables:
  # pp = indicates the number of GIRM parameters of interest.
  # ss = number of simulated data sets to generate
  # sample = number of subjects in simulated data sets
  # items = number of dichotomous items in simulated data sets
  ss<-500
  sample<-300
  N= sample
  items<-10
  T= items

##################################################################################################
bsim1.mean =matrix(nrow = T, ncol = ss)
bsim1.sd =matrix(nrow = T, ncol = ss)
tsim1.mean =matrix(nrow = N, ncol = ss)
tsim1.sd =matrix(nrow = N, ncol = ss)
################################################ Original GIRM  #################################
sig2.p.0 = c();
sig2.i.0 = c();
sig2.pi.0 = c();
sig2.e.0 = c();
sig2.pie.0 = c();   ## sig2.pi.0 + sig2.e.0
rho2.0 = c();phi.0 = c()
############################################ AGIRM 1 ############################################
sig2.p.1 = c();
sig2.i.1 = c();
sig2.pie.1 = c();
rho2.1 = c();phi.1 = c()
############################################ AGIRM 2 ############################################
sig2.p.2 = c() ;
sig2.i.2 = c() ;
```

```
sig2.pi.2 = c();
sig2.e.2 = c() ;
sig2.pie.2 = c() ; ## sig2.pi.0 + sig2.e.0
rho2.2 = c();phi.2 = c()
########################################################################################################
###############################  Estimate Variance components using GT Approach #######################
#####################################Call Data################################################### #######
library(R2WinBUGS)
 for (i in 1:ss){
filename = paste("data",i,".txt")
 r = scan(filename)
r = matrix(r,nrow = N, ncol = T)
#}
########################################################################################################
ragrim1 <- function()
{
        for (j in 1:N)  {
                theta[j]~dnorm(0,tau)
                for (k in 1:T) {
                        logit(p[j,k]) <- theta[j] - delta[k]
            r[j,k]~dbern(p[j,k])
                }
        }
        sigma2 <- 1/tau
         Priors
        for (k in 1:T) {
                delta[k]~dnorm(0,0.0001)
                }
                tau~dgamma(.001,.001)
}
########################################################################################################
filename1 <- file.path(tempdir(), "ragrim1.bug")
write.model(ragrim1,filename1)
#file.show(filename1)
data<-list("N", "T","r")
    inits <- function() {
        list(delta = rep(0,T), tau=1)
        }
    parameters <- list("delta", "theta","tau","p")
```

```
##########################################################################################
  girm1.sim <- bugs(data, inits, parameters, model.file= "ragrim1.bug", n.chains=1  n.burnin=1000 n.iter=10000,n.thin
= 1,debug=F,
bugs.directory = "C:/Program Files/WinBUGS14",codaPkg = FALSE)
##########################################################################################
p  <-     girm1.sim$mean$p
######################################  ORIGINAL  ######################################
# GT TERMS
pi.p.0 = c();sum2.p.0 = c();
pi.i.0 = c();sum2.i.0 = c();
pi.pi.0 = matrix(nrow = N, ncol = T);sum2.pi.0 =matrix(nrow = N, ncol = T);
ss.e.0 =  matrix(nrow = N, ncol = T);
        # Grand mean
        mu.g.0 <- mean(p[,])
        # Person specific mean item response  (vector length 500)
        for (j in 1:N) {
                pi.p.0[j] <- (sum(p[j,])/T)- mu.g.0
        }
        # Item specific mean person response (vector length of 5)
        for (k in 1:T) {
                pi.i.0[k] <- (sum(p[,k])/N) - mu.g.0
        }
        # Person by item interaction effect (500 by 5 matrix)
        for (j in 1:N) {
                for (k in 1:T) {
                        pi.pi.0[j,k] <-p[j,k] - pi.p.0[j] - pi.i.0[k] - mu.g.0
                }
        }
        # Variance Components
        for (j in 1:N) {
        for (k in 1:T) {
                        ss.e.0[j,k] <- p[j,k]*(1-p[j,k])
                }
        }
        sig2.e.0[i] <- mean(ss.e.0[,])
        for (j in 1:N) {
                sum2.p.0[j] <- pi.p.0[j]^2
        }
        sig2.p.0[i] <- mean(sum2.p.0[])
```

```
            for (k in 1:T) {
                        sum2.i.0[k] <- pi.i.0[k]^2
            }
            sig2.i.0[i] <- mean(sum2.i.0[])
            for (j in 1:N) {
            for (k in 1:T) {
                                    sum2.pi.0[j,k] <- pi.pi.0[j,k]^2
                        }
            }
            sig2.pi.0[i] <- mean(sum2.pi.0[,])
             sig2.pie.0[i] = sig2.pi.0[i]+sig2.e.0[i]
            # GT Coefficients original GIRM
            rho2.0[i] <- sig2.p.0[i]/(sig2.p.0[i] + sig2.pi.0[i]/T + sig2.e.0[i]/T)
            phi.0[i] <-  sig2.p.0[i]/(sig2.p.0[i] + sig2.pi.0[i]/T +  sig2.e.0[i]/T + sig2.i.0[i]/T)
############################################  AGIRM1  ############################################
pi.p.1 = c();ss.p.1 = c();MS.p.1 =c();
pi.i.1 = c();  ss.i.1 = c(); MS.i.1 = c();
pi.pi.1 = matrix(nrow = N, ncol = T);
ss.e.1 =  matrix(nrow = N, ncol = T);
ss.pie.1 =matrix(nrow = N, ncol = T);MS.pie.1 =c()
################################################################################################
nexmn=sample
 try<-matrix(runif(nexmn*N,0,1),nrow=N,ncol=T)
rtemp<-matrix(0,N,T)
for(x in 1:N){
            for(y in 1:T){
                        if  (try[x,y] < p[x,y]) {
                                    rtemp[x,y]=1 }
                        else {rtemp[x,y]=0}
                  } # loop l2
              } # loop l1
################################################################################################
# Grand mean
            mu.g.1 <- mean(rtemp[,])
            # Person specific mean item response  (vector length 500)
            for (j in 1:N) {
                        pi.p.1[j] <- (sum(rtemp[j,])/T)- mu.g.1
            }
            # Item specific mean person response (vector length of 5)
```

```
        for (k in 1:T) {
                pi.i.1[k] <- (sum(rtemp[,k])/N) - mu.g.1
        }
        # Person by item interaction effect (500 by 5 matrix)
        for (j in 1:N) {
        for (k in 1:T) {
                        pi.pi.1[j,k] <- p[j,k] - pi.p.1[j] - pi.i.1[k] - mu.g.1
                }
        }
##variance components (random effects)
        #1) overall errors variance
        for (j in 1:N){
                for (k in 1:T){
                        ss.e.1[j,k] <- (rtemp[j,k]-mu.g.1)^2
                 }
        }
        #2) person variance
        for (j in 1:N){
                ss.p.1[j] <-  (pi.p.1[j])^2
        }
                MS.p.1[i] <- (T*sum(ss.p.1[]))/(N-1)
        #3) item variance
        for (k in 1:T){
                ss.i.1[k] <-  pi.i.1[k]^2
        }
                MS.i.1[i]<- (N*sum(ss.i.1[]))/(T-
        #4) interaction between person and item and residual
        for (j in 1:N){
                for (k in 1:T){
                        ss.pie.1[j,k] <-  (rtemp[j,k]-(sum(rtemp[j,])/T)-(sum(rtemp[,k])/N)+mu.g.1)^2
                          }
           }
                        MS.pie.1[i] <- sum(ss.pie.1[,])/((N-1)*(T-1))
                    sig2.pie.1[i] <-MS.pie.1[i]
                   sig2.p.1[i] <-(MS.p.1[i] - MS.pie.1[i])/T
                    sig2.i.1[i] <-(MS.i.1[i] - MS.pie.1[i])/N
        # GT Coefficients AGIRM1
        rho2.1[i] <- sig2.p.1[i]/(sig2.p.1[i] + sig2.pie.1[i]/T )
        phi.1[i] <-  sig2.p.1[i]/(sig2.p.1[i] + sig2.pie.1[i]/T  + sig2.i.1[i]/T)
```

```
############################################ AGIRM2 ##########################################
sig2.e.2[i] = sig2.pie.1[i]-sig2.pi.0[i]
sig2.p.2[i] = sig2.p.1[i]
sig2.i.2[i] = sig2.i.1[i]
sig2.pi.2[i] = sig2.pi.0[i]
sig2.pie.2[i] = sig2.pi.2[i]+sig2.e.2[i]
 # GT Coefficients AGIRM2
        rho2.2[i] <- sig2.p.2[i]/(sig2.p.2[i] + sig2.pi.2[i]/T + sig2.e.2[i]/T)
        phi.2[i] <-  sig2.p.2[i]/(sig2.p.2[i] + sig2.i.2[i]/T+ sig2.pi.2[i]/T + sig2.e.2[i]/T)
###############################################################################################
   bsim1.mean.0 <-girm1.sim$mean$delta
for (k in 1:T){
     bsim1.mean[k,i] = bsim1.mean.0[k]
}
###############################################################################################
      tsim1.mean.0 <-girm1.sim$mean$theta
for (j in 1:N){
     tsim1.mean[j,i] = tsim1.mean.0[j]
}
      #AGIRM1
##############################################################################################
 cat(c("loop :",i),fill=T)
}
###############################################################################################
bsim.overall.mean = c();tsim.overall.mean = c()
bsim.overall.sd = c();tsim.overall.sd = c()

########################################## Original ############################################
mean.sig2.e.0 = mean(sig2.e.0[])
sd.sig2.e.0 = sd(sig2.e.0[])
mean.sig2.p.0 = mean(sig2.p.0[])
sd.sig2.p.0 = sd(sig2.p.0[])
mean.sig2.i.0 = mean(sig2.i.0[])
sd.sig2.i.0 = sd(sig2.i.0[])
mean.sig2.pi.0 = mean(sig2.pi.0[])
sd.sig2.pi.0 = sd(sig2.pi.0[])
mean.sig2.pie.0 = mean(sig2.pie.0[])
sd.sig2.pie.0 = sd(sig2.pie.0[])
mean.rho2.0 = mean(rho2.0[])
```

```
sd.rho2.0 = sd(rho2.0[])
mean.phi.0 = mean(phi.0[])
sd.phi.0 = sd(phi.0[])
################################################## AGRIM 1 ######################################
mean.sig2.pie.1 = mean(sig2.pie.1[])
sd.sig2.pie.1 = sd(sig2.pie.1[])
mean.sig2.p.1 = mean(sig2.p.1[])
sd.sig2.p.1 = sd(sig2.p.1[])
mean.sig2.i.1 = mean(sig2.i.1[])
sd.sig2.i.1 = sd(sig2.i.1[])
mean.rho2.1 = mean(rho2.1[])
sd.rho2.1 = sd(rho2.1[])
mean.phi.1 = mean(phi.1[])
sd.phi.1 = sd(phi.1[])
################################################# AGIRM 2 ########################################
mean.sig2.p.2 = mean.sig2.p.1
sd.sig2.p.2 = sd.sig2.p.1
mean.sig2.i.2 = mean.sig2.i.1
sd.sig2.i.2 = sd.sig2.i.1
mean.sig2.pi.2 = mean.sig2.pi.0
sd.sig2.pi.2 = sd.sig2.pi.0
mean.sig2.pie.2 = mean(sig2.pie.2[])
sd.sig2.pie.2 = sd(sig2.pie.2[])
mean.sig2.e.2 = mean(sig2.e.2[])
sd.sig2.e.2 = sd(sig2.e.2[])
mean.rho2.2 = mean(rho2.2[])
sd.rho2.2 = sd(rho2.2[])
mean.phi.2 = mean(phi.2[])
sd.phi.2 = sd(phi.2[])
##################################################################################################
for (k in 1:T){
bsim.overall.mean[k] = mean(bsim1.mean[k,])
bsim.overall.sd[k] = sd(bsim1.mean[k,])
}
for (j in 1:N){j
tsim.overall.mean[j] = mean(tsim1.mean[j,])
tsim.overall.sd[j] = sd(tsim1.mean[j,])
}
##################################  Delta and Theta ###############################################
```

```
output_bsim= data.frame(bsim1.mean)
write.table(output_bsim,file = "bsim.txt", sep = "\t",row.name = FALSE)
output_tsim= data.frame(tsim1.mean)
write.table(output_tsim,file = "tsim.txt", sep = "\t",row.name = FALSE)
##############################################################################################
output_b_sim= data.frame(bsim.overall.mean, bsim.overall.sd)
write.table(output_b_sim,file = "b_sim.txt", sep = "\t",row.name = FALSE)
##############################################################################################
output_t_sim= data.frame(tsim.overall.mean,tsim.overall.sd)
write.table(output_t_sim,file = "t_sim.txt", sep = "\t",row.name = FALSE)
##############################################################################################
output_grim= data.frame(sig2.e.0,sig2.p.0,
sig2.i.0,sig2.pi.0 ,rho2.0 ,phi.0 ,
sig2.pie.1,sig2.p.1,sig2.i.1,rho2.1,
phi.1,sig2.e.2,rho2.2,phi.2)
write.table(output_grim,file = "grim.txt", sep = "\t",row.name = FALSE)
##############################################################################################
output_mean_sd= data.frame(mean.sig2.e.0,sd.sig2.e.0,
mean.sig2.p.0 ,sd.sig2.p.0 ,
mean.sig2.i.0 , sd.sig2.i.0 ,
mean.sig2.pi.0 , sd.sig2.pi.0 ,
mean.sig2.pie.0 , sd.sig2.pie.0 ,
mean.rho2.0 , sd.rho2.0 ,
mean.phi.0 , sd.phi.0 ,
mean.sig2.pie.1 , sd.sig2.pie.1 ,
mean.sig2.p.1 ,sd.sig2.p.1 ,
mean.sig2.i.1 ,sd.sig2.i.1 ,
mean.rho2.1 , sd.rho2.1 ,
mean.phi.1 ,sd.phi.1 ,
mean.sig2.e.2 ,sd.sig2.e.2,
mean.sig2.p.2 ,sd.sig2.p.2 ,
mean.sig2.i.2 , sd.sig2.i.2 ,
mean.sig2.pi.2 , sd.sig2.pi.2 ,
mean.sig2.pie.2 , sd.sig2.pie.2 ,
mean.rho2.2 , sd.rho2.2 ,
mean.phi.2 ,sd.phi.2 )
write.table(output_mean_sd,file = "mean_sd.txt", sep = "\t",row.name = FALSE)
##############################################################################################
```

```
#par(ask = T)
cat(c("----------------------------- ORIGINAL  METHOD ----------------"),fill = T)
cat(c(" mean.sig2.e.0   = ", round(mean.sig2.e.0,6),  " sd.sig2.e.0   = ",round(sd.sig2.e.0,6),
 " mean.sig2.p.0   = ", round(mean.sig2.p.0,6), " sd.sig2.p.0   = ",round(sd.sig2.p.0,6),
 " mean.sig2.i.0 = ", round(mean.sig2.i.0,6), " sd.sig2.i.0 = " ,round(sd.sig2.i.0,6),
 " mean.sig2.pi.0 = ", round(mean.sig2.pi.0,6), "sd.sig2.pi.0 = ", round(sd.sig2.pi.0,6),
" mean.sig2.pie.0 = ", round(mean.sig2.pie.0,6), "sd.sig2.pie.0 = ", round(sd.sig2.pie.0,6),
" mean.rho2.0 = ", round(mean.rho2.0,6), "sd.rho2.0 = " ,round(sd.rho2.0,6),
  "mean.phi.0 = ", round(mean.phi.0,6), "sd.phi.0 = " , round(sd.phi.0,6)),fill = T)
cat(c("----------------------------- AGRIM 1  METHOD ----------------"),fill = T)
cat(c(" mean.sig2.pie.1   = ", round(mean.sig2.pie.1,6),  " sd.sig2.p.0   = ",round(sd.sig2.p.1,6),
 " mean.sig2.p.1   = ", round(mean.sig2.p.1,6), " sd.sig2.p.1   = ",round(sd.sig2.p.1,6),
 " mean.sig2.i.1 = ", round(mean.sig2.i.1,6), " sd.sig2.i.1 = " ,round(sd.sig2.i.1,6),
" mean.rho2.1 = ", round(mean.rho2.1,6), "sd.rho2.1 = " ,round(sd.rho2.1,6),
  "mean.phi.1 = ", round(mean.phi.1,6), "sd.phi.1 = " , round(sd.phi.1,6)),fill = T)
cat(c("----------------------------- AGIRM 2 METHOD  ----------------"),fill = T)
cat(c(" mean.sig2.e.2   = ", round(mean.sig2.e.2,6),  " sd.sig2.e.2   = ",round(sd.sig2.e.2,6),
 " mean.sig2.p.2   = ", round(mean.sig2.p.2,6), " sd.sig2.p.2   = ",round(sd.sig2.p.2,6),
 " mean.sig2.i.2 = ", round(mean.sig2.i.2,6), " sd.sig2.i.2 = " ,round(sd.sig2.i.0,6),
 " mean.sig2.pi.2 = ", round(mean.sig2.pi.2,6), "sd.sig2.pi.2 = ", round(sd.sig2.pi.2,6),
" mean.sig2.pie.2 = ", round(mean.sig2.pi.2,6), "sd.sig2.pi.2 = ", round(sd.sig2.pi.2,6),
" mean.rho2.2 = ", round(mean.rho2.2,6), "sd.rho2.0 = " ,round(sd.rho2.2,6),
  "mean.phi.2 = ", round(mean.phi.2,6), "sd.phi.0 = " , round(sd.phi.2,6)),fill = T)
################################################################################################
##################################END of Method1-3#############################################
################################################################################################
n=N
i=T
################################################################################################
#GENERATE TRUE PARAMETERS
################################################################################################
theta_true = matrix(nrow = N, ncol = ss)
beta_true =  matrix(nrow = T, ncol = ss)
sigma2.e_t = c(); sigma2.p_t = c() ;
sigma2.i_t = c(); sigma2.pi_t = c();
sigma2.pie_t = c() ; rho2_t = c(); phi_t = c()
for (j in 1:ss){
theta_t<-rnorm(n,0,1)
beta_t<-numeric(i)
```

```
for (k in 1:i)
{
beta_t[k]<-rnorm(1,0.5,0.5)
}
prob_t<-matrix(1,n,i)
x<-matrix(0,n,i)
score<-numeric(n)
for (p in 1:n)
{
for (k in 1:i)
  {
    prob_t[p,k]<-exp(theta_t[p]-beta_t[k])/(1+exp(theta_t[p]-beta_t[k]))
  }
 }
mu.p<-numeric(n)
mu.i<-numeric(i)
 theta<-numeric(n)
 beta<-numeric(i)
v.theta.beta<-matrix(1,n,i)
ss.e<-matrix(1,n,i)
ss.p<-numeric(n)
ss.i<-numeric(i)
ss.pi<-matrix(1,n,i)
#1) true grand mean
mu<-mean(prob_t[,])
#2) person effect
for (p in 1:n)
{
mu.p[p]<-mean(prob_t[p,])
 theta[p]<-(sum(prob_t[p,])/i)-mu
}
#3) item effect
for (k in 1:i)
{
mu.i[k]<-mean(prob_t[,k])
 beta[k]<-(sum(prob_t[,k])/n)-mu
}
##################################################################################################
```

```
 for(k in 1:i){beta_true[k,j] =  beta[k]}
  for(p in 1:n){theta_true[p,j] = theta[p]}
##############################################################################################
#4) iteraction effect (between person and item)
for (p in 1:n)
{
 for (k in 1:i)
  {
    v.theta.beta[p,k]<-prob_t[p,k]-theta[p]-beta[k]-mu
  }
}
##variance components (random effects)
        #1) errors variance
                for (p in 1:n)
                  {
                        for (k in 1:i)
                        {
                          ss.e[p,k]<-prob_t[p,k]*(1-prob_t[p,k])
                        }
                  }

                sigma2.e_t[j] <-mean(ss.e[,])
        #2) person variance
                for (p in 1:n)
                  {
                        ss.p[p]<-(theta[p])^2
                  }
                sigma2.p_t[j] <-mean(ss.p[])
        #3) item variance
                for (k in 1:i)
                  {
                ss.i[k]<-(beta[k])^2
                  }
                sigma2.i_t[j] <-mean(ss.i[])
        #4) interaction variance
                for (p in 1:n)
                  {
                for (k in 1:i)
                  {
```

```
                    ss.pi[p,k]<-(v.theta.beta[p,k])^2
                      }
            }
                    sigma2.pi_t[j] <-mean(ss.pi[,])
                    sigma2.pie_t[j] <- sigma2.pi_t[j] +sigma2.e_t[j]
                    rho2_t[j] <- sigma2.p_t[j]/(sigma2.p_t[j] + sigma2.pi_t[j]/i + sigma2.e_t[j]/i)
                    phi_t[j] <-  sigma2.p_t[j]/(sigma2.p_t[j] + sigma2.pi_t[j]/i +  sigma2.e_t[j]/i + sigma2.i_t[j]/i)
###################### Loop j #####################
}
######### MAD : Mean Absolute Deviation ###########
d_beta =   matrix(nrow = T, ncol = ss)
d_theta = matrix(nrow = N, ncol = ss)
MAD_beta = c();MAD_theta = c()
#1) MAD for beta
        for (k in 1:T){
                for (d in 1:ss){
                        d_beta[k,d]<-abs((bsim1.mean[k,d]-beta_true[k,d])/beta_true[k,d])
                                }
                  }
        for (k in 1:T){
                MAD_beta[k] <- mean(d_beta[k,])
                  }
#2) MAD for theta
        for (k in 1:N){
                for (d in 1:ss){
                        d_theta[k,d]<-abs((tsim1.mean[k,d]-theta_true[k,d])/theta_true[k,d])
                            }
                  }
        for (k in 1:N){
                        MAD_theta[k] <- mean(d_theta[k,])
                  }

##############################################################################################
mean_MAD_beta = mean(MAD_beta[])
mean_MAD_theta = mean(MAD_theta[])
##############################################################################################
#############   Original Method  ##################
##############################################################################################
MAD_sig2.e.0 = mean(abs((sigma2.e_t[]-sig2.e.0[])/sigma2.e_t[]))
```

```
MAD_sig2.p.0 = mean(abs((sigma2.p_t[]-sig2.p.0[])/sigma2.p_t[]))
MAD_sig2.i.0 = mean(abs((sigma2.i_t[]-sig2.i.0[])/sigma2.i_t[]))
MAD_sig2.pi.0 = mean(abs((sigma2.pi_t[]-sig2.pi.0[])/sigma2.pi_t[]))
MAD_sig2.pie.0 = mean(abs((sigma2.pie_t[]-sig2.pie.0[])/sigma2.pie_t[]))
MAD_rho2.0 = mean(abs((rho2_t[]-rho2.0[])/rho2_t[]))
MAD_phi.0 = mean(abs((phi_t[]-phi.0[])/phi_t[]))
EUC.0 = mean(sqrt((sigma2.e_t[]-sig2.e.0[])^2+(sigma2.p_t[]-sig2.p.0[])^2+
(sigma2.i_t[]-sig2.i.0[])^2+(sigma2.pi_t[]-sig2.pi.0[])^2+
(sigma2.pie_t[]-sig2.pie.0[])^2))
##########################################################################################
AD_sig2.e.0 = (abs((sigma2.e_t[]-sig2.e.0[])/sigma2.e_t[]))
AD_sig2.p.0 =  (abs((sigma2.p_t[]-sig2.p.0[])/sigma2.p_t[]))
AD_sig2.i.0 =  (abs((sigma2.i_t[]-sig2.i.0[])/sigma2.i_t[]))
AD_sig2.pi.0 =  (abs((sigma2.pi_t[]-sig2.pi.0[])/sigma2.pi_t[]))
AD_sig2.pie.0 =  (abs((sigma2.pie_t[]-sig2.pie.0[])/sigma2.pie_t[]))
AD_rho2.0 =  (abs((rho2_t[]-rho2.0[])/rho2_t[]))
AD_phi.0 =  (abs((phi_t[]-phi.0[])/phi_t[]))
##################################   AGRIM1   ################################################
MAD_sig2.p.1 = mean(abs((sigma2.p_t[]-sig2.p.1[])/sigma2.p_t[]))
MAD_sig2.i.1 = mean(abs((sigma2.i_t[]-sig2.i.1[])/sigma2.i_t[]))
MAD_sig2.pie.1 = mean(abs((sigma2.pie_t[]-sig2.pie.1[])/sigma2.pie_t[]))
MAD_rho2.1 = mean(abs((rho2_t[]-rho2.1[])/rho2_t[]))
MAD_phi.1 = mean(abs((phi_t[]-phi.1[])/phi_t[]))
EUC.1 = mean(sqrt((sigma2.p_t[]-sig2.p.0[])^2+
(sigma2.i_t[]-sig2.i.0[])^2+ (sigma2.pie_t[]-sig2.pie.0[])^2))
##########################################################################################
AD_sig2.p.1 =  (abs((sigma2.p_t[]-sig2.p.1[])/sigma2.p_t[]))
AD_sig2.i.1 =  (abs((sigma2.i_t[]-sig2.i.1[])/sigma2.i_t[]))
AD_sig2.pie.1 =  (abs((sigma2.pie_t[]-sig2.pie.1[])/sigma2.pie_t[]))
AD_rho2.1 =  (abs((rho2_t[]-rho2.1[])/rho2_t[]))
AD_phi.1 =  (abs((phi_t[]-phi.1[])/phi_t[]))
###################################   AGRIM  2 #############################################
MAD_sig2.e.2 = mean(abs((sigma2.e_t[]-sig2.e.2[])/sigma2.e_t[]))
MAD_sig2.p.2 = mean(abs((sigma2.p_t[]-sig2.p.2[])/sigma2.p_t[]))
MAD_sig2.i.2 = mean(abs((sigma2.i_t[]-sig2.i.2[])/sigma2.i_t[]))
MAD_sig2.pi.2 = mean(abs((sigma2.pi_t[]-sig2.pi.2[])/sigma2.pi_t[]))
MAD_sig2.pie.2 = mean(abs((sigma2.pie_t[]-sig2.pie.2[])/sigma2.pie_t[]))
MAD_rho2.2 = mean(abs((rho2_t[]-rho2.2[])/rho2_t[]))
MAD_phi.2 = mean(abs((phi_t[]-phi.2[])/phi_t[]))
```

```
EUC.2 = mean(sqrt((sigma2.e_t[]-sig2.e.2[])^2+(sigma2.p_t[]-sig2.p.2[])^2+
(sigma2.i_t[]-sig2.i.2[])^2+(sigma2.pi_t[]-sig2.pi.2[])^2+
(sigma2.pie_t[]-sig2.pie.2[])^2))
##########################################################################################
AD_sig2.e.2 =  (abs((sigma2.e_t[]-sig2.e.2[])/sigma2.e_t[]))
AD_sig2.p.2 =  (abs((sigma2.p_t[]-sig2.p.2[])/sigma2.p_t[]))
AD_sig2.i.2 =  (abs((sigma2.i_t[]-sig2.i.2[])/sigma2.i_t[]))
AD_sig2.pi.2 =  (abs((sigma2.pi_t[]-sig2.pi.2[])/sigma2.pi_t[]))
AD_sig2.pie.2 =  (abs((sigma2.pie_t[]-sig2.pie.2[])/sigma2.pie_t[]))
AD_rho2.2 =  (abs((rho2_t[]-rho2.2[])/rho2_t[]))
AD_phi.2 =  (abs((phi_t[]-phi.2[])/phi_t[]))
########################################Delta and Theta ####################################
output_MAD_beta= data.frame(MAD_beta)
write.table(output_MAD_beta,file = "MAD_beta.txt", sep = "\t",row.name = FALSE)
output_MAD_theta= data.frame(MAD_theta)
write.table(output_MAD_theta,file = "MAD_theta.txt", sep = "\t",row.name = FALSE)
##########################################################################################
output_MAD_0= data.frame(MAD_sig2.e.0, MAD_sig2.p.0,MAD_sig2.i.0,
MAD_sig2.pi.0,MAD_sig2.pie.0,MAD_rho2.0,MAD_phi.0,EUC.0)
write.table(output_MAD_0,file = "MAD_0.txt", sep = "\t",row.name = FALSE)
##########################################################################################
output_MAD_1= data.frame(MAD_sig2.p.1,MAD_sig2.i.1,
 MAD_sig2.pie.1,MAD_rho2.1,MAD_phi.1,EUC.1)
write.table(output_MAD_1,file = "MAD_1.txt", sep = "\t",row.name = FALSE)
#########################################################################################
output_MAD_2= data.frame(MAD_sig2.e.2, MAD_sig2.p.2,MAD_sig2.i.2,
 MAD_sig2.pi.2,MAD_sig2.pie.2,MAD_rho2.2,MAD_phi.2,EUC.2)
write.table(output_MAD_2,file = "MAD_2.txt", sep = "\t",row.name = FALSE)
##########################################################################################
############### Absolute deviation ############################
##########################################################################################
output_AD_0= data.frame(AD_sig2.e.0, AD_sig2.p.0,AD_sig2.i.0,
 AD_sig2.pi.0,AD_sig2.pie.0,AD_rho2.0,AD_phi.0)
write.table(output_AD_0,file = "AD_0.txt", sep = "\t",row.name = FALSE)
#########################################################################################
output_AD_1= data.frame(AD_sig2.p.1,AD_sig2.i.1,
 AD_sig2.pie.1,AD_rho2.1,AD_phi.1)
write.table(output_AD_1,file = "AD_1.txt", sep = "\t",row.name = FALSE)
#########################################################################################
```

```
output_AD_2= data.frame(AD_sig2.e.2, AD_sig2.p.2,AD_sig2.i.2,
 AD_sig2.pi.2,AD_sig2.pie.2,AD_rho2.2,AD_phi.2)
write.table(output_AD_2,file = "AD_2.txt", sep = "\t",row.name = FALSE)
##########################################################################################
#par(ask = T)
cat(c("------------- MAD  ORIGINAL  METHOD ------------"),fill = T)
cat(c(" MAD_sig2.e.0   = ", round(MAD_sig2.e.0,6),
 " MAD_sig2.p.0   = ", round(MAD_sig2.p.0,6),
 " MAD_sig2.i.0 = ", round(MAD_sig2.i.0,6),
 " MAD_sig2.pi.0 = ", round(MAD_sig2.pi.0,6),
" MAD_sig2.pie.0 = ", round(MAD_sig2.pie.0,6),
" MAD_rho2.0 = ", round(MAD_rho2.0,6),
  "MAD_phi.0 = ", round(MAD_phi.0,6),
   "EUC.0 = " , round(EUC.0,6)), fill = T)
cat(c("---------------- AGRIM1  METHOD ----------------"),fill = T)
cat(c( " MAD_sig2.p.1   = ", round(MAD_sig2.p.1,6),
 " MAD_sig2.i.1 = ", round(MAD_sig2.i.1,6),
" MAD_sig2.pie.1 = ", round(MAD_sig2.pie.1,6),
" MAD_rho2.1 = ", round(MAD_rho2.1,6),
  "MAD_phi.1 = ", round(MAD_phi.1,6),
 "EUC.1 = " , round(EUC.1,6)), fill = T)
cat(c("----------------   AGRIM2  METHOD ----------------"),fill = T)
cat(c(" MAD_sig2.e.2   = ", round(MAD_sig2.e.2,6),
 " MAD_sig2.p.2   = ", round(MAD_sig2.p.2,6),
 " MAD_sig2.i.2 = ", round(MAD_sig2.i.2,6),
 " MAD_sig2.pi.2 = ", round(MAD_sig2.pi.2,6),
" MAD_sig2.pie.2 = ", round(MAD_sig2.pie.2,6),
" MAD_rho2.2 = ", round(MAD_rho2.2,6),
  "MAD_phi.2 = ", round(MAD_phi.2,6),
 "EUC.2 = " , round(EUC.2,6)), fill = T)
cat(c("-----------------------------   MAD Mean of theta and beta ----------------"),fill = T)
cat(c(" mean_MAD_beta  = ", round(mean_MAD_beta,6),
 " mean_MAD_theta   = ", round(mean_MAD_theta,6)), fill = T)
######################################## ORIGiNAL  ##########################################
t.test(AD_sig2.e.0, mu=0)
sd(AD_sig2.e.0)
t.test(AD_sig2.p.0, mu=0)
sd(AD_sig2.p.0)
t.test(AD_sig2.i.0, mu=0)
```

```
sd(AD_sig2.i.0)
t.test(AD_sig2.pi.0, mu=0)
sd(AD_sig2.pi.0)
t.test(AD_sig2.pie.0, mu=0)
sd(AD_sig2.pie.0)
t.test(AD_rho2.0, mu=0)
sd(AD_rho2.0)
t.test(AD_phi.0, mu=0)
sd(AD_phi.0)
######################################## AGIRM1 #####################################################
t.test(AD_sig2.p.1, mu=0)
sd(AD_sig2.p.1)
t.test(AD_sig2.i.1, mu=0)
sd(AD_sig2.i.1)
t.test(AD_sig2.pie.1, mu=0)
sd(AD_sig2.pie.1)
t.test(AD_rho2.1, mu=0)
sd(AD_rho2.1)
t.test(AD_phi.1, mu=0)
sd(AD_phi.1)
######################################## AGIRM2 ####################################################
t.test(AD_sig2.e.2, mu=0)
sd(AD_sig2.e.2)
t.test(AD_sig2.p.2, mu=0)
sd(AD_sig2.p.2)
t.test(AD_sig2.i.2, mu=0)
sd(AD_sig2.i.2)
t.test(AD_sig2.pi.2, mu=0)
sd(AD_sig2.pi.2)
t.test(AD_sig2.pie.2, mu=0)
sd(AD_sig2.pie.2)
t.test(AD_rho2.2, mu=0)
sd(AD_rho2.2)
t.test(AD_phi.2, mu=0)
sd(AD_phi.2)
################################################################################################
```

ตัวอย่างคำสั่งที่ใช้ประมาณค่าพารามิเตอร์ในรูปแบบที่ 4 Numerical Bayesian GIRM ด้วยโปรแกรม R สำหรับกรณี Normal-Normal-300-10

```
####################################    1ST  STEP  ##########################################
   ss<-500
   sample<-300
   N= sample
   items<-10
   T= items
#############################################################################################
sig2.e.3 = c(); sig2.p.3 = c();sig2.i.3 = c() ;
sig2.pie.3 = c(); sd.sig2.pie.3 = c()
sig2.pi.3 = c() ; rho2.3 = c(); phi.3 = c()
sd.sig2.e.3 = c(); sd.sig2.p.3 = c(); sd.sig2.i.3 = c() ;
sd.sig2.pi.3 = c() ; sd.rho2.3 = c(); sd.phi.3 = c()
 beta.mean.i = matrix(nrow = T, ncol = ss)
  theta.mean.i = matrix(nrow = N, ncol = ss)
################################### Specify parameters########################################
# Sets the convergence tolerance criterion
tol = 10^-3;
# Sets the number of maximum number of times the updating procedure will run
numit = 10,000;
# Specify priors
theta.prior <- "Normal"
b.prior <- "Normal"
# Sets the value of a in Gamma(a,b)
theta.shape <- 1
theta.rate <- 1
# Sets the value of a in Beta(a,b)
b.shape1 <- 0.5
b.shape2 <- 0.5
# Sets the value of sigma2 beta
lambda.b = 10
v.b = 10
##############################################################################################
for (i in 1:ss){
filename = paste("data",i,".txt")
X = scan(filename)
```

```
X = matrix(X,nrow = N, ncol = T)
data = X
###############################################################################################
# rowsum is total score per person
  row = nrow(data)
# column is item and row is person
# colsum is total score of all person per item
  col = ncol(data)
  R = numeric(col)
# Compute the correlation between question score and total score
  for(j in 1:col){
      R[j] <- cor(X[,j], rowSums(data))
   }
#compute the proportion of the class that got each question right
  temp.tot.row <- pmin(pmax(colSums(data), 0.5), row-0.5)
  Q = temp.tot.row/row
#compute the proportion of the class that got score per person
  temp.tot.col <- pmin(pmax(rowSums(data), 0.5), col-0.5)
# initial value for theta i(0)
  theta.i.0 = log(temp.tot.col/(col-temp.tot.col))
# initial value for beta b(0)
  beta.i.0 = (qnorm(Q))/R
###############################################################################################
theta = theta.i.0
  beta  = beta.i.0
  theta.matrix = matrix(0, nrow = numit+1 , ncol = row)
  beta.matrix = matrix(0, nrow = numit+1 , ncol = col)
  theta.matrix.n = matrix(0, nrow = numit  , ncol = row)
  beta.matrix.n = matrix(0, nrow = numit  , ncol = col)
  beta.matrix[1,] = beta.i.0
  theta.matrix[1,] = theta.i.0
   diff.theta = 2*tol
   diff.beta  = 2*tol
   count = 0
  Xbar = c() ; XbarP = c() ;XbarI = c() ;
 Xbarinteraction = c() ; ss.e = c() ; sig2.e = c() ;
sum2.p = c() ;        sig2.p = c() ; sum2.i = c() ;
sig2.i= c() ;             sum2.pi = c() ; sig2.pi = c() ;
 rho2 = c() ; phi = c() ;sig2.pie = c()
```

```
################################################################################################
p = matrix(nrow=row, ncol=col)
 for (z in 1:numit)  {

  tol = tol;  diff.theta = 2*tol;    diff.beta  = 2*tol
   while (diff.theta >= tol || diff.beta >= tol) {
     count = count + 1
              P = matrix(0, nrow = row, ncol = col)
     for(r in 1:row){
       for(c in 1:col){
         P[r,c] <- exp(theta[r]-beta[c])/( 1 + exp(theta[r]-beta[c]) )
               }
              }
   ##for theta
   # Likelihood founction of theta for 1st derivative
     lik.theta.dot    =  rowSums(data - P, na.rm=TRUE)
   # Likelihood founction of theta for 2st derivative
     lik.theta.dotdot =  -1*rowSums(P*(1-P))
   # prior of theta
     if(theta.prior == "Normal"){
                          f.theta.dot      =  -theta
                          f.theta.dotdot   =  -1
              }
              if(theta.prior == "Gamma"){
                          f.theta.dot      =  (((theta.shape - 1)/(theta)) - (1/theta.rate))
                          f.theta.dotdot  =  -(theta.shape - 1)/(theta^2)
     }
   ##prior of beta
   # Likelihood founction of beta for 1st derivative
     lik.beta.dot    =  -1*colSums(data - P, na.rm=TRUE)
   # Likelihood founction of beta for 2st derivative
     lik.beta.dotdot =  -1*colSums(P*(1-P))
   # prior of beta
     bdot           = (sum(beta))/col
     sigma.square     = (lambda.b + sum((beta - bdot)^2) )/( v.b + col - 1)
     if(b.prior == "Normal"){
                          f.beta.dot       =  -(beta-bdot)^2/sigma.square
                          f.beta.dotdot    =  -( sigma.square*(1-(1/col)) - (2*(beta-bdot)^2)/(v.b+col-1) )/(sigma.square^2)
     }
```

```
    if(b.prior == "Beta"){
                      f.beta.dot      = (    ( (b.shape1-1)*(1/beta) )  -  ( (b.shape2-1)*(1/(1-beta)) )   )
          f.beta.dotdot   =  (    ( -(b.shape1-1)*(beta^2) )  -  ( (b.shape2-1)*(1/(1-beta)^2) )    )


    }
    # update
    g.theta = lik.theta.dot    + f.theta.dot
    h.theta = lik.theta.dotdot + f.theta.dotdot
    g.beta  = lik.beta.dot     + f.beta.dot
    h.beta  = lik.beta.dotdot  + f.beta.dotdot
    temp.theta = theta - g.theta/h.theta
    temp.beta  = beta  - g.beta/h.beta
    diff.theta = sum(abs(theta-temp.theta))
    diff.beta  = sum(abs(beta -temp.beta))
    theta = temp.theta
    beta  = temp.beta
}
   beta.matrix[z+1,] = beta
   theta.matrix[z+1,] = theta
  for( c in 1:col ){  beta.matrix.n[z,c] = beta.matrix[z+1,c] }
   for( r in 1:row ){   theta.matrix.n[z,r] =  theta.matrix[z+1,r] }
###############  2 ND STEP   ########################################################################
##############  preparation the expected response matrix  ##############################################
p = matrix(0,nrow=row, ncol=col)
   for(r in 1:row){
     for(c in 1:col){
          p[r,c] <- exp(theta[r]-beta[c])/(1+exp(theta[r]-beta[c]))
                      }
            }
#filename = paste("data_p",z,".txt")
# write.table(p, filename,row.names=FALSE,col.names=FALSE)
####################################    3 RD STEP   ##############################################
#############  estimated variance components  ######################################################
# GT TERMS
            # Grand mean
            Xbar[z] <- mean(p[,])
            # Person specific mean item response  (vector length 500)
            for (r in 1:row) {
                        XbarP[z] <- (sum(p[r,])/col)- Xbar[z]
```

```
            }
            # Item specific mean person response (vector length of 5)
            for (k in 1:col) {
                        XbarI[z] <- (sum(p[,k])/row) - Xbar[z]
            }

            # Person by item interaction effect (500 by 5 matrix)
            for (r in 1:row) {
                        for (k in 1:col) {
                                    Xbarinteraction[z] <-p[r,k] - XbarP[z] - XbarI[z] - Xbar[z]
                        }
             }
            # Variance Components
            for (r in 1:row) {
            for (k in 1:col) {
                                    ss.e[z] <- p[r,k]*(1-p[r,k])
                        }
            }
            sig2.e[z] <- mean(ss.e[z])
             sum2.p[z] <-(XbarP[z]^2 )
            sig2.p[z] <- mean(sum2.p[z])
            sum2.i[z] <-(XbarI[z]^2)
            sig2.i[z] <- mean(sum2.i[z])
            sum2.pi[z] <-(Xbarinteraction[z]^2)
            sig2.pi[z] <- mean(sum2.pi[z])
             sig2.pie[z] <- sig2.pi[z] +sig2.e[z]
            # GT Coefficients
            rho2[z] <- sig2.p[z]/(sig2.p[z] + sig2.pi[z]/col + sig2.e[z]/col)
            phi[z] <-  sig2.p[z]/(sig2.p[z] + sig2.pi[z]/col +  sig2.e[z]/col + sig2.i[z]/col)
}
 # close loop z
##########################################Start Loop i ####################################
   sig2.e.3[i] = mean(sig2.e[])
 #   sd.sig2.e.3[i] = sd(sig2.e[])
    sig2.p.3[i] = mean(sig2.p[])
 #   sd.sig2.p.3[i] = sd(sig2.p[])
    sig2.i.3[i] = mean(sig2.i[])
 #   sd.sig2.i.3[i] = sd(sig2.i[])
    sig2.pi.3[i] = mean(sig2.pi[])
```

```
  #   sd.sig2.pi.3[i] = sd(sig2.pi[])
      sig2.pie.3[i] = mean(sig2.pie[])
  #   sd.sig2.pie.3[i] = sd(sig2.pie[])
      rho2.3[i] = mean(rho2[])
  #   sd.rho2.3[i] = sd(rho2[])
      phi.3[i] = mean(phi[])
  #   sd.phi.3[i] = sd(phi[])
##########################################################################################
for (k in 1:T){
      beta.mean.i[k,i] = mean(beta.matrix.n[,k])
}
##########################################################################################
for (j in 1:N){
      theta.mean.i[j,i] = mean(theta.matrix.n[,j])
}
############LOOP ss######################################################################
cat(c("loop :",i),fill=T)
}
#######################################theta and beta #####################################
beta.mean = c() ; theta.mean = c();
beta.sd = c() ; theta.sd = c();
for (k in 1:T){
      beta.mean[k] =  mean(beta.mean.i[k,])
      beta.sd[k] = sd(beta.mean.i[k,])
}
##########################################################################################
for (j in 1:N){
      theta.mean[j] = mean(theta.mean.i[j, ])
      theta.sd[j] = sd(theta.mean.i[j, ])
}
##########################################################################################
     mean.sig2.e.3   = mean(sig2.e.3[])
     sd.sig2.e.3  = sd(sig2.e.3[])
     mean.sig2.p.3 = mean(sig2.p.3[])
     sd.sig2.p.3 = sd(sig2.p.3[])
     mean.sig2.i.3 = mean(sig2.i.3[])
     sd.sig2.i.3 = sd(sig2.i.3[])
     mean.sig2.pi.3 = mean(sig2.pi.3[])
      sd.sig2.pi.3 = sd(sig2.pi.3[])
```

```
mean.sig2.pie.3 = mean(sig2.pie.3[])
 sd.sig2.pie.3 = sd(sig2.pie.3[])
mean.rho2.3 = mean(rho2.3[])
sd.rho2.3 = sd(rho2.3[])
mean.phi.3 = mean(phi.3[])
 sd.phi.3 = sd(phi.3[])

##############################################Delta and Theta ###################################
output_beta3 = data.frame(beta.mean,beta.sd)
write.table(output_beta3,file = "beta3.txt", sep = "\t",row.name = FALSE)
output_theta3= data.frame(theta.mean, theta.sd)
write.table(output_theta3,file = "theta3.txt", sep = "\t",row.name = FALSE)
################################################################################################
output_mean_sd_3= data.frame(mean.sig2.e.3,sd.sig2.e.3,
mean.sig2.p.3 ,sd.sig2.p.3 ,
mean.sig2.i.3 , sd.sig2.i.3 ,
mean.sig2.pi.3 , sd.sig2.pi.3 ,
mean.sig2.pie.3 , sd.sig2.pie.3 ,
mean.rho2.3 , sd.rho2.3 ,
mean.phi.3 , sd.phi.3 )
write.table(output_mean_sd_3,file = "mean_sd_3.txt", sep = "\t",row.name = FALSE)
#############################################################################################
#par(ask = T)
cat(c("----------------------------- numerical  METHOD ----------------"),fill = T)
cat(c(" mean.sig2.e.3   = ", round(mean.sig2.e.3,6),  " sd.sig2.e.3   = ",round(sd.sig2.e.3,6),
 " mean.sig2.p.3   = ", round(mean.sig2.p.3,6), " sd.sig2.p.3   = ",round(sd.sig2.p.3,6),
 " mean.sig2.i.3 = ", round(mean.sig2.i.3,6), " sd.sig2.i.3 = " ,round(sd.sig2.i.3,6),
 " mean.sig2.pi.3 = ", round(mean.sig2.pi.3,6), "sd.sig2.pi.3 = ", round(sd.sig2.pi.3,6),
 " mean.sig2.pie.3 = ", round(mean.sig2.pie.3,6), "sd.sig2.pie.3 = ", round(sd.sig2.pie.3,6),
" mean.rho2.3 = ", round(mean.rho2.3,6), "sd.rho2.0 = " ,round(sd.rho2.3,6),
  "mean.phi.3 = ", round(mean.phi.3,6), "sd.phi.0 = " , round(sd.phi.3,6)),fill = T)
########################################END of Method1-3########################################
n=N
i=T
################################################################################################
#GENERATE TRUE PARAMETERS
################################################################################################
theta_true = matrix(nrow = N, ncol = ss)
beta_true =  matrix(nrow = T, ncol = ss)
```

```
sigma2.e_t = c(); sigma2.p_t = c() ;
sigma2.i_t = c(); sigma2.pi_t = c();
sigma2.pie_t = c() ; rho2_t = c(); phi_t = c()
for (j in 1:ss){
theta_t<-rnorm(n,0,1)
beta_t<-numeric(i)

for (k in 1:i)
{
beta_t[k]<-rnorm(1,0.5,0.5)
}
prob_t<-matrix(1,n,i)
x<-matrix(0,n,i)
score<-numeric(n)
for (p in 1:n)
{
for (k in 1:i)
  {
    prob_t[p,k]<-exp(theta_t[p]-beta_t[k])/(1+exp(theta_t[p]-beta_t[k]))
  }
 }
mu.p<-numeric(n)
mu.i<-numeric(i)
 theta<-numeric(n)
 beta<-numeric(i)
v.theta.beta<-matrix(1,n,i)
ss.e<-matrix(1,n,i)
ss.p<-numeric(n)
ss.i<-numeric(i)
ss.pi<-matrix(1,n,i)
#1) true grand mean
mu<-mean(prob_t[,])
#2) person effect
for (p in 1:n)
{
mu.p[p]<-mean(prob_t[p,])
 theta[p]<-(sum(prob_t[p,])/i)-mu
}
#3) item effect
```

```
for (k in 1:i)
{
mu.i[k]<-mean(prob_t[,k])
 beta[k]<-(sum(prob_t[,k])/n)-mu
}
##########################################################################################################
 for(k in 1:i){beta_true[k,j] =  beta[k]}
   for(p in 1:n){theta_true[p,j] = theta[p]}
##########################################################################################################
#4) iteraction effect (between person and item)
for (p in 1:n)
{
  for (k in 1:i)
   {
     v.theta.beta[p,k]<-prob_t[p,k]-theta[p]-beta[k]-mu
   }
}
##variance components (random effects)
#1) errors variance
for (p in 1:n)
{
 for (k in 1:i)
 {
  ss.e[p,k]<-prob_t[p,k]*(1-prob_t[p,k])
 }
}
sigma2.e_t[j] <-mean(ss.e[,])
#2) person variance
for (p in 1:n)
{
ss.p[p]<-(theta[p])^2
}
sigma2.p_t[j] <-mean(ss.p[])
#3) item variance
for (k in 1:i)
{
ss.i[k]<-(beta[k])^2
}
sigma2.i_t[j] <-mean(ss.i[])
```

```
#4) interaction variancesigma2.pi_t[j] <-mean(ss.pi[,])
sigma2.pie_t[j] <- sigma2.pi_t[j] +sigma2.e_t[j]
rho2_t[j] <- sigma2.p_t[j]/(sigma2.p_t[j] + sigma2.pi_t[j]/i + sigma2.e_t[j]/i)
phi_t[j] <-  sigma2.p_t[j]/(sigma2.p_t[j] + sigma2.pi_t[j]/i +  sigma2.e_t[j]/i + sigma2.i_t[j]/i)
#####################################Loop j#####################################################
}
###############################Mean Absolute Deviation############################################
d_beta =   matrix(nrow = T, ncol = ss)
d_theta = matrix(nrow = N, ncol = ss)
MAD_beta = c();MAD_theta = c()
################################################################################################
##########1) MAD for beta
for (k in 1:T){
for (d in 1:ss){
d_beta[k,d]<-abs((beta.mean.i[k,d]-beta_true[k,d])/beta_true[k,d])
}
}
for (k in 1:T){
MAD_beta[k] <- mean(d_beta[k,])
}
##########2) MAD for theta
for (k in 1:N){
for (d in 1:ss){
d_theta[k,d]<-abs((theta.mean.i[k,d]-theta_true[k,d])/theta_true[k,d])
}
}
for (k in 1:N){
MAD_theta[k] <- mean(d_theta[k,])
}
##############################################################################################
mean_MAD_beta = mean(MAD_beta[])
mean_MAD_theta = mean(MAD_theta[])
################################################################################################
############################################  Numerical  Method  ##################################
################################################################################################
MAD_sig2.e.3 = mean(abs((sigma2.e_t[]-sig2.e.3[])/sigma2.e_t[]))
MAD_sig2.p.3 = mean(abs((sigma2.p_t[]-sig2.p.3[])/sigma2.p_t[]))
MAD_sig2.i.3 = mean(abs((sigma2.i_t[]-sig2.i.3[])/sigma2.i_t[]))
MAD_sig2.pi.3 = mean(abs((sigma2.pi_t[]-sig2.pi.3[])/sigma2.pi_t[]))
```

```
MAD_sig2.pie.3 = mean(abs((sigma2.pie_t[]-sig2.pie.3[])/sigma2.pie_t[]))
MAD_rho2.3 = mean(abs((rho2_t[]-rho2.3[])/rho2_t[]))
MAD_phi.3 = mean(abs((phi_t[]-phi.3[])/phi_t[]))
EUC.3 = mean(sqrt((sigma2.e_t[]-sig2.e.3[])^2+(sigma2.p_t[]-sig2.p.3[])^2+
(sigma2.i_t[]-sig2.i.3[])^2+(sigma2.pi_t[]-sig2.pi.3[])^2+
(sigma2.pie_t[]-sig2.pie.3[])^2))
##########################################################################################
AD_sig2.e.3 = mean(abs((sigma2.e_t[]-sig2.e.3[])/sigma2.e_t[]))
AD_sig2.p.3 = mean(abs((sigma2.p_t[]-sig2.p.3[])/sigma2.p_t[]))
AD_sig2.i.3 = mean(abs((sigma2.i_t[]-sig2.i.3[])/sigma2.i_t[]))
AD_sig2.pi.3 = mean(abs((sigma2.pi_t[]-sig2.pi.3[])/sigma2.pi_t[]))
AD_sig2.pie.3 = mean(abs((sigma2.pie_t[]-sig2.pie.3[])/sigma2.pie_t[]))
AD_rho2.3 = mean(abs((rho2_t[]-rho2.3[])/rho2_t[]))
AD_phi.3 = mean(abs((phi_t[]-phi.3[])/phi_t[]))
#########################################Delta and Theta ##################################
output_MAD_beta= data.frame(MAD_beta)
write.table(output_MAD_beta,file = "MAD_beta.txt", sep = "\t",row.name = FALSE)
output_MAD_theta= data.frame(MAD_theta)
write.table(output_MAD_theta,file = "MAD_theta.txt", sep = "\t",row.name = FALSE)
##########################################################################################
output_MAD_3= data.frame(MAD_sig2.e.3, MAD_sig2.p.3,MAD_sig2.i.3,
 MAD_sig2.pi.3,MAD_sig2.pie.3,MAD_rho2.3,MAD_phi.3,EUC.3)
write.table(output_MAD_3,file = "MAD_3.txt", sep = "\t",row.name = FALSE)
##########################################################################################
output_AD_3= data.frame(AD_sig2.e.3, AD_sig2.p.3,AD_sig2.i.3,
 AD_sig2.pi.3,AD_sig2.pie.3,AD_rho2.3,AD_phi.3)
write.table(output_AD_3,file = "AD_3.txt", sep = "\t",row.name = FALSE)
##########################################################################################
#par(ask = T)
cat(c("-------------------------MAD  NUMERICAL  METHOD ----------------"),fill = T)
cat(c(" MAD_sig2.e.3   = ", round(MAD_sig2.e.3,6),
 " MAD_sig2.p.3   = ", round(MAD_sig2.p.3,6),
 " MAD_sig2.i.3 = ", round(MAD_sig2.i.3,6),
 " MAD_sig2.pi.3 = ", round(MAD_sig2.pi.3,6),
 " MAD_sig2.pie.3 = ", round(MAD_sig2.pie.3,6),
 " MAD_rho2.3 = ", round(MAD_rho2.3,6),
 " MAD_phi.3 = ", round(MAD_phi.3,6),
   "EUC.3 = " , round(EUC.3,6)), fill = T)
##########################################################################################
```

```
cat(c("-----------------------------   MAD Mean of theta and beta ----------------"),fill = T)
cat(c(" mean_MAD_beta  = ", round(mean_MAD_beta,6),
 " mean_MAD_theta   = ", round(mean_MAD_theta,6)), fill = T)
####################################################################################################
t.test(AD_sig2.e.3, mu=0)
sd(AD_sig2.e.3)
t.test(AD_sig2.p.3, mu=0)
sd(AD_sig2.p.3)
t.test(AD_sig2.i.3, mu=0)
sd(AD_sig2.i.3)
t.test(AD_sig2.pi.3, mu=0)
sd(AD_sig2.pi.3)
t.test(AD_sig2.pie.3, mu=0)
sd(AD_sig2.pie.3)
t.test(AD_rho2.3, mu=0)
sd(AD_rho2.3)
t.test(AD_phi.3, mu=0)
sd(AD_phi.3)
####################################################################################################
```

ประวัติผู้เขียนวิทยานิพนธ์

นายชนะศึก นิชานนท์ เป็นบุตรชายคนเดียวของนายระพินทร์-นางจรรยา นิชานนท์ เกิดเมื่อวันพฤหัสบดีที่ 6 เดือนธันวาคม พ.ศ. 2522 สำเร็จการศึกษาระดับปริญญาตรี คณะครุศาสตร์ เกียรตินิยมอันดับ 2 จากจุฬาลงกรณ์มหาวิทยาลัย จากนั้นเข้าศึกษาต่อในระดับปริญญาโท สาขาวิชาวิจัยการศึกษา คณะครุศาสตร์ เกรดเฉลี่ย 3.97 จากจุฬาลงกรณ์มหาวิทยาลัย หลังจากนั้นเข้าทำงานรับราชการที่กองวิจัยทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ ในตำแหน่งนักวิชาการศึกษา 4 ก่อนโอนย้ายข้าราชการ มาเป็นอาจารย์ประจำ สาขาวิชาวิจัยการศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏสวนดุสิต และดำรงตำแหน่งรองผู้อำนวยการสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏสวนดุสิต ในปี 2550 ได้รับทุนโครงการปริญญาเอกกาญจนาภิเษก (คปก.) ของสำนักงานกองทุนสนับสนุนการวิจัย (สกว.) โดยเข้ารับศึกษาในระดับปริญญาเอก สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ที่จุฬาลงกรณ์มหาวิทยาลัย นอกจากนี้ยังได้รับทุนให้ไปศึกษาและวิจัย ที่ University of California, Berkeley ประเทศสหรัฐอเมริกา และสำเร็จการศึกษาด้วยเกรดเฉลี่ย 4.00 ปัจจุบันดำรงตำแหน่งเป็นอาจารย์ ประจำคณะครุศาสตร์ และดำรงตำแหน่ง ผู้อำนวยการสถาบันวิจัยและพัฒนา มหาวิทยาลัยราชภัฏสวนดุสิต