ประสิทธิภาพของการประมาณค่าพารามิเตอร์ผู้สอบแบบสุ่มข้อสอบสำหรับโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์

นางทิพวัลย์ ปัญจมะวัต

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2555

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย
บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ที่ส่งผ่านทางบัณฑิตวิทยาลัย
The abstract and full text of theses from the academic year 2011 in Chulalongkorn University Intellectual Repository(CUIR)
are the thesis authors' files submitted through the Graduate School.

EFFICIENCY OF EXAMINEE PARAMETER ESTIMATION WITH RANDOM ITEM FOR ONE-PARAMETER ITEM RESPONSE MODEL

Mrs. Tippawan Panjamawat

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 2012

Copyright of Chulalongkorn University

หัวข้อวิทยานิพนธ์	ประสิทธิภาพของการประมาณค่าพารามิเตอร์ผู้สอบแบบสุ่มข้อสอบสำหรับโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์
โดย	นางทิพวัลย์ ปัญจมะวัต
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	ดร.ชูศักดิ์ ขัมภลิขิต

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

……คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร.ชนิตา รักษ์พลเมือง)

คณะกรรมการสอบวิทยานิพนธ์

……ประธานกรรมการ
(รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ)

……..อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

……..อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(ดร.ชูศักดิ์ ขัมภลิขิต)

……..กรรมการ
(รองศาสตราจารย์ ดร.โชติกา ภาษีผล)

……..กรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.ณัฏฐภรณ์ หลาวทอง)

……กรรมการภายนอกมหาวิทยาลัย
(รองศาสตราจารย์ ดร.ดิเรก ศรีสุโข)

ทิพวัลย์ ปัญจมะวัต: ประสิทธิภาพของการประมาณค่าพารามิเตอร์ผู้สอบแบบสุ่มข้อสอบสำหรับโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์. (EFFICIENCY OF EXAMINEE PARAMETER ESTIMATION WITH RANDOM ITEM FOR ONE-PARAMETER ITEM RESPONSE MODEL)
อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ศ.ดร.ศิริชัย กาญจนวาสี,
อ.ที่ปรึกษาวิทยานิพนธ์ร่วม : ดร.ชูศักดิ์ ขัมภลิขิต, 145 หน้า.

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างการประมาณค่าโดยโมเดลประมาณค่าพารามิเตอร์แบบอิทธิพลเจาะจงข้อสอบกับแบบอิทธิพลสุ่มข้อสอบ 3 รูปแบบ คือ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ด้วยโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ โดยศึกษากับข้อมูลจำลอง (simulation) จากโปรแกรม R และประมวลผลภายใต้คำสั่งการประมวลผลด้วยโปรแกรม WinBUGS ด้วยการเชื่อมโยงผ่านคำสั่ง R2 WinBUGS Package โดยจำลองข้อมูลที่แตกต่างกัน 48 เงื่อนไข (4 x 3 x 4) ประกอบด้วย 1) ลักษณะการแจกแจงก่อนหน้าของความยากข้อสอบ 4 แบบ คือ Normal, Uniform, Negative skew และ Positive skew 2) ความยาวของแบบทดสอบ 3 เงื่อนไข คือ 30, 60 และ 90 ข้อ และ 3) จำนวนผู้สอบ 4 เงื่อนไข คือ 100, 300, 500 และ 1,000 คน กำหนดจำนวนการทำซ้ำจำนวน 1,000 รอบในแต่ละเงื่อนไข การประเมินประสิทธิภาพของการประมาณค่าพารามิเตอร์ผู้สอบพิจารณาจากความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์และค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง นอกจากนี้ได้ศึกษากับข้อมูลผลการสอบ O-Net ปีการศึกษา 2553 ของนักเรียนมัธยมศึกษาปีที่ 6 รายวิชาภาษาอังกฤษ คณิตศาสตร์ และภาษาไทย โดยสุ่มผู้สอบจำนวน 100, 300, 500 และ 1,000 คน ในแต่ละรายวิชา การประเมินประสิทธิภาพของการประมาณค่าพารามิเตอร์ผู้สอบพิจารณาจากค่าสถิติ Akaike criterion (AIC)

ผลการวิจัยพบว่า

1) โมเดลการประมาณค่าพารามิเตอร์ผู้สอบรูปแบบ FPRI ที่ลักษณะการแจกแจงความยากข้อสอบแบบ Normal มีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์และค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง มีค่าต่ำที่สุดสำหรับข้อมูลจำลอง และเป็นโมเดลที่มีประสิทธิภาพดีที่สุดในเกือบทุกเงื่อนไขที่ศึกษา

2) โมเดลการประมาณค่าพารามิเตอร์ผู้สอบรูปแบบ FPRI มีประสิทธิภาพในการประมาณค่าสอดคล้องกับข้อมูลมากที่สุด รองลงมาคือ รูปแบบ RPRI และ รูปแบบ RPFI ตามลำดับ สำหรับข้อมูลผลการสอบ O-NET

ภาควิชา วิจัยและจิตวิทยาการศึกษา ลายมือชื่อนิสิต .....
สาขาวิชา การวัดและประเมินผลการศึกษา ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์หลัก .....
ปีการศึกษา 2555 ลายมือชื่ออ.ที่ปรึกษาวิทยานิพนธ์ร่วม .....

5184224627: MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEYWORDS: ESTIMATION EFFICIENCY / PARAMETER ESTIMATION / RANDOM ITEM EFFECT

TIPPAWAN PANJAMAWAT: EFFICIENCY OF EXAMINEE PARAMETER ESTIMATION WITH RANDOM ITEM FOR ONE-PARAMETER ITEM RESPONSE MODEL. ADVISOR: PROF. SIRICHAI KARNJANAWASEE, Ph.D., CO-ADVISOR: CHOOSAK KHAMPALIKIT, Ph.D., 145 pp.

The purpose of this research was to investigate the efficiency of examinee's parameter estimation by one-parameter item response model. The parameter consists of the evaluation of specific item response and of three random item response models: Random Persons-Fixed Items (RPFI), Fixed Persons-Random Items (FPRI), and Random Persons-Random Items (RPRI).The data was formulated by using R Program to run the simulation assessed by WinBUGS program and connected through R2 WinBUGS Package code. The data model consisted of 48 conditions: including the initial distribution of the test in 4 distributions of difficulty – Normal, Uniform, Negative skew and Positive skew; 3 conditions of the test length (30, 60, 90 questions); and 4 conditions of the amount of examinees (100, 300, 500 and 1,000 examinees). The efficiency evaluation of the examinees' parameter estimation was considered from the standard error and the bias of estimation. In addition, this research used the data from the 3 subjects (English, Mathematics and Thai) of 2003's O-Net test results of Mathayom 6 students by random examinees (100, 300, 500 and 1,000 examinees) and on three subjects by assessing the parameter estimation efficiency of the examinees from Akaike criterion (AIC) statistics.

The results of research were as follows:

1) The FPRI parameter estimation model shown the highest estimation efficiency, followed by the RPFI and RPRI parameter estimation model respectively. In addition, the initial normal distribution of the test's difficulty level had the least discrepancy and inequity of parameter estimation according to the simulation. Also, all three types of parameter estimation models tend to gain more estimation efficiency when the sample size is larger or when the test is longer.

2) The FPRI parameter estimation model had the least AIC in every subject and every sample size. The study revealed the most efficient of estimation by FPRI, RPFI and RPRI respectively.

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

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

Academic Year ..2012 ..... Co-Advisor's Signature.....

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้สำเร็จได้ด้วยดี ด้วยการดูแลให้คำปรึกษาและความเมตตากรุณาของศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาวิทยานิพนธ์หลักที่ได้ให้ความรู้ ช่วยปรับปรุงแก้ไข ดิฉันรู้สึกซาบซึ้งและขอกราบขอบพระคุณท่านอาจารย์มา ณ โอกาสนี้

ขอกราบขอบพระคุณ อ.ดร. ชูศักดิ์ ขัมภลิขิต อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ที่ให้ความรู้ความเมตตา รวมทั้งแนะแนวทางในการปรับปรุงแก้ไข ทำให้สามารถดำเนินการวิจัยในครั้งนี้สำเร็จลุล่วง

ขอกราบขอบพระคุณ รศ.ดร.ดิเรก ศรีสุโข รศ.ดร. ศิริเดช สุชีวะ รศ.ดร.โชติกา ภาษีผล และ ผศ.ดร.ณัฏฐภรณ์ หลาวทอง คณะกรรมการวิทยานิพนธ์ที่ช่วยขัดเกลางานวิจัยให้มีความชัดเจนและมีความสมบูรณ์เพิ่มมากขึ้น

ขอขอบคุณ ดร.ชนะศึก นิชานนท์ ที่จุดประกายความสนใจในเรื่องนี้และให้ความช่วยเหลือข้อมูลต่างๆ ขอขอบคุณ ดร.ธเกียรติกมล ทองงอก ที่ให้แนวทางในการดำเนินการวิจัย ขอขอบคุณ ดร. สิวะโชติ ศรีสุทธิยากร ที่ให้ความช่วยเหลือและให้คำปรึกษาในการตรวจสอบคำสั่งสำหรับการวิเคราะห์ข้อมูล ตลอดจน ดร.ทัศน์ศิรินทร์ สว่างบุญ ดร.ชุตินันท์ จันทรเสนานนท์ คุณสรัญญา จันทร์ชูสกุล คุณสุกัญญา จันทวาลย์ คุณนัยนา จันตะเสน และเพื่อนๆ พี่ๆ น้องๆ ในภาควิชาวิจัยและจิตวิทยาการศึกษาทุกท่านที่ได้ช่วยเหลือและให้กำลังใจผู้วิจัยเสมอมา

ผู้วิจัยขอขอบพระคุณ สำนักงานกองทุนสนับสนุนการวิจัย (สทศ.)ที่ได้ให้การสนับสนุนด้านข้อมูลเพื่อการวิจัย ทำให้ได้แนวทางในการประยุกต์ใช้กับข้อมูลเชิงประจักษ์

ท้ายที่สุดนี้ ขอขอบคุณ ผศ.ชูพงศ์ ปัญจมะวัต คู่ชีวิต คู่คิด ที่คอยสนับสนุน กระตุ้นและเป็นกำลังใจที่ยิ่งใหญ่แก่ผู้วิจัยให้สามารถสำเร็จการศึกษาทั้งในระดับปริญญาบัณฑิต มหาบัณฑิต และดุษฎีบัณฑิตในที่สุด

สารบัญ

หน้า

บทคัดย่อภาษาไทย..... ง

บทคัดย่อภาษาอังกฤษ..... จ

กิตติกรรมประกาศ..... ฉ

สารบัญ..... ช

สารบัญตาราง..... ฌ

สารบัญภาพ..... ฎ

บทที่

1 บทนำ..... 1

ความเป็นมาและความสำคัญของปัญหา..... 1

คำถามการวิจัย..... 3

วัตถุประสงค์การวิจัย..... 4

สมมติฐานของการวิจัย..... 4

ขอบเขตของการวิจัย..... 5

นิยามศัพท์เฉพาะ..... 6

ประโยชน์ที่คาดว่าจะได้รับ..... 7

2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... 8

ตอนที่ 1 แนวคิดทฤษฎีเกี่ยวกับโมเดลทฤษฎีการตอบสนองข้อสอบ..... 8

ตอนที่ 2 การประมาณค่าพารามิเตอร์..... 27

ตอนที่ 3 โมเดลการตอบสนองข้อสอบเชิงสุ่ม..... 32

ตอนที่ 4 โปรแกรมการจำลองข้อมูลและวิเคราะห์ข้อมูล..... 44

ตอนที่ 5 บทความและงานวิจัยที่เกี่ยวข้อง..... 54

ตอนที่ 6 กรอบแนวคิดในการวิจัย..... 60

3 วิธีดำเนินการวิจัย..... 62

การจำลองข้อมูล..... 62

ขั้นตอนการวิจัย..... 62

การวิเคราะห์ข้อมูล..... 66

การดำเนินการวิจัย..... 67

บทที่		หน้า
4	ผลการวิเคราะห์ข้อมูล.....	69
	ตอนที่ 1 การวิเคราะห์ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบจากข้อมูลจำลอง : ความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์.....	69
	ตอนที่ 2 การวิเคราะห์ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบจากข้อมูลจำลอง : ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าพารามิเตอร์.....	83
	ตอนที่ 3 การวิเคราะห์ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบจากข้อมูลเชิงประจักษ์.....	98
5	สรุปผลการวิจัย อภิปรายผลการวิจัย และข้อเสนอแนะ.....	101
	สรุปผลการวิจัย.....	101
	อภิปรายผลการวิจัย.....	104
	ข้อเสนอแนะ.....	106
รายการอ้างอิง.....		109
ภาคผนวก.....		115
ประวัติผู้เขียนวิทยานิพนธ์.....		145

สารบัญตาราง

ตารางที่		หน้า
2.1	ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ.....	18
2.2	สูตรการคำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ.....	26
2.3	โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ จำแนกตามการสุ่มหรือการเจาะจงบุคคลและข้อสอบ.....	42
2.4	สรุปแนวคิดเกี่ยวกับโมเดลการตอบสนองข้อสอบเชิงสุ่ม.....	56
4.1	ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก.....	71
4.2	ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก.....	73
4.3	ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก.....	74
4.4	ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก.....	75
4.5	ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ.....	77

ตารางที่ หน้า

4.6 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ..... 78

4.7 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ..... 80

4.8 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ..... 81

4.9 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก..... 84

4.10 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก..... 86

4.11 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก..... 87

ตารางที่ หน้า

4.12 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก..... 88

4.13 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ..... 90

4.14 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ..... 91

4.15 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ..... 93

4.16 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ..... 94

4.17 สรุปประสิทธิภาพของโมเดลในการประมาณค่าพารามิเตอร์ จำแนกตามรูปแบบการประมาณค่าพารามิเตอร์ และเงื่อนไขการศึกษา 48 เงื่อนไข..... 96

4.18 ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบวิชาภาษาอังกฤษ (70 ข้อ)..... 98

4.19 ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบวิชาคณิตศาสตร์ (40 ข้อ)..... 99

4.20 ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบวิชาภาษาไทย (90 ข้อ)..... 99

สารบัญภาพ

ภาพที่		หน้า
1	โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ของตัวอย่างข้อสอบ 3 ข้อ.....	20
2	โค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์ของตัวอย่างข้อสอบ 3 ข้อ.....	21
3	โค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ของตัวอย่างข้อสอบ 3 ข้อ.....	22
4	ความเป็นตัวแทนของโมเดลอิทธิพลเจาะจงกับอิทธิพลสุ่ม.....	34
5	ขั้นตอนการสร้างแบบสอบ.....	36
6	กรอบแนวคิดในการวิจัย.....	61
7	ขั้นตอนการดำเนินการวิจัยกรณีการจำลองข้อมูล.....	65
8	ขั้นตอนการดำเนินการวิจัยกรณีข้อมูลจริง.....	66

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวัดผลทางการศึกษาเป็นกระบวนการเชิงระบบเพื่อวัดความสามารถหรือคุณลักษณะแฝงของบุคคลตามวัตถุประสงค์ของการวัด โดยใช้สิ่งเร้าที่มีความเป็นตัวแทนคุณลักษณะแฝงที่ต้องการวัด เช่น ข้อความ ข้อคำถาม หรือข้อสอบ มากระตุ้นให้บุคคลแสดงพฤติกรรมตอบสนอง โดยส่วนใหญ่แล้วผู้สอนจะเขียนข้อสอบเพื่อวัดความสามารถหรือคุณลักษณะอันใดอันหนึ่ง ที่ประกอบด้วยองค์ประกอบย่อยๆหลายองค์ประกอบ ผู้สอนจะเขียนข้อสอบมา 1 ชุด และใช้ข้อสอบทั้งหมดในการทดสอบความสามารถของผู้เรียน ข้อสอบชุดนี้เปรียบเสมือนเป็นตัวแทนของประชากรข้อสอบเพื่อวัดคุณลักษณะที่ต้องการ ดังนั้นการวัดผลทุกครั้งจึงเป็นเพียงตัวแทนพฤติกรรม แล้วนำผลการตอบมาผ่านกระบวนการแปลผลโดยการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบด้วยทฤษฎีการทดสอบต่างๆ เช่น ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ทฤษฎีการสรุปอ้างอิง (Generalizability Theory: GT) และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) เป็นต้น เพื่อให้การตัดสินผลมีความถูกต้องแม่นยำน่าเชื่อถือ

ความถูกต้องแม่นยำในการประมาณค่าพารามิเตอร์ด้วยการใช้โมเดลทฤษฎีการตอบสนองข้อสอบในการประมาณค่าพารามิเตอร์ความสามารถและพารามิเตอร์ข้อสอบ ยังคงมีประเด็นที่น่าสนใจศึกษาเกี่ยวกับการประมาณค่าด้วยโมเดลอิทธิพลเจาะจงข้อสอบหรือโมเดลอิทธิพลสุ่มข้อสอบ ลักษณะของโมเดลอิทธิพลเชิงสุ่มและโมเดลอิทธิพลเจาะจงมีความแตกต่างกันในเชิงแนวคิด โมเดลอิทธิพลเจาะจงจะสมมติให้ข้อมูลของตัวแปรที่นำมาวิเคราะห์มีค่าคงที่หรือมีค่าเจาะจง ทำให้เมื่อได้ผลวิเคราะห์แล้วจะสรุปอ้างอิงได้ในขอบเขตจำกัดเท่ากับข้อมูลที่นำมาวิเคราะห์เท่านั้น ในขณะที่โมเดลอิทธิพลเชิงสุ่มจะสมมติให้ข้อมูลของตัวแปรที่นำมาวิเคราะห์นั้นได้มาอย่างสุ่มจากกลุ่มประชากร จึงทำให้มีความคลาดเคลื่อนจากค่าเฉลี่ยอย่างสุ่มเป็น

องค์ประกอบในโมเดล และเมื่อได้ผลวิเคราะห์แล้วสามารถสรุปอ้างอิงไปยังกลุ่มประชากรของตัวแปรที่สนใจนั้นได้

โดยที่การใช้โมเดลทฤษฎีการตอบสนองข้อสอบในการประมาณค่าพารามิเตอร์ความสามารถและพารามิเตอร์ข้อสอบในปัจจุบันมักใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าปลายทาง (Marginal maximum likelihood method) ดังสมการ

$$n_{pi}|\theta_p = \theta_p - \beta_i \qquad \text{เมื่อ} \quad \theta_p \approx N(\mu_\theta, \sigma_\theta^2)$$

เมื่อ $\mu_\theta, \sigma_\theta^2$ คือ ค่าเฉลี่ยของความสามารถ และค่าความแปรปรวนของความสามารถ ซึ่งวิธีการประมาณค่าลักษณะนี้เป็นโมเดลอิทธิพลสุ่มผู้สอบ-เจาะจงข้อสอบ ที่ทำให้การประมาณค่าความสามารถผู้สอบที่ได้จะเป็นความสามารถที่สรุปอ้างอิงไปยังประชากรของผู้สอบ และได้ค่าความยากของข้อสอบแต่ละข้อ ซึ่งมิได้เป็นค่าความสามารถของผู้สอบแต่ละคนและไม่สามารถสรุปอ้างอิงไปยังประชากรของข้อสอบได้ แต่สิ่งที่ผู้สอบและผู้สอนต้องการอย่างแท้จริงในสถานการณ์ของการทดสอบ คือ การสรุปอ้างอิงไปยังประชากรข้อสอบเพื่อวัดความสามารถของบุคคลแต่ละคน โมเดลที่จะตอบสนองความต้องการนี้ได้คือ โมเดลอิทธิพลแบบเจาะจงบุคคลและอิทธิพลข้อสอบเชิงสุ่ม สามารถเขียนแทนด้วยสมการของโมเดลการวัดความสามารถบุคคลได้ดังนี้

$$n_{pi}|\beta_i = \theta_p - \beta_i \qquad \text{เมื่อ} \quad \beta_i \approx N(\mu_\beta, \sigma_\beta^2) or U(0,1)$$

เมื่อ μ_β คือ ค่าเฉลี่ยของความยากของข้อสอบ และ σ_β^2 คือ ค่าความแปรปรวนความยากของข้อสอบ

โมเดลการวัดบุคคลที่ใช้อิทธิพลแบบเจาะจงบุคคลและอิทธิพลข้อสอบเชิงสุ่มนี้ จะทำให้สามารถประมาณค่าความสามารถของบุคคลได้โดยตรงและเพียงพอที่จะสรุปอ้างอิงผลการวัดนี้ให้ครอบคลุมข้อสอบข้อต่างๆ ซึ่งทำให้การวัดคุณลักษณะแฝงหรือความสามารถนั้นครอบคลุมคุณลักษณะที่ต้องการวัดและสามารถสรุปได้อย่างมั่นใจน่าเชื่อถือ อีกทั้งการใช้โมเดลการตอบสนองข้อสอบแบบอิทธิพลสุ่มข้อสอบ มีแนวโน้มที่จะให้ความคลาดเคลื่อนในการประมาณค่า

ลดลง คือ มีแนวโน้มที่จะให้ประสิทธิภาพในการประมาณค่าดีขึ้น รวมทั้งประสิทธิภาพของโมเดลแบบอิทธิพลสุ่มข้อสอบมีความสอดคล้องกับข้อมูลดีกว่าโมเดลแบบอิทธิพลเจาะจงข้อสอบ (De Boeck, 2008)

การศึกษาวิจัยครั้งนี้ผู้วิจัยจึงสนใจศึกษาเกี่ยวกับประสิทธิภาพของการวัดในด้านการประมาณค่าความสามารถผู้สอบผ่านทางทฤษฎีการตอบสนองข้อสอบแบบอิทธิพลสุ่มข้อสอบ โดยใช้โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ซึ่งเป็นโมเดลที่มีผู้นิยมนำมาใช้ เนื่องจากมีจำนวนพารามิเตอร์ที่ต้องประมาณค่าน้อยกว่าโมเดลทฤษฎีการตอบสนองข้อสอบอื่นๆ ทำให้ง่ายต่อการวิเคราะห์ ไม่ยุ่งยากซับซ้อน (Hambleton, 1989) อีกทั้งโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์นี้ยังให้ค่าประมาณของพารามิเตอร์ที่ถูกต้องกว่าโมเดลแบบอื่นๆ เมื่อนำไปใช้กับกลุ่มตัวอย่างขนาดเล็ก (Vijver, 1986; Lord, 1980) ด้วยเหตุดังกล่าวผู้วิจัยจึงสนใจนำโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ มาเป็นตัวแบบในการศึกษาประมาณค่าความสามารถผู้สอบผ่านทางทฤษฎีการตอบสนองข้อสอบแบบอิทธิพลสุ่มข้อสอบครั้งนี้ เพื่อเป็นพื้นฐานการศึกษาในขั้นต้นได้อย่างมีประสิทธิภาพ

คำถามวิจัย

โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบการสุ่มข้อสอบ มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ผู้สอบได้ดีขึ้นกว่าโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบการเจาะจงข้อสอบหรือไม่ โดยพิจารณาจาก

1. ความคลาดเคลื่อนในการประมาณค่าระหว่างโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) กับโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) และความคลาดเคลื่อนในการประมาณค่าระหว่างโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) กับโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ที่มีขนาดกลุ่มตัวอย่างเท่ากันในระดับขนาดกลุ่มตัวอย่างหลายระดับ และจำนวนข้อสอบที่แตกต่างกัน โดยใช้ข้อมูลจำลอง

2. ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) กับโมเดลการตอบสนองข้อสอบ 1

พารามิเตอร์ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองระหว่างโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) กับโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ที่มีขนาดกลุ่มตัวอย่างเท่ากันในระดับขนาดกลุ่มตัวอย่างหลายระดับ และจำนวนข้อสอบที่แตกต่างกัน โดยใช้ข้อมูลจำลอง

3. ประสิทธิภาพของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ 3 รูปแบบ ได้แก่ รูปแบบสุ่มผู้สอบ-เจาะจง ข้อสอบ (RPFI) รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) พิจารณาจากความสอดคล้องกลมกลืนของโมเดลกับข้อมูล โดยใช้ข้อมูลเชิงประจักษ์

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาประสิทธิภาพของการประมาณค่าของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ระหว่างการประมาณค่าโดยใช้รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ กับ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ กับ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ ด้วยข้อมูลจำลอง

2. เพื่อศึกษาความสอดคล้องของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ที่มีการประมาณค่าโดยใช้รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบกับ รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ กับ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ ด้วยข้อมูลเชิงประจักษ์

สมมติฐานของการวิจัย

วิธีการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ (IRT) จะมีความคลาดเคลื่อนในการประมาณค่าเกิดขึ้นด้วยเสมอ ทำให้มีการพัฒนาวิธีการประมาณค่าโดยการใช้ความแปรปรวนที่เป็นได้ของพารามิเตอร์นั้นๆ มาเป็นค่าตั้งต้น เรียกว่า การกระจายก่อนหน้าของพารามิเตอร์ (prior distribution) จะทำให้ความคลาดเคลื่อนลดลงได้ ซึ่งการประมาณค่าพารามิเตอร์ส่วนใหญ่มักใช้รูปแบบการเจาะจงข้อสอบ (fixed item) ในการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าปลายทาง (Marginal maximum likelihood) โดยการกำหนดค่าการกระจายก่อนหน้าของผู้สอบเป็นค่าตั้งต้นในการประมาณค่าพารามิเตอร์ความยากข้อสอบแต่ละข้อ มิได้เป็นการประมาณค่าพารามิเตอร์ความสามารถผู้สอบแต่ละคน แต่เป็นการสรุปอ้างอิงไปยังประชากรของผู้สอบ ดังนั้นการวิจัยครั้งนี้จึงใช้รูปแบบการสุ่มข้อสอบ (random item) เพื่อ

ประมาณค่าพารามิเตอร์ผู้สอบแต่ละคนได้โดยตรงและสามารถสรุปอ้างอิงผลการวัดไปยังข้อสอบข้อต่างๆว่าสามารถวัดได้ครอบคลุมคุณลักษณะที่ต้องการวัด อีกทั้งโมเดลการประมาณค่าพารามิเตอร์รูปแบบการสุ่มข้อสอบ มีแนวโน้มจะทำให้มีประสิทธิภาพในการประมาณค่าดีขึ้นและมีแนวโน้มความสอดคล้องกับข้อมูลดีกว่ารูปแบบเจาะจงข้อสอบ (De Boeck, 2008) ดังนั้นผู้วิจัยจึงมีสมมติฐานของการวิจัย ดังนี้

1. โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ จะมีความคลาดเคลื่อนในการประมาณค่าน้อยกว่า โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ ที่มีขนาดกลุ่มตัวอย่างเท่ากันในระดับขนาดกลุ่มตัวอย่างหลายระดับ และจำนวนข้อสอบที่แตกต่างกันจากข้อมูลจำลอง

2. โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง น้อยกว่า โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ ที่มีขนาดกลุ่มตัวอย่างเท่ากันในระดับขนาดกลุ่มตัวอย่างหลายระดับ และจำนวนข้อสอบที่แตกต่างกันจากข้อมูลจำลอง

3. โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ จะมีประสิทธิภาพของโมเดลสูงกว่า โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบจากข้อมูลเชิงประจักษ์

ขอบเขตการวิจัย

การศึกษาใช้วิธีจำลองข้อมูล โดยใช้โปรแกรม R และวิเคราะห์ข้อมูลด้วยโปรแกรม WinBUGS และการวิเคราะห์ข้อมูลเชิงประจักษ์ ตัวแปรที่ใช้ในการศึกษา ประกอบด้วย

ตัวแปรต้น 4 ตัวแปร ได้แก่

1) โมเดลตอบสนองข้อสอบ 1 พารามิเตอร์ ที่มีการสุ่ม 3 แบบ คือ แบบสุ่มผู้สอบ-เจาะจงข้อสอบ แบบเจาะจงผู้สอบ-สุ่มข้อสอบ และแบบสุ่มผู้สอบ-สุ่มข้อสอบ

2) ขนาดกลุ่มตัวอย่าง 4 ระดับ ได้แก่ 100, 300, 500 และ 1,000 คน

3) จำนวนข้อสอบ 3 ระดับ ได้แก่ 30, 60 และ 90 ข้อ และ

4) รูปแบบการแจกแจง Prior Distribution ความยากของข้อสอบ 4 แบบ คือ Normal, Uniform, การแจกแจงเบ้ซ้าย (Negative skew) และการแจกแจงเบ้ขวา (Positive skew)

รวมเงื่อนไขในการจำลองข้อมูล 144 เงื่อนไขที่ต้องการศึกษา โดยแต่ละเงื่อนไขทำซ้ำ 1,000 รอบ

ตัวแปรตาม 2 ตัวแปร ได้แก่

1) ค่าประมาณพารามิเตอร์ความสามารถผู้สอบ (θ)

2) ค่าประมาณพารามิเตอร์ความยากข้อสอบ (β)

เกณฑ์การประเมินประสิทธิภาพ ในกรณีข้อมูลจำลองพิจารณาจากค่าความคลาดเคลื่อนของการประมาณค่าความสามารถ; SE(θ) และค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง; MSE และในกรณีข้อมูลเชิงประจักษ์พิจารณาจากค่า Akaike information criterion; AIC ซึ่งค่าที่น้อยกว่าแสดงว่ามีความกลมกลืนระหว่างข้อมูลกับการประมาณค่ามากกว่า

นิยามศัพท์เฉพาะ

โมเดลเชิงสุ่ม (Random-effect model) หมายถึง โมเดลเชิงสถิติที่สมมติให้ตัวแปรอิสระเป็นตัวแปรสุ่ม โดยสมมติให้กลุ่มตัวอย่างข้อมูลของตัวแปรที่นำมาวิเคราะห์ได้มาอย่างสุ่มจากประชากรข้อมูลหลายๆค่า จึงมีค่าความคลาดเคลื่อนจากค่าเฉลี่ยซึ่งมีการแจกแจงแบบสุ่มเป็นส่วนประกอบหนึ่งในโมเดล

โมเดลเจาะจง (Fixed-effect model) หมายถึง โมเดลเชิงสถิติที่สมมติให้ตัวแปรอิสระเป็นตัวแปรคงที่ โดยสมมติให้กลุ่มตัวอย่างข้อมูลของตัวแปรที่นำมาวิเคราะห์ประกอบด้วยข้อมูลหลายๆค่าชุดหนึ่ง ในโมเดลเจาะจงจะไม่มีเทอมของความคลาดเคลื่อนจากค่าเฉลี่ยเป็นส่วนประกอบในโมเดล

ประสิทธิภาพของการประมาณค่าพารามิเตอร์ หมายถึง ความถูกต้องแม่นยำของการประมาณค่าพารามิเตอร์ พิจารณาได้จากดัชนีค่าความคลาดเคลื่อนของการประมาณค่า (SE(θ)) และความลำเอียงของการประมาณค่าพิจารณาจากค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean squared error ; MSE)

ค่าฟังก์ชั่นสารสนเทศของแบบสอบ $I(\theta)$ หมายถึง ความแม่นยำของการประมาณค่าความสามารถ ซึ่งเกิดจากผลรวมเชิงพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อรวมกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกัน

ความคลาดเคลื่อนของการประมาณค่าความสามารถ SE(θ) หมายถึง ความถูกต้องแม่นยำของการประมาณค่าความสามารถ ซึ่งเป็นสัดส่วนผกผันกับค่าฟังก์ชันสารสนเทศของแบบสอบ ถ้าค่า SE(θ) มีค่าสูง แสดงว่ามีความแม่นยำต่ำในการประมาณค่าความสามารถของผู้สอบ

ณ θ นั้น แต่ถ้า SE(θ) มีค่าต่ำ แสดงว่ามีความแม่นยำสูงในการประมาณค่าความสามารถของผู้สอบ ณ θ นั้น

ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean squared error; MSE) หมายถึง คุณสมบัติของตัวประมาณค่าที่มีค่าเฉลี่ยของการแจกแจงการสุ่มของตัวประมาณค่าต่างจากค่าพารามิเตอร์ หรือความแตกต่างระหว่างค่าที่ได้จากตัวประมาณค่ากับค่าจริง โดยพิจารณาจาก ค่าที่เข้าใกล้ศูนย์ แสดงว่าตัวประมาณค่ามีประสิทธิภาพในการประมาณค่าได้แม่นยำมาก

ประสิทธิภาพของโมเดลที่ใช้ในการประมาณค่า หมายถึง โมเดลที่ใช้ในการวิเคราะห์ประมาณค่ามีความเหมาะสมกับข้อมูล พิจารณาได้จากความสอดคล้องกลมกลืนของโมเดลกับข้อมูล ด้วยค่า Akaike information criterion (AIC) ที่ใช้เป็นเกณฑ์ในการคัดเลือกโมเดล ค่า AIC ที่มีค่าต่ำกว่าแสดงว่าโมเดลมีความสอดคล้องกับข้อมูลดีกว่า

Akaike information criterion (AIC) หมายถึง ค่าสถิติที่แสดงถึงความกลมกลืนระหว่างข้อมูลกับโมเดล โดยมีสูตรการคำนวณ AIC = 2k-2ln(L) โดยที่ k คือ จำนวนพารามิเตอร์ที่ต้องการประมาณค่า และ L คือ Likelihood อธิบายความว่า Likelihood เป็นการทดสอบความน่าจะเป็นของความกลมกลืนระหว่างพฤติกรรมการตอบของผู้สอบกับการประมาณค่าพารามิเตอร์ ดังนั้น ถ้ามีจำนวนพารามิเตอร์ที่ต้องการประมาณค่ามาก เช่น สมมติว่ามีข้อสอบ 20 ข้อ และสมมติว่ามีผู้สอบ 200 คน พารามิเตอร์ที่ต้องประมาณค่า 20x200=4000 ตัว ถ้าไม่มี Likelihood เลย คือ Likelihood=0 จะมีค่า AIC = 8000 ถ้า Likelihood เพิ่มขึ้น ค่า AIC จะน้อยลง ดังนั้นเกณฑ์ในการคัดเลือกโมเดล คือ ค่า AIC ที่น้อยกว่า แสดงว่ามีความกลมกลืนระหว่างพฤติกรรมการตอบของผู้สอบกับการประมาณค่าพารามิเตอร์มากกว่า

ประโยชน์ที่คาดว่าจะได้รับ

1. ขยายองค์ความรู้ของการประมาณค่าพารามิเตอร์ผู้สอบ ให้มีความแม่นยำถูกต้องยิ่งขึ้น
2. นำผลการศึกษาไปใช้ในการตัดสินใจเลือกวิธีการประมาณค่าที่เหมาะสมต่อสถานการณ์การนำไปใช้ประโยชน์และเกิดความยุติธรรมสำหรับผู้สอบมากที่สุด
3. เป็นแนวทางในการประยุกต์ใช้กับสถานการณ์ของการทดสอบจริงที่จำเป็นต้องใช้โมเดลการตอบสนองข้อสอบ

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยแบ่งการนำเสนอออกเป็น 5 ตอน คือ **ตอนที่** 1 แนวคิดทฤษฎีเกี่ยวกับโมเดลของทฤษฎีการตอบสนองข้อสอบ **ตอนที่** 2 การประมาณค่าพารามิเตอร์ **ตอนที่** 3 โมเดลการตอบสนองข้อสอบเชิงสุ่ม **ตอนที่** 4 โปรแกรมการจำลองข้อมูลและวิเคราะห์ข้อมูล **ตอนที่** 5 บทความและงานวิจัยที่เกี่ยวข้อง **ตอนที่** 6 กรอบแนวคิดในการวิจัย ดังต่อไปนี้

ตอนที่ 1 แนวคิดทฤษฎีเกี่ยวกับโมเดลของทฤษฎีการตอบสนองข้อสอบ

1.1 ประวัติความเป็นมาของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบมีพัฒนาการมาอย่างต่อเนื่องยาวนาน เริ่มตั้งแต่ ค.ศ. 1904 ที่ Charles Spearman ได้ตีพิมพ์บทความ 2 ชิ้น ที่ใช้ข้อมูลชุดเดียวกันแต่ใช้การวิเคราะห์แตกต่างกัน บทความแรกนำเสนอการประมาณค่าความคลาดเคลื่อนของคะแนนจากแบบสอบ และสรุปว่าคะแนนจากแบบสอบประกอบด้วยคะแนนจริงกับความคลาดเคลื่อนของการวัด ซึ่งเป็นพื้นฐานความคิดของทฤษฎีการทดสอบแบบดั้งเดิม ส่วนบทความชิ้นที่สองมีความเกี่ยวข้องกับทฤษฎีการตอบสนองข้อสอบอย่างมาก เนื่องจากบทความชิ้นนี้ได้เสนอทฤษฎีเชาวน์ปัญญาขึ้นเป็นครั้งแรก เรียกว่า “common factor model” มีองค์ประกอบ จากแนวคิดของ Spearman เกี่ยวกับ common factor model ต่อมาได้พัฒนามาเป็น “item factor analysis” หรือ “latent trait theory” ซึ่งปัจจุบันกลายมาเป็น “item response theory” (McDonald, 1999)

มีการสร้างแบบวัดเชาวน์ปัญญาขึ้นมาใช้ได้สำเร็จครั้งแรก โดย Alfred Binet และ Victor Henri โดยเริ่มแรกที่ปี ค.ศ. 1895 ต่อมามีการปรับปรุงแบบวัดนี้ในปี ค.ศ. 1905 โดย Binet และ Simon (McDonald, 1999) และในปี ค.ศ.1916 Binet และ Simon ได้เขียนกราฟขึ้นเป็นครั้งแรกระหว่างระดับความสามารถกับตัวแปรอิสระซึ่งคืออายุ และใช้กราฟนี้ในการพัฒนาแบบวัด (Hambleton and Swaminathan, 1985, Bock, 1997)

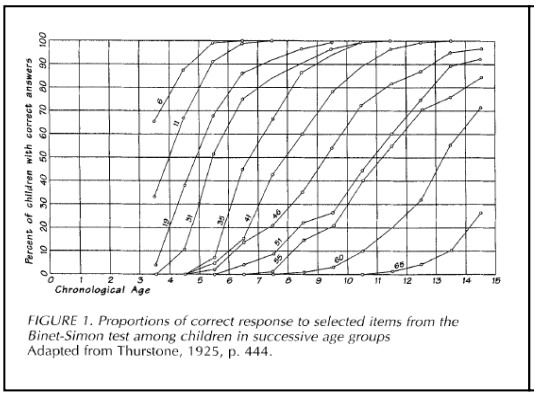


FIGURE 1. Proportions of correct response to selected items from the Binet-Simon test among children in successive age groups
Adapted from Thurstone, 1925, p. 444.

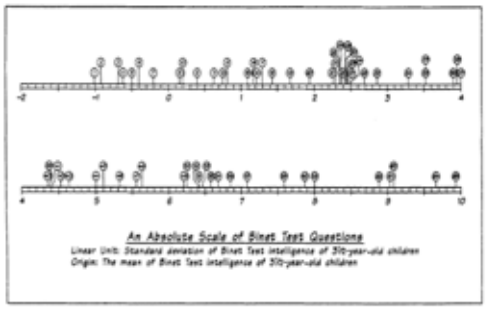


FIGURE 2. Estimated location of items on age scale with arbitrary origin and unit of measurement
Adapted from Thurstone, 1925, p. 449.

ปี ค.ศ.1925 Thurstone ได้วิเคราะห์กราฟที่ได้จากแบบวัดเชาวน์ปัญญาของ Binet และ Simon ข้างต้น ซึ่งการวิเคราะห์นี้มีส่วนเกี่ยวข้องอย่างมากกับลักษณะของ item response theory ทั้งข้อสันนิษฐานว่า โมเดลการตอบสนองที่แสดงความน่าจะเป็นของการตอบได้ถูกต้องว่า เป็นฟังก์ชันของตัวแปรต่อเนื่องของการวัดคุณลักษณะซึ่งคืออายุของผู้ตอบ และโมเดลนี้รวมเอาพารามิเตอร์ที่อธิบายคุณลักษณะของข้อคำถามเอาไว้ด้วยซึ่งคือพารามิเตอร์ตำแหน่ง หรือ ความยากข้อสอบ และพารามิเตอร์คะแนน หรือความสามารถผู้ตอบ และ Thurstone ได้ตีพิมพ์ในบทความเรื่อง " A Method of Scaling Psychological and Educational Tests" เพื่อเสนอวิธีการแก้ปัญหาในการเทียบข้อคำถามจากแบบวัดพัฒนาการทางสมอง (mental development) ของเด็ก ซึ่งเป็นแบบวัดเชาวน์ปัญญาของ Binet และ Simon กับแบบวัด age-graded scale (Bock, 1997) Thurstone จึงได้รับการยกย่องว่าเป็นผู้วางรากฐานของทฤษฎีการทดสอบแนวใหม่ (Modern Test Theory)

ปี ค.ศ. 1936 Richardson สร้างสูตรความสัมพันธ์ระหว่างพารามิเตอร์ของโมเดล IRT กับพารามิเตอร์ข้อสอบแบบดั้งเดิม ซึ่งให้แนวทางเริ่มแรกในการประมาณค่าพารามิเตอร์ของ IRT (Hambleton & Swaminathan, 1985) ปี ค.ศ. 1943, 1944 Lawley สร้างวิธีการใหม่ในการประมาณค่าพารามิเตอร์ (Hambleton & Swaminathan, 1985) ว่าทำอย่างไรจึงจะได้ค่าประมาณความน่าจะเป็นสูงสุดของพารามิเตอร์ของโค้งคุณลักษณะข้อสอบ รวมทั้งให้คำนิยามของคะแนนจริงในเทอมของข้อสอบในแบบสอบ และแสดงให้เห็นว่าสัมประสิทธิ์ความเที่ยงแบบดั้งเดิม (classical reliability coefficient) สามารถแสดงอยู่ในรูปฟังก์ชันของพารามิเตอร์แบบสอบได้ (Baker, 1992)

ปี ค.ศ. 1950 Larzarsfeld เป็นผู้ที่เริ่มใช้คำว่า latent trait และเสนอทฤษฎี latent structure anaylsis แนวคิดหลักส่วนใหญ่ของเขาต่อมาเป็นที่รู้จักกันอย่างกว้างขวาง เรียกว่า item

response theory ; IRT (McDonold, 1999, Baker, 1992) ปี ค.ศ. 1952 Frederic Lord ได้นำ single factor model ของ Spearman มาใช้กับการให้คะแนนข้อสอบแบบ 2 ค่า (binary item scores) เขาได้อธิบายเกี่ยวกับโมเดล normal ogive สำหรับ 2 พารามิเตอร์ สร้างโมเดลการประมาณค่าพารามิเตอร์ และวิธีการนำโมเดลไปใช้ (Hambleton and Swaminathan, 1985, McDonold, 1999) Lord ได้ขยายงานจากของ Lawley ออกไป เขานำเสนอเกี่ยวกับโครงสร้างทฤษฎีการทดสอบแบบดั้งเดิมเพิ่มเติมว่าสามารถแสดงอยู่ในรูปฟังก์ชันพารามิเตอร์ของโค้งคุณลักษณะข้อสอบ (Baker, 1992)

ปี ค.ศ. 1957, 1958 Allen Birnbaum ได้นำสูตรทางคณิตศาสตร์มาช่วยในการนำโมเดลโลจิสติกมาใช้แทนโมเดล normal ogive และพัฒนาหลักการทางสถิติสำหรับโมเดลโลจิสติกที่สร้างขึ้น และปี ค.ศ.1968 ข้อเขียนเกี่ยวกับทฤษฎีของคุณลักษณะแฝง (the theory of latent traits) ของเขา ซึ่งรวบรวมไว้ในหนังสือ “Statistical Theories of Mental Test Scores” ของ Lord และ Novick จำนวน 4 บท จึงทำให้ IRT เป็นที่รู้จักกันอย่างแพร่หลาย (Hambleton and Swaminathan, 1985, Baker, 1992)

ปี ค.ศ. 1960 George Rasch นักคณิตศาตร์ชาวเดนมาร์ก พัฒนาโมเดลตอบสนองข้อสอบขึ้น 3 โมเดล โมเดลที่เป็นที่รู้จักของเขาคือ Rasch model ซึ่งเป็นโมเดล 1 พารามิเตอร์ และตีพิมพ์ไว้ในหนังสือ “Probabilistic Models for Some Intelligence and Attainment Tests” ผลงานของเขามีอิทธิพลต่อ Benjamin Wright ชาวอเมริกัน และ Andersen กับ Fischer ซึ่งเป็นนักจิตวิทยาชาวยุโรป

ปี ค.ศ. 1967 Benjamin Wright เป็นผู้นำและมีบทบาทในการกระตุ้นให้มีการวิจัยจำนวนมากเกี่ยวกับโมเดลราสช์ (the Rasch model) ในอเมริกาตลอดทศวรรษ 1960 และจากการนำเสนอของ Wright ในการประชุมสัมมนาของ ETS เรื่องปัญหาในการทดสอบ จึงเป็นการกระตุ้นงานด้าน IRT ที่สำคัญ โดยเฉพาะอย่างยิ่งกับโมเดลราสช์ (Hambleton and Swaminathan, 1985)

ปี ค.ศ. 1968 Lord และ Novick ตีพิมพ์หนังสือ “Statistical Theories of Mental Test Scores” ของ Lord และ Novick ซึ่งรวบรวมข้อเขียนเกี่ยวกับทฤษฎีของคุณลักษณะแฝง (the theory of latent traits จำนวน 5 บท (4 บท เป็นข้อเขียนของ Birnbaum) จึงกระตุ้นให้เกิดความสนใจในการทำวิจัยเกี่ยวกับ IRT เป็นจำนวนมาก (Hambleton and Swaminathan, 1985, Baker, 1992)

ปี ค.ศ. 1969 Wright และ Panchapakesan อธิบายวิธีการประมาณค่าพารามิเตอร์สำหรับโมเดลราสช์และสร้างโปรแกรมคอมพิวเตอร์ BICAL ที่นำมาใช้วิเคราะห์ข้อสอบจากโมเดลราสช์

ปี ค.ศ. 1969 Samejima ตีพิมพ์ชุดชิ้นงานที่อธิบายถึงโมเดลตอบสนองข้อสอบและวิธีการนำไปใช้ โมเดลของ Samejima ครอบคลุมทั้งข้อมูลการตอบสนองแบบหลายค่าและแบบต่อเนื่อง (polytomous and continuous response data) และขยายจากโมเดลมิติเดียว (unidimensional models) ไปสู่โมเดลหลายมิติ (multidimensional models)

ปี ค.ศ. 1970 Bock และ Lieberman และ ปี ค.ศ. 1972 Bock ได้ขยายแนวคิดใหม่ในการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบไปพร้อมๆกันด้วยกระบวนการ marginalization ปี ค.ศ. 1974 Lord อธิบายวิธีการประมาณค่าพารามิเตอร์แบบใหม่ที่นำไปใช้กับโปรแกรมคอมพิวเตอร์ เรียกว่า LOGIST

ปี ค.ศ. 1974 Fischer ให้คำอธิบายเกี่ยวกับโปรแกรมการวิจัยของโมเดลโลจิสติกเชิงเส้นตรง (linear logistic models) ปี ค.ศ. 1976 Lord และคณะ สร้างโปรแกรมคอมพิวเตอร์ ชื่อ LOGIST เพื่อใช้ในการประมาณค่าพารามิเตอร์ของโมเดลโลจิสติก ซึ่งเป็นหนึ่งในสองโปรแกรมที่ใช้กันมากในปัจจุบัน (อีกโปรแกรมหนึ่งคือ BICAL) ปี ค.ศ. 1977 Baker เสนอวิธีการประมาณค่าพารามิเตอร์ต่างๆ เข้าไว้ด้วยกัน

ปี ค.ศ. 1977 นักวิจัยหลายท่าน เช่น Bashaw, Lord, Marco, Rentz, Urry และ Wright ได้อธิบายความก้าวหน้าของการวัดผลที่สำคัญจำนวนมากที่เกี่ยวกับการนำ IRT ไปใช้ ในวารสารชื่อ “Journal of Educational Measurement” ปี ค.ศ. 1979 Wright and Stone เขียนบทความอธิบายทฤษฎีและการประยุกต์ใช้ที่เน้นโมเดลราสช์ไว้ในหนังสือ Best Test Design ปี ค.ศ. 1980 Lord เขียนเกี่ยวกับพัฒนาการใหม่และการนำโมเดล 3 พารามิเตอร์ไปใช้ ในหนังสือ Applications of Item Response Theory to Practical Testing Problems ปี ค.ศ. 1980 Weiss เรียบเรียงหนังสือ Proceeding of the 1979 Computerized Adaptive Testing Conference ซึ่งรวบรวมบทความใหม่ๆของการทดสอบแบบปรับเหมาะ ปี ค.ศ. 1982 Lord และคณะ ร่วมกันปรับปรุงโปรแกรม LOGIST ทำให้ทำงานได้เร็วขึ้น ติดตั้งได้ง่ายขึ้น และเพิ่มเติมผลการวิเคราะห์ที่มีประโยชน์ยิ่งขึ้นกว่าโปรแกรมในเวอร์ชั่นปี 1976 (Edwards, 2008; Hambleton, Swaminathan and Rogers, 1991; Hambleton and Swaminathan, 1985)

ในยุคแรกมีการพัฒนาองค์ความรู้ของ IRT ไปในการทดสอบที่มีการตอบแบบให้คะแนน 2 ค่า หรือแบบทวิวิภาค (binary response) ทั้งในทางแนวคิด ทฤษฎีและประยุกต์ใช้กับการตอบ

ข้อสอบในการทดสอบทั่วไป รวมทั้งพัฒนาโปรแกรมคอมพิวเตอร์เพื่อช่วยวิเคราะห์ข้อสอบและประมาณค่าความสามารถของผู้สอบ ดังที่กล่าวข้างต้น

ในยุคต่อมามีการพัฒนาองค์ความรู้ของ IRT ขยายไปสู่รูปแบบการตอบประเภทต่าง ๆ เช่นการตอบที่ให้คะแนนมากกว่า 2 ค่า (Polytomous) การตอบที่มีตัวเลือกแบบมาตรประมาณค่า (rating scale) (Andrich, 1978) การตอบที่ตรวจให้คะแนนบางส่วน (Partial Credit) (Masters, 1982) การให้คะแนนแบบหลายกลุ่ม (multiple category scoring) (Thrissen และ Steinberg, 1984) การตรวจให้คะแนนตามลำดับขั้นของรายการหลายคำตอบ (Ordered-response categories) อีกทั้งมีการสร้างโปรแกรมคอมพิวเตอร์เพื่อช่วยในการวิเคราะห์หลายโปรแกรม เช่น โปรแกรม MULTILOG โดย Thissen ค.ศ.1986 โปรแกรม NOHARM โดย Fraser และ McDonald ค.ศ.1988 โปรแกรม PARSCALE โดย Muraki และ Bock ค.ศ.1993 โปรแกรม BIGSTEPS โดย Linacre และ Wright ค.ศ.1994 และโปรแกรม RUMN โดย Sheridan, Andrich และ Luo ค.ศ.1996 เป็นต้น (Edwards, 2008; ศิริชัย กาญจนวาสี, 2555) ซึ่งมีส่วนผลักดันให้ทฤษฎีการทดสอบเป็นกระแสหลักในการวัดผลทางการศึกษาและทางจิตวิทยา รวมทั้งนำไปสู่การพัฒนาเครื่องมือ ระบบการทดสอบอย่างกว้างขวางมากในปัจจุบัน

1.2 ทฤษฎีการตอบสนองข้อสอบ

1.2.1 แนวคิดพื้นฐานและวิธีการ

โมเดลการวัดที่เป็นแนวคิดพื้นฐาน คือโมเดลเชิงคณิตศาสตร์ (Mathematical Model) ที่รวมเอาชุดของข้อตกลงเบื้องต้นเกี่ยวกับข้อมูลที่นำมาใช้กับโมเดล และความสัมพันธ์เฉพาะระหว่างโครงสร้างที่สังเกตได้และโครงสร้างที่สังเกตไม่ได้ ที่รู้จักกันดีในโมเดลการทดสอบแบบดั้งเดิม (Classical Test Model) (Hambleton, & Jones, 1993)

$$X_p = T_p + E_p$$

เมื่อ X_p คือ คะแนนที่สังเกตได้ T_p คือคะแนนจริงของผู้สอบ และ E_p คือ คะแนนความคลาดเคลื่อน โดยมีข้อตกลงเบื้องต้น คือ คะแนนความคลาดเคลื่อนไม่มีความสัมพันธ์กับคะแนนจริงของผู้สอบ และคะแนนความคลาดเคลื่อนที่ได้จากการสอบแต่ละครั้งไม่มีความสัมพันธ์กัน (Lord, 1980) แต่โมเดลการทดสอบแบบดั้งเดิมนี้มีข้อจำกัดคือ คะแนนความคลาดเคลื่อนมาตรฐาน และค่าพารามิเตอร์ของข้อสอบและแบบสอบขึ้นอยู่กับกลุ่มผู้สอบ กล่าวคือเมื่อเปลี่ยน

กลุ่มผู้สอบจะทำให้ค่าพารามิเตอร์ของข้อสอบและแบบสอบเปลี่ยนไป คะแนนรายข้อจึงไม่เชื่อมโยงกับพฤติกรรมการตอบและคะแนนจริงของผู้สอบ (Fan, 1998)

ด้วยข้อจำกัดของทฤษฎีการทดสอบแบบดั้งเดิมดังกล่าว นักวัดผลการการศึกษาจึงพัฒนาวิธีการวัดแนวใหม่ขึ้น ที่จะช่วยขยายแนวคิดเกี่ยวกับข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบดั้งเดิมให้สมเหตุสมผลเป็นที่ยอมรับมากขึ้น ซึ่งริเริ่มโดยเทอร์สโตน ซึ่งพัฒนาเทคนิคการวิเคราะห์ตัวประกอบเพื่อศึกษาคุณลักษณะทางจิตวิทยาและเป็นรากฐานสำคัญของทฤษฎีการทดสอบแนวใหม่ 2 ทฤษฎี คือ (1) ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Theory ; G-Theory) ที่พัฒนาขึ้นมาโดยครอนบาคและคณะ เพื่อศึกษาความเที่ยงทั่วไปของแบบสอบภายใต้เงื่อนไขต่างๆของการทดสอบ และ (2) ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory ; IRT) ที่นักทฤษฎีการทดสอบยุคใหม่หลายท่านมีส่วนร่วมในการพัฒนาขึ้น เพื่อศึกษาความคลาดเคลื่อนจากการวัดที่มีค่าแตกต่างกันไปตามลักษณะของข้อสอบและแบบสอบโดยขึ้นกับความสามารถของผู้สอบและคุณลักษณะของข้อสอบแต่ละข้อ เพื่อที่จะวัดคุณลักษณะภายในที่เป็นความสามารถที่แท้จริงของผู้สอบแต่ละคน (Hambleton & Swaminathan, 1985)

ในปี ค.ศ.1997 Cooke และ Michie ได้ให้ความเห็นว่าทฤษฎีตอบสนองข้อสอบมีประโยชน์ที่มากขึ้นไปจากทฤษฎีการทดสอบแบบดั้งเดิม โดยมีวิธีการวิเคราะห์ข้อสอบและแบบสอบที่แตกต่างออกไปจากวิธีการของทฤษฎีการทดสอบแบบดั้งเดิม อีกทั้งยังช่วยเสริมให้มีความสมบูรณ์ในการวิเคราะห์ยิ่งขึ้น โมเดลทฤษฎีการตอบสนองข้อสอบใช้ฟังก์ชันทางคณิตศาสตร์ที่ระบุถึงความสัมพันธ์ระหว่างการตอบสนองข้อสอบกับคุณลักษณะแฝงหรือความสามารถผู้สอบที่แสดงออกมาด้วยผลการตอบข้อสอบ (Hare, 1999)

โมเดลทฤษฎีการตอบสนองข้อสอบได้เสนอขึ้นตั้งแต่ ค.ศ. 1943 โดย Lawley ในระยะนั้นมีเพียงการเสนอแนวคิดและหลักการ แต่ไม่มีการนำโมเดลไปใช้ในทางปฏิบัติ ต่อมาในช่วง ค.ศ. 1970 จึงเริ่มมีการประยุกต์ใช้ทฤษฎีการตอบสนองรายข้อ ทำให้มีการพัฒนาของวิธีวิทยาการด้านการวัดขึ้นอย่างรวดเร็ว ทั้งในเรื่องการปรับเทียบข้อสอบ (test equating) การทำหน้าที่ต่างกัน (differential item functioning) หรือความลำเอียงของข้อสอบ (test bias) การบริหารการสอบด้วยคอมพิวเตอร์ (computerized test administration) หรือการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (Adaptive Testing) การสร้างมาตรวัด และการหาปกติวิสัย (scaling and norming) ที่ได้รับการพัฒนาใหม่ ให้มีประสิทธิภาพดีมากขึ้นเรื่อยๆ (Linn, R. L., 1989)

ลักษณะของโมเดลการตอบสนองรายข้อที่ได้รับการพัฒนาที่ผ่านมา นอกจากจะมีโมเดลโลจิสติกแบบเอกมิติหนึ่ง สอง สาม และสี่พารามิเตอร์ และโมเดลโอไจว์ปกติซึ่งใช้กับผลสัมฤทธิ์ทางการเรียน และความถนัดที่เป็นข้อมูลทวิภาค (Hambleton, R.K., 1989) แล้ว ยังมีโมเดลอีกหลายแบบที่นักวัดผลการศึกษาได้พัฒนาจากโมเดลการตอบสนองรายข้อ โดยสรุปดังนี้

(1) Nominal Response Model ; NRM เป็นโมเดลที่พัฒนาโดย Bock สำหรับใช้กับข้อมูลแบบพหุวิภาค เพื่อประมาณค่าความสามารถของผู้สอบให้มีความถูกต้องสูงสุด โดยใช้สารสนเทศจากโค้งลักษณะการสอบแต่ละตัวเลือก โดยมีเงื่อนไขว่าผลรวมความน่าจะเป็นในการตอบทุกตัวเลือกของผู้สอบที่ระดับความสามารถที่กำหนดให้มีค่าเท่ากับหนึ่ง หลังจากนั้น Thissen พัฒนาแนวคิดนี้ต่อไปโดยกำหนดน้ำหนักสำหรับตัวเลือกแต่ละตัว เพื่อการประมาณค่าความสามารถผู้สอบที่มีระดับความสามารถต่ำได้ถูกต้องมากยิ่งขึ้น

(2) Graded- Response Model ; GRM หรือ Difference Model เป็นโมเดลที่พัฒนาโดย Samejima นำมาใช้กับมาตรวัดแบบประมาณค่า เช่น มาตรวัดแบบ Likert และมีการใช้สารสนเทศจากโค้งลักษณะการตอบแต่ละตัวเลือก การเรียกชื่อโมเดลนี้ว่า difference models เพราะการคำนวณความน่าจะเป็นของผู้สอบที่ตอบตัวเลือกที่ K ได้ถูกต้องนั้นต้องคำนวณจากผลต่างระหว่างความน่าจะเป็นจากตัวเลือกที่ I และตัวเลือกที่ k-1 นั่นเอง การพัฒนาต่อมาของ GRM โดย Andrich, D. (1995) คือ multiplicative Poisson Model ; MPM เพื่อแก้ปัญหาในกรณีที่ไม่สามารถใช้ unidimensional Rasch Model ; URM ได้ เนื่องด้วยมีข้อจำกัดเมื่อมีการรวมตัวเลือกการตอบสนองของแบบสอบ

(3) Binomial Trials Model และ Rating Scale Model ; RSM เป็นโมเดลที่พัฒนาโดย Andrich เพื่อใช้ในการวัดตัวแปรพหุวิภาค (polytomous variable) และมีการพัฒนารวมทั้งการนำไปประยุกต์ใช้อย่างต่อเนื่อง Fischer, G.H, และ Parzer, P. (1991) ได้ประยุกต์ RSM มาเป็นโมเดลที่พัฒนาขึ้นใหม่เรียกว่า linear rating scale model (LRSM) ที่พารามิเตอร์ข้อคำถามจะอยู่ในรูปของฟังก์ชั่นเชิงเส้นของพารามิเตอร์ อื่นๆ โมเดลนี้นำมาใช้ในการวัดคะแนนความเปลี่ยนแปลง (change) ซึ่งทฤษฎีการวัดแบบดั้งเดิมใช้ไม่ได้ผลดีเท่าที่ควร เพราะมีปัญหาเนื่องจากอิทธิพลของเพดานและพื้น (ceiling and floor effects) และลักษณะการแจกแจงเบ้ นอกจากนี้มีการพัฒนาวิธีการสร้างมาตร (scaling methods) เรียกว่า dual scaling โดย Nishisato และ Cheung, K.C. กับ Mooi, L.C. เป็นวิธีที่สามารถใช้กับมาตรวัดแบบ Likert เช่นเดียวกับ RSM ทั้งสองวิธีการนี้ได้ผลคล้ายคลึงกัน แต่ dual scaling ผ่อนคลายข้อตกลงเบื้องต้นได้มากกว่า RSM

(4) Partial Credit Model ; PCM พัฒนาโดย Masters เพื่อนำมาใช้กับแบบสอบที่มีหลายตัวเลือก และการให้คะแนนตัวเลือกแต่ละตัวแตกต่างกัน ต่อมาได้พัฒนาโมเดลให้ดีขึ้นโดย Wright และ Masters (1984) นักวัดผลการศึกษาทั้งสองยังได้ประมวลโมเดลโลจิสติกหนึ่งพารามิเตอร์ หรือ Rasch model แบบต่างๆที่มีการพัฒนาขึ้นรวมเรียกว่า โมเดลการวัดขั้นพื้นฐาน (fundamental measurement model) ซึ่งมีคุณสมบัติสำคัญคือ การแยกพารามิเตอร์ผู้สอบ และพารามิเตอร์ข้อคำถาม และมีการใช้สถิติที่เพียงพอในการประมาณค่าพารามิเตอร์ของโมเดล หลังจากนั้น Muraki, E. (1993) ได้พัฒนาโมเดล generalized partial credit model ; GPCM ซึ่งรวมค่าพารามิเตอร์ความชัน (slope) ของโค้งลักษณะข้อคำถามแต่ละข้อไว้ในโมเดล

(5) Linear Logistic Latent Trait Model พัฒนาโดย Fischer และ Formann (1972) กับ Multicomponent Latent Trait Model พัฒนาโดย Embretson (1984) จัดว่าเป็นโมเดลที่รวมองค์ประกอบด้านปัญญา (cognitive component) เข้าเป็นพารามิเตอร์ในโมเดลด้วย นอกจากนี้มี linear logistic model ซึ่ง Fischer พัฒนาโดยกำหนดให้มีพารามิเตอร์ความยากของข้อสอบในรูปของผลรวมเชิงเส้นของ องค์ประกอบที่มีอิทธิพลต่อค่าความยากอยู่ในโมเดลด้วย ผลงานของ Embretson และ Fischer นับว่าเป็นจุดเริ่มต้นของการผสมผสานแนวความคิดทาง cognitive psychology กับ psychometric เข้าด้วยกัน โมเดลของ Fischer ใช้ในการวัดตัวแปรเอกมิติแต่โมเดลของ Embretson สามารถใช้กับปัญญาทักษะ (cognitive skill) หลายองค์ประกอบได้โดยที่แต่ละองค์ประกอบมีความสำคัญต่อการที่ผู้สอบจะตอบข้อ คำถามแต่ละข้อได้ถูกต้อง

(6) Unfolding Models พัฒนาโดย Andrich และ Hyperbolic Cosine Unfolding Model พัฒนาโดย Andrich, D. และ Luo, G. (1993) ซึ่งมีการพัฒนาควบคู่กับ probabilistic models for the cumulative models in pair comparison design ตามแนวคิดของ Thurstone โมเดลในกลุ่มนี้ ได้แก่ Squared Simple Logistic Model ; SSLM พัฒนาโดย Andrich ซึ่งให้ผลการวิเคราะห์เทียบเท่ากับ Simple Hyperbolic Cosine Model ; SHCM ของ Andrich และ Lou และ two-parameter hyperbolic cosine model ; 2PHCM เป็นต้น

(7) โมเดลที่พัฒนาจากทฤษฎีการตอบสนองข้อสอบเพื่อใช้ในสภาพการณ์ที่ต่างจากการสอบปกติ ได้แก่ โมเดลที่พัฒนาขึ้นเพื่อใช้ในการสอบที่มีหน่วยการวิเคราะห์เป็นกลุ่ม เช่น ชั้นเรียน โรงเรียน จังหวัด เขตการศึกษา โดย Bock, R.D., Mislevy, R. Woodson, C. (1982) นอกจากนี้ยังมี random coefficients multinomial logit model ; RCMLM ที่พัฒนาจาก Rasch model โดย Wilson, M. (1995) เพื่อใช้กับแบบสอบที่มีข้อสอบรวมเป็นชุดข้อสอบหลายชุด (item bundles) ซึ่ง

แต่ละชุดเกี่ยวข้องกับคุณลักษณะของผู้สอบ และมีส่วนที่เชื่อมโยงกันระหว่างชุดข้อสอบ เช่น โครงสร้างข้อสอบร่วมกัน ตัวคำถามร่วมกัน หรือเนื้อหาข้อสอบร่วมกัน

โมเดลทฤษฎีการตอบสนองข้อสอบมีประโยชน์ต่อการวัดผลการศึกษา เมื่อข้อมูลสอดคล้องกับข้อตกลงเบื้องต้นของโมเดลจะสามารถประมาณค่าความสามารถของผู้สอบได้ โดยที่พารามิเตอร์ความสามารถเป็นอิสระไม่ขึ้นกับการเปลี่ยนแปลงของข้อสอบ ทำให้ได้ค่าพารามิเตอร์ของข้อสอบที่ไม่ขึ้นกับกลุ่มตัวอย่างผู้สอบ ค่าสถิติที่ได้แสดงถึงความถูกต้องในการประมาณค่าความสามารถผู้สอบที่ขึ้นอยู่กับค่าความสามารถผู้สอบ จำนวนและคุณสมบัติทางสถิติของข้อคำถาม และได้มาตรร่วม (common scale) ซึ่งใช้บรรยายคุณสมบัติผู้สอบและข้อสอบได้

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างลักษณะหรือความสามารถที่มีอยู่ภายในตัวบุคคล (Latent trait or Ability) กับผลการตอบสนองข้อสอบหรือข้อคำถามของบุคคลนั้น ทฤษฎีนี้มีความเชื่อว่าผลการตอบสนองข้อสอบของผู้สอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรงสามารถทำนายหรืออธิบายคุณลักษณะ (Latent trait) หรือความสามารถ (ability) ที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้โดยตรง โดยใช้ฟังก์ชันการตอบสนองข้อสอบ ซึ่งเป็นฟังก์ชันคณิตศาสตร์หรือโมเดลในการอธิบายความสัมพันธ์ระหว่างระดับความสามารถคุณลักษณะของข้อสอบและโอกาสของการตอบข้อสอบได้ถูก ที่มีลักษณะความสัมพันธ์เป็นแบบฟังก์ชันโลจิสหรือฟังก์ชันปกติสะสม ที่เรียกว่า ฟังก์ชันลักษณะข้อสอบ (Item Characteristic Function) หรือ โค้งคุณลักษณะของข้อสอบ (Item Response Function) (ศิริชัย กาญจนวาสี, 2555) ประกอบด้วยค่าพารามิเตอร์ต่างๆของข้อสอบ คือ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และโอกาสการเดาข้อสอบถูก (c) ของข้อสอบ แต่ละโมเดลนี้จะมีจำนวนพารามิเตอร์ของข้อสอบแตกต่างกันไป กล่าวคือ แบบหนึ่งพารามิเตอร์จะแสดงเฉพาะพารามิเตอร์ความยาก (difficulty) แบบสองพารามิเตอร์จะเพิ่มพารามิเตอร์อำนาจจำแนก (discrimination) อีกตัวหนึ่ง และแบบสามพารามิเตอร์จะประกอบด้วยค่าความยาก อำนาจจำแนก และค่าการเดา สำหรับการเลือกโมเดลของทฤษฎีการตอบสนองข้อสอบมาใช้นั้น จะต้องเลือกให้เหมาะสมกับสถานการณ์ต่างๆ ตามข้อตกลงเบื้องต้นของแต่ละโมเดล จุดเด่นของทฤษฎีตอบสนองข้อสอบมีหลายประการ ดังนี้ (Hambleton and Swaminathan, 1985)

ประการแรก ค่าพารามิเตอร์ของข้อสอบเป็นอิสระจากกลุ่มผู้สอบที่ใช้ในการประมาณค่า กล่าวคือ ค่าพารามิเตอร์ของข้อสอบแต่ละข้อ ได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) เป็นค่าคงที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ แม้ว่าจะเปลี่ยนกลุ่มผู้สอบแต่

ค่าพารามิเตอร์ของข้อสอบจะมีค่าคงเดิมเสมอ และในทำนองเดียวกันค่าพารามิเตอร์ความสามารถของผู้สอบเป็นอิสระจากข้อสอบที่ใช้ในการประมาณค่า จะไม่แปรเปลี่ยนตามค่าพารามิเตอร์ของข้อสอบ ไม่ว่าผู้สอบจะทำข้อสอบข้อใด หรือได้ทำข้อสอบเมื่อใดก็ตาม

ประการที่สอง การเปรียบเทียบความสามารถของผู้สอบ จะไม่ขึ้นอยู่กับคำถามในแบบสอบ ถึงแม้ว่าผู้สอบจะใช้ข้อคำถามต่างกัน ก็สามารถนำค่าความสามารถมาเปรียบเทียบกันได้ ทั้งนี้เนื่องจากค่าความสามารถที่ประมาณค่ามาได้นั้น เป็นคะแนนโลจิท (Logit) ซึ่งอยู่ในมาตรวัดเดียวกัน

ประการที่สาม การรายงานคุณภาพของข้อสอบในรูปค่าสารสนเทศ (information) สามารถรายงานได้ทั้งเป็นรายข้อและทั้งฉบับ ค่าสารสนเทศเป็นตัวบ่งชี้ถึงความถูกต้องแม่นยำ (accuracy) ในการประมาณค่าความสามารถ สามารถนำมาใช้แทนค่าความเที่ยง และค่าความคลาดเคลื่อนมาตรฐานในการวัดได้

นอกจากนี้ Cooke และ Michie, 1997; Embretson, 1996; Nunnally & Bernstein, 1994 และ Steinberg & Thissen, 1996 (Hare, 1999) ที่กล่าวถึงข้อดีของทฤษฎีการตอบสนองข้อสอบในทำนองเดียวกันไว้ 5 ประการ **ประการแรก** พารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบมีความไม่แปรเปลี่ยนไปตามกลุ่มของผู้สอบ พารามิเตอร์ข้อสอบเป็นอิสระจากการทดสอบอย่างแท้จริง ในอีกแง่หนึ่งความตรงและความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม เช่น ค่าความสัมพันธ์ระหว่างข้อสอบที่ตอบถูกกับคะแนนรวม หรือค่าสัมประสิทธิ์อัลฟ่า จะมีความไวต่อความแปรปรวนระหว่างกลุ่มผู้สอบโดยขึ้นอยู่กับชุดของแบบสอบที่ใช้หรือช่วงคะแนนที่สังเกตได้จากแบบสอบ **ประการที่สอง** การวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบมีความถูกต้องแม่นยำในการวัดทุกๆค่าของความสามารถ ในขณะที่ทฤษฎีการทดสอบแบบดั้งเดิมให้เพียงแค่ค่าประมาณที่ถูกต้องแม่นยำของการวัดระดับคุณลักษณะของผู้สอบ แต่ทฤษฎีการตอบสนองข้อสอบได้ให้ข้อมูลที่ละเอียดกว่านั้น ทั้งสารสนเทศของข้อสอบและสารสนเทศของแบบสอบที่มีความถูกต้องแม่นยำในทุกช่วงของความสามารถผู้สอบ **ประการที่สาม** การวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบสามารถบอกค่าอำนาจจำแนกของข้อสอบแต่ละข้อและคะแนนการสอบโดยภาพรวมในทุกๆ ช่วงของความสามารถ ถ้าใช้ทฤษฎีการตอบสนองข้อสอบวิเคราะห์แบบสอบวินิจฉัย จะทำให้สามารถเลือกข้อสอบที่ให้สารสนเทศสูงสุดที่ใกล้เคียงกับจุดตัดในการวินิจฉัยนั้น แต่ถ้าใช้ทฤษฎีการตอบสนองข้อสอบวิเคราะห์แบบสอบที่ใช้วัดความเข้มของคุณลักษณะหรือความสามารถ จะบ่งบอกได้ว่าข้อสอบเหล่านั้นเพียงพอที่จะให้สารสนเทศในทุกระดับของความสามารถหรือไม่ **ประการที่สี่** การวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบยินยอมให้

เปรียบเทียบได้โดยตรงระหว่างข้อสอบคู่ขนานที่วัดคุณลักษณะเดียวกันในกลุ่มตัวอย่างเดียวกัน เช่น แบบวัดต้นฉบับกับแบบวัดที่แปลเป็นภาษาอื่น หรือแบบวัดฉบับเต็มกับแบบวัดฉบับย่อ ซึ่งทฤษฎีการตอบสนองข้อสอบเหมาะสมอย่างยิ่งในการใช้หาคะแนนข้อสอบและคะแนนแบบสอบของแบบสอบคู่ขนาน **ประการที่ห้า** วิธีการของทฤษฎีการตอบสนองข้อสอบ แม้ว่าจะวัดด้วยเครื่องมือ 2 แบบที่สามารถเปรียบเทียบกันได้ก็ตาม พบว่า ความสามารถของผู้สอบไม่แปรเปลี่ยนไปตามแบบสอบหรือเครื่องมือวัดความสามารถ

จากคุณลักษณะดังกล่าว จึงทำให้นักวัดผลนำทฤษฎีการตอบสนองข้อสอบไปประยุกต์ใช้ในการวัดผลทางการศึกษาและจิตวิทยาในหลายเรื่องด้วยกัน เช่น การสร้างคลังข้อสอบ (Item Banking) การพัฒนาแบบทดสอบ (Test Development) การหาความลำเอียงของข้อสอบ (Item Bias) การทดสอบอิงเกณฑ์ (Criterion-Reference Testing) การปรับเทียบคะแนนการสอบ (Test Score Equating) การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (Adaptive Testing) การสร้างมาตรวัด และการหาปกติวิสัย (Scaling and Norming)

1) คุณลักษณะของโมเดลการตอบสนองข้อสอบ

โมเดลการตอบสนองข้อสอบเป็นการอธิบายความสัมพันธ์ระหว่างโอกาสของการตอบข้อสอบได้ถูก (P_i) กับความสามารถของผู้ตอบข้อสอบ (θ) โดยมีฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อประกอบด้วยพารามิเตอร์แบบ 1, 2, 3 พารามิเตอร์และค่าคงที่ มีโค้งลักษณะข้อสอบ (Item Characteristic Curve; ICC) ที่เขียนอยู่ในรูปของฟังก์ชันปกติสะสมหรือฟังก์ชันโลจิส ดังตารางข้างล่างนี้

ตารางที่ 2.1 ฟังก์ชันทางคณิตศาสตร์ของโมเดลการตอบสนองข้อสอบ (ศิริชัย กาญจนวาสี, 2555)

โมเดล	ฟังก์ชันปกติสะสม (Normal Ogive Function)	ฟังก์ชันโลจิส (Logistic Function)
1 พารามิเตอร์	$P_i(\theta)=\int_{-\infty}^{\theta-b_i}\frac{1}{\sqrt{2\pi}}e^{-Z^{\frac{2}{2}}}dz$	$P_i(\theta)=\frac{1}{1+e^{-D(\theta-b_i)}}$
2 พารามิเตอร์	$P_i(\theta)=\int_{-\infty}^{a_i(\theta-b_i)}\frac{1}{\sqrt{2\pi}}e^{-Z^{\frac{2}{2}}}dz$	$P_i(\theta)=\frac{1}{1+e^{-Da_i(\theta-b_i)}}$
3 พารามิเตอร์	$P_i(\theta)=c_i+(1+c_i)\int_{-\infty}^{a_i(\theta-b_i)}\frac{1}{\sqrt{2\pi}}e^{-Z^{\frac{2}{2}}}dz$	$P_i(\theta)=c_i+\frac{1-c_i}{1+e^{-Da_i(\theta-b_i)}}$

โมเดลการตอบสนองข้อสอบประกอบด้วยค่าพารามิเตอร์และค่าคงที่ดังนี้

(1) พารามิเตอร์ของผู้สอบ (θ) คือระดับความสามารถของผู้สอบ ซึ่งประมาณค่าได้จากโมเดลตามทฤษฎีการตอบสนองข้อสอบ โดยปรับให้เป็นคะแนนมาตรฐานที่มีค่าเฉลี่ยเป็น 0 กล่าวคือผู้ที่มีความสามารถระดับปานกลาง และมีส่วนเบี่ยงเบนมาตรฐานเป็น 1 ค่าพารามิเตอร์ของผู้สอบนี้มีพิสัยอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่การวิเคราะห์ส่วนใหญ่จะมีค่าอยู่ในช่วง -3 ถึง +3 โดยค่า $P_i(\theta)$ คือ ค่าความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

(2) พารามิเตอร์ของข้อสอบ ประกอบด้วย ค่าความยาก(b_i) ค่าอำนาจจำแนก(a_i) และโอกาสในการเดาข้อสอบได้ถูก (c_i)

พารามิเตอร์ความยาก (b_i) คือ ความยากของข้อสอบข้อที่ i (Difficulty parameter) ซึ่งอยู่ในตำแหน่งของโค้ง ICC บนสเกลความสามารถ (θ) ที่ทำให้โอกาสของการตอบข้อสอบได้ถูกต้องสำหรับโมเดล 1-พารามิเตอร์กับ 2-พารามิเตอร์เท่ากับ 0.50 และโอกาสของการตอบข้อสอบได้ถูกต้องสำหรับโมเดล 3-พารามิเตอร์ เท่ากับ $\frac{1+c_i}{2}$ ค่าพารามิเตอร์ความยากนี้มีค่าระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัตินิยมใช้ค่าที่อยู่ระหว่าง -2.50 ถึง +2.50 โดยค่าที่อยู่ใกล้ -2.50 เป็นข้อสอบที่ง่าย ค่าที่อยู่ใกล้ +2.50 เป็นข้อสอบที่ยาก

พารามิเตอร์อำนาจจำแนก (a_i) คือ ความสามารถในการจำแนกความแตกต่างของโอกาสในการตอบข้อสอบข้อที่ i ได้ถูกต้องระหว่างผู้สอบที่มีความสามารถน้อยกว่าหรือเท่ากับ θ กับผู้สอบที่มีความสามารถมากกว่า θ พารามิเตอร์อำนาจจำแนกนี้มีค่าเป็นสัดส่วนโดยตรงกับค่าความชันของโค้ง ICC ที่ตำแหน่งของค่าความยาก (b_i) ถ้าค่าอำนาจจำแนกสูง แสดงถึงการจำแนกผู้สอบที่มีความสามารถแตกต่างกันได้ดี ค่าอำนาจจำแนกมีค่าระหว่าง $-\infty$ ถึง $+\infty$ ควรมีค่าเป็นบวกแต่ไม่เกิน +2.50 ในทางปฏิบัตินิยมใช้ค่าที่อยู่ระหว่าง +0.50 ถึง +2.50

พารามิเตอร์โอกาสในการเดาข้อสอบได้ถูก (c_i) คือ โอกาสที่ผู้ที่มีความสามารถต่ำสามารถตอบข้อสอบข้อที่ i ได้ถูก เป็นค่ากำกับต่ำสุด (lower assymtote) ของโค้ง ICC ซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1 แต่โดยทั่วไปนิยมใช้ข้อสอบที่มีค่าโอกาสในการเดาข้อสอบได้ถูก ไม่เกิน 0.30 และควรมีค่าต่ำกว่าโอกาสในการตอบถูกโดยการเดาตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT)

(3) ค่าคงที่ ได้แก่

e เป็นค่าคงที่ของลอการิทึมธรรมชาติ (Natural log) มีค่าเท่ากับ 2.71828

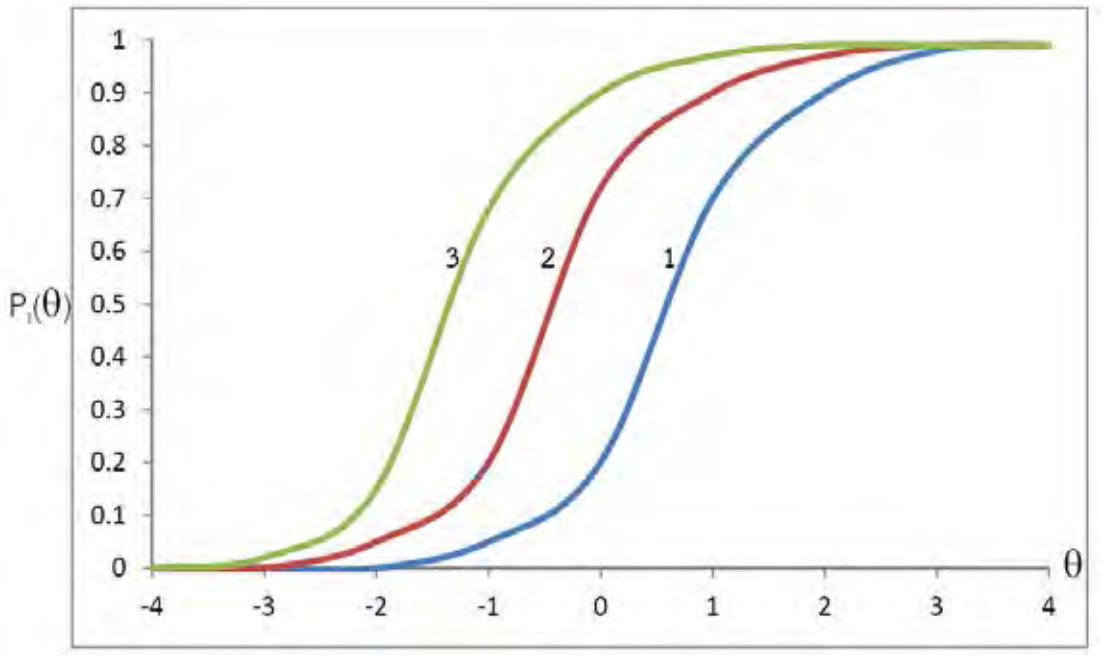
D เป็นค่าองค์ประกอบของการปรับสเกล (Scaling factor) เพื่อทำให้ฟังก์ชันโลจิส (Logistic Function) และฟังก์ชันปกติสะสม (Normal Ogive Function) มีค่าใกล้เคียงกัน หรือมีค่า θ ที่ประมาณได้ แตกต่างกันไม่เกิน 0.01 โดยกำหนดให้ค่า D มีค่าเท่ากับ 1.70

2) โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์

เป็นโมเดลที่มีลักษณะสำคัญคือ ค่าพารามิเตอร์ความยาก (b_i) มีค่าแปรเปลี่ยนไปตามลักษณะของข้อสอบแต่ละข้อ ในตำแหน่ง θ ที่มีโอกาสตอบถูก 0.50 ส่วนค่าพารามิเตอร์อำนาจจำแนก (a_i) มีค่าคงที่ และพารามิเตอร์โอกาสในการเดาข้อสอบได้ถูก (c_i) มีค่าเป็น 0 โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ สามารถเขียนด้วยฟังก์ชันโลจิก ได้ดังนี้

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

และมีโค้งลักษณะข้อสอบ ดังภาพ



ภาพที่ 1 โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ของตัวอย่างข้อสอบ 3 ข้อ

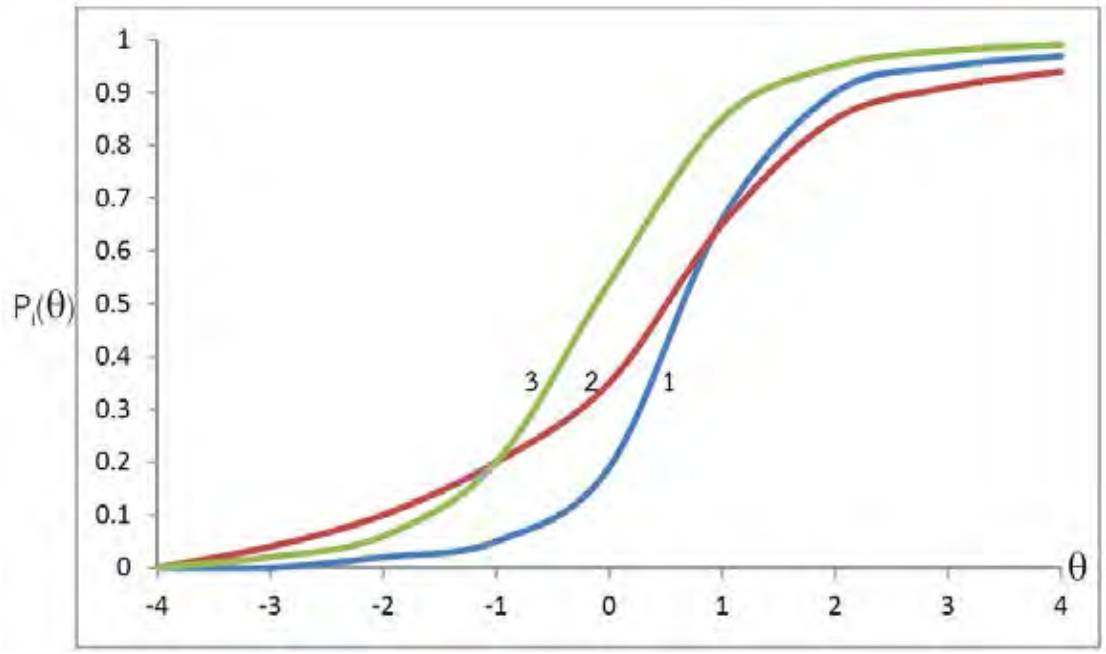
3) โมเดลการตอบสนองข้อสอบ 2 พารามิเตอร์

เป็นโมเดลซึ่งมีลักษณะสำคัญคือ ค่าพารามิเตอร์ความยาก (b_i) มีค่าแปรเปลี่ยนไปตามลักษณะของข้อสอบแต่ละข้อ ในตำแหน่ง θ ที่มีโอกาสตอบถูก 0.50 ส่วนค่าพารามิเตอร์

อำนาจจำแนก (a_i) มีค่าแปรเปลี่ยนไปตามข้อสอบแต่ละข้อ เป็นค่าความชันบนโค้ง ICC ณ ตำแหน่งของค่าความยากของข้อที่ i และพารามิเตอร์โอกาสในการเดาข้อสอบได้ถูก (c_i) มีค่าเป็น 0 โดยมีค่าคงที่ e เท่ากับ 2.718 และ D เท่ากับ 1.70 โมเดลการตอบสนองข้อสอบ 2 พารามิเตอร์ สามารถเขียนด้วยฟังก์ชันโลจิก ได้ดังนี้

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

และมีโค้งลักษณะข้อสอบ ดังภาพ



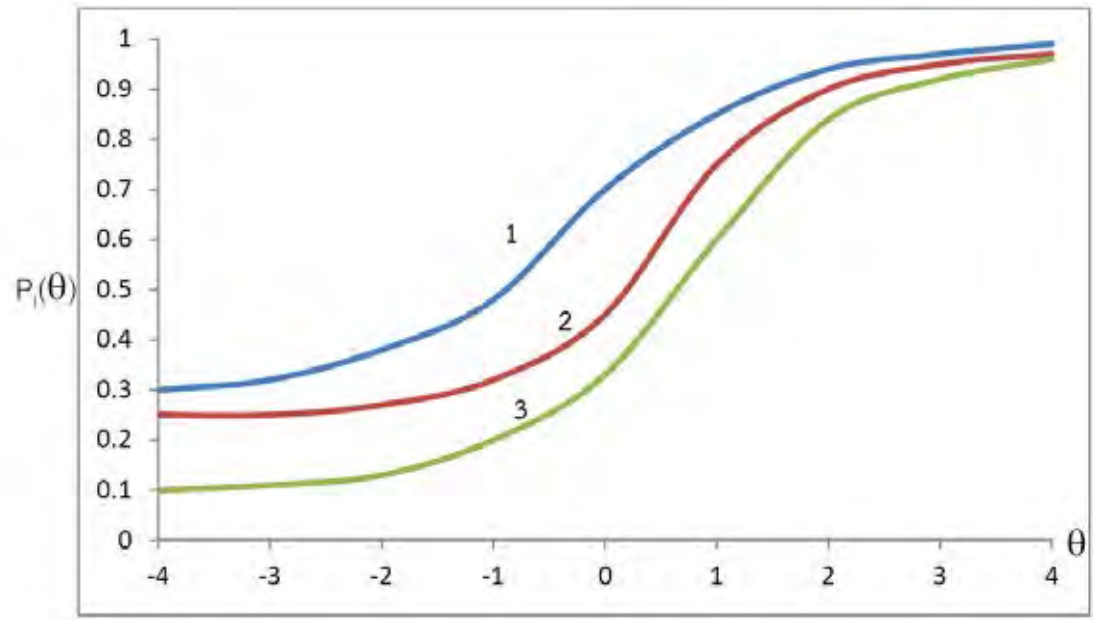
ภาพที่ 2 โค้งลักษณะข้อสอบแบบ 2 พารามิเตอร์ของตัวอย่างข้อสอบ 3 ข้อ

4) โมเดลการตอบสนองข้อสอบ 3 พารามิเตอร์

เป็นโมเดลซึ่งมีลักษณะสำคัญคือ ค่าพารามิเตอร์ความยาก (b_i) มีค่าแปรเปลี่ยนไปตามลักษณะของข้อสอบแต่ละข้อ ในตำแหน่ง θ ที่มีโอกาสตอบถูก $\frac{1+c_i}{2}$ ส่วนค่าพารามิเตอร์อำนาจจำแนก (a_i) มีค่าแปรเปลี่ยนไปตามข้อสอบแต่ละข้อ เป็นค่าความชันบนโค้ง ICC ณ ตำแหน่งของค่าความยากของข้อที่ i และพารามิเตอร์โอกาสในการเดาข้อสอบได้ถูก (c_i) มีค่าแปรเปลี่ยนไปตามลักษณะของข้อสอบแต่ละข้อ โดยมีค่าคงที่ e เท่ากับ 2.718 และ D เท่ากับ 1.70 โมเดลการตอบสนองข้อสอบ 3 พารามิเตอร์ สามารถเขียนด้วยฟังก์ชันโลจิก ได้ดังนี้

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

และมีโค้งลักษณะข้อสอบ ดังภาพ



ภาพที่ 3 โค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ของตัวอย่างข้อสอบ 3 ข้อ

เนื่องจากโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ เป็นโมเดลที่มีจำนวนพารามิเตอร์ที่ต้องประมาณค่าน้อยกว่าโมเดลอื่น ทำให้ง่ายต่อการวิเคราะห์ ไม่ยุ่งยากซับซ้อน (Hambleton, 1989) และโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ยังเป็นโมเดลที่ให้ค่าประมาณของพารามิเตอร์ที่ถูกต้องกว่าโมเดลแบบอื่นๆเมื่อนำไปใช้กับกลุ่มตัวอย่างขนาดเล็ก (van de Vijver, 1986 อ้างถึงใน Baker, 1992; Lord, 1980) ด้วยเหตุนี้ โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์จึงเป็นโมเดลที่ผู้วิจัยสนใจนำมาเป็นตัวแบบในการศึกษาประมาณค่าพารามิเตอร์ในครั้งนี้

1.2.2 ข้อตกลงเบื้องต้น

ทฤษฎีการตอบสนองข้อสอบ (IRT) มีข้อตกลงเบื้องต้น ดังต่อไปนี้ (Hambleton and Swaminathan, 1985)

1) ความเป็นเอกมิติ (Unidimensionality) คุณลักษณะภายในหรือความสามารถของผู้สอบมีอยู่หลายประการ ซึ่งคุณลักษณะภายในหรือความสามารถแต่ละอย่างสามารถกำหนดพฤติกรรมการตอบสนองข้อสอบ ถ้าผลการตอบสนองข้อสอบหรือคะแนนของผู้สอบสามารถอธิบายได้ด้วยคุณลักษณะเดียว ก็ถือว่ามีความเป็นมิติเดียว โมเดลของการตอบสนองข้อสอบมีข้อตกลงว่า ข้อสอบแต่ละข้อในแบบสอบนั้นวัดความสามารถหรือคุณลักษณะเดียวกัน แต่ข้อตกลงข้อนี้ไม่เข้มงวดนัก ถ้ามีลักษณะเด่นที่จะวัดองค์ประกอบใดองค์ประกอบหนึ่งก็ถือว่าเป็นมิติเดียวเช่นกัน

การตรวจสอบความเป็นเอกมิติของแบบสอบ มีวิธีการตรวจสอบได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) ของข้อสอบทั้งฉบับ ถ้าค่าไอเกน (Eigen Value) ของตัวประกอบแรกแตกต่างจากตัวประกอบถัดไปอย่างชัดเจน โดยมีอัตราส่วนที่สูง แสดงว่าแบบสอบนั้นวัดมิติเดียวหรือวัดคุณลักษณะเดียว (Single Dominant Factor) (Hambleton and Swaminathan, 1985) และสามารถวิเคราะห์ยืนยันว่าแบบสอบนั้นวัดคุณลักษณะเดียวได้โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis)

2) ความเป็นอิสระ (Local Independence) หมายถึง ความเป็นอิสระระหว่างข้อสอบและผู้สอบ

(1) ความเป็นอิสระระหว่างข้อสอบ เมื่อผลการตอบข้อสอบรายข้อในแบบสอบของผู้สอบคนเดียวกันเป็นอิสระจากกัน การตอบข้อสอบข้อหนึ่งๆจะไม่มีผลกระทบต่อการตอบข้อสอบข้ออื่นๆในแบบสอบ เนื้อหาของข้อสอบข้อหนึ่งๆจะต้องไม่มีเงื่อนไขในการตอบข้อสอบข้ออื่นๆ ถ้าสุ่มผู้สอบที่มีความสามารถ θ มา 1 คน เพื่อตอบข้อสอบ k ข้อ และให้ U_j เป็นคะแนนข้อสอบข้อที่ j ที่ได้จากการตอบหลังจากควบคุมความสามารถ θ ของผู้สอบ คะแนนผลการตอบในแต่ละข้อของผู้สอบคนเดียวกันนั้นจะไม่สัมพันธ์กัน ดังนี้

$$P(U_1,U_2,\ldots,U_k/\theta) = P(U_1/\theta)\,P(U_2/\theta)\ldots P(U_k/\theta)$$

$$= \prod_{j=1}^{k} P(U_k/\theta)$$

(2) ความเป็นอิสระระหว่างผู้สอบ เมื่อผลการตอบข้อสอบข้อเดียวกันของผู้สอบต่างคนกันเป็นอิสระจากกัน การตอบข้อสอบข้อหนึ่งๆของผู้สอบแต่ละคนจะไม่มีผลกระทบต่อผู้สอบอื่นที่ตอบข้อสอบข้อนั้น ถ้าสุ่มข้อสอบมา 1 ข้อ ในการตอบข้อสอบของผู้สอบ n คน และให้ U_i เป็นคะแนนข้อสอบของผู้สอบคนที่ i ที่ได้จากการตอบหลังจากควบคุมความสามารถ θ ของผู้สอบ คะแนนผลการตอบข้อนั้นของผู้สอบแต่ละคนไม่สัมพันธ์กัน ดังนี้

$$P(U_1,U_2,\ldots,U_n/\theta) = P(U_1/\theta)\,P(U_2/\theta)\ldots P(U_n/\theta)$$

$$= \prod_{i=1}^{n} P(U_i/\theta)$$

3) โมเดลการตอบสนองข้อสอบ (Item Response Models) เสนอฟังก์ชันโลจิสมีลักษณะเป็นโค้งลักษณะของข้อสอบ (Item Characteristic Curves: ICC) ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบนั้นได้ถูกต้องกับระดับความสามารถที่วัดได้โดยชุดของข้อสอบหรือแบบสอบ จะเห็นได้ว่าโอกาสที่ผู้สอบจะตอบข้อสอบถูก (probability) ไม่ขึ้นอยู่กับจำนวนของผู้สอบที่มีความสามารถเหมือนกัน แต่ขึ้นอยู่กับโค้งลักษณะของข้อสอบ (ICC) ที่เป็นอิสระจากการกระจายของความสามารถของผู้สอบ รูปร่างของโค้งลักษณะของข้อสอบ (ICC) แต่ละข้อมีคุณสมบัติที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ดังนั้นจึงทำให้โอกาสในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยน

โมเดลการตอบสนองข้อสอบแบบตรวจให้คะแนน 2 ค่า มีอยู่หลากหลายโมเดล แต่ที่นิยมใช้กันแพร่หลาย ได้แก่ โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์, 2 พารามิเตอร์ และ 3 พารามิเตอร์ ไม่ว่าจะเลือกใช้โมเดลใดก็ตาม ควรมีการตรวจสอบความเหมาะสมของโมเดลการตอบสนองข้อสอบกับข้อมูล (Model-Data Fit) เพื่อให้ผลการวิเคราะห์มีความถูกต้องและน่าเชื่อถือ โดยตรวจสอบความสอดคล้องทั้งความไม่แปรเปลี่ยนของค่าประมาณความสามารถ (Invariance of ability parameter estimates) และความไม่แปรเปลี่ยนของค่าประมาณพารามิเตอร์ของข้อสอบ (Invariance of item parameter estimates)

การตรวจสอบความไม่แปรเปลี่ยนของค่าประมาณความสามารถ ทำได้โดยการเปรียบเทียบค่าประมาณความสามารถของผู้สอบที่ได้จากกลุ่มข้อสอบที่แตกต่างกัน เช่น กลุ่มข้อสอบยาก กลุ่มข้อสอบง่าย หรือกลุ่มข้อสอบจากคลังข้อสอบเดียวกันแต่มีความครอบคลุมของเนื้อหาแตกต่างกัน เป็นต้น เมื่อความแตกต่างเกิดขึ้นไม่เกินความคลาดเคลื่อนมาตรฐานของการประมาณค่า จะถือว่าค่าประมาณความสามารถไม่แปรเปลี่ยน (Wright, 1968 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) การตรวจสอบความไม่แปรเปลี่ยนของค่าประมาณพารามิเตอร์ของข้อสอบทำได้โดยเปรียบเทียบค่าประมาณพารามิเตอร์แต่ละตัวของข้อสอบที่ได้จาก กลุ่มตัวอย่างประชากรผู้สอบหลายกลุ่ม เช่น กลุ่มผู้สอบชาย/หญิง กลุ่มผู้สอบจำแนกตามภูมิภาค เป็นต้น เมื่อผลการพล๊อตกราฟออกมาเป็นเส้นตรง โดยมีการกระจายไม่แตกต่างจากผลที่ได้จากกลุ่มตัวอย่าง 2 กลุ่ม ซึ่งเป็นกลุ่มสุ่มที่ทัดเทียมกัน จะถือว่าค่าประมาณพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยน (Shepard, Camilli & Williams, 1984)

4) การสอบที่ไม่แข่งขันด้านเวลา (Nonspeeded Test Administration) ตามทฤษฎีตอบสนองข้อสอบ ความเร็วในการตอบต้องไม่มีอิทธิพลต่อผลการตอบ ดังนั้นการจัดสอบจึงต้องคำนึงถึงเวลาที่กำหนดให้ในการสอบจะต้องเพียงพอที่ผู้สอบซึ่งมีความสามารถจะทำข้อสอบได้ทันเวลา การตรวจสอบความเหมาะสมของเวลาในการสอบ พิจารณาจากสัดส่วนหรือร้อยละของผู้สอบที่ทำข้อสอบได้ครบทุกข้อ เช่น จำนวนผู้สอบส่วนใหญ่ ร้อยละ 80 สามารถตอบข้อสอบได้ครบหรือเกือบครบทุกข้อ อีกทั้งควรตรวจสอบเปรียบเทียบความแปรปรวนของจำนวนข้อที่เว้นกับความแปรปรวนของจำนวนข้อที่ตอบผิด (Gulliksen, 1950 อ้างถึงใน ศิริชัย กาญจนวาสี, 2555) ถ้าอัตราส่วนของความแปรปรวนเข้าใกล้ 0 แสดงว่าเวลาที่กำหนดในการจัดสอบนี้เพียงพอต่อการทำข้อสอบเป็นไปตามข้อตกลงเบื้องต้น (Hambleton, Swaninathan & Rogers, 1991)

1.2.3 การวัดคุณภาพตามแนวทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ สามารถบ่งบอกคุณภาพของข้อสอบรายข้อและคุณภาพของข้อสอบทั้งฉบับ จากฟังก์ชันสารสนเทศของข้อสอบและแบบสอบ (Item and Test Information) นอกจากนี้ยังสามารถเปรียบเทียบประสิทธิภาพระหว่างแบบสอบสองฉบับ ณ ตำแหน่งความสามารถเดียวกัน โดยพิจารณาจากประสิทธิภาพสัมพัทธ์ (Relative Efficiency)

1) ฟังก์ชันสารสนเทศของข้อสอบ (Item Information; $I_i(\theta)$)

ฟังก์ชันสารสนเทศของข้อสอบ ที่ใช้บ่งชี้คุณภาพของข้อสอบรายข้อ ประกอบด้วยพารามิเตอร์ความยาก อำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อ ดังนี้

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \qquad ; i = 1, 2, \ldots, k$$

เมื่อกำหนดให้ $I_i(\theta)$ คือ ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้รับจากข้อสอบข้อที่ i สำหรับผู้ตอบที่มีความสามารถ θ

$P'_i(\theta)$ คือ P'_i เป็นค่าความชันของฟังก์ชันการตอบสนองข้อสอบข้อที่ i ณ ตำแหน่งความสามารถ θ

$P_i(\theta)$ คือ P_i เป็นค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$Q_i(\theta)$ คือ Q_i หรือ $1- P_i(\theta)$

โดยมีสูตรการคำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ ดังนี้

ตารางที่ 2.2 สูตรการคำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ (ศิริชัย กาญจนวาสี, 2555)

ค่าประมาณ	1 พารามิเตอร์	2 พารามิเตอร์	3 พารามิเตอร์
$I_i(\theta)$	$D^2P_iQ_i$	$D^2a_i^2P_iQ_i$	$D^2a_i^2Q_i(P_i-c_i)^2 / (1- c_i)^2$
P'_i	DP_iQ_i	$D\ a_i\ P_iQ_i$	$D\ a_iQ_i(P_i-c_i) / (1- c_i)$
$I_i(\theta)_{max}$	$\frac{1}{4}D^2$	$\frac{1}{4}D^2\ a_i^2$	$D^2a_i^2\ \frac{[1-20c_i-8c_i^2+(1+8c_i)^{3/2}]}{8(1-c_i^2)}$
θ_{max}	b_i	b_i	$b_i + \frac{1}{D_i}[\ln\frac{(1+(1+8c_i)^{1/2}}{2}]$

ค่าสารสนเทศของข้อสอบจะมีค่าสูง เมื่อผู้สอบมีความสามารถ θ ใกล้กันกับค่าพารามิเตอร์ความยาก b ของข้อสอบ แต่จะมีค่าลดลงเมื่อผู้สอบมีความสามารถ ห่างออกไปจากค่าพารามิเตอร์ความยากของข้อสอบ ค่าสารสนเทศของข้อสอบจะมีค่าสูงขึ้น ถ้าค่าอำนาจจำแนก a ของข้อสอบ มีค่ามากขึ้น และข้อสอบ ค่าสารสนเทศของข้อสอบจะมีค่าสูงขึ้น เมื่อโอกาสในการเดาข้อสอบได้ถูก c มีค่ามากขึ้น นั่นคือ ค่าสารสนเทศของข้อสอบ $I_i(\theta)$ จะมีค่าสูงสุด ในตำแหน่ง θ_{max} เมื่อ P_i และ Q_i เท่ากับ .5 ถ้า $c_i = 0$ แล้ว $\theta_{max} = b$ แต่ถ้า $c_i > 0$ แล้ว $\theta_{max} > b$

2) ฟังก์ชันสารสนเทศของแบบสอบ (Test Information; $I(\theta)$)

คือ สารสนเทศของแบบสอบ ที่เกิดจากผลรวมพีชคณิตของค่าฟังก์ชันสารสนเทศของข้อสอบทุกข้อในแบบสอบหนึ่งๆ ณ ตำแหน่งความสามารถ θ เดียวกัน ค่าฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อเป็นอิสระจากกัน โดยแต่ละข้อมีส่วนต่อค่าสารสนเทศของแบบสอบ ดังนี้

$$I(\theta) = \sum_{i=1}^{k} I_i(\theta)$$

3) ประสิทธิภาพสัมพัทธ์ (Relative Efficiency; $RE(\theta)$)

เป็นการเปรียบเทียบประสิทธิภาพของแบบสอบ 2 ฉบับที่วัดคุณลักษณะเดียวกัน สำหรับใช้สอบกับผู้สอบที่มีความสามารถ θ สามารถคำนวณได้จากอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบสอบต่างฉบับกัน ที่ตำแหน่ง θ เดียวกัน ดังนี้

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

เมื่อกำหนดให้ $I_A(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของแบบสอบ A ที่ตำแหน่งความสามารถร่วมกันระดับ θ และ

$I_B(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของแบบสอบ B ที่ตำแหน่งความสามารถร่วมกันระดับ θ เดียวกัน

4) ความคลาดเคลื่อนมาตรฐาน (Standard Error; SE(θ)

คือ ค่าส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็นของค่าประมาณความสามารถที่แท้จริง หรือเป็นค่าสัดส่วนผกผันกับค่าฟังก์ชันสารสนเทศของแบบสอบ คุณภาพของแบบสอบจะมีความแม่นยำสูงในการประมาณค่าความสามารถของผู้สอบ ณ ระดับ θ นั้น เมื่อค่า SE(θ) มีค่าต่ำ หรือเมื่อมีค่า I(θ) สูง ดังสูตร

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

ความคลาดเคลื่อนมาตรฐานของการประมาณค่า SE(θ) นี้มีค่าผันแปรไปตามตำแหน่งหรือระดับความสามารถของผู้ตอบข้อสอบแต่ละคน ซึ่งเป็นค่าที่คล้ายกับความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Estimation; SEM)ในทฤษฎีการทดสอบแบบดั้งเดิม แต่มีความแตกต่างกันที่ SEM เป็นค่าคงที่สำหรับผู้ตอบทุกคน

ตอนที่ 2 การประมาณค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) และพารามิเตอร์ของข้อสอบ (β; a, b, c) ซึ่งทั้งสองค่าต่างก็เป็นพารามิเตอร์ที่ไม่สามารถสังเกตได้โดยตรง จึงต้องทำการประมาณค่าความสามารถและพารามิเตอร์ของข้อสอบจากข้อมูลผลการตอบของผู้สอบที่สามารถสังเกตได้ วิธีประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบมีหลายวิธีการ ซึ่งมีการพัฒนาขึ้นมาตามลำดับ ดังนี้

2.1 การประมาณค่าที่เป็นไปได้สูงสุด (Maximum Likelihood Estimation)

การนำทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้ ได้นำเอาวิธีการประมาณค่าที่เป็นไปได้สูงสุด (Maximum Likelihood Estimation ; MLE) มาประมาณค่าพารามิเตอร์ได้ทั้งในกรณีที่ทราบค่าพารามิเตอร์ข้อสอบแล้วต้องการประมาณค่าพารามิเตอร์ความสามารถ และกรณีที่ทราบพารามิเตอร์ความสามารถแล้วต้องการประมาณค่าพารามิเตอร์ข้อสอบ

(1) กรณีที่ต้องการประมาณค่าพารามิเตอร์ความสามารถ เมื่อทราบค่าพารามิเตอร์ข้อสอบ โดยสมมติว่าในการสอบครั้งหนึ่งมีข้อสอบ k ข้อ มีกลุ่มผู้สอบ n คน เมื่อสุ่มผู้สอบ p ที่มีความสามารถ θ และมีผลการตอบข้อสอบเป็น $U_1, U_2, \ldots, U_k$ ซึ่งมีค่าเป็น 0 ถ้าตอบผิด หรือ 1 ถ้า

ตอบถูก จะได้ความน่าจะเป็นร่วมของค่าสังเกตได้ตามผลการตอบที่มีค่าเท่ากับผลคูณระหว่างความน่าจะเป็นของค่าสังเกตได้รายข้อ ดังนี้

$$P(U_1,U_2,..., U_j ,...,U_k/\theta) = P(U_1/\theta)\, P(U_2/\theta) \ldots P(U_n/\theta) \ldots P(U_k/\theta)$$
$$= \prod_{j=1}^{k} P(U_j/\theta)$$
$$= \prod_{j=1}^{k} [P(U_j/\theta)]^{U_j} [1-P(U_j/\theta)]^{1-U_j}$$
$$= \prod_{j=1}^{k} P^{U_j} Q^{1-U_j}$$

หรือเรียกค่าความน่าจะเป็นร่วมของค่าสังเกตได้ นี้ว่า ฟังก์ชันความน่าจะเป็น (Likelihood function) โดยใช้สัญญลักษณ์แทนด้วย

$$L(u_1,u_2,..., u_j ,...,u_k/\theta) = \prod_{j=1}^{k} P^{U_j} Q^{1-U_j} \qquad \text{เมื่อ } U_j = u_j$$

แต่วิธีการประมาณค่าที่เป็นไปได้สูงสุด (MLE) นี้ ยังไม่สามารถให้ค่าประมาณพารามิเตอร์ความสามารถได้ ต้องใช้วิธีของฟิชเชอร์ (Fisher's method) หรืออีกวิธีที่นิยมใช้กันคือ วิธีนิวตัน-ราฟสัน (Newton-Raphson procedure) เป็นวิธีการคำนวณทวนซ้ำบนพื้นฐานของ Taylor series expansion เพื่อช่วยในการแก้สมการอนุพันธ์ที่หนึ่ง (first derivative) ของฟังก์ชัน ln L ซึ่งกำหนดค่าอนุพันธ์เป็น 0 (Hambleton and Swaminatan, 1985)

(2) กรณีที่ต้องการประมาณค่าพารามิเตอร์ข้อสอบ จะใช้วิธีการของนิวตัน-ราฟสัน ในรูปแบบตัวแปรพหุ เนื่องจากฟังก์ชันความน่าจะเป็นสำหรับค่าพารามิเตอร์ข้อสอบแต่ละข้อมีลักษณะเป็นพหุมิติ คือเป็นฟังก์ชันของพารามิเตอร์ 3 ตัว (กรณีโมเดล 3-พารามิเตอร์; a b c) หาค่าอนุพันธ์ที่หนึ่งของฟังก์ชันความน่าจะเป็นสัมพัทธ์กับค่าพารามิเตอร์ข้อสอบแต่ละตัว แล้วกำหนดให้มีค่าอนุพันธ์เท่ากับ 0 จากนั้นจึงแก้สมการ 3 สมการไปพร้อมๆกัน การประมาณค่าข้อสอบจะทำซ้ำเท่ากับจำนวนข้อสอบ

การประมาณค่าด้วยวิธีการประมาณค่าที่เป็นไปได้สูงสุดนี้ จำเป็นต้องใช้ในกลุ่มตัวอย่างที่มีขนาดใหญ่ เพื่อให้ค่าพารามิเตอร์ที่ประมาณได้มีความถูกต้อง และเมื่อแบบสอบมีความยาวมากขึ้น จะทำให้ค่าพารามิเตอร์ความสามารถที่ประมาณได้ ไม่มีความลำเอียง (unbiased) โดยมีการแจงแจงปกติ มีค่าเฉลี่ยเท่ากับค่าความสามารถ ค่าความสามารถที่ระดับความเชื่อมั่น (1-α)% มีค่าอยู่ในช่วง

$$\hat{\theta} - Z_{\alpha/2} SE(\hat{\theta}) < \theta < \hat{\theta} + Z_{\alpha/2} SE(\hat{\theta})$$

2.2 การประมาณค่าร่วมกันของพารามิเตอร์ข้อสอบและความสามารถ (Joint Maximum Likelihood Estimation ; JMLE) ซึ่งเป็นรูปแบบที่เสนอโดย Birnbaum เนื่องจากในสถานการณ์โดยทั่วไปมักจะไม่ทราบค่าทั้งพารามิเตอร์ความสามารถและพารามิเตอร์ข้อสอบ ดังนั้น ในปี ค.ศ. 1968 Birnbaum จึงได้พัฒนาวิธีการในการประมาณค่าพารามิเตอร์ทั้ง 2 ไปพร้อมๆกันโดยใช้กระบวนการ 2 ขั้นตอนสลับไปมา เพื่อแก้สมการสำหรับค่าประมาณของพารามิเตอร์จำนวน k + n สำหรับโมเดล 1-พารามิเตอร์ และจำนวน 2k + n หรือ 3k + n สำหรับโมเดล 2-พารามิเตอร์ และ 3-พารามิเตอร์ ตามลำดับ การประมาณค่าวิธี JMLE ทำได้โดยการกำหนดค่าเริ่มต้นของพารามิเตอร์ มี 2 ขั้นตอน คือ **ขั้นตอนแรก** ต้องกำหนดค่าพารามิเตอร์ความสามารถเริ่มต้นขึ้นมาก่อน โดยใช้ค่า log ของอัตราส่วนจำนวนข้อที่ตอบถูก ต่อจำนวนข้อที่ตอบผิด สำหรับผู้สอบแต่ละคน แปลงเป็นคะแนนมาตรฐานมาใช้เป็นค่าความสามารถเริ่มต้น เสมือนกับทราบค่าความสามารถ เพื่อนำไปประมาณค่าพารามิเตอร์ข้อสอบ **ขั้นตอนที่สอง** ค่าประมาณของข้อสอบที่ได้มาจากขั้นตอนแรก เสมือนกับทราบค่าพารามิเตอร์ข้อสอบ นำมาใช้ในการประมาณค่าพารามิเตอร์ความสามารถ วิธีการนี้จะดำเนินการประมาณค่าสลับไปมาระหว่าง 2 ขั้นตอนนี้จนกระทั่งค่าที่ประมาณได้เปลี่ยนแปลงน้อยมาก พอที่จะยอมรับได้ (Baker, 1992)

วิธีการนี้มีข้อจำกัดคือ ไม่สามารถประมาณค่าในกรณีที่ผู้สอบมีคะแนนเต็มหรือไม่ได้คะแนนเลย หรือข้อสอบที่ง่ายมากผู้สอบตอบได้ทุกคนหรือข้อสอบยากมากผู้สอบตอบผิดทุกคนต้องขจัดออกไปก่อนจะวิเคราะห์ วิธีการนี้จะต้องใช้จำนวนข้อสอบมากและกลุ่มผู้สอบขนาดใหญ่จึงจะทำให้ค่าประมาณที่ได้จะคงเส้นคงวาสำหรับโมเดล 2 และ 3 พารามิเตอร์ นอกจากนี้สำหรับโมเดล 3 พารามิเตอร์ อาจมีปัญหาในการประมาณค่าได้ หากไม่จำกัดค่าพารามิเตอร์ความสามารถและพารามิเตอร์ข้อสอบ การแก้ปัญหาข้อจำกัดเหล่านี้ได้โดยใช้วิธีของเบส์ (Bayesian estimates) ช่วยในการประมาณค่าพารามิเตอร์ โดยกำหนดการแจกแจงเบื้องต้นความน่าเป็นของข้อสอบและความสามารถ (Baker, 1992)

2.3 การประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าปลายทาง (Marginal Maximum Likelihood Estimation ; MMLE) เป็นวิธีการที่พัฒนาขึ้นโดย Bock และ Lieberman (1970) ได้พัฒนาสูตรการประมาณค่าต่อมาโดย Bock และ Aitkin (1981) วิธีนี้ช่วยแก้ปัญหาความไม่คงเส้นคงวาของการประมาณค่าพารามิเตอร์ โดยการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถไปพร้อมๆกัน ด้วยการประมาณการแจกแจงความสามารถผู้สอบขึ้นมาก่อน ทำให้ได้ฟังก์ชันความน่าจะเป็นปลายทางของพารามิเตอร์ข้อสอบ ประมาณค่าพารามิเตอร์ข้อสอบ แล้วนำมาใช้

ประมาณค่าความสามารถผู้สอบ (Van Der Linden and Hambliton,1997; ศิริชัย กาญจนวาสี, 2555)

ข้อจำกัดของวิธีการนี้คือ การวิเคราะห์จำเป็นต้องมีกลุ่มผู้สอบขนาดใหญ่ และจำนวนข้อสอบมาก เพื่อให้ประมาณค่าพารามิเตอร์ข้อสอบได้คงที่และประมาณค่าความสามารถได้ถูกต้อง นอกจากนี้ในกรณีโมเดล 3 พารามิเตอร์ ควรจะมีผู้สอบที่ความสามารถต่ำพอที่จะประมาณค่าพารามิเตอร์การเดาข้อสอบได้ถูก (c) ได้อย่างน่าเชื่อถือ เพราะจะส่งผลต่อคุณภาพของค่าประมาณของพารามิเตอร์ a และ b การแก้ปัญหาข้อจำกัดนี้ได้โดยใช้วิธีของเบย์ (Bayesian estimation) ช่วยในการกำหนดค่าแจกแจงเบื้องต้นของพารามิเตอร์ c (ศิริชัย กาญจนวาสี, 2555)

2.4 การประมาณค่าที่เป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood Estimation ; CMLE) ใช้ประมาณค่าพารามิเตอร์ความสามารถ ในกรณีที่ทราบค่าพารามิเตอร์ข้อสอบ แต่มีข้อจำกัด ถ้าผู้สอบตอบถูกทุกข้อหรือผิดทุกข้อ จะไม่สามารถประมาณค่าพารามิเตอร์ความสามารถได้ นอกจากนี้วิธี CMLE สามารถใช้ในการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถไปพร้อมๆกันได้ ในกรณีที่ไม่ทราบค่าพารามิเตอร์ทั้งสองค่า โดยบทความของ Andersen ในปี 1972 และ 1973 ได้กล่าวถึงกรณีที่ความลำเอียงในตัวประมาณค่าไม่หายไป เมื่อมีผู้สอบเพิ่มจำนวนมากขึ้น ทำให้ได้ตัวประมาณค่าที่เป็นไปได้สูงสุดของพารามิเตอร์ข้อสอบไม่คงเส้นคงวา Andersen จึงเสนอวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบมีเงื่อนไขที่ทำให้ได้ตัวประมาณค่าที่เป็นไปได้สูงสุดของพารามิเตอร์ข้อสอบที่มีความคงเส้นคงวา วิธีการแบบมีเงื่อนไขนี้ถูกยืนยันด้วยการได้มาของสถิติที่เพียงพอสำหรับพารามิเตอร์ความสามารถ ซึ่งเกิดขึ้นโดยบังเอิญ (Hambleton and Swaminathan, 1985) จากราสช์โมเดล เมื่อจำนวนคะแนนที่ตอบถูกคือ r_a เป็นค่าสถิติที่เพียงพอสำหรับพารามิเตอร์ความสามารถ θ_a เราสามารถแสดงฟังก์ชันความน่าจะเป็นของ $L(u|\theta_a, b_i)$ ในเทอมของ r_a และ θ_a โดยในโมเดลราสช์ เราสามารถตัดตัวห้อยในเทอมออกได้เลย เป็น r และ θ

$$P(U_i|\theta, b_i) = \exp DU_i(\theta - b_i)/[1 + \exp D(\theta - b_i)]$$

ดังนั้น

$$P(U_1, U_2, ..., U_n|\theta, b) = \prod_{i=1}^{n} P[U_i|\theta, b_i] =$$

$$[\exp(D\theta\sum U_i)\exp(-D\sum U_i b_i)] \Big/ \prod_{i=1}^{n}[1 + \exp D(\theta - b_i)]$$

$$= [\exp(D\theta r)\exp(-D\sum U_i b_i)]/g(\theta, b)$$

จะได้ค่าความน่าจะเป็นจากคะแนนดิบ r ดังนี้

$$P[r|\theta,b] = [\exp(D\theta x][\sum \exp(-D\sum U_i b_i)]/g(\theta,b)$$

$$P[r|\theta,b] = P[U|\theta,b]/P[r|\theta,b]$$

$$= \exp(-D\sum U_i b_i)/[\sum_r \exp(-D\sum U_i b_i)]$$

$$= \exp(-D\sum U_i b_i)/\gamma_r$$

เมื่อกำหนดให้ $\gamma_r = \sum_r \exp(-D\sum_{i=1}^{n} U_i b_i)$

ฟังก์ชันข้างต้นเป็นฟังก์ชันความยากของข้อสอบ b ซึ่งเป็นองค์ประกอบหนึ่งของฟังก์ชันคะแนนที่ตอบถูก r ดังนั้นเมื่อมีผลการตอบแบบสอบออกมา ความน่าจะเป็น $P(U_i|\theta,b_i)$ จะถูกแปลไปเป็นฟังก์ชันความน่าจะเป็น $L(u|r,b)$ ซึ่งมีความเป็นอิสระของค่าพารามิเตอร์สามารถ

วิธีการ CMLE จะให้ตัวประมาณค่าที่มีคุณสมบัติคงเส้นคงวาสูงเท่าที่จะเป็นได้ แต่วิธีการนี้มีข้อจำกัดคือ จะใช้ได้ผลดีกับจำนวนข้อสอบไม่เกิน 40 ข้อ ถ้าข้อสอบสูงเกิน 60 ข้อ วิธีการนี้จะคำนวณได้ช้า และถ้ามีข้อสอบ 80 หรือ 100 ข้อ ก็จะทำให้ไม่สามารถคำนวณผลได้ด้วย CMLE (Hambleton and Swaminathan, 1985)

2.5 การประมาณค่าแบบเบย์ (Bayesian estimation)

การประมาณค่าแบบเบย์จะให้การประมาณค่าที่ถูกต้องกว่าวิธีอื่น (วินัย วงศ์ฤทัยวัฒนา, 2533; วิชุดา บัวคง, 2533; สุนทร เทียนงาม, 2551) โดยที่จะประมาณค่าพารามิเตอร์ความสามารถผู้สอบ $\theta_i (i=1,2,...,N)$ และพารามิเตอร์ข้อสอบ $\beta_j (j=1,2,...,n)$ ที่ทำให้ฟังก์ชันการแจกแจงภายหลัง $f(\theta,\beta|u)$ มีค่าสูงสุด แต่เนื่องจาก $\ln f(\theta,\beta|u)$ ฟังก์ชันที่หาอนุพันธ์ได้ และสามารถหา θ_i และ β_j ได้ จึงใช้ $\ln f(\theta,\beta|u)$ โดยกำหนดให้มีค่าเท่ากับศูนย์ แล้วหาค่ารากของอนุพันธ์ของ $\ln f(\theta,\beta|u)$

สมการอนุพันธ์ของ $\ln f(\theta,\beta|u)$ หรือ สมการโมดัล (Modal Equation) เมื่อถอดรากของสมการนี้ จะได้ค่าความสามารถ θ และค่าพารามิเตอร์ข้อสอบ β ที่ทำให้ได้ฟังก์ชันการแจกแจงภายหลังมีค่าสูงสุด โดยอาจจะใช้วิธีของ Newton-Raphson ที่ทำการคำนวณทวนซ้ำ (Iterative) เพื่อประมาณค่า การประมาณค่าแบบเบย์นี้จะกำหนดค่าเริ่มต้นสำหรับใช้ประมาณค่าความสามารถ และค่าเริ่มต้นสำหรับใช้ประมาณค่าพารามิเตอร์ข้อสอบ และดำเนินการคำนวณทวนซ้ำ จนกระทั่งได้ค่าประมาณความสามารถและค่าพารามิเตอร์ข้อสอบ ที่ลู่เข้าหา (convergence) ค่าคงที่ค่าใดค่าหนึ่ง ค่าที่ได้ครั้งสุดท้าย คือ ค่าที่ประมาณได้จากการประมาณค่าแบบเบย์

การประมาณค่าแบบเบย์นี้จะใช้ค่าเฉลี่ยภายหลังไปใช้ประมาณค่า ค่าเฉลี่ยภายหลังหรือตัวประมาณค่าที่ดีจะต้องช่วยลดค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่า (MSE) ลง ตัวประมาณค่าจะช่วยทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าลดลง โดยจะให้การแจกแจงที่เหมาะสมมาใช้ในการประมาณค่า ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าความสามารถ คำนวณได้จากสมการ

$$\mathrm{MSE} = \mathrm{E}(\hat{\theta}_{\mathrm{a}} - \theta_{\mathrm{a}})^2$$

เมื่อ $\hat{\theta}$ เป็นตัวประมาณค่าของ θ และ E คือ ค่าคาดหวังของผู้สอบทั้งหมดจาก N และค่าเฉลี่ยคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ข้อสอบ คำนวณได้จากสมการ

$$\mathrm{MSE}_{\mathrm{i}} = \frac{\sum_{k=1}^{R}(t_{ki} - \overline{\tau}_i)^2}{R}$$

เมื่อ τ_i คือ ค่าจริง (True value) ของพารามิเตอร์ข้อสอบ ข้อที่ i และ t_{ik} คือ ค่าประมาณที่ได้จากการจำลองครั้งที่ k (k = 1,..,R)

เนื่องจากวิธีการประมาณค่าที่ให้ค่าประมาณที่ดีที่สุดที่จะช่วยลดความคลาดเคลื่อนในการประมาณค่าได้ คือ การประมาณค่าแบบเบย์ ดังนั้น ผู้วิจัยจึงสนใจเลือกนำวิธีการประมาณค่าพารามิเตอร์แบบเบย์มาใช้ประมาณค่าพารามิเตอร์ในการวิจัยครั้งนี้

ตอนที่ 3 โมเดลการตอบสนองข้อสอบเชิงสุ่ม

คำว่า “สุ่ม” และ “เจาะจง” ได้นำมาใช้ในการศึกษาวิจัยเพื่อสร้างโมเดลเชิงเส้นระดับลดหลั่น แต่มีการใช้คำทั้งสองนี้ ในวัตถุประสงค์ของการใช้ต่างๆกัน เช่น การนำไปใช้กับตัวแปร ได้แก่ “ตัวแปรเจาะจง” ซึ่งสมมติว่าเป็นตัวแปรที่วัดได้ค่าที่ไม่มีความคลาดเคลื่อน และสมมติให้ค่าทุกๆค่าของตัวแปรเจาะจงในการศึกษาครั้งนั้นมีค่าเดียวกันกับค่าทุกๆค่าของตัวแปรเจาะจงในการศึกษาครั้งอื่นๆต่อไป หรือการใช้กับ “ตัวแปรสุ่ม” ซึ่งสมมติว่าค่าต่างๆของตัวแปรนี้ได้มาจากกลุ่มตัวอย่างขนาดใหญ่ จึงมีความเป็นตัวแทน หรือค่าต่างๆของตัวแปรสุ่มเป็นตัวแทนของกลุ่มตัวอย่างที่ได้จากการสุ่มค่าที่เป็นไปได้ทั้งหมดของตัวแปรนั้น จึงทำให้คาดหวังได้ว่าจะสรุปอ้างอิงผลที่ได้จากตัวแปรสุ่มนั้นไปยังค่าอื่นๆที่เป็นไปได้ทั้งหมดของตัวแปรสุ่ม (Newsom, 2009)

นอกจากนี้มีการนำคำว่า “สุ่ม” และ “เจาะจง” ไปใช้ในบริบทของโมเดลการวิเคราะห์ความแปรปรวน โมเดลการวิเคราะห์ถดถอย และโมเดลการวิเคราะห์พหุระดับ ซึ่งมีรูปแบบโมเดลที่

เป็นไปได้ 3 รูปแบบ คือ โมเดลอิทธิพลสุ่ม (Random-effect model) โมเดลอิทธิพลเจาะจง (Fixed-effect model) และโมเดลอิทธิพลผสม (Mixed-effect model) (Newsom, 2009)

โมเดลอิทธิพลสุ่ม (Random-effect model) หรืออาจเรียกอีกอย่างว่า โมเดลส่วนประกอบความแปรปรวน (variance components model) หมายถึง โมเดลเชิงสถิติที่สมมติให้ตัวแปรอิสระเป็นตัวแปรสุ่ม โดยสมมติให้กลุ่มตัวอย่างข้อมูลของตัวแปรที่นำมาวิเคราะห์ได้มาอย่างสุ่มจากประชากรข้อมูลหลายๆค่า จึงมีค่าความคลาดเคลื่อนจากค่าเฉลี่ยซึ่งมีการแจกแจงแบบสุ่มเป็นส่วนประกอบหนึ่งในโมเดล และทำให้สามารถสรุปอ้างอิงไปยังประชากรของข้อมูลทั้งหมด ซึ่งครอบคลุมไปได้มากกว่าในกลุ่มตัวอย่างข้อมูลที่นำมาวิเคราะห์ (Newsom, 2009; Glass &Stanley, 1970) ดังตัวอย่างในภาพที่ 4

โมเดลอิทธิพลเจาะจง (Fixed-effect model) หมายถึง โมเดลเชิงสถิติที่สมมติให้ตัวแปรอิสระเป็นตัวแปรคงที่ โดยสมมติให้กลุ่มตัวอย่างข้อมูลของตัวแปรที่นำมาวิเคราะห์ประกอบด้วยข้อมูลหลายๆค่าชุดหนึ่ง ในโมเดลเจาะจงจะไม่มีเทอมของความคลาดเคลื่อนจากค่าเฉลี่ยเป็นส่วนประกอบในโมเดล และทำให้สรุปอ้างอิงไปได้เพียงในกลุ่มชุดข้อมูลที่นำมาวิเคราะห์เท่านั้น (Newsom, 2009; Glass &Stanley, 1970) ดังตัวอย่างในภาพที่ 4

โมเดลอิทธิพลผสม (Mixed-effect model) หมายถึง โมเดลเชิงสถิติที่มีทั้งอิทธิพลเจาะจงและอิทธิพลสุ่มอยู่ภายในโมเดล กล่าวคือ มีตัวแปรมากกว่า 1 ตัว โดยสมมติให้ตัวแปรอิสระตัวหนึ่งเป็นตัวแปรคงที่ด้วยอิทธิพลเจาะจง และสมมติให้ตัวแปรอิสระอีกตัวหนึ่งเป็นตัวแปรสุ่มด้วยอิทธิพลสุ่ม (Newsom, 2009; Glass &Stanley, 1970)

Fixed-effect model (J=3)		Random-effect model (J=3)	
$X_{ij}=\mu+\alpha_j+e_{ij}$		$X_{ij}=\mu+\boldsymbol{a}_j+e_{ij}$	
Main effects: $\alpha_1, \alpha_2, \alpha_3$		Main effects: $\boldsymbol{a}_1, \boldsymbol{a}_2,\ldots, \boldsymbol{a}_\infty$	
Replication of experiment	Main effects present in replication	Replication of experiment	Main effects present in replication
1	$\alpha_1, \alpha_2, \alpha_3$	1	$\boldsymbol{a}_5, \boldsymbol{a}_{31}, \boldsymbol{a}_8$
2	$\alpha_1, \alpha_2, \alpha_3$	2	$\boldsymbol{a}_{16}, \boldsymbol{a}_3, \boldsymbol{a}_9$
3	$\alpha_1, \alpha_2, \alpha_3$	3	$\boldsymbol{a}_{21}, \boldsymbol{a}_{11}, \boldsymbol{a}_{50}$

ภาพที่ 4 ความเป็นตัวแทนของโมเดลอิทธิพลเจาะจงกับอิทธิพลสุ่ม (Glass & Stanley, 1970)

เมื่อ X_{ij} คือ คะแนนของผู้สอบคนที่ i ของตัวแปรระดับ j ; μ คือ คะแนนเฉลี่ยที่แท้จริง; α_j และ $\boldsymbol{a}_j$ คือ ผลต่างระหว่างคะแนนเฉลี่ยของผู้สอบทุกคนกับคะแนนเฉลี่ยที่แท้จริง ($\mu_j - \mu$) ของโมเดลอิทธิพลเจาะจง ; $\boldsymbol{a}_j$ คือ ผลต่างระหว่างคะแนนเฉลี่ยของผู้สอบทุกคนกับคะแนนเฉลี่ยที่แท้จริง ($\mu_j - \mu$) ของโมเดลอิทธิพลสุ่ม ; และ e_{ij} คือ ความคลาดเคลื่อนของคะแนนผู้สอบคนที่ j จากคะแนนเฉลี่ยของผู้สอบทั้งหมด

3.1 แนวคิดเบื้องต้นในการใช้วิธีการข้อสอบเชิงสุ่ม

แนวคิดเบื้องต้นของโมเดลการตอบสนองข้อสอบเชิงสุ่ม โดยใช้โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ที่ให้ผลการตอบแบบ 2 ค่า โดยกำหนดให้ค่าความน่าจะเป็นของผลการตอบ 1 ข้อ คือ ผลรวมของความสามารถบุคคล และค่าความง่ายของข้อสอบ หรือเป็นค่าความยากติดลบ ดังนี้

$$n_{pi} = \theta_p - \beta_i$$

เมื่อกำหนดให้ n_{pi} คือ ค่า $\ln(\Pr(Y_{pi}=1)/\Pr(Y_{pi}=0))$; $Y_{pi}=1$ เมื่อผู้สอบตอบถูกและ $Y_{pi}=0$ เมื่อตอบผิด

Y_{pi} คือ ผลการตอบสนองข้อสอบข้อที่ i ของผู้สอบ p (i= 1,…,I และ p = 1,…,P)

θ_p คือ ความสามารถของผู้สอบ p

β_i คือ ค่าความยากของข้อสอบข้อที่ i

การนำวิธีการข้อสอบเชิงสุ่มมาใช้ ด้วยเหตุผลที่สำคัญ 2 ประการ คือ

ประการแรก การจัดกระทำข้อสอบเป็นแบบสุ่ม (Treating Elements as Random)

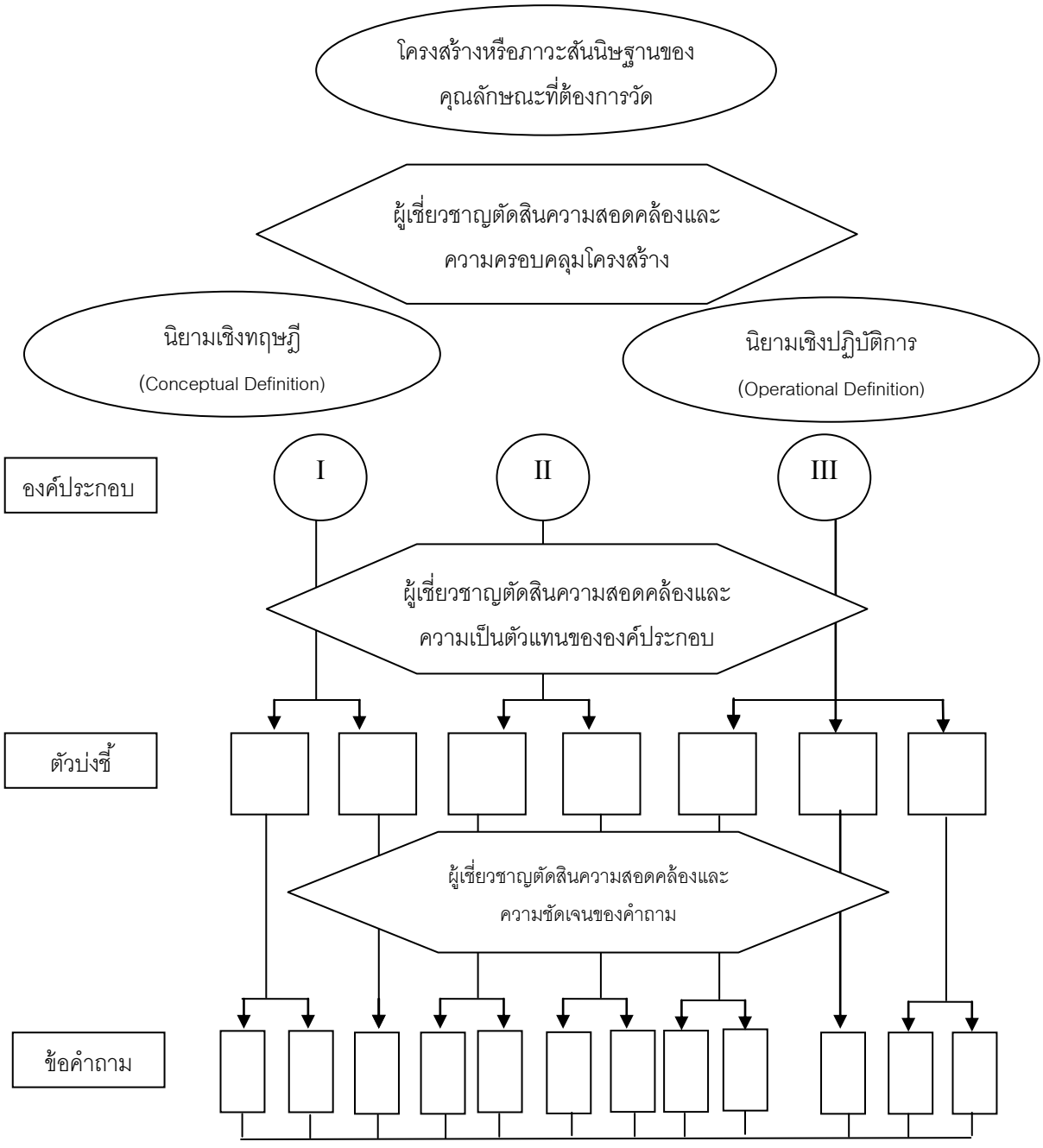
คำว่า “ประชากร” มักจะใช้ในความหมายของประชากรผู้สอบ ประชากรส่วนใหญ่ที่เรารู้จักมักแบ่งออกเป็นกลุ่มประชากรย่อยๆ ที่กำหนดคุณลักษณะเฉพาะของกลุ่มประชากรย่อยเพื่อให้เป็นไปตามข้อตกลงเบื้องต้นของการกระจายข้อมูลแบบปกติ เช่น การกำหนดคุณลักษณะต่างๆ บางครั้งจะขึ้นอยู่กับ เชื้อชาติ เพศ การศึกษา ความชำนาญ และลักษณะอื่นๆของบุคคล เมื่อมีการออกแบบการสุ่มอย่างดีจะทำให้ค่าความแปรปรวนเชิงสุ่มยังเหลืออยู่บ้างแต่ก็น้อยมากจนไม่ค่อยสำคัญนัก นั่นคือ ช่วงระหว่างประชากรต่างๆมีขนาดค่อนข้างเล็ก ถ้าพิจารณาในเชิงเปรียบเทียบกับข้อสอบแล้ว ประชากรของผู้สอบก็มีการกำหนดเหมือนกันกับกลุ่มข้อสอบ (item families) ที่มีลักษณะเหมือนกันภายในกลุ่ม

การจัดกระทำข้อสอบเป็นแบบสุ่ม (Treating Elements as Random) คือ การสุ่มเอาข้อสอบจากประชากรข้อสอบย่อยๆหลายๆกลุ่มออกมาเป็นกลุ่มเฉพาะ กล่าวคือในการออกข้อสอบเพื่อวัดคุณลักษณะแฝงหนึ่งๆนั้น คุณลักษณะดังกล่าวจะประกอบด้วยคุณลักษณะย่อยหลายๆด้าน ดังภาพที่ 5

องค์ประกอบของคุณลักษณะที่ต้องการวัดแต่ละด้านเทียบได้กับประชากรย่อย ในกลุ่มประชากรย่อยของข้อสอบนี้จะมีข้อสอบที่จะวัดคุณลักษณะที่คล้ายคลึงกันภายในกลุ่มแต่จะแตกต่างกันระหว่างกลุ่มประชากรย่อย เมื่อสุ่มข้อสอบจากกลุ่มประชากรย่อยของข้อสอบ ข้อสอบเหล่านี้เป็นองค์ประกอบที่ถูกดึงออกจากประชากรข้อสอบแบบสุ่ม ทำให้ข้อมูลมีลักษณะของการกระจายแบบปกติเป็นไปตามข้อตกลงเบื้องต้นทั่วไปของการกระจายข้อมูล

การสุ่มผู้สอบกับข้อสอบมีความแตกต่างที่สำคัญประการหนึ่งคือ ประชากรของผู้สอบมีอยู่ก่อนการสุ่ม แต่ประชากรข้อสอบมิได้มีอยู่ก่อนการสุ่ม ยกเว้นกรณีการสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ที่มีการสร้างคลังข้อสอบขึ้นมาเป็นประชากรของข้อสอบ นั่นคือ สถานการณ์การทดสอบโดยทั่วไปเมื่อผู้สอนต้องการตรวจสอบคุณสมบัติของผู้เรียนหลังเรียน ก็จะออกข้อสอบขึ้นมาชุดหนึ่งเพื่อวัดผลการเรียนรู้ที่เกิดขึ้นหลังเรียน ประชากรผู้สอบจึงมีอยู่ก่อนการสอบ ในขณะที่ข้อสอบที่เขียนจะขึ้นมาเพื่อวัดคุณสมบัติผู้เรียนตามวัตถุประสงค์การเรียนรู้ และนำข้อสอบที่เขียนขึ้นมาใช้ทั้งหมดในการสอบครั้งนั้นๆ โดยที่ผู้สอนอาจมิได้นำมาตรวจสอบหาคุณภาพของ

ข้อสอบก่อนที่จะนำไปใช้ ซึ่งเป็นการใช้อิทธิพลแบบเจาะจงข้อสอบ ทำให้ไม่สามารถสรุปอ้างอิงไปยังประชากรข้อสอบได้ หากแต่ถ้ามีการสุ่มข้อสอบจากกลุ่มประชากรย่อยของข้อสอบในคลังข้อสอบที่ผ่านการตรวจสอบคุณภาพแล้วมาจัดเป็นแบบทดสอบ แบบสอบชุดที่ใช้อิทธิพลแบบสุ่มข้อสอบนี้ ก็จะมีความเป็นตัวแทนของประชากรข้อสอบ และสามารถสรุปอ้างอิงไปยังประชากรข้อสอบได้ รวมทั้งสรุปอ้างอิงคุณลักษณะแฝงที่ต้องการวัดได้อย่างน่าเชื่อถือ



ภาพที่ 5 ขั้นตอนการสร้างแบบสอบ (ศิริชัย กาญจนวาสี, 2552)

การสร้างข้อสอบอาจพิจารณาได้ว่าเป็นการสุ่มมาจากประชากรทางทฤษฎี เนื่องจากในการสร้างข้อสอบแต่ละครั้งคือการสุ่มหนึ่งครั้ง ทั้งนี้การออกข้อสอบเพื่อการวัดคุณลักษณะแฝงใดๆ การเขียนข้อสอบชุดหนึ่งๆ อาจมาจากผู้ออกข้อสอบหลายๆ คนซึ่งแต่ละคนมีแนวคิดบางอย่างอยู่ในใจของผู้เขียนข้อสอบแต่ละคนที่อาจจะไม่ได้แสดงให้เห็นอย่างชัดเจน แต่ในอีกกรณีหนึ่งเมื่อข้อสอบถูกสร้างขึ้นโดยอัตโนมัติ เช่น ในระหว่างกระบวนการสร้างข้อสอบของการทดสอบแบบปรับเหมาะจะใช้แนวคิดที่มีความชัดเจนในการสร้าง การกำหนดคุณลักษณะต่างๆในแนวคิดแสดงถึงประชากรเหล่านั้น แนวคิดหนึ่งๆอาจนำไปสู่ประชากรที่แตกต่างกันขึ้นอยู่กับคุณลักษณะของแนวคิดที่บางครั้งจะระบุเฉพาะเจาะจงมาก จนเกือบจะเป็นข้อสอบเฉพาะข้อเดียว เช่น ในการสร้างสำเนาข้อสอบ (item cloning) ชุดของข้อสอบที่ถูกสร้างขึ้นอย่างเป็นไปได้ คือ ประชากรของข้อสอบ

ประการที่สอง สาเหตุของการจัดกระทำองค์ประกอบเป็นแบบสุ่ม

ในกระบวนการการประมาณค่าพารามิเตอร์ จะมีความคลาดเคลื่อนในการประมาณค่าเกิดขึ้นด้วย พัฒนากระบวนการประมาณค่าพารามิเตอร์ได้ถูกพัฒนาอย่างต่อเนื่องเพื่อให้ความคลาดเคลื่อนในการประมาณค่าลดลง โดยนำแนวคิดการกระจายก่อนหน้า (prior distribution) ของพารามิเตอร์นั้นมาใช้เป็นค่าตั้งต้น การกระจายก่อนหน้าของพารามิเตอร์ คือ ค่าความแปรปรวนที่เป็นไปได้ของพารามิเตอร์นั้น ซึ่งได้จากข้อมูลในอดีตของพารามิเตอร์ตัวนั้น ความคลาดเคลื่อนจะลดลงหลังจากเมื่อเราได้สารสนเทศของข้อมูลมาแล้ว แต่เมื่อการกระจายก่อนหน้าถูกนำมาใช้กับพารามิเตอร์ที่ไม่ทราบค่า เช่น ความแปรปรวนของความสามารถผู้สอบจะทำให้สามารถประมาณค่าการกระจายของความแปรปรวนนั้นได้ การกระจายนี้เทียบเท่ากับการกระจายของความแปรปรวนก่อนหน้า อีกทั้งยังเทียบเท่ากับการกระจายของประชากร ซึ่งเป็นองค์ประกอบที่เสมือนได้จากการสุ่ม ในวิธีแบบเบย์นั้นการกระจายของพารามิเตอร์ต่างๆนี้ได้ถูกกำหนดค่าการกระจายก่อนหน้าเอาไว้

ดังนั้นโมเดลข้อสอบเชิงสุ่มที่มีการจัดกระทำข้อสอบแบบสุ่มจึงมีความเหมาะสมทั้งด้านหลักการและเชิงทฤษฎีที่จะนำมาใช้ในการประมาณค่าพารามิเตอร์ความสามารถผู้สอบ

3.2 แนวคิดและวิธีการข้อสอบเชิงสุ่ม

การนำโมเดลข้อสอบเชิงสุ่มนำมาวิเคราะห์การวัดทางจิตมิติในเชิงปฏิบัติใน 3 ประเด็นหลักๆ คือ การอธิบายความยากของข้อสอบ การทำหน้าที่ต่างกันของข้อสอบ และการวัดความสามารถของผู้สอบ ดังนี้

3.2.1 การอธิบายความยากของข้อสอบ (Explanation of Item Difficulties)

ตามปกติแล้วจะใช้โมเดลการทดสอบโลจิสเชิงเส้น (Linear Logistic Test Model ; LLTM) ที่มีการนิยามถึงความยากไว้ว่าเป็นฟังก์ชันเชิงเส้นตรงของคุณลักษณะข้อสอบที่แสดงอยู่ในรูปเมตริกคุณลักษณะเพื่ออธิบายความยากของข้อสอบ สิ่งเหล่านี้อาจจะเป็นปัจจัยในการออกแบบข้อสอบ แต่ตามปกติแล้วจะมีตัวแปรร่วมประเภทอื่นๆอยู่บ้าง ดังสมการ

$$\beta_i = \sum_{q=1}^{Q} \beta_q X_{iq}$$

เมื่อ X_{iq} คือ ค่าผลการตอบของข้อสอบข้อที่ i ที่มีคุณลักษณะ q

และ β_q คือ น้ำหนักของคุณลักษณะ q (q = 1,…,Q)

แต่โมเดลข้างต้น ค่อนข้างไม่เป็นไปตามความเป็นจริง เนื่องจากโมเดลดังกล่าวแสดงถึงการอธิบายค่าความยากของข้อสอบได้อย่างสมบูรณ์โดยไม่มีความคลาดเคลื่อนเกิดขึ้นเลย ซึ่งในความเป็นจริงแล้วกรณีเช่นนี้ไม่สามารถเกิดขึ้นได้ ดังนั้น ถ้าเราเพิ่มตัวแปรความคลาดเคลื่อนเข้าไปในโมเดล ก็จะทำให้ได้โมเดลที่มีความสมบูรณ์น่าเชื่อถือยิ่งขึ้น

$$\beta_i = \sum_{q=1}^{Q} \beta_q X_{iq} + \varepsilon_i$$

เมื่อ ε_i คือ ค่าความคลาดเคลื่อนที่มีการกระจายแบบปกติ $\varepsilon_i \approx N(\mu_\varepsilon, \sigma_\varepsilon^2)$

ดังนั้น ในความเป็นจริงแล้วข้อสอบต่างๆจะถูกจัดกระทำเป็นแบบสุ่ม ซึ่งการรวมกันแต่ละครั้งของค่าผลการตอบของข้อสอบ หมายถึง ประชากรของข้อสอบ

3.2.2 การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning; DIF)

การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความน่าจะเป็นของการตอบข้อสอบข้อใดข้อหนึ่งได้ถูกต้อง ไม่เพียงขึ้นอยู่กับความสามารถของบุคคลแต่ยังขึ้นอยู่กับคุณลักษณะกลุ่มของบุคคลนั้นด้วย การทำหน้าที่ต่างกันของข้อสอบ (DIF) มีสมการ ดังนี้

$$\beta_i = \beta_{i0} + g_p \delta_i$$

เมื่อ β_{i0} คือ ค่าความยากเฉพาะของข้อสอบข้อที่ i ในกลุ่มอ้างอิง (reference group)

β_{i1} คือ ค่าความยากเฉพาะของข้อสอบข้อที่ i ในกลุ่มเปรียบเทียบ (focal group)

g_p คือ ค่าดัชนีประจำกลุ่มที่บุคคลนั้นอยู่ กลุ่มอ้างอิง = 0 กลุ่มเปรียบเทียบ = 1

และ δ_i คือ ค่าเฉพาะของการทำหน้าที่ต่างกันของข้อสอบ เป็นความแตกต่างระหว่าง β_{i0} กับ β_{i1}

ถ้า มีค่าเท่ากับ 0 แสดงว่าข้อสอบข้อนั้นไม่ได้ทำหน้าที่ต่างกัน (No DIF)

การหาค่า DIF สามารถใช้ได้ทั้งดัชนีหรือใช้โมเดลการตอบสนองข้อสอบ ด้วยกลวิธีที่นิยมใช้กันแพร่หลาย 2 วิธี คือ (1) กลวิธีข้อสอบร่วม (Anchoring Strategy) ที่ใช้ทั้งดัชนีและโมเดลการตอบสนองข้อสอบ (2) กลวิธีที่ไม่ยึดพารามิเตอร์ (Free Parameters Strategy) ที่ใช้กับโมเดลการตอบสนองข้อสอบเท่านั้น แต่กลวิธีทั้งสองนี้ต่างก็มีข้อเสีย คือ การวิเคราะห์ค่า DIF ของกลวิธีข้อสอบร่วม จะขึ้นอยู่กับคุณภาพของแบบสอบร่วมและสอดคล้องกับข้อตกลงเบื้องต้น ในขณะที่กลวิธีที่ไม่ยึดพารามิเตอร์ จะมีพารามิเตอร์ที่ต้องการจำนวนมากมาย และยังมีโอกาสเสี่ยงต่อการลงทุนมากกว่า

วิธีการอื่นที่จะนำเสนอ 2 วิธี ซึ่งใช้เทคนิคข้อสอบเชิงสุ่ม คือ (1) วิธีสองตัวแปรด้วยความยากเชิงสุ่มในทั้งกลุ่มควบคุมและกลุ่มเปรียบเทียบ การสรุปผลโดยรวมของ DIF ด้วยโมเดล 3 พารามิเตอร์ ซึ่งประกอบด้วย ความแปรปรวน 2 พารามิเตอร์ และความแปรปรวนร่วม 1 พารามิเตอร์ ซึ่งอธิบายได้ด้วย random item profiles model (RIP Model) (2) ตัวแปรข้อสอบแฝงแบบให้คะแนน 2 ค่า (A Binary Latent Item Variable) ซึ่งใช้ในการหาความแตกต่างระหว่างข้อสอบที่มี DIF กับข้อสอบที่ไม่มี DIF ซึ่งสามารถอธิบายได้ด้วย random item mixture model (RIM Model) ดังสมการ

$$\beta_{ig} = (1 - \alpha_i g_p)\beta_{i0} + \alpha_i g_p \beta_{i1}$$

เมื่อ $(\beta_{i0}, \beta_{i1}) \approx BVN(\mu_{\beta_0}, \mu_{\beta_1}, \sum \beta_0 \beta_1), \alpha_i \approx Bernoulli(\pi_\alpha)$

เมื่อ β_{ig} คือ ความยากของข้อสอบข้อที่ i ในกลุ่ม g

α_i คือ ตัวแปรแฝงแบบให้คะแนน 2 ค่า โดยมีค่า = 1 เมื่อมี DIF และมีค่า = 0 เมื่อไม่มี DIF

$\mu_{\beta_0}, \mu_{\beta_1}$ คือ ค่าความยากเฉลี่ยของกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ

$\sum \beta_0 \beta_1$ คือ เมตริกซ์ความแปรปรวนร่วมของค่าความยาก

นั่นคือ จะมีการกระจายของตัวแปรแบบให้คะแนน 2 ค่า ก็ต่อเมื่อ ค่า α_i = 1

3.2.3 การวัดความสามารถของผู้สอบ

โดยทั่วไปแล้ว เราจะทำการวัดความสามารถของผู้สอบโดยใช้ผลการตอบสนองข้อสอบของผู้สอบ ซึ่งมักจะใช้วิธีการเจาะจงข้อสอบและสุ่มผู้สอบ และใช้วิธีการประมาณค่าความสามารถของผู้สอบด้วยวิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าปลายทาง (Marginal maximum likelihood method) ดังสมการ

$$n_{pi}|\theta_p = \theta_p - \beta_i \qquad \text{เมื่อ} \quad \theta_p \approx N(\mu_\theta, \sigma_\theta^2)$$

เมื่อ $\mu_\theta, \sigma_\theta^2$ คือ ค่าเฉลี่ยของความสามารถ และค่าความแปรปรวนของความสามารถ

การประมาณค่าโมเดลนี้ได้มาจากค่าประมาณความยากของข้อสอบ และค่าประมาณค่าเฉลี่ยและความแปรปรวนของการกระจายของบุคคล แต่ด้วยวิธีการดังกล่าวยังมีข้อบกพร่องที่มากับวัตถุประสงค์ของการใช้วิธีการนี้อยู่ 2 ประการ คือ 1) วิธีการนี้ไม่ได้วัดบุคคลแต่ละคน แต่เป็นการวัดข้อสอบแต่ละข้อ เนื่องจากวิธีการนี้เป็นการสุ่มบุคคลแต่เจาะจงข้อสอบ จึงทำให้เมื่อประมาณค่าแล้วจะได้ค่าความยากของข้อสอบแต่ละข้อ ซึ่งไม่สามารถสรุปอ้างอิงไปยังประชากรข้อสอบได้ อีกทั้งการใช้อิทธิพลแบบสุ่มบุคคลทำให้ผลการประมาณค่าที่ได้ไม่ใช่เป็นความสามารถของผู้สอบแต่ละคน แต่เป็นการสรุปอ้างอิงไปยังความสามารถของประชากรผู้สอบ 2) วิธีการนี้จะใช้การยืมค่าตัวมันเองเพื่อการสรุปอ้างอิงความสามารถของบุคคลเพื่อการวัดข้อสอบ โดยนำค่าการกระจายก่อนหน้าของพารามิเตอร์ความสามารถผู้สอบมาเป็นค่าตั้งต้นเพื่อประมาณค่าความยากของข้อสอบ แต่ความจริงแล้วสิ่งที่เราต้องการคือการสรุปอ้างอิงข้อสอบเพื่อวัดความสามารถของบุคคล วิธีการที่แก้ไขข้อบกพร่องนี้คือการใช้อิทธิพลแบบเจาะจงบุคคลและอิทธิพลข้อสอบเชิงสุ่ม ซึ่งอาจเขียนเป็นสมการของโมเดลการวัดบุคคลได้ ดังนี้

$$n_{pi}|\beta_i = \theta_p - \beta_i \qquad \text{เมื่อ} \quad \beta_i \approx N(\mu_\beta, \sigma_\beta^2) or U(0,1)$$

เมื่อ μ_β คือ ค่าเฉลี่ยของความยากของข้อสอบ
และ σ_β^2 คือ ค่าความแปรปรวนความยากของข้อสอบ

โมเดลการวัดบุคคลที่ใช้อิทธิพลแบบเจาะจงบุคคล และอิทธิพลข้อสอบเชิงสุ่มนี้ จะทำให้สามารถประมาณค่าความสามารถของบุคคลได้โดยตรงและเพียงพอที่จะสรุปอ้างอิงผลการวัดนี้ให้ครอบคลุมข้อสอบข้อต่างๆ

ประเด็นทั้ง 3 ข้อข้างต้น เป็นประเด็นที่สำคัญในการวัดทางจิตมิติ ที่ประกอบด้วย การวัดเพื่อการอธิบาย การวัดเพื่อระบุปัญหา และการวัดเพื่อให้ได้ผลที่แท้จริง **การวัดเพื่อการอธิบาย** เป็นประเด็นหนึ่งที่มีความสำคัญ เพราะการอธิบายเป็นวัตถุประสงค์หลักทางวิทยาศาสตร์ และเป็นพื้นฐานของการปฏิบัติทุกๆประเภท เพราะว่าถ้าสิ่งที่เราสร้างขึ้นมาไม่สามารถนำมาอธิบายได้ เราก็อาจหันมาใช้วิธีการลองผิดลองถูกแทน การอธิบายการวัดคือการรวมการวัดและการอธิบายเพื่อที่จะบอกให้ได้ว่าวัดอะไรและอธิบายได้อย่างไรในเวลาเดียวกัน (De Boeck & Wilson, 2005) **การวัดเพื่อระบุปัญหา** หรือการระบุปัญหาผ่านการวัด และปัญหาที่สำคัญมากที่สุดโดยใช้คุณภาพของการวัด วิธีการระบุปัญหาส่วนใหญ่เป็นสิ่งที่ต้องค้นคว้าด้วยตนเอง โดยใช้ตัวบ่งชี้ที่ทราบคุณสมบัติทางสถิติและไม่ทราบคุณสมบัติทางสถิติ ซึ่งเป็นตัวบ่งชี้ที่สำคัญในการนำมาใช้วิเคราะห์ DIF การสร้างรูปแบบที่บ่อยครั้งไม่ได้เป็นทางเลือกที่ดีสำหรับการระบุปัญหาเกิดขึ้น เนื่องจากโมเดลที่ใช้ส่วนใหญ่ไม่สามารถระบุปัญหาได้โดยตรง โดยมักจะขึ้นอยู่กับแต่ละกรณี แต่เมื่อโมเดลการวัดถูกสร้างขึ้นภายหลังจากระบุปัญหาแล้ว โมเดลการวัดยังอาจช่วยในการระบุปัญหาได้อีกด้วย **การวัดเพื่อให้ได้ผลที่แท้จริง**เป็นประเด็นที่สำคัญอีกประเด็นหนึ่งเพราะการตัดสินใจต่างๆจะอยู่บนพื้นฐานของผลที่ได้จากการวัด เช่น การวัดผลทางการศึกษา ในการศึกษาครั้งนี้จึงสนใจประเด็นการวัดผลเพื่อให้ได้ผลที่แท้จริงทางการศึกษา เพื่อจะให้ทราบถึงความสามารถที่แท้จริงของบุคคล นำไปสู่การประเมินผลความสามารถของบุคคลเพื่อการตัดสินผลการศึกษาหรือการคัดเลือกเข้าศึกษาต่อ ซึ่งเป็นการตัดสินที่มีผลกระทบต่อบุคคลเป็นอย่างมาก

3.3 โมเดลการวิเคราะห์

โมเดลการตอบสนองข้อสอบสามารถจัดกระทำข้อสอบและผู้สอบได้ทั้งแบบอิทธิพลเจาะจงหรืออิทธิพลสุ่ม จึงสามารถสร้างโมเดลการตอบสนองข้อสอบได้เป็น 4 แบบ ที่แตกต่างกันดังนี้

แบบที่ 1 โมเดลการตอบสนองข้อสอบแบบเจาะจงผู้สอบ-เจาะจงข้อสอบ (FPFI)

แบบที่ 2 โมเดลการตอบสนองข้อสอบแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI)

แบบที่ 3 โมเดลการตอบสนองข้อสอบแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI)

แบบที่ 4 โมเดลการตอบสนองข้อสอบแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI)

ตารางที่ 2.3 โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ จำแนกตามการสุ่มหรือการเจาะจงผู้สอบและข้อสอบ

	ข้อสอบ	
ผู้สอบ	เจาะจง	สุ่ม
เจาะจง	FPFI	FPRI
สุ่ม	RPFI	RPRI

โมเดลที่ใช้โดยทั่วไป 2 โมเดล คือ โมเดล FPFI และ โมเดล RPFI โมเดลแรกใช้วิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าร่วมกัน (Joint Maximum Likelihood ; JML) และโมเดล RPFI ใช้วิธีการประมาณค่าที่เป็นไปได้สูงสุดแบบประมาณค่าปลายทาง (Marginal Maximum Likelihood ; MML) แต่วิธีการประมาณค่า แบบ JML นี้ ยังคงมีปัญหาที่เกิดขึ้นอย่างต่อเนื่อง จึงได้มีการพัฒนาวิธีการประมาณค่าแบบ MML ขึ้นมาช่วยในการประมาณค่าได้ดียิ่งขึ้น วิธีการประมาณค่าแบบ JML และ MML เป็นวิธีที่นิยมใช้ในการประมาณค่า เนื่องจากความสนใจที่จะประมาณค่าพารามิเตอร์ข้อสอบ ดังนั้นข้อตกลงเบื้องต้นของผู้สอบที่นำมาพิจารณา จะถูกใช้เป็นเครื่องมือในการประมาณค่าพารามิเตอร์ข้อสอบ นอกจากนี้มีวิธีการประมาณค่าแบบดั้งเดิมอีกแบบหนึ่งคือ การประมาณค่าที่เป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood ; CML) เป็นวิธีที่ไม่มีข้อตกลงเบื้องต้นของบุคคลที่ต้องคำนึงถึง แต่ขึ้นอยู่กับคะแนนรวมของพารามิเตอร์ผู้สอบที่ได้มาจากโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ในความเป็นจริงแล้วข้อมูลส่วนใหญ่มีสารสนเทศเกี่ยวกับพารามิเตอร์ข้อสอบมากกว่าพารามิเตอร์ผู้สอบ จึงอาจเป็นเหตุผลสำคัญของการมุ่งความสนใจไปที่การประมาณค่าข้อสอบ

ส่วนโมเดลอีก 2 โมเดลที่ไม่ค่อยนิยมใช้ คือ โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (The Fixed Persons-Rondom Items Response Model; FPRI) และโมเดลสุ่มผู้สอบ-สุ่มข้อสอบ (The Rondom Persons-Rondom Items Response Model; RPRI) โดยที่โมเดล FPRI มีการประมาณค่าโดยใช้วิธีการประมาณค่าแบบ MML เช่นเดียวกันกับโมเดล RPFI แต่มีการเปลี่ยนบทบาทระหว่างผู้สอบกับข้อสอบ ส่วนโมเดล RPRI เป็นโมเดลที่มีอิทธิพลเชิงสุ่มแบบไขว้ คือมีทั้งอิทธิพลของผู้สอบเชิงสุ่มและข้อสอบเชิงสุ่ม ซึ่งจะต้องใช้วิธีการประมาณค่าที่แตกต่างออกไปจากโมเดลอื่นๆ

วิธีการประมาณค่าที่พัฒนาขึ้นเพื่อใช้กับโมเดล RPRI คือวิธีการประมาณค่าแบบเชื่อมโยงสลับภายหลัง (Alternating Imputation Posterior ; AIP) เป็นการหาพื้นที่แบบปรับเหมาะของ

Gauss-Hermite (Adaptive Gauss-Hermite Quadrature) วิธีการคำนวณด้วย AIP โดยเริ่มจากการแบ่งโมเดลรวมออกเป็นโมเดลย่อย เรียกว่า ปีก (wings) สำหรับผู้สอบและข้อสอบ เพื่อการคำนวณสลับกันระหว่างโมเดลปีกทั้งสองจนกระทั่งค่าที่ได้เบนเข้าหากันหรือมาบรรจบกัน ณ จุดหนึ่ง การคำนวณทวนซ้ำระหว่างโมเดลปีกของข้อสอบข้อหนึ่ง ซึ่งค่าเฉลี่ยและความแปรปรวนของข้อสอบข้อนั้นได้ประมาณค่าเพื่อให้ได้ค่าอิทธิพลของผู้สอบ กับการคำนวณทวนซ้ำโมเดลปีกความสามารถผู้สอบคนหนึ่ง ซึ่งค่าเฉลี่ยและความแปรปรวนของผู้สอบนั้นได้ประมาณค่าเพื่อให้ได้ค่าอิทธิพลของข้อสอบ การประมาณค่าแบบนี้เป็นการประมาณค่าที่เป็นไปได้สูงสุด (Maximum Likelihood) และเป็นการหาค่าพื้นที่แบบปรับเหมาะ (Adaptive Quadrature) อิทธิพลของผู้สอบที่นำมาใช้กับโมเดลปีกของข้อสอบได้มาจากการสุ่มของการกระจายภายหลัง (Posterior Distribution) ที่ประมาณค่ามาจากโมเดลปีกของผู้สอบ และอิทธิพลของข้อสอบที่นำมาใช้กับโมเดลปีกของผู้สอบก็ได้มากจากการสุ่มของการกระจายภายหลังที่ประมาณค่ามาจากโมเดลปีกของข้อสอบเช่นเดียวกัน

แม้ว่ารูปแบบที่เป็นไปได้ในการจัดกระทำข้อสอบและผู้สอบ จะมี 4 รูปแบบ ดังตารางที่ 2.3 แต่ในการวิจัยครั้งนี้ผู้วิจัยสนใจศึกษาเพียง 3 โมเดล คือ โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ แบบสุ่มบุคคล-เจาะจงข้อสอบ (RPFI) โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ แบบเจาะจงบุคคล-สุ่มข้อสอบ (FPRI) และโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ แบบสุ่มบุคคล--สุ่มข้อสอบ (RPRI) เนื่องจากเป็นโมเดลที่สอดคล้องกับบริบทการทดสอบในประเทศ น่าจะเป็นประโยชน์ในการนำไปใช้กับการทดสอบในห้องเรียนและการทดสอบระดับชาติต่อไป

ในการวิจัยครั้งนี้จะใช้โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ มีรูปฟังก์ชันโลจิสเป็น $P_i(\theta)=\frac{1}{1+e^{-D(\theta-b_i)}}$ ในการประมาณค่าพารามิเตอร์ความสามารถผู้สอบและข้อสอบจากโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ที่มีการสุ่มที่แตกต่างกัน 3 โมเดล คือ 1) แบบ Random person - Fixed item (RPFI) 2) แบบ Fixed person - Random item (FPRI) 3) แบบ Random person - Random item (RPRI) โดยกำหนดให้ลักษณะการแจกแจงของพารามิเตอร์ θ_p มีการแจกแจงปกติ ~N (0,1) และ β_i มีการแจกแจง 4 ลักษณะ คือ Normal (0,1), Uniform (-2,2) การแจกแจงเบ้ซ้าย และการแจกแจงเบ้ขวา

ตอนที่ 4 โปรแกรมการจำลองข้อมูลและวิเคราะห์ข้อมูล

การดำเนินการวิจัยครั้งนี้ดำเนินการวิจัยโดยใช้ข้อมูลที่ได้จากการจำลองด้วยโปรแกรมคอมพิวเตอร์ ซึ่งปัจจุบันนี้มีโปรแกรมหลากหลายโปรแกรมที่สามารถจำลองข้อมูลเชิงปริมาณได้ อาทิเช่น Excel , MATLAB, R เป็นต้น ในงานวิจัยครั้งนี้ผู้วิจัยสนใจนำโปรแกรม R มาใช้ในการจำลองข้อมูล เนื่องจากความเหมาะสมและสะดวกในการนำไปวิเคราะห์ต่อเนื่องด้วยโปรแกรม WINBUGS หลังจากการจำลองข้อมูลด้วยโปรแกรม R

4.1 โปรแกรม R

โปรแกรม R เป็นโปรแกรมสำเร็จรูปฟรีแวร์ ที่เปิดให้บุคคลทั่วไปใช้ได้โดยไม่ต้องเสียค่าใช้จ่าย โดยมูลนิธิ R Development Core Team ได้เผยแพร่โปรแกรม R ผ่าน http://www.r-project.org โปรแกรม R ถูกเขียนโดย Robert Gentlemen และ Ross Ihaka จาก University of Auckland ประเทศนิวซีแลนด์ และตั้งแต่ปี ค.ศ.1997 เป็นต้นมา ก็ได้รับการพัฒนาอย่างต่อเนื่องโดยกลุ่มนักสถิติและนักคอมพิวเตอร์ นำโดย John Chambers ที่ Bell Laboratories ที่ในขณะนั้นปัจจุบันเป็นที่รู้จักกันในนามของ Lucent Technologies

โปรแกรม R นี้มีความสามารถไม่ด้อยไปกว่าโปรแกรมสำเร็จรูปทางสถิติอื่นๆ อีกทั้งสามารถดาวน์โหลด และติดตั้งได้ง่าย รวมถึงการปรับปรุงความสามารถทางการวิเคราะห์ของโปรแกรมก็สามารถทำได้โดยการดาวน์โหลดแพกเก็จ (package) ของคำสั่งที่ต้องการมาติดตั้งในคอมพิวเตอร์

โปรแกรม R ได้รับการพัฒนาโดยนักสถิติเป็นจำนวนมากอย่างต่อเนื่อง มีรุ่น (version) ใหม่ออกมาอย่างสม่ำเสมอ นักสถิติได้เขียนฟังก์ชันต่างๆที่ใช้วิเคราะห์ข้อมูลทางสถิติเพิ่มเติมให้กับโปรแกรมนี้ตลอดเวลา และผู้ใช้สามารถปรับปรุงโปรแกรม R รุ่นเก่าได้โดยการเพิ่มเติมแพกเก็จ (package) ของคำสั่งใหม่ๆได้เรื่อยๆ สามารถอัพเดทได้ค่อนข้างบ่อย ซึ่งเป็นการอัพเดทแบบย่อย จึงไม่กระทบต่อโครงสร้างหลักของโปรแกรมมากนัก จึงไม่จำเป็นต้องเปลี่ยนโปรมแกรมรุ่นใหม่ตลอด ยกเว้นว่าแพกเก็จคำสั่งที่ต้องการใช้นั้นบังคับให้ต้องใช้โปรแกรม R รุ่นนั้นๆ

ความสามารถของโปรแกรม R ในการวิเคราะห์ข้อมูลทางสถิติ ได้แก่ การวิเคราะห์ข้อมูลเบื้องต้นด้วยสถิติบรรยาย (การแจกแจงความถี่, ค่าสถิติเบื้องต้น, กราฟ) การประมาณค่าและการทดสอบสมมติฐานทางสถิติ การตรวจสอบคุณภาพเครื่องมือ การทดสอบความแตกต่างของค่าเฉลี่ย (สำหรับตัวอย่าง 1 กลุ่ม ถึง หลายกลุ่ม) การทดสอบความแตกต่างของค่าสัดส่วน

(สำหรับตัวอย่าง 1 กลุ่ม ถึง หลายกลุ่ม ; ข้อมูลมีค่าเป็นไปได้ 2 ค่า หรือมากกว่า 2 ค่า ; กลุ่มตัวอย่าง 2 กลุ่ม เป็นอิสระจากกัน หรือไม่เป็นอิสระจากกัน) การทดสอบความแตกต่างด้วยวิธีการนันพาราเมตริก การตรวจสอบความสัมพันธ์ระหว่าง 2 ตัวแปร การวิเคราะห์ถดถอยอย่างง่าย การวิเคราะห์ถดถอยพหุคูณ เป็นต้น (ศิริชัย พงษ์วิชัย, 2552) โปรแกรม R สามารถอ่านข้อมูลจากโปรแกรมอื่นได้ เช่น Excel SAS SPSS โดยอ่านเป็น ASCII format (Paradis, 2005) อีกทั้งสามารถนำไปเชื่อมโยงใช้กับโปรแกรมอื่นได้ เช่น WINBUGS นอกจากนี้ โปรแกรม R ยังสามารถใช้จำลองข้อมูลเพื่อนำไปใช้ในการวิเคราะห์ต่อไปได้

4.2 การจำลองข้อมูลด้วยโปรแกรม R

ในการศึกษาวิเคราะห์ระบบหรือการแก้ปัญหาโดยใช้วิธีเชิงปริมาณ มี 2 วิธีการ คือ วิธีการเชิงวิเคราะห์ (Analytical methods) หรือวิธีการเชิงคณิตศาสตร์ (Mathematical methods) และวิธีการจำลอง (Simulation methods) ซึ่งเป็นวิธีการเชิงตัวเลข (Numerical method) ที่มีการศึกษาด้วยการทดลองซ้ำๆ กับตัวแบบจำลอง (มานพ วราภักดิ์, 2550) ซึ่งในการศึกษานี้จะเป็นตัวแบบนามธรรมที่เป็นตัวแบบคณิตศาสตร์ (Mathematical models) ในการทดลองซ้ำๆ จะไม่ใช้ข้อมูลเชิงประจักษ์ที่ต้องเก็บข้อมูลหลายครั้ง หลายเงื่อนไข แต่จะใช้วิธีที่สะดวกและง่ายกว่า คือ การสุ่มตัวเลข โดยใช้ตัวแบบจำลองเชิงคณิตศาสตร์ โดยที่ตัวแบบจำลองมี 3 ประเภท คือ ตัวแบบจำลองแบบสถิต (Static model) ตัวแบบจำลองแบบพลวัต (Dynamic model) ตัวแบบจำลองเชิงกำหนด (Deterministic model) และตัวแบบความน่าจะเป็น (Probabilistic model) ซึ่งการศึกษาครั้งนี้เป็นตัวแบบจำลองแบบความน่าจะเป็น ซึ่งเป็นตัวแบบจำลองระบบที่เกี่ยวข้องกับความไม่แน่นอน (Probabilistic หรือ Stochatic simulation model) ที่จะให้ผลลัพธ์เป็นเชิงสุ่ม (random outputs) ผลลัพธ์หรือค่าต่างๆ ที่ได้นี้มีความไม่แน่นอน จึงต้องหาผลลัพธ์ในรูปของค่าประมาณ เช่น ค่าเฉลี่ย ในการจำลองข้อมูลนั้นไม่จำเป็นที่จะต้องทราบค่าพารามิเตอร์ที่แท้จริง เพียงแต่มีข้อมูลในอดีตเกี่ยวกับพารามิเตอร์ที่ต้องการจำลองว่ามีลักษณะการแจกแจงของข้อมูลในลักษณะใด เพื่อนำมาเป็นค่าเริ่มต้นในฟังก์ชันของการจำลองข้อมูลที่ต้องการ

สำหรับผู้เริ่มใช้โปรแกรม R สามารถหาฟังก์ชันต่างๆที่ต้องการได้อย่างง่ายที่สุดโดยใช้คำสั่ง “help.search(‘keyword’)” และถ้าต้องการรายละเอียดเพิ่มเติมในแต่ละฟังก์ชันโดยผ่านคำสั่ง “?topic” หรือคำสั่ง “help(‘topic’)” ในโปรแกรม R บรรจุค่าเริ่มต้นของการแจกแจงลักษณะต่างๆไว้จำนวนมาก เช่น “Normal” ซึ่งเป็นค่าเริ่มต้นฟังก์ชันที่เกี่ยวข้องกับการแจกแจงปกติ 4 ฟังก์ชัน ได้แก่ (www.stat.berkeley.edu/~vigre/activities/.../2006/wickham_R)

- pnorm – คือ ฟังก์ชันการแจกแจง (the distribution function)
- rnorm – คือ ฟังก์ชันที่ใช้สร้างตัวแปรเชิงสุ่มปกติ (nomal random variates)
- qnorm – คือ ฟังก์ชันควอไทล์ (the quantile function)
- dnorm – คือ ฟังก์ชันความเข้ม (the density function)

ฟังก์ชันต่างๆ เหล่านี้สามารถกำหนดค่าเฉลี่ยและความแปรปรวนต่างๆ ของการแจกแจงได้ โดยทั่วไปแล้วตัวอักษรตัวแรกแสดงถึงข้อมูลเกี่ยวกับการแจกแจงที่ต้องการ ซึ่งอาจเขียนในรูปย่อๆได้ เช่น

คำสั่งเต็ม	**คำสั่งย่อ**	**การแจกแจง**
?Normal	norm	The normal distribution
?Possion	pois	The poisson distribution
?Binomial	binom	The binomial distribution
?Geometric	geom	The geometric distribution
?TDist	t	The Students t distribution
?FDist	f	The Chi-Squared Distribution
?Chisquare	chisq	

กรณีที่ผู้ใช้ต้องการผลิตตัวเลขสุ่มซ้ำที่ได้ค่าเหมือนเดิม เพื่อนำไปใช้ในการศึกษาเปรียบเทียบภายใต้สภาพแวดล้อมหรือสถานการณ์เดียวกัน (มานพ วราภักดิ์, 2550) จะใช้คำสั่ง “set.seed” ที่แสดงถึงจุดเริ่มต้นของการสร้างตัวเลขสุ่ม หากกำหนด seed เหมือนกัน ก็จะได้ค่าตัวเลขสุ่มชุดเดียวกัน ตัวอย่างเช่น

```
> rnorm(10)
[1] -0.30125091 -0.23552783 -0.61879922 0.23110361 1.31541443 0.41800526
[7] 0.39434739 1.53440933 -0.37052444 0.03481197

> rnorm(10)
[1] -1.25067676 -1.60956084 0.29984995 -0.58450931 -1.03388333 1.40477092
[7] -0.04511211 1.21148672 0.15405686 -0.42036729

> set.seed(18749)
> rnorm(10)
[1] -0.54555908 1.22938955 2.58249311 1.57057115 -0.03186376 0.32983807
[7] -0.44899542 0.56599635 -0.23902963 0.79350037

> set.seed(18749)
> rnorm(10)
[1] -0.54555908 1.22938955 2.58249311 1.57057115 -0.03186376 0.32983807
[7] -0.44899542 0.56599635 -0.23902963 0.79350037
```

สองตัวอย่างแรกใช้คำสั่ง rnorm(10) เหมือนกัน ตัวเลขในวงเล็บแสดงถึงจำนวนเลขสุ่มที่ต้องการ ผลที่ได้คือตัวเลขสุ่ม 10 ค่า ที่ไม่เหมือนกัน สำหรับสองตัวอย่างหลังใช้คำสั่ง set.seed(18749) ซึ่งกำหนดตัวเลขในวงเล็บเหมือนๆกัน ก่อนใช้คำสั่ง rnorm(10) ทำให้ได้ตัวเลขสุ่ม 10 ค่าที่เหมือนกัน

กรณีที่ผู้ใช้มีตัวเลขชุดหนึ่งอยู่แล้ว เช่น เลข 1-10 แล้วต้องการสุ่มตัวเลขดังกล่าว โดยใช้คำสั่ง sample(1:10) โปรแกรมจะดำเนินการสุ่มตัวเลข 1-10 โดยไม่ใส่คืนให้ ชุดตัวเลขที่ได้จึงไม่มีค่าซ้ำกันเลย แต่ถ้าเรากำหนดขนาดของกลุ่มตัวอย่าง เช่น 5 โดยใช้คำสั่ง sample(1:10, size = 5) โปรแกรมจะดำเนินการสุ่มตัวเลขจากเวกเตอร์ตัวเลข 1-10 มา 5 จำนวน โดยไม่มีการใส่คืน ถ้าต้องการให้มีการสุ่มแบบใส่คืน ต้องใช้คำสั่ง replace = TRUE ดังตัวอย่างต่อไปนี้

```
> sample(1:10)
[1] 10 7 8 9 3 2 6 1 4 5

> sample(1:10, size = 5)
[1] 5 8 6 1 10

> sample(1:10, size = 15, replace = TRUE)
[1] 9 5 1 4 5 10 1 2 1 8 1 2 7 5 10
```

นอกจากนี้สามารถกำหนดความน่าจะเป็นของการแจกแจงอื่นๆมากกว่าที่เป็น uniform ได้ด้วยคำสั่ง prob ตัวอย่างการหาค่าสถิติเบื้องต้นของตัวแปร x โดยกำหนดตัวแปรสุ่มปกติ 20 จำนวน โดยกำหนด mean = 10 , sd = 20 โดยใช้คำสั่ง ดังนี้

```
> x <- rnorm(20, mean = 10, sd = 20)
> mean(x)
[1] 8.144194
> sd(x)
[1] 17.36403
> sqrt(var(x))
[1] 17.36403
> median(x)
[1] 8.448636
> quantile(x, probs = c(0.05, 0.95))
5%          95%
-18.35355   31.73562
> summary(x)
   Min.    1st Qu.   Median    Mean   3rd Qu.    Max.
-18.630   -6.728    8.449     8.144  25.220     36.880
```

อีกทั้งมีฟังก์ชันเบื้องต้นที่มีประโยชน์ เช่น จำนวนของกลุ่มตัวอย่าง y ใช้คำสั่ง length(y) หาค่าสูงสุดของ y ค่าต่ำสุดของ y ค่าลอกการิทึ่มของ y ผลรวมของ y ผลผลิตของ y ดังตัวอย่าง

```
> y <- 1:20
> length(y)
[1] 20
> max(y)
[1] 20
> min(y)
[1] 1
> log(y)
[1] 0.0000000 0.6931472 1.0986123 1.3862944 1.6094379 1.7917595 1.9459101
[8] 2.0794415 2.1972246 2.3025851 2.3978953 2.4849066 2.5649494 2.6390573
[15] 2.7080502 2.7725887 2.8332133 2.8903718 2.9444390 2.9957323
> sum(y)
[1] 210
> prod(y)
[1] 2.432902e+18
```

เมื่อจำลองข้อมูลตามเงื่อนไขที่ต้องการแล้วด้วยโปรแกรม R สามารถเข้าสู่โปรแกรม WINBUGS เพื่อวิเคราะห์ผลต่อไปผ่านใช้คำสั่งพิเศษของโปรแกรม R ที่สามารถนำข้อมูลดังกล่าวเข้าไปใช้กับโปรแกรม WINBUGS ได้อย่างสะดวกรวดเร็ว

4.3 การเชื่อมโยงโปรแกรม R ไปสู่ โปรแกรม WINBUGS

การนำข้อมูลจำลองตามเงื่อนไขที่ต้องการด้วยโปรแกรม R เข้าไปใช้กับโปรแกรม WINBUGS โดยไม่ต้องเปิดเข้าตัวโปรแกรม WINBUGS ต้องใช้คำสั่งพิเศษที่สามารถ แพกเก็จคำสั่งที่ใช้เรียกว่า R2WINBUGS ซึ่งผู้ใช้งานสามารถดาว์นโหลดเข้ามาติดตั้งในโปรแกรม R ฟังก์ชันของคำสั่ง R2WINBUGS นี้ทำการเขียนข้อมูลและคำสั่งรูปแบบเฉพาะที่ WinBUGS สามารถอ่านได้เพื่อนำไปวิเคราะห์ด้วยชุดคำสั่งของ WinBUGS เรียกว่า “WinBUGS จาก R” เมื่อดำเนินการด้วยโปรแกรม WinBUGS เสร็จสิ้นแล้ว สามารถอ่านผลการวิเคราะห์ได้จากโปรแกรม R ด้วยแพจเก็จเดียวกันนี้ได้ หรือจะใช้สิ่งอำนวยความสะดวกจากแพจเก็จ “coda” สำหรับการวิเคราะห์ผลที่แสดงออกมาต่อไปอีก (Sturtz, S., Ligges, U., Gelman, A., 2005)

ฟังก์ชันหลักที่สำคัญในการเริ่มดำเนินการของแพจเก็จ R2WINBUGS คือ ฟังก์ชัน “bug ( )” ซึ่งประกอบด้วยกลุ่มฟังก์ชันนี้หลายๆฟังก์ชันด้วยกัน ทีละขั้น คือ

1) ฟังก์ชัน bug.data.inits ( ) เป็นฟังก์ชันในการเขียนแฟ้มข้อมูล ‘data.txt’ และ ‘inits1.txt’, ‘inits2.txt’,… ไปสู่ฐานข้อมูลของการทำงาน แฟ้มข้อมูลเหล่านี้จะถูกนำไปใช้กับโปรแกรม WinBUGS ในกระบวนการวิเคราะห์ ข้อสำคัญคือข้อมูลนำเข้าสู่โปรแกรม WinBUGS ต้องไม่มากเกินกว่าจำนวนตัวเลขที่กำหนดไว้ และจำเป็นต้องใช้ตัวอักษร E แทน e ซึ่งเป็นสัญญลักษณ์ทางวิทยาศาสตร์ เนื่องจากข้อจำกัดของจำนวนตัวเลขที่กำหนดไว้ เช่น ค่าเริ่มต้นกำหนดจำนวนตัวเลขไว้ digits= 5 จึงต้องเปลี่ยนรูปจำนวน 123456.789 ไปเป็น 1.234567E+05

2) ฟังก์ชัน bug.script ( ) เป็นฟังก์ชันในการเขียนแฟ้ม ‘script.txt’ ซึ่งเป็นชุดคำสั่งสำหรับใช้กับการวิเคราะห์ของโปรแกรม WinBUGS

3) ฟังก์ชัน bugs.run ( ) เป็นฟังก์ชันในการอัพเดทความยาวของเฟสปรับเหมาะ (The adaptive phases)ในทะเบียนของโปรแกรม WinBUGS (โดยการใช้ฟังก์ชัน bug.update.settings( )) เพื่อเรียกใช้โปรแกรม WinBUGS และวิเคราะห์ข้อมูลด้วยชุดคำสั่ง ‘script.txt’

4) ฟังก์ชัน bugs.sims ( ) เป็นฟังก์ชันที่จะถูกเรียกใช้เมื่อแพจเก็จคำสั่ง codaPkg ถูกตั้งค่าเริ่มต้นเป็น FALSE มิฉะนั้นแล้ว bugs ( ) จะกลายเป็นชื่อแฟ้มของข้อมูลที่ถูกจัดเก็บ ฟังก์ชัน bugs.sims ( ) นี้จะอ่านการจำลองจาก โปรแกรม WinBUGS ไปสู่ R

4.4 โปรแกรม WINBUGS

WINBUGS เป็นโปรแกรมที่ได้รับความนิยมสูง เนื่องจากเป็นซอฟแวร์ที่สามารถวิเคราะห์โมเดลเชิงสถิติที่ซับซ้อนโดยใช้วิธีมาร์คอฟเชนมอนติคาร์โล (Markov chain Monte Carlo; MCMC) วิธีการนี้ใช้การสุ่มตัวอย่างแบบกิ๊บส์ (Gibbs Sampling) และอัลกอริทึ่มแบบเมโทรโพลิส (Metropolis algorithm) เพื่อสร้างมาร์คอฟเชนด้วยการสุ่มจากการแจกแจงแบบมีเงื่อนไขเต็มรูป (Sturtz, S., Ligges, U., Gelman, A., 2005; Goldstein, H., 2002)

ผู้ใช้ WINBUGS ต้องกำหนดโมเดลเพื่อดำเนินการวิเคราะห์ และเพื่อนำข้อมูลเข้าและกำหนดค่าเริ่มต้นของจำนวนมาร์คอฟเชน จึงดำเนินการวิเคราะห์มาร์คอฟเชนเหล่านั้นและเก็บรักษาผลการวิเคราะห์สำหรับพารามิเตอร์ที่ผู้ใช้สนใจได้ ซึ่งจะให้ค่าสถิติเบื้องต้น การวินิจฉัยการลู่เข้า ฯลฯ อีกด้วย ผู้ใช้โปรแกรม WINBUGS บางส่วนอาจต้องการเก็บผลการวิเคราะห์และนำไปอ่านด้วยโปรแกรม R เพื่อนำไปวิเคราะห์ต่อไป ซึ่ง WinBUGS รุ่น 1.4 เป็นรุ่นที่สามารถดำเนินการวิเคราะห์ได้ด้วยชุดของคำสั่งที่อยู่บนโปรแกรม R ซึ่งเรียกว่า R2WinBUGS ซึ่งทำให้สามารถนำเสนอและจัดหาเครื่องมือที่เรียกโปรแกรม WinBUGS ได้โดยตรงหลังจากที่ได้จัดกระทำข้อมูลในโปรแกรม R แล้ว นอกจากนี้ ยังทำให้สามารถกลับมาใช้งานโปรแกรม R หลังจากที่ได้ผลลัพธ์

จากโปรแกรม WinBUGS แล้วได้อีก เมื่อต้องมีการเปลี่ยนข้อมูลหรือการวิเคราะห์กลุ่มของชุดข้อมูลหลายกลุ่ม เพราะมีความสะดวกในการใช้ฟังก์ชันบางอย่างในโปรแกรม R ที่จะทำงานวนลูปมากกว่าที่จะต้องสำเนาข้อมูลมาวางไว้ในโปรแกรม WinBUGS แต่ละครั้งเพื่อทำการวิเคราะห์ ถึงกระนั้นก็ยังพบว่ามีความยากลำบากในเรื่องนี้เพราะทั้ง โปรแกรม R และโปรแกรม WinBUGS อาจถูกเก็บไว้ในหน่วยความจำสำรอง (RAM) ของระบบปฏิบัติการวินโดว์ส

โปรแกรม WinBUGS เป็นแพจเก็จคำสั่งของมาร์คอฟเชน มอนติคาร์โล ด้วยวิธีของเบย์ ซึ่งเป็นผลมาจากการพัฒนาอย่างต่อเนื่องหลายปีโดยทีมนักสถิติและนักโปรแกรมเมอร์แห่งหน่วยวิจัยชีวสถิติของสภาวิจัยทางการแพทย์ในแคมบริดจ์ โมเดลหลายโมเดลถูกเสนอด้วยภาษาที่มีความยืดหยุ่น และยังมีการนำเสนอในเชิงรูปภาพ เรียกว่า DOODLEBUGS ที่ยอมให้ผู้ใช้กำหนดโมเดลของตนเองในเทอมของรูปภาพได้โดยตรงและมีประโยชน์อย่างมากสำหรับโมเดลที่มีความซับซ้อน (Goldstein, H., 2002)

ประเด็นของการประมาณค่าวิธีของเบย์โดยใช้วิธีมาร์คอฟเชนมอนติคาร์โลมีเนื้อหาสาระที่กว้างมาก ก่อนที่จะนำวิธีการนี้ไปใช้ ผู้ใช้จะต้องเรียนรู้เนื้อหาเบื้องต้นเหล่านี้มาเป็นอย่างดีด้วยตนเองจากคู่มือการใช้งานโปรแกรม WinBUGS จะให้ความรู้ที่จะเริ่มต้นการใช้งาน ผู้ใช้ที่จะจำลองข้อมูลด้วยวิธีมาร์คอฟเชนมอนติคาร์โลควรตระหนักว่า วิธีนี้ไม่เหมือนกับกระบวนการความน่าจะเป็นสูงสุด (maximum likelihood procedures) วิธีการ MCMC นี้มีองค์ประกอบของความไม่แน่นอนในค่าประมาณที่ได้มาและกระบวนการดังกล่าวที่ใช้สำหรับตัดสินการลู่เข้าและความถูกต้องแม่นยำมีความซับซ้อนอย่างมากและบางส่วนต้องประกอบด้วยการตัดสินและประสบการณ์ โดยสามารถศึกษาได้จากคู่มือการใช้โปรแกรม WinBUGS

Data interface: **การเชื่อมต่อข้อมูล** ข้อมูลนำเข้าสู่ WinBUGS มีส่วนประกอบ 3 ส่วนด้วยกันที่เป็นไฟล์ที่แตกต่างกัน ไฟล์แรกคือชุดของโมเดลที่กำหนดโครงสร้าง (a series of specification instructions) ไฟล์ที่สองประกอบด้วยข้อมูลที่สังเกตได้ (the observed data) และไฟล์ที่สามประกอบด้วยกลุ่มของค่าเริ่มต้น (a set of starting values) โดยที่จริงแล้ว WinBUGS จะกำหนดค่าเริ่มต้นให้ถ้าเราต้องการ แต่ตามปกติแล้วผู้ใช้มักจะกำหนดค่าเริ่มต้นเอง ซึ่งจะแสดงวิธีการใช้ไฟล์ทั้งสามนี้ในตัวอย่างที่จะนำเสนอต่อไป WinBUGS ทำงานภายใต้วินโดว์สหลายๆรุ่น รุ่นที่เก่ากว่าจะยังสามารถทำงานได้ในระบบปฏิบัติการอื่นเช่น LINUX (ดูรายละเอียดเพิ่มเติมในเวบไซด์ของ WinBUGS) ในขณะที่ WinBUGS รุ่น 1.4 มีเครื่องมือกราฟฟิกที่มีประโยชน์มากมาย อีกทั้งยังมีชุดเครื่องมือที่ใช้ประจำ เช่น CODA ที่สามารถนำเสนอการวาดกราฟเพิ่มเติม

ของผลลัพธ์จาก WinBUGS ในโปรแกรม SPLUS หรือ โปรแกรม R หลักการนำเสนอต่อไปในตัวอย่าง

Model types: **ประเภทของโมเดล** ประเภทของโมเดลที่จะสอดคล้องกับ WinBUGS มีช่วงกว้างมาก ซึ่งเป็นได้ทั้งแบบเชิงเส้นตรงและไม่เป็นเส้นตรง รวมทั้งชุดมาตรฐานของโมเดลเส้นตรงเชิงสรุปนัยทั่วไป (the standard set of generalized linear models) สำหรับอิทธิพลเชิงสุ่ม (the random effect) ซึ่งเป็นรูปแบบการกระจายที่หลากหลายจะถูกกำหนด รวมถึงการกระจายปกติหลายตัวแปรและการกระจายแบบที (the multivariate Normal and T-distributions) โมเดลเชิงมิติ (Spatial models) และโมเดลตัวแปรแฝง (latent variable models) เป็นโมเดลที่มีความสอดคล้องกับ WinBUGS

ตัวอย่าง

โมเดลองค์ประกอบของความแปรปรวน 2 ระดับโดยใช้ชุดข้อมูลชื่อไฟล์ว่า EXAM และจะใช้ข้อมูลนี้นำเสนอด้วยภาษาในรูปแบบของ WinBUGS โมเดล คือ

$$y_{ij} = (X\beta)_{ij} + u_{0j} + e_{0ij}$$
$$u_{0j} \sim N(0,\sigma^2_{u0}) \quad e_{0ij} \sim N(0,\sigma^2_{e0}) \qquad (1)$$

เมื่อ i,j เป็นดัชนีระดับ 1 และระดับ 2 และ $(X\beta)_{ij}$ เป็นตัวแทนของส่วนที่กำหนดให้สำหรับรายการที่ ij ส่วนที่กำหนดนี้รวมทั้งจุดตัดβ_0 และความแปรปรวนร่วม 4 ตัว คือ β_1, β_2, β_3, β_4 ได้แก่ คะแนน LRT เพศ และองค์ประกอบของเพศในโรงเรียน (มีตัวแปรดัมมี่ 2 ตัว คือ โรงเรียนชาย และโรงเรียนหญิง เปรียบเทียบระหว่างโรงเรียนสหศึกษา กับโรงเรียนหญิงล้วน-ชายล้วน)

รหัสที่กำหนดสำหรับโมเดล WinBUGS มีดังนี้

```
model
{
# Level 1 definition
for(i in 1:N) {
normexam[i] ~ dnorm(mu[i],tau)
mu[i]<- beta[1] * cons[i]
+ beta[2] * standlrt[i]
+ beta[3] * gender[i]
+ beta[4] * boysch[i]
+ beta[5] * girlsch[i]
+ u2[school[i]] * cons[i]
}
# Higher level definitions
for (j in 1:n2) {
u2[j] ~ dnorm(0,tau.u2)
}
# Priors for fixed effects
for (k in 1:5) { beta[k] ~ dflat() }
# Priors for random terms
tau ~ dgamma(0.001,0.001)
sigma2 <- 1/tau
tau.u2 ~ dgamma(0.001,0.001)
sigma2.u2 <- 1/tau.u2

}
```

- Comment (เริ่มด้วยสัญญลักษณ์ #) ใส่เพื่อให้อ่านได้ง่าย
- บรรทัดที่ 5 กำหนดว่าการตอบสนองเป็นแบบปกติ และมีค่าเฉลี่ย เท่ากับ mu(i) และมีความแม่นยำ/ความเที่ยงตรง เท่ากับ tau ค่าความแม่นยำ/ความเที่ยงตรง นี้ได้มากจากส่วนกลับของความแปรปรวนและความสัมพันธ์นี้ถูกกำหนดในบรรทัดก่อนสุดท้ายด้วยความสัมพันธ์ในเชิงตรรก (logical) หรือข้อเท็จจริง (deterministic) คือ sigma2.u2 <- 1/tau.u2
- บรรทัดที่ 6-10 กำหนดค่า mu(i) เป็นฟังก์ชันเพิ่มเชิงเส้นตรงที่มีจุดตัดและความแปรปรวนร่วม 4 ตัว และบรรทัดที่ 11 เพิ่มชุดของผลเชิงสุ่ม เมื่อ school(i) คือ รหัสโรงเรียน ลำดับที่ 1-65 สำหรับข้อมูลรายการที่ i แล้วคูณด้วย u2 ที่เป็นตัวแปรสุ่ม แล้วนำไปคูณกับ cons(i) ซึ่งอาจไม่จำเป็นมากนักในที่นี้แต่ยังคงเอาไว้เพื่อรักษาความคงเส้นคงวาของโมเดลสัมประสิทธิ์เชิงสุ่มทั่วไปที่ตัวแปรที่อธิบายได้มีความสัมพันธ์กับสัมประสิทธิ์เชิงสุ่ม
- ค่าคงที่ N, n2 ที่ใช้ในที่นี้มีค่าตามข้อมูลนำเข้า คือ 4059 และ 65
- ค่าก่อนหน้า (prior) สำหรับอิทธิพลกำหนด (fixed effects) จะใช้ ‘flat’ เช่น ค่าก่อนหน้าแบบ uniform ตลอดทั้งเส้นตรงของแต่ละสัมประสิทธ์ถดถอย (fixed effects) หรืออีกทางหนึ่งกำหนดค่าก่อนหน้าที่เหมาะสม ‘proper prior’ ซึ่งเทียบเท่ากันในเชิงหน้าที่ คือ แบบ normal เมื่อมีความแปรปรวนขนาดใหญ่มาก เช่น $N(0,1*10^{6})$ ที่ใช้ในการสอนโปรแกรม WinBUGS
- ค่าก่อนหน้า (prior) สำหรับอิทธิพลเชิงสุ่ม (random effects) เรากำหนดความถูกต้องแม่นยำ ค่าก่อนหน้า เท่ากับ Gamma (0.001,0.001) เป็นค่าเริ่มต้นสำหรับโปรแกรม WinBUGS
- ในหัวข้อ คำนิยามระดับสูงกว่า ‘higher level definitions’ การกระจายของ u2(j) (j=1….n2) ถูกกำหนดเป็นการกระจายปกติที่มีค่าเฉลี่ย 0 และมีความแม่นยำ/ความเที่ยงตรง เท่ากับ tau.u2 ซึ่งในบรรทัดสุดท้ายเกี่ยวข้องกับความแปรปรวนระดับ 2 ของพารามิเตอร์ความแม่นยำ/ความเที่ยงตรงนี้
- ผู้ใช้สามารถกำหนดพารามิเตอร์เพิ่มเติมได้อีก เช่น สัมประสิทธิ์บางส่วนของความแปรปรวน (variance partition coefficient; VPC) ซึ่งในที่นี้มีค่าเทียบเท่ากับความสัมพันธ์ภายในกลุ่ม (intra-cluster correlation) โดยกำหนดให้มีความสัมพันธ์เชิงตรรกะ คือ vpc <-sigma2.u2/(sigma2.u2+sigma2) ซึ่งผู้ใช้สามารถควบคุมและได้ค่าประมาณที่ต้องการ

4.5 การวิเคราะห์ประสิทธิภาพของการประมาณค่าพารามิเตอร์

ประสิทธิภาพของการประมาณค่าพารามิเตอร์ พิจารณาได้จาก ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ความสามารถผู้สอบ; SE(θ) และค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง Mean squared error; MSE

(1) ความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ความสามารถผู้สอบ ; SE(θ)

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

ค่า SE(θ) ที่ต่ำกว่า แสดงว่า มีประสิทธิภาพในการประมาณค่าพารามิเตอร์สูงกว่าหรือดีกว่า

(2) ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean squared error ; MSE) (Swaminathan, Hambleton, Sireci, Xing และ Rizavi, 2003) คือ ความถูกต้องของการประมาณค่าพารามิเตอร์ ได้จากความแตกต่างระหว่างค่าที่ได้จากตัวประมาณค่ากับค่าจริง ถ้ามีค่าเป็นศูนย์ แสดงว่าตัวประมาณค่ามีประสิทธิภาพในการประมาณค่าได้แม่นยำมากที่สุด

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

$$MSE(\hat{\theta}) = (Bias(\hat{\theta}, \theta))^2 + Variance(\hat{\theta})$$

นอกจากนี้ประสิทธิภาพของโมเดลที่ใช้ในการประมาณค่า พิจารณาได้จากความสอดคล้องกลมกลืนระหว่างโมเดลกับข้อมูล โดยดูจากค่า Akaike information criterion (AIC) (deLeeuw, 1992; Klein Entink, 2009)

ค่า AIC หรือ Akaike information criterion คือ ค่าสถิติที่ใช้เป็นเกณฑ์ในการคัดเลือกโมเดล แสดงถึงความกลมกลืนระหว่างข้อมูลกับโมเดล ค่า AIC ที่ต่ำที่สุดคือโมเดลที่ดีที่สุด คำนวณได้จาก

$$AIC = 2k - 2\ln(L)$$

เมื่อ k เป็นจำนวนพารามิเตอร์ในโมเดล และ L คือค่าสูงสุดของฟังก์ชันความน่าจะเป็น (Likelihood function) ของโมเดลที่ประมาณค่า อธิบายความว่า Likelihood เป็นการทดสอบความน่าจะเป็นของความกลมกลืนระหว่างพฤติกรรมการตอบของผู้สอบกับการประมาณค่าพารามิเตอร์ ดังนั้น ถ้ามีจำนวนพารามิเตอร์ที่ต้องการประมาณค่ามาก เช่น สมมติว่ามีข้อสอบ 20 ข้อ และสมมติว่ามีผู้สอบ 200 คน พารามิเตอร์ที่ต้องประมาณค่า 20 x 200 = 4,000 ตัว แต่ถ้าไม่มี Likelihood เลย คือ Likelihood = 0 จะมีค่า AIC = 8,000 และเมื่อ Likelihood เพิ่มขึ้น ค่า

AIC จะน้อยลง ดังนั้นเกณฑ์ในการคัดเลือกโมเดล คือ ค่า AIC ที่น้อยกว่า แสดงว่ามีความกลมกลืนระหว่างพฤติกรรมการตอบของผู้สอบกับการประมาณค่าพารามิเตอร์มากกว่า

ตอนที่ 5 บทความและงานวิจัยที่เกี่ยวข้อง

ในตอนที่ 5 นำเสนอเอกสารและงานวิจัยที่เกี่ยวข้องกับโมเดลทฤษฎีการตอบสนองข้อสอบเชิงสุ่มและการประมาณค่าพารามิเตอร์ เพื่อนำไปสู่การสร้างกรอบแนวทางการวิจัย ผู้วิจัยแบ่งการนำเสนอออกเป็น 2 ส่วน คือ

5.1 งานวิจัยเกี่ยวกับโมเดลการตอบสนองเชิงสุ่ม

Glas และ Van der Linden (2003) ศึกษาเรื่อง Computerized adaptive testing with item cloning โดยการสุ่มข้อสอบจากตระกูลข้อสอบที่ได้รับการคัดเลือกจากค่าประมาณของพารามิเตอร์ความสามารถผู้สอบในขณะนั้นมาใช้ในการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ เป็นการศึกษาด้วยข้อมูลจำลอง ประมาณค่าพารามิเตอร์ด้วยโมเดลตอบสนองข้อสอบ 3 พารามิเตอร์ แบ่งการศึกษาเป็น 2 ส่วน คือ (1) การปรับเทียบมาตรฐานของคลังข้อสอบ และ (2) การทดสอบแบบปรับเหมาะที่ถูกต้องแม่นยำ โครงสร้างตระกูลข้อสอบของคลังข้อสอบของการวิจัยนี้ได้มาจากข้อสอบ LSAT ที่แตกต่างกัน 4 คลัง จำนวน 753 ข้อ การปรับเทียบมาตรฐานของคลังข้อสอบด้วยการจัดกระทำความยาวแบบสอบ 20 และ 40 ข้อ และขนาดกลุ่มตัวอย่างผู้สอบ 100, 400 และ 1,000 คน ทำซ้ำ 100 รอบในแต่ละเงื่อนไข การจำลองข้อมูลความสามารถผู้สอบโดยสุ่มจากการแจกแจงข้อมูลแบบปกติมาตรฐาน และในการทดสอบแบบปรับเหมาะ มีการเปรียบเทียบระหว่างความคลาดเคลื่อนของตัวประมาณค่าความสามารถผู้สอบในเงื่อนไขการศึกษาเดียวกันของผู้สอบ 1,000 คน ณ ค่าความสามารถผู้สอบ (θ) ที่ -2.0, -1.0, 0.0, 1.0, และ 2.0 ผลการวิจัยการปรับเทียบมาตรฐานของคลังข้อสอบได้รับการยืนยันความถูกต้องแม่นยำ และการทดสอบแบบปรับเหมาะโดยใช้เทคนิคสำเนาข้อสอบ (item-cloning) ทำให้ความผันแปรภายในตระกูลข้อสอบมีค่าน้อยลง จึงเป็นแบบสอบที่ปรับได้เหมาะสมกับระดับความสามารถผู้สอบได้ดียิ่งขึ้น

Jean-Paul Fox (2005) วิจัยเรื่อง Ramdomized Item Response Theory Models (RIRT models) ศึกษาวิธีการที่ใช้เทคนิคการตอบสนองเชิงสุ่ม (randomized response ; RR) ซึ่งพัฒนาขึ้น เพื่อวัดตัวแปรแฝงที่การถามอย่างตรงไปตรงมาอาจทำให้ได้ผลการตอบที่ไม่เป็นจริงหรือเกิดความลำเอียงขึ้นได้ เช่น การถามเกี่ยวกับเจตคติของบุคคล วิธีดำเนินการวิจัยโดยการจำลองข้อมูลผลการตอบจากผู้ตอบ 1,000 คน แบ่งออกเป็น 20 กลุ่มเท่าๆกัน และมีข้อคำถามที่วัดคุณลักษณะเดียวกัน จำนวน 20 ข้อ รวมทั้งสิ้นมีชุดข้อมูล 100 ชุด การประมาณค่าพารามิเตอร์

ด้วยโมเดล RIRT ระดับลดหลั่น ที่ประกอบด้วย 3 องค์ประกอบ คือ โมเดลการตอบเชิงสุ่ม (randomized response model) โมเดลการตอบสนองข้อสอบ (item response model) และโมเดลเชิงโครงสร้างระดับลดหลั่น (structural hierarchical model); หรือ ANOVA ทำการคำนวณทวนซ้ำ 50,000 รอบและช่วง burn-in 5,000 รอบ ด้วย Markov chain Monte Carlo ประมาณค่าพารามิเตอร์ข้อสอบจากข้อมูลการตอบข้อสอบจริงและข้อมูลการตอบข้อสอบเชิงสุ่ม ผลการวิจัยพบว่า ค่าพารามิเตอร์จริงใกล้เคียงกับค่าประมาณเฉลี่ย ค่าพารามิเตอร์ข้อสอบทั้งค่าอำนาจจำแนกและค่าความยาก ที่ประมาณจากข้อมูลการตอบข้อสอบจริงมีช่วงที่แคบกว่าค่าที่ประมาณจากข้อมูลการตอบข้อสอบเชิงสุ่ม ความแปรปรวนภายหลังของการประมาณค่าจากข้อมูลการตอบข้อสอบเชิงสุ่มมีช่วงกว้างกว่าเพราะสารสนเทศจากผู้ตอบแต่ละคนมีน้อยกว่า

De Jong, Steenkamp, J-P. Fox และ Baumgartner (2008) ศึกษาเรื่อง Using item response theory to measure extreme response style in marketing research: A global investigation เพื่อศึกษารูปแบบการตอบแบบสุดโต่ง โดยให้พารามิเตอร์ความยากและอำนาจจำแนกของข้อสอบแตกต่างกันไปข้ามประเทศ ผลการวิจัยกับข้อมูลจำลอง 20 ประเทศ ผู้ตอบประเทศละ 300 คน ข้อคำถามทั้งหมด 50 ข้อ ที่ประกอบกันเป็นแบบสอบถามย่อย 5 ชุด ชุดละ 10 ข้อ (เป็นแบบสอบย่อยไม่ลำเอียงที่แตกต่างกัน 3 ชุด และแบบสอบย่อยที่ลำเอียง (DIF) 2 ชุด) รวมเงื่อนไขที่ศึกษา 6 เงื่อนไข เปรียบเทียบระหว่าง Traditional model กับ IRT model พบว่า การประมาณค่าด้วย Traditional model มีความถูกต้องแม่นยำลดลงเมื่อมีความสัมพันธ์ระหว่างข้อสอบและ DIF ของข้อสอบเพิ่มขึ้น และการประมาณค่าด้วย IRT model มีความถูกต้องแม่นยำมากกว่า Traditional model ในทุกเงื่อนไขที่ศึกษา โดยพิจารณาจากค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE)

Derek C. Briggs และ Mark Wilson (2007) วิจัยเรื่อง Generalizability in Item Response Modeling (GIRM) ศึกษาการสรุปอ้างอิงในรูปแบบของโมเดลตอบสนองข้อสอบ โดยสร้างข้อตกลงเบื้องต้นของการกระจายของฟาเซทการวัดที่เกี่ยวข้อง ให้เป็นโมเดลการวัดอิทธิพลเชิงสุ่ม (relevant measurement facets) เพื่อเชื่อมโยงการสุ่มของทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ไปสู่ scaling model ของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ทำให้สามารถประมาณค่าองค์ประกอบของความแปรปรวนในทฤษฎีการสรุปอ้างอิงได้พร้อมๆกันกับการประมาณค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบแบบเดิม

De Boeck (2008) วิจัยเรื่อง Random Item IRT Models ศึกษาเกี่ยวกับการใช้โมเดล IRT แบบตรวจให้คะแนน 2 ค่า ที่เป็นโมเดลแบบ 1 พารามิเตอร์ ในการศึกษาโมเดลการตอบสนอง

ข้อสอบเชิงสุ่ม โดยการปรับเปลี่ยนจากการประมาณค่าพารามิเตอร์ของข้อสอบที่ใช้รูปแบบเจาะจงข้อสอบ (fixed item) และสุ่มบุคคล (random person) มาเป็นการประมาณค่าความสามารถของบุคคลโดยใช้รูปแบบสุ่มข้อสอบ (random item) และเจาะจงบุคคล (fixed person) แทน เพื่อที่จะได้สามารถสรุปอ้างอิงค่าความสามารถของบุคคลได้โดยตรง

ตารางที่ 2.4 สรุปแนวคิดเกี่ยวกับโมเดลการตอบสนองข้อสอบเชิงสุ่ม

ผู้วิจัย / ผู้เขียน	ปี ค.ศ.	ประเด็นในการวิจัย	สาระสำคัญ
Glas และ Van der Linden	2003	นำ Random Effect มาใช้ในการสุ่มข้อสอบในการทดสอบแบบปรับเหมาะ	เทคนิค item-cloning สุ่มข้อสอบในการทดสอบแบบปรับเหมาะเพื่อให้ได้ข้อสอบที่เหมาะสมกับความสามารถผู้สอบมากที่สุด
Jean-Paul Fox	2005	วิธีการเทคนิคการตอบสนองเชิงสุ่ม (Randomized Response ; RR)	วัดตัวแปรแฝงที่การถามอย่างตรงไปตรงมาอาจทำให้ได้ผลการตอบที่ไม่เป็นจริงหรือเกิดความลำเอียงขึ้นได้ เช่น การถามเกี่ยวกับเจตคติของบุคคล โดยวิธีการเทคนิคการตอบสนองเชิงสุ่ม
Derek C. Briggs และ Mark Wilson	2007	Generalizability in Item Response Modeling (GIRM)	สร้างข้อตกลงเบื้องต้นของการกระจายของฟาเซทการวัดที่เกี่ยวข้อง ให้เป็นโมเดลการวัดอิทธิพลเชิงสุ่ม (relevant measurement facets) เพื่อเชื่อมโยงการสุ่มของ G-Theory ไปสู่ scaling model ของ IRT
De Jong, Steenkamp, J-P. Fox และ Baumgartner	2008	การวัด Extreme Response Style โดยใช้อิทธิพลเชิงสุ่มกับพารามิเตอร์ข้อสอบ	นำ Random Effect มาใช้กับ ERS และพารามิเตอร์ข้อสอบทั้งความยากและอำนาจจำแนก

ตารางที่ 2.4 (ต่อ)

ผู้วิจัย / ผู้เขียน	ปี ค.ศ.	ประเด็นในการวิจัย	สาระสำคัญ
De Boeck	2008	Random Item IRT Models	การประมาณค่าความสามารถของบุคคลโดยใช้รูปแบบสุ่มข้อสอบ (random item) และเจาะจงบุคคล (fixed person) ด้วยโมเดล IRT แบบ 1 พารามิเตอร์

จากตารางที่ 5 แสดงให้เห็นว่างานวิจัยส่วนใหญ่จะใช้อิทธิพลเชิงสุ่มมาใช้ประโยชน์ในงานวิจัยในลักษณะต่างๆกัน เช่น นำมาใช้กับเทคนิคการสำเนาข้อสอบ (item cloning) หรือการใช้โมเดลเชิงโครงสร้างระดับลดหลั่น (HLM SEM) เป็นต้น แต่จากงานวิจัยที่ผ่านมาพบว่า แม้จะใช้โมเดลการตอบสนองข้อสอบเชิงสุ่มแต่ส่วนใหญ่ก็ยังศึกษาแบบ Fixed item หรือมีการใช้ข้อมูลจริงในการการศึกษาแต่มีจำนวนข้อสอบน้อย กลุ่มตัวอย่างขนาดเล็ก และใช้โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงสนใจนำโมเดลการตอบสนองข้อสอบแบบอิทธิพลสุ่มมาใช้กับข้อมูลจำลองโดยใช้จำนวนข้อสอบขนาดต่างๆกัน ขนาดกลุ่มตัวอย่างขนาดเล็ก กลาง ใหญ่ และลักษณะการแจกแจงของพารามิเตอร์ข้อสอบแบบต่างๆกัน

5.2 งานวิจัยเกี่ยวกับการประมาณค่าพารามิเตอร์

De Boeck (2008) ศึกษาอิทธิพลสุ่มข้อสอบของโมเดลประมาณค่าพารามิเตอร์แบบ 1 พารามิเตอร์ ด้วยแบบวัดพฤติกรรมความก้าวร้าวที่ตอบสนองทางภาษา 24 ข้อ เลือกตอบ 0, 1 เก็บข้อมูลจากนักศึกษาปริญญาตรี ปี 1 จำนวน 316 คน เปรียบเทียบประสิทธิภาพของโมเดลประมาณค่า 4 โมเดล คือ รูปแบบเจาะจงผู้สอบ-เจาะจงข้อสอบ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ ผลการวิจัยพบว่า ค่าสถิติ BIC ของรูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบมีค่าต่ำที่สุด รองลงมาคือ รูปแบบเจาะจงผู้สอบ-เจาะจงข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ ตามลำดับ แต่สำหรับค่าสถิติ DIC และ AIC ให้ผลที่แตกต่างออกไป คือ รูปแบบเจาะจงผู้สอบ-เจาะจงข้อสอบ มีค่า DIC และ AIC ต่ำที่สุด รองลงมาคือ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ ตามลำดับ ผู้วิจัยเรื่องนี้ได้สรุปไว้ว่าผลการวิจัยที่ได้ยังไม่แน่นอน อาจเป็นเพราะมีจำนวนข้อสอบน้อยเกินไป

วิชุดา บัวคง (2533) ศึกษาเปรียบเทียบประสิทธิผลของวิธีประมาณค่าพารามิเตอร์ของแบบจำลองโลจีสติก 3 พารามิเตอร์ ระหว่างวิธีแมกซิมัมไลค์ลิฮูด วิธีฮิวริสติก และวิธีของเบส์ ในแบบสอบวัดผลสัมฤทธิ์ และแบบสอบความถนัด ในวิชาภาษาไทย จำนวน 40 และ 80 ข้อ

ตามลำดับ เก็บข้อมูลจากนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 1,028 คน โดยพิจารณาเปรียบเทียบประสิทธิภาพสัมพัทธ์ของแบบสอบด้วยการหาค่าอัตราส่วนค่าฟังก์ชันสารสนเทศของแบบสอบระหว่างวิธีประมาณค่าพารามิเตอร์ 2 วิธีที่ระดับความสามารถเดียวกัน ผลการวิจัยพบว่า (1) การประมาณค่าพารามิเตอร์ข้อสอบในแบบสอบวัดผลสัมฤทธิ์ในกลุ่มผู้มีความสามารถสูง วิธีประมาณค่าที่มีประสิทธิภาพสูงสุด คือ วิธีแมกซิมัมไลค์ลิฮูด รองลงมาคือ วิธีของเบย์ และวิธีฮิวริสติก ตามลำดับ แต่ในกลุ่มผู้มีความสามารถปานกลางและต่ำที่วัด วิธีประมาณค่าที่มีประสิทธิภาพสูงสุด คือ วิธีของเบย์ รองลงมาคือ วิธีแมกซิมัมไลค์ลิฮูด และวิธีฮิวริสติก ตามลำดับ และ (2) การประมาณค่าพารามิเตอร์ข้อสอบในแบบสอบความถนัดในกลุ่มผู้มีความสามารถสูง วิธีประมาณค่าที่มีประสิทธิภาพสูงสุด คือ วิธีแมกซิมัมไลค์ลิฮูด รองลงมาคือ วิธีของเบย์ และวิธีฮิวริสติก ตามลำดับ สอดคล้องกับแบบวัดผลสัมฤทธิ์ ส่วนกลุ่มผู้มีความสามารถปานกลาง วิธีประมาณค่าที่มีประสิทธิภาพสูงสุด คือ วิธีของเบย์ รองลงมาคือ วิธีแมกซิมัมไลค์ลิฮูด และวิธีฮิวริสติก ตามลำดับ แต่กลุ่มผู้มีความสามารถต่ำ วิธีประมาณค่าที่มีประสิทธิภาพสูงสุด คือ วิธีของเบย์ รองลงมาคือ วิธีฮิวริสติก และวิธีแมกซิมัมไลค์ลิฮูด ตามลำดับ

วินัย วงศ์ฤทัยวัฒนา (2533) เปรียบเทียบผลการประมาณค่าพารามิเตอร์ของโมเดลโลจิสติกแบบสองพารามิเตอร์ระหว่างวิธีของเบย์กับวิธีแมกซิมัมไลค์ลิฮูด ในแบบสอบวิชาคณิตศาสตร์ เก็บรวบรวมข้อมูลจากนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 323 คน โดยพิจารณาเปรียบเทียบจากค่าความแตกต่างของค่าประมาณความยากและอำนาจจำแนกแต่ละคู่ด้วยสถิติทดสอบ t สำหรับกลุ่มตัวอย่างที่มีความสัมพันธ์กัน ผลการวิจัยพบว่า การประมาณค่าความยากค่าด้วยวิธีของเบย์มีค่าแตกต่างกันกับวิธีแมกซิมัมไลค์ลิฮูด แต่การประมาณค่าอำนาจจำแนกด้วยวิธีของเบย์มีค่าไม่แตกต่างกันกับวิธีแมกซิมัมไลค์ลิฮูด

ชนะศึก นิชานนท์ (2553) ศึกษาประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบเบย์โดยใช้การสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ (GIRM) 4 รูปแบบ ได้แก่ (1) Original GIRM (2) AGIRM-A (3) AGIRM-B และ (4) Numerical Bayesian GIRM และอิทธิพลของข้อมูลจำลองเงื่อนไขขนาดกลุ่มตัวอย่างและจำนวนข้อสอบขนาดต่างๆ และการแจกแจงก่อนหน้าของผู้สอบและข้อสอบต่างๆกัน รวม 64 เงื่อนไข โดยศึกษาจากประสิทธิภาพของวิธีการประมาณค่า 4 แบบ คือ ความลำเอียงในการประมาณค่า ความไม่แน่นอนในการประมาณค่า ประสิทธิภาพขององค์ประกอบความแปรปรวน และความไวของรูปแบบการประมาณค่า ผลการวิจัยพบว่า (1) ขนาดกลุ่มตัวอย่างมากขึ้น ความลำเอียงในการประมาณค่าในรูปแบบ 1, 2 และ 3 จะลดลง และมีความแน่นอนในการประมาณค่าในทุกรูปแบบเพิ่มขึ้น แต่ไม่ส่งผลต่อ

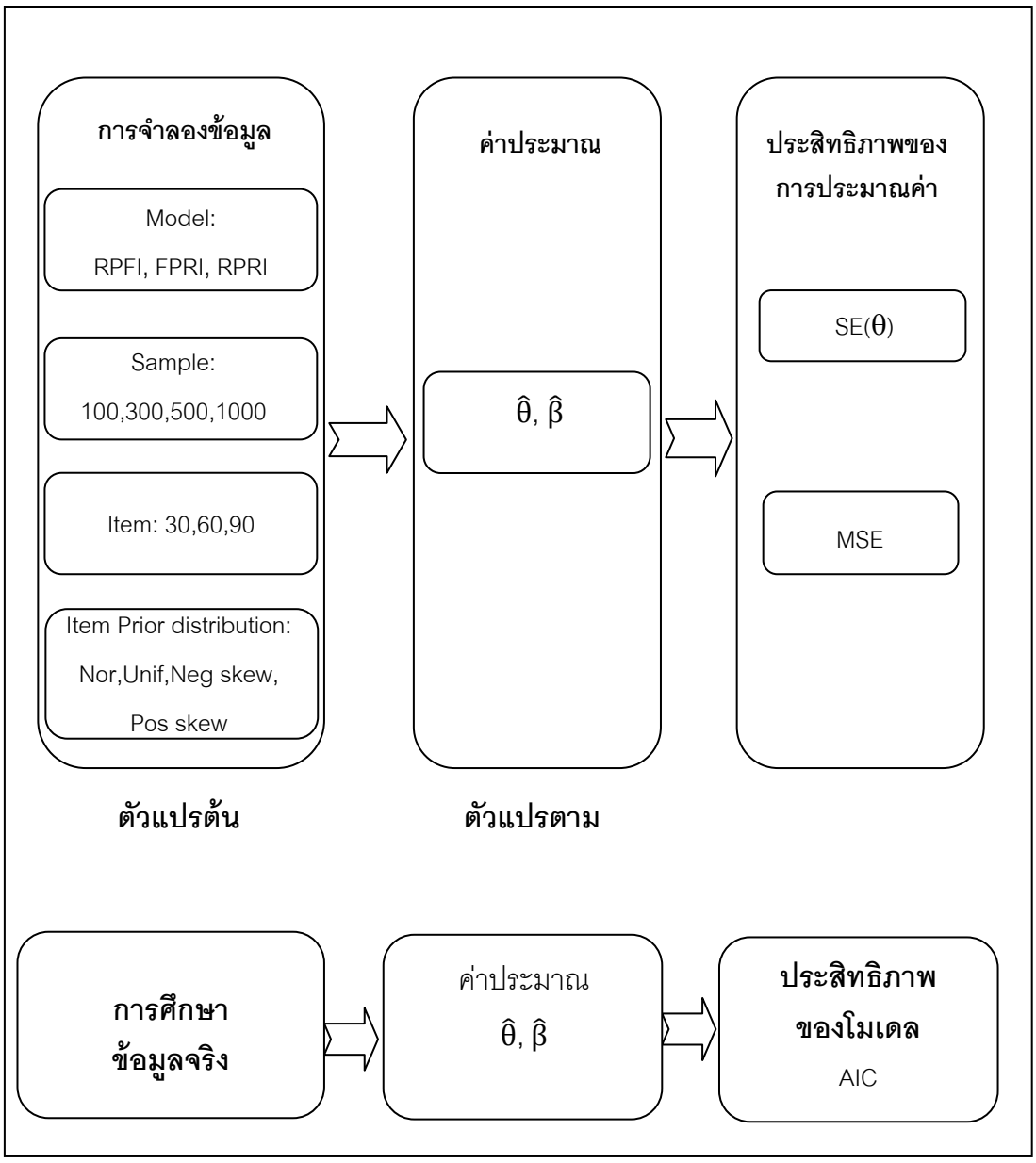
ประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนในทุกรูปแบบ (2) ความยาวแบบสอบมากขึ้น ความลำเอียงในการประมาณค่าในรูปแบบ 1, 2 และ 3 จะลดลง มีประสิทธิภาพในการประมาณค่าองค์ประกอบความแปรปรวนในรูปแบบ 1, 2 และ 3 เพิ่มขึ้น และมีความแน่นอนในการประมาณค่าเพิ่มขึ้นในทุกรูปแบบ (3) ความไวของรูปแบบ จากการแจกแจงก่อนหน้าของผู้สอบแบบแกมมามีความแน่นอนในการประมาณค่ามากกว่าการแจกแจงก่อนหน้าของผู้สอบแบบปกติ และการแจกแจงก่อนหน้าของผู้สอบส่งผลประสิทธิภาพขององค์ประกอบความแปรปรวนในการประมาณค่าในทุกรูปแบบ แต่การแจกแจงก่อนหน้าของผู้สอบไม่ส่งผลต่อความลำเอียงในการประมาณค่า (4) ความไวของรูปแบบ จากการแจกแจงก่อนหน้าของข้อสอบแบบปกติ มีความลำเอียงในการประมาณค่าต่ำกว่าการแจกแจงก่อนหน้าของข้อสอบแบบเบต้า ยกเว้นกรณีที่มีการแจกแจงก่อนหน้าของผู้สอบแบบแกมมาจะให้ผลตรงข้ามกัน และการแจกแจงก่อนหน้าของข้อสอบแบบเบต้ามีความแน่นอนในการประมาณค่าในทุกรูปแบบต่ำกว่าการแจกแจงก่อนหน้าของข้อสอบแบบปกติ และการแจกแจงก่อนหน้าของข้อสอบแบบปกติมีประสิทธิภาพขององค์ประกอบความแปรปรวนในการประมาณค่าสูงกว่าการแจกแจงก่อนหน้าของข้อสอบแบบเบต้าเฉพาะเมื่อมีการแจกแจงก่อนหน้าของผู้สอบแบบแกมมาเท่านั้น

ชลี ภัทรพิชญธรรม (2553) ศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง ด้วยโมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น ด้วยการประมาณค่าพารามิเตอร์ ได้เปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบ 4 โมเดล ด้วยข้อมูลจำลองเงื่อนไขจำนวนผู้สอบ ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม และขนาดปฏิสัมพันธ์แบบ 2 ทาง รวม 18 เงื่อนไข โดยพิจารณาประสิทธิภาพของโมเดลจาก ความลำเอียง ความคลาดเคลื่อนมาตรฐาน และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย ผลการวิจัยพบว่า ขนาดกลุ่มตัวอย่างมากขึ้น จะมีค่าความลำเอียง ค่าความคลาดเคลื่อนมาตรฐานและค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ยลดลง

ตอนที่ 6 กรอบแนวคิดในการวิจัย

จากการทบทวนเอกสาร ศึกษาทฤษฎี และรายงานการวิจัยที่เกี่ยวข้อง ผู้วิจัยจึงได้กำหนดกรอบความคิดจากการศึกษาดังกล่าว โดยมีกระบวนการดังนี้

การใช้โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์อิทธิพลสุ่มข้อสอบมาใช้ในการประมาณค่า เนื่องจากโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ มีจำนวนพารามิเตอร์น้อยทำให้ไม่ยุ่งยากซับซ้อนและง่ายต่อการวิเคราะห์ (Hambleton, 1989) และยังให้ค่าประมาณของพารามิเตอร์ที่ถูกต้องกว่าโมเดลอื่นๆเมื่อนำไปใช้กับกลุ่มตัวอย่างขนาดเล็ก (van de Vijver, 1986 อ้างถึงใน Baker, 1992 ; Lord, 1980) โดยเปรียบเทียบประสิทธิภาพของการประมาณค่าระหว่างโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์อิทธิพลสุ่มข้อสอบกับโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์อิทธิพลเจาะจงข้อสอบ เนื่องจากมีแนวโน้มว่าโมเดลอิทธิพลสุ่มข้อสอบจะมีประสิทธิภาพดีกว่าโมเดลอิทธิพลเจาะจงข้อสอบ (De Boeck, 2008) โดยศึกษาทั้งวิธีการจำลองข้อมูล (Simulation) เนื่องจากการศึกษาครั้งนี้เป็นการทดสอบทฤษฎีซึ่งจะดำเนินการได้สะดวกยิ่งขึ้นเมื่อศึกษาด้วยวิธีการจำลองข้อมูล (ศิริชัย กาญจนวาสี และคณะ, 2551) และศึกษาโดยใช้ข้อมูลเชิงประจักษ์ (Real data) สำหรับการศึกษาข้อมูลเชิงประจักษ์ ผู้วิจัยสนใจเปรียบเทียบประสิทธิภาพของโมเดลอิทธิพลสุ่มข้อสอบและโมเดลอิทธิพลเจาะจงข้อสอบ เนื่องจากมีแนวโน้มว่าโมเดลอิทธิพลสุ่มข้อสอบจะมีประสิทธิภาพของโมเดลสูงกว่า (De Boeck, 2008) โดยมีกรอบแนวคิด ดังนี้



ภาพที่ 6 กรอบแนวคิดในการวิจัย (conceptual framework)

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาประสิทธิภาพของการประมาณค่าของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ระหว่างการประมาณค่าโดยใช้รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบด้วยข้อมูลจำลอง และศึกษาความสอดคล้องของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ที่มีการประมาณค่าโดยโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ระหว่างการประมาณค่าโดยใช้รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบกับข้อมูลเชิงประจักษ์ วิเคราะห์ด้วยโปรแกรม WinBUGS โดยมีรายละเอียดวิธีดำเนินการวิจัย ประกอบด้วย 3 ส่วน คือ สถานการณ์การจำลองข้อมูล ขั้นตอนการวิจัย และการวิเคราะห์ข้อมูล ดังนี้

3.1 การจำลองข้อมูล

การจำลองข้อมูลโดยใช้โปรแกรม R จำลองประชากรในการศึกษาครั้งนี้ กำหนดให้การแจกแจงก่อนหน้าของพารามิเตอร์ผู้สอบ (prior distribution of person) เป็น Standard Normal (0,1) โดยการจำลองข้อมูลในสถานการณ์ต่างๆ ประกอบด้วย

1) รูปแบบของการสุ่มหรือการเจาะจงผู้สอบและข้อสอบ 3 แบบ คือ แบบสุ่มผู้สอบ-เจาะจงข้อสอบ แบบเจาะจงผู้สอบ-สุ่มข้อสอบ และแบบสุ่มผู้สอบ-สุ่มข้อสอบ
2) ขนาดกลุ่มตัวอย่างผู้สอบ (sample size) 4 ขนาด คือ 100, 300, 500,และ1,000 คน
3) จำนวนข้อสอบ (item) 30, 60, 90 ข้อ
4) การแจกแจงก่อนหน้าของพารามิเตอร์ความยาก (prior distribution of item difficulty) 4 แบบ คือ Normal, Uniform, Negative skew และ Positive skew

3.2 ขั้นตอนการวิจัย

การศึกษาในครั้งนี้มีขั้นตอนการวิจัย ตามลำดับดังนี้

1) ศึกษาค้นคว้ามโนทัศน์ของทฤษฎีการตอบสนองข้อสอบ การประมาณค่าพารามิเตอร์ ประสิทธิภาพของการประมาณค่า ประสิทธิภาพของโมเดล จากเอกสาร หนังสือ วารสารวิชาการ และงานวิจัยที่ผ่านมา

2) ศึกษาโปรแกรมที่จะนำมาใช้จำลองข้อมูล และวิเคราะห์ข้อมูล ได้แก่ โปรแกรม R และ WinBUGS

3) การจำลองข้อมูล ด้วยการเขียนคำสั่ง algorithm ให้จำลองข้อมูลผ่านโปรแกรม R โดยจำลองข้อมูลตามเงื่อนไข และตรวจสอบความสมบูรณ์ของข้อมูลที่จำลองขึ้น

4) การวิเคราะห์ประมาณค่าพารามิเตอร์ความสามารถผู้สอบและพารามิเตอร์ความยากของข้อสอบจากข้อมูลที่จำลองขึ้น ด้วยโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ที่มีการสุ่ม 3 แบบ คือ แบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) แบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) จากเงื่อนไขของการจำลองข้อมูลกลุ่มตัวอย่าง 100, 300, 500 และ 1,000 คน จำนวนข้อสอบ 30, 60 และ 90 ข้อ และลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบ 4 แบบ คือ Normal, Uniform, Negative skew และ Positive skew โดยการส่งผ่านข้อมูลที่จำลองแล้วด้วยชุดคำสั่ง R2WINBUGS เพื่อวิเคราะห์ข้อมูลด้วยโปรแกรม WINBUGS

5) การวิเคราะห์ประสิทธิภาพของการประมาณค่าจากข้อมูลจำลอง โดยพิจารณาจากค่าความคลาดเคลื่อนในการประมาณค่า และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่า

6) การศึกษากับข้อมูลเชิงประจักษ์ โดยจะใช้ข้อมูลจากผลการสอบ O-NET มาวิเคราะห์ประมาณค่าพารามิเตอร์ความสามารถผู้สอบและพารามิเตอร์ความยากของข้อสอบโดยใช้โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ที่มีการสุ่ม 3 แบบ คือ แบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) แบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI)

7) เปรียบเทียบประสิทธิภาพของการประมาณค่าของข้อมูลจำลอง โดยพิจารณาจากค่าความคลาดเคลื่อนในการประมาณค่า และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่า และเปรียบเทียบประสิทธิภาพของโมเดล โดยพิจารณาความสอดคล้องของโมเดลกับข้อมูลเชิงประจักษ์ จากค่าสถิติ Akaike information criterion (AIC) ที่ใช้เป็นเกณฑ์ในการคัดเลือกโมเดล ค่า AIC ที่ต่ำที่สุดคือโมเดลที่ดีที่สุด

8) สรุปผลการวิเคราะห์ สรุปผลการวิจัย อภิปรายผล และเขียนรายงานการวิจัย

การจำลองข้อมูล ในกรณีศึกษาการจำลองข้อมูล (Simulation data)

ผู้วิจัยศึกษาในสถานการณ์จำลอง โดยใช้โปรแกรม R เวอร์ชั่น 2.13.0 ในการจำลองข้อมูล และใช้โปรแกรม WinBUGS ในการประมาณค่าพารามิเตอร์ มีลำดับขั้นตอนดังนี้

(1) กำหนดลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ผู้สอบเป็น Normal (0, 1)

(2) กำหนดลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากของข้อสอบ 4 แบบ คือ Normal, Uniform, Negative skew และ Positive skew

(3) ความยาวของแบบสอบทั้งฉบับ (Test length) 3 ระดับ คือ 30, 60 และ 90 ข้อ

(4) ขนาดกลุ่มตัวอย่าง (sample size) 4 ระดับ คือ 100, 300, 500และ 1,000 คน

(5) ประมาณค่าพารามิเตอร์ความสามารถ θ และพารามิเตอร์ความยาก b ด้วย โปรแกรม WinBUGS

(6) ทำซ้ำ 1,000 รอบ (จากข้อ (1) ถึงข้อ (6))

(7) คำนวณค่าความคลาดเคลื่อนในการประมาณค่า และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่า โดยนำเสนอขั้นตอนการดำเนินการวิจัยกรณีการจำลองข้อมูล ดังภาพที่ 4

การศึกษาข้อมูลจริง (Real data)

(1) ทำหนังสือขอความอนุเคราะห์ข้อมูลการสอบ O-NET ปีการศึกษา 2553 จากสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน)

(2) เลือกศึกษาข้อมูลผลการสอบ O-NET ปีการศึกษา 2553 ระดับมัธยมศึกษาปีที่ 6 รายวิชาภาษาอังกฤษ คณิตศาสตร์ และภาษาไทย เนื่องจากมีการให้คะแนนแบบสองค่า

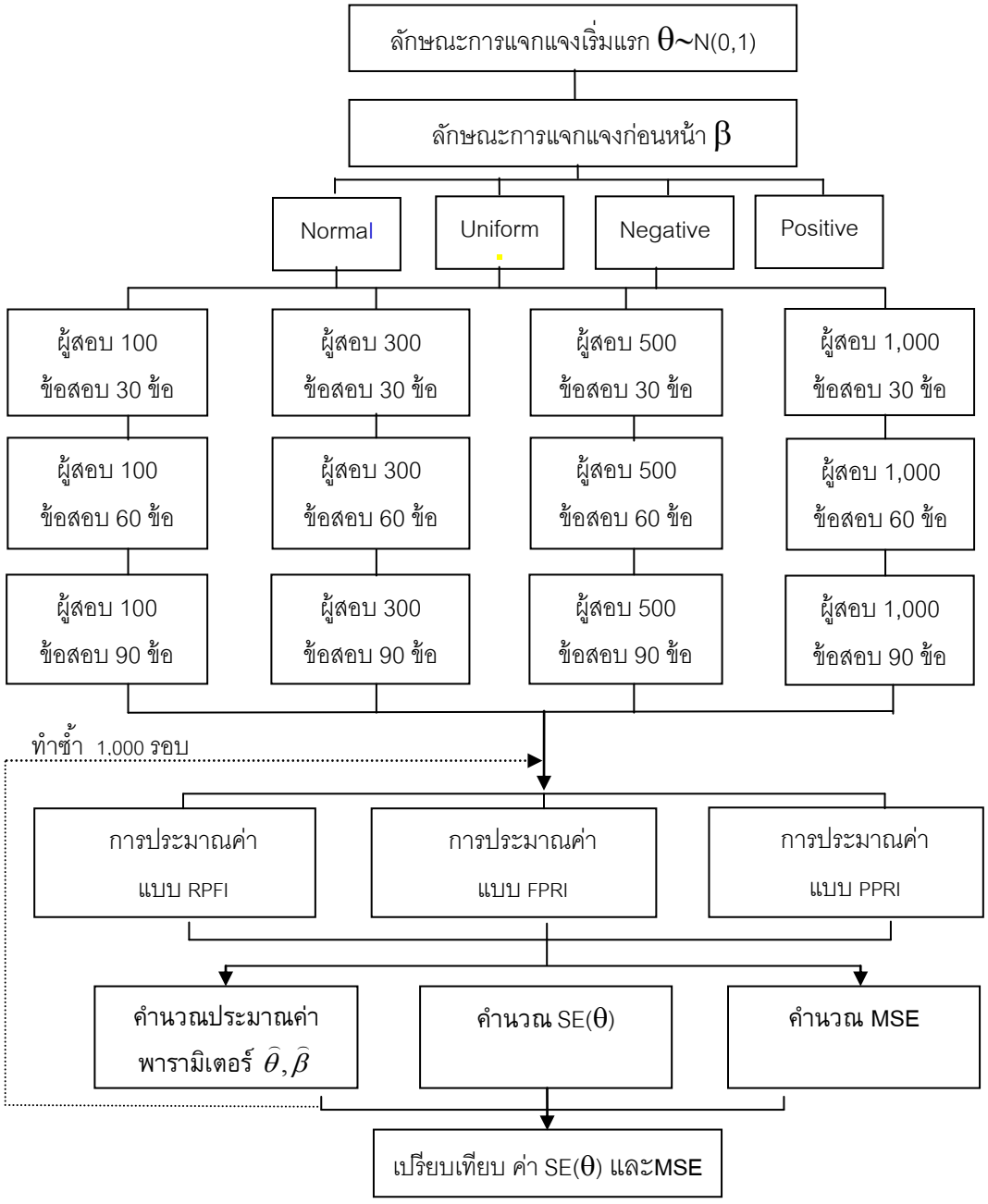
(3) นำข้อมูลคะแนนรวมมาเขียนการกระจายด้วย Histogram และจัดกระทำข้อมูลโดยตัดข้อมูลของผู้สอบที่มีคะแนนต่ำกว่าหรือเท่ากับคะแนนที่น่าจะเกิดจากการเดา โดยคำนวณจาก Guessing = [(1/4) – 0.05] x จำนวนข้อสอบ; เมื่อมีโอกาสตอบถูก 1 ใน 4 ตัวเลือก

(4) ใช้โปรแกรม R เขียนคำสั่งสุ่มจำนวนกลุ่มตัวอย่าง 100, 300, 500 และ 1,000 คน จากข้อมูลที่เหลือจากข้อ (3)

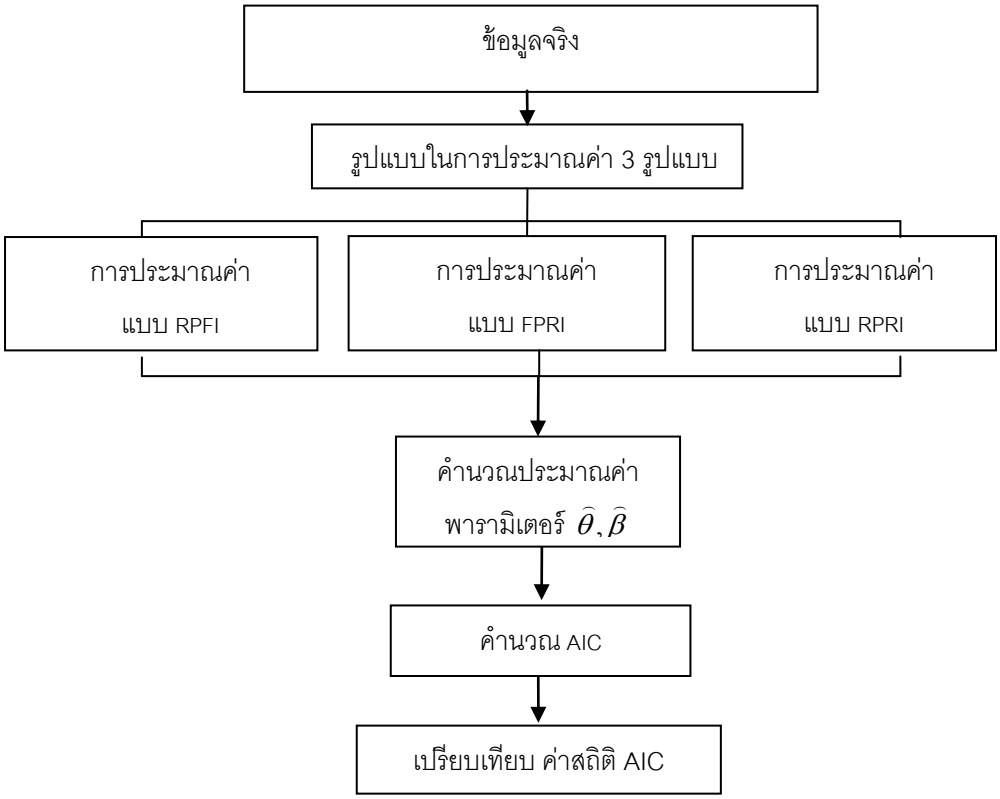
(5) ประมาณค่าพารามิเตอร์ความสามารถ (θ)

(6) คำนวณค่าสถิติ Akaike information criterion (AIC)

(7) ศึกษาเปรียบเทียบประสิทธิภาพความสอดคล้องของโมเดล



ภาพที่ 7 ขั้นตอนการดำเนินการวิจัยกรณีการจำลองข้อมูล



ภาพที่ 8 ขั้นตอนการดำเนินการวิจัยกรณีข้อมูลจริง

3.3 การวิเคราะห์ข้อมูล

1) ศึกษาและเปรียบเทียบประสิทธิภาพของการประมาณค่าของโมเดลราสช์ 1 พารามิเตอร์ 3 รูปแบบ คือ รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบจากเงื่อนไขจำลองข้อมูลแต่ละเงื่อนไข การวัดประสิทธิภาพของการประมาณค่า โดยพิจารณาจาก

(1) ความคลาดเคลื่อนมาตรฐานในการประมาณค่า ; SE(θ)

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

(2) ค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean squared error ; MSE) (Swaminathan, Hambleton, Sireci, Xing และ Rizavi, 2003)

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$$

$$MSE(\hat{\theta}) = (Bias(\hat{\theta}, \theta))^2 + Variance(\hat{\theta})$$

2) ศึกษาและเปรียบเทียบประสิทธิภาพของโมเดลด้านความสอดคล้องของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ 3 รูปแบบ คือ รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบ จากข้อมูลจริง โดยพิจารณาจากค่าสถิติ Akaike information criterion (AIC) เป็นเกณฑ์ในการคัดเลือกโมเดล ค่า AIC ที่ต่ำที่สุดคือโมเดลที่ดีที่สุด คำนวณได้จากสูตร

$$AIC = 2k - 2\ln(L)$$

เมื่อ k เป็นจำนวนพารามิเตอร์ในโมเดล และ L คือค่าสูงสุดของฟังก์ชันความน่าจะเป็น (Likelihood function) ของโมเดลที่ประมาณค่า

การวิเคราะห์เปรียบเทียบประสิทธิภาพ

1) การเปรียบเทียบประสิทธิภาพของการประมาณค่า พิจารณาจากดัชนีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่า โดยเปรียบเทียบค่าที่ได้ ถ้าเงื่อนไขการศึกษาใดมีค่าที่น้อยกว่า แสดงว่ามีประสิทธิภาพของการประมาณค่าพารามิเตอร์ได้ดีกว่า และเปรียบเทียบประสิทธิภาพของการประมาณค่าด้วยสถิติทดสอบ F-test

2) การเปรียบเทียบประสิทธิภาพของโมเดล พิจารณาจากความสอดคล้องของโมเดลกับข้อมูล โดยดูจากค่าสถิติ AIC ถ้ามีค่าต่ำกว่า แสดงว่าโมเดลมีความสอดคล้องกับข้อมูลมากกว่าประสิทธิภาพของโมเดลสูงกว่า

3.4 การดำเนินการวิจัย

การดำเนินการวิจัยครั้งนี้ แบ่งออกเป็น 2 ช่วง คือ

1) การดำเนินการวิจัยโดยใช้ข้อมูลจำลอง โดยใช้โปรแกรม R และวิเคราะห์ข้อมูลด้วยโปรแกรม WinBUGS ตัวแปรที่ใช้ในการศึกษา ประกอบด้วย

ตัวแปรต้น ได้แก่ (1) โมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ที่มีการสุ่ม 3 แบบ คือ แบบสุ่มผู้สอบ-เจาะจงข้อสอบ แบบเจาะจงผู้สอบ-สุ่มข้อสอบ และแบบสุ่มผู้สอบ-สุ่มข้อสอบ (2) ขนาดกลุ่มตัวอย่าง 4 ระดับ คือ 100, 300, 500 และ 1,000 คน (3) จำนวนข้อสอบ 3 ระดับ คือ 30, 60 และ 90 ข้อ และ4) การวิเคราะห์ Prior Distribution ความยากของข้อสอบ 4

แบบ คือ Normal, Uniform, Negative skew และ Positive skew รวมเงื่อนไขที่จำลองข้อมูล 96 เงื่อนไข ในแต่ละเงื่อนไขทำซ้ำ 1,000 รอบ

ตัวแปรตาม คือ (1) ค่าประมาณความสามารถของผู้สอบ ($\hat{\theta}$) และ (2) ค่าประมาณความยากของข้อสอบ ($\hat{\beta}$)

เปรียบเทียบประสิทธิภาพในการประมาณค่า โดยคำนวณและเปรียบเทียบค่าความคลาดเคลื่อนในการประมาณค่า (SE(θ)) และค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Mean squared error; MSE) ด้วยสถิติ F-test

2) การวิเคราะห์ข้อมูลผลสอบ O-NET ปีการศึกษา 2553 ระดับมัธยมศึกษาปีที่ 6 รายวิชาภาษาอังกฤษ คณิตศาสตร์ และภาษาไทย โดยใช้โปรแกรม R สุ่มขนาดกลุ่มตัวอย่าง 4 ระดับ ได้แก่ 100, 300, 500 และ 1,000 คน ในแต่ละรายวิชา สุ่มแบบใส่กลับคืนและทำซ้ำ (bootstrap sampling) 100 รอบ วิเคราะห์ค่าประมาณความสามารถของผู้สอบ ($\hat{\theta}$) ค่าประมาณความยากของข้อสอบ ($\hat{\beta}$) และค่า AIC ของแต่ละโมเดลโดยใช้โปรแกรม WinBUGS เปรียบเทียบค่าเฉลี่ย AIC ของโมเดลการประมาณค่า 3 รูปแบบ ด้วยสถิติ F-test

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มุ่งศึกษาประสิทธิภาพของการประมาณค่าของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ระหว่างการประมาณค่าโดยใช้รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบด้วยข้อมูลจำลอง และศึกษาความสอดคล้องของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ที่มีการประมาณค่าโดยใช้รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบกับข้อมูลเชิงประจักษ์ โดยศึกษาตามขนาดกลุ่มตัวอย่างและจำนวนข้อสอบต่างๆกัน ในบทนี้ผู้วิจัยจึงนำเสนอแบ่งออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 การวิเคราะห์ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบจากข้อมูลจำลอง: ความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์

ตอนที่ 2 การวิเคราะห์ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบจากข้อมูลจำลอง: ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าพารามิเตอร์

ตอนที่ 3 การวิเคราะห์ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบจากข้อมูลเชิงประจักษ์

ตอนที่ 1 การวิเคราะห์ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบจากข้อมูลจำลอง: ความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์

คุณภาพของการประมาณค่าพารามิเตอร์เป็นสิ่งสำคัญยิ่งต่อกระบวนการวัดและประเมินผลการศึกษา เนื่องจากคุณภาพของการประมาณค่านำไปสู่ความน่าเชื่อถือของค่าประมาณพารามิเตอร์ที่ได้ และสร้างความเชื่อมั่นให้กับผู้นำผลการประมาณค่าไปใช้ รวมถึงการให้การยอมรับผลสอบของผู้สอบและผู้มีส่วนเกี่ยวข้อง ตัวบ่งชี้หนึ่งที่สำคัญซึ่งบ่งชี้ถึงคุณภาพของการประมาณค่าพารามิเตอร์ คือ ค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ (SE (θ)) ถ้ามีค่าความคลาดเคลื่อนในการประมาณค่าต่ำกว่าแสดงถึงประสิทธิภาพในการประมาณค่าพารามิเตอร์ที่ดีกว่า

ในการวิเคราะห์เปรียบเทียบคุณภาพของการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลการประมาณค่า 3 รูปแบบ คือ (1) รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2) รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ตามทฤษฎีการตอบสนองข้อสอบ (IRT) โดยการเปรียบเทียบค่าความคลาดเคลื่อนของการประมาณ

ค่าพารามิเตอร์ผู้สอบที่ได้จากโมเดลการประมาณค่า 3 รูปแบบดังกล่าว ดังรายละเอียดของผลการวิเคราะห์ ดังนี้

1.1 การวิเคราะห์ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ

การประมาณค่าพารามิเตอร์ผู้สอบเปรียบเทียบระหว่างโมเดลประมาณค่าพารามิเตอร์ผู้สอบ 3 รูปแบบ คือ คือ (1) รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2) รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ตามทฤษฎีการตอบสนองข้อสอบ (IRT) โดยใช้ข้อมูลจำลองด้วยโปรแกรม R ตามเงื่อนไขกรณีศึกษา 48 เงื่อนไข มีผลการวิเคราะห์ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ดังนี้

1.1.1 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ จำแนกตามจำนวนข้อสอบและการแจกแจงพารามิเตอร์ความยากข้อสอบ

ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบจากโมเดล 3 รูปแบบ คือ (1) รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2) รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก (prior distribution of item difficulty) แจกแจงแบบ Normal, Uniform Positive Skew และ Negative Skew วิเคราะห์เปรียบเทียบเงื่อนไขกรณีศึกษา 12 เงื่อนไข พบว่า

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน พบว่า โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในทุกระดับของจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ ที่มีลักษณะการแจกแจงของพารามิเตอร์ความยากข้อสอบทั้ง Normal, Uniform และ Negative skew และโมเดลการประมาณค่าที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาคือ โมเดลสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) และโมเดลสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ตามลำดับ

ยกเว้นเมื่อมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Positive Skew ในทุกระดับของจำนวนข้อสอบทั้ง 30 และ 60 ข้อ พบว่า โมเดลสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่สำหรับในทุกระดับของจำนวนข้อสอบ 90 ข้อ พบว่า โมเดลสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างไม่มีนัยสำคัญทางสถิติ และโมเดลการประมาณค่าที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์

ผู้สอบต่ำรองลงมาคือ โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และโมเดลสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ตามลำดับ ดังตารางที่ 4.1

ตารางที่ 4.1 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	Normal	297.476	2.908	294.065	2.700	294.699	2.789	41.974	.000**
	Uniform	343.118	0.856	341.866	0.768	341.991	0.787	73.45	.000**
	Negative	331.747	1.184	330.346	1.087	330.490	1.100	46.943	.000**
	Positive	353.124	2.593	356.064	2.474	356.547	2.528	53.556	.000**
Post Hoc	LSD: Normal: FPRI < RPFI**; RPRI < RPFI **; FPRI < RPRI LSD: Uniform: FPRI < RPFI**; RPRI < RPFI **; FPRI < RPRI LSD: Negative: FPRI < RPFI**; RPRI < RPFI **; FPRI < RPRI LSD: Positive: RPFI < FPRI **; RPFI < RPRI **; FPRI < RPRI								
60	Normal	213.058	2.648	210.961	2.507	211.415	2.58	18.301	.000**
	Uniform	245.673	0.683	244.732	0.621	244.772	0.621	68.693	.000**
	Negative	240.219	1.366	239.447	1.273	239.448	1.27	11.679	.000**
	Positive	255.44	2.068	261.466	2.133	261.652	2.128	280.632	.000**
Post Hoc	LSD: Normal: FPRI < RPFI**; RPRI < RPFI **; FPRI < RPRI LSD: Uniform: FPRI < RPFI**; RPRI < RPFI **; FPRI < RPRI LSD: Negative: FPRI < RPFI**; RPRI < RPFI **; FPRI < RPRI LSD: Positive: RPFI< FPRI **; RPFI < RPRI **; FPRI < RPRI								
90	Normal	174.21	2.288	172.944	2.159	173.153	2.198	9.376	.000**
	Uniform	209.202	0.413	208.802	0.36	208.835	0.368	33.908	.000**
	Negative	198.209	1.104	197.986	1.014	197.982	1.021	1.54	0.216
	Positive	201.46	0.96	203.716	0.928	203.774	0.941	195.832	.000**
Post Hoc	LSD: Normal: FPRI < RPFI**; RPRI < RPFI **; FPRI < RPRI LSD: Uniform: FPRI < RPFI**; RPRI < RPFI **; FPRI < RPRI LSD: Negative: FPRI < RPRI < RPFI LSD: Positive: RPFI< FPRI **; RPFI < RPRI **; FPRI < RPRI								

*p<.05 **p<.01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน ในระดับของจำนวนข้อสอบ 30 ข้อ ที่มีลักษณะการแจกแจงของพารามิเตอร์ความยากข้อสอบทั้ง Normal, Uniform และ Negative skew พบว่า โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่สำหรับลักษณะการแจกแจงของพารามิเตอร์ความยากข้อสอบ Positive skew พบว่า โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุดอย่างไม่มีนัยสำคัญ และโมเดลการประมาณค่าที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาคือ โมเดลสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) และโมเดลสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ตามลำดับ

สำหรับขนาดกลุ่มตัวอย่างผู้สอบ 300 คน ในระดับของจำนวนข้อสอบ 60 ข้อ ที่มีลักษณะการแจกแจงของพารามิเตอร์ความยากข้อสอบทั้ง Normal, Negative skew และ Positive skew พบว่า พบว่า โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่เงื่อนไขที่มีลักษณะการแจกแจงของพารามิเตอร์ความยากข้อสอบแบบ Uniform พบว่าโมเดลสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

สำหรับขนาดกลุ่มตัวอย่างผู้สอบ 300 คน ในระดับของจำนวนข้อสอบ 90 ข้อ ที่มีลักษณะการแจกแจงของพารามิเตอร์ความยากข้อสอบทั้ง Normal และ Positive skew พบว่า โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่เงื่อนไขที่มีลักษณะการแจกแจงของพารามิเตอร์ความยากข้อสอบแบบ Uniform พบว่าโมเดลสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเงื่อนไขที่มีลักษณะการแจกแจงของพารามิเตอร์ความยากข้อสอบแบบ Negative skew พบว่าโมเดลสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังตารางที่ 4.2

ตารางที่ 4.2 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	Normal	295.819	3.323	295.036	3.240	295.391	3.275	4.287	.014*
	Uniform	350.281	0.730	350.061	0.707	350.098	0.713	8.058	.000**
	Negative	373.449	0.756	373.123	0.736	373.175	0.747	16.564	.000**
	Positive	380.068	1.099	378.553	21.956	378.602	21.959	0.691	0.501
Post Hoc	LSD: Normal : FPRI < RPFI**; FPRI < RPRI; RPRI < RPFI LSD: Uniform : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Negative : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Positive : FPRI < RPRI < RPFI								
60	Normal	210.683	2.587	210.005	2.532	210.073	2.546	6.406	.002**
	Uniform	258.725	0.593	259.034	0.581	259.047	0.587	28.906	.000**
	Negative	257.220	0.951	256.965	0.933	256.988	0.932	6.763	.001**
	Positive	242.592	0.857	242.250	0.822	242.253	0.827	16.6	.000**
Post Hoc	LSD: Normal : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Uniform : RPFI < FPRI **; RPFI < RPRI**; FPRI < RPRI LSD: Negative : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Positive : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI								
90	Normal	173.136	2.129	172.560	2.074	172.684	2.086	6.271	.002**
	Uniform	208.934	0.413	209.004	0.400	209.017	0.403	3.592	.028*
	Negative	205.859	0.857	205.626	0.838	205.620	0.837	7.812	.000**
	Positive	200.760	0.831	200.440	0.806	200.454	0.809	14.802	.000**
Post Hoc	LSD: Normal : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Uniform : RPFI < FPRI *; RPFI < RPRI *; FPRI < RPRI LSD: Negative : FPRI < RPFI**; RPRI < RPFI**; RPRI < FPRI LSD: Positive : FPRI < RPFI*; RPRI < RPFI*; FPRI < RPRI								

*p<.05 **p<.01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 ในระดับของจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ ที่มีลักษณะการแจกแจงของพารามิเตอร์ความยากข้อสอบทั้ง Normal, Uniform, Negative skew และ Positive skew พบว่า โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าความคลาดเคลื่อน

ในการประมาณค่าพารามิเตอร์ผู้สอบ ต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ.05 ยกเว้นเมื่อมีลักษณะการแจกแจงพารามิเตอร์ความยากข้อสอบ Normal ที่มีจำนวนข้อสอบ 60 ข้อ Negative skew ที่มีจำนวนข้อสอบ 30 และ 60 ข้อ และ Positive skew ที่มีจำนวนข้อสอบ 90 ข้อ ที่พบว่าโมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ต่ำที่สุด อย่างไม่มีนัยสำคัญทางสถิติ

และโมเดลการประมาณค่าที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาคือ โมเดลสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) และโมเดลสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ตามลำดับ ดังตารางที่ 4.3

ตารางที่ 4.3 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	Normal	293.483	2.968	292.752	2.897	292.779	2.914	10.037	.000**
	Uniform	326.048	0.972	325.778	0.940	325.807	0.956	12.045	.000**
	Negative	344.260	1.529	344.236	1.502	344.303	1.513	0.250	0.779
	Positive	328.420	1.924	327.940	1.867	328.085	1.888	8.463	.000**
Post Hoc	LSD: Normal : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Uniform : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Negative : FPRI < RPFI < RPRI LSD: Positive : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI								
60	Normal	211.077	2.520	210.715	2.499	210.817	2.512	2.769	0.063
	Uniform	237.282	0.703	237.121	0.690	237.145	0.695	7.823	.000**
	Negative	237.584	0.935	237.499	0.921	237.503	0.924	1.35	0.26
	Positive	245.407	1.098	245.015	1.080	245.015	1.080	21.786	.000**
Post Hoc	LSD: Normal : FPRI < RPRI < RPFI LSD: Uniform : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Negative : FPRI < RPRI < RPFI LSD: Positive : FPRI < RPFI**; RPRI < RPFI**; FPRI = RPRI								
90	Normal	173.002	2.315	172.606	2.276	172.659	2.288	4.398	.012*
	Uniform	196.977	0.482	196.841	0.470	196.846	0.473	13.205	.000**

ตารางที่ 4.3 (ต่อ)

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
	Negative	193.505	0.669	193.343	0.652	193.347	0.655	9.885	.000**
	Positive	191.631	0.781	191.542	0.770	191.557	0.771	1.929	0.146
Post Hoc	LSD: Normal : FPRI < RPFI*; RPRI < RPFI**; FPRI < RPRI LSD: Uniform : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Negative : FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Positive : FPRI < RPRI < RPFI								

*p<.05 **p<.01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ พบว่า โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ยกเว้นเงื่อนไขที่มีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Normal ที่จำนวนข้อสอบ 30 และ 90 ข้อ และการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Uniform ที่จำนวนข้อสอบ 30, 60 และ 90 ข้อ ที่พบว่า โมเดลเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบด้วยต่ำที่สุด อย่างไม่มีนัยสำคัญทางสถิติ และยกเว้นการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Negative Skew ที่จำนวนข้อสอบ 60 ข้อ ที่พบว่า โมเดลสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบด้วยต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังตารางที่ 4.4

ตารางที่ 4.4 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	Normal	294.558	2.998	294.348	2.987	294.455	2.993	1.235	0.291
	Uniform	326.935	1.224	326.831	1.213	326.850	1.213	2.08	0.125
	Negative	328.869	1.271	328.666	1.258	328.699	1.262	7.429	.001**
	Positive	336.806	1.027	336.668	1.021	336.686	1.021	5.364	.005**
Post Hoc	LSD: Normal: FPRI < RPRI < RPFI LSD: Uniform: FPRI < RPRI < RPFI LSD: Negative: FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI								

ตารางที่ 4.4 (ต่อ)

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
	LSD: Positive: FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI								
60	Normal	210.563	2.660	210.274	2.645	210.351	2.651	3.191	.041*
	Uniform	236.630	0.795	236.594	0.790	236.611	0.792	0.499	0.607
	Negative	230.255	0.981	230.118	0.971	230.106	0.972	7.203	.001**
	Positive	230.279	0.927	230.136	0.916	230.142	0.917	7.766	.000**
Post Hoc	LSD: Normal: FPRI < RPFI*; RPRI < RPFI ; FPRI < RPRI LSD: Uniform: FPRI < RPRI < RPFI LSD: Negative: FPRI < RPFI**; RPRI < RPFI**; RPRI < FPRI LSD: Positive: FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI								
90	Normal	172.644	2.281	172.434	2.261	172.442	2.267	2.751	0.064
	Uniform	199.343	0.404	199.303	0.398	199.304	0.400	3.341	0.036
	Negative	191.451	0.757	191.332	0.747	191.333	0.750	8.296	.000**
	Positive	192.437	0.706	192.336	0.698	192.344	0.698	6.403	.002**
Post Hoc	LSD: Normal: FPRI < RPRI < RPFI LSD: Uniform: FPRI < RPRI < RPFI LSD: Negative: FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI LSD: Positive: FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI								

*p<.05 **p<.01

สรุปการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลการประมาณค่าพารามิเตอร์ จำแนกตามจำนวนข้อสอบ การแจกแจงก่อนหน้าของพารามิเตอร์ความยาก และขนาดกลุ่มตัวอย่างผู้สอบ พบว่า เมื่อมีขนาดกลุ่มตัวอย่างผู้สอบ 100, 300, 500 และ 1,000 คน ส่วนใหญ่โมเดลการประมาณค่าพารามิเตอร์ผู้สอบที่มีค่าความคลาดเคลื่อนในการประมาณค่าต่ำที่สุดคือ โมเดลการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) รองลงมาคือ โมเดลการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) และโมเดลโมเดลการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ตามลำดับ ในทุกลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบ

1.1.2 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลการประมาณค่า 3 รูปแบบ คือ (1) รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2)

รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ในการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก 4 ลักษณะ คือ Normal, Uniform, Positive Skew และ Negative Skew วิเคราะห์เปรียบเทียบเงื่อนไขกรณีศึกษา 12 เงื่อนไข ดังนี้

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ พบว่า เมื่อมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก Normal โมเดลประมาณค่าทั้ง 3 โมเดล มีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากที่มีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาคือ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก Negative Skew, Uniform และ Positive Skew ตามลำดับ ยกเว้นที่จำนวนข้อสอบ 90 ข้อ ที่ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าต่ำรองลงมา คือ Negative Skew, Positive Skew และ Uniform ตามลำดับ ดังตารางที่ 4.5

ตารางที่ 4.5 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^3$)	SD ($x10^3$)	Mean ($x10^3$)	SD ($x10^3$)	Mean ($x10^3$)	SD ($x10^3$)	Mean ($x10^3$)	SD ($x10^3$)		
30	RPFI	297.476	2.908	343.118	0.856	331.747	1.184	353.124	2.593	13555.987	.000**
	FPRI	294.065	2.700	341.866	0.768	330.346	1.087	356.064	2.474	18532.685	.000**
	RPRI	294.699	2.789	341.991	0.787	330.49	1.100	356.547	2.528	17433.778	.000**
Post Hoc	Dunnett: RPFI: Nor<Neg < Uniform < Pos** Dunnett: FPRI: Nor<Neg < Uniform < Pos** Dunnett: RPRI: Nor<Neg < Uniform < Pos**										
60	RPFI	213.058	2.648	245.673	0.683	240.219	1.366	255.44	2.068	9680.236	.000**
	FPRI	210.961	2.507	244.732	0.621	239.447	1.273	261.466	2.133	13746.419	.000**
	RPRI	211.415	2.58	244.772	0.621	239.448	1.27	261.652	2.128	13223.514	.000**
Post Hoc	Dunnett: RPFI: Nor<Neg < Uniform < Pos** Dunnett: FPRI: Nor<Neg < Uniform < Pos** Dunnett: RPRI: Nor<Neg < Uniform < Pos**										

ตารางที่ 4.5 (ต่อ)

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^3$)	SD ($x10^3$)	Mean ($x10^3$)	SD ($x10^3$)	Mean ($x10^3$)	SD ($x10^3$)	Mean ($x10^3$)	SD ($x10^3$)		
90	RPFI	174.210	2.288	209.202	0.413	198.209	1.104	201.460	0.960	12080.909	.000**
	FPRI	172.944	2.159	208.802	0.360	197.986	1.014	203.716	0.928	15147.288	.000**
	RPRI	173.153	2.198	208.835	0.368	197.982	1.021	203.774	0.941	14526.357	.000**
Post Hoc	Dunnett: RPFI: Nor<Neg < Pos < Uniform **										
	Dunnett: FPRI: Nor<Neg < Pos < Uniform **										
	Dunnett: RPRI: Nor<Neg < Pos < Uniform **										

**p<.01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ พบว่า เมื่อมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Normal โมเดลประมาณค่าทั้ง 3 โมเดล มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุดอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 รองลงมาคือ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Positive Skew, Negative Skew และ Uniform ตามลำดับ ยกเว้นที่จำนวนข้อสอบ 30 ข้อ การแจกแจงก่อนหน้าของพารามิเตอร์ความยากที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าต่ำรองลงมา คือ Uniform, Negative Skew และ Positive Skew ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังตารางที่ 4.6

ตารางที่ 4.6 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน จำแนกตามจำนวนข้อสอบและรูปแบบโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^3$)	SD ($x10^3$)	Mean ($x10^3$)	SD ($x10^3$)	Mean ($x10^3$)	SD ($x10^3$)	Mean ($x10^3$)	SD ($x10^3$)		
30	RPFI	295.819	3.323	350.281	0.730	373.449	0.756	380.068	1.099	131486.34	.000**
	FPRI	295.036	3.240	350.061	0.707	373.123	0.736	378.553	21.956	3540.063	.000**
	RPRI	295.391	3.275	350.098	0.713	373.175	0.747	378.602	21.959	3510.628	.000**
Post Hoc	Dunnett: RPFI : Nor< Uniform<Neg< Pos**										
	Dunnett: FPRI : Nor< Uniform<Neg< Pos**										
	Dunnett: RPRI : Nor< Uniform<Neg< Pos**										
60	RPFI	210.683	2.587	258.725	0.593	257.220	0.951	242.592	0.857	68732.256	.000**

ตารางที่ 4.6 (ต่อ)

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
	FPRI	210.005	2.532	259.034	0.581	256.965	0.933	242.250	0.822	74179.793	.000**
	RPRI	210.073	2.546	259.047	0.587	256.988	0.932	242.253	0.827	73285.973	.000**
Post Hoc	Dunnett: RPFI : Nor<Pos<Neg<Uniform** Dunnett: FPRI : Nor<Pos<Neg<Uniform** Dunnett: RPRI : Nor<Pos<Neg<Uniform**										
90	RPFI	173.136	2.129	208.934	0.413	205.859	0.857	200.760	0.831	52483.994	.000**
	FPRI	172.560	2.074	209.004	0.400	205.626	0.838	200.440	0.806	56958.854	.000**
	RPRI	172.684	2.086	209.017	0.403	205.62	0.837	200.454	0.809	56024.882	.000**
Post Hoc	Dunnett: RPFI : Nor<Pos<Neg<Uniform** Dunnett: FPRI : Nor<Pos<Neg<Uniform** Dunnett: RPRI : Nor<Pos<Neg<Uniform**										

**p<.01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ พบว่า เมื่อมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal โมเดลประมาณค่าทั้ง 3 โมเดล มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบอื่นๆที่จำนวนข้อสอบแตกต่างกัน จะมีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาจากลักษณะการแจกแจงแบบ Normal แตกต่างกัน คือ จำนวนข้อสอบ 30 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ Uniform, Positive Skew และ Negative Skew ตามลำดับ จำนวนข้อสอบ 60 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ Uniform, Negative Skew และ Positive Skew ตามลำดับ และจำนวนข้อสอบ 90 ข้อ การแจกแจงก่อนหน้าของพารามิเตอร์ความยากที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ Positive Skew, Negative Skew และ Uniform ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังตารางที่ 4.7

ตารางที่ 4.7 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน จำแนกตามจำนวนข้อสอบและรูปแบบโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	RPFI	293.483	2.968	326.048	0.972	344.260	1.529	328.420	1.924	57488.649	.000**
	FPRI	292.752	2.897	325.778	0.940	344.236	1.502	327.940	1.867	62030.701	.000**
	RPRI	292.779	2.914	325.807	0.956	344.303	1.513	328.085	1.888	61175.937	.000**
Post Hoc	Dunnett: RPFI : Nor< Uniform< Pos < Neg **										
	Dunnett: FPRI : Nor< Uniform< Pos < Neg **										
	Dunnett: RPRI : Nor< Uniform< Pos < Neg **										
60	RPFI	211.077	2.520	237.282	0.703	237.584	0.935	245.407	1.098	50345.908	.000**
	FPRI	210.715	2.499	237.121	0.690	237.499	0.921	245.015	1.080	51704.707	.000**
	RPRI	210.817	2.512	237.145	0.695	237.503	0.924	245.015	1.080	50953.733	.000**
Post Hoc	Dunnett: RPFI : Nor < Uniform < Neg < Pos **										
	Dunnett: FPRI : Nor < Uniform < Neg < Pos **										
	Dunnett: RPRI : Nor < Uniform < Neg < Pos **										
90	RPFI	173.002	2.315	196.977	0.482	193.505	0.669	191.631	0.781	34757.129	.000**
	FPRI	172.606	2.276	196.841	0.470	193.343	0.652	191.542	0.770	36865.421	.000**
	RPRI	172.659	2.288	196.846	0.473	193.347	0.655	191.557	0.771	36361.772	.000**
Post Hoc	Dunnett: RPFI : Nor < Pos < Neg < Uniform **										
	Dunnett: FPRI : Nor < Pos < Neg < Uniform **										
	Dunnett: RPRI : Nor < Pos < Neg < Uniform **										

**p<.01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ พบว่า เมื่อมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Normal โมเดลประมาณค่าทั้ง 3 โมเดล มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุดอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่ลักษณะการแจกแจงของก่อนหน้าของพารามิเตอร์ความยากแบบอื่นๆ ที่จำนวนข้อสอบแตกต่างกัน จะมีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาจากการแจกแจงแบบ Normal แตกต่างกัน คือ เมื่อมีจำนวนข้อสอบ 30 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากที่ให้ค่าความคลาดเคลื่อนในการ

ประมาณค่าต่ำรองลงมาจากแบบ Normal คือ Uniform, Negative Skew และ Positive Skew ตามลำดับ เมื่อมีจำนวนข้อสอบ 60 และ 90 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากที่ให้ค่าความคลาดเคลื่อนในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ Negative Skew, Positive Skew และ Uniform ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังตารางที่ 4.8

ตารางที่ 4.8 ผลการวิเคราะห์เปรียบเทียบความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน จำแนกตามจำนวนข้อสอบและรูปแบบโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	RPFI	294.558	2.998	326.935	1.224	328.869	1.271	336.806	1.027	105766.26	.000**
	FPRI	294.348	2.987	326.831	1.213	328.666	1.258	336.668	1.021	107260.2	.000**
	RPRI	294.455	2.993	326.85	1.213	328.699	1.262	336.686	1.021	106385.97	.000**
Post Hoc	Dunnett: RPFI : Nor < Uniform < Neg < Pos**										
	Dunnett: FPRI : Nor < Uniform < Neg < Pos**										
	Dunnett: RPRI : Nor < Uniform < Neg < Pos**										
60	RPFI	210.563	2.660	236.630	0.795	230.255	0.981	230.279	0.927	53761.97	.000**
	FPRI	210.274	2.645	236.594	0.790	230.118	0.971	230.136	0.916	55484.184	.000**
	RPRI	210.351	2.651	236.611	0.792	230.106	0.972	230.142	0.917	54935.83	.000**
Post Hoc	Dunnett: RPFI : Nor < Neg < Pos < Uniform **										
	Dunnett: FPRI : Nor < Neg < Pos < Uniform **										
	Dunnett: RPRI : Nor < Neg < Pos < Uniform **										
90	RPFI	172.644	2.281	199.343	0.404	191.451	0.757	192.437	0.706	81278.458	.000**
	FPRI	172.434	2.261	199.303	0.398	191.332	0.747	192.336	0.698	83836.122	.000**
	RPRI	172.442	2.267	199.304	0.400	191.333	0.750	192.344	0.698	83382.261	.000**
Post Hoc	Dunnett: RPFI : Nor < Neg < Pos < Uniform **										
	Dunnett: FPRI : Nor < Neg < Pos < Uniform **										
	Dunnett: RPRI : Nor < Neg < Pos < Uniform **										

**p< .01

โดยสรุป ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Normal จะมีค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุดด้วยโมเดลการประมาณค่าทั้ง 3 โมเดล ในทุกเงื่อนไขจำนวนข้อสอบ 30, 60 และ 90 ข้อ แสดงว่า โมเดลการประมาณ

ค่าพารามิเตอร์ทั้ง 3 โมเดล มีประสิทธิภาพดีที่สุด เมื่อข้อสอบมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Normal แต่เมื่อข้อสอบมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบอื่นๆ จะมีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาจากแบบ Normal แตกต่างกันไป ระหว่างการแจกแจงมีก่อนหน้าของพารามิเตอร์ความยากแบบ Uniform, Positive Skew หรือ Negative Skew ขึ้นอยู่กับจำนวนข้อสอบและจำนวนผู้สอบ

ตอนที่ 2 การวิเคราะห์ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบจากข้อมูลจำลอง: ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าพารามิเตอร์

ตัวบ่งชี้ที่แสดงถึงคุณภาพของการประมาณค่าพารามิเตอร์ นอกเหนือจากความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์แล้ว ยังสามารถพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Mean Squared Error; MSE (θ)) ของการประมาณค่าพารามิเตอร์ ซึ่งเป็นความแตกต่างระหว่างค่าที่ได้จากตัวประมาณค่ากับค่าจริง เมื่อค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองเข้าใกล้ศูนย์มากเท่าใดแสดงถึงตัวประมาณค่านั้นมีประสิทธิภาพมาก

ในการวิเคราะห์เปรียบเทียบคุณภาพของการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลการประมาณค่า 3 รูปแบบ คือ (1) รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2) รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ตามทฤษฎีการตอบสนองข้อสอบ (IRT) โดยการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าพารามิเตอร์ผู้สอบที่ได้จากโมเดลการประมาณค่า 3 รูปแบบดังกล่าว ดังรายละเอียดของผลการวิเคราะห์ ดังนี้

2.1 การวิเคราะห์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ

การประมาณค่าพารามิเตอร์ผู้สอบเปรียบเทียบระหว่างโมเดลประมาณค่าพารามิเตอร์ผู้สอบ 3 รูปแบบ คือ (1) รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2) รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ตามทฤษฎีการตอบสนองข้อสอบ (IRT) โดยใช้ข้อมูลจำลองด้วยโปรแกรม R ตามเงื่อนไขกรณีศึกษา 48 เงื่อนไข มีผลการวิเคราะห์ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ดังนี้

2.1.1 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างโมเดลประมาณค่าพารามิเตอร์ผู้สอบ จำแนกตามจำนวนข้อสอบและการแจกแจงพารามิเตอร์ความยากข้อสอบ

ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบจากโมเดล 3 รูปแบบ คือ (1) รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2) รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ที่มีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Normal, Uniform และ Negative Skew วิเคราะห์เปรียบเทียบเงื่อนไขกรณีศึกษา 12 เงื่อนไข พบว่า

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ โมเดลการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด และมีค่าใกล้เคียงกับค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ด้วยโมเดลการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ในขณะที่การประมาณค่าพารามิเตอร์ด้วยโมเดลการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์สูงกว่าโมเดลอื่นๆ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกจากนี้พบว่า เมื่อมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Positive Skew ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ด้วยโมเดลการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่าต่ำกว่าโมเดลอื่นๆอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 รองลงมาคือ ค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ด้วยโมเดลการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และโมเดลการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ตามลำดับ ดังตารางที่ 4.9

ตารางที่ 4.9 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	Normal	95.832	5.782	86.675	5.456	87.594	5.571	80.938	.000**
	Uniform	116.384	5.305	112.254	4.930	113.025	5.093	18.461	.000**
	Negative	110.396	5.002	105.650	4.860	106.522	4.837	26.590	.000**
	Positive	122.243	5.590	124.274	5.210	125.568	5.414	9.612	.000**
Post Hoc	LSD: Normal: FPRI<RPFI**; RPRI<RPFI**; FPRI<RPRI LSD: Uniform: FPRI<RPFI**, RPRI<RPFI**; FPRI<RPRI LSD: Negative: FPRI<RPFI**, RPRI<RPFI**; FPRI<RPRI LSD: Positive: RPFI<FPRI**, RPFI<RPRI**; FPRI<RPRI								
60	Normal	56.772	3.411	46.557	2.950	46.762	2.883	357.073	.000**
	Uniform	67.442	3.373	62.794	3.089	62.917	3.184	67.777	.000**
	Negative	65.426	3.392	61.447	3.119	61.760	3.169	46.972	.000**
	Positive	73.694	4.271	86.287	4.584	87.339	4.658	283.742	.000**
Post Hoc	LSD: Normal: FPRI<RPFI**; RPRI<RPFI**; FPRI<RPRI LSD: Uniform: FPRI<RPFI**, RPRI<RPFI**; FPRI<RPRI								

ตารางที่ 4.9 (ต่อ)

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
	LSD: Negative: FPRI<RPFI**, RPRI<RPFI**; FPRI<RPRI LSD: Positive: RPFI<FPRI**, RPFI<RPRI**; FPRI<RPRI								
90	Normal	43.548	2.588	32.503	2.310	32.734	2.292	691.231	.000**
	Uniform	53.265	2.846	47.424	2.477	47.543	2.481	163.973	.000**
	Negative	50.079	2.889	45.440	2.569	45.587	2.537	97.562	.000**
	Positive	51.543	2.445	56.061	2.510	56.582	2.529	123.335	.000**
Post Hoc	LSD: Normal: FPRI<RPFI**; RPRI<RPFI**; FPRI<RPRI LSD: Uniform: FPRI<RPFI**, RPRI<RPFI**; FPRI<RPRI LSD: Negative: FPRI<RPFI**, RPRI<RPFI**; FPRI<RPRI LSD: Positive: RPFI<FPRI**, RPFI<RPRI**; FPRI<RPRI								

**p< .01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ โมเดลการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) ส่วนใหญ่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด และมีค่าใกล้เคียงกับค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ด้วยโมเดลการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ในขณะที่ส่วนใหญ่ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ด้วยโมเดลการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่าสูงกว่าโมเดลอื่นๆอย่างไม่มีนัยสำคัญทางสถิติ ยกเว้นในเงื่อนไขจำนวนข้อสอบ 30, 60 ข้อที่มีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Normal และเงื่อนไขจำนวนข้อสอบ 60 ข้อที่มีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Positive Skew มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าด้วยโมเดลการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) สูงกว่าโมเดลอื่นๆอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเงื่อนไขจำนวนข้อสอบ 60 ข้อที่มีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Negative Skew มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของการประมาณค่าด้วยโมเดลการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) สูงกว่าโมเดลอื่นๆอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกจากนี้พบว่า เมื่อมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Uniform ในจำนวนข้อสอบ 60 ข้อ ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ด้วยโมเดลการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่าเท่ากันกับโมเดลการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) ดังตารางที่ 4.10

ตารางที่ 4.10 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	Normal	88.015	4.655	86.624	4.669	87.070	4.647	6.981	.001**
	Uniform	113.490	5.023	112.789	4.952	113.152	4.972	1.485	.227
	Negative	125.180	6.379	124.490	6.415	124.867	6.413	.875	.417
	Positive	128.941	6.330	128.118	6.390	128.527	6.301	1.265	.283
Post Hoc	LSD: Normal: FPRI<RPFI**; RPRI<RPFI*; FPRI<RPRI LSD: Uniform: FPRI < RPRI < RPFI LSD: Negative: FPRI < RPRI < RPFI LSD: Positive: FPRI < RPRI < RPFI								
60	Normal	49.087	3.164	46.947	3.052	47.017	3.116	45.842	.000**
	Uniform	66.294	3.303	66.346	3.310	66.451	3.312	.177	.838
	Negative	65.704	3.208	65.065	3.190	65.147	3.187	3.558	.029**
	Positive	59.415	2.832	58.736	2.850	58.803	2.845	5.195	.006**
Post Hoc	LSD: Normal: FPRI<RPFI**; RPRI<RPFI*; FPRI<RPRI LSD: Uniform: RPFI < FPRI < RPRI LSD: Negative: FPRI < RPFI**; RPRI < RPFI*; FPRI < RPRI LSD: Positive: FPRI < RPFI**; RPRI < RPFI*; FPRI < RPRI								
90	Normal	483.138	256.334	462.159	233.449	462.075	233.299	0.228	0.797
	Uniform	6326.157	2811.764	6205.113	2683.872	6205.339	2684.807	0.059	0.943
	Negative	7696.569	7882.182	6724.298	4678.440	6723.972	4676.874	0.804	0.449
	Positive	7247.412	7195.536	6409.414	4457.613	6409.449	4459.566	0.69	0.502
Post Hoc	Normal: RPRI < FPRI < RPFI Uniform: FPRI < RPRI < RPFI Negative: RPRI < FPRI < RPFI Positive: FPRI < RPRI < RPFI								

**p< .01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ โมเดลการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) ส่วนใหญ่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด และมีค่าใกล้เคียงกับค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ด้วยโมเดลการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) อย่างไม่มี

นัยสำคัญทางสถิติ ยกเว้นในเงื่อนไขจำนวนข้อสอบ 90 ข้อ ที่มีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal, Uniform และ Negative Skew มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบด้วยโมเดลการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) ต่ำที่สุด อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังตารางที่ 4.11

ตารางที่ 4.11 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	Normal	85.401	5.556	84.655	5.556	84.788	5.590	2.558	0.078
	Uniform	99.422	5.128	99.044	5.104	99.225	5.106	0.684	0.505
	Negative	109.384	6.009	109.085	6.001	109.323	6.023	0.346	0.708
	Positive	101.940	5.417	101.519	5.419	101.656	5.441	0.786	0.456
Post Hoc	Normal: FPRI < RPRI < RPFI Uniform: FPRI < RPRI < RPFI Negative: FPRI < RPRI < RPFI Positive: FPRI < RPRI < RPFI								
60	Normal	472.289	363.468	457.229	334.207	457.213	334.334	0.038	0.962
	Uniform	5538.770	2644.306	5478.464	2555.988	5480.343	2558.311	0.011	0.99
	Negative	5987.255	5803.575	5554.726	3323.994	5554.972	3324.530	0.201	0.818
	Positive	8151.511	10154.435	6927.219	5930.057	6927.640	5931.636	0.518	0.596
Post Hoc	Normal: RPRI < FPRI < RPFI Uniform: RPRI < FPRI < RPFI Negative: FPRI < RPRI < RPFI Positive: FPRI < RPRI < RPFI								
90	Normal	34.326	2.470	32.640	2.450	32.692	2.460	75.98	.000**
	Uniform	39.787	2.077	39.415	2.031	39.433	2.058	5.205	.006**
	Negative	38.391	1.896	38.015	1.874	38.049	1.898	6.064	.002**
	Positive	37.779	1.855	37.650	1.858	37.712	1.851	0.608	0.545
Post Hoc	LSD: Normal: FPRI < RPFI**; RPRI <RPFI**; FPRI < RPRI LSD: Uniform: FPRI < RPFI**; RPRI <RPFI**; FPRI < RPRI LSD: Negative: FPRI < RPFI**; RPRI <RPFI**; FPRI < RPRI Positive: FPRI < RPRI <RPFI								

**p< .01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ ส่วนใหญ่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบใกล้เคียงกัน ทั้ง 3 โมเดล ยกเว้นในเงื่อนไขจำนวนข้อสอบ 60 และ 90 ข้อ ที่มีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากแบบ Normal ที่พบว่าโมเดลการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และโมเดลการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบเท่ากัน และมีค่าต่ำกว่าค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบด้วยโมเดลการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังตารางที่ 4.12

ตารางที่ 4.12 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน จำแนกตามจำนวนข้อสอบและลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	Normal	85.307	5.376	85.396	5.342	85.460	5.368	0.205	0.815
	Uniform	99.160	5.243	99.094	5.231	99.207	5.262	0.116	0.891
	Negative	100.097	5.192	100.016	5.230	100.152	5.218	0.171	0.843
	Positive	103.932	5.490	103.806	5.455	103.896	5.475	0.139	0.87
Post Hoc	Normal: RPFI < FPRI < RPRI Uniform: FPRI < RPFI < RPRI Negative: FPRI < RPFI < RPRI Positive: FPRI < RPRI < RPFI								
60	Normal	46.995	3.418	46.488	3.416	46.511	3.397	7.06	.001**
	Uniform	54.875	2.845	54.910	2.852	54.974	2.845	0.313	0.731
	Negative	52.361	2.908	52.315	2.907	52.381	2.897	0.133	0.876
	Positive	52.200	2.706	52.046	2.710	52.079	2.707	0.905	0.405
Post Hoc	LSD: Normal: FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI Uniform: RPFI < FPRI < RPRI Negative: FPRI < RPFI < RPRI Positive: FPRI < RPRI < RPFI								
90	Normal	32.946	2.522	32.430	2.515	32.431	2.511	13.963	.000**
	Uniform	39.573	2.046	39.502	2.045	39.513	2.041	0.35	0.705
	Negative	37.014	1.920	37.012	1.915	37.049	1.910	0.117	0.889

ตารางที่ 4.12 (ต่อ)

จำนวนข้อสอบ	การแจกแจงข้อสอบ	RPFI Model		FPRI Model		RPRI Model		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
	Positive	37.180	1.883	37.073	1.867	37.106	1.867	0.858	0.424
Post Hoc	LSD: Normal: FPRI < RPFI**; RPRI < RPFI**; FPRI < RPRI Uniform: FPRI < RPRI < RPFI Negative: FPRI < RPFI < RPRI Positive: FPRI < RPRI < RPFI								

**p< .01

สรุปการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างโมเดลการประมาณค่าพารามิเตอร์ จำแนกตามจำนวนข้อสอบ การแจกแจงพารามิเตอร์ความยากของข้อสอบ และขนาดกลุ่มตัวอย่างผู้สอบ พบว่า เมื่อมีขนาดกลุ่มตัวอย่างผู้สอบ 100 และ 300 คน ส่วนใหญ่โมเดลการประมาณค่าพารามิเตอร์ผู้สอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำที่สุดคือ โมเดลการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) รองลงมาคือ โมเดลการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) และโมเดลการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ตามลำดับ และเมื่อมีขนาดกลุ่มตัวอย่างผู้สอบ 500 และ 1,000 คน ส่วนใหญ่ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบที่ได้จากทั้ง 3 โมเดลมีค่าใกล้เคียงกัน

2.1.2 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงพารามิเตอร์ความยากข้อสอบ จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบระหว่างลักษณะการแจกแจงก่อนหน้าของความยากข้อสอบ 4 ลักษณะ คือ Normal, Uniform, Positive Skew และ Negative Skew วิเคราะห์เปรียบเทียบเงื่อนไขกรณีศึกษา 12 เงื่อนไข พบว่า

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุดในทั้ง 3 โมเดลที่ใช้ในการประมาณค่า อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำ

รองลงมาคือ แบบ Negative Skew, Uniform และ Positive Skew ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ดังตารางที่ 4.13

ตารางที่ 4.13 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 100 คน จำแนกตามจำนวนข้อสอบและโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	RPFI	95.832	5.782	116.384	5.305	110.396	5.002	122.243	5.590	436.396	.000**
	FPRI	86.675	5.456	112.254	4.930	105.650	4.860	124.274	5.210	942.127	.000**
	RPRI	87.594	5.571	113.025	5.093	106.522	4.837	125.568	5.414	914.567	.000**
Post Hoc	LSD: RPFI : Nor < Neg< Uniform < Pos** LSD: FPRI : Nor < Neg < Uniform < Pos** LSD: RPRI : Nor < Neg < Uniform < Pos**										
60	RPFI	56.772	3.411	67.442	3.373	65.426	3.392	73.694	4.271	370.655	.000**
	FPRI	46.557	2.950	62.794	3.089	61.447	3.119	86.287	4.584	2201.115	.000**
	RPRI	46.762	2.883	62.917	3.184	61.760	3.169	87.339	4.658	2247.887	.000**
Post Hoc	Dunnett: RPFI : Nor<Neg<Uniform<Pos** Dunnett: FPRI : Nor<Neg<Uniform<Pos** Dunnett: RPRI : Nor<Neg**; Nor<Uniform**; Nor<Pos**; Uniform<Pos**; Neg<Uniform										
90	RPFI	43.548	2.588	53.265	2.846	50.079	2.889	51.543	2.445	247.53	.000**
	FPRI	32.503	2.310	47.424	2.477	45.440	2.569	56.061	2.510	1554.093	.000**
	RPRI	32.734	2.292	47.543	2.481	45.587	2.537	56.582	2.529	1594.825	.000**
Post Hoc	LSD: RPFI : Nor<Neg<Pos<Uniform** LSD: FPRI : Nor<Neg<Uniform<Pos** LSD: RPRI : Nor<Neg<Uniform<Pos**										

**p< .01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุดในทั้ง 3 โมเดลที่ใช้ในการประมาณค่า อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในเงื่อนไขจำนวนข้อสอบ 30 ข้อ ลักษณะการ

แจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ แบบ Uniform, Negative Skew และ Positive Skew ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในเงื่อนไขจำนวนข้อสอบ 60 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ แบบ Positive Skew, Negative Skew และ Uniform ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และในเงื่อนไขจำนวนข้อสอบ 90 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ แบบ Uniform, Positive Skew และ Negative Skew ตามลำดับ อย่างไม่มีนัยสำคัญทางสถิติ ดังตารางที่ 4.14

ตารางที่ 4.14 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 300 คน จำแนกตามจำนวนข้อสอบและรูปแบบโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	RPFI	88.015	4.655	113.490	5.023	125.180	6.379	128.941	6.330	3207.605	.000**
	FPRI	86.624	4.669	112.789	4.952	124.490	6.415	128.118	6.390	3293.293	.000**
	RPRI	87.070	4.647	113.152	4.972	124.867	6.413	128.527	6.301	3314.985	.000**
Post Hoc	LSD: RPFI : Nor< Uniform < Neg <Pos** LSD: FPRI : Nor< Uniform < Neg <Pos** LSD: RPRI : Nor< Uniform < Neg <Pos**										
60	RPFI	49.087	3.164	66.294	3.303	65.704	3.208	59.415	2.832	1952.633	.000**
	FPRI	46.947	3.052	66.346	3.310	65.065	3.190	58.736	2.850	2445.278	.000**
	RPRI	47.017	3.116	66.451	3.312	65.147	3.187	58.803	2.845	2429.511	.000**
Post Hoc	LSD: RPFI : Nor< Pos < Neg < Uniform ** LSD: FPRI : Nor< Pos < Neg < Uniform ** LSD: RPRI : Nor< Pos < Neg < Uniform **										
90	RPFI	483.138	256.334	6326.157	2811.764	7696.569	7882.182	7247.412	7195.536	33.196	.000**
	FPRI	462.159	233.449	6205.113	2683.872	6724.298	4678.440	6409.414	4457.613	66.087	.000**
	RPRI	462.075	233.299	6205.339	2684.807	6723.972	4676.874	6409.449	4459.566	66.077	.000**

ตารางที่ 4.14 (ต่อ)

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
Post Hoc	Dunnett: RPFI : Nor<Uniform**; Nor<Pos**; Nor< Neg**; Uniform<Pos<Neg Dunnett: FPRI : Nor<Uniform**; Nor<Pos**; Nor< Neg**; Uniform<Pos<Neg Dunnett: RPRI : Nor<Uniform**; Nor<Pos**; Nor< Neg**; Uniform<Pos<Neg										

**p< .01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุดในทั้ง 3 โมเดลที่ใช้ในการประมาณค่า อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในเงื่อนไขจำนวนข้อสอบ 30 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ แบบ Uniform, Positive Skew และ Negative Skew ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในเงื่อนไขจำนวนข้อสอบ 60 ข้อ ลักษณะการแจกแจงพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ แบบ Uniform, Negative Skew และ Positive Skew ตามลำดับ อย่างไม่มีนัยสำคัญทางสถิติ และในเงื่อนไขจำนวนข้อสอบ 90 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ แบบ Positive Skew, Negative Skew และ Uniform ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ยกเว้นการประมาณค่าด้วยโมเดล RPRI ที่ลักษณะการแจกแจงพารามิเตอร์ความยากข้อสอบมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำรองลงมาจากแบบ Normal คือ แบบ Positive Skew, Negative Skew และ Uniform ตามลำดับ อย่างไม่มีนัยสำคัญทางสถิติ ดังตารางที่ 4.15

ตารางที่ 4.15 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 500 คน จำแนกตามจำนวนข้อสอบและรูปแบบโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	RPFI	85.401	5.556	99.422	5.128	109.384	6.009	101.940	5.417	1639.637	.004**
	FPRI	84.655	5.556	99.044	5.104	109.085	6.001	101.519	5.419	1706.829	.000**
	RPRI	84.788	5.590	99.225	5.106	109.323	6.023	101.656	5.441	1706.801	.000**
Post Hoc	Dunnett: RPFI : Nor<Uniform<Pos <Neg ** Dunnett: FPRI : Nor<Uniform<Pos <Neg ** Dunnett: RPRI : Nor<Uniform<Pos <Neg **										
60	RPFI	472.289	363.468	5538.770	2644.306	5987.255	5803.575	8151.511	10154.435	17.616	.000**
	FPRI	457.229	334.207	5478.464	2555.988	5554.726	3323.994	6927.219	5930.057	36.719	.000**
	RPRI	457.213	334.334	5480.343	2558.311	5554.972	3324.530	6927.640	5931.636	36.704	.000**
Post Hoc	Dunnett: RPFI : Nor<Uniform**; Nor<Neg**; Nor<Pos**; Uniform<Neg< Pos Dunnett: FPRI : Nor<Uniform**; Nor<Neg**; Nor<Pos**; Uniform<Neg< Pos Dunnett: RPRI : Nor<Uniform**; Nor<Neg**; Nor<Pos**; Uniform<Neg< Pos										
90	RPFI	34.326	2.470	39.787	2.077	38.391	1.896	37.779	1.855	617.136	.000**
	FPRI	32.640	2.450	39.415	2.031	38.015	1.874	37.650	1.858	1024.557	.000**
	RPRI	32.692	2.460	39.433	2.058	38.049	1.898	37.712	1.851	1004.395	.000**
Post Hoc	Dunnett: RPFI : Nor<Pos <Neg <Uniform** Dunnett: FPRI : Nor<Pos <Neg<Uniform ** Dunnett: RPRI : Nor<Pos**; Nor<Neg**; Nor<Uniform**; Neg<Uniform**; Pos<Neg*										

**p< .01

เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน ในจำนวนข้อสอบทั้ง 30, 60 และ 90 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุดในทั้ง 3 โมเดลที่ใช้ในการประมาณค่า อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในขณะที่เมื่อมีจำนวนข้อสอบ 30 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาจากแบบ Normal คือ Uniform, Negative

skew และ Positive skew ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อมีข้อสอบ 60 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาจากแบบ Normal คือ Positive skew, Uniform ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาจากแบบ Normal คือ Negative skew, Positive skew ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเมื่อมีข้อสอบ 90 ข้อ ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบที่มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาจากแบบ Normal คือ Negative skew, Uniform ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ Positive skew, Uniform ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังตารางที่ 4.16

ตารางที่ 4.16 ผลการวิเคราะห์เปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ระหว่างลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยาก เมื่อกำหนดขนาดกลุ่มตัวอย่างผู้สอบ 1,000 คน จำแนกตามจำนวนข้อสอบและรูปแบบโมเดลในการประมาณค่าพารามิเตอร์ผู้สอบ

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
30	RPFI	85.307	5.376	99.160	5.243	100.097	5.192	103.932	5.490	2337.463	.000**
	FPRI	85.396	5.342	99.094	5.231	100.016	5.230	103.806	5.455	2294.082	.000**
	RPRI	85.460	5.368	99.207	5.262	100.152	5.218	103.896	5.475	2291.06	.000**
Post Hoc	LSD : RPFI : Nor< Uniform<Neg< Pos**										
	LSD : FPRI: Nor< Uniform<Neg< Pos**										
	LSD : RPRI: Nor< Uniform<Neg< Pos**										
60	RPFI	46.995	3.418	54.875	2.845	52.361	2.908	52.200	2.706	1232.798	.000**
	FPRI	46.488	3.416	54.910	2.852	52.315	2.907	52.046	2.710	1412.066	.000**
	RPRI	46.511	3.397	54.974	2.845	52.381	2.897	52.079	2.707	1435.965	.000**
Post Hoc	Dunnett: RPFI:Nor<Pos<Uniform**; Nor<Neg <Uniform**; Pos<Neg										
	Dunnett: FPRI: Nor<Pos<Uniform**; Nor<Neg <Uniform**; Pos<Neg										
	Dunnett: RPRI: Nor<Pos<Uniform**; Nor<Neg <Uniform**; Pos<Neg										
90	RPFI	32.946	2.522	39.573	2.046	37.014	1.920	37.180	1.883	1700.556	.000**
	FPRI	32.430	2.515	39.502	2.045	37.012	1.915	37.073	1.867	1975.928	.000**
	RPRI	32.431	2.511	39.513	2.041	37.049	1.910	37.106	1.867	1991.324	.000**

ตารางที่ 4.16 (ต่อ)

จำนวนข้อสอบ	โมเดลประมาณค่า	Normal Distribution		Uniform Distribution		Negative Distribution		Positive Distribution		F-test	p-value
		Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)	Mean ($x10^{-3}$)	SD ($x10^{-3}$)		
Post Hoc	Dunnett: RPFI: Nor<Neg<Uniform**; Nor<Pos <Uniform**; Neg < Pos Dunnett: FPRI: Nor<Neg<Uniform**;Nor<Pos <Uniform**; Neg < Pos Dunnett: RPRI: Nor<Neg<Uniform**;Nor<Pos <Uniform**; Neg < Pos										

**p< .01

โดยสรุป ลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ต่ำที่สุดในทุกเงื่อนไขจำนวนข้อสอบที่ 30, 60 และ 90 ข้อ ในทุกโมเดลการประมาณค่าทั้ง 3 โมเดล แสดงว่า โมเดลการประมาณค่าพารามิเตอร์ทั้ง 3 โมเดล มีประสิทธิภาพดีที่สุด เมื่อข้อมูลมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal ในขณะที่เมื่อข้อมูลมีลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบอื่นๆ จะมีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำรองลงมาจากลักษณะการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal แตกต่างกันไป ระหว่างการแจกแจงก่อนหน้าของความยากข้อสอบแบบ Uniform, Positive Skew หรือ Negative Skew ขึ้นอยู่กับจำนวนข้อสอบและจำนวนผู้สอบ

สรุปได้ว่า โมเดลการประมาณค่ารูปแบบสุ่มข้อสอบ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ผู้สอบสูงกว่าโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) และโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ผู้สอบสูงกว่าโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) แสดงสรุปประสิทธิภาพของโมเดลในการประมาณค่าพารามิเตอร์จำแนกตามรูปแบบการประมาณค่าพารามิเตอร์ในตารางที่ 4.17

ตารางที่ 4.17 สรุปประสิทธิภาพของโมเดลในการประมาณค่าพารามิเตอร์ จำแนกตามรูปแบบการประมาณค่าพารามิเตอร์ และเงื่อนไขการศึกษา 48 เงื่อนไข

เงื่อนไขการศึกษา	โมเดลที่มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ดีที่สุด					
	SE(θ)			MSE		
	RPFI	FPRI	RPRI	RPFI	FPRI	RPRI
Normal-100-30		*			*	
100-60		*			*	
100-90		*			*	
300-30		*			*	
300-60		*			*	
300-90		*				*
500-30		*			*	
500-60		*				*
500-90		*			*	
1000-30		*		*		
1000-60		*			*	
1000-90		*			*	
Uniform-100-30		*			*	
100-60		*			*	
100-90		*			*	
300-30		*			*	
300-60	*			*		
300-90	*				*	
500-30		*			*	
500-60		*			*	
500-90		*			*	
1000-30		*			*	
1000-60		*		*		
1000-90		*			*	

ตารางที่ 4.17 (ต่อ)

เงื่อนไขการศึกษา	โมเดลที่มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ดีที่สุด					
	SE(θ)			MSE		
	RPFI	FPRI	RPRI	RPFI	FPRI	RPRI
Negative-100-30		*			*	
100-60		*			*	
100-90		*			*	
300-30		*			*	
300-60		*			*	
300-90			*			*
500-30		*			*	
500-60		*			*	
500-90		*			*	
1000-30		*			*	
1000-60			*		*	
1000-90		*			*	
Positive-100-30	*			*		
100-60	*			*		
100-90	*			*		
300-30		*			*	
300-60		*			*	
300-90		*			*	
500-30		*			*	
500-60		*			*	
500-90		*			*	
1000-30		*			*	
1000-60		*			*	
1000-90		*			*	

หมายเหตุ: * หมายถึง โมเดลที่มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ดีที่สุด โดยมีค่า SE(θ) หรือ MSE ต่ำที่สุด

ตอนที่ 3 การวิเคราะห์ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบจากข้อมูลเชิงประจักษ์

การศึกษาประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบพิจารณาจากค่าสถิติ Akaike Information Criterion (AIC) เปรียบเทียบระหว่างโมเดลประมาณค่าพารามิเตอร์ 3 โมเดล คือ (1) รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2) รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) โดยใช้ข้อมูลผลการสอบ O-NET ปี พ.ศ. 2553 ระดับชั้นมัธยมศึกษาปีที่ 6 จำนวน 3 รายวิชา คือ ภาษาอังกฤษ คณิตศาสตร์ และภาษาไทย ในแต่ละขนาดกลุ่มตัวอย่าง 100, 300, 500 และ 1,000 คน โดยใช้การสุ่มแบบใส่คืนและทำซ้ำ (bootstrap sampling) 100 รอบ

ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบวิชาภาษาอังกฤษ พบว่า โมเดลประมาณค่าพารามิเตอร์เจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่า AIC ต่ำกว่า โมเดลประมาณค่าพารามิเตอร์สุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างไม่มีนัยสำคัญทางสถิติ ในกลุ่มตัวอย่างทุกขนาด และโมเดลประมาณค่าพารามิเตอร์สุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่า AIC ต่ำกว่า โมเดลประมาณค่าพารามิเตอร์สุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) อย่างไม่มีนัยสำคัญทางสถิติ ในกลุ่มตัวอย่างทุกขนาด ดังตารางที่ 4.18

ตารางที่ 4.18 ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบวิชาภาษาอังกฤษ (70 ข้อ)

sample size	AIC						F-test	p
	RPFI		FPRI		RPRI			
	Mean	S.D.	Mean	S.D.	Mean	S.D.		
100	7,643.82	83.30	7,643.81	83.58	7,645.89	83.12	0.021	0.980
300	22,761.72	138.40	22,760.13	138.89	22,764.45	138.11	0.025	0.975
500	37,920.58	177.72	37,916.69	178.47	37,923.68	177.59	0.039	0.962
1,000	75,818.36	274.40	75,808.32	274.86	75,819.41	274.43	0.050	0.952

ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบวิชาคณิตศาสตร์ พบว่า โมเดลประมาณค่าพารามิเตอร์เจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่า AIC ต่ำกว่า โมเดลประมาณค่าพารามิเตอร์สุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างไม่มีนัยสำคัญทางสถิติ ในกลุ่มตัวอย่างทุกขนาด และโมเดลประมาณค่าพารามิเตอร์สุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มี

ค่า AIC ต่ำกว่า โมเดลประมาณค่าพารามิเตอร์สุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) อย่างไม่มีนัยสำคัญทางสถิติ ในกลุ่มตัวอย่างทุกขนาด แสดงในตารางที่ 4.19

ตารางที่ 4.19 ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบวิชาคณิตศาสตร์ (40 ข้อ)

sample size	AIC						F-test	p
	RPFI		FPRI		RPRI			
	Mean	S.D.	Mean	S.D.	Mean	S.D.		
100	4,632.61	66.79	4,629.06	67.20	4,633.88	66.73	0.139	0.870
300	13,864.87	109.93	13,853.39	110.89	13,867.04	109.65	0.444	0.642
500	23,090.06	161.91	23,071.40	162.46	23,092.86	161.71	0.518	0.596
1,000	46,126.43	220.01	46,086.57	221.78	46,127.45	220.51	1.115	0.329

ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบวิชาภาษาไทย พบว่า โมเดลประมาณค่าพารามิเตอร์เจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่า AIC ต่ำกว่า โมเดลประมาณค่าพารามิเตอร์สุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างไม่มีนัยสำคัญทางสถิติในกลุ่มตัวอย่างทุกขนาด และโมเดลประมาณค่าพารามิเตอร์สุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีค่า AIC ต่ำกว่า โมเดลประมาณค่าพารามิเตอร์สุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) อย่างไม่มีนัยสำคัญทางสถิติ ในกลุ่มตัวอย่างทุกขนาด ดังตารางที่ 4.20

ตารางที่ 4.20 ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบวิชาภาษาไทย (90 ข้อ)

sample size	AIC						F-test	p
	RPFI		FPRI		RPRI			
	Mean	S.D.	Mean	S.D.	Mean	S.D.		
100	10,877.42	84.25	10,876.91	84.16	10,880.97	84.19	0.069	0.934
300	32,459.31	149.80	32,455.11	149.71	32,463.36	149.83	0.076	0.927
500	54,104.71	177.73	54,096.37	177.81	54,107.81	177.51	0.111	0.895
1,000	107,977.76	242.75	107,965.19	242.85	107,983.04	242.92	0.142	0.867

สรุปได้ว่า ประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบพิจารณาจากค่าสถิติ Akaike Information Criterion (AIC) เปรียบเทียบระหว่างโมเดลประมาณค่าพารามิเตอร์ 3 โมเดล คือ (1)

รูปแบบการสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2) รูปแบบการเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบการสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) โดยใช้ข้อมูลผลการสอบ O-NET ปี พ.ศ. 2553 ระดับชั้นมัธยมศึกษาปีที่ 6 จำนวน 3 รายวิชา คือ ภาษาอังกฤษ คณิตศาสตร์ และภาษาไทย ในแต่ละขนาดกลุ่มตัวอย่าง 100, 300, 500 และ 1,000 คน โดยใช้การสุ่มแบบใส่คืนและทำซ้ำ (bootstrap sampling) 100 รอบ พบว่า โมเดลการประมาณค่าพารามิเตอร์ผู้สอบทั้ง 3 โมเดล มีความสอดคล้องระหว่างโมเดลกับข้อมูลไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

แต่เมื่อพิจารณาจากเกณฑ์การคัดเลือกโมเดลที่มีความสอดคล้องกับข้อมูลที่สุดจะมีค่า AIC ต่ำที่สุด พบว่า โมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ผู้สอบสูงกว่าโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) และโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ผู้สอบสูงกว่าโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI)

บทที่ 5

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มุ่งศึกษาประสิทธิภาพของการประมาณค่าของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ระหว่างการประมาณค่าโดยใช้รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) โดยศึกษาตามขนาดกลุ่มตัวอย่างและจำนวนข้อสอบต่างๆกันด้วยข้อมูลจำลอง นอกจากนี้ยังศึกษาความสอดคล้องของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ที่มีการประมาณค่าโดยโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ ระหว่างการประมาณค่าโดยใช้รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบกับข้อมูลเชิงประจักษ์ ศึกษาโดยสุ่มขนาดกลุ่มตัวอย่างต่างๆ กันจากข้อมูลผลการสอบ O-NET ปีการศึกษา 2553 วิชาภาษาอังกฤษ คณิตศาสตร์ และภาษาไทย ในระดับมัธยมศึกษาปีที่ 6 การจำลองข้อมูลในการศึกษาครั้งนี้ได้จากการจำลองข้อมูลด้วยโปรแกรม R และส่งผ่านเข้าไปประมวลผลในโปรแกรม WinBUGS จากนั้นจึงนำเสนอผลกลับมายังโปรแกรม R การทำงานร่วมกันระหว่างโปรแกรมทั้งสองดำเนินการผ่านคำสั่งใน Package R2 WinBUGS

ประสิทธิภาพของรูปแบบการประมาณค่าจากข้อมูลจำลองพิจารณาจากความคลาดเคลื่อนในการประมาณค่า และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่า และประสิทธิภาพของรูปแบบการประมาณค่าจากข้อมูลเชิงประจักษ์พิจารณาจากค่าสถิติ Akaike Information Criterion (AIC) ในบทนี้ผู้วิจัยจึงนำเสนอสรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ ดังต่อไปนี้

สรุปผลการวิจัย

1. ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพของโมเดลการประมาณค่าพารามิเตอร์ผู้สอบโดยใช้โมเดลการประมาณค่าพารามิเตอร์ผู้สอบ 3 รูปแบบ ที่มีขนาดกลุ่มตัวอย่างผู้สอบน้อย 100 คน ขนาดกลุ่มตัวอย่างผู้สอบปานกลาง 300 และ 500 คน และขนาดกลุ่มตัวอย่างผู้สอบมาก 1,000 คน จำนวนข้อสอบน้อย 30 ข้อ จำนวนข้อสอบปานกลาง 60 ข้อ และจำนวนข้อสอบมาก 90 ข้อ และมีการแจกแจงเริ่มต้นของพารามิเตอร์ความยากของข้อสอบแบบ Normal, Uniform, Negative skew และ Positive skew พบว่า

1.1 เมื่อมีการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal, Uniform และ Negative skew และจำนวนข้อ 30, 60 และ 90 ข้อ ส่วนใหญ่โมเดลการประมาณค่าพารามิเตอร์ผู้สอบรูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำกว่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างมีนัยสำคัญทางสถิติ และต่ำกว่ารูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) อย่างไม่มีนัยสำคัญ ส่วนใหญ่โมเดลการประมาณค่าพารามิเตอร์ผู้สอบรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีค่าเฉลี่ยความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำกว่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างมีนัยสำคัญทางสถิติ

1.2 เมื่อมีการแจกแจงก่อนหน้าของพารามิเตอร์ความยากข้อสอบแบบ Normal, Uniform และ Negative skew และจำนวนข้อ 30, 60 และ 90 ข้อ ส่วนใหญ่โมเดลการประมาณค่าพารามิเตอร์ผู้สอบรูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำกว่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างมีนัยสำคัญทางสถิติ และต่ำกว่ารูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) อย่างไม่มีนัยสำคัญ ส่วนใหญ่โมเดลการประมาณค่าพารามิเตอร์ผู้สอบรูปแบบรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบต่ำกว่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างมีนัยสำคัญทางสถิติ

สรุปได้ว่า โมเดลการประมาณค่าพารามิเตอร์ผู้สอบทั้ง 3 โมเดล มีค่าความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบต่ำที่สุด เมื่อมีการแจกแจงเริ่มต้นของพารามิเตอร์ความยากของข้อสอบแบบ Normal ขนาดกลุ่มตัวอย่างผู้สอบมาก 1,000 คน จำนวนข้อสอบมาก 90 ข้อ สอดคล้องกับค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบ ในขณะที่ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ผู้สอบมีค่าสูงสุด เมื่อมีการแจกแจงเริ่มต้นของพารามิเตอร์ความยากของข้อสอบแบบ Positive skew และมีขนาดกลุ่มตัวอย่างผู้สอบปานกลาง 300 คน จำนวนข้อสอบน้อย 30 ข้อ ทั้ง 3 โมเดลการประมาณค่าพารามิเตอร์ผู้สอบ และค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าพารามิเตอร์ผู้สอบมีค่าสูงสุด เมื่อมีการแจกแจงเริ่มต้นของพารามิเตอร์ความยากของข้อสอบแบบ Positive

skew และมีขนาดกลุ่มตัวอย่างผู้สอบปานกลาง 500 คน จำนวนข้อสอบปานกลาง 60 ข้อ ทั้ง 3 โมเดลการประมาณค่าพารามิเตอร์ผู้สอบ

Model	SE(θ)		MSE	
	ค่าต่ำสุด (เงื่อนไข)	ค่าสูงสุด (เงื่อนไข)	ค่าต่ำสุด (เงื่อนไข)	ค่าสูงสุด (เงื่อนไข)
FPRI	.172434 (I=90; P=1,000; Nor)	.378553 (I=30; P=300; Pos)	.032430 (I=90; P=1,000; Nor)	6.927219 (I=60; P=500; Pos)
RPRI	.172442 (I=90; P=1,000; Nor)	.378602 (I=30; P=300; Pos)	.032431 (I=90; P=1,000; Nor)	6.927640 (I=60; P=500; Pos)
RPFI	.172644 (I=90; P=1,000; Nor)	.380068 (I=30; P=300; Pos)	.032946 (I=90; P=1,000; Nor)	8.151511 (I=60; P=500; Pos)

หมายเหตุ: I=ความยาวข้อสอบ; P=ขนาดกลุ่มตัวอย่าง

2. ผลการวิเคราะห์เปรียบเทียบประสิทธิภาพการประมาณค่าพารามิเตอร์ผู้สอบพิจารณาจากค่าสถิติ Akaike Information Criterion (AIC) เปรียบเทียบระหว่างโมเดลประมาณค่าพารามิเตอร์ 3 โมเดล คือ (1) รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) (2) รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และ (3) รูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) โดยใช้ข้อมูลผลการสอบ O-NET ปี พ.ศ. 2553 ระดับชั้นมัธยมศึกษาปีที่ 6 จำนวน 3 รายวิชา คือ ภาษาอังกฤษ คณิตศาสตร์ และภาษาไทย เมื่อมีจำนวนผู้สอบ 100, 300, 500 และ 1,000 คน พบว่า โมเดลการประมาณค่าพารามิเตอร์ผู้สอบทั้ง 3 โมเดล มีความสอดคล้องระหว่างโมเดลกับข้อมูลไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่เมื่อเปรียบเทียบตามเกณฑ์ความสอดคล้องของโมเดลการประมาณค่ากับข้อมูล พบว่า ในรายวิชาภาษาอังกฤษ คณิตศาสตร์ และภาษาไทย โมเดลประมาณค่าพารามิเตอร์รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีค่า AIC ต่ำที่สุดในกลุ่มตัวอย่างทุกขนาด รองลงมาคือ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ตามลำดับ

อภิปรายผลการวิจัย

การประมาณค่าความสามารถของผู้สอบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory: IRT) ในการวัดผลการศึกษา มักใช้โมเดลการประมาณค่าพารามิเตอร์ผู้สอบในรูปแบบการเจาะจงข้อสอบ ซึ่งทำให้ได้ค่าความสามารถที่จะสรุปอ้างอิงไปยังประชากรผู้สอบ

แต่ไม่ได้เป็นค่าความสามารถของผู้สอบ De Boeck (2008) จึงได้ศึกษาการประมาณค่าพารามิเตอร์ผู้สอบในรูปแบบการสุ่มข้อสอบ และเนื่องจากการศึกษาเริ่มแรกนั้น ประสิทธิภาพที่ได้จากการประมาณค่ายังไม่คงที่ อีกทั้งยังไม่ได้ศึกษาอิทธิพลของขนาดกลุ่มผู้สอบ จำนวนข้อสอบ และการแจกแจงก่อนหน้าของค่าพารามิเตอร์ความยากข้อสอบ จึงเป็นที่มาของงานวิจัยในครั้งนี้ ที่มุ่งศึกษาประสิทธิภาพของการประมาณค่าของโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ 3 โมเดล ระหว่างการประมาณค่าโดยใช้รูปแบบสุ่มผู้สอบ- เจาะจงข้อสอบ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ โดยศึกษาอิทธิพลของขนาดกลุ่มตัวอย่างผู้สอบ จำนวนข้อสอบ และการแจกแจงก่อนหน้าของค่าพารามิเตอร์ความยากของข้อสอบต่างๆ กันด้วยข้อมูลจำลอง รวมทั้งยังศึกษาความสอดคล้องของโมเดลในการประมาณค่าพารามิเตอร์ทั้ง 3 โมเดล จากข้อมูลผลการสอบ O-NET ปีการศึกษา 2553 ระดับมัธยมศึกษาปีที่ 6 ใน 3 รายวิชา คือ ภาษาอังกฤษ คณิตศาสตร์ และภาษาไทย ในการอภิปรายผลการวิจัย จะนำเสนอตามวัตถุประสงค์และผลการวิจัย มีรายละเอียดดังนี้

1. ผลการวิจัยด้วยข้อมูลจำลองส่วนใหญ่เป็นไปตามสมมติฐาน ข้อที่ 1 คือ โมเดลประมาณค่าพารามิเตอร์ผู้สอบ 1 พารามิเตอร์รูปแบบการสุ่มข้อสอบ ทั้ง 2 รูปแบบ ได้แก่ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีความคลาดเคลื่อนในการประมาณค่าน้อยกว่า โมเดลประมาณค่าพารามิเตอร์ผู้สอบ 1 พารามิเตอร์รูปแบบการเจาะจงข้อสอบ คือ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ณ ขนาดกลุ่มตัวอย่างและความยาวแบบสอบเดียวกัน และผลการวิจัยด้วยข้อมูลจำลองส่วนใหญ่เป็นไปตามสมมติฐาน ข้อที่ 2 คือ โมเดลประมาณค่าพารามิเตอร์ผู้สอบ 1 พารามิเตอร์รูปแบบการสุ่มข้อสอบ ทั้ง 2 รูปแบบ ได้แก่ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าน้อยกว่า โมเดลประมาณค่าพารามิเตอร์ผู้สอบ 1 พารามิเตอร์รูปแบบการเจาะจงข้อสอบ คือ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ณ ขนาดกลุ่มตัวอย่างและความยาวแบบสอบเดียวกัน โดยที่โมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีความคลาดเคลื่อนในการประมาณค่าน้อยกว่าโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างมีนัยสำคัญทางสถิติ และโมเดลการประมาณค่ารูปแบบสุ่ม

ผู้สอบ-สุ่มข้อสอบ (RPRI) มีความคลาดเคลื่อนในการประมาณค่าน้อยกว่าโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) อย่างมีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับงานวิจัยของ De Boeck (2008) จะเห็นได้ว่าโมเดลการประมาณค่ารูปแบบการสุ่มข้อสอบทั้ง 2 โมเดล (FPRI และ RPRI) มีประสิทธิภาพในการประมาณค่ามากกว่าโมเดลการประมาณค่ารูปแบบเจาะจงข้อสอบ (RPFI) แต่โมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีความคลาดเคลื่อนในการประมาณค่าใกล้เคียงกับโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI)

นอกจากนี้ เมื่อขนาดกลุ่มตัวอย่างมากขึ้น จะทำให้ค่าความคลาดเคลื่อนในการประมาณค่าและค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่ามีค่าลดลง สอดคล้องกับงานวิจัยของ ชลี ภัทรพิชญธรรม (2553) และ ชนะศึก นิชานนท์ (2553) เมื่อขนาดกลุ่มตัวอย่างมากขึ้น จะทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าลดลงสอดคล้องกับงานวิจัยของ de la Torre, Stark and Chernyshenko (2006) และเมื่อความยาวแบบสอบมากขึ้น จะทำให้ค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าลดลง สอดคล้องกับงานวิจัยของ de la Torre, Stark and Chernyshenko (2006) และ ชนะศึก นิชานนท์ (2553)

2. ประสิทธิภาพของโมเดล พิจารณาจากความสอดคล้องกลมกลืนของโมเดลกับข้อมูลผลสอบ O-NET ปีการศึกษา 2553 ระดับชั้นมัธยมศึกษาปีที่ 6 วิชาภาษาอังกฤษ วิชาภาษาไทย และวิชาคณิตศาสตร์ ที่มีจำนวนข้อสอบ 70, 90 และ 40 ข้อ ตามลำดับ ผลการวิจัย พบว่า ไม่เป็นไปตามสมมติฐาน ข้อที่ 3 กล่าวคือ โมเดลการประมาณค่าพารามิเตอร์ผู้สอบทั้ง 3 โมเดล มีความสอดคล้องระหว่างโมเดลกับข้อมูลไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ แต่เมื่อพิจารณาเปรียบเทียบประสิทธิภาพความสอดคล้องกลมกลืนของโมเดลตามเกณฑ์ที่ว่า โมเดลที่มีค่า AIC ต่ำกว่า เป็นโมเดลที่มีประสิทธิภาพสูงกว่า พบว่า โมเดลประมาณค่าพารามิเตอร์ผู้สอบ 1 พารามิเตอร์รูปแบบการสุ่มข้อสอบ มีค่าเฉลี่ย AIC ต่ำกว่า โมเดลประมาณค่าพารามิเตอร์ผู้สอบ 1 พารามิเตอร์รูปแบบการเจาะจงข้อสอบ แสดงถึงประสิทธิภาพของโมเดลประมาณค่าพารามิเตอร์ผู้สอบ 1 พารามิเตอร์รูปแบบการสุ่มข้อสอบสูงกว่ารูปแบบการเจาะจงข้อสอบ กล่าวคือ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีความสอดคล้องกลมกลืนของโมเดลกับข้อมูลดีที่สุด รองลงมาคือ รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) และ รูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ตามลำดับ แสดงว่า รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีประสิทธิภาพของโมเดลดีที่สุด รองลงมาคือ

รูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ตามลำดับ สอดคล้องกับงานวิจัยของ De Boeck (2008) สำหรับรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ที่มีความสอดคล้องกลมกลืนระหว่างโมเดลกับข้อมูลต่ำที่สุด เป็นเพราะจำนวนพารามิเตอร์ที่ต้องประมาณค่าในโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีค่าน้อยกว่า โมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) เนื่องจาก ค่า AIC ขึ้นอยู่กับจำนวนพารามิเตอร์อิสระ ตามสูตรการคำนวณ ซึ่งมีจำนวนค่าพารามิเตอร์ในโมเดลเป็นตัวคูณ; AIC =2k – 2ln(L) โดยที่ k คือ จำนวนพารามิเตอร์ (De Leeuw, 1992; Klein Entink, 2009) ค่าองศาความเป็นอิสระของแต่ละโมเดลที่แตกต่างกัน คือ รูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) มีองศาความเป็นอิสระเท่ากับ 2 x ความยาวแบบสอบ ส่วนรูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) มีองศาความเป็นอิสระเท่ากับ 1+ ความยาวแบบสอบ และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) มีองศาความเป็นอิสระเท่ากับ 3 (ประกอบด้วย ค่าเฉลี่ยของผู้สอบหรือข้อสอบ 1 ค่า และค่าความแปรปรวน 2 ค่า) ดังนั้นค่าสถิติที่คำนวณได้จากโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) จึงมีค่าน้อยกว่าโมเดลการประมาณค่ารูปแบบอื่นๆ (De Boeck, 2008)

ข้อเสนอแนะ

ข้อเสนอแนะในเชิงนโยบาย

1) โมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) เหมาะสมในการนำมาใช้สำหรับสถานการณ์การทดสอบที่มีผลกระทบสูง เช่น การทดสอบระดับชาติของนักเรียน ที่มีลักษณะข้อมูลสอดคล้องกลมกลืนกับโมเดลการตอบสนองข้อสอบ 1 พารามิเตอร์ เนื่องจากโมเดลทั้งสองมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ความสามารถผู้สอบสูงกว่าโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) ทั้งนี้จากการพิจารณาค่าความคลาดเคลื่อนในการประมาณค่าและค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าที่มีค่าต่ำกว่าแสดงถึงประสิทธิภาพการประมาณค่าที่สูงกว่า

2) โมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) และรูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) สามารถในการนำมาใช้สำหรับสถานการณ์การทดสอบระดับ

โรงเรียนหรือระดับเขตพื้นที่ ที่มีจำนวนผู้สอบน้อย (100 คน) และจำนวนข้อสอบน้อย (30 ข้อ) เนื่องจากประสิทธิภาพของการประมาณค่าพารามิเตอร์ความสามารถผู้สอบสูงกว่าโมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) และเมื่อมีจำนวนผู้สอบเพิ่มขึ้น และ/หรือมีจำนวนข้อสอบมากขึ้นประสิทธิภาพการประมาณค่าพารามิเตอร์ความสามารถผู้สอบจะสูงขึ้น

ข้อเสนอแนะในเชิงปฏิบัติ

1) โมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) เหมาะกับการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อสถานการณ์การแจกแจงก่อนหน้าของความยากข้อสอบเป็นการแจกแจงแบบปกติ เนื่องจากค่าพารามิเตอร์ที่ได้มีความคลาดเคลื่อนต่ำและค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำ นอกจากนี้ยังเหมาะกับการประมาณค่าพารามิเตอร์ที่มีการแจกแจงเริ่มต้นของความยากข้อสอบแบบ Uniform, Negative skew และ Positive skew โดยส่วนใหญ่ ยกเว้นการแจกแจงเริ่มต้นของความยากข้อสอบแบบเบ้ขวา (Positive skew) ที่มีจำนวนผู้สอบน้อย (100 คน)

โมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-สุ่มข้อสอบ (RPRI) ให้ค่าพารามิเตอร์ที่ได้มีความคลาดเคลื่อนต่ำและค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำในการประมาณค่าใกล้เคียงกับโมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) จึงเหมาะกับการประมาณค่าพารามิเตอร์ผู้สอบในสถานการณ์เดียวกัน

โมเดลการประมาณค่ารูปแบบสุ่มผู้สอบ-เจาะจงข้อสอบ (RPFI) เหมาะกับการประมาณค่าพารามิเตอร์ผู้สอบ เมื่อสถานการณ์การแจกแจงก่อนหน้าของความยากข้อสอบเป็นการแจกแจงแบบเบ้ขวา (Positive skew) ที่มีจำนวนผู้สอบน้อย (100 คน) จำนวนข้อสอบมากตั้งแต่ 30 ข้อขึ้นไป

ดังนั้น ในสถานการณ์การทดสอบในชั้นเรียนทั่วไป ข้อสอบมีลักษณะการแจกแจง Normal ที่มีลักษณะการแจกแจงแบบปกติ โมเดลที่เหมาะกับการประมาณค่าพารามิเตอร์ผู้สอบ คือ โมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) เนื่องจากค่าพารามิเตอร์ที่ได้มีความคลาดเคลื่อนต่ำและค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำกว่าโมเดลอื่นๆ และในสถานการณ์การทดสอบที่ต้องการคัดเลือกผู้มีความสามารถสูง เช่น การสอบชิงทุน

ข้อสอบมีลักษณะการแจกแจง Negative skew ที่มีลักษณะการแจกแจงแบบเบ้ซ้าย โมเดลที่เหมาะกับการประมาณค่าพารามิเตอร์ผู้สอบ คือ โมเดลการประมาณค่ารูปแบบเจาะจงผู้สอบ-สุ่มข้อสอบ (FPRI) เนื่องจากค่าพารามิเตอร์ที่ได้มีความคลาดเคลื่อนต่ำและค่าเฉลี่ยความคลาดเคลื่อนกำลังสองในการประมาณค่าต่ำกว่าโมเดลอื่นๆ

2) ขนาดกลุ่มตัวอย่าง และความยาวแบบสอบ เป็นสิ่งที่ต้องคำนึงถึงสำหรับการประมาณค่าพารามิเตอร์ผู้สอบ จากผลการวิจัยเมื่อมีจำนวนผู้สอบมากขึ้น หรือจำนวนข้อสอบมากขึ้น จะยิ่งทำให้ค่าความคลาดเคลื่อนในการประมาณค่าและค่าเฉลี่ยความคลาดเคลื่อนกำลังสองลดลง แสดงถึงประสิทธิภาพในการประมาณค่าที่เพิ่มขึ้น

ข้อเสนอแนะสำหรับการทำวิจัยครั้งต่อไป

1) การศึกษาครั้งนี้เป็นการศึกษาประสิทธิภาพในการประมาณค่าพารามิเตอร์ผู้สอบของโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ จึงมีเพียงพารามิเตอร์เดียวในการประมาณค่าเท่านั้น ยังไม่ได้พิจารณาพารามิเตอร์อำนาจจำแนกและพารามิเตอร์การเดาข้อสอบได้ถูก ดังนั้นการวิจัยครั้งต่อไปควรศึกษาเพิ่มเติมในโมเดลตอบสนองข้อสอบแบบ 2 และ 3 พารามิเตอร์ เพื่อให้ครอบคลุมสถานการณ์การทดสอบที่เป็นไปได้ในสภาพจริง

2) การศึกษาครั้งนี้มีขอบเขตการวิจัยโดยกำหนดให้ การแจกแจงก่อนหน้าเป็นการแจกแจงแบบปกติของความสามารถของผู้สอบ ยังไม่ได้พิจารณาการแจกแจงแบบอื่นๆของความสามารถผู้ตอบ ดังนั้นการวิจัยครั้งต่อไปควรมีการศึกษาการแจกแจงแบบอื่นๆของความสามารถผู้สอบเพิ่มเติม เช่น การกำหนดให้การแจกแจงก่อนหน้าของความสามารถผู้สอบมีการแจกแจงแบบ Uniform ที่มีลักษณะการแจกแจงใกล้เคียงกับรูปสี่เหลี่ยมผืนผ้า หรือมีการแจกแจงแบบ Positive Skew ที่มีลักษณะการแจกแจงแบบเบ้ขวา เป็นต้น

3) การศึกษาครั้งนี้ ศึกษาโมเดลการตอบสนองข้อสอบแบบเอกมิติ (unidimension) เท่านั้น ยังไม่ได้พิจารณากรณีข้อสอบแบบพหุมิติ (multidimension) ดังนั้นการวิจัยครั้งต่อไปควรศึกษาในโมเดลตอบสนองข้อสอบแบบ multidimension เพิ่มเติม เพราะอาจจะให้ประสิทธิภาพการประมาณค่าด้วยโมเดลสุ่มข้อสอบที่แตกต่างออกไป

รายการอ้างอิง

ภาษาไทย

ชลี ภัทรพิชญธรรม. (2553). *การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทางด้วยโมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น : การประมาณค่าพารามิเตอร์.* วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ชนะศึก นิชานนท์. (2553). *ประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบเบส์โดยใช้การสรุปอ้างอิงความน่าเชื่อถือของโมเดลการตอบสนองข้อสอบ*. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ภัทรสุดา สุดแสน. (2548). *การเปรียบเทียบเกณฑ์การคัดเลือกตัวแบบการถดถอยเชิงเส้นเมื่อตัวอย่างมีขนาดเล็ก*. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาสถิติ ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.

มานพ วราภักดิ์. (2550). *การจำลอง*. พิมพ์ครั้งที่ 1. กรุงเทพฯ : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

วิชุดา บัวคง. (2533). *การเปรียบเทียบประสิทธิผลของวิธีการประมาณค่าพารามิเตอร์ของแบบจำลองโลจิสติก 3 พารามิเตอร์ ระหว่างวิธีแมกซิมัมไลค์ลิฮูด วิธีฮิวริสติก และวิธีของเบย์ในแบบสอบวัดผลสัมฤทธิ์ และแบบสอบความถนัด*. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

วินัย วงศ์ฤทัยวัฒนา. (2533). *การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ของโมเดลโลจิสติกแบบสองพารามิเตอร์ระหว่างวิธีของเบย์กับวิธีแมกซิมัมไลดค์ลิฮูด*. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

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

ศิริชัย กาญจนวาสี. (2552). *ทฤษฎีการทดสอบแบบดั้งเดิม*. พิมพ์ครั้งที่ 6. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี และคณะ. (2551). *การเลือกใช้สถิติที่เหมาะสมสำหรับการวิจัย*. พิมพ์ครั้งที่ 5. กรุงเทพฯ : โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย พงษ์วิชัย. (2552). สถิติเพื่อการวิจัยด้วยโปรแกรม R: เล่มที่ 1 การทดสอบความแตกต่าง. พิมพ์ครั้งที่ 1. กรุงเทพฯ : บริษัท สุพีเรียพริ้นติ้งเฮาส์ จำกัด.

ศิริชัย พงษ์วิชัย. (2552). สถิติเพื่อการวิจัยด้วยโปรแกรม R: เล่มที่ 2 การทดสอบความสัมพันธ์. พิมพ์ครั้งที่ 1. กรุงเทพฯ : บริษัท สุพีเรียพริ้นติ้งเฮาส์ จำกัด.

สุนทร เทียนงาม. (2551). *ผลของความไม่เป็นอิสระของข้อสอบที่มีต่อค่าความเที่ยง ค่าพารามิเตอร์ของข้อสอบ ค่าความสามารถของผู้สอบและค่าสารสนเทศของแบบสอบ เมื่อมีเงื่อนไขการทดสอบที่แตกต่างกัน.* วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน, กระทรวงศึกษาธิการ. [ออนไลน์]. (2551). แหล่งที่มา : www.obec.go.th [14 มิถุนายน 2552]

สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน). (2551). รายงานประเมินตนเองสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) ประจำปีการศึกษา 2551.

ภาษาอังกฤษ

Aiken, Lewis R. (1998). *Tests and examinations : measuring abilities and performance*. New York : Wiley.

Baker, F. B. (1992). *Item Response Theory : Parameter Estimation Techniques*. N.Y. : Marcel Dekker.

Baker, R. (1997). *Classical test theory and item response theory in test analysis*. Lancaster : Centre for Research in Language Education.

Böchenholt, U., & van der Heijden. (2007). Item Randomized-Response Models for Measuring Noncompliance: Risk-Return Perceptions, Social Influences, and Self-Protective. *Psychometrika*, 72(2) : 245-262.

Briggs, D. and Wilson, M. (2007). Generalizability in Item Response Modeling. *Journal of Educational Measurement*, 44(2) : 131-155.

Briggs, D.C. (2008). Using Explanatory Item Response Models to Analyze Group Difference in Science Achievement. *Applied Measurement in Education*, 21 : 89-118.

De la Torre, J., Stark, S. and Chernyshenko, O. S. (2006). Marginal Maximum A Posteriori Item Parameter Estimation for the Generalized Graded Unfolding Model. *Applied Psychological Measurement*, 30(3) : 216-232.

De Leeuw, J. (1992). Introduction to Akaike (1973) Information Theory and an Extension of the Maximum Likelihood Principle. Pages 599-609 In: Kotz, S., and N.L.Johmson, editors. *Breakthoughs in Statistics Volume 1*. Foundations and Basic Theory. Springer Series in Statistics, Perspectives in Statistics. Springer-Verlag: New York.

De Boeck, P. (2008). Random Item IRT Models. *Psychometrika*, 73(4) : 533-559.

De Boeck, P., & Wilson, M. editor (2005). Explanatory item response models: A Generalized Linear and Nonlinear Approach. reviewed by Cees A. W. Glas. *Journal of Educational Measurement*, 42(3) : 303-307.

De Jong, M. G., Steenkamp, J. B. E. M., Fox, J-P., & Baumgartner, H. (2008). Using item response theory to measure extreme response style in marketing research: A global investigation. *Journal of Marketing Research,* 104 : 104–115.

Edwards M.C. (2008). Item factor analysis: Where we've been and where we might be going. Paper presented at the 73rd annual meeting of the Psychometric Society. [Online]. 2010. Available from http://faculty.psy.ohio-state.edu/edwards/ documents/IMPS6.30.08.pdf/ [2010, November 19].

Embretson, S.E., & Reise, S. P. (2000). *Item Response Theory for Psychologists.* Mahwah, N.J. : Lawrence Erlbaum Associates.

Fan, X. (1998). Item response theory and classical test theory: an empirical comparison of their item/person statistics. *Educational and Psychological Measurement*, 58(3), 357(25).

Farmer, W. L., Thompson, R. C., Heil, S. K. R., & Heil, M. C. (2001). Latent Trait Theory Analysis of Changes in Item Response Anchors. (Final Report) *U.S. Department of Transportation Federal Aviation Administration*. Office of Aviation Medicine: Washington, D.C. 20591.

Fox, J.-P. (2005). Ramdomized Item Response Theory Models. *Journal of Educational and Behavioral Statistics*, 30(2), 189-212.

Fox, J.-P. and Wyrick, C. (2008). A Mixed Effects Randomized Item Response Model. *Journal of Educational and Behavioral Statistics*, 33(4) : 389-415.

Frederiksen, N., Mislevy, R. J., Bejar, I. I. (1993). *Test theory for a new generation of tests*. Hillsdale, New Jersey : Lawrence erlbaum Associates.

Glas, C. A. W., & Van der Linden, W. J. (2003). Computerized adaptive testing with item cloning. *Applied Psychological Measurement,* 27 : 247–261.

Glass, G. V., & Stanley, J. C. (1970). *Statistical Methods in Education and Psychology*. Englewood Cliffs, N.J. : Prentice-Hall, Inc.

Goldstein, H. (2002). *Review of WINBUGS*. [Online]. 2010. Available from http://scholar.google.co.th/scholar?q=related:87l5vgxkvRYJ:scholar.google.com/&hl=th&as_sdt=2000. [2010, September 29].

Hambleton, R.K. (1989). Principles and selected applications of item response theory. In R. L. Linn (Ed.), *Educational measurement* (3rd ed). New York: Macmillian.

Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: principles and applications*. Boston: Kluwer-Nijh off Publishing.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA:Sage.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3) : 38-47.

Hare, R. D. (1999). Evaluating the Screening Version of the Hare Psychopathy Checklist-Revised (PCL:SV): An Item Response Theory Analysis. *Psychological Assessment*, 11 (1) : 3-13.

Janssen, R., Tuerlinckx, F., Meulders, M., & De Boeck, P. (2000). A hierarchical IRT model for criterion-referenced measurement. *Journal of Educational and Behavioral Statistics,* 25 : 285–306.

Klein Entink, R.H. (2009). *Statistical Models for Responses and Response Times*. University of Twente, Enschede: Nederlands.

Linn, R. L.(1989). *Educational measurement* (3ed ed.). New York: Macmillian.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale, N.J. : Lawrence Erlbaum Associates.

McDonald, R. P. (1999). *Test Theory: A Unified Treatment*. Mahwah, N.J. : Lawrence erlbaum associated.

Newsom. (2009). *Distinguishing Between Random and Fixed: Variables, Effects, and Coefficients.* [Online]. 2011. Available from http://www.upa.pdx.edu/IOA/newsom/mlrclass/ho_randfixd.pdf. [2011, September 30].

Paek, I., & Young, M. J. (2005). Investigation of Student Growth Recovery in a Fixed-Item Linking Procedure With a Fixed-Person Prior Distribution for Mixed Format Test Data. *Applied Measurement in Education*, 18(2) : 199-215.

Paradis, E. (2005). *R for Beginners.* [Online]. 2010. Available from http://mirror.kapook.com/cran/ [2010, November 24].

Reckase, M. D. (2009). Multidimentional Item Response Theory. *Statistics for Social and Behavioral Sciences*. N.Y. : Springer Science+Business Media.

Rovine, M. J., & Molenaar, P. C. M. (2000). A Structural Modeling Approach to a Multilevel Random Coefficients Model. *Multivariate Behavioral Research*, 35(1) : 51-88.

Sturtz, S., Ligges, U., Gelman, A. (2005). R2WinBUGS: A Package for Running WinBUGS from R. *Journal of Statistical Software*. [Online]. 2010. Available from citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.161. [2010, September 29].

Suen, Hoi K. (1990). *Principles of test theories*. Hillsdate, N.J. : Lawrence Erlbaum Associates.

Van der Linden, W. J., Hambliton, R. K. (1997). *Handbook of Modern Item Response Theory*. N.Y. : Springer.

Van den Norrtgate, W., De Boeck, P., & Meulders, M. (2003). Cross-Classification Multilevel Logistic Models in Psychometrics. *Journal of Educational and Behavioral Statistics*, 28(4) : 369-386.

Venables,W. N., Smith, D. M. and the R Development Core Team. [Online]. 2009. Available from *An Introduction to R*. http://mirror.kapook.com/cran/. [2009, November 24].

Vijver, F. J. R. van de (1986). The robustness of Rasch estimates. *Applied Psychological Measurement,* 10(1), 45-57. [Online]. 2008. Available from http://arno.uvt.nl/show.cgi?fid=26717/. [2009, November 24].

Wiberg, M. (2004). Classical Test Theory vs. Item Response Theory. *An evaluation of the theory test in the Swedish driving-license test*. (EM No. 50): Umeå Universitet.

Wilson, M., & Engelhard, G, Jr. (editor). (2000). *Objective Measurement: Theory into Practice*. Vol.5. Norwood, New Jersdy : Ables.

Yi, H. S. (2009). Feasibility of Using Prior Information about Predicted Item Difficulty in Increasing the Accuracy of Item Parameter Estimation and IRT Equating. *The Journal of Curriculum and Evaluation*, 12(1) : 103-122.

ภาคผนวก

ภาคผนวก ก

ตัวอย่างคำสั่งโปรแกรม R และ WinBUGS

คำสั่งการทำงานในโปรแกรม R

```
########Code Run Normal Distribution#########
library("psych")
library("ltm")
library("R2WinBUGS")
N<-1000 #number of examinees
I<-90 #number of items
rep=800 #Monte Caelo replications
iter=1000 #number of MCMC iterations

t<-matrix(nrow=N,ncol=rep) #true theta per replication
b<-matrix(nrow=I,ncol=rep) #true beta per replication

t1.mean<-matrix(nrow=N,ncol=rep) #theta mean FIRP
b1.mean<-matrix(nrow=I,ncol=rep) #beta mean FIRP
sigma2.t1.mean<-matrix(nrow=rep) #sigma.theta mean FIRP
t1.sd<-matrix(nrow=N,ncol=rep) #theta sd FIRP
b1.sd<-matrix(nrow=I,ncol=rep) #beta sd FIRP
sigma2.t1.sd<-matrix(nrow=rep) #sigma.theta sd FIRP

t2.mean<-matrix(nrow=N,ncol=rep) #theta mean RIFP
b2.mean<-matrix(nrow=I,ncol=rep) #beta mean RIFP
sigma2.b2.mean<-matrix(nrow=rep) #sigma.theta mean RIFP
t2.sd<-matrix(nrow=N,ncol=rep) #theta sd RIFP
b2.sd<-matrix(nrow=I,ncol=rep) #beta sd RIFP
sigma2.b2.sd<-matrix(nrow=rep) #sigma.theta sd RIFP

t3.mean<-matrix(nrow=N,ncol=rep) #theta mean RIRP
b3.mean<-matrix(nrow=I,ncol=rep) #beta mean RIRP
sigma2.t3.mean<-matrix(nrow=rep) #sigma.theta mean RIRP
sigma2.b3.mean<-matrix(nrow=rep) #sigma.theta mean RIRP
t3.sd<-matrix(nrow=N,ncol=rep) #theta sd RIRP
b3.sd<-matrix(nrow=I,ncol=rep) #beta sd RIRP
sigma2.t3.sd<-matrix(nrow=rep) #sigma.theta sd RIRP
sigma2.b3.sd<-matrix(nrow=rep) #sigma.theta sd RIRP

#1) function to generate probability for given theta and beta
```

```
# logistic cumulative normal function
logis.irt<-function(t,b,prob)
{
prob<-1/(1+exp(-1.7*(t-b)))
}

#seed<-93848
#set.seed(seed, kind=NULL, normal.kind=NULL)

iterate=0
for (k in 1:rep)
{

## generate trait score using normal(0,1)
t[,k]<-rnorm(N,0,1) #true trait score

#gen.beta<-function(I)
#{
#item<-1:I
#hi=3 #difficulty of most item difficulties
#range<-2*hi #range of item difficulties
#b<-(item-mean(item))*range/(I-1)
#return(b)
#}

#b<-gen.beta(I) #true difficult item parameters

b[,k]<-rnorm(I,0,0.5)

prob<-matrix(nrow=N,ncol=I)   ### Matrix for item response probabilities
y<-matrix(nrow=N,ncol=I)
for (i in 1:N)
{
  for (j in 1:I)
   {
prob[i,j]<-logis.irt(t[i,k],b[j],prob)
y[i,j]<-ifelse(runif(1,0,1)<prob[i,j],1,0)
```

```
  }
}

data<-list("y","N","I")

inits<-function(){list(theta=rnorm(N,0,1),tau.theta<-runif(1,0,10),beta=rnorm(I,0,1),mu.theta<-
runif(1,0,10),tau.beta<-rnorm(1,0,10))}

start<-Sys.time()
output1<-bugs(data,inits,model.file="C:/Users/Admin/Documents/my
thesis/trial2/IRT_FIRP.txt",parameters=c("theta","beta","sigma2.theta"),n.chains=1,n.iter=iter,n.burnin=
100,n.thin=3,bugs.directory="C:/Program Files/WinBUGS14",debug=FALSE)
time1<-Sys.time()-start

t1.mean[,k]<-output1$mean$theta
b1.mean[,k]<-output1$mean$beta
sigma2.t1.mean[k]<-output1$mean$sigma2.theta

t1.sd[,k]<-output1$sd$theta
b1.sd[,k]<-output1$sd$beta
sigma2.t1.sd[k]<-output1$sd$sigma2.theta

start<-Sys.time()
output2<-bugs(data,inits,model.file="C:/Users/Admin/Documents/my
thesis/trial2/IRT_RIFP.txt",parameters=c("theta","beta","sigma2.beta"),n.chains=1,n.iter=iter,n.burnin=
100,n.thin=3,bugs.directory="C:/Program Files/WinBUGS14",debug=FALSE)
time2<-Sys.time()-start


t2.mean[,k]<-output2$mean$theta
b2.mean[,k]<-output2$mean$beta
sigma2.b2.mean[k]<-output2$mean$sigma2.beta

t2.sd[,k]<-output2$sd$theta
b2.sd[,k]<-output2$sd$beta
sigma2.b2.sd[k]<-output2$sd$sigma2.beta
```

```
start<-Sys.time()
output3<-bugs(data,inits,model.file="C:/Users/Admin/Documents/my
thesis/trial2/IRT_RIRP.txt",parameters=c("theta","beta","sigma2.beta","sigma2.theta"),n.chains=1,n.ite
r=iter,n.burnin=100,n.thin=3,bugs.directory="C:/Program Files/WinBUGS14",debug=FALSE)
time2<-Sys.time()-start

t3.mean[,k]<-output3$mean$theta
b3.mean[,k]<-output3$mean$beta
sigma2.t3.mean[k]<-output3$mean$sigma2.theta
sigma2.b3.mean[k]<-output3$mean$sigma2.beta
t3.sd[,k]<-output3$sd$theta
b3.sd[,k]<-output3$sd$beta
sigma2.t3.sd[k]<-output3$sd$sigma2.theta
sigma2.b3.sd[k]<-output3$sd$sigma2.beta

iterate=iterate+1
cat("replication= ",iterate,"\n")
}

################### end of looping ########################

###### ITEM INFORMATION ######
item.inf<-function(t.est,b.est,N,I) #calculate item informations
{

P<-matrix(nrow=N,ncol=I)
I.i<-matrix(nrow=N,ncol=I) #item information

for (i in 1:N)
{
for (j in 1:I)
{
P[i,j]<-1/(1+exp((-1.7)*(t.est[i]-b.est[j])))
I.i[i,j]<-(1.7)^2*(P[i,j])*(1-P[i,j])
}
}
return(I.i)
```

```
}
################################

####### TEST INFORMATION #######
test.inf<-function(I.ti) #calculate test informations
{
N<-dim(I.ti)[1]
I.t<-matrix(nrow=N,ncol=1)
I.t[]<-rowSums(I.ti)

SE<-1/sqrt(I.t)
#out<-list(SE.t,I.t)
#names(out)<-c("SE.theta","Test information")
return(SE)
}
################################
############ CALCULATE SE(THETA) and SE(BETA) ###############
# SE of each method
SE1.rep<-matrix(nrow=N,ncol=rep) #SE of FIRP
SE2.rep<-matrix(nrow=N,ncol=rep) #SE of RIFP
SE3.rep<-matrix(nrow=N,ncol=rep) #SE of RIRP
SE1<-matrix(nrow=N,ncol=1) # mean of SE1.rep per person
SE2<-matrix(nrow=N,ncol=1) # mean of SE2.rep per person
SE3<-matrix(nrow=N,ncol=1) # mean of SE3.rep per person

for (k in 1:rep)
{
I1<-item.inf(t1.mean[,k],b1.mean[,k],N,I)
SE1.rep[,k]<-test.inf(I1)

I2<-item.inf(t2.mean[,k],b1.mean[,k],N,I)
SE2.rep[,k]<-test.inf(I2)

I3<-item.inf(t3.mean[,k],b1.mean[,k],N,I)
SE3.rep[,k]<-test.inf(I3)
}
```

```
SE1[,1]<-rowMeans(SE1.rep)
SE2[,1]<-rowMeans(SE2.rep)
SE3[,1]<-rowMeans(SE3.rep)

#sd of theta and beta
combine.sd.t1<-matrix(nrow=N,ncol=1)
combine.sd.t2<-matrix(nrow=N,ncol=1)
combine.sd.t3<-matrix(nrow=N,ncol=1)

combine.sd.b1<-matrix(nrow=I,ncol=1)
combine.sd.b2<-matrix(nrow=I,ncol=1)
combine.sd.b3<-matrix(nrow=I,ncol=1)

n<-(iter-1000)/3
for (k in 1:rep)
{

for (i in 1:N)
{
combine.sd.t1[i,1]<-(sum(t1.sd[i,]^2)*(n-1))/(rep*n-rep)
combine.sd.t2[i,1]<-(sum(t2.sd[i,]^2)*(n-1))/(rep*n-rep)
combine.sd.t3[i,1]<-(sum(t3.sd[i,]^2)*(n-1))/(rep*n-rep)

}
for (i in 1:I)
{
combine.sd.b1[i,1]<-(sum(b1.sd[i,]^2)*(n-1))/(rep*n-rep)
combine.sd.b2[i,1]<-(sum(b2.sd[i,]^2)*(n-1))/(rep*n-rep)
combine.sd.b3[i,1]<-(sum(b3.sd[i,]^2)*(n-1))/(rep*n-rep)
}
}

combine.sd.t<-data.frame(combine.sd.t1,combine.sd.t2,combine.sd.t3)
names(combine.sd.t)<-c("sd.t.FIRP","sd.t.RIFP","sd.t.RIRP")
combine.sd.b<-data.frame(combine.sd.b1,combine.sd.b2,combine.sd.b3)
names(combine.sd.b)<-c("sd.b.FIRP","sd.b.RIFP","sd.b.RIRP")
```

```
plot(combine.sd.t1,type="l")
lines(combine.sd.t2,type="l",col="red")
lines(combine.sd.t3,type="l",col="blue")

plot(combine.sd.b1,type="l",ylim=c(0.005,0.016))
lines(combine.sd.b2,type="l",col="red")
lines(combine.sd.b3,type="l",col="blue")

################### END OF SE #######################################


################### MSE CALCULATION#################################

##### MSE for each method

### MSE for THETA ####
MSE1.t<-matrix(nrow=N,ncol=1) #MSE for FIRP (theta)
MSE2.t<-matrix(nrow=N,ncol=1) #MSE for RIFP (theta)
MSE3.t<-matrix(nrow=N,ncol=1) #MSE for RIRP (theta)

MSE1.b<-matrix(nrow=I,ncol=1) #MSE for FIRP (beta)
MSE2.b<-matrix(nrow=I,ncol=1) #MSE for RIFP (beta)
MSE3.b<-matrix(nrow=I,ncol=1) #MSE for RIRP (beta)

for (i in 1:N)
{
MSE1.t[i,1]<-t(t1.mean[i,]-t[i,])%*%(t1.mean[i,]-t[i,])/rep
MSE2.t[i,1]<-t(t2.mean[i,]-t[i,])%*%(t2.mean[i,]-t[i,])/rep
MSE3.t[i,1]<-t(t3.mean[i,]-t[i,])%*%(t3.mean[i,]-t[i,])/rep
}

for (i in 1:I)
{
MSE1.b[i,1]<-t(b1.mean[i,]-b[i,])%*%(b1.mean[i,]-b[i,])/rep
MSE2.b[i,1]<-t(b2.mean[i,]-b[i,])%*%(b2.mean[i,]-b[i,])/rep
MSE3.b[i,1]<-t(b3.mean[i,]-b[i,])%*%(b3.mean[i,]-b[i,])/rep
}
```

```
MSE.theta<-data.frame(MSE1.t,MSE2.t,MSE3.t) #MSE for theta

MSE.beta<-data.frame(MSE1.b,MSE2.b,MSE3.b) #MSE for beta

##############################################################################
```

คำสั่งการทำงานในโปรแกรม WinBUGS

```
##### FPRI Estimate Model#####

model{

for (j in 1:I)
{
beta[j]~dnorm(0,tau.beta)
for (i in 1:N)
{

y[i,j]~dbern(prob[i,j])
prob[i,j]<-1/(1+exp(-1.7*(theta[i]-beta[j])))

}
}

for (i in 1:N)
{
theta[i]~dnorm(0,1)
}
#prior for hyperparameters
#mu.beta~dnorm(0,0.01)
tau.beta~dgamma(0.1,0.1)
sigma2.beta<-1/tau.beta
}

###########END###########
```

คำสั่งที่ใช้ในการสุ่มข้อมูลจริง ประมาณค่าพารามิเตอร์ และคำนวณค่า AIC

```
#############   SAMPLING REAL DATA  ####################
sampling<-function(response,size)
{
person<-size
number<-seq(from=1,to=dim(response)[1],by=1)
sample<-sample(number,size=person)
response<-as.matrix(response)

y<-matrix(nrow=person,ncol=dim(response)[2])

for (i in 1:person)
{
y[i,]<-response[sample[i],]
}
return(y)
}
#####################################################
library("R2WinBUGS")
z<-read.csv("C:/Users/Admin/Documents/my thesis/run real data/thai6_cut guess.csv",header=TRUE)

boot<-1
n<-100
AIC.boot<-matrix(nrow=boot,ncol=3)
colnames(AIC.boot)<-c("AIC1","AIC2","AIC3")

start<-Sys.time()
for (b in 1:boot)
{
y<-sampling(z,n)

N<-dim(y)[1]
I<-dim(y)[2]
y<-as.matrix(y)
data<-list("y","N","I")

#N<-dim(z)[1]
```

```
#I<-dim(z)[2]
#z<-as.matrix(z)
#data<-list("z","N","I")

inits<-function(){list(theta=rnorm(N,0,1),tau.theta<-runif(1,0,10),beta=rnorm(I,0,1),mu.theta<-runif(1,0,10),tau.beta<-
rnorm(1,0,10))}
iter=1000
start<-Sys.time()
output1<-bugs(data,inits,model.file="C:/Users/Admin/Documents/my
thesis/trial2/IRT_FIRP.txt",parameters=c("theta","beta","sigma2.theta"),n.chains=1,n.iter=iter,n.burnin=100,n.thin=3,bug
s.directory="C:/Program Files/WinBUGS14",debug=FALSE)
time1<-Sys.time()-start

start<-Sys.time()
output2<-bugs(data,inits,model.file="C:/Users/Admin/Documents/my
thesis/trial2/IRT_RIFP.txt",parameters=c("theta","beta","sigma2.beta"),n.chains=1,n.iter=iter,n.burnin=100,n.thin=3,bug
s.directory="C:/Program Files/WinBUGS14",debug=FALSE)
time2<-Sys.time()-start

start<-Sys.time()
output3<-bugs(data,inits,model.file="C:/Users/Admin/Documents/my
thesis/trial2/IRT_RIRP.txt",parameters=c("theta","beta","sigma2.beta","sigma2.theta"),n.chains=1,n.iter=iter,n.burnin=1
00,n.thin=3,bugs.directory="C:/Program Files/WinBUGS14",debug=FALSE)
time3<-Sys.time()-start

#DIC
DIC1<-output1$DIC
DIC2<-output2$DIC
DIC3<-output3$DIC

DIC<-data.frame(DIC1,DIC2,DIC3)

## AIC calculation
num.par1<-N+I+1
num.par2<-N+I+1
num.par3<-N+I+2

dev1<-output1$summary[num.par1+1,1]
```

```
dev2<-output2$summary[num.par2+1,1]
dev3<-output3$summary[num.par3+1,1]

AIC1<-num.par1+dev1
AIC2<-num.par2+dev2
AIC3<-num.par3+dev3
AIC.boot[b,]<-c(AIC1,AIC2,AIC3)
}
time<-Sys.time()-start
time
write.csv(AIC.boot,file="AIC.boot.csv")
############################END#########################################
```

ภาคผนวก ข

ค่าความยากข้อสอบของแบบสอบ O-NET

วิชาภาษาอังกฤษ ภาษาไทย และคณิตศาสตร์ ม.6

ค่าความยากวิชาภาษาอังกฤษ 70

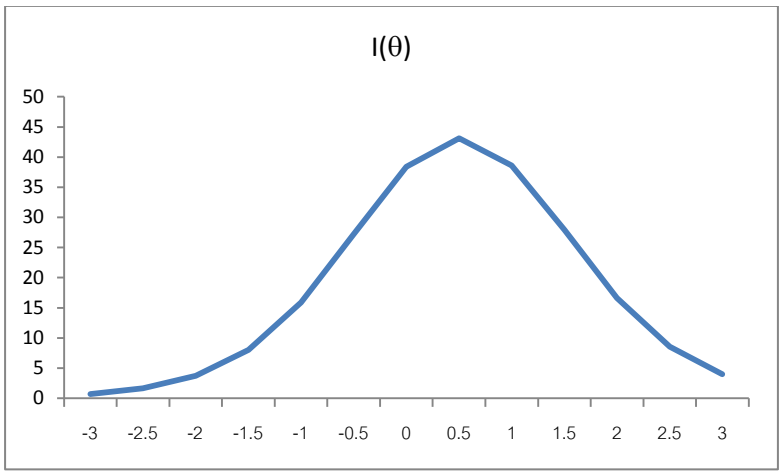
ข้อ	Beta	ข้อ	Beta
1	0.52	36	2.23
2	0.66	37	2.29
3	0.71	38	2.34
4	0.48	39	2.4
5	0.28	40	2.46
6	0.57	41	2.51
7	-0.04	42	2.57
8	0.25	43	2.63
9	-0.11	44	2.68
10	0.59	45	2.74
11	1.02	46	2.8
12	0.76	47	2.85
13	1.13	48	2.91
14	1.14	49	2.97
15	1.41	50	3.02
16	1.26	51	3.08
17	1.29	52	3.14
18	1.41	53	3.19
19	0.97	54	3.25
20	1.32	55	3.31
21	1.38	56	3.36
22	1.43	57	3.42
23	1.49	58	3.48
24	1.55	59	3.53
25	1.6	60	3.59
26	1.66	61	3.65
27	1.72	62	3.7
28	1.77	63	3.76
29	1.83	64	3.82
30	1.89	65	3.88
31	1.94	66	3.93
32	2	67	3.99
33	2.06	68	4.05
34	2.12	69	4.1
35	2.17	70	4.16

ตัวอย่างค่า Test Information ของแบบสอบวิชาภาษาอังกฤษ 70 ข้อ จากผู้สอบ 1,000 คน (โมเดล RPFI)

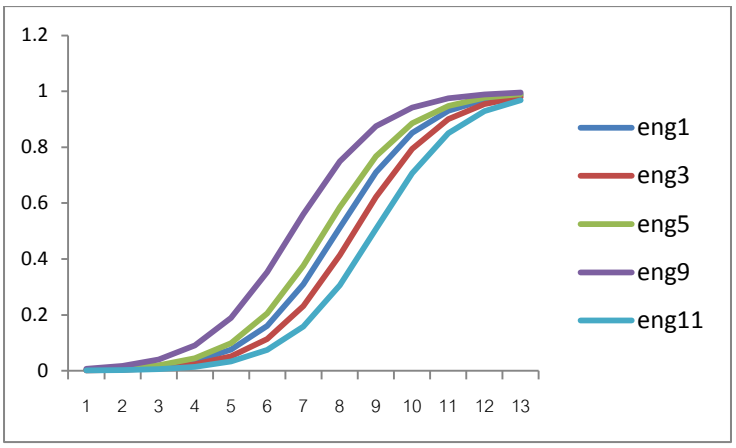
	$I_i(\theta)$												
Item	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3
eng1	0.01	0.02	0.04	0.09	0.20	0.39	0.62	0.72	0.59	0.36	0.19	0.09	0.04
eng2	0.01	0.01	0.03	0.07	0.15	0.31	0.54	0.71	0.66	0.45	0.24	0.12	0.05
eng3	0.01	0.01	0.03	0.06	0.14	0.29	0.51	0.70	0.68	0.47	0.26	0.12	0.06
eng4	0.01	0.02	0.04	0.10	0.21	0.41	0.63	0.72	0.58	0.35	0.18	0.08	0.04
eng5	0.01	0.02	0.06	0.12	0.26	0.47	0.68	0.70	0.52	0.29	0.14	0.07	0.03
eng6	0.01	0.02	0.04	0.09	0.19	0.37	0.60	0.72	0.61	0.38	0.20	0.09	0.04
eng7	0.02	0.04	0.10	0.21	0.40	0.63	0.72	0.59	0.36	0.18	0.08	0.04	0.02
eng8	0.01	0.03	0.07	0.15	0.30	0.52	0.70	0.68	0.47	0.26	0.12	0.06	0.02
eng9	0.02	0.05	0.11	0.24	0.44	0.66	0.71	0.54	0.32	0.16	0.07	0.03	0.01
eng10	0.01	0.02	0.04	0.08	0.18	0.35	0.58	0.72	0.63	0.40	0.21	0.10	0.04
eng11	0.00	0.01	0.02	0.04	0.09	0.20	0.39	0.61	0.72	0.60	0.37	0.19	0.09
eng12	0.00	0.01	0.03	0.06	0.13	0.27	0.49	0.69	0.69	0.50	0.28	0.14	0.06
eng13	0.00	0.01	0.01	0.03	0.07	0.16	0.31	0.54	0.71	0.66	0.45	0.24	0.11
eng14	0.00	0.01	0.01	0.03	0.07	0.14	0.30	0.52	0.70	0.68	0.47	0.26	0.12
eng15	0.00	0.00	0.01	0.02	0.05	0.12	0.25	0.45	0.67	0.71	0.53	0.31	0.15
eng16	0.00	0.01	0.01	0.03	0.06	0.14	0.29	0.51	0.70	0.68	0.48	0.26	0.13
eng17	0.00	0.00	0.01	0.02	0.05	0.11	0.23	0.44	0.65	0.71	0.55	0.32	0.16
eng18	0.00	0.00	0.01	0.02	0.05	0.11	0.23	0.43	0.65	0.71	0.55	0.32	0.16
eng19	0.00	0.01	0.02	0.04	0.10	0.21	0.41	0.63	0.72	0.58	0.35	0.18	0.08
eng20	0.00	0.01	0.01	0.03	0.07	0.15	0.30	0.53	0.71	0.67	0.46	0.25	0.12
eng21	0.00	0.00	0.01	0.03	0.06	0.13	0.28	0.49	0.69	0.69	0.49	0.27	0.13
eng22	0.00	0.00	0.01	0.02	0.05	0.11	0.24	0.44	0.66	0.71	0.54	0.32	0.16
eng23	0.00	0.00	0.01	0.02	0.05	0.12	0.25	0.46	0.67	0.70	0.52	0.30	0.15
eng24	0.00	0.01	0.02	0.04	0.09	0.20	0.38	0.61	0.72	0.60	0.37	0.19	0.09
eng25	0.00	0.01	0.02	0.04	0.08	0.17	0.34	0.57	0.72	0.64	0.41	0.22	0.10
eng26	0.00	0.01	0.02	0.04	0.09	0.18	0.36	0.59	0.72	0.62	0.39	0.20	0.10
eng27	0.00	0.00	0.01	0.02	0.05	0.10	0.22	0.41	0.64	0.72	0.57	0.34	0.17
eng28	0.00	0.00	0.01	0.02	0.05	0.11	0.24	0.44	0.66	0.71	0.54	0.32	0.16
eng29	0.00	0.01	0.01	0.03	0.07	0.16	0.32	0.55	0.71	0.65	0.43	0.23	0.11
eng30	0.00	0.01	0.02	0.04	0.08	0.18	0.36	0.59	0.72	0.63	0.40	0.21	0.10
eng31	0.01	0.02	0.04	0.08	0.18	0.35	0.58	0.72	0.63	0.40	0.21	0.10	0.04

	$I_i(\theta)$												
Item	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3
eng32	0.03	0.08	0.17	0.34	0.57	0.72	0.64	0.42	0.22	0.10	0.05	0.02	0.01
eng33	0.01	0.03	0.06	0.13	0.27	0.49	0.69	0.69	0.50	0.28	0.13	0.06	0.03
eng34	0.01	0.01	0.03	0.06	0.14	0.29	0.51	0.70	0.68	0.47	0.26	0.12	0.06
eng35	0.01	0.02	0.06	0.12	0.26	0.47	0.68	0.70	0.52	0.29	0.14	0.07	0.03
eng36	0.01	0.03	0.06	0.14	0.28	0.50	0.70	0.68	0.48	0.27	0.13	0.06	0.03
eng37	0.01	0.02	0.04	0.08	0.18	0.36	0.59	0.72	0.63	0.40	0.21	0.10	0.04
eng38	0.01	0.03	0.06	0.13	0.26	0.48	0.68	0.70	0.51	0.29	0.14	0.06	0.03
eng39	0.02	0.04	0.09	0.20	0.39	0.61	0.72	0.60	0.37	0.19	0.09	0.04	0.02
eng40	0.01	0.01	0.03	0.07	0.15	0.31	0.53	0.71	0.67	0.45	0.25	0.12	0.05
eng41	0.03	0.08	0.16	0.33	0.56	0.72	0.65	0.43	0.23	0.11	0.05	0.02	0.01
eng42	0.04	0.08	0.17	0.34	0.57	0.72	0.64	0.41	0.22	0.10	0.05	0.02	0.01
eng43	0.02	0.04	0.09	0.19	0.37	0.60	0.72	0.61	0.38	0.20	0.09	0.04	0.02
eng44	0.02	0.05	0.11	0.22	0.42	0.64	0.72	0.56	0.34	0.17	0.08	0.03	0.01
eng45	0.04	0.08	0.18	0.35	0.58	0.72	0.63	0.41	0.21	0.10	0.04	0.02	0.01
eng46	0.02	0.04	0.10	0.21	0.41	0.63	0.72	0.58	0.35	0.18	0.08	0.04	0.02
eng47	0.01	0.02	0.05	0.11	0.22	0.42	0.64	0.72	0.57	0.34	0.17	0.08	0.03
eng48	0.01	0.02	0.04	0.09	0.20	0.39	0.62	0.72	0.60	0.37	0.19	0.09	0.04
eng49	0.01	0.02	0.04	0.09	0.20	0.39	0.61	0.72	0.60	0.37	0.19	0.09	0.04
eng50	0.01	0.03	0.06	0.13	0.27	0.49	0.69	0.69	0.50	0.28	0.14	0.06	0.03
eng51	0.01	0.02	0.04	0.09	0.19	0.38	0.61	0.72	0.61	0.38	0.19	0.09	0.04
eng52	0.01	0.02	0.04	0.10	0.21	0.40	0.62	0.72	0.59	0.36	0.18	0.08	0.04
eng53	0.01	0.02	0.04	0.09	0.19	0.37	0.60	0.72	0.61	0.38	0.20	0.09	0.04
eng54	0.01	0.02	0.06	0.13	0.26	0.48	0.68	0.70	0.51	0.29	0.14	0.06	0.03
eng55	0.02	0.04	0.09	0.19	0.37	0.60	0.72	0.61	0.38	0.20	0.09	0.04	0.02
eng56	0.03	0.07	0.15	0.30	0.53	0.71	0.67	0.46	0.25	0.12	0.05	0.02	0.01
eng57	0.02	0.04	0.10	0.21	0.41	0.63	0.72	0.58	0.35	0.18	0.08	0.04	0.02
eng58	0.01	0.03	0.07	0.15	0.31	0.53	0.71	0.67	0.46	0.25	0.12	0.05	0.02
eng59	0.01	0.03	0.06	0.14	0.28	0.50	0.70	0.68	0.48	0.27	0.13	0.06	0.03
eng60	0.01	0.01	0.03	0.08	0.17	0.34	0.57	0.72	0.64	0.42	0.22	0.10	0.05
eng61	0.01	0.02	0.04	0.08	0.17	0.35	0.57	0.72	0.64	0.41	0.22	0.10	0.05
eng62	0.00	0.01	0.02	0.04	0.10	0.21	0.40	0.63	0.72	0.58	0.35	0.18	0.08
eng63	0.01	0.02	0.05	0.10	0.22	0.42	0.64	0.72	0.57	0.34	0.17	0.08	0.03
eng64	0.01	0.03	0.08	0.17	0.34	0.57	0.72	0.64	0.42	0.22	0.10	0.05	0.02

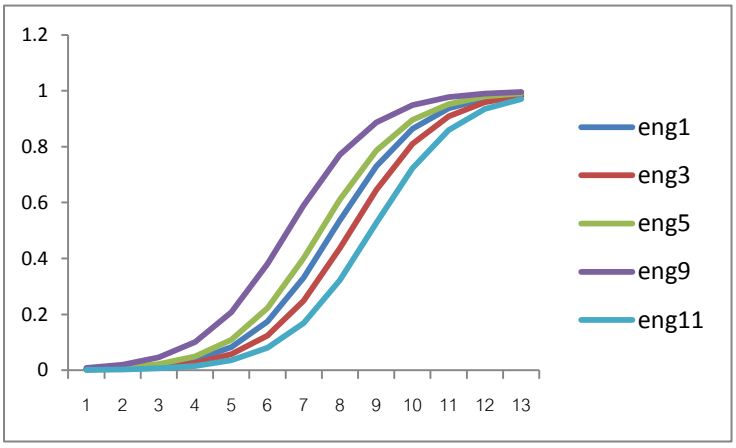
	$I_i(\theta)$												
Item	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3
eng65	0.02	0.05	0.10	0.22	0.41	0.64	0.72	0.57	0.34	0.17	0.08	0.04	0.02
eng66	0.02	0.04	0.10	0.20	0.39	0.62	0.72	0.59	0.36	0.18	0.09	0.04	0.02
eng67	0.01	0.02	0.04	0.10	0.21	0.40	0.63	0.72	0.58	0.35	0.18	0.08	0.04
eng68	0.01	0.03	0.07	0.15	0.30	0.53	0.71	0.67	0.46	0.25	0.12	0.05	0.02
eng69	0.01	0.02	0.05	0.12	0.25	0.45	0.67	0.71	0.53	0.31	0.15	0.07	0.03
eng70	0.01	0.01	0.03	0.07	0.16	0.32	0.55	0.71	0.65	0.44	0.23	0.11	0.05
$I(\theta)$	0.71	1.64	3.71	8.01	15.86	27.23	38.38	43.11	38.60	27.93	16.64	8.53	3.98
$SE(\theta)$	1.18	0.78	0.52	0.35	0.25	0.19	0.16	0.15	0.16	0.19	0.25	0.34	0.50



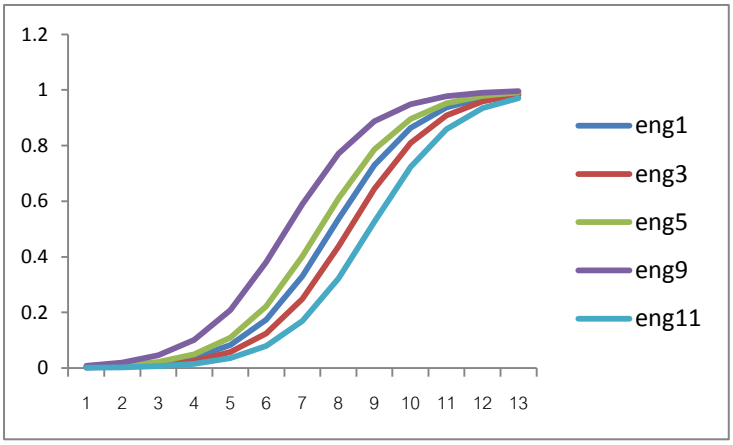
ตัวอย่าง Test Information วิชาภาษาอังกฤษ 70 ข้อ (โมเดล RPFI)



ตัวอย่างกราฟ ICC ข้อสอบวิชาภาษาอังกฤษ 5 ข้อ (โมเดล RPFI)



ตัวอย่างกราฟ ICC ข้อสอบวิชาภาษาอังกฤษ 5 ข้อ (โมเดล FPRI)



ตัวอย่างกราฟ ICC ข้อสอบวิชาภาษาอังกฤษ 5 ข้อ (โมเดล RPRI)

ค่าความยากของข้อสอบวิชาภาษาไทย 90 ข้อ

ข้อที่	Beta	ข้อที่	Beta	ข้อที่	Beta
1	0.359	41	-0.103	81	0.603
2	1.468	42	0.024	82	0.276
3	1.023	43	0.136	83	-0.091
4	1.220	44	-0.086	84	0.698
5	1.325	45	-0.153	85	0.082
6	1.179	46	-0.603	86	0.035
7	0.686	47	-0.681	87	0.338
8	0.499	48	0.357	88	0.304
9	0.969	49	-0.271	89	0.270
10	0.576	50	-0.611	90	-0.042
11	-0.482	51	-0.417		
12	-0.890	52	-1.216		
13	-0.578	53	0.033		
14	0.494	54	0.135		
15	0.154	55	0.168		
16	-0.436	56	0.566		
17	-0.196	57	0.241		
18	0.130	58	-0.159		
19	0.203	59	-0.041		
20	-0.655	60	0.491		
21	-0.616	61	0.489		
22	-0.093	62	1.059		
23	-0.237	63	-0.623		
24	0.028	64	1.091		
25	0.164	65	0.225		
26	-0.321	66	0.241		
27	-0.068	67	0.117		
28	-0.120	68	0.709		
29	0.118	69	0.076		
30	0.351	70	0.491		
31	0.198	71	0.089		
32	-0.175	72	-0.618		
33	0.015	73	-0.076		
34	0.210	74	-0.058		
35	0.501	75	0.069		
36	0.205	76	-0.407		
37	0.013	77	-0.229		
38	0.349	78	0.446		
39	0.488	79	0.078		
40	0.065	80	0.025		

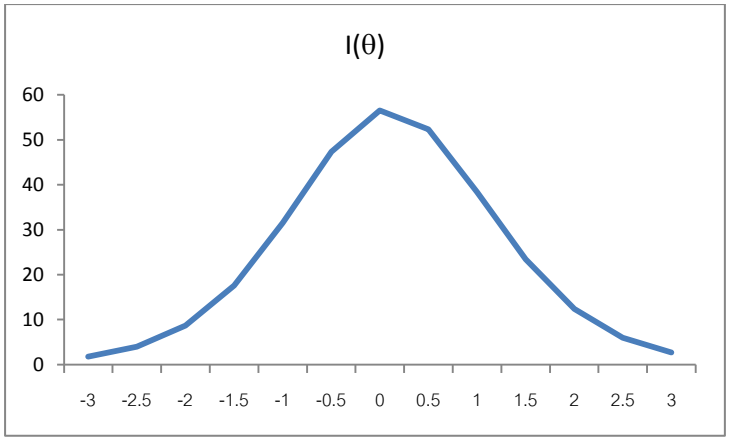
ตัวอย่างค่า Test Information ของแบบสอบวิชาภาษาไทย 90 ข้อ จากผู้สอบ 1,000 คน (โมเดล RPFI)

	$I_i(\theta)$												
Item	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3
th1	0.01	0.02	0.05	0.12	0.24	0.45	0.66	0.71	0.54	0.31	0.15	0.07	0.03
th2	0.00	0.00	0.01	0.02	0.04	0.09	0.20	0.39	0.62	0.72	0.60	0.37	0.19
th3	0.00	0.01	0.02	0.04	0.08	0.18	0.36	0.59	0.72	0.62	0.40	0.21	0.10
th4	0.00	0.01	0.01	0.03	0.07	0.15	0.31	0.54	0.71	0.67	0.45	0.24	0.12
th5	0.00	0.00	0.01	0.03	0.06	0.13	0.27	0.48	0.69	0.70	0.50	0.28	0.14
th6	0.00	0.01	0.02	0.04	0.08	0.18	0.36	0.59	0.72	0.62	0.39	0.21	0.10
th7	0.01	0.01	0.03	0.06	0.14	0.29	0.51	0.70	0.68	0.48	0.27	0.13	0.06
th8	0.01	0.02	0.04	0.09	0.20	0.39	0.61	0.72	0.60	0.37	0.19	0.09	0.04
th9	0.00	0.01	0.02	0.04	0.09	0.20	0.39	0.62	0.72	0.59	0.36	0.19	0.09
th10	0.01	0.01	0.03	0.08	0.17	0.34	0.57	0.72	0.64	0.42	0.22	0.10	0.05
th11	0.04	0.09	0.19	0.37	0.60	0.72	0.61	0.38	0.20	0.09	0.04	0.02	0.01
th12	0.07	0.14	0.29	0.52	0.70	0.68	0.47	0.26	0.12	0.06	0.02	0.01	0.00
th13	0.05	0.10	0.22	0.41	0.63	0.72	0.58	0.35	0.17	0.08	0.04	0.02	0.01
th14	0.01	0.02	0.04	0.10	0.21	0.41	0.63	0.72	0.58	0.35	0.18	0.08	0.04
th15	0.01	0.03	0.07	0.16	0.32	0.55	0.71	0.66	0.44	0.23	0.11	0.05	0.02
th16	0.03	0.07	0.16	0.32	0.55	0.71	0.66	0.44	0.23	0.11	0.05	0.02	0.01
th17	0.03	0.06	0.13	0.27	0.48	0.68	0.70	0.51	0.29	0.14	0.06	0.03	0.01
th18	0.01	0.03	0.07	0.16	0.31	0.54	0.71	0.66	0.45	0.24	0.11	0.05	0.02
th19	0.01	0.03	0.07	0.15	0.31	0.53	0.71	0.67	0.46	0.25	0.12	0.05	0.02
th20	0.05	0.11	0.23	0.43	0.65	0.71	0.55	0.32	0.16	0.07	0.03	0.01	0.01
th21	0.05	0.11	0.23	0.43	0.65	0.71	0.55	0.32	0.16	0.07	0.03	0.01	0.01
th22	0.02	0.05	0.11	0.23	0.43	0.65	0.72	0.56	0.33	0.16	0.08	0.03	0.01
th23	0.03	0.06	0.14	0.28	0.50	0.70	0.69	0.48	0.27	0.13	0.06	0.03	0.01
th24	0.02	0.04	0.09	0.20	0.39	0.62	0.72	0.59	0.37	0.19	0.09	0.04	0.02
th25	0.01	0.03	0.07	0.15	0.31	0.53	0.71	0.67	0.45	0.25	0.12	0.05	0.02
th26	0.03	0.06	0.13	0.27	0.49	0.69	0.69	0.50	0.28	0.14	0.06	0.03	0.01
th27	0.02	0.04	0.10	0.21	0.40	0.63	0.72	0.59	0.36	0.18	0.08	0.04	0.02

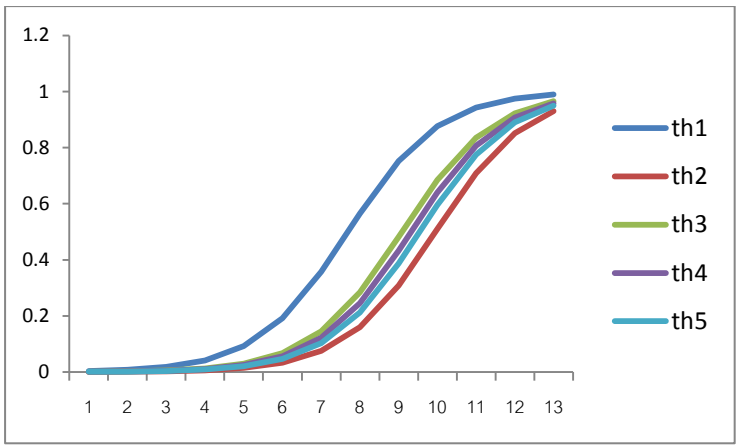
	$I_i(\theta)$												
Item	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3
th28	0.02	0.05	0.11	0.23	0.43	0.65	0.72	0.56	0.33	0.16	0.07	0.03	0.01
th29	0.02	0.03	0.08	0.17	0.34	0.57	0.72	0.64	0.41	0.22	0.10	0.05	0.02
th30	0.01	0.02	0.05	0.12	0.24	0.45	0.66	0.71	0.54	0.31	0.15	0.07	0.03
th31	0.01	0.03	0.07	0.15	0.30	0.53	0.71	0.67	0.46	0.25	0.12	0.05	0.02
th32	0.02	0.06	0.13	0.26	0.47	0.68	0.70	0.51	0.29	0.14	0.06	0.03	0.01
th33	0.02	0.04	0.09	0.19	0.37	0.60	0.72	0.61	0.39	0.20	0.09	0.04	0.02
th34	0.01	0.03	0.06	0.13	0.28	0.50	0.69	0.69	0.49	0.27	0.13	0.06	0.03
th35	0.01	0.01	0.03	0.07	0.16	0.33	0.55	0.72	0.65	0.43	0.23	0.11	0.05
th36	0.01	0.03	0.07	0.15	0.30	0.52	0.70	0.67	0.47	0.25	0.12	0.05	0.02
th37	0.02	0.04	0.10	0.21	0.40	0.63	0.72	0.58	0.35	0.18	0.08	0.04	0.02
th38	0.01	0.02	0.06	0.12	0.26	0.47	0.68	0.70	0.51	0.29	0.14	0.06	0.03
th39	0.01	0.02	0.04	0.10	0.21	0.41	0.63	0.72	0.58	0.35	0.18	0.08	0.04
th40	0.02	0.04	0.08	0.17	0.34	0.57	0.72	0.64	0.41	0.22	0.10	0.05	0.02
th41	0.02	0.05	0.12	0.25	0.46	0.67	0.71	0.53	0.30	0.15	0.07	0.03	0.01
th42	0.02	0.04	0.08	0.18	0.36	0.59	0.72	0.62	0.40	0.21	0.10	0.04	0.02
th43	0.01	0.03	0.07	0.15	0.31	0.54	0.71	0.66	0.45	0.24	0.12	0.05	0.02
th44	0.02	0.04	0.10	0.21	0.40	0.63	0.72	0.58	0.35	0.18	0.08	0.04	0.02
th45	0.02	0.05	0.11	0.24	0.44	0.66	0.71	0.55	0.32	0.16	0.07	0.03	0.01
th46	0.05	0.11	0.22	0.42	0.64	0.72	0.56	0.33	0.17	0.08	0.03	0.01	0.01
th47	0.06	0.13	0.27	0.49	0.69	0.69	0.50	0.28	0.14	0.06	0.03	0.01	0.01
th48	0.01	0.03	0.06	0.13	0.28	0.49	0.69	0.69	0.49	0.27	0.13	0.06	0.03
th49	0.03	0.07	0.14	0.29	0.52	0.70	0.68	0.47	0.26	0.12	0.06	0.02	0.01
th50	0.05	0.12	0.25	0.46	0.67	0.71	0.53	0.30	0.15	0.07	0.03	0.01	0.01
th51	0.03	0.07	0.16	0.32	0.54	0.71	0.66	0.44	0.24	0.11	0.05	0.02	0.01
th52	0.12	0.25	0.46	0.67	0.70	0.52	0.30	0.15	0.07	0.03	0.01	0.01	0.00
th53	0.02	0.04	0.08	0.18	0.36	0.59	0.72	0.62	0.40	0.21	0.10	0.04	0.02
th54	0.01	0.03	0.07	0.16	0.32	0.54	0.71	0.66	0.44	0.24	0.11	0.05	0.02
th55	0.01	0.03	0.07	0.15	0.31	0.54	0.71	0.66	0.45	0.24	0.12	0.05	0.02

	$I_i(\theta)$												
Item	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3
th56	0.01	0.02	0.04	0.09	0.20	0.39	0.62	0.72	0.60	0.37	0.19	0.09	0.04
th57	0.01	0.03	0.06	0.13	0.28	0.49	0.69	0.69	0.49	0.28	0.13	0.06	0.03
th58	0.03	0.06	0.13	0.27	0.49	0.69	0.69	0.50	0.28	0.14	0.06	0.03	0.01
th59	0.02	0.04	0.10	0.21	0.41	0.63	0.72	0.58	0.35	0.18	0.08	0.04	0.02
th60	0.01	0.02	0.04	0.10	0.20	0.39	0.62	0.72	0.59	0.36	0.18	0.09	0.04
th61	0.01	0.02	0.04	0.09	0.20	0.39	0.62	0.72	0.59	0.36	0.19	0.09	0.04
th62	0.00	0.01	0.02	0.03	0.08	0.17	0.34	0.57	0.72	0.64	0.42	0.22	0.10
th63	0.05	0.12	0.25	0.46	0.67	0.71	0.53	0.30	0.15	0.07	0.03	0.01	0.01
th64	0.00	0.01	0.01	0.03	0.07	0.16	0.32	0.55	0.71	0.66	0.44	0.24	0.11
th65	0.01	0.03	0.07	0.15	0.30	0.52	0.70	0.67	0.46	0.25	0.12	0.05	0.02
th66	0.01	0.03	0.06	0.14	0.28	0.50	0.69	0.69	0.49	0.27	0.13	0.06	0.03
th67	0.02	0.04	0.09	0.19	0.36	0.59	0.72	0.62	0.39	0.20	0.09	0.04	0.02
th68	0.01	0.01	0.03	0.07	0.16	0.33	0.55	0.71	0.65	0.43	0.23	0.11	0.05
th69	0.02	0.04	0.09	0.19	0.37	0.60	0.72	0.62	0.39	0.20	0.09	0.04	0.02
th70	0.01	0.02	0.04	0.09	0.19	0.38	0.61	0.72	0.61	0.38	0.19	0.09	0.04
th71	0.02	0.04	0.09	0.19	0.37	0.60	0.72	0.62	0.39	0.20	0.09	0.04	0.02
th72	0.05	0.12	0.25	0.45	0.67	0.71	0.53	0.31	0.15	0.07	0.03	0.01	0.01
th73	0.02	0.05	0.10	0.22	0.41	0.64	0.72	0.57	0.34	0.17	0.08	0.04	0.02
th74	0.02	0.05	0.10	0.22	0.41	0.64	0.72	0.57	0.34	0.17	0.08	0.04	0.02
th75	0.01	0.03	0.07	0.15	0.31	0.53	0.71	0.67	0.45	0.25	0.12	0.05	0.02
th76	0.04	0.08	0.18	0.36	0.58	0.72	0.63	0.40	0.21	0.10	0.04	0.02	0.01
th77	0.03	0.06	0.14	0.28	0.50	0.69	0.69	0.49	0.27	0.13	0.06	0.03	0.01
th78	0.01	0.02	0.05	0.11	0.24	0.44	0.66	0.71	0.55	0.32	0.16	0.07	0.03
th79	0.02	0.04	0.08	0.17	0.35	0.57	0.72	0.64	0.41	0.22	0.10	0.05	0.02
th80	0.02	0.04	0.10	0.21	0.41	0.63	0.72	0.58	0.35	0.18	0.08	0.04	0.02
th81	0.01	0.01	0.03	0.07	0.16	0.32	0.54	0.71	0.66	0.44	0.24	0.11	0.05
th82	0.01	0.02	0.06	0.12	0.26	0.47	0.68	0.70	0.52	0.29	0.14	0.07	0.03
th83	0.02	0.05	0.11	0.23	0.43	0.65	0.71	0.55	0.33	0.16	0.07	0.03	0.01

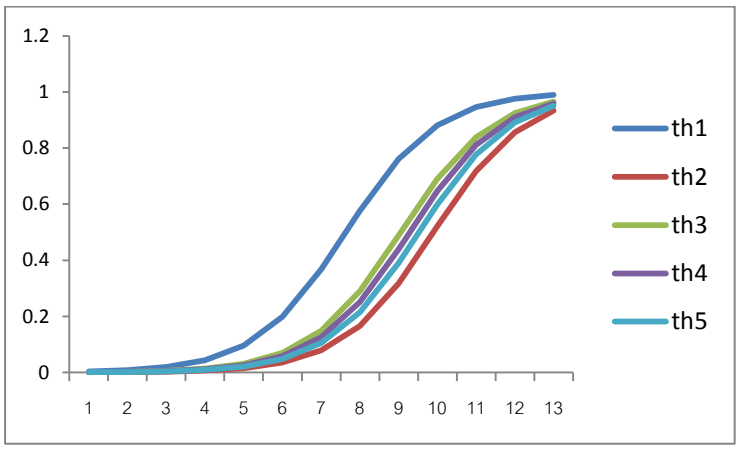
	$I_i(\theta)$												
Item	-3	-2.5	-2	-1.5	-1	-0.5	0	0.5	1	1.5	2	2.5	3
th84	0.01	0.01	0.03	0.07	0.14	0.29	0.52	0.70	0.68	0.47	0.26	0.12	0.06
th85	0.01	0.03	0.07	0.16	0.31	0.54	0.71	0.66	0.45	0.24	0.11	0.05	0.02
th86	0.01	0.03	0.08	0.17	0.33	0.56	0.72	0.65	0.43	0.23	0.11	0.05	0.02
th87	0.01	0.02	0.05	0.12	0.25	0.46	0.67	0.71	0.53	0.30	0.15	0.07	0.03
th88	0.01	0.02	0.05	0.11	0.23	0.44	0.66	0.71	0.55	0.32	0.16	0.07	0.03
th89	0.01	0.03	0.06	0.14	0.28	0.50	0.70	0.69	0.48	0.27	0.13	0.06	0.03
th90	0.02	0.04	0.10	0.21	0.40	0.62	0.72	0.59	0.36	0.18	0.08	0.04	0.02
$I(\theta)$	1.76	3.99	8.69	17.60	31.56	47.35	56.51	52.31	38.36	23.39	12.41	5.96	2.70
$SE(\theta)$	0.75	0.50	0.34	0.24	0.18	0.15	0.13	0.14	0.16	0.21	0.28	0.41	0.61



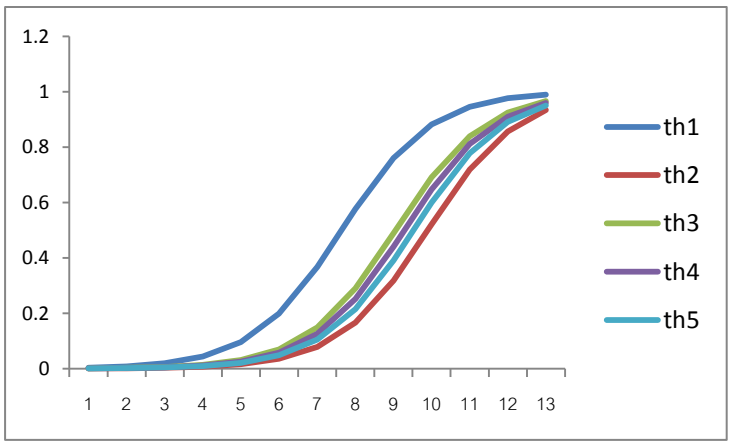
ตัวอย่าง Test Information วิชาภาษาไทย 90 ข้อ (โมเดล RPFI)



ตัวอย่างกราฟ ICC ข้อสอบวิชาภาษาไทย 5 ข้อ (โมเดล RPFI)



ตัวอย่างกราฟ ICC ข้อสอบวิชาภาษาไทย 5 ข้อ (โมเดล FPRI)



ตัวอย่างกราฟ ICC ข้อสอบวิชาภาษาไทย 5 ข้อ (โมเดล RPRI)

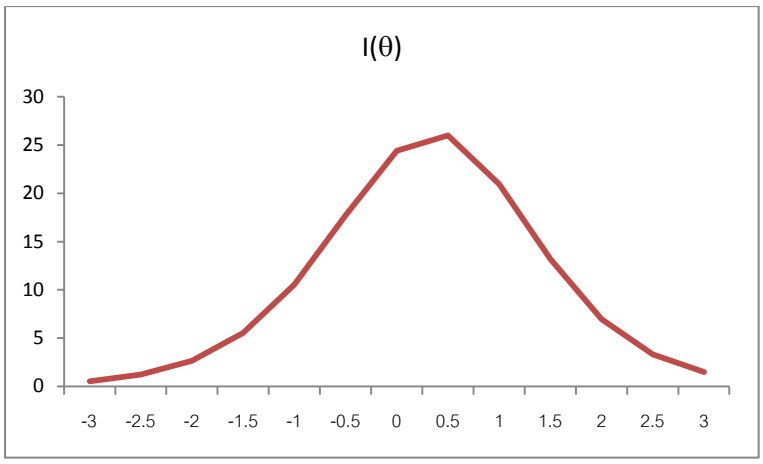
ค่าความยากวิชาคณิตศาสตร์ 40 ข้อ

ข้อ	Beta	ข้อ	Beta
1	0.05	21	0.41
2	-0.21	22	0.47
3	0.76	23	0.60
4	0.70	24	0.47
5	0.28	25	0.75
6	0.10	26	0.57
7	-0.09	27	0.85
8	-1.13	28	0.60
9	0.32	29	0.49
10	0.12	30	1.10
11	0.12	31	0.20
12	0.01	32	0.22
13	0.59	33	0.57
14	0.80	34	0.56
15	0.60	35	0.03
16	-0.20	36	0.61
17	-0.18	37	0.57
18	0.02	38	1.13
19	0.27	39	0.37
20	-0.26	40	0.27

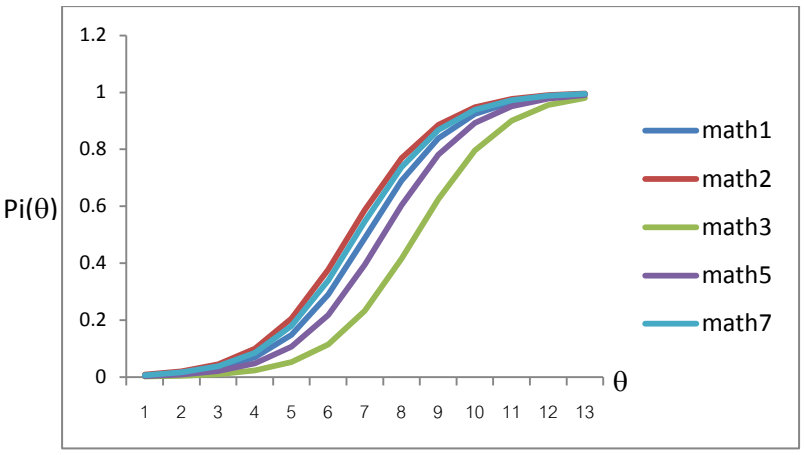
ตัวอย่างค่า Test Information ของแบบสอบวิชาคณิตศาสตร์ 40 ข้อ จากผู้สอบ 1,000 คน (โมเดล RPFI)

	$I_i(\theta)$												
item	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0
math1	0.02	0.04	0.09	0.19	0.37	0.59	0.72	0.62	0.39	0.20	0.09	0.04	0.02
math2	0.02	0.06	0.12	0.26	0.47	0.68	0.70	0.51	0.29	0.14	0.06	0.03	0.01
math3	0.01	0.01	0.03	0.07	0.14	0.29	0.52	0.70	0.68	0.47	0.26	0.12	0.06
math4	0.01	0.01	0.03	0.07	0.15	0.31	0.54	0.71	0.66	0.45	0.24	0.12	0.05
math5	0.01	0.03	0.06	0.13	0.27	0.49	0.69	0.69	0.49	0.28	0.13	0.06	0.03
math6	0.02	0.03	0.08	0.17	0.34	0.57	0.72	0.64	0.42	0.22	0.10	0.05	0.02
math7	0.02	0.05	0.11	0.23	0.43	0.65	0.72	0.56	0.33	0.17	0.08	0.03	0.01
math8	0.10	0.22	0.42	0.64	0.72	0.57	0.34	0.17	0.08	0.03	0.02	0.01	0.00
math9	0.01	0.02	0.06	0.12	0.26	0.47	0.68	0.70	0.52	0.30	0.15	0.07	0.03
math10	0.01	0.03	0.08	0.17	0.34	0.57	0.72	0.64	0.42	0.22	0.10	0.05	0.02
math11	0.01	0.03	0.07	0.16	0.32	0.54	0.71	0.66	0.44	0.24	0.11	0.05	0.02

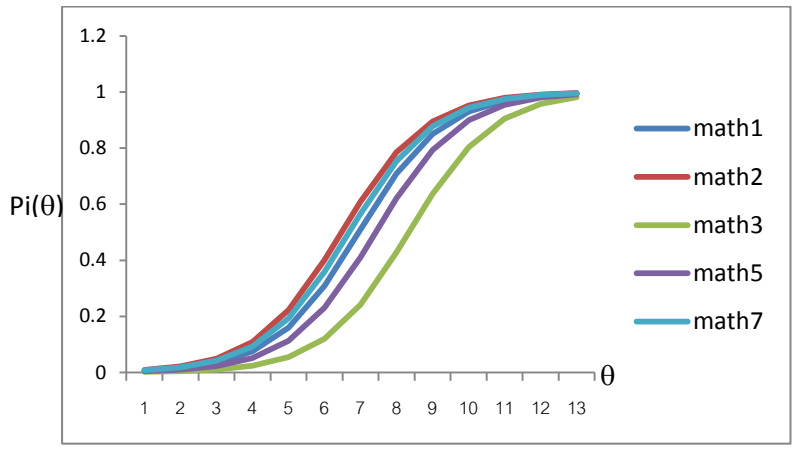
	$I_i(\theta)$												
item	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0
math12	0.02	0.04	0.09	0.19	0.37	0.60	0.72	0.62	0.39	0.20	0.09	0.04	0.02
math13	0.01	0.01	0.03	0.08	0.17	0.33	0.56	0.72	0.65	0.42	0.22	0.11	0.05
math14	0.00	0.01	0.03	0.06	0.13	0.27	0.48	0.68	0.70	0.51	0.29	0.14	0.06
math15	0.01	0.01	0.03	0.07	0.15	0.30	0.53	0.71	0.67	0.46	0.25	0.12	0.05
math16	0.02	0.05	0.11	0.23	0.44	0.66	0.71	0.55	0.32	0.16	0.07	0.03	0.01
math17	0.02	0.05	0.12	0.25	0.46	0.67	0.71	0.52	0.30	0.15	0.07	0.03	0.01
math18	0.02	0.04	0.08	0.17	0.35	0.57	0.72	0.64	0.41	0.22	0.10	0.05	0.02
math19	0.01	0.03	0.06	0.13	0.27	0.49	0.69	0.69	0.50	0.28	0.14	0.06	0.03
math20	0.03	0.07	0.15	0.30	0.53	0.71	0.67	0.46	0.25	0.12	0.05	0.02	0.01
math21	0.01	0.02	0.04	0.10	0.21	0.40	0.63	0.72	0.58	0.35	0.18	0.08	0.04
math22	0.01	0.02	0.04	0.09	0.19	0.37	0.60	0.72	0.61	0.38	0.20	0.09	0.04
math23	0.01	0.01	0.03	0.07	0.16	0.33	0.55	0.72	0.65	0.43	0.23	0.11	0.05
math24	0.01	0.02	0.04	0.09	0.19	0.38	0.61	0.72	0.61	0.38	0.19	0.09	0.04
math25	0.01	0.01	0.03	0.07	0.15	0.30	0.53	0.71	0.67	0.46	0.25	0.12	0.05
math26	0.01	0.02	0.04	0.08	0.17	0.34	0.57	0.72	0.64	0.41	0.22	0.10	0.05
math27	0.00	0.01	0.03	0.06	0.13	0.27	0.49	0.69	0.69	0.50	0.28	0.14	0.06
math28	0.01	0.02	0.04	0.08	0.18	0.35	0.58	0.72	0.63	0.40	0.21	0.10	0.04
math29	0.01	0.02	0.04	0.09	0.20	0.38	0.61	0.72	0.60	0.37	0.19	0.09	0.04
math30	0.00	0.01	0.02	0.03	0.08	0.17	0.34	0.57	0.72	0.64	0.42	0.22	0.10
math31	0.01	0.03	0.06	0.13	0.27	0.49	0.69	0.69	0.50	0.28	0.14	0.06	0.03
math32	0.01	0.03	0.06	0.14	0.28	0.50	0.70	0.69	0.48	0.27	0.13	0.06	0.03
math33	0.01	0.02	0.04	0.08	0.18	0.35	0.58	0.72	0.63	0.41	0.21	0.10	0.04
math34	0.01	0.01	0.03	0.08	0.17	0.33	0.56	0.72	0.65	0.43	0.23	0.11	0.05
math35	0.02	0.04	0.09	0.20	0.39	0.62	0.72	0.60	0.37	0.19	0.09	0.04	0.02
math36	0.01	0.01	0.03	0.07	0.16	0.32	0.55	0.71	0.65	0.44	0.23	0.11	0.05
math37	0.01	0.02	0.04	0.09	0.19	0.38	0.60	0.72	0.61	0.38	0.20	0.09	0.04
math38	0.00	0.01	0.01	0.03	0.07	0.15	0.30	0.53	0.71	0.67	0.46	0.25	0.12
math39	0.01	0.02	0.05	0.11	0.24	0.44	0.66	0.71	0.54	0.32	0.16	0.07	0.03
math40	0.01	0.03	0.06	0.13	0.27	0.48	0.69	0.70	0.50	0.28	0.14	0.06	0.03
$I(\theta)$	0.53	1.21	2.65	5.52	10.54	17.70	24.40	25.99	20.96	13.21	6.97	3.31	1.48
$SE(\theta)$	1.37	0.91	0.61	0.43	0.31	0.24	0.20	0.20	0.22	0.28	0.38	0.55	0.82



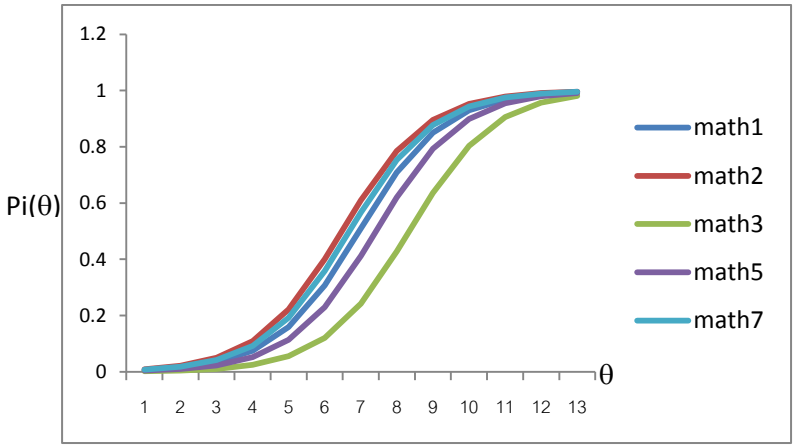
ตัวอย่าง Test Information วิชาคณิตศาสตร์ 40 ข้อ (โมเดล RPFI)



ตัวอย่างกราฟ ICC ข้อสอบวิชาคณิตศาสตร์ 5 ข้อ (โมเดล RPFI)



ตัวอย่างกราฟ ICC ข้อสอบวิชาคณิตศาสตร์ 5 ข้อ (โมเดล FPRI)



ตัวอย่างกราฟ ICC ข้อสอบวิชาคณิตศาสตร์ 5 ข้อ (โมเดล RPRI)

ประวัติผู้เขียนวิทยานิพนธ์

นางทิพวัลย์ ปัญจมะวัต เกิดเมื่อวันที่ 14 พฤษภาคม 2512 สำเร็จการศึกษา วิทยาศาสตรบัณฑิต สาขาเทคโนโลยีภาพถ่ายและการพิมพ์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2538 จากนั้นเข้าทำงานที่ บริษัท อิมเมจควอลิตี้แล็บ จำกัด ตำแหน่ง หัวหน้างานควบคุมคุณภาพ ตั้งแต่ระหว่างการทำงานได้สำเร็จการศึกษาศิลปศาสตรบัณฑิต สาขาสารสนเทศทั่วไป มหาวิทยาลัยสุโขทัยธรรมาธิราช เมื่อปีการศึกษา 2542 จากนั้นได้ศึกษาต่อและสำเร็จการศึกษาครุศาสตรมหาบัณฑิต สาขาวิชาวิจัยการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษาคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย เมื่อปีการศึกษา 2548 และระหว่างปี 2549-2550 ได้เข้าทำงานในสำนักงานรับรองมาตรฐานและประเมินคุณภาพการศึกษา (องค์การมหาชน) และในปีการศึกษา 2551 เข้าศึกษาต่อในหลักสูตรครุศาสตร ดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษาคณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย