การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทางด้วยโมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น: การประมาณค่าพารามิเตอร์

นางสาวชลี ภัทรพิชญธรรม

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2553
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

TWO-WAY DIFFERENTIAL ITEM FUNCTIONING ANALYSIS USING HIERARCHICAL GENERALIZED LINEAR MODEL: A PARAMETER ESTIMATION

MISS CHALIE PATARAPICHAYATHAM

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

หัวข้อวิทยานิพนธ์	การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง ด้วยโมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น: การประมาณค่าพารามิเตอร์
โดย	นางสาวชลี ภัทรพิชญธรรม
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก	ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี
อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม	Professor Akihito Kamata, Ph.D

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

.....คณบดีคณะครุศาสตร์
(ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

คณะกรรมการสอบวิทยานิพนธ์

.....ประธานกรรมการ
(รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ)

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

....อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม
(Professor Akihito Kamata, Ph.D)

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

.....กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.ชูศักดิ์ ขัมภลิขิต)

ชลี ภัทรพิชญธรรม: การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทางด้วยโมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น: การประมาณค่าพารามิเตอร์ (TWO-WAY DIFFERENTIAL ITEM FUNCTIONING ANALYSIS USING HIERARCHICAL GENERALIZED LINEAR MODEL: A PARAMETER ESTIMATION) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: ศ.ดร.ศิริชัย กาญจนวาสี, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม: Professor Akihito Kamata, Ph.D, 306 หน้า

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มและเพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) โดยการจำลองข้อมูลด้วยเงื่อนไขที่แตกต่างกัน 18 เงื่อนไข (2 x 3 x 3) นั่นคือ 1) จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) 2 ระดับ คือ โรงเรียนละ 50 คนและโรงเรียนละ 100 คน 2) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 3 ระดับ คือ 0.2, 0.4 และ 0.6 และ 3) ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ คือ 0.1, 0.2 และ 0.3 มีโมเดลการทำหน้าที่ต่างกันของข้อสอบ 4 โมเดลประกอบไปด้วย 1) โมเดลเต็มรูป 2) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง 3) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม และ 4) โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม และกำหนดจำนวนการทำซ้ำในแต่ละเงื่อนไขจำนวน 200 รอบ ผู้วิจัยมีเกณฑ์การคัดสรรโมเดล 5 เกณฑ์ ได้แก่ ดัชนี AIC, ดัชนี BIC, ดัชนี ABIC, ดัชนี 2 of 3 และดัชนี p-value และมีเกณฑ์การประเมินในการประมาณค่าพารามิเตอร์ 3 เกณฑ์ คือ ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) โดยจำลองข้อมูลด้วยโปรแกรม R ประมาณค่าพารามิเตอร์ด้วยโปรแกรม Mplus

ผลการศึกษาสรุปได้ดังนี้

1. ค่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริงในทุกเงื่อนไข

2. เมื่อใช้ดัชนี AIC และดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล พบว่า โมเดลเต็มรูปเป็นโมเดลที่ดีที่สุด ในขณะที่เมื่อใช้ดัชนี BIC ดัชนี ABIC และดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดลได้ผลที่สอดคล้องกันคือ โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม เป็นโมเดลที่ดีที่สุดในทุกเงื่อนไข

ข้อเสนอแนะ: นักวิจัยควรเลือกใช้โมเดลเต็มรูป เมื่อต้องการศึกษาทุกพารามิเตอร์ในโมเดล HGLM หรือเลือกใช้โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม เมื่อต้องการศึกษาด้วยโมเดลที่ประหยัดกว่า

ภาควิชา วิจัยและจิตวิทยาการศึกษา ลายมือชื่อนิสิต .....

สาขาวิชา การวัดและประเมินผลการศึกษา ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2553 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม .....

5084210727: MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEYWORDS: TWO-WAY DIFFERENTIAL ITEM FUNCTIONING (TWO-WAY DIF) /PARAMETER ESTIMATION / HIERARCHICAL GENERALIZED LINEAR MODELING (HGLM)

CHALIE PATARAPICHAYATHAM: TWO-WAY DIFFERENTIAL ITEM FUNCTIONING ANALYSIS USING HIERARCHICAL GENERALIZED LINEAR MODEL: A PARAMETER ESTIMATION. THESIS ADVISOR: PROF. SIRICHAI KANJANAWASEE, Ph.D., THESIS CO-ADVISOR: PROF. AKIHITO KAMATA, Ph.D., 306 pp.

The objectives of this study were to 1) investigate the quality of parameter estimates for the two-way differential item functioning (two-way DIF) model, when one of the DIF factors was a group characteristics variable, and 2) compare the efficiency of 4 differential item functioning models, which one was the best model in order to investigate the quality of parameter estimates. This study was based on the different 18 simulation conditions (2 x 3 x 3) consisted of 1) the number of student in each school (50, 100) 2) the cluster-level DIF (0.2, 0.4, 0.6), and 3) the two-way interaction (0.1, 0.2, 0.3). 4 differential item functioning models were investigated: 1) the complete model, 2) the incomplete model without two-way interaction, 3) the incomplete model without cluster-level DIF, and 4) the incomplete model without both two-way interaction and cluster-level DIF. The simulation was replicated 200 times for each condition. 5 model selection criteria were used in this simulation study: AIC (Akaike's information criterion), BIC (Bayesian information criterion), ABIC (Sample-Size Adjusted Bayesian information criterion), 2 of 3, and p-value. Also, 3 evaluation criteria were examined: BIAS, SE, and RMSE. R software was used to generate the data, and fitted the model by Mplus.

The results were as follows:

1) The parameter estimate of the three-way interaction (3WAYINT) and the individual-level DIF (ILDIF) were different from the true value across conditions.

2) The complete model was the best model when AIC and p-value were selected as the model selection criteria. However, the incomplete model without both two-way interaction and cluster-level DIF was the best model when BIC, ABIC, and 2 of 3 were assumed as the model selection criteria.

Suggesstion: Researches should use the complete model in case they would like to study all effects of HGLM moldel. On the other hand, they might use the incomplete model without both two-way interaction and cluster-level DIF which is simplier.

Department: Educational Research and Psychology Student's Signature .....
Field of Study: Educational Measurement and Evaluation Advisor's Signature .....
Academic Year: 2010 Co-Advisor's Signature .....

กิตติกรรมประกาศ

การทำวิทยานิพนธ์ระดับดุษฎีบัณฑิตในครั้งนี้สำเร็จลุล่วงไปได้ด้วยดี เนื่องจากได้รับความเมตตาเป็นอย่างยิ่งจาก ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี อาจารย์ที่ปรึกษาวิทยานิพนธ์หลักที่ดูแลอย่างใกล้ชิดตลอดระยะเวลาที่ศึกษาในระดับดุษฎีบัณฑิตทั้งในด้านวิชาการ การทำงานที่เป็นระบบและการดำเนินชีวิตที่มีคุณค่า รวมไปถึงการให้โอกาสผู้วิจัยในการช่วยงานของศูนย์ทดสอบทางการศึกษา คณะครุศาสตร์ ซึ่งผู้วิจัยได้รับความรู้และทักษะที่เป็นประโยชน์เป็นอย่างยิ่ง

ขอขอบพระคุณ Professor Akihito Kamata, Ph.D ในฐานะอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ซึ่งเป็นทั้งอาจารย์ที่ปรึกษา เป็นทั้ง Host family เป็นทั้งญาติและเป็นทั้งเพื่อน กว่า 1 ปี ที่ผู้วิจัยพัฒนาโครงร่างวิทยานิพนธ์ ผู้วิจัยได้รับความช่วยเหลืออย่างดียิ่ง โดยการสื่อสารทางอีเมล์ทุกสัปดาห์ ทำให้การพัฒนาโครงร่างวิทยานิพนธ์มีความก้าวหน้าและสามารถสอบโครงร่างได้ตามระยะเวลาที่กำหนด ผู้วิจัยได้รับความช่วยเหลือเป็นอย่างดียิ่งตลอดระยะเวลาที่ศึกษาและทำวิทยานิพนธ์ ณ ประเทศสหรัฐอเมริกาทั้งเรื่องวิชาการและชีวิตความเป็นอยู่ประหนึ่งผู้วิจัยเป็นคนในครอบครัว จากการทำงานร่วมกันเกือบทุกวัน ทำให้ผู้วิจัยได้เรียนรู้ถึงคุณค่าของการทำงานหนัก ความอดทนและไม่ยอมแพ้ แนวทางในการพัฒนาตัวเองอย่างเป็นระบบ วิธีคิด วิธีแก้ปัญหาและแนวทางในการดำเนินชีวิตที่มุ่งสร้างคุณค่าและองค์ความรู้ใหม่แก่สังคม

ขอขอบพระคุณสำนักงานคณะกรรมการอุดมศึกษา (สกอ) ที่ให้ทุนการศึกษาตลอดหลักสูตรดุษฎีบัณฑิตและให้โอกาสผู้วิจัยไปทำวิทยานิพนธ์และศึกษาค้นคว้าเพิ่มเติม ณ Department of Educational Methodology, Policy, and Leadership (EMPL), College of Education, University of Oregon ประเทศสหรัฐอเมริกา ผู้วิจัยขอขอบพระคุณคณบดี คณาจารย์และเจ้าหน้าที่ EMPL ที่เมตตาผู้วิจัยเป็นอย่างยิ่ง รวมไปถึงเพื่อนนักเรียนไทยและเพื่อนชาวต่างชาติ

ขอขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ อันได้แก่ รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ รองศาสตราจารย์ ดร. โชติกา ภาษีผลและอาจารย์ ดร. ชูศักดิ์ ขำภลิขิต ที่กรุณาชี้แนะประเด็นที่เป็นประโยชน์ในการทำวิทยานิพนธ์ในครั้งนี้

ขอขอบพระคุณคณาจารย์ประจำภาควิชาวิจัยและจิตวิทยาการศึกษา รวมไปถึง พี่ๆ เพื่อนๆ น้องๆ สาขาการวัดและประเมินผลการศึกษาที่คอยให้คำแนะนำที่เป็นประโยชน์และเป็นกำลังใจที่ดีตลอดระยะเวลาการศึกษา รวมไปถึงคุณกาญจนา ตลึงจิตร คุณชัยวิชิต เชียรชนะ คุณกิติยา ภมรคล คุณสุพัตรา ชะมะบูรณ์ คุณวัชราภรณ์ เกียรติบุญญาฤทธิ์ คุณศักดิ์สิทธิ์ ฤทธิลัน คุณวราพร เอราวรรณ์ คุณสาธิดา สกุลรัตนกุลชัย คุณชนะศึก นิชานนท์และคุณทัศน์สิรินทร์ สว่างบุญ ผู้เป็นมหามิตรที่คอยดูแลและช่วยเหลือผู้วิจัยมาโดยตลอด

สุดท้ายนี้ผู้วิจัยขอบคุณพ่อแม่ น้องสาว หลาน ญาติๆ และคุณอิทธิพัทธ์ ตระการสิริพล ที่คอยดูแลชีวิตความเป็นอยู่และเป็นกำลังใจที่ยิ่งใหญ่สำหรับผู้วิจัยเสมอมาจนผู้วิจัยสำเร็จการศึกษาดังที่ตั้งใจไว้

สารบัญ

		หน้า
บทคัดย่อภาษาไทย		ง
บทคัดย่อภาษาอังกฤษ		จ
กิตติกรรมประกาศ		ฉ
สารบัญ		ช
สารบัญตาราง		ฌ
สารบัญแผนภาพ		ญ
บทที่		
1	บทนำ	1
	ความเป็นมาและความสำคัญของปัญหาวิจัย	1
	คำถามการวิจัย	8
	วัตถุประสงค์การวิจัย	8
	ขอบเขตของการวิจัย	9
	ข้อจำกัดของการวิจัย	10
	นิยามคำศัพท์	10
	ประโยชน์ที่คาดว่าจะได้รับ	16
2	เอกสารและงานวิจัยที่เกี่ยวข้อง	18
	ตอนที่ 1 มโนทัศน์ของทฤษฎีการตอบสนองข้อสอบ	18
	ตอนที่ 2 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ	27
	ตอนที่ 3 มโนทัศน์ของการวิเคราะห์พหุระดับ	39
	ตอนที่ 4 โมเดลเชิงเส้นตรงทั่วไปแบบลดหลั่น (HGLM) สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า	46
	ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง	55
	กรอบแนวคิดในการวิจัย	87
3	วิธีดำเนินการวิจัย	88
	การศึกษานำร่องและผลการศึกษานำร่อง	88
	วิธีดำเนินการวิจัย	101

สารบัญ (ต่อ)

บทที่		หน้า
3	เงื่อนไขในการจำลองข้อมูล……..	101
	โมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง……..	103
	เกณฑ์การคัดสรรโมเดล…….	108
	พารามิเตอร์เป้าหมายในการประมาณค่า…….	112
	ขั้นตอนการจำลองข้อมูลและเกณฑ์การประเมิน……..	113
	การออกแบบการวิจัย…….	115
4	ผลการวิเคราะห์ข้อมูล……..	118
	ตอนที่ 1 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการประมาณค่าพารามิเตอร์จากทั้ง 4 โมเดล……..	120
	ตอนที่ 2 โมเดลที่ดีที่สุดของแต่ละการทำซ้ำเมื่อใช้เกณฑ์การคัดสรรโมเดลทั้ง 5 เกณฑ์…….	139
	ตอนที่ 3 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ที่สกัดจากโมเดลที่ดีที่สุด เมื่อใช้เกณฑ์การคัดสรรโมเดลทั้ง 5 เกณฑ์……..	186
	ตอนที่ 4 ค่าความลำเอียง ค่าความคลาดเคลื่อนมาตรฐานและค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ยของแต่ละพารามิเตอร์……..	207
5	สรุปผล อภิปรายผลและข้อเสนอแนะ……..	217
	สรุปผลการศึกษา…….	218
	อภิปรายผลการศึกษา…….	220
	ข้อเสนอแนะ…….	230
รายการอ้างอิง…….		236
ภาคผนวก…….		248
	ภาคผนวก ก คำสั่งในการจำลองข้อมูลและการวิเคราะห์ข้อมูล……..	249
	ภาคผนวก ข คำสั่งในการประมาณค่าพารามิเตอร์ของโปรแกรม Mplus…….	265
	ภาคผนวก ค ตัวอย่างผลการสกัดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานจากโมเดลที่ดีที่สุดของแต่ละเกณฑ์การประเมิน……..	270
ประวัติผู้เขียนวิทยานิพนธ์……..		306

สารบัญตาราง

หน้า

ตารางที่

2.1 การคำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ ค่าสารสนเทศสูงสุดของข้อสอบ และตำแหน่ง θ ที่มีสารสนเทศของข้อสอบสูงสุด สำหรับโมเดลการตอบสนองข้อสอบโลจิสแบบ 1, 2 และ 3 พารามิเตอร์..... 23

2.2 นิยามของการทำหน้าที่ต่างกันของข้อสอบ..... 30

2.3 วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 36

2.4 การเปรียบเทียบค่าความสามารถของผู้สอบและความยากของข้อสอบจากการประมาณค่าด้วยโมเดล HGLM แบบ 3 ระดับกับโมเดลราสช์..... 51

2.5 การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ..... 67

2.6 งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับด้วยโมเดล HGLM..... 81

3.1 ค่าความยากของข้อสอบแต่ละข้อ..... 97

3.2 ค่าจริงของค่าพารามิเตอร์แต่ละค่า..... 98

3.3 ค่าพารามิเตอร์ที่ได้จากการประมาณค่าด้วยโมเดลการทำหน้าทำต่างกันของข้อสอบแบบ 2 ทาง เมื่อเมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม..... 99

3.4 การจำลองข้อมูลตามเงื่อนไขจำนวนนักเรียนในแต่ละโรงเรียน ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ที่แตกต่างกัน..... 104

4.1 ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของการประมาณค่าพารามิเตอร์ก่อนการใช้เกณฑ์การคัดสรรโมเดล..... 121

4.2 โมเดลที่ดีที่สุดของแต่ละการทำซ้ำแต่ละครั้ง..... 139

4.3 โมเดลที่ดีที่สุดของแต่ละการทำซ้ำ เมื่อใช้เกณฑ์การคัดสรรโมเดลทั้ง 5..... 163

4.4 ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ที่สกัดจากโมเดลที่ดีที่สุดเมื่อใช้เกณฑ์การคัดสรรโมเดลทั้ง 5 เกณฑ์..... 186

4.5 ค่าความลำเอียง ค่าความคลาดเคลื่อนมาตรฐานและค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ยของแต่ละพารามิเตอร์..... 208

สารบัญแผนภาพ

แผนภาพที่		หน้า
2.1	การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป.....	31
2.2	การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป.....	32
2.3	ลักษณะโครงสร้างของข้อมูลที่จะนำมาวิเคราะห์พหุระดับ.....	42
2.4	กรอบแนวคิดในการวิจัย.....	87
3.1	กรอบแนวคิดในการวิจัย.....	95
4.1	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 1 ก่อนการใช้เกณฑ์การคัดสรรโมเดล.....	126
4.2	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 2 ก่อนการใช้เกณฑ์การคัดสรรโมเดล.....	126
4.3	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 3 ก่อนการใช้เกณฑ์การคัดสรรโมเดล.....	127
4.4	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 4 ก่อนการใช้เกณฑ์การคัดสรรโมเดล.....	128
4.5	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 5 ก่อนการใช้เกณฑ์การคัดสรรโมเดล.....	128
4.6	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 6 ก่อนการใช้เกณฑ์การคัดสรรโมเดล.....	129
4.7	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 7 ก่อนการใช้เกณฑ์การคัดสรรโมเดล.....	130
4.8	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 8 ก่อนการใช้เกณฑ์การคัดสรรโมเดล.....	130
4.9	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 9 ก่อนการใช้เกณฑ์การคัดสรรโมเดล.....	131
4.10	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 10 ก่อนการใช้เกณฑ์การสรรโมเดล.....	132

สารบัญแผนภาพ (ต่อ)

แผนภาพที่		หน้า
4.11	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 11 ก่อนการใช้เกณฑ์การคัดสรรโมเดล	132
4.12	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 12 ก่อนการใช้เกณฑ์การคัดสรรโมเดล	133
4.13	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 13 ก่อนการใช้เกณฑ์การคัดสรรโมเดล	134
4.14	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 14 ก่อนการใช้เกณฑ์การคัดสรรโมเดล	134
4.15	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 15 ก่อนการใช้เกณฑ์การคัดสรรโมเดล	135
4.16	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 16 ก่อนการใช้เกณฑ์การคัดสรรโมเดล	136
4.17	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 17 ก่อนการใช้เกณฑ์การคัดสรรโมเดล	136
4.18	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 18 ก่อนการใช้เกณฑ์การคัดสรรโมเดล	137
4.19	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 1 หลังการใช้เกณฑ์การคัดสรรโมเดล	167
4.20	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 2 หลังการใช้เกณฑ์การคัดสรรโมเดล	168
4.21	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 3 หลังการใช้เกณฑ์การคัดสรรโมเดล	169
4.22	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 4 หลังการใช้เกณฑ์การคัดสรรโมเดล	170
4.23	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 5 หลังการใช้เกณฑ์การคัดสรรโมเดล	171

สารบัญแผนภาพ (ต่อ)

แผนภาพที่		หน้า
4.24	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 6 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	172
4.25	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 7 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	173
4.26	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 8 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	174
4.27	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 9 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	175
4.28	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 10 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	176
4.29	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 11 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	177
4.30	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 12 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	178
4.31	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 13 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	179
4.32	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 14 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	180
4.33	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 15 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	181
4.34	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 16 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	182
4.35	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 17 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	183
4.36	ผลการสกัดโมเดลที่ดีที่สุดของเงื่อนไขที่ 18 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	184

สารบัญแผนภาพ (ต่อ)

แผนภาพที่		หน้า
4.37	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 1 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	190
4.38	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 2 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	191
4.39	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 3 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	192
4.40	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 4 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	193
4.41	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 5 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	194
4.42	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 6 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	194
4.43	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 7 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	195
4.44	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 8 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	196
4.45	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 9 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	197
4.46	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 10 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	198
4.47	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 11 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	199
4.48	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 12 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	200
4.49	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 13 หลังการใช้เกณฑ์การคัดสรรโมเดล.....	201

สารบัญแผนภาพ (ต่อ)

แผนภาพที่		หน้า
4.50	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 14 หลังการใช้เกณฑ์การคัดสรรโมเดล……	202
4.51	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 15 หลังการใช้เกณฑ์การคัดสรรโมเดล……	203
4.52	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 16 หลังการใช้เกณฑ์การคัดสรรโมเดล……	204
4.53	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 17 หลังการใช้เกณฑ์การคัดสรรโมเดล……	205
4.54	ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 18 หลังการใช้เกณฑ์การคัดสรรโมเดล……	206
4.55	ค่าความลำเอียง ค่าความคลาดเคลื่อนมาตรฐานและค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ยของทุกเงื่อนไข เมื่อใช้ดัชนี AIC เป็นเกณฑ์การคัดสรรโมเดล……....	211
4.56	ค่าความลำเอียง ค่าความคลาดเคลื่อนมาตรฐานและค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ยของทุกเงื่อนไข เมื่อใช้ดัชนี BIC และดัชนี ABIC เป็นเกณฑ์การคัดสรรโมเดล……....	213
4.57	ค่าความลำเอียง ค่าความคลาดเคลื่อนมาตรฐานและค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ยของทุกเงื่อนไข เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล……..	214
4.58	ค่าความลำเอียง ค่าความคลาดเคลื่อนมาตรฐานและค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ยของทุกเงื่อนไข เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล…….....	216

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

ความลำเอียงของข้อสอบ (item bias) เป็นประเด็นสำคัญมากทางการศึกษาและจิตวิทยาประเด็นหนึ่ง การตรวจสอบความลำเอียงของข้อสอบจึงมีความสำคัญไม่ยิ่งหย่อนไปกว่าการตรวจสอบความตรง (validity) หรือการตรวจสอบความเที่ยง (reliability) ทั้งนี้ก็เพราะว่าหากข้อสอบขาดความยุติธรรมแล้ว ข้อสอบนั้นจะเกิดความลำเอียงที่อาจเข้าข้างกลุ่มผู้สอบกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ ซึ่งมีผลทำให้ผู้สอบกลุ่มหนึ่งได้เปรียบในขณะที่ผู้สอบอีกกลุ่มหนึ่งเสียเปรียบ ทั้งๆ ที่ผู้สอบทั้งสองกลุ่มมีระดับความสามารถเท่ากัน ซึ่งความลำเอียงของข้อสอบดังกล่าวนี้ถือเป็นความคลาดเคลื่อนอย่างเป็นระบบ (systematic error) (Camilli and Shepard, 1994) ดังนั้นเพื่อขจัดปัญหานี้ให้หมดไป การตรวจสอบว่าข้อสอบเกิดความลำเอียงหรือไม่ จึงมีความจำเป็นอย่างยิ่งเพื่อใช้เป็นหลักฐานที่แสดงถึงความยุติธรรมของแบบสอบและพิจารณาตัดข้อสอบที่มีความลำเอียงออกจากแบบสอบ โดยกระบวนการตรวจสอบดังกล่าวถือเป็นส่วนหนึ่งของการพัฒนาแบบสอบ ซึ่งในปัจจุบันเรียกว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ใช้แทนคำว่า การตรวจสอบความลำเอียงของข้อสอบ (item bias) เนื่องจากคำว่า ความลำเอียงของข้อสอบ (item bias) ที่ใช้กันมาในอดีตนั้นมี 2 ความหมาย คือ ความหมายในเชิงสถิติและความหมายในเชิงสังคม ซึ่งความหมายทั้งสองประเด็นนี้มีการโต้เถียงกันมาโดยตลอดว่ามีความขัดแย้งกัน ต่อมาจึงใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (DIF) ซึ่งเน้นหลักฐานเชิงประจักษ์ที่ได้จากการทดสอบทางสถิติโดยแยกออกจากความหมายทางสังคม (Hambleton, Swaminathan and Rogers, 1991) โดยคำดังกล่าวมีความเหมาะสมมากกว่าในการอธิบายความหมายในเชิงสถิติ ดังนั้นนักวิจัยจึงใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (DIF) แทนคำว่าความลำเอียงของข้อสอบ (item bias) มาจนปัจจุบันนี้ (ศิริชัย กาญจนวาสี, 2550)

การศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) ไม่ได้เป็นเรื่องใหม่ทางการศึกษาหรือทางการวัดผลแต่อย่างใด หากแต่เป็นเรื่องที่นักวัดผลให้ความสนใจมาทุกยุคทุกสมัยและมีการพัฒนาวิธีในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมาอย่างต่อเนื่อง ซึ่งแม้ว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีฐานคิดมาจากทฤษฎีการทดสอบแบบดั้งเดิม (CTT) จะมีข้อได้เปรียบหลายประการ เป็นต้นว่า 1) สามารถใช้ได้จริงในเชิงปฏิบัติ 2) มีกระบวนการตรวจสอบที่ไม่ยุ่งยากซับซ้อนไม่จำเป็นต้องใช้โปรแกรมคอมพิวเตอร์ที่ซับซ้อน 3) สามารถแปลผลได้ง่าย และ 4) ใช้กลุ่มตัวอย่างขนาดเล็ก ซึ่งทำให้ประหยัดเวลาและเสียค่าใช้จ่าย อย่างไรก็ตาม วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีฐานคิดมาจากทฤษฎีการทดสอบแบบดั้งเดิม (CTT) นี้ มีข้อจำกัดที่

สำคัญคือค่าสถิติของข้อสอบไม่คงที่ เช่น ค่าความยาก (difficulty) และค่าอำนาจจำแนก (discrimination) ของข้อสอบมีค่าแปรเปลี่ยนไปตามกลุ่มผู้สอบ (Camilli and Shepard, 1994; ศิริชัย กาญจนวาสี, 2550) ดังนั้นจึงทำให้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีฐานคิดมาจากทฤษฎีการทดสอบแบบดั้งเดิม (CTT) นี้ เกิดความคลาดเคลื่อนประเภทที่ 1 (type I error) ซึ่งปัญหาดังกล่าวนี้ สามารถแก้ได้ด้วยการใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) เนื่องจากคุณสมบัติของความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบ (parameter invariance) (Hambleton, Swaminathan and Rogers, 1991) นั่นคือ ค่าพารามิเตอร์ของข้อสอบจะไม่ขึ้นอยู่กับการแจกแจงความสามารถของกลุ่มผู้สอบ ดังนั้นฟังก์ชันการตอบข้อสอบจะมีค่าเท่ากันเมื่อผู้สอบมีระดับความสามารถเท่ากัน โดยไม่คำนึงว่าผู้สอบมาจากกลุ่มใด ทำให้สามารถเปรียบเทียบผลการตอบข้อสอบที่ระดับความสามารถเดียวกันได้ ซึ่งเป็นหลักการสำคัญของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

หากใช้เกณฑ์การวิเคราะห์ข้อมูลในการแบ่งการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จะสามารถแบ่งออกเป็น 2 กลุ่มสำคัญ คือ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบ 1 ระดับ (single level DIF) และการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) โดยการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบ 1 ระดับ (single level DIF) เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เมื่อนำตัวแปรทั้งหมดที่ต้องการศึกษามาวิเคราะห์ในระดับเดียวกัน เช่น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธีแมนเทล-แฮนส์เซล (Mantel–Haenszel) หรือวิธีดีเอฟไอที (DFIT) เป็นต้น ส่วนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ให้ความสำคัญกับระดับชั้นของข้อมูลที่นำมาวิเคราะห์ นั่นคือ มีการแบ่งข้อมูลออกเป็น 2 ระดับ หรือ 3 ระดับ ตามลักษณะของข้อมูล เช่น การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธี HLM แบบ 2 ระดับ หรือ 3 ระดับหรือการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยโมเดล HGLM แบบ 2 ระดับ หรือ 3 ระดับ เป็นต้น ซึ่งสามารถวิเคราะห์ได้ด้วยโปรแกรมสำเร็จรูป HLM โปรแกรม Mplus และ/หรือโปรแกรม WinBUGS เป็นต้น

การวิเคราะห์ข้อมูลแบบพหุระดับเป็นการวิเคราะห์ข้อมูลที่ให้ค่าที่ถูกต้องและน่าเชื่อถือได้มากกว่าเมื่อข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลแบบระดับขั้น ศิริชัย กาญจนวาสี (2550) กล่าวสอดคล้องกับ Kamata (1998, 2000, 2005, 2009) ว่าหากข้อมูลมีลักษณะเป็นระดับขั้น แต่นักวิจัยบางคนละเลยลักษณะของข้อมูลที่เป็นหลายระดับ โดยการจัดกระทำตัวแปรต่างระดับกันให้มาอยู่ในระดับที่ผู้วิจัยสนใจ แล้ววิเคราะห์เสมือนหนึ่งข้อมูลเหล่านั้นอยู่ในระดับเดียวกัน ผลการวิจัยที่ได้อาจจะบิดเบือนจากความเป็นจริง ซึ่งจากการศึกษาของ Cronbach และ Webb ที่ได้วิเคราะห์ซ้ำกับผลการศึกษาในอดีตที่ได้รายงานความสัมพันธ์ของตัวแปรความถนัดกับผลสัมฤทธิ์ทางการเรียน

คณิตศาสตร์ ว่ามีปฏิสัมพันธ์ (interaction) กันกับผู้เรียนโดยการศึกษานี้ Cronbach และ Webb พบว่า ผู้วิจัยก่อนหน้ามิได้สนใจการศึกษาตัวแปรทางสังคม บริบทของโรงเรียนหรือห้องเรียนแต่อย่างใด Cronbach และ Webb จึงได้ใช้เทคนิคการศึกษาที่เรียกว่า aggregation techiques เมื่อวิเคราะห์ความสัมพันธ์ดังกล่าวใหม่ พบว่าความคลาดเคลื่อนของการสรุปผลการวิจัยนั้นคือความถนัดมีอิทธิผลปฏิสัมพันธ์กับตัวแปรความต่างระหว่างห้องเรียนมากกว่าในระดับบุคคล Cronbach และ Webb จึงสรุปว่าด้วยข้อจำกัดทางสถิติวิเคราะห์อาจจะทำให้ผลการวิเคราะห์ที่ได้ผิดพลาดได้ แม้ว่านักวิจัยจะออกแบบการวิจัยด้วยความรอบคอบก็ตาม ดังนั้นการใช้โมเดลทางสถิติที่เหมาะสมตามระดับของหน่วยการวิเคราะห์ที่สอดคล้องกับข้อมูลจึงน่าจะมีความเหมาะสมมากกว่า (Cronbach, 1976; Burstein,1978,1980; Goldstien, 1976, 2007,2008; Rogisa, 1978; Aitkin and Longford , 1986; Kamata,1998, 2000, 2005, 2009)

ดังนั้นหากข้อมูลที่จะวิเคราะห์เป็นข้อมูลแบบระดับขั้น การวิเคราะห์แบบพหุระดับจึงเป็นการวิเคราะห์ที่น่าจะมีความถูกต้องและเหมาะสมมากกว่า Binici (2007) ได้กล่าวถึงข้อดีของการวิเคราะห์พหุระดับ (Multilevel analysis) ไว้ดังนี้ 1) การวิเคราะห์พหุระดับ (Multilevel analysis) จะให้ผลการวิเคราะห์ที่ถูกต้องมากกว่าหากข้อมูลที่นำมาวิเคราะห์มีลักษณะเป็นลำดับขั้น (hierarchical data) ซึ่ง Lee (2000) กล่าวว่าทั้งนี้เนื่องมาจากการบังคับข้อมูลหลายระดับเข้าสู่การวิเคราะห์ในระดับเดียวกันจะทำให้ตัวแปรตามบางตัวที่สำคัญถูกละเลยไม่นำเข้าสู่การวิเคราะห์ 2) การวิเคราะห์พหุระดับ (Multilevel analysis) สามารถศึกษาผลของปฏิสัมพันธ์ข้ามระดับได้ ซึ่งการศึกษาในลักษณะนี้จะไม่สามารถศึกษาได้ในการศึกษาแบบ 1 ระดับ (single level) และ 3) การวิเคราะห์พหุระดับ (Multilevel analysis) จะให้ค่าความแปรปรวน (variance) และความแปรปรวนร่วม (covariance) ของจุดตัดแบบสุ่ม (random intercept) และ/หรือสัมประสิทธิ์ความชัน (slope coefficient) ของโครงสร้างแบบระดับขั้น (level of hierarchical structure) Guo and Zhao (2000) และ Heck and Thomas (2000) ได้กล่าวถึงข้อดีของการวิเคราะห์พหุระดับ (Multilevel analysis) ว่าเป็นการวิเคราะห์ที่เป็นที่นิยมกันอย่างแพร่หลายทางการศึกษา ทางการแพทย์และในทางธุรกิจ ทั้งนี้ก็เพราะว่าเป็นการวิเคราะห์ที่เหมาะสมกับข้อมูลที่มีลักษณะแทรกสอด (nested structure data) นอกจากนี้การวิเคราะห์พหุระดับ (Multilevel analysis) ยังศึกษาได้ในข้อมูลแบบตัดขวาง (cross-sectional data) และข้อมูลตามลำดับเวลา (longitudinal data) เช่น การศึกษาพัฒนาการ (growth study) เป็นต้น (Kamata, 2005, 2009) จากที่กล่าวมาจะเห็นได้อย่างชัดเจนว่าการวิเคราะห์แบบพหุระดับนั้นเป็นการศึกษาที่ให้สารสนเทศมากกว่าการวิเคราะห์แบบ 1 ระดับ

การศึกษาเกี่ยวกับโมเดล HGLM (Hierarchical Generalized Linear Model) มีมาตั้งแต่ปี ค.ศ. 1995 จากการศึกษาของ Raudenbush (1995 อ้างถึงใน Cheong, 2006; Kamata, 1998, 2001, 2002, 2003) โดยโมเดล HGLM (Hierarchical Generalized Linear Model) เกิดจากการ

รวมกันของโมเดล GLM (Generalized Linear Model) และโมเดล HLM (Hierarchical Linear Model) ซึ่งสามารถวิเคราะห์ได้จากโปรแกรม HLM (Cheong, 2006; Kamata, 1998, 2001, 2002, 2003; Beretvas and Williams, 2004) แต่โมเดล HGLM มาเป็นที่รู้จักกันมากขึ้นใน ปี ค.ศ. 1998 จากการศึกษาของ Kamata (1998) ที่ศึกษาโมเดล HGLM แบบ 1 พารามิเตอร์ แต่เป็นการประยุกต์โมเดล HGLM สู่ทฤษฎีการตอบสนองข้อสอบ (IRT) โดย Kamata ได้นำเสนอผลการพิสูจน์ทางคณิตศาสตร์ว่า โมเดลราสช์ (Rasch model) เป็นกรณีหนึ่งของโมเดลการตอบสนองข้อสอบ (IRT model) หรืออาจกล่าวอีกนัยหนึ่งว่า โมเดล HGLM มีฐานคิดมาจากโมเดลราสช์แบบ 1 พารามิเตอร์ (1 parameter Rasch model) นอกจากนี้ Kamata เป็นคนแรกที่ค้นพบว่าสามารถวิเคราะห์โมเดลราสช์ด้วยโปรแกรม HLM ได้ จากนั้นในปี ค.ศ. 2001 Kamata (2001) ได้ขยายแนวคิดของโมเดล HGLM สู่การวิเคราะห์พหุระดับแบบ 3 ระดับ (3 level Multilevel analysis) โดยเพิ่มการวิเคราะห์ในระดับที่ 3 คือ ระดับโรงเรียน (school level) หรือระดับกลุ่ม (cluster level) เข้าไปในโมเดลการวิเคราะห์และในปี ค.ศ. 2002 Kamata (2002) ได้นำเสนอผลการศึกษากระบวนการในการจัดกระทำกับการวิเคราะห์การตอบสนองข้อสอบด้วยโมเดล HGLM ซึ่งได้ข้อค้นพบว่าการวิเคราะห์พหุระดับด้วยโมเดลราสช์ (Rasch model) มีความทัดเทียม (equivalent) กับการวิเคราะห์ด้วยโมเดลการตอบสนองข้อสอบ (IRT model)

ปัจจุบันนี้โมเดล HGLM เป็นที่รู้จักในนามของโมเดล Kamata (Kamata's Model) และ Kamata and Cheong (2007) สรุปว่า โมเดล HGLM สามารถใช้ได้กับข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) และข้อสอบที่ตรวจให้คะแนนมากกว่า 2 ค่า (polytomous) ใช้ได้กับแบบสอบแบบเอกมิติ (Unidimensional test) และแบบสอบแบบพหุมิติ (Multidimensional test) และสามารถวิเคราะห์พหุระดับได้ทั้งแบบ 2 ระดับ (2 level HGLM) และแบบ 3 ระดับ (3 level HGLM) รวมไปถึงการประยุกต์สู่การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วย

หากใช้กรอบแนวคิดของโมเดล HGLM เมื่อระดับที่ 1 เป็นการวิเคราะห์ระดับข้อสอบ (item level) ระดับที่ 2 เป็นการวิเคราะห์ระดับบุคคล (individual level) และระดับที่ 3 เป็นการวิเคราะห์ระดับกลุ่มหรือระดับโรงเรียน (cluster level or group level or school level) จะได้สมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ i ได้ถูกต้องของนักเรียนคนที่ j ซึ่งมาจากโรงเรียนหรือกลุ่มที่ k I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์

อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 โดยในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

โดย Kamata เป็นผู้ริเริ่มในการประยุกต์โมเดล HGLM สู่การวิเคราะห์การทำหน้าที่ต่างกัน (DIF) นั่นก็หมายความว่าเป็นคนแรกที่วิเคราะห์การทำหน้าที่ต่างกันแบบพหุระดับ (Multilevel DIF) โดยใช้โมเดล HGLM เป็นกรอบแนวคิดในการศึกษาและได้พัฒนาสูตรและสมการรวมทั้งขยายแนวคิดของโมเดล HGLM ให้มีคุณค่าต่อศาสตร์แห่งการวัดผลมากยิ่งขึ้น จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องอย่างครอบคลุม พบว่า การศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF Model) ที่ผ่านมาเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM ทั้งสิ้น โดยโมเดล HGLM นี้สามารถประยุกต์ใช้ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ทั้งในการวิเคราะห์แบบ 2 ระดับและแบบ 3 ระดับ (Binici, 2007; Chu and Kamata, 2003; Chaimongkol, 2005; Cheong, 2006; Kamata and al, 2005; Vaughn, 2006; Williams and Beretvas, 2006) Chaimongkol (2005) กล่าวว่าโมเดล HGLM หรือ โมเดล Kamata (Kamata's Model) เป็นโมเดลที่มีคุณค่า/เหมาะสมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เพราะว่า 1) โมเดล HGLM สามารถวินิจฉัยอิทธิพลของข้อสอบ (item impact) และอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ (DIF effect) ได้ 2) โมเดล HGLM สามารถประมาณค่าพารามิเตอร์

ขนาดของการทำหน้าที่ต่างกันของข้อสอบ (Magnitude of DIF) ได้โดยตรง และ 3) โมเดล HGLM มีกระบวนการทดสอบระดับนัยสำคัญทางสถิติของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

แต่อย่างไรก็ตาม จากการศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยโมเดล HGLM ที่ผ่านมานั้น นักวิจัยจะเลือกเพียงบางพารามิเตอร์เท่านั้นเข้าสู่โมเดลการวิเคราะห์ นั่นคือนักวิจัยจะละเลยอิทธิพลของพารามิเตอร์บางพารามิเตอร์ ตัวอย่างเช่น การศึกษาของ Kamata and Cheong (2007) ที่ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับด้วยโมเดล HGLM (multilevel DIF detection by Hierarchical Generalized Linear Modeling (HLGM)) แต่การศึกษาในครั้งนั้น กำหนดให้ค่าพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์ นอกจากนี้จากการศึกษาของ Chaimongkol, Huffer and Kamata (2006) และ Vaughn (2006) ก็ศึกษาโดยกำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์ด้วยเช่นกัน เหตุผลสำคัญที่นักวิจัยเหล่านี้กำหนดค่าพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์ก็เพื่อให้โมเดลการวิเคราะห์เป็นโมเดลที่ง่ายและไม่ซับซ้อน ซึ่งจะทำให้การตีความหมายง่ายขึ้นและเหตุผลที่สำคัญมากอีกข้อหนึ่ง คือ เนื่องจากยังไม่มีการศึกษาค้นคว้ามาก่อนว่าพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) มีอิทธิพลต่อการผันแปรของการประมาณค่าพารามิเตอร์ต่างๆ ของโมเดลการทำหน้าที่ต่างกันของข้อสอบหรือไม่เพียงใด รวมไปถึงยังไม่มีหลักฐานว่าควรจะกำหนดขนาดของพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นเท่าใดเพื่อให้ผลการประมาณค่าพารามิเตอร์ต่างๆ ของโมเดลการทำหน้าที่ต่างกันของข้อสอบมีความน่าเชื่อถือมากที่สุด

ดังนั้นการศึกษาในครั้งนี้จึงมีจุดมุ่งหมายที่จะตรวจสอบว่าสมมุติฐานของการกำหนดให้ค่าพารามิเตอร์บางพารามิเตอร์เป็นศูนย์นั้นไม่ค่อยถูกต้องนักและเพื่อเป็นการขยายขายแดนความรู้ของโมเดล HGLM และการศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) ให้กว้างขวางมากยิ่งขึ้น ผู้วิจัยมีวัตถุประสงค์ในการศึกษาในครั้งนี้ 2 ประการ คือ 1) เพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มและ 2) เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)

เพื่อให้ได้มาซึ่งคำตอบตามวัตถุประสงค์ดังกล่าวข้างต้น ซึ่งเป็นการศึกษาค้นคว้าในเชิงทฤษฎีที่ยังไม่มีการศึกษาค้นคว้ามาก่อน การศึกษาในครั้งนี้จึงเป็นการศึกษาโดยการจำลองข้อมูลด้วยเงื่อนไขที่แตกต่างกันจำนวน 18 เงื่อนไข (2 x 3 x 3) อันประกอบไปด้วย 1) จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school or the number of individual in each cluster) 2 ระดับ คือ โรงเรียนละ 50 คนและโรงเรียนละ 100 คน 2) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม

(CLDIF) 3 ระดับ คือ 0.2, 0.4 และ 0.6 และ 3) ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ คือ 0.1, 0.2 และ 0.3 ส่วนจำนวนโรงเรียนนั้นกำหนดคงที่ทุกเงื่อนไขที่ 50 โรงเรียน โดยในแต่ละเงื่อนไขกำหนดการทำซ้ำเงื่อนไขละ 200 รอบ

ผู้วิจัยประยุกต์แนวคิดของโมเดล HGLM เป็นโมเดลการทำหน้าที่ต่างกันของข้อสอบ โดยการศึกษาในครั้งนี้มีที่ต้องการศึกษา 4 โมเดล ได้แก่ 1) โมเดลเต็มรูป (complete model) ซึ่งเป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) ทั้ง 7 อิทธิพลและอิทธิพลสุ่ม (random effect) 2 อิทธิพล 2) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล 3) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพลและ 4) โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 5 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล

การศึกษาในครั้งนี้มีพารามิเตอร์เป้าหมายในการประมาณค่า 6 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 2) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) 3) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 6) พารามิเตอร์ความสามารถระดับกลุ่ม (CA)

ผู้วิจัยมีเกณฑ์การคัดสรรโมเดล (Model selection criteria) 5 เกณฑ์ ได้แก่ 1) ดัชนี AIC (Akaike's information criterion), 2) ดัชนี BIC (Bayesian information criterion), 3) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), 4) ดัชนี 2 of 3 และ 5) ดัชนี p-value และมีเกณฑ์การประเมิน (evaluation criteria) ในการประมาณค่าพารามิเตอร์ 3 เกณฑ์ คือ ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) โดยการศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลด้วยโปรแกรม R และประมาณค่าพารามิเตอร์ด้วยวิธี Maximum likelihood ด้วยโปรแกรม Mplus ภายใต้การเขียนคำสั่ง การนำข้อมูลไปวิเคราะห์และการแสดงผลในโปรแกรม R

เนื่องจากการศึกษาในครั้งนี้ มุ่งเน้นการศึกษาในเชิงทฤษฎี เพื่อบุกเบิกองค์ความรู้ใหม่ที่ยังไม่มีการศึกษาค้นคว้าใดๆ มาก่อน ดังนั้นจึงอาจกล่าวได้ว่าสารสนเทศจากการศึกษาค้นคว้าในครั้งนี้จะมีความสำคัญและเป็นประโยชน์เป็นอย่างยิ่งในการขยายองค์ความรู้ของการศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) องค์ความรู้ของการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) องค์ความรู้ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF Model) และองค์ความรู้ของโมเดล HGLM ให้กว้างขวางมากยิ่งขึ้น โดยข้อค้นพบที่ได้จากการศึกษาในครั้งนี้จะช่วยขยายศาสตร์ของการวัดผล (measurement) และศาสตร์ของจิตมิติ (psychometric) ให้มีอาณาเขตที่กว้างขึ้นอย่างมีคุณค่า

คำถามวิจัย

การศึกษาครั้งนี้มีคำถามวิจัยที่จะนำไปสู่การศึกษาค้นคว้าต่อไป ดังนี้

1) คุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มเป็นอย่างไร

2) โมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดล โมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)

วัตถุประสงค์ของการวิจัย

การศึกษาในครั้งนี้มีวัตถุประสงค์ในการศึกษา คือ

1) เพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม

2) เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)

ขอบเขตของการวิจัย

การศึกษาในครั้งนี้มีขอบเขตของการศึกษา ดังนี้

1) การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูล (simulation) แบบ 3 ระดับ โดยข้อมูลจำลองระดับที่ 1 เป็นข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) จำนวน 12 ข้อ โดยความยาวของข้อสอบไม่ได้เป็นปัจจัยที่ผู้วิจัยต้องการศึกษา

2) การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลด้วยเงื่อนไขที่แตกต่างกันจำนวน 18 (2 x 3 x 3) เงื่อนไข อันได้แก่ 1) จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) 2 ระดับ คือ โรงเรียนละ 50 คนและโรงเรียนละ 100 คน 2) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 3 ระดับ คือ 0.2, 0.4 และ 0.6 และ 3) ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ คือ 0.1, 0.2 และ 0.3 โดยในแต่ละเงื่อนไขกำหนดการทำซ้ำ 200 รอบ

3) จำนวนโรงเรียนไม่ได้เป็นตัวแปรที่ผู้วิจัยต้องการศึกษา ดังนั้น ผู้วิจัยจึงกำหนดให้จำนวนโรงเรียนเป็นค่าคงที่ นั่นคือกำหนด 50 โรงเรียนในทุกเงื่อนไข

4) โมเดลการทำหน้าที่ต่างกันของข้อสอบที่ศึกษาในครั้งนี้มี 4 โมเดล ได้แก่ 1) โมเดลเต็มรูป (complete model) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) ทั้ง 7 อิทธิพลและอิทธิพลสุ่ม (random effect) 2 อิทธิพล 2) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล 3) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพลและ 4) โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 5 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล

5) การศึกษาในครั้งนี้มีพารามิเตอร์เป้าหมายในการประมาณค่า 6 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 2) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) 3) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 6) พารามิเตอร์ความสามารถระดับกลุ่ม (CA)

6) เกณฑ์การคัดสรรโมเดล (Model selection criteria) ในการศึกษาครั้งนี้ ประกอบไปด้วย 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ได้แก่ 1) ดัชนี AIC (Akaike's information criterion), 2) ดัชนี BIC (Bayesian information criterion), 3) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), 4) ดัชนี 2 of 3 ซึ่งเป็นการใช้เกณฑ์ 2 ใน 3 ของดัชนี AIC (Akaike's information criterion) ดัชนี BIC (Bayesian information criterion) หรือดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และ 5) ดัชนี p-value โดยพิจารณาดัชนี p-value ของพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และดัชนี p-value ของพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF)

7) การศึกษาในครั้งนี้มีเกณฑ์การประเมิน (evaluation criteria) ในการประมาณค่าพารามิเตอร์ 3 เกณฑ์ คือ 1) ค่าความลำเอียง (BIAS) 2) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และ3) ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE)

ข้อจำกัดของการวิจัย

จากวัตถุประสงค์ข้อ 1 ที่ผู้วิจัยต้องการศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม ซึ่งผู้วิจัยต้องการศึกษาว่าค่าประมาณของพารามิเตอร์เป้าหมายแต่ละตัวที่ได้จากการประมาณค่าเป็นอย่างไรเท่านั้น ดังนั้นผู้วิจัยจึงไม่มีการทดสอบว่าขนาดอิทธิพลเท่าไหร่ที่ส่งผลต่อการประมาณค่าพารามิเตอร์ หรือมีปฏิสัมพันธ์ระหว่างตัวแปรที่ผู้วิจัยศึกษาหรือไม่

นิยามศัพท์ที่ใช้ในการวิจัย

การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) หมายถึง การที่ผู้สอบที่มีระดับความสามารถเดียวกันแต่มีโอกาสในการตอบข้อสอบข้อนั้นๆ ได้ถูกต้องแตกต่างกันเมื่อมาจากกลุ่มที่แตกต่างกัน

การทำหน้าที่ต่างกันของข้อสอบแบบ 1 ระดับ (single level Differential Item Functioning: single level DIF) หมายถึง การที่ผู้สอบที่มีระดับความสามารถเดียวกันแต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน เมื่อมาจากกลุ่มที่แตกต่างกัน โดยวิเคราะห์ข้อมูลแบบระดับเดียว

การทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (multilevel Differential Item Functioning; Multilevel DIF) หมายถึง การที่ผู้สอบที่มีระดับความสามารถเดียวกันแต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน เมื่อมาจากกลุ่มที่แตกต่างกัน โดยวิเคราะห์ข้อมูลแบบ 2 ระดับหรือ 3 ระดับ

โมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น (Hierarchical Generalized Linear Model: HGLM) หมายถึง ระเบียบวิธีในการวิเคราะห์ข้อมูลเชิงเส้นทั่วไปที่มีการประยุกต์โมเดลเชิงเส้นทั่วไป (Generalized Linear Model: GLM) และโมเดลเชิงเส้นตรงระดับลดหลั่น (Hierarchical Linear Model: HLM) เข้าด้วยกัน มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

โมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF model) หมายถึง โมเดลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อกำหนดตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม นั่นคือ เป็นโมเดลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบข้ามระดับที่ 3 กับระดับที่ 1 และระดับที่ 3 กับระดับที่ 2 โดยผู้วิจัยประยุกต์ใช้โมเดล HGLM เป็นโมเดลในการวิเคราะห์

โมเดลเต็มรูป (complete model) หมายถึง โมเดลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยผู้วิจัยประยุกต์ใช้โมเดล HGLM เป็นโมเดลในการวิเคราะห์ ซึ่งเป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) ทั้ง 7 อิทธิพลและอิทธิพลสุ่ม (random effect) 2 อิทธิพล มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) หมายถึง โมเดลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่ผู้วิจัยประยุกต์ใช้โมเดล HGLM เป็นโมเดลในการวิเคราะห์ เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) หมายถึง โมเดลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่ผู้วิจัยประยุกต์ใช้โมเดล HGLM ในการวิเคราะห์ เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction, and cluster-level DIF) หมายถึง โมเดลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่ผู้วิจัยประยุกต์ใช้โมเดล HGLM ในการวิเคราะห์ เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 5 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_{5i} X_j I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม

(cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 5 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (Root-Mean Square Error; RMSE) หมายถึง ขนาดของความแตกต่างระหว่างการประมาณค่าพารามิเตอร์และค่าจริง เป็นค่าที่สะท้อนให้เห็นถึงความคงที่ (Stability) ของการประมาณค่าพารามิเตอร์ หากพารามิเตอร์ใดมีค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) น้อยกว่า แสดงว่าความคงที่ในการประมาณค่าพารามิเตอร์สูงกว่า สะท้อนให้เห็นถึงความคลาดเคลื่อนของการประมาณค่าน้อย แต่ถ้าค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) มากกว่า แสดงว่ามีความคงที่ของการประมาณค่าพารามิเตอร์ต่ำ สะท้อนให้ว่ามีความคลาดเคลื่อนของการประมาณค่าพารามิเตอร์มากกว่า

ค่าความลำเอียง (BIAS) หมายถึง ค่าความคลาดเคลื่อนการประมาณค่าพารามิเตอร์ นั่นคือการประมาณค่าที่ได้ต่างจากค่าพารามิเตอร์ที่แท้จริงเพียงใด ค่าความลำเอียง (BIAS) สะท้อนให้เห็นถึงความถูกต้อง (Accuracy) ของการประมาณค่าพารามิเตอร์ หากค่าความลำเอียง (BIAS) ใกล้ศูนย์มากกว่าสะท้อนให้เห็นถึงการประมาณค่าใกล้เคียงกับค่าที่แท้จริงมากกว่า โดยเครื่องหมายที่ได้จะสะท้อนถึงความคลาดเคลื่อนของการประมาณค่าที่ห่างจากค่าที่แท้จริงในลักษณะใด ยกตัวอย่างเช่น หากค่าความลำเอียง (BIAS) ติดลบ (-) แสดงให้เห็นว่าการประมาณค่าต่ำกว่าความเป็นจริง (Underestimate) แต่ถ้ามีค่าเป็นบวก (+) แสดงว่าการประมาณค่าสูงกว่าความเป็นจริง (Overestimate)

ค่าความคลาดเคลื่อนมาตรฐาน (SE) หมายถึง ค่าที่แสดงว่าโดยเฉลี่ยแล้วค่าเฉลี่ยของค่าพารามิเตอร์ที่ประมาณได้แต่ละตัวแตกต่างจากค่าจริงมากน้อยเพียงใด ซึ่งค่าความคลาดเคลื่อนมาตรฐาน (SE) น้อยๆ แสดงให้เห็นว่าค่าประมาณของพารามิเตอร์นั้นๆ ใกล้เคียงกับค่าจริง

ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ทางวิชาการ

1) เป็นการขยายองค์ความรู้ของการศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) ด้วยใช้โมเดล HGLM เป็นกรอบแนวคิดให้กว้างขวาง ครอบคลุมบริบททางการศึกษาและสอดคล้องกับสภาพจริงมากยิ่งขึ้น รวมไปถึงเป็นการบุกเบิกองค์ความรู้ใหม่ที่ยังไม่มีการศึกษาค้นคว้ามาก่อน ทั้งนี้ก็เนื่องมาจาก จากการศึกษาค้นคว้าที่ผ่านมา ยังไม่การศึกษาค้นคว้าโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มมาก่อนแต่อย่างใด นั่นคือ นักวิจัยจะกำหนดพารามิเตอร์บางพารามิเตอร์ให้เป็นศูนย์ เพื่อให้ง่ายต่อการตีความหมายและเป็นโมเดลที่ประหยัด ดังนั้นจากผลการศึกษาในครั้งนี้จะเป็นการนำเสนอสารสนเทศหรือทางเลือกใหม่แก่ศาสตร์แห่งการวัดผลว่า เมื่อกำหนดค่าพารามิเตอร์ต่างๆ กัน จะได้ค่าประมาณของพารามิเตอร์เป้าหมายเป็นอย่างไร และการกำหนดให้ค่าพารามิเตอร์บางพารามิเตอร์เป็นศูนย์เป็นการกระทำที่เหมาะสมแล้วหรือยัง นอกจากนี้หากพบว่าค่าพารามิเตอร์บางพารามิเตอร์แตกต่างจากค่าจริง ก็จะเป็นแนวทางในการศึกษาค้นคว้าต่อไปในอนาคตว่าการที่ค่าพารามิเตอร์นั้นแตกต่างจากค่าจริงมีสาเหตุมาจากปัจจัยใด เป็นต้น

2) จากการศึกษาประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ของโมเดล HGLM ได้ดีที่สุด ซึ่งยังไม่มีการศึกษาค้นคว้าเลยในปัจจุบัน ว่าโมเดลไหนเป็นโมเดลการทำหน้าที่ต่างกันของข้อสอบที่ดีที่สุด ดังนั้นสารสนเทศที่ได้จากการศึกษาค้นคว้าในครั้งนี้ จะเป็นการนำเสนอสารสนเทศหรือแนวทางในการเลือกใช้โมเดลในอนาคตแต่นักวิจัยว่าควรจะเลือกใช้โมเดลไหนในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบถึงจะเป็นโมเดลที่ดีที่สุดทั้งนี้ก็เพราะว่าการเลือกใช้โมเดลที่เหมาะสมในการวิเคราะห์จะทำให้ผลการวิเคราะห์ที่ได้มีความน่าเชื่อถือ

3) จากการศึกษาค้นคว้าที่ผ่านๆ มา นักวิจัยจะเลือกใช้เพียงเกณฑ์ใดเกณฑ์หนึ่งเท่านั้นในการตัดสินว่าโมเดลไหนเป็นโมเดลที่ดีที่สุด ตัวอย่างเช่น การศึกษาของ Kamata and Cheong (2007) ที่ใช้เพียงดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล แต่สำหรับการศึกษาในครั้งนี้ ผู้วิจัยใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ถึง 5 เกณฑ์ อันได้แก่ 1) ดัชนี AIC (Akaike's information criterion) 2) ดัชนี BIC (Bayesian information criterion) 3) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), 4) ดัชนี 2 of 3 และ 5) ดัชนี p-value ซึ่งทั้ง 5 เกณฑ์นี้ ล้วนเป็นเกณฑ์ที่นักวิจัยนิยมใช้ในการคัดเลือกโมเดลทั้งสิ้น แล้วล้วนเป็นเกณฑ์ที่ผ่านการศึกษาค้นคว้ามาแล้วว่ามีความถูกต้องแม่นยำสูงและให้ผลที่เชื่อถือได้ เหตุผลที่ผู้วิจัยใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ถึง 5 เกณฑ์ก็เพื่อเป็นการยืนยันและตรวจสอบว่าผลที่ได้จากการศึกษาในครั้งนี้ มีความน่าเชื่อถือและได้มาซึ่งโมเดลที่ดีที่สุด (best

model) จริงๆ รวมไปถึงการได้สารสนเทศว่าการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่แตกต่างกันส่งผลต่อการได้มาซึ่งโมเดลที่ดีที่สุด (best model) ที่ต่างกันหรือไม่

ประโยชน์ในการนำไปใช้

1) ผลที่ได้จากการศึกษาในครั้งนี้ ทำให้ทราบถึงคุณภาพของค่าพารามิเตอร์ที่ประมาณได้แต่ละตัวว่าเหมือนหรือแตกต่างจากค่าจริงมากน้อยเพียงใดและ/หรือเกิดขึ้นในเงื่อนไขใดบ้าง ซึ่งผลการศึกษาในครั้งนี้จะเป็นการนำเสนอสารสนเทศใหม่ เพื่อให้ผู้ที่สนใจสามารถนำผลการศึกษาไปอ้างอิงหรือนำไปศึกษาค้นคว้าต่อไปในอนาคต โดยเฉพาะอย่างยิ่งหากผลการศึกษาพบว่ามีค่าพารามิเตอร์บางพารามิเตอร์หรือในบางสถานการณ์ที่ได้ค่าประมาณแตกต่างออกไปจากค่าจริง

2) สำหรับประโยชน์ในทางปฏิบัติ (practical value) นั้น หากหน่วยงานที่เกี่ยวข้องกับการศึกษา เป็นต้นว่า หน่วยงานระดับโรงเรียน หน่วยงานระดับเขตพื้นที่การศึกษาหรือระดับกระทรวงตระหนักและเห็นความสำคัญของการจัดการทดสอบที่มีความยุติธรรมสำหรับนักเรียนทุกคนแล้วหน่วยงานที่เกี่ยวข้องสามารถนำผลจากการศึกษาในครั้งนี้ไปใช้ในการวางแผนการทดสอบ การนำผลไปใช้ในระดับนโยบายหรือใช้ในการบริหารจัดการ เพื่อให้อิทธิพลของตัวแปรระดับกลุ่มหรือปฏิสัมพันธ์ข้ามระดับเมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม นั่นคือ ปฏิสัมพันธ์ของตัวแปรระดับที่ 3 กับตัวแปรระดับที่ 2 และปฏิสัมพันธ์ของตัวแปรระดับที่ 3 กับตัวแปรระดับที่ 1 ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ (DIF) น้อยที่สุด นั่นคือ เพื่อให้การบริหารจัดการสอบเกิดความยุติธรรมแก่นักเรียนมากที่สุด นั่นเอง แม้ว่าการศึกษาในครั้งนี้เป็นการศึกษาในเชิงทฤษฎีแต่ว่าสามารถนำผลการศึกษาที่ได้ไปประยุกต์ใช้ในทางปฏิบัติได้อย่างมีประสิทธิภาพในเชิงนโยบายหรือในทางปฏิบัติ

3) เป็นแนวทางในการศึกษาค้นคว้าการศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) การศึกษาค้นคว้าการทำหน้าที่ต่างกันของแบบสอบแบบพหุระดับ (Multilevel DIF) และการประยุกต์แนวคิดของโมเดล HGLM สู่การวิเคราะห์ด้วยข้อมูลจริงหรือในบริบททางการศึกษาจริงต่อไป

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องทั้งในประเทศและต่างประเทศ รวมทั้งจากฐานข้อมูลต่างๆ ผู้วิจัยจึงได้เขียนรายงานการศึกษาค้นคว้า โดยแบ่งออกเป็น 5 ตอน ซึ่งประกอบไปด้วย ตอนที่ 1 มโนทัศน์ของทฤษฎีการตอบสนองข้อสอบ (IRT) ตอนที่ 2 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ (DIF) ตอนที่ 3 มโนทัศน์ของการวิเคราะห์พหุระดับ (Multilevel analysis) แบ่งออกเป็น 3.1) มโนทัศน์ของการวิเคราะห์พหุระดับ และ 3.2) การวิเคราะห์พหุระดับด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น (HLM) ตอนที่ 4 โมเดลเชิงเส้นตรงทั่วไปแบบลดหลั่น (HGLM) สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง แบ่งออกเป็น 5.1) งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และ 5.2) งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM และในส่วนท้ายของบทที่ 2 เป็นการสรุปเนื้อหาที่ได้จากการศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องรวมไปถึงการเสนอกรอบแนวคิดของการวิจัย โดยมีรายละเอียด ดังนี้

ตอนที่ 1 มโนทัศน์ของทฤษฎีการตอบสนองข้อสอบ (IRT)

1.1) มโนทัศน์ของทฤษฎีการตอบสนองข้อสอบ (IRT)

ขอนำเสนอมโนทัศน์ของทฤษฎีการตอบสนองข้อสอบ (IRT) โดยเริ่มต้นจาก 1.1.1) แนวคิดพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) 1.1.2) หลักการและข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ (IRT) 1.1.3) ฟังก์ชันสารสนเทศของข้อสอบและฟังก์ชันสารสนเทศของแบบสอบ (Item Information Function and Test Information Function) และ 1.1.4) โมเดลการตอบสนองข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (Dichotomous Item response Theory) และ 1.1.5) สรุปสาระสำคัญของทฤษฎีการตอบสนองข้อสอบ (IRT) ดังรายละเอียดต่อไปนี้

1.1.1) แนวคิดพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT)

ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีที่เกิดขึ้นเพื่อแก้ไขข้อจำกัดของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) (Hambleton & Swaminatan, 1985; ศิริชัย กาญจนวาสี, 2550) อันประกอบไปด้วย 1) ค่าสถิติของข้อสอบภายใต้ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) อันได้แก่ ค่าความ

ยากของข้อสอบจะแปรเปลี่ยนตามลักษณะของกลุ่มผู้สอบ หากกลุ่มผู้สอบมีความสามารถสูง ค่าความยาก (difficulty) ของข้อสอบจะสูง ข้อสอบจะกลายเป็นข้อสอบที่ง่าย แต่ถ้าผู้สอบมีระดับความสามารถที่ต่ำ ค่าความยากที่วิเคราะห์ได้จะมีค่าต่ำ ข้อสอบข้อดังกล่าวจะกลายเป็นข้อสอบที่ยากไปทั้งๆ ที่เป็นข้อสอบข้อเดียวกัน แต่นำไปวัดกับกลุ่มผู้สอบที่มีระดับความสามารถแตกต่างกัน ส่วนค่าอำนาจจำแนก (discrimination) ก็จะขึ้นอยู่กับความเป็นเอกพันธ์ของความสามารถของผู้สอบ หากผู้สอบมีความสามารถแตกต่างกันมาก ข้อสอบก็จะมีค่าอำนาจจำแนก (discrimination) สูง ซึ่งจะส่งผลให้แบบสอบมีค่าความเที่ยง (reliability) สูงตามไปด้วย ในขณะที่หากนำแบบสอบฉบับเดียวกันนี้ไปใช้กับกลุ่มผู้ตอบต่างกลุ่ม ค่าอำนาจจำแนก (discrimination) และค่าความเที่ยง (reliability) ก็จะแปรเปลี่ยนไปตามกลุ่มผู้สอบ 2) การเปรียบเทียบความสามารถของผู้สอบนั้น จะต้องใช้แบบสอบฉบับเดียวกันหรือแบบสอบคู่ขนาน ปัญหาที่เกิดขึ้นก็คือแบบสอบวัดผลสัมฤทธิ์และแบบสอบวัดความถนัดนั้น ส่วนใหญ่จะเหมาะกับผู้สอบที่มีความสามารถในระดับปานกลาง ดังนั้นความถูกต้องแม่นยำของการวัดสำหรับผู้สอบที่มีความสามารถสูงและผู้สอบที่มีความสามารถต่ำจึงลดลง 3) ค่าความเที่ยง (reliability) ของแบบสอบถูกนิยามในรูปของผลที่ได้จากการใช้แบบสอบคู่ขนาน ซึ่งในทางปฏิบัติแล้ว นับเป็นเรื่องยากที่จะควบคุมให้การสอบทั้ง 2 ครั้งมีสภาพที่เหมือนกัน ถึงแม้ว่าแบบสอบคู่ขนานนั้นจะขนานกันจริง แต่ผู้สอบอาจจะมีคุณลักษณะที่แตกต่างไปจากการสอบในครั้งแรก เช่น ความวิตกกังวล การลืม หรือการพัฒนาตนเองเพิ่มมากขึ้นในบางทักษะความสามารถที่อาจส่งผลต่อการตอบข้อสอบ เป็นต้น 4) ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ไม่สามารถบอกได้ว่าผู้สอบจะตอบข้อสอบอย่างไร ยกเว้นแต่ว่าจะเคยใช้ข้อสอบข้อนั้นกับผู้สอบที่มีลักษณะคล้ายคลึงกันมาก่อนแล้ว และ 5) ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ใช้ค่าความแปรปรวนของความคลาดเคลื่อนในการวัด (variance of error of measurement) เหมือนกันในผู้สอบทุกคน ซึ่งตามสภาพแห่งความเป็นจริงแล้ว ผู้สอบที่มีความสามารถสูงและผู้สอบที่มีความสามารถต่ำจะมีค่าความแปรปรวนของความคลาดเคลื่อนในการวัดแตกต่างจากผู้สอบที่มีความสามารถในระดับปานกลาง

Master & Keeves (1999) ได้กล่าวถึงแนวคิดพื้นฐานที่สำคัญสำหรับทฤษฎีการตอบสนองข้อสอบ (IRT) 3 ประการ ดังนี้

1) หลักการสัมพันธ์ (relative principle) เป็นการประเมินแบบเชื่อมโยงประสาน ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) มุ่งประเมินผลงานของบุคคลตามชุดของข้อสอบ แต่อย่างไรก็ตาม การวัดทางจิตวิทยาและทางการศึกษา อาจมีปฏิสัมพันธ์ระหว่างบุคคลกับข้อสอบ การวัดจึงไม่ได้มีความหมายเพียงการดูข้อสอบรายข้อเท่านั้น แต่เป็นการพิจารณาถึงการตอบสนองบุคคลกับข้อสอบด้วย กล่าวอีกนัยหนึ่ง คือ ถ้า β_n คือ ดัชนีความสามารถของบุคคล *n* คือ คุณลักษณะที่ต้องการวัด และ ถ้า δ_i คือ ดัชนีความยากของข้อสอบซึ่งเกี่ยวข้องกับคุณลักษณะภายในของบุคคลด้วย จะได้ว่าดัชนี

ความสามารถของบุคคลจะเกี่ยวข้องกับดัชนีความยากของข้อสอบ โดยเป็นความแตกต่างระหว่างความสามารถของบุคคลกับความยากของข้อสอบ หรือ $\beta_n - \delta_i$ หรือเป็นสัดส่วนของค่าความสามารถของบุคคลกับความยากของข้อสอบ ถ้าความสามารถของบุคคลสูงกว่าค่าความยากของข้อสอบ คำตอบที่คาดหวังน่าจะถูกต้องในทำนองเดียวกันหากค่าความสามารถของบุคคลต่ำกว่าค่าความยากของข้อสอบ คำตอบที่คาดหวังน่าจะไม่ถูกต้อง ดังนั้นค่าความสามารถของบุคคลและความยากของข้อสอบจึงควรนำมาพิจารณาร่วมกันในการวิเคราะห์คำตอบทั้งหมด

2) หลักความน่าจะเป็น (probability principle) เป็นความน่าจะเป็นของคำถามที่ตอบสนองต่อข้อสอบ คำตอบในการตอบข้อสอบมักมีความไม่แน่นอนอันเนื่องมาจากความไม่ระมัดระวังหรือเกิดจากการเดา หลักของความน่าจะเป็นของคำตอบจะใช้ในการพิจารณาในการวัดองค์ประกอบที่เกี่ยวข้องกับความคลาดเคลื่อน ซึ่งมักจะเกิดขึ้นไม่แน่นอนในการวัดพฤติกรรมมนุษย์จึงต้องใช้วิธีการของความน่าจะเป็นมาใช้วิเคราะห์ ซึ่งจะมีแนวคิดที่สอดคล้องกับทฤษฎีการทดสอบแบบดั้งเดิม (CTT) อย่างไรก็ตาม ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) จะใช้ความน่าจะเป็นการในการวิเคราะห์คำตอบของผู้สอบว่ามีโอกาสตอบข้อสอบได้ถูกมากน้อยแค่ไหน

3) หลักการวัด (measurement principle) ความน่าจะเป็นของคำตอบทั้งผู้สอบและกลุ่มผู้สอบจะมีพิสัยอยู่ระหว่าง 0.0 - 1.0 ซึ่งจะเป็นตัวควบคุมข้อมูลที่แสดงในรูปของสัดส่วนให้อยู่ในระดับช่วงคะแนน (interval scale) การวัดค่าเป็นสัดส่วนหรือคะแนนดิบไม่สามารถเขียนอยู่ในรูปของมาตรระดับช่วงคะแนนได้ แต่อย่างไรก็ตาม ถ้าแบบสอบที่สร้างขึ้นไม่มีนักเรียนที่มีคะแนนสูงติดเพดานหรือต่ำติดพื้นและค่าคะแนนเฉลี่ยจะอยู่ใกล้จุดกึ่งกลางของพิสัย คะแนนดิบและสัดส่วนที่มาจากแบบสอบที่มีจำนวนข้อสอบมากมักจะใกล้เคียงกับมาตรช่วงคะแนน ปัญหานี้สามารถแก้โดยการใช้การแปลงสมการโลจิสติก

ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายใน (latent trait or ability) กับผลการตอบข้อสอบ โดยใช้โค้งลักษณะข้อสอบ (ICC) ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดา (c) โดยทฤษฎีการตอบสนองข้อสอบ (IRT) มีแนวคิดพื้นฐานที่สำคัญ 2 ประการคือ 1) ผลการตอบข้อสอบสามารถอธิบายได้ด้วยความสามารถที่มีอยู่ภายในของผู้ตอบและ 2)ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายในสามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (ICC) ซึ่งมีลักษณะเป็นฟังก์ชันทางคณิตศาสตร์เรียกว่าฟังก์ชันโลจิส (Logistic function) หรือใกล้เคียงกับฟังก์ชันสะสมปกติ (Normal ogive function) (ศิริชัย กาญจนวาสี, 2550)

1.1.2) หลักการและข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ (IRT)

ทฤษฎีการตอบสนองข้อสอบ (IRT) มีความเชื่อเกี่ยวกับค่าพารามิเตอร์ของข้อสอบ (item parameter) คือ ค่าความยาก (difficulty: b) ค่าอำนาจจำแนก (discrimination: a) ค่าโอกาสในการเดา (guessing: c) ของข้อสอบแต่ละข้อว่าเป็นคุณลักษณะที่ไม่แปรเปลี่ยน (parameter invariance) ในตัวข้อสอบนั้น ดังนั้นค่าพารามิเตอร์เหล่านี้จึงเป็นค่าคงที่ที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ และในทำนองเดียวกัน ค่าพารามิเตอร์ของผู้สอบ (person parameter) หรือค่าความสามารถที่แท้จริงของผู้สอบซึ่งเป็นคุณลักษณะที่มีอยู่ภายในตัวผู้สอบก็จะไม่แปรเปลี่ยนไปตามชุดข้อสอบที่เลือกใช้ด้วย แต่เนื่องจากความสามารถของผู้สอบเป็นคุณลักษณะแฝงไม่สามารถสังเกตหรือวัดได้โดยตรง (unobservable) จึงจำเป็นต้องใช้การทำนาย (predict) หรืออธิบาย (explain) คุณลักษณะดังกล่าวโดยอาศัยผลที่ได้จากการตอบแบบสอบหรือคะแนนที่ได้จากการตอบข้อสอบ ซึ่งเป็นสิ่งที่สามารถสังเกตและวัดได้โดยตรง (observable) (Hambleton & Cook, 1977; Hambleton & Swaminatan, 1985)

นักวัดผลได้พยายามหาความสัมพันธ์ระหว่างผลที่ได้จากการตอบแบบสอบหรือคะแนนที่ได้จากการตอบแบบสอบ (Test performance or score) กับระดับความสามารถ (ability) ของผู้ตอบแต่ละคน เพื่อเขียนเป็นโมเดลทางคณิตศาสตร์ (mathematical model) (Hambleton & Swaminatan, 1985) ของความสัมพันธ์ระหว่างผลการตอบแบบสอบกับระดับความสามารถของผู้สอบ โดยสามารถเขียนในรูปของความสัมพันธ์ ได้ดังสมการ

$$P = f(U_i / \theta_1, \theta_2, \theta_3, \ldots . \theta_k ; \beta_k)$$

เมื่อ P เป็นผลจากการตอบแบบสอบ (test performance)

f เป็นฟังก์ชัน (function)

U_i เป็นผลการตอบข้อสอบข้อที่ i (ตอบถูก $U_i = 1$ ตอบผิด $U_i = 0$)

$\theta_1, \theta_2, \theta_3, \ldots . \theta_k$ เป็นระดับความสามารถ (ability) ที่ 1, 2, 3,…,k

β_k เป็นค่าพารามิเตอร์ของข้อสอบข้อที่ j

เนื่องจากความสัมพันธ์ดังกล่าวเป็นเพียงฟังก์ชันความสัมพันธ์ในลักษณะทั่วๆ ไป เท่านั้นนักวัดผลจึงพยายามหาโมเดลทางคณิตศาสตร์ที่เหมาะสมกว่ามาใช้แทนฟังก์ชันความสัมพันธ์ดังกล่าว โดยอาศัยข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ (IRT) ซึ่งประกอบไปด้วย 1) ความเป็นเอกมิติ (Unidimentionality) ในที่นี้หมายถึง แบบสอบหรือข้อสอบทุกข้อในแบบสอบมุ่งวัดคุณลักษณะเพียงคุณลักษณะเดียวหรือความสามารถเดียวเท่านั้น วิธีการตรวจสอบความเป็นเอกมิติของแบบสอบนั้นสามารถทำได้หลายวิธี เช่น การวิเคราะห์องค์ประกอบ (factor analysis) โดยพิจารณา

ค่าไอเกน (eigen value) ค่าสูงสุดแตกต่างจากค่าอื่นๆ อย่างชัดเจนหรือไม่ 2) ความเป็นอิสระ (local independence) ความเป็นอิสระในที่นี้หมายถึง เมื่อกำหนด θ ให้คงที่แล้ว ผลการตอบข้อสอบแต่ละข้อจะเป็นอิสระต่อกัน ความเป็นอิสระในที่นี้จำแนกออกเป็น ความเป็นอิสระระหว่างข้อสอบและความเป็นอิสระระหว่างผู้สอบ 3) โมเดลการตอบสนองข้อสอบ (Item Response Models) ของทฤษฎีการตอบสนองข้อสอบ (IRT) ตั้งอยู่บนฐานคิดที่ว่า ฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะข้อสอบ (ICC) สามารถสะท้อนความสัมพันธ์ระหว่างความสามารถของผู้สอบกับลักษณะของข้อสอบและผลการตอบข้อสอบ ซึ่งโมเดลการตอบสนองข้อสอบที่เป็นที่นิยม ได้แก่ 1) โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ ซึ่งเป็นโมเดลที่มีเพียงพารามิเตอร์ความยาก (difficulty: b) เท่านั้นที่แปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ ส่วนพารามิเตอร์อำนาจจำแนก (discrimination: a) ถูกกำหนดให้คงที่และพารามิเตอร์โอกาสในการเดา (guessing: c) ถูกกำหนดให้เป็น 0 2) โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์เป็นโมเดลที่ค่าพารามิเตอร์ความยาก (difficulty: b) และพารามิเตอร์อำนาจจำแนก (discrimination: a) แปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ แต่พารามิเตอร์โอกาสในการเดา (guessing: c) ถูกกำหนดให้เป็น 0 และ 3) โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ เป็นโมเดลที่ค่าพารามิเตอร์ทั้ง 3 ตัว อันได้แก่ พารามิเตอร์ความยาก (difficulty: b) พารามิเตอร์อำนาจจำแนก (discrimination: a) และพารามิเตอร์โอกาสในการเดา (guessing: c)แปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ สำหรับการเลือกใช้โมเดลนั้นจะขึ้นอยู่กับจุดมุ่งหมายของงานและธรรมชาติของข้อมูลและข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ (IRT) ส่วนข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ (IRT) ข้อสุดท้าย คือ 4) การสอบที่ไม่แข่งขันด้านเวลา (Nonspeeded Test Administration) นั่นคือ เป็นการสอบที่ผู้สอบทุกคนมีเวลาในการทำข้อสอบอย่างเพียงพอหรือเป็นการสอบที่ผู้สอบสามารถแสดงความสามารถได้อย่างเต็มที่ในเวลาที่กำหนดนั่นเอง (ศิริชัย กาญจนวาสี, 2550)

1.1.3) **ฟังก์ชันสารสนเทศของข้อสอบและฟังก์ชันสารสนเทศของแบบสอบ (Item Information Function and Test Information Function)**

1.1.3.1) **ฟังก์ชันสารสนเทศของข้อสอบ** (Item Information Function)

ศิริชัย กาญจนวาสี (2550) ได้อธิบายถึงฟังก์ชันสารสนเทศของข้อสอบ (Item Information) ว่าเป็นดัชนีผสมเพื่อบ่งชี้คุณภาพของข้อสอบ ประกอบด้วยค่าพารามิเตอร์ a, b และ c ดังสมการ

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

เมื่อ $I_i(\theta)$ เป็นค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้รับจากข้อสอบข้อที่ i สำหรับผู้ตอบที่มีความสามารถ θ

$P_i'(\theta) = P_i'$ เป็นความชันของฟังก์ชันการตอบสนองข้อสอบข้อที่ i ณ ตำแหน่งความสามารถ θ

$P_i(\theta) = P_i$ เป็นความน่าจะเป็นที่ผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง และ $Q_i(\theta) = Q_i = 1 - P_i(\theta)$

ตารางที่ 2.1 การคำนวณค่าฟังก์ชันสารสนเทศของข้อสอบ ค่าสารสนเทศสูงสุดของข้อสอบ ณ ตำแหน่ง θ ที่มีสารสนเทศของข้อสอบสูงสุด สำหรับโมเดลการตอบสนองข้อสอบโลจิสแบบ 1 พารามิเตอร์ 2 พารามิเตอร์และ 3 พารามิเตอร์

ค่าประมาณ	1-parameter	2-parameter	3-parameter
$I_i(\theta)$	$D^2P_iQ_i$	$D^2a_i^2P_iQ_i$	$D^2a_i^2Q_i(P_i+c_i)^2/(1-c_i)^2$
$P_i'(\theta)$	DP_iQ_i	$Da_iP_iQ_i$	$Da_iQ_i(P_i+c_i)/(1-c_i)$
$I_i(\theta)_{max}$	$\frac{1}{4}D^2$	$\frac{1}{4}D^2a_i^2$	$D^2a_i^2\frac{\left[1-20c_i-8c_i^2+(1+8c_i)^{\frac{3}{2}}\right.}{8-(1-c_i^2)}$
θ_{max}	b_i	b_i	$b_i+\frac{1}{Da_i}\left[Ln1+\frac{(1+c_i)^{\frac{1}{2}}}{2}\right.$

ที่มา: ศิริชัย กาญจนวาสี (2550)

1.1.3.2) **ฟังก์ชันสารสนเทศของแบบสอบ** (Test Information Function)

การวิเคราะห์ตามแนวคิดทฤษฎีการตอบสนองข้อสอบ (IRT) จะใช้แบบแผนการตอบสนองแบบสอบเป็นรายข้อในการประมาณค่าความสามารถของผู้สอบ ดังนั้น การประเมินคุณภาพของแบบสอบจึงสามารถพิจาณาความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้ตอบ โดยใช้สารสนเทศของแบบสอบ (Test Information; $I(\theta)$) ซึ่งเป็นค่าฟังก์ชันสารสนเทศของแบบ

สอบอันเกิดจากผลรวมเชิงพีชคณิตของฟังก์ชันสารสนเทศของข้อสอบ (Item Information) แต่ละข้อรวมเข้าด้วยกันทั้งฉบับ ณ ตำแหน่ง θ เดียวกัน ดังสมการ

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

เมื่อ $I(\theta)$ เป็นค่าฟังก์ชันสารสนเทศที่ได้รับจากแบบสอบสำหรับผู้ตอบที่มีความสามารถ θ

1.1.4) โมเดลการตอบสนองข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (Dichotomous Item Response Theory)

ข้อสอบหรือแบบสอบที่ตรวจให้คะแนนแบบ 2 ค่า (Dichotomous) เป็นวิธีการให้คะแนนที่นิยมใช้กันอย่างแพร่หลายเพราะมีความสะดวก รวดเร็วและประหยัดเวลาในการตรวจ การตรวจให้คะแนนแบบ 2 ค่า (dichotomous) เหมาะกับข้อสอบประเภทเลือกตอบ เช่น ข้อสอบแบบจับคู่ (matching) ข้อสอบแบบถูก – ผิด (true false) และข้อสอบแบบหลายตัวเลือก (multiple choice) (Wiersma & Jurs, 1990) การตรวจให้คะแนนแบบ 2 ค่า (dichotomous) จะกำหนดตัวเลือกหรือคำตอบที่เหมาะสมและถูกต้องที่สุดไว้ 1 คำตอบ หากผู้ตอบข้อสอบเลือกคำตอบถูกนี้ก็จะได้ 1 คะแนน แต่หากผู้ตอบข้อสอบเลือกคำตอบข้ออื่นๆ ก็จะได้ 0 คะแนน หรือหากข้อสอบมีเพียง 2 ตัวเลือก หากผู้ตอบข้อสอบเลือกคำตอบที่ถูกต้องก็จะได้ 1 คะแนนและหากผู้ตอบข้อสอบเลือกคำตอบอีกตัวเลือกหนึ่งที่เหลือก็จะได้ 0 คะแนน

สำหรับข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (Dichotomous) นี้ โมเดลการตอบสนองข้อสอบ (IRT) ที่นิยมใช้เป็นโมเดลโลจิส (Logistic model) ซึ่งมีวิธีคำนวณที่ง่ายและสะดวกกว่าโมเดลปกติสะสม (Normal ogive model) ซึ่งโมเดลการตอบสนองข้อสอบ (IRT Model) ที่เป็นที่นิยม ได้แก่ โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์และโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ดังรายละเอียดต่อไปนี้

1.1.4.1) โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (One-Parameter IRT Model)

โมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ หรือ Rasch Model มีโค้งลักษณะข้อสอบ (ICC) ที่เขียนด้วยฟังก์ชันโลจิส ดังสมการ

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

เมื่อ $P_i(\theta)$ เป็นความน่าจะเป็นที่ผู้ตอบที่มีความสามารถ θ จะตอบข้อสอบข้อ i ได้ถูกต้อง

b_i เป็นค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ ตำแหน่ง θ ที่มีโอกาสตอบข้อสอบถูก 0.50

e มีค่าเป็น 2.718

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ มีลักษณะที่สำคัญ คือ มีเพียงพารามิเตอร์ความยาก (difficulty: b) เท่านั้นที่แปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ ส่วนพารามิเตอร์อำนาจจำแนก (discrimination: a) ถูกกำหนดให้คงที่และพารามิเตอร์โอกาสในการเดา (guessing: c) นั้นถูกกำหนดให้เป็น 0

1.1.4.2) **โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์** (Two-Parameter Model)

โมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ มีโค้งลักษณะข้อสอบ (ICC) ที่เขียนด้วยฟังก์ชันโลจิส ดังสมการ

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

เมื่อ $P_i(\theta)$ เป็นความน่าจะเป็นที่ผู้ตอบที่มีความสามารถ θ จะตอบข้อสอบข้อ i ได้ถูกต้อง

b_i เป็นค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ ตำแหน่ง θ ที่มีโอกาสตอบข้อสอบถูก 0.50

a_i เป็นค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นค่าความชันของ ICC ณ ตำแหน่ง b_i

e มีค่าเป็น 2.718

D มีค่าเป็น 1.70

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 2 พารามิเตอร์ มีลักษณะที่สำคัญ คือ ค่าพารามิเตอร์ความยาก (difficulty: b) และพารามิเตอร์อำนาจจำแนก (discrimination: a) จะแปรเปลี่ยนตามลักษณะของข้อสอบแต่ละข้อ แต่พารามิเตอร์โอกาสในการเดา (guessing: c) นั้นถูกกำหนดให้เป็น 0

1.1.4.3) โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ (Three-Parameter Model)

โมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ มีโค้งลักษณะข้อสอบ (ICC) ที่เขียนด้วยฟังก์ชันโลจิส ดังสมการ

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

เมื่อ $P_i(\theta)$ เป็นความน่าจะเป็นที่ผู้ตอบที่มีความสามารถ θ จะตอบข้อสอบข้อ i ได้ถูกต้อง

b_i เป็นค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ ตำแหน่ง θ ที่มีโอกาสตอบข้อสอบถูก $\frac{1+c_i}{2}$

a_i เป็นค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นค่าความชันของ ICC ณ ตำแหน่ง b_i

c_i เป็นค่าพารามิเตอร์โอกาสการเดาข้อสอบได้ถูก

e มีค่าเป็น 2.718

D มีค่าเป็น 1.70

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ มีลักษณะที่สำคัญ คือ ค่าพารามิเตอร์ทั้ง 3 ตัว อันได้แก่ พารามิเตอร์ความยาก (b) พารามิเตอร์อำนาจจำแนก (a) และพารามิเตอร์โอกาสในการเดา (c) จะแปรเปลี่ยนไปตามลักษณะของข้อสอบแต่ละข้อ

1.1.5) สรุปสาระสำคัญของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นทฤษฎีที่เกิดขึ้นเพื่อแก้ไขข้อจำกัดของทฤษฎีการทดสอบแบบดั้งเดิม (CTT) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายใน (latent trait or ability) กับผลการตอบข้อสอบ โดยใช้โค้งลักษณะข้อสอบ (ICC) ซึ่งกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (difficulty: b) อำนาจจำแนก (discrimination: a) และโอกาสการเดา (guessing: c) โดยทฤษฎีการตอบสนองข้อสอบ (IRT) นี้มีฐานคิดที่สำคัญ 2 ประการคือ 1) ผลการตอบข้อสอบ สามารถอธิบายได้ด้วยความสามารถที่มีอยู่ภายในของผู้ตอบและ 2)ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายในสามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบหรือโค้งลักษณะ

ข้อสอบ (ICC) ซึ่งมีลักษณะเป็นฟังก์ชันโลจิส (Logistic function) ที่สามารถสะท้อนความสัมพันธ์ระหว่างความสามารถของผู้สอบกับลักษณะของข้อสอบและผลการตอบข้อสอบของผู้สอบได้

ตอนที่ 2 มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ (DIF)

2.1) มโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ (DIF)

ขอนำเสนอมโนทัศน์ของการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยเริ่มต้นจาก 2.1.1) ประวัติของการทำหน้าที่ต่างกันของข้อสอบ (DIF) 2.1.2) ความหมายของการทำหน้าที่ต่างกันของข้อสอบ (DIF) 2.1.3) รูปแบบของการทำหน้าที่ต่างกันของข้อสอบ (DIF) และ 2.1.4) วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ดังรายละเอียดต่อไปนี้

2.1.1) ประวัติของการทำหน้าที่ต่างกันของข้อสอบ (DIF)

การศึกษาเกี่ยวกับความยุติธรรมของแบบสอบ (Test fairness) มีมากว่าศตวรรษแล้ว(McNamara & Roever, 2004) โดย McNamara และ Roever กล่าวไว้ในบทความ Psychometric Approaches to Fairness ว่า การศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ (DIF) เริ่มต้นขึ้นในปี ค.ศ. 1905 จากการที่ Binet และคณะ (Binet & Simon, 1916 อ้างถึงใน McNamara & Roever, 2004) ได้มีการโต้เถียงเกี่ยวกับแบบวัดเชาวน์ปัญญาว่าขาดความยุติธรรมสำหรับผู้เข้าสอบทุกคน นั่นคือ นักเรียนที่มีสถานะทางเศรษฐกิจและสังคมสูงมีผลการสอบดีกว่านักเรียนที่มีสถานะทางเศรษฐกิจและสังคมต่ำอย่างมาก Binet จึงปรับปรุงแก้ไขแบบสอบดังกล่าว โดยพิจารณาข้อสอบที่ขาดความยุติธรรมสำหรับผู้เข้าสอบแล้วตัดออกจากแบบสอบนั้น ต่อมาในปี ค.ศ. 1914 Stern (Stern, 1914 อ้างถึงใน McNamara & Roever, 2004) เป็นบุคคลแรกที่ศึกษาเกี่ยวกับความยุติธรรมของข้อสอบ (test fairness) เพื่อใช้ในการคัดเลือกข้อสอบที่มีความยากมากหรือง่ายมากออกจากแบบสอบสำหรับผู้สอบที่มาจากครอบครัวที่มีสถานภาพทางเศรษฐานะสูงและผู้สอบที่มาจากครอบครัวที่มีสถานภาพทางเศรษฐานะต่ำ จากนั้นในปี ค.ศ. 1912 Weintrob และ Weintrob (Weintrob & Weintrob, 1912 อ้างถึงใน McNamara & Roever, 2004) เป็นบุคคลแรกที่ใช้เชื้อชาติ (race) เป็นตัวแปรหนึ่งในการศึกษาความยุติธรรมของแบบทดสอบ (test fairness) ซึ่งหลังจากนั้น งานวิจัยนี้ได้กลายเป็นประเด็นสำคัญในการศึกษาเรื่องความลำเอียง (bias) ของข้อสอบหรือแบบสอบในประเทศสหรัฐอเมริกา ทั้งนี้ก็เพราะว่าประเทศสหรัฐอเมริกาเป็นประเทศที่มีความแตกต่างของประชากรเกี่ยวกับเชื้อชาติ เผ่าพันธุ์ ศาสนา ภาษา วัฒนธรรม ประสบการณ์ สถานะทางเศรษฐกิจและสังคม หรือภูมิหลังอื่นๆ และวงการศึกษาและหน่วยงานต่างๆ ได้ให้ความสำคัญกับการทดสอบและนำผลการทดสอบมาใช้ในการพิจารณาตัดสินในเรื่องต่างๆ

มากยิ่งขึ้น เช่น การคัดเลือกนักเรียนเข้าศึกษาต่อ การคัดเลือกบุคคลเข้าทำงาน การเลื่อนขั้น การปรับตำแหน่ง การออกใบอนุญาตหรือใบรับรองต่างๆ ซึ่งแบบสอบที่นำมาใช้จึงน่าจะเป็นแบบสอบมาตรฐานที่มีความตรง (validity) ความเที่ยง (reliability) มีค่าความยากพอเหมาะและมีอำนาจจำแนกสูง นั่นคือ ต้องเป็นแบบสอบที่สามารถจำแนกผู้สอบได้ดีและเรียงลำดับความสามารถของผู้สอบแต่ละคนได้อย่างถูกต้องและแม่นยำ (Ackerman, 1992) ทั้งนี้ก็เพราะว่าหากข้อสอบหรือแบบสอบขาดความยุติธรรมหรือเกิดความลำเอียง โดยข้อสอบหรือแบบทดสอบอาจเข้าข้างผู้สอบกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ ซึ่งทำให้ผู้สอบกลุ่มนั้นได้เปรียบในการตอบข้อสอบ โดยมีโอกาสตอบข้อสอบถูกได้มากกว่าอีกกลุ่มหนึ่ง ทั้งๆ ที่ผู้สอบทั้งสองกลุ่มมีความสามารถระดับเดียวกัน ซึ่งความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มดังกล่าวนี้เป็นผลมาจากความแตกต่างของวัฒนธรรม ซึ่งเป็นภูมิหลังของผู้สอบมากกว่าความสามารถของผู้สอบ (Angoff, 1993 อ้างถึงใน McNamara & Roever, 2004; สุมาลี แกล้วทนงค์, 2547) ความลำเอียงของข้อสอบหรือแบบสอบ (item bias/test bias) นี้เป็นความคลาดเคลื่อนอย่างเป็นระบบ (systematic error) (Camilli & Shepard, 1994) ซึ่งจะส่งผลให้แบบสอบขาดความตรง (validity) ได้

ในช่วงปลายศตวรรษที่ 20 ประเด็นการศึกษาเกี่ยวกับความลำเอียง (bias) ของข้อสอบได้ปรับเปลี่ยนจากการศึกษาเกี่ยวกับแบบวัดเชาวน์ปัญญา (IQ test) มาเป็นการศึกษาเกี่ยวกับความเท่าเทียมทางสังคม (social equity) และความเท่าเทียมทางการศึกษามากยิ่งขึ้น นอกจากนี้ยังมีการศึกษาตัวแปรอื่นๆ ที่เกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ (DIF) มากขึ้น เป็นต้นว่า ตัวแปรเชื้อชาติ (race) ตัวแปรเพศ (gender) ตัวแปรภาษา ตัวแปรอายุ ตัวแปรประสบการณ์ ตัวแปรความแตกต่างทางภูมิลำเนา รวมไปถึงตัวแปรความแตกต่างทางวัฒนธรรม (cultural difference) (McNamara & Roever, 2004; ศิริชัย กาญจนวาสี, 2550) และการศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้รับความนิยมตั้งแต่ปี ค.ศ. 1960 เป็นต้นมาจนถึงปัจจุบัน (Mapuranga, Dorans & Middleton, 2008)

2.1.2) ความหมายของการทำหน้าที่ต่างกันของข้อสอบ (DIF)

เอกสาร Standards for Educational and Psychological Testing (AERA, APA, NCME, 1999) ไม่ได้ให้ความหมายของคำว่าการทำหน้าที่ต่างกันของข้อสอบ (DIF) ไว้แต่อย่างใด แต่ได้ให้ความหมายของความยุติธรรมของแบบสอบ (test fairness) ไว้ว่า ความยุติธรรมของแบบสอบ (test fairness) ไม่ได้มีความหมายเดียว (single meaning) ดังนั้นจึงไม่สามารถกำหนดเป็นนิยามที่มีความหมายเฉพาะได้ แต่ได้ให้ข้อสังเกตแนวคิดของความยุติธรรมของแบบสอบไว้ 4 ลักษณะ คือ 1) แบบทดสอบที่มีความยุติธรรมจะต้องปราศจากความลำเอียง (no bias) ความลำเอียงเกิดขึ้นเมื่อแปลคะแนนของแบบสอบมีความหมายแตกต่างกัน ซึ่งเป็นผลมาจากความแตกต่างของสมาชิกในกลุ่ม 2) แบบ

สอบที่มีความยุติธรรมเป็นแบบสอบที่ให้ผู้สอบทั้งหมดได้รับการปฏิบัติในกระบวนการทดสอบเท่ากัน (equitable treatment in process) ซึ่งเป็นเงื่อนไขของการสอบที่ทำให้ผู้สอบมีโอกาสเท่ากัน เช่น รูปแบบของแบบสอบ วัสดุที่ใช้ในการฝึก ความสะดวกของผู้สอบที่ทุพพลภาพ ตลอดจนการดำเนินการสอบ เป็นต้น 3) แบบสอบที่มีความยุติธรรมเป็นแบบสอบที่ให้ผลของการทดสอบที่มีความเสมอภาค (equal outcomes) ซึ่งผู้สอบทุกคนมีโอกาสแสดงศักยภาพตามโครงสร้างของแบบสอบ เมื่อผู้สอบที่มีความสามารถเท่ากันและถูกวัดด้วยแบบสอบเดียวกันแล้วย่อมได้คะแนนเท่ากัน และ 4) แบบสอบที่มีความยุติธรรมจะต้องให้ผู้สอบมีโอกาสได้เรียนรู้เนื้อหาของแบบทดสอบอย่างครอบคลุม (equal opportunity to learn content) แล้ว

ในปัจจุบันงานวิจัยโดยทั่วไปใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (DIF) แทนคำว่า ความลำเอียงของข้อสอบ (item bias) และใช้คำว่า การทำหน้าที่ต่างกันของแบบสอบ (differential test functioning: DTF) แทนคำว่า ความลำเอียงของแบบสอบ (test bias) เนื่องจากความลำเอียงของข้อสอบหรือแบบทดสอบนั้นมีความหมายในเชิงสถิติและความหมายเชิงสังคม ซึ่งทำให้สับสนในการนำไปใช้ (Hambleton, Swaminathan & Rogers, 1991) สำหรับความหมายของการทำหน้าที่ต่างกันของข้อสอบ (DIF) นั้น มีนักวิจัยหลายคนได้กำหนดนิยามไว้ ดังตารางที่ 2.2

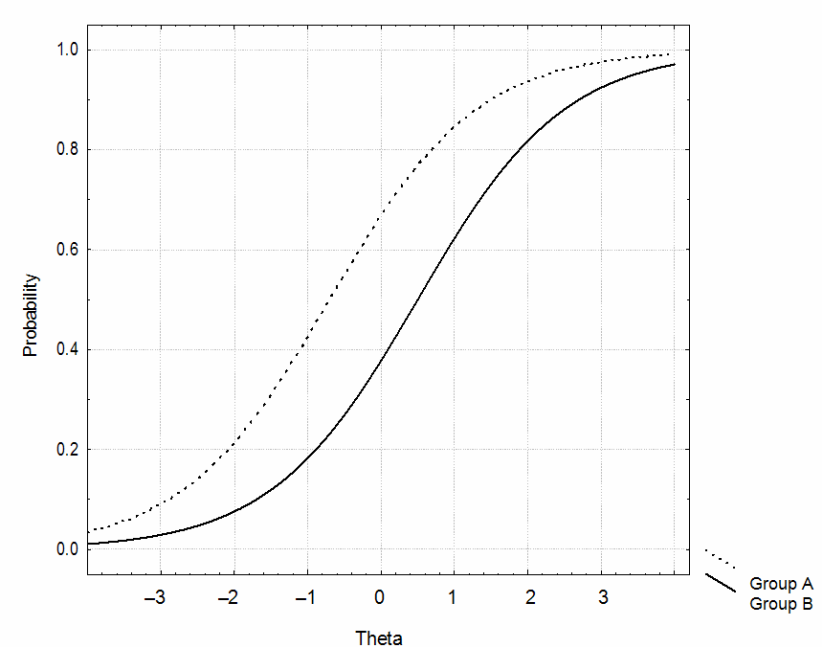
ตารางที่ 2.2 นิยามของการทำหน้าที่ต่างกันของข้อสอบ (DIF)

นักวิจัย (ปี)	นิยาม
Hambleton, Swaminathan & Rogers (1991)	ผู้สอบที่มีความสามารถระดับเดียวกัน แต่มาจากกลุ่มผู้สอบที่แตกต่างกัน มีความน่าจะเป็นในการตอบข้อสอบถูกไม่เท่ากัน
Angoff (1993)	ข้อสอบแสดงคุณลักษณะทางสถิติที่แตกต่างกันจากการตอบข้อสอบจากกลุ่มผู้สอบแตกต่างกัน
Camilli (1993)	เป็นการเปรียบเทียบความสามารถระหว่างผู้สอบสองกลุ่มหรือมากกว่า โดยความน่าจะเป็นในการตอบข้อสอบถูกมีความสัมพันธ์กับการเป็นสมาชิกของกลุ่มผู้สอบ
Dorans & Holland (1993)	ความแตกต่างในการทำหน้าที่ของข้อสอบภายหลังการจับคู่ความสามารถของกลุ่มผู้สอบ ซึ่งเป็นความสามารถตามที่ข้อสอบต้องการวัด
Narayanan & Swaminathan (1996)	ฟังก์ชันการตอบสนองข้อสอบ ซึ่งคำนวณจากสมาชิกของผู้สอบกลุ่มย่อยที่แตกต่างกัน มีค่าไม่เท่ากัน
Roussos & Stout (1996)	ผู้สอบที่มีความสามารถระดับเดียวกันแต่มาจากประชากรหลายกลุ่มที่แตกต่างกันมีความน่าจะเป็นในการตอบข้อสอบถูกแตกต่างกัน
Clauser & Mazor (1998)	ผู้สอบจากกลุ่มที่แตกต่างกัน มีความน่าจะเป็นที่จะตอบข้อสอบถูกแตกต่างกันภายหลังการจับคู่ความสามารถตามที่ข้อสอบต้องการวัด
Bolt (2002)	ผู้สอบจากกลุ่มประชากรสองกลุ่มที่แตกต่างกันภายใต้ระดับความสามารถเดียวกัน มีความน่าจะเป็นที่จะตอบข้อสอบได้ถูกไม่เท่ากัน
Kamata (2009)	ผู้สอบที่มาจากสองกลุ่มย่อยที่ต่างกันและมีระดับความสามารถที่มุ่งวัดเดียวกัน แต่มีความน่าจะเป็นที่จะตอบข้อสอบได้ถูกไม่เท่ากัน
ศิริชัย กาญจนวาสี (2550)	การที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกัน

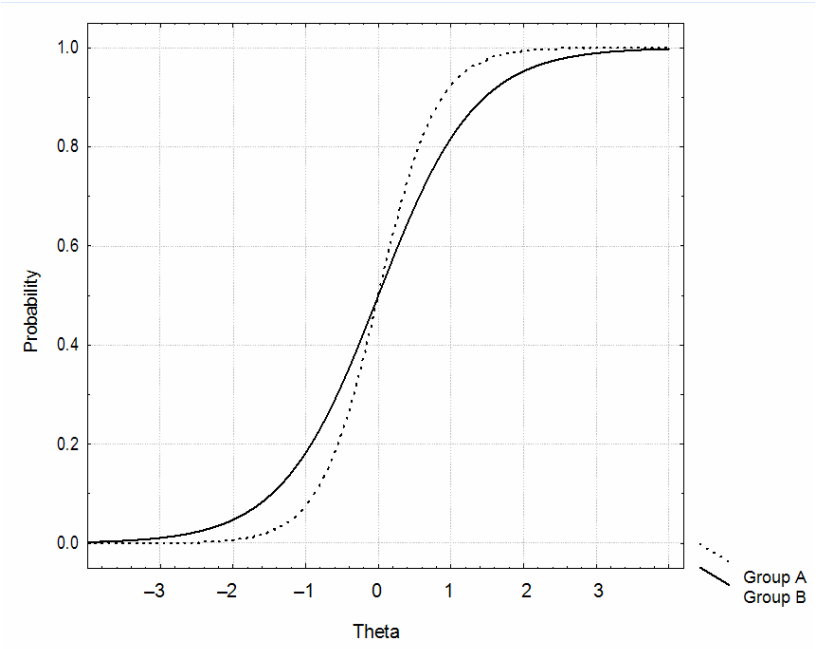
จากนิยามทั้งหมดข้างต้น สรุปได้ว่าการทำหน้าที่ต่างกันของข้อสอบ หมายถึง การที่ผู้สอบมีระดับความสามารถเดียวกันแต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน เนื่องจากมาจากกลุ่มที่แตกต่างกัน

2.1.3) รูปแบบของการทำหน้าที่ต่างกันของข้อสอบ (DIF)

Mellenbergh (1982) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) โดยใช้ log-linear model และ logit model ในการทำนายผลการตอบข้อสอบ ภายใต้ตารางการณ์จร (contingency table) แบบสามมิติ ในการศึกษาสามปัจจัยหลัก อันได้แก่ 1) การเป็นสมาชิกของกลุ่ม (group membership) 2) ระดับความสามารถ (ability level) และ3) ปฏิสัมพันธ์ (interaction) ระหว่างการเป็นสมาชิกของกลุ่มกับระดับความสามารถ จากการศึกษาในครั้งนี้ Mellenbergh ได้จำแนกข้อสอบทำหน้าที่ต่างกันออกเป็นสองรูปแบบ คือ ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) และข้อสอบที่หน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF) จะเกิดขึ้นเมื่อไม่มีปฏิสัมพันธ์ (interaction) ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่ม หมายความว่า ความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่งอย่างคงที่ในทุกระดับความสามารถ ส่วนการทำหน้าที่ต่างกันของข้อสอบแบบเอนกรูป (Nonuniform DIF) เกิดขึ้นเมื่อมีปฏิสัมพันธ์ (interaction) ระหว่างระดับความสามารถกับการเป็นสมาชิกของกลุ่ม แสดงว่าความแตกต่างของความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบสองกลุ่มไม่เหมือนกันทุกระดับความสามารถ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF) การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF) แสดงได้ดังแผนภาพที่ 2.1 และ 2.2 ตามลำดับ



แผนภาพที่ 2.1 การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF)



แผนภาพที่ 2.2 การทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF)

โดยข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) นั้น ฟังก์ชันการตอบข้อสอบระหว่างกลุ่มผู้สอบซึ่งไม่ขนานกันนั้นอาจจะตัดกันหรือไม่ตัดกันก็ได้ โดยจากศึกษาของ Swaminathan & Rogers (1990) พบว่า ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) มี 2 ลักษณะ คือ ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปแบบมีปฏิสัมพันธ์เป็นลำดับ (ordinal interaction) และ ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปแบบมีปฏิสัมพันธ์ไม่เป็นลำดับ (disordinal interaction) นั่นคือ ในช่วงความสามารถตามมาตรของทฤษฎีการตอบสนองข้อสอบ เมื่อฟังก์ชันการตอบข้อสอบระหว่างกลุ่มผู้สอบตัดกันตรงกึ่งกลางของช่วงความสามารถ ซึ่งเมื่อวิเคราะห์ด้วยการวิเคราะห์ความแปรปรวน (ANOVA) จะพบว่า มีลักษณะเป็นปฏิสัมพันธ์ไม่เป็นลำดับ (disordinal interaction) และเมื่อฟังก์ชันการตอบข้อสอบระหว่างกลุ่มตัดกันนอกช่วงความสามารถหรือฟังก์ชันการตอบข้อสอบระหว่างกลุ่มซึ่งไม่ขนานกันแต่ไม่ตัดกันจะเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปแบบมีปฏิสัมพันธ์เป็นลำดับ (ordinal interaction) ต่อมา Stout และ Li (1993, อ้างถึงใน Narayanan & Swaminathan, 1996) เรียกข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปแบบมีปฏิสัมพันธ์เป็นลำดับ (ordinal interaction) ว่า ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทาง (Unidirectional DIF) และเรียกข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปแบบมีปฏิสัมพันธ์ไม่เป็นลำดับ (disordinal interaction) ว่า ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (nondirectional DIF)

2.1.4) วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

จากอดีตถึงปัจจุบัน นักวัดผลได้พยายามคิดค้น ปรับปรุงและพัฒนาศาสตร์ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) มาอย่างต่อเนื่องและสม่ำเสมอ เพื่อให้ผลการตรวจสอบที่ได้มีความถูกต้องและแม่นยำมากที่สุด โดยวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) เริ่มต้นจากฐานคิดของการตรวจสอบในแบบสอบที่มุ่งวัดความสามารถแบบเอกมิติ (Unidimensional) จากข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) โดยใช้วิธีการตรวจสอบที่ง่ายและไม่ซับซ้อน ไปถึงการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ตรวจให้คะแนนแบบหลายค่า (polytomous) ที่มีวิธีการตรวจสอบที่ซับซ้อนและละเอียดอ่อนมากยิ่งขึ้น ต่อมามีการพัฒนาโมเดลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มุ่งวัดความสามารถพหุมิติ (Multidimensional) ขึ้น ประกอบกับความเจริญก้าวหน้าทางวิทยาศาสตร์และเทคโนโลยีจึงมีการพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการวิเคราะห์และตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่มีประสิทธิภาพในการตรวจสอบและเป็นที่นิยมกันอย่างแพร่หลายมาถึงปัจจุบันนี้

แม้ว่าแนวคิดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จะมีมาตั้งแต่ปี ค.ศ. 1905 (McNamara & Roever, 2004) แต่วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) อย่างเป็นระบบเพิ่งมีขึ้นในปี ค.ศ. 1964 จากการศึกษาของ Cardall & Coffman (อรินทร์ น่วมถนอม, 2549) โดยประยุกต์กระบวนการวิเคราะห์ความแปรปรวน (analysis of variance) ทดสอบปฏิกิริยาร่วมของข้อสอบที่ใช้กับผู้ตอบข้อสอบผิวดำและผิวขาวที่เข้ารับการทดสอบ SAT และพัฒนามาเรื่อยๆ จนถึงปัจจุบันนี้ กระบวนการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสามารถทำได้ 2 ขั้นตอน ดังนี้ (อุทุมพร จามรมาน, 2537) ขั้นตอนแรกเป็นการตรวจสอบกระบวนการสร้างข้อสอบโดยให้ผู้เชี่ยวชาญพิจารณาว่าข้อสอบข้อใดตรงตามวัตถุประสงค์และมีเนื้อหาสอดคล้องกับตารางวิเคราะห์ข้อสอบหรือไม่ รวมทั้งพิจารณาความชัดเจนของภาษาที่ใช้และขจัดความคลุมเครือของภาษาที่อาจทำให้เกิดประโยชน์แก่ผู้ตอบข้อสอบกลุ่มใดกลุ่มหนึ่ง และขั้นตอนที่สองเป็นการตรวจสอบโดยใช้วิธีการทางสถิติ ซึ่งจะดำเนินการภายหลังการนำแบบทดสอบไปใช้แล้ว โดยสามารถทำได้ 2 แนวทาง คือ 1) การใช้เกณฑ์ภายนอก (external criterion) เป็นการใช้ข้อมูลการตอบแบบสอบที่สร้างขึ้นแล้วหาความสัมพันธ์กับเกณฑ์ที่เป็นมาตรฐาน เช่น เกรดเฉลี่ยหรือผลการปฏิบัติงาน เป็นต้น และ 2) การใช้เกณฑ์ภายใน (internal criterion) ศึกษาจากการตรวจสอบโครงสร้างภายในของแบบสอบ โดยพิจารณาจากคะแนนของผู้ตอบข้อสอบแต่ละกลุ่มว่าวัดคุณลักษณะตามโครงสร้างเดียวกันหรือไม่ ถือเป็นกระบวนการในการสนับสนุนการตรวจสอบความตรงเชิงโครงสร้าง (construct validity) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยวิธีการทางสถิติที่อาศัยเกณฑ์ภายในแบ่งออกเป็น 3 กลุ่มสำคัญ (Camilli & Shepard, 1994) คือ 1) การ

ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ได้แก่ 1.1) วิธีการแปลงค่าความยาก (transformation item difficulty) 1.2) วิธีการวิเคราะห์ความแปรปรวน (analysis of variance) และ 1.3) วิธีพอยท์ไบซีเรียล (point biserial) 2) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้ตารางการณ์จร (contingency table) ได้แก่ 2.1) วิธีแมนเทล-แฮนส์เซล (Mantel–Haenszel) 2.2) วิธีไคสแควร์ (Chi–Square) และ 2.3) วิธีทำให้เป็นมาตรฐาน (standardization Method) และ 3) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามแนวคิดของทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นแนวคิดที่พิจารณาผลการตอบข้อสอบกับโอกาสการตอบข้อสอบถูกของผู้ที่มีความสามารถระดับเดียวกันแต่อยู่ต่างกลุ่มกัน เช่น กลุ่มเพศหญิงและกลุ่มเพศชาย ถ้าโค้งคุณลักษณะข้อสอบ (ICC) มีความแตกต่างกันแสดงว่าข้อสอบนั้นทำหน้าที่ต่างกันทำให้เกิดการได้เปรียบหรือเสียเปรียบกันขึ้นระหว่างกลุ่ม วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีนักการศึกษาได้นำเสนอไว้หลายวิธี ดังนี้

ETS โดย Mapuranga, Dorans และ Middleton (2008) สรุปวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยแบ่งออกเป็น 4 วิธี ดังนี้ 1) วิธีการหาคะแนนรายข้อที่คาดหวัง (expected item score methods) ประกอบด้วย วิธีซิบเทสท์ (SIBTEST) วิธี Kernel Smoothed SIBTEST และ วิธีมัลติซิบ (MULTISIB) 2) วิธีการใช้นันพาราเมทริกแบบอัตราแต้มต่อ (nonparametric odds ratio methods) โดยการวิเคราะห์ด้วยวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) 3) วิธีการใช้โมเดลเชิงเส้นทั่วไป (generalized linear model methods) ซึ่งประกอบด้วย 1) วิธีการถดถอยโลจิสติก (Logistic Regression) 2) วิธี logistic mixed model 3) วิธี mixture models 4) วิธี hierarchical generalized linear model (HGLM) 5) วิธี hierarchical logistic regression (HLR) และ 4) วิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบเป็นฐาน (IRT based methods) ซึ่งประกอบด้วย 1) วิธี TESTGRAF และ 2) วิธี RCML

Wiberg (2007) สรุปวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ไว้ 2 วิธี คือ 1) การใช้วิธีการของสถิตินันพาราเมทริก (nonparametric methods) ประกอบด้วย 1) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และ 2) วิธีซิบเทสท์ (SIBTEST) และ 2) การใช้วิธีการของสถิติพาราเมทริก (parametric methods) ประกอบด้วย 1) วิธีการถดถอยโลจิสติก (Logistic Regression) 2) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) 3) วิธีที่ใช้ระเบียบวิธีของทฤษฎีการตอบสนองข้อสอบ (IRT methods) ซึ่งประกอบด้วย 4 วิธี คือ 3.1) general IRT-LR 3.2) loglinear IRT-LR 3.3)

limited information IRT-LR และ 3.4) IRT-D2 4) วิธี Log linear models และ 5) วิธี Mixed Effect Models

McNamara และ Roever (2004) สรุปวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ไว้ 4 วิธี คือ 1) การวิเคราะห์ค่าความยากของข้อสอบ (Analyses based on item difficulty) 2) การใช้กระบวนการของสถิตินันพาราเมทริก (nonparametric approaches) ประกอบด้วย 1) ไคว์-สแควร์ (chi-square) 2) การใช้อัตราแต้มต่อ (odds ratios) 3) การใช้กระบวนการของทฤษฎีการตอบสนองข้อสอบ (Item response theory based approaches) ซึ่งประกอบด้วยการวิเคราะห์ 1, 2 หรือ 3 พารามิเตอร์ และ 4) การใช้กระบวนการอื่นๆ (other approaches) เช่น การวิเคราะห์การถดถอยโลจิสติก (Logistic Regression)

ศิริชัย กาญจนวาสี (2550) กล่าวว่า วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) นั้น จำแนกได้หลายลักษณะตามเกณฑ์ที่ใช้ เช่น ใช้เกณฑ์การตรวจให้คะแนนของข้อสอบจะแบ่งได้ 2 แบบ คือ การตรวจให้คะแนนแบบ 2 ค่า (dichotomous DIF methods) และการตรวจให้คะแนนแบบหลายค่า (polytomous DIF methods) หากใช้เกณฑ์ทฤษฎีของการวิเคราะห์จะแบ่งได้ 2 แบบ คือ กลุ่ม Non-IRT (non item response theory methods) ซึ่งจะวิเคราะห์ดัชนีการทำหน้าที่ต่างกันของข้อสอบโดยใช้คะแนนที่สังเกตได้ภายใต้ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) และ กลุ่ม IRT (IRT methods) จะวิเคราะห์โดยใช้ตัวแปรแฝงภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) แต่หากใช้ข้อตกลงเบื้องต้นของโมเดลเป็นเกณฑ์จะสามารถแบ่งออกเป็น 2 แบบ คือ รูปแบบพาราเมทริก (parametric form) ซึ่งจะวิเคราะห์ดัชนีการทำหน้าที่ต่างกันของข้อสอบที่มีข้อตกลงเบื้องต้นของโมเดลสำหรับอธิบายความสัมพันธ์ระหว่างคะแนนของข้อสอบและการจับคู่ตัวแปรและรูปแบบนันพาราเมทริก (nonparametric form) ซึ่งจะไม่มีข้อตกลงเบื้องต้นดังกล่าว สรุปได้ดังตารางที่ 2.3

ตารางที่ 2.3 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

วิธีการตรวจสอบ	รูปแบบพาราเมทริก (parametric form)	รูปแบบนันพาราเมทริก (nonparametric form)
กลุ่ม Non-IRT	1) วิธีวิเคราะห์ความแปรปรวน 2) วิธีการถดถอยโลจิสติก	1) วิธีแปลงค่าความยากของข้อสอบ 2) วิธีไค-สแควร์ 3) วิธีลอก-ลิเนียร์ 4) วิธีแมนเทล-แฮนส์เซล
กลุ่ม IRT	1) วิธีการวัดพื้นที่ 1.1) วิธีการวัดพื้นที่ของ Raju 1.2) วิธีการวัดพื้นที่ของ Kim และ Cohen 2) วิธีเปรียบเทียบค่าพารามิเตอร์ 2.1) วิธีเปลี่ยนค่าความยาก 2.2) วิธีการทดสอบ F 2.3) วิธีทดสอบไค-สแควร์ของ Lord 2.4) วิธี IRT เทียม 2.5) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด	1) วิธีซิบเทสท์

ที่มา: ศิริชัย กาญจนวาสี (2550)

จากแนวคิดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (Dichotomous) สามารถแบ่งออกเป็นสองกลุ่มใหญ่ๆ คือ กลุ่มวิธีที่ใช้ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) และกลุ่มวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) (ศิริชัย กาญจนวาสี, 2550) โดยกลุ่มที่ใช้ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) จะใช้คะแนนที่ได้จากการสอบเป็นตัวแปรการจับคู่ (matching variable) ความสามารถระหว่างผู้สอบกลุ่มอ้างอิง (reference group) และกลุ่มเปรียบเทียบ (focal group) ซึ่งจะใช้คะแนนรวมของแบบสอบแทนระดับความสามารถของผู้สอบ วิธีที่สำคัญในกลุ่มนี้ ได้แก่ 1) วิธีการวิเคราะห์ความแปรปรวน (ANOVA) 2) วิธีแปลงค่าความยากของข้อสอบ (transformed item difficulty; TID) 3) วิธีไค-สแควร์ (Chi-square) 4) วิธีล็อก-ลิเนียร์ (Log-linear) 5) วิธีแมนเทล-แฮนส์เซล

(Mantel-Heanszel; MH) 6) วิธีการทำให้เป็นมาตรฐาน (Standardization; STD) 7) วิธีการถดถอยโลจิสติก (Logistic regression; LR) เป็นต้น ส่วนกลุ่มวิธีที่ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) จะไม่ใช้คะแนนรวมเป็นเกณฑ์การจับคู่แต่จะใช้ค่าประมาณความสามารถของผู้สอบ (θ) ซึ่งเป็นตัวแปรแฝง (latent variable) เป็นเกณฑ์การจับคู่ วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในกลุ่มนี้ ได้แก่ 1) วิธีไค-สแควร์ของลอร์ด 2) วิธีอัตราส่วนความน่าจะเป็นของ Thissen Steinberge และ Wainer 3) วิธีการวัดพื้นที่ของ Shepard, Camilli และ Williams 4) วิธีการวัดพื้นที่ของ Raju 5) วิธีการวัดพื้นที่ของ Kim และ Cohen และ 6) วิธีดีเอฟไอที (DFIT) เป็นต้น แต่หากใช้ข้อตกลงเบื้องต้นของโมเดลเป็นเกณฑ์จะสามารถแบ่งออกเป็น 2 กลุ่มที่สำคัญ คือ รูปแบบพาราเมทริก (parametric) และรูปแบบนันพาราเมทริก (nonparametric)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ภายใต้ทฤษฎีการตอบสนองข้อสอบ (IRT) ได้รับความสนใจจากนักวิจัยและนักวัดผลเสมอมา มีการพัฒนามาวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) มาโดยตลอด โดยในยุคแรกๆ จะใช้โมเดลการวัดความสามารถแบบเอกมิติ (Unidimensional) จนถึงปัจจุบันที่นักวิจัยและนักประเมินให้ความสำคัญกับการวัดความสามารถแบบพหุมิติ (Multidimensional) โดยใช้โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นฐานคิดของวิธีการตรวจสอบ

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้โมเดลการตอบสนองข้อสอบ (IRT) แบ่งได้เป็น 2 แนวคิดที่สำคัญ คือ แนวคิดแรกเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยการเปรียบเทียบความแตกต่างของค่าพารามิเตอร์ของข้อสอบระหว่างกลุ่มผู้สอบที่มีความสามารถระดับเดียวกัน โดยวิธีที่เป็นที่นิยม คือ 1) วิธีการทดสอบไค-สแควร์ของ Lord และ 2) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (likelihood ratio statistic) ซึ่งจะเปรียบเทียบความแตกต่างของค่าพารามิเตอร์ระหว่างกลุ่ม (Thissen Steinberge & Wainer, 1993) ส่วนแนวคิดที่สองจะมุ่งเปรียบเทียบความแตกต่างของพื้นที่ใต้โค้งระหว่างกลุ่มผู้สอบที่มีความสามารถระดับเดียวกัน วิธีที่เป็นที่นิยม คือ 1) วิธีการวัดพื้นที่ของ Shepard Camilli & Williams (1984) 2) วิธีการวัดพื้นที่ของ Kim & Cohen และ 3) วิธีการวัดพื้นที่ของ Raju เป็นต้น ซึ่งวิธีการวัดพื้นที่ของ Raju จะให้ผลการตรวจสอบมีความถูกต้องและแม่นยำสูงกว่าวิธีอื่นๆ (อรินทร์ น่วมถนอม, 2549)

กล่าวโดยสรุป การศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) เป็นประเด็นทางการศึกษาประเด็นหนึ่งที่นักวัดผลให้ความสำคัญในการศึกษาค้นคว้ามาเป็นเวลานานทั้งในประเทศและต่างประเทศ โดยศึกษาทั้งในเชิงทฤษฎีจากการจำลองข้อมูลขึ้นหรือศึกษาในสภาพจริงจากบริบททางการศึกษา ดังนั้น ขอสรุปผลการสังเคราะห์วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ดังนี้ วิธีที่

นิยมใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ประกอบไปด้วย 1) วิธีแมนเทล(Mantel) 2) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) 3) วิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) 4) วิธีซิบเทสท์ (SIBTEST) 5) วิธีโพลีซิบเทสท์ (Poly-SIBTEST) 6) วิธีซิบเทสท์ปรับใหม่ (Modified SIBTEST) 7) วิธีดีเอฟไอที (DFIT) 8) วิธีโครซิบ (CRO SIB) 9) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด 10) วิธีทดสอบไค-สแควร์ของ Lord (Lord' 9s $\chi 2$) 11) วิธีการวัดพื้นที่ของ Raju (Raju' s Area Measures) 12) วิธีการวิเคราะห์องค์ประกอบจำกัด 13) วิธี Closed Interval Area (ญาณภัทร สีหะมงคล, 2540) 14) วิธีการถดถอยโลจิสติก (Logistic Regression) และ 15) วิธีขอความเห็นของผู้เชี่ยวชาญ (Expert Judgement)

ตอนที่ 3 มโนทัศน์ของการวิเคราะห์พหุระดับ (Multilevel analysis)

ในตอนที่ 3 นี้แบ่งการนำเสนอออกเป็น 2 ส่วนสำคัญ คือ 3.1) มโนทัศน์ของการวิเคราะห์พหุระดับ และ 3.2) การวิเคราะห์พหุระดับด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น (HLM) ดังนี้

3.1) มโนทัศน์ของการวิเคราะห์พหุระดับ

3.1.1) แนวคิดของการวิเคราะห์พหุระดับ

การวิจัยทางการศึกษาโดยส่วนใหญ่นั้น มักเกี่ยวกับข้อมูลหลายระดับ ทั้งนี้เป็นเพราะโครงสร้างและธรรมชาติของข้อมูลทางการศึกษาจะมีลักษณะลดหลั่นเป็นระดับ (Hierarchy) โดยข้อมูลซึ่งเป็นค่าสังเกต (Observation) นี้ มีโครงสร้างเกี่ยวพันกันในลักษณะเป็นลำดับขั้น กล่าวคือ นอกเหนือจากค่าสังเกตจะได้รับอิทธิพลจากตัวแปรอิสระแล้วยังอาจได้รับอิทธิพลจากปัจจัยที่อยู่ในระดับอื่นๆ ร่วมด้วย

การวิจัยทางการศึกษาที่เกี่ยวข้องกับข้อมูลหลายระดับดังกล่าว หากผู้วิจัยละเลยและไม่ให้ความสนใจต่อโครงสร้างของระดับข้อมูล ผลการวิจัยจะบิดเบือนจากความเป็นจริง ตัวอย่างเช่น การศึกษาอิทธิพลของบรรยากาศในการสอนที่มีผลสัมฤทธิ์ทางการเรียน หากทำการวิเคราะห์แบบระดับเดียว โดยยึดนักเรียนหรือชั้นเรียนเป็นหน่วยของการวิเคราะห์ คือ การปรับตัวแปรต่างระดับให้เป็นข้อมูลระดับเดียวกันแล้วศึกษาความสัมพันธ์ระหว่างตัวแปร โดยใช้วิธีการวิเคราะห์การถดถอยแบบพหุคูณ (Multiple Regression Analysis) จะทำให้ผลที่ได้เกิดความคลาดเคลื่อนจากความเป็นจริงในการประมาณค่าพารามิเตอร์ (ศิริชัย กาญจนวาสี, 2550)

ในปี ค.ศ. 1976 ได้มีการประชุมเกี่ยวกับปัญหาของการวิจัยทางการศึกษา ซึ่งนักวิจัยที่เข้าร่วมประชุมได้ร่วมกันนำเสนอเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรของการวิเคราะห์ต่างระดับและความเหมาะสมของหน่วยวิเคราะห์ ซึ่ง Cronbach (Cronbach, 1976 อ้างถึงใน พิทักษ์ วงแหวน, 2546) ได้นำเสนอผลงานวิจัยเรื่อง การวิจัยระดับห้องเรียนและระดับโรงเรียน : การกำหนดรูปแบบของปัญหาและการวิเคราะห์ (Research or Classroom and school : formulation of Question Design and Analysis) และแสดงความคิดเห็นว่า การศึกษาในเรื่องอิทธิพลของตัวแปรในระบบการศึกษามีการเก็บรวบรวมและนำเสนอข้อมูลในแนวทางที่คลุมเครือ อีกทั้งวิธีการศึกษาที่ใช้กันอยู่ ได้ก่อให้เกิดข้อสรุปที่ผิดพลาดหลายประการที่สำคัญ คือ ความผิดพลาดในการตีความจากการวิเคราะห์อิทธิพลของตัวแปรในภาพรวมด้วยระเบียบวิธีวิเคราะห์แบบดั้งเดิม จากความผิดพลาดดังกล่าว Cronbach ได้เสนอแนวทางในการวิเคราะห์ข้อมูลด้วยการแบ่งอิทธิพลของตัวแปรทางการศึกษาออกเป็นอิทธิพลภายในกลุ่มและอิทธิพลระหว่างกลุ่มที่สนใจศึกษา ต่อมา Buretein และคณะ (Buretein et al,1978 อ้างถึงใน พิทักษ์ วงแหวน,

2546) ได้ศึกษาค้นคว้าขยายผลแนวคิดของ Cronbach ที่แสดงถึงความแตกต่างกัน (heterogeneity) ภายในกลุ่มของข้อมูลทางการศึกษา และได้นำเสนอเทคนิควิธีการวิเคราะห์ข้อมูลพหุระดับ เรียกว่า เทคนิคการใช้ความชันเป็นผลลัพธ์หรือเป็นตัวแปรตาม (Slop as outcome) ซึ่งเป็นเทคนิคการประมาณค่าพารามิเตอร์ในโมเดลการวิเคราะห์ข้อมูลพหุระดับแบบกำลังสองน้อยที่สุดแบ่งสองสมการ (OLS separate Equation Approach) ต่อมาเทคนิคการวิเคราะห์ข้อมูลพหุระดับได้รับความสนใจอย่างแพร่หลาย นักวิจัยทางการศึกษาต่างก็เสนอเทคนิคการประมาณค่าพารามิเตอร์ ตลอดจนมีการพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการวิเคราะห์ข้อมูลพหุระดับขึ้นโดยมีวิธีการประมาณค่าพารามิเตอร์ตลอดจนมีการพัฒนาโปรแกรมคอมพิวเตอร์ที่ใช้ในการวิเคราะห์ข้อมูลพหุระดับขึ้นและมีวิธีการประมาณค่าพารามิเตอร์ที่สำคัญ เช่น วิธีการประมาณค่าส่วนประกอบความแปรปรวน (Analysis of Variance Component Estimation) วิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum Likelihood) และวิธีการประมาณค่าพารามิเตอร์ของเบย์ส (Bayesian Estimation) เป็นต้น

3.1.2) ความหมายของการวิเคราะห์พหุระดับ

นักวิชาการได้ให้ความหมายของการวิเคราะห์พหุระดับไว้ดังนี้

Leeuw (2008) กล่าวว่า การวิเคราะห์พหุระดับ (multilevel analysis) เป็นการวิเคราะห์ทางสถิติของข้อมูลแบบลดหลั่นและแบบไม่ลดหลั่นแต่เป็นข้อมูลแบบ nested ตัวอย่างที่ง่ายที่สุดของการวิเคราะห์แบบพหุระดับ เช่น การที่กลุ่มของนักเรียนรวมอยู่ในห้องเรียนเดียวกัน ซึ่งโมเดลที่ใช้ในการวิเคราะห์ข้อมูลพหุระดับนั้นมีทั้งโมเดลเชิงเส้นตรงและโมเดลแบบไม่เป็นเส้นตรง

Bryk & Raudenbush (1992) กล่าวว่า การวิเคราะห์พหุระดับ (multilevel analysis) เป็นการจัดกระทำกับข้อมูลแบบ nested ตัวอย่างที่เห็นได้ชัด คือ ข้อมูลทางการศึกษาที่นักเรียน nested อยู่ในชั้นเรียนและชั้นเรียน nested อยู่ในเขตการศึกษา ซึ่งข้อมูลที่อยู่ระดับระดับล่างได้รับอิทธิพลจากข้อมูลในระดับที่สูงกว่า

ศิริชัย กาญจนวาสี (2550) กล่าวว่า การวิเคราะห์พหุระดับ (multilevel analysis) เป็นเทคนิคทางสถิติที่ใช้วิเคราะห์อิทธิพลของตัวแปรทำนายที่มีโครงสร้างเป็นระดับลดหลั่น (hierarchical) อย่างน้อย 2 ระดับ โดยตัวแปรทำนายและตัวแปรตามที่อยู่ระดับล่างมีความสัมพันธ์ซึ่งกันและกันและได้รับอิทธิพลร่วมกันจากตัวแปรทำนายที่อยู่ระดับบน

สำราญ มีแจ้ง (2544) กล่าวว่า การวิเคราะห์พหุระดับ หมายถึง เทคนิควิธีการทางสถิติที่ใช้วิเคราะห์อิทธิพลของตัวแปรเป็นระดับลดหลั่นกันอย่างน้อย 2 ระดับต่อตัวแปรตามซึ่งอยู่ในโมเดลการวัดซ้ำ โดยตัวแปรอิสระระดับบุคคลจะมีอิทธิพลต่อตัวแปรตามผ่านตัวแปรอิสระที่อยู่ในโมเดลการวัดซ้ำและตัวแปรอิสระระดับเดียวกันจะมีปฏิสัมพันธ์ภายในด้วยกัน

สรุปได้ว่า การวิเคราะห์พหุระดับ เป็นเทคนิควิธีทางสถิติที่ใช้วิเคราะห์ข้อมูลที่มีตัวแปรอิสระหลายตัวและตัวแปรอิสระเหล่านี้สามารถแบ่งเป็นระดับได้อย่างน้อย 2 ระดับขึ้นไป โดยตัวแปรระดับเดียวกันต่างมีความสัมพันธ์ซึ่งกันและกันและได้รับผลร่วมกันจากตัวแปรในระดับบน

3.1.3) **โครงสร้างของข้อมูล**

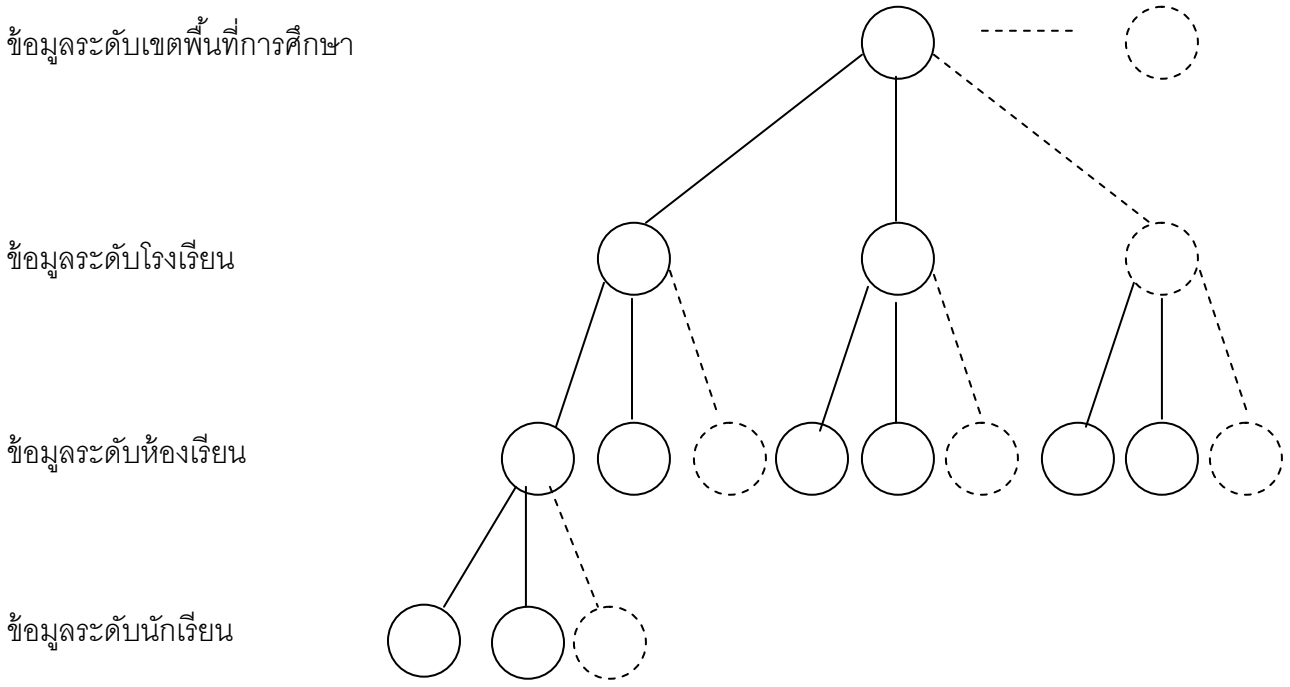
Burstein (1980) ได้กล่าวว่าหลักสำคัญที่จะทำให้เกิดความก้าวหน้าในการวิจัยเกี่ยวกับประสิทธิภาพการจัดการศึกษาของโรงเรียนจำเป็นต้องมีการพัฒนาโมเดลและวิธีการวิเคราะห์ข้อมูลที่มีความเหมาะสมกับโครงสร้างองค์กรที่มีลักษณะเป็นข้อมูลแบบพหุระดับ ทั้งนี้ก็เพราะว่าลักษณะทั่วไปของการจัดการองค์กรทางการศึกษานั้นนักเรียนจะถูกสอนรวมกันเป็นชั้นเรียนภายในโรงเรียน ซึ่งโรงเรียนก็จะอยู่ในแต่ละสังกัดหรืออำเภอที่แตกต่างกันไป ข้อมูลทางการศึกษาจึงมักเป็นข้อมูลที่มีหลายระดับสอดแทรกกันอยู่ ทั้งนี้เพราะระบบทางการศึกษาเป็นระบบที่มีความซับซ้อน การบริหารงานมีโครงสร้างเป็นหน่วยที่ซ้อนกันเป็นลำดับ โดยหน่วยที่อยู่ระดับล่างสุดจะได้รับอิทธิผลร่วมกันจากตัวแปรที่อยู่ในระดับสูงกว่า

อย่างไรก็ตาม ในการวิเคราะห์ข้อมูลที่ผ่านมา นักวิจัยมักจะละเลยลักษณะของข้อมูลที่เป็นหลายระดับ โดยการจัดกระทำตัวแปรต่างระดับกันให้มาอยู่ในระดับที่ผู้วิจัยสนใจ แล้ววิเคราะห์เสมือนหนึ่งข้อมูลเหล่านั้นอยู่ในระดับเดียวกัน ซึ่งผลการวิจัยที่ได้อาจจะบิดเบือนจากความเป็นจริง (ศิริชัย กาญจนวาสี 2548ก, 2550) จากงานวิจัยของ Cronbach และ Webb ที่ได้ทำการวิเคราะห์ซ้ำกับผลการศึกษาในอดีตที่ได้รายงานความสัมพันธ์ของตัวแปรความถนัดกับผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ว่ามีปฏิสัมพันธ์ (interaction) กันกับผู้เรียนโดยงานวิจัยนี้ Cronbach และ Webb พบว่าผู้วิจัยก่อนหน้ามิได้สนใจการศึกษาตัวแปรทางสังคม บริบทของโรงเรียนหรือห้องเรียนแต่อย่างใด Cronbach และ Webb จึงได้ใช้เทคนิคการศึกษาที่เรียกว่า aggregation techiques เมื่อวิเคราะห์ความสัมพันธ์ดังกล่าวใหม่ พบว่าความคลาดเคลื่อนของการสรุปผลการวิจัยนั้นคือความถนัดมีอิทธิผลปฏิสัมพันธ์กับตัวแปรความต่างระหว่างห้องเรียนมากว่าในระดับบุคคล Cronbach และ Webb จึงสรุปว่าด้วยข้อจำกัดทางสถิติวิเคราะห์อาจจะทำให้ผลการวิเคราะห์ที่ได้ผิดพลาดได้ แม้ว่านักวิจัยจะออกแบบการวิจัยด้วยความรอบคอบก็ตาม ดังนั้นการใช้โมเดลทางสถิติที่เหมาะสมตามระดับของหน่วยการวิเคราะห์ที่

สอดคล้องกับข้อมูลจึงน่าจะมีความเหมาะสมมากกว่า ซึ่งมีนักวิจัยหลายท่านได้นำเสนอเทคนิคการวิเคราะห์แบบพหุระดับ เช่น Cronbach (1976), Burstein (1978,1980), Golodstien (1976), Rogisa (1978) และ Aitkin & Longford (1986) เป็นต้น

3.1.4) ลักษณะของข้อมูลที่จะนำมาวิเคราะห์พหุระดับ

การวิเคราะห์ข้อมูลแบบพหุระดับมีประเด็นที่สำคัญที่ควรพิจารณาประการหนึ่ง คือ ลักษณะของข้อมูลที่จะนำมาวิเคราะห์จะต้องมีลักษณะเป็นพหุระดับ (Multilevel data) หรือเป็นข้อมูลสอดแทรกแบบลดหลั่น (Hierarchical nested data) ซึ่งเมื่อพิจารณาบริบททางการศึกษา จะพบว่าข้อมูลทางการศึกษามักจะเป็นข้อมูลแบบพหุระดับ นั่นคือ ข้อมูลระดับนักเรียน (student level) เช่น ความถนัดทางการเรียน พฤติกรรมการเรียน หรือเจตคติต่อการเรียน เป็นต้น ข้อมูลระดับนักเรียนจะอยู่ภายใต้หรือได้รับอิทธิพลมาจากข้อมูลระดับห้องเรียน (classroom level) เช่นคุณภาพการสอนของครู วุฒิการศึกษาของครู หรือบรรยากาศในชั้นเรียน เป็นต้น ข้อมูลระดับห้องเรียนอยู่ภายใต้หรือได้รับอิทธิพลมาจากข้อมูลระดับโรงเรียน (school level) เช่น ความเป็นผู้นำของผู้บริหารโรงเรียน หรือลักษณะการให้ความร่วมมือของผู้บริหารโรงเรียน เป็นต้น ข้อมูลระดับโรงเรียนก็จะอยู่ภายใต้หรือได้รับอิทธิพลจากข้อมูลระดับสูงกว่า (H) ต่อๆ กันไปดังปรากฏ ดังแผนภาพที่ 2.3



แผนภาพที่ 2.3 ลักษณะโครงสร้างของข้อมูลที่จะนำมาวิเคราะห์พหุระดับ

ลักษณะข้อมูลที่จะนำมาวิเคราะห์พหุระดับที่กล่าวมาข้างต้น แสดงให้เห็นว่า การที่จะวิเคราะห์ข้อมูลเพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรต่างๆ ด้วยการนำตัวแปรที่อยู่ต่างระดับกันมาวิเคราะห์ให้อยู่ในระดับเดียวกันนั้นเป็นการวิเคราะห์ข้อมูลที่ไม่สอดคล้องกับธรรมชาติและโครงสร้างของข้อมูล

3.1.5) **วิธีการประมาณค่าพารามิเตอร์ของการวิเคราะห์พหุระดับ**

วิธีการประมาณค่าพารามิเตอร์ของการวิเคราะห์พหุระดับมีเทคนิค ดังนี้

1) การประมาณค่าส่วนประกอบความแปรปรวน
2) วิธีกำลังสองน้อยที่สุด
3) วิธีกำลังสองน้อยที่สุดแบ่งสองสมการ
4) วิธีการประมาณค่าความเป็นไปได้สูงสุด
5) การประมาณค่าพารามิเตอร์ด้วยวิธีการของเบย์

ศิริชัย กาญจนวาสี (2550) ได้อธิบายหลักการสำคัญของการวิเคราะห์พหุระดับไว้ 3 ประการคือ

1) โครงสร้างตามลำดับชั้นของข้อมูลถูกนำมาพิจารณา เพื่อให้ความสำคัญต่อข้อมูลต่างระดับโดยการศึกษาความสัมพันธ์ของตัวแปรที่อยู่ในระดับเดียวกัน

2) หลักการของตัวแปรสุ่มจะนำมาใช้ในการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม โดยถือว่าตัวแปรตามระดับชั้นเรียน/โรงเรียน น่าจะมีอิทธิพลที่แตกต่างกัน

3) การเลือกใช้สถิติวิเคราะห์ที่มีอยู่หลายวิธีการให้เหมาะสมต่อการคำนวณหาค่าสัมประสิทธิ์การถดถอย

3.1.6) **หลักการวิเคราะห์พหุระดับ**

เนื่องจากข้อมูลทางการศึกษาเป็นข้อมูลหลายระดับที่สอดแทรกลดหลั่นกัน การวิเคราะห์ข้อมูลจึงควรทราบถึงค่าความแปรปรวนของตัวแปรในระดับต่างๆ ว่าแต่ละส่วนแตกต่างกันอย่างไรตามระดับของข้อมูลหรือความแปรปรวนจากความต่างระหว่างระดับนักเรียน ชั้นเรียน โรงเรียนหรือไม่ เพียงใด ซึ่งตัวแปรที่วัดได้ในระดับนักเรียนมีความแปรปรวนสามารถแยกส่วนประกอบได้ตามระดับที่ลดหลั่นกัน เช่น หากศึกษาแบบสามระดับก็สามารถแยกศึกษาความแปรปรวนได้ตั้งแต่ระดับนักเรียน

ระดับห้องเรียนและระดับโรงเรียน แสดงสมการส่วนประกอบความแปรปรวนได้ดังนี้ (นงลักษณ์ วิรัชชัย, 2535; ศิริชัย กาญจนวาสี, 2548)

$$\sigma^2 y = \sigma^2_{pupil} + \sigma^2_{class} + \sigma^2_{school}$$

เมื่อ $\sigma^2 y$ เป็นความแปรปรวนของตัวแปรตาม

σ^2_{pupil} เป็นความแปรปรวนระหว่างนักเรียนภายในห้องเรียน

σ^2_{clas} เป็นความแปรปรวนระหว่างห้องเรียนภายในโรงเรียน

σ^2_{school} เป็นความแปรปรวนระหว่างโรงเรียน

ดังนั้นในการวิจัยทางการศึกษา ผู้วิจัยควรให้ความสำคัญกับการประมาณค่าส่วนประกอบความแปรปรวนที่มีในแต่ละระดับของข้อมูล ซึ่งวิธีการประมาณค่าส่วนประกอบความแปรปรวนสามารถคำนวณหาได้ 4 วิธี (ศิริชัย กาญจนวาสี, 2541) ดังนี้

1) การใช้วิธีการวิเคราะห์ความแปรปรวน โดยการเลือกโมเดลที่มีความเหมาะสมกับโครงสร้างของแหล่งความแปรปรวนของข้อมูลเพื่อคำนวณค่าคาดหมายของกำลังสองเฉลี่ยของแต่ละแหล่งความแปรปรวน จากนั้นจึงหาค่าความแปรปรวนของแต่ละส่วนที่ต้องการศึกษา

2) การใช้วิธีความเป็นไปได้สูงสุด เพื่อประมาณค่าส่วนประกอบความแปรปรวนแต่ละส่วนที่มีความเป็นไปได้สูงสุด

3) การใช้วิธีกำลังสองที่ไม่ลำเอียงสูงสุด เพื่อประมาณค่าส่วนประกอบความแปรปรวนแต่ละส่วนที่ให้ค่าความคลาดเคลื่อนกำลังสองน้อยที่สุด

4) การใช้วิธีการของเบย์ เป็นวิธีการประมาณค่า b_{ij} ใช้ผลรวมถ่วงน้ำหนักด้วยค่าความเที่ยงของค่า OLS Slope กับค่าเฉลี่ยประชากรผลลัพธ์ที่ได้จึงน่าเชื่อถือ

การประมาณค่าพารามิเตอร์ของการวิเคราะห์พหุระดับ โดยการใช้ความชันเป็นผลลัพธ์หรือตัวแปรตาม (slope as outcome) มีหลักในการวิเคราะห์คือ การตรวจสอบหรือพิจารณาความสัมพันธ์ระหว่างตัวแปรภายในชั้นเรียนโดยใช้เทคนิคกำลังสองที่น้อยที่สุดซึ่งมีข้อตกลงเบื้องต้นของการวิเคราะห์ คือ 1) ตัวแปรอิสระในแต่ละระดับต้องไม่มีความคลาดเคลื่อนในการวัด และ 2) ในแต่ละระดับของตัวแปรที่นำมาศึกษาค่าของตัวแปรตาม (y) มีการแจกแจงเป็นโค้งปกติในแต่ละค่าของตัวแปรอิสระ (x) โดยมีค่าความแปรปรวนเท่ากันในทุกค่าของตัวแปรอิสระด้วย แต่ความแปรปรวนที่อยู่ต่างระดับ

กันไม่จำเป็นต้องเท่ากัน หากศึกษาตัวแปรสองระดับ ได้แก่ ระดับนักเรียน และระดับชั้นเรียน จะต้องสามารถวิเคราะห์ถดถอยตัวแปรระดับชั้นเรียนเป็นระดับมหภาค (macro level) และตัวแปรระดับนักเรียนเป็นระดับจุลภาค (micro level) แสดงโมเดลสรุปแบบการวิเคราะห์ได้ดังนี้

1) <u>การวิเคราะห์ระดับนักเรียน (student level)</u>

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรตามระดับนักเรียน(Y_{ij}) กับตัวแปรอิสระระดับนักเรียน (X_{ij}) โดยการวิเคราะห์ถดถอยในแต่ละชั้นเรียนได้ดังสมการ

$$Y_{ij} = b_{oj}+b_{ij}X_{ij}+e_{ij}$$

เมื่อ Y_{ij} เป็นตัวแปรตามระดับนักเรียน เช่น คะแนนผลสัมฤทธิ์ทางการเรียนของนักเรียนคนที่ i ชั้นที่ j

X_{ij} เป็นตัวแปรอิสระระดับนักเรียน เช่น พื้นฐานของความรู้เดิมของนักเรียน คนที่ i ชั้นที่ j

b_{oj} เป็นจุดตัดแกนของตัวแปรระดับนักเรียน ในชั้นที่ j

b_{ij} เป็นค่าความชันของการถดถอย ซึ่งเป็นความสัมพันธ์ของ X_{ij} และ Y_{ij} ในชั้นที่ j

e_{ij} เป็นค่าความคลาดเคลื่อนในการทำนาย ในระดับนักเรียนโดยที่ e ~ N โดยที่ห้องเรียนแต่ละห้องเป็นอีสระต่อกัน

จากนั้นจึงใช้ b_{oj} และ b_{ij} ของแต่ละชั้นเป็นตัวแปรตามสำหรับการวิเคราะห์ในระดับชั้นเรียนต่อไปและกำหนดให้ทั้งสองค่าเป็นอิทธิพลกำหนด (fixed effect) นั้นคือให้มีค่าคงที่ภายในแต่ละห้องเรียนและไม่มีความคลาดเคลื่อนในการประมาณค่าระหว่างห้องเรียน

2) <u>การวิเคราะห์ระดับชั้นเรียน (classroom level)</u>

เป็นการศึกษาความสัมพันธ์ระหว่างตัวแปรอิสระระดับชั้นเรียน (Zi) กับตัวแปร b_{ij} และ b_{oj} ที่นำมาจากการวิเคราะห์ระดับนักเรียนโดยการวิเคราะห์ถดถอยมีสมการดังนี้

$$b_{0j} = \gamma_{oo +} \gamma_{o1Zj} +u_{oj}$$
$$b_{ij} = \gamma_{1o +} \gamma_{11}Zj +U_{ii}$$

เมื่อ Z_j เป็นตัวแปรอิสระระดับชั้นเรียน เช่น เพศของครูผู้สอนในชั้นเรียนที่ j

γ_{oo} เป็นจุดตัดแกน (intercept) ของ b_{oj}

γ_{o1} เป็นค่าความชัน (slope) ที่แสดงอิทธิผลของ Z_j ต่อ b_{oj}

u_{oj} ค่าความคลาดเคลื่อนในการทำนาย Y_{ij} ระดับชั้นเรียนโดยที่ $e \sim N(0, \sigma^2)$ และห้องเรียนแต่ละห้องเป็นอิสระต่อกัน

γ_{1o} เป็นจุดตัดแกน (intercept) ของ b_{ij}

γ_{11} เป็นค่าความชัน (slope)ที่แสดงความสัมพันธ์ของ Z_j ต่อ b_{ij} ในชั้นที่ j

u_{ij} เป็นค่าความคลาดเคลื่อนระดับชั้นเรียนในการทำนาย b_{ij} ห้องเรียนที่ j

อย่างไรก็ตาม การประมาณค่าพารามิเตอร์ด้วยวิธีการดังกล่าวก็ยังไม่มีความสมบูรณ์มากนักเพราะยังมีข้อจำกัดคือ หากกลุ่มตัวอย่างที่ศึกษามีขนาดเล็ก จะทำให้ค่าสัมประสิทธิ์การถดถอยที่ได้จากการวิเคราะห์ระดับนักเรียน มีค่าต่ำซึ่งจะทำให้ความคลาดเคลื่อนของกลุ่มมีค่ามากขึ้น อันจะทำให้ความสัมพันธ์ที่ได้จากการวิเคราะห์ระดับชั้นเรียนมีค่าน้อยตามลงไปด้วย ประสิทธิภาพในการประมาณค่าพารามิเตอร์ระดับชั้นเรียนมีค่าต่ำลง และที่สำคัญเทคนิคกำลังสองน้อยที่สุดแบบสองสมการก็มีข้อจำกัดด้านความไม่เหมาะสมของโมเดลที่ใช้ในการวิเคราะห์ ความยุ่งยากในการเตรียมแฟ้มข้อมูลสำหรับการวิเคราะห์และไม่สามารถตรวจสอบนัยสำคัญความแปรปรวนของตัวแปรที่ศึกษาก่อนการวิเคราะห์ในแต่ละระดับได้

ตอนที่ 4 โมเดลเชิงเส้นตรงทั่วไปแบบลดหลั่น (HGLM) สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous)

การศึกษาเกี่ยวกับโมเดล HGLM มีมาตั้งแต่ปี ค.ศ. 1995 จากการศึกษาของ Raudenbush (1995 อ้างถึงใน Cheong, 2006; Kamata, 1998, 2001, 2002, 2003) โดยโมเดล HGLM เกิดจากการรวมกันของโมเดล GLM (Generalized Linear Model) และโมเดล HLM (Hierarchical Linear Model) ซึ่งสามารถวิเคราะห์ได้จากโปรแกรม HLM (Cheong, 2006; Kamata, 1998, 2001, 2002, 2003; Beretvas & Williams, 2004) ในปี ค.ศ. 1996 Adams และ Wilson (Adams & Wilson, 1996 อ้างถึงใน Kamata, 1998, 2001, 2002, 2003) ได้นำเสนอ A random coefficient multinomial logit model (RCMLM) ซึ่งเป็นโมเดลราสช์ (Rasch model) ที่ใช้ได้ทั้งข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) และข้อสอบที่ตรวจให้คะแนนมากกว่า 2 ค่า (polytomous) และในปี ค.ศ. 1997 Adams

Wilsons และ Wang (Adams, Wilson & Wang, 1997 อ้างถึงใน Kamata, 1998, 2001, 2002, 2003) ได้ขยาย A random coefficient multinomial logit model (RCMLM) สู่โมเดลพหุระดับ (multilevel model) แต่เป็นการวิเคราะห์โมเดลพหุระดับ (multilevel model) แบบ 2 ระดับเท่านั้น (Kamata, 1998, 2001, 2002, 2003) แต่โมเดล HGLM มาเป็นที่รู้จักกันมากขึ้นใน ปี ค.ศ. 1998 จากการศึกษาของ Kamata (1998) ที่ศึกษาโมเดล HGLM แบบ 1 พารามิเตอร์ แต่เป็นการประยุกต์โมเดล HGLM สู่ทฤษฎีการตอบสนองข้อสอบ (IRT) ข้อค้นพบที่ได้จากการศึกษาในครั้งนี้ พบว่า โมเดลราสช์ (Rasch model) เป็นกรณีหนึ่งของโมเดลการตอบสนองข้อสอบ (IRT model) อาจกล่าวอีกนัยหนึ่งว่า โมเดล HGLM มีฐานคิดมาจากโมเดลราสช์แบบ 1 พารามิเตอร์ (1 parameter Rasch model) จากนั้นในปี ค.ศ. 2001 Kamata (2001) ได้ขยาย (extend) แนวคิดของโมเดล HGLM สู่การวิเคราะห์พหุระดับแบบ 3 ระดับ (3 level Multilevel analysis) โดยเพิ่มการวิเคราะห์ในระดับที่สาม คือ ระดับโรงเรียน (school level) หรือระดับกลุ่ม (group level) เข้าไปในโมเดลการวิเคราะห์และในปี ค.ศ. 2002 Kamata (2002) ได้นำเสนอผลการศึกษากระบวนการในการจัดกระทำกับการวิเคราะห์การตอบสนองข้อสอบด้วยโมเดล HGLM ซึ่งได้ข้อค้นพบว่าการวิเคราะห์พหุระดับด้วยโมเดลราสช์ (Rasch model) มีความทัดเทียมกัน (equivalent) กับการวิเคราะห์ด้วยโมเดลการตอบสนองข้อสอบ (IRT model)

ปัจจุบันนี้โมเดล HGLM เป็นที่รู้จักในนามของโมเดล Kamata (Kamata's Model) และ Kamata & Cheong (2007) สรุปว่า โมเดล HGLM สามารถใช้ได้กับข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) และข้อสอบที่ตรวจให้คะแนนมากกว่า 2 ค่า (polytomous) ใช้ได้กับแบบสอบแบบเอกมิติ (Unidimensional test) และแบบสอบแบบพหุมิติ (Multidimensional test) และสามารถวิเคราะห์พหุระดับได้ทั้งแบบ 2 ระดับ (2 level HGLM) และแบบ 3 ระดับ (3 level HGLM) รวมไปถึงการประยุกต์สู่การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF)

4.1) โมเดล HGLM สำหรับข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous)

Kamata (2001) ได้เสนอวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยโมเดล HGLM โดยใช้แนวคิดของโมเดลราสช์ (Rasch model) ร่วมกับหลักการของโมเดลการตอบสนองข้อสอบ (IRT model) โดยการวิเคราะห์ในระดับที่ 1 นั้น เป็นการวิเคราะห์ระดับข้อสอบ (item level) การวิเคราะห์ในระดับที่ 2 เป็นการวิเคราะห์ระดับผู้สอบ (person level) และการวิเคราะห์ในระดับที่ 3 เป็นการวิเคราะห์ระดับโรงเรียน (school level) หรือระดับกลุ่ม (group level) โดยวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบพหุระดับ (multilevel) สามารถใช้หลักการของการถดถอยโลจิสติก (Logistic regression) มาใช้ในการตรวจสอบได้

4.1.1) ความสอดคล้องกันของโมเดล HGLM สำหรับข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่ากับโมเดลราสช์ (Rasch Model)

โมเดลราสช์ (Rasch Model) มีแนวคิดเช่นเดียวกันกับแนวคิดของทฤษฎีการตอบสนองข้อสอบแบบ 1 พารามิเตอร์ (1 parameter IRT Model) ซึ่งประกอบด้วยพารามิเตอร์ของผู้สอบหรือความสามารถของผู้สอบและพารามิเตอร์ของข้อสอบ นั่นคือ ค่าความยากของข้อสอบ (difficulty) การใช้โมเดลเชิงเส้นตรงทั่วไปแบบลดหลั่น (HGLM) ในการวิเคราะห์ แสดงให้เห็นถึงความสอดคล้องกันของพารามิเตอร์ในการวิเคราะห์ 3 ระดับ ดังนี้

<u>การวิเคราะห์ระดับที่ 1 ระดับข้อสอบ (item level)</u>

เป็นการวิเคราะห์อิทธิพลของข้อสอบที่แทรกสอด (nested) อยู่ในผู้สอบแต่ละคนในแต่ละโรงเรียน ซึ่งอิทธิพลของข้อสอบจะเป็นค่าคงที่สำหรับผู้สอบทุกคน แต่จะมีความผันแปรแบบสุ่มในข้อสอบแต่ละข้อ แสดงได้ดังสมการ

$$\eta_{ijm} = \log\left(\frac{p_{ijm}}{1-p_{ijm}}\right) = \beta_{0jm} + \beta_{1jm}x_{1ijm} + \beta_{2jm}x_{2ijm} + \ldots + \beta_{kjm}x_{kijm}$$
$$= \beta_{0j} + \sum_{q=1}^{k}\beta_{qim}x_{qijm}$$
$$= \beta_{oj} + \beta_{qjm}$$

เมื่อ p_{ijm} เป็นความน่าจะเป็นในการตอบข้อสอบที่ i ของผู้สอบ j ในโรงเรียน m ได้ถูกต้อง

η_{ijm} เป็นค่าฟังก์ชันโลจิสกับ log odd ในการตอบข้อสอบที่ i ของผู้สอบ j ในโรงเรียน m

x_{qijm} เป็นตัวแปรดัมมี่ตัวที่ q ในการทำข้อสอบที่ i ของผู้สอบที่ j ในโรงเรียน m โดยค่า x_{qijm} จะเป็น 0 เมื่อ $q \neq i$ และ x_{qijm} มีค่าเป็น 1 เมื่อ $q = i$

$\beta_{0\,jm}$ เป็นสัมประสิทธิ์ที่เป็นค่าจุดตัดแกนและเป็นค่าที่ใช้แปลความหมายของอิทธิพลของข้อสอบที่ตัดออกหรือข้อสอบข้ออ้างอิง (reference item)

β_{qjm} เป็นอิทธิพลของตัวแปรดัมมี่ที่ q เช่นข้อสอบข้อที่ i เมื่อ i = 1, …, k–1) เมื่อเปรียบเทียบกับข้อสอบข้ออ้างอิง

การวิเคราะห์ระดับที่ 2 ระดับผู้สอบ (person level)

เป็นการวิเคราะห์ค่าความสามารถในการตอบข้อสอบระหว่างผู้สอบและค่าอิทธิพลของข้อสอบรายข้อของผู้สอบที่อยู่ในโรงเรียนเดียวกัน ซึ่งค่าอิทธิพลของข้อสอบจะเป็นค่าคงที่ระหว่างโรงเรียนแต่มีความผันแปรแบบสุ่มไปตามข้อสอบแต่ละข้อ แสดงได้ดังสมการ

$$\begin{aligned}\beta_{0jm} &= \gamma_{00m} + r_{0jm} \\ \beta_{1jm} &= \gamma_{10m} \\ \beta_{2jm} &= \gamma_{20m} \\ &\vdots \\ \beta_{(k-1)} &= \gamma_{(k-1)0m}\end{aligned}$$

เมื่อ γ_{00m} เป็นค่าจุดตัดแกนของ β_{0jm} หรือค่าเฉลี่ยของอิทธิพลข้อสอบข้อสอบข้ออ้างอิงต่อโอกาสในการตอบข้อสอบถูกในโรงเรียน m

r_{ojm} เป็นค่าเศษเหลือของ β_{0jm} หรือค่าที่เป็นส่วนเบี่ยงเบนของโอกาสในการตอบข้อสอบของผู้สอบ j ในโรงเรียน m ซึ่งการแจกแจงแบบโค้งปกติ มีค่าเฉลี่ยเป็นศูนย์และมีความแปรปรวนเท่ากับ τ

การวิเคราะห์ระดับที่ 3 ระดับโรงเรียน (school level)

เป็นการวิเคราะห์ค่าความสามารถในการตอบข้อสอบของผู้สอบในระดับโรงเรียนและค่าเฉลี่ยรวมอิทธิพลของข้อสอบรายข้อของผู้สอบต่อโอกาสในการตอบข้อสอบถูกระหว่างโรงเรียน แสดงได้ดังสมการ

$$\begin{aligned}\gamma_{00m} &= \pi_{000} + u_{00m} \\ \gamma_{10m} &= \pi_{100} \\ \gamma_{20m} &= \pi_{200} \\ &\vdots \\ \gamma_{(k-1)0m} &= \pi_{(k-1)00}\end{aligned}$$

เมื่อ π_{000} เป็นค่าจุดตัดแกนของ γ_{00m} หรือเป็นค่าเฉลี่ยรวมของอิทธิพลของข้อสอบข้ออ้างอิงต่อโอกาสในการตอบถูกของทุกโรงเรียน

u_{00m} เป็นค่าเศษเหลือของ γ_{00m} หรือค่าที่เป็นส่วนเบี่ยงเบนของโอกาสในการตอบข้อสอบของของโรงเรียนที่ m หรือเป็นความสามารถเฉลี่ยของนักเรียนโรงเรียนที่ m ซึ่งการแจกแจงแบบโค้งปกติมีค่าเฉลี่ยเป็นศูนย์และมีความแปรปรวนเท่ากับ τ_{π}

π_{q00} เป็นค่าอิทธิพลของข้อสอบข้อที่ 1 ถึงข้อที่ k – 1 ต่อโอกาสในการตอบข้อสอบถูกต้องเมื่อเทียบกับข้ออ้างอิง

เมื่อรวมสมการในระดับที่ 1 ถึงระดับที่ 3 จะได้สมการดังนี้

$$\begin{aligned}\eta_{ijm} &= \beta_{oj} + \beta_{qjm} \\ &= \pi_{000} + r_{0jm} + u_{00m} + \pi_{q00}\end{aligned}$$

ความน่าจะเป็นที่จะตอบข้อสอบที่ i ถูกของผู้สอบที่ j ในโรงเรียนที่ m ดังสมการ

$$P_{ijm} = \frac{1}{1+\exp\{-\eta_{ijm}\}}$$

$$P_{ijm} = \frac{1}{1+\exp\{-(u_{00m} + r_{ojm}) - (-\pi_{q00} - \pi_{000})\}}$$

เมื่อเทียบค่าพารามิเตอร์ของโมเดล HGLM กับโมเดลราสซ์ (Rasch model) ได้ดังนี้

โมเดล HGLM-3L; $$P_{ij} = \frac{1}{1+\exp\{-(u_{00m} + r_{0jm}) - (-\pi_{q00} - \pi_{000})\}}$$

โมเดลราสซ์ (Rasch model); $$P_{ij} = \frac{1}{1+\exp\{-(\theta_j - \delta_i)\}}$$

เมื่อ u_{00m} เป็นค่าความสามารถเฉลี่ยของนักเรียนในโรงเรียนที่ m

r_{0jm} เป็นค่าความสามารถเฉพาะของนักเรียนที่ j ในโรงเรียนที่ m

π_{000} เป็นค่าความยากของข้อสอบข้ออ้างอิง

π_{q00} เป็นค่าอิทธิพลของข้อสอบข้อที่ q (เมื่อ qเป็นข้อสอบที่ i ถึง $k-1$) เมื่อเทียบกับข้อสอบข้ออ้างอิง

θ_j เป็นค่าความสามารถของผู้สอบที่ j

δ_i เป็นค่าความยากของข้อสอบข้อที่ i

สรุปค่าพารามิเตอร์ของโมเดล HGLM แบบ 3 ระดับกับโมเดลราสช์ (Rasch model) ได้ดังตารางที่ 2.4

ตารางที่ 2.4 การเปรียบเทียบค่าความสามารถของผู้สอบและความยากของข้อสอบจากการประมาณค่าด้วยโมเดล HGLM แบบ 3 ระดับกับโมเดลราสช์ (Rasch model)

โมเดล	ความสามารถของผู้สอบ	ความยากของข้อสอบ
โมเดล HGLM-3L	$u_{00m} + r_{0jm}$	$(-\pi_{q00} - \pi_{000})$
โมเดลราสช์ (Rasch model)	θ_j	δ_i

4.1.2) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ตรวจให้คะแนนแบบ 2 ค่าด้วยโมเดล HGLM

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบพหุระดับ (multilevel) สามารถประยุกต์ใช้หลักการของการถดถอยโลจิสติก (Logistic regression) มาใช้ในการตรวจสอบได้ Kamata (1998) ได้เสนอวิธีการวิเคราะห์พหุระดับแบบ 2 ระดับเพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ใช้หลักการของโมเดลราสช์ (Rasch Model) โดยการวิเคราะห์ในระดับที่ 1 เป็นระดับข้อสอบ (item level) และการวิเคราะห์ในระดับที่ 2 เป็นการวิเคราะห์ระดับผู้สอบ (person level) ต่อมา Binici (2007) และ Kamata, Chaimongkol, Genc & Bilir (2005) ได้ขยายโมเดลการวิเคราะห์พหุระดับด้วยโมเดล HGLM แบบ 2 ระดับมาเป็นโมเดลการวิเคราะห์แบบ 3 ระดับ ซึ่งการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยด้วยโมเดล HGLM แบบ 3 ระดับนี้จะมีประโยชน์สำหรับโมเดลอิทธิพลสุ่ม (random-effect model) ซึ่งเป็นตัวบ่งชี้ถึงความไม่สอดคล้องกันของการเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) จากการจัดกลุ่มในโมเดลระดับที่ 3 และเป็นโมเดลอิทธิพลอธิบาย (explanatory effect model) เพื่อระบุตัวแปรคุณลักษณะในโมเดลแบบ 3 ระดับ

ในการวิเคราะห์ตามแนวคิดของโมเดล HGLM ในการวิเคราะห์ระดับที่ 1 ซึ่งเป็นระดับข้อสอบ (Item level) นั้นจะกำหนดให้ค่าพารามิเตอร์ความยาก (difficulty) ของข้อสอบคงที่กับผู้สอบแต่ละกลุ่มและการวิเคราะห์ระดับที่ 2 ซึ่งเป็นระดับผู้สอบ (person level) นั้นจะใช้ตัวแปรบ่งชี้ของกลุ่มเพื่อทำนายโอกาสในการตอบข้อสอบได้ถูก ถ้าพบว่ามีนัยสำคัญทางสถิติจะแสดงได้ว่าตัวแปรบ่งชี้ของกลุ่มมีอิทธิพลต่อโอกาสในการตอบข้อสอบได้ถูก นั่นคือ ข้อสอบข้อนั้นเป็นข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ (DIF) และนอกจากนี้ยังสามารถวิเคราะห์ในระดับที่ 3 ซึ่งเป็นระดับโรงเรียน (school level) หรือเป็นระดับที่ผู้สอบแทรกสอด (nested) อยู่ในกลุ่มข้อมูลที่อยู่ในระดับใหญ่ขึ้นเพื่อใช้คุณลักษณะของกลุ่มเป็นตัวแปรในการทำนายคุณลักษณะของกลุ่มผู้สอบที่ส่งผลต่อโอกาสในการตอบข้อสอบได้ถูก เพื่อระบุแหล่งของการทำหน้าที่ต่างกันของข้อสอบ (DIF) เขียนสมการแสดงการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ตามแนวคิดของโมเดล HGLM-3L ได้ดังนี้

<u>การวิเคราะห์ระดับที่ 1 ระดับข้อสอบ (item level)</u>

เป็นการวิเคราะห์อิทธิพลของสอบที่แทรกสอด (nested) อยู่ในผู้สอบแต่ละคนในแต่ละโรงเรียน ดังสมการ

$$\begin{aligned}\eta_{ijm} &= \beta_{0jm} + \beta_{1jm}x_{1ijm} + \beta_{2jm}x_{2ijm} + \ldots + \beta_{kjm}x_{kijm} \\ &= \beta_{0j} + \sum_{q=1}^{k}\beta_{qim}x_{qijm} \\ &= \beta_{oj} + \beta_{qjm}\end{aligned}$$

เมื่อ η_{ijm} เป็นค่าฟังก์ชันโลจิสกับ log odd ในการตอบข้อสอบที่ i ของผู้สอบ j ในโรงเรียน m

$\beta_{0\,jm}$ เป็นสัมประสิทธิ์ที่เป็นค่าจุดตัดแกนและเป็นค่าที่ใช้แปลความหมายของอิทธิพลของข้อสอบที่ตัดออกหรือข้อสอบข้ออ้างอิง (reference item)

β_{qjm} เป็นอิทธิพลของตัวแปรดัมมี่ที่ q เช่นข้อสอบข้อที่ i เมื่อ i = 1, …, k–1) เมื่อเปรียบเทียบกับข้อสอบข้ออ้างอิง (reference item)

x_{qijm} เป็นตัวแปรอิทธิพลของข้อสอบรายข้อซึ่งเป็นตัวแปรดัมมี่ตัวที่ q ในการทำข้อสอบที่ i ของผู้สอบที่ j ในโรงเรียน m โดยค่า x_{qijm} จะเป็น 0 เมื่อ $q \neq i$ และ x_{qijm} มีค่าเป็น 1 เมื่อ $q = i$

เพื่อให้โมเดลในสมการพหุระดับเป็นเมตริกซ์เอกลักษณ์ (identity matrix) จึงต้องกำหนดให้ x_{qijm} ตัวใดตัวหนึ่งเป็นศูนย์ เพื่อทำให้เป็นเมตริกซ์เต็มอันดับ (full rank)

<u>การวิเคราะห์ระดับที่ 2 ระดับผู้สอบ (person level)</u>

เป็นการวิเคราะห์ที่นำตัวแปรคุณลักษณะของกลุ่มผู้สอบเข้าไปในสมการเพื่อทำนายโอกาสในการตอบข้อสอบถูกต้อง โดยค่าอิทธิพลของข้อสอบรายข้อของผู้สอบจะเป็นตัวบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ (DIF) ดังสมการ

$$\begin{aligned}
\beta_{0jm} &= \gamma_{00m} + \gamma_{01m}G_{jm} + r_{0jm} \\
\beta_{1jm} &= \gamma_{10m} + \gamma_{11m}G_{jm} \\
\beta_{2jm} &= \gamma_{20m} + \gamma_{21m}G_{jm} \\
&\vdots \\
\beta_{qjm} &= \gamma_{q0m} + \gamma_{q1m}G_{jm}
\end{aligned}$$

เมื่อ γ_{q0m} เป็นค่าจุดตัดแกนของ β_{qjm} หรือค่าเฉลี่ยของโอกาสในการตอบข้อสอบข้อที่ q ได้ถูกต้องของผู้สอบในโรงเรียนที่ m

γ_{q1m} เป็นค่าอิทธิพลของตัวแปรของกลุ่มผู้สอบ G_{jm} ที่มีผลต่อโอกาสในการตอบข้อสอบถูกในข้อที่ q ในโรงเรียน m

r_{0jm} เป็นค่าเศษเหลือของ β_{0jm} หรือโอกาสของการตอบข้อสอบถูกของนักเรียนที่ j ในโรงเรียน m

<u>การวิเคราะห์ระดับที่ 3 ระดับโรงเรียน (school level)</u>

เป็นขั้นตอนการวิเคราะห์ที่นำตัวแปรทำนายระดับโรงเรียนเข้าไปในสมการเพื่อทำนายโอกาสในการตอบถูกของผู้สอบที่สอดแทรก (nested) ในโรงเรียนที่มีคุณลักษณะต่างกัน ดังสมการ

$$\gamma_{00m} = \pi_{000} + \pi_{001}SCH_m + u_{00m}$$
$$\gamma_{01m} = \pi_{010} + \pi_{011}SCH_m + u_{01m}$$
$$\gamma_{10m} = \pi_{100}$$
$$\gamma_{20m} = \pi_{200}$$
$$\vdots$$
$$\gamma_{q0m} = \pi_{q00}$$

เมื่อ π_{000} เป็นค่าจุดตัดแกนของ γ_{00m} หรือเป็นค่าเฉลี่ยรวมของอิทธิพลของข้อสอบข้ออ้างอิงต่อโอกาสในการตอบถูกของทุกโรงเรียน

π_{001} เป็นค่าอิทธิพลของตัวแปรในระดับโรงเรียนที่มีผลต่อโอกาสในการตอบข้อสอบข้ออ้างอิงได้ถูกในโรงเรียนที่ m

π_{010} เป็นค่าจุดตัดแกนของ γ_{01m} หรือเป็นค่าเฉลี่ยรวมของอิทธิพลของตัวแปรในระดับผู้สอบที่ส่งผลต่อโอกาสในการตอบถูกของทุกโรงเรียน

π_{011} เป็นค่าจุดตัดแกนของ γ_{00m} หรือเป็นค่าเฉลี่ยรวมของอิทธิพลของตัวแปรระดับโรงเรียนต่อโอกาสในการตอบถูกของทุกโรงเรียน

π_{q00} เป็นค่าเฉลี่ยของโอกาสในการตอบข้อสอบข้อ q ของผู้สอบได้ถูกต้อง

ตอนที่ 5 งานวิจัยที่เกี่ยวข้อง

ในตอนที่ 5 นี้แบ่งการนำเสนอออกเป็น 2 ส่วนสำคัญ คือ 5.1) งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และ 5.2) งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM ดังนี้

5.1) งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

จากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ (DIF) ทั้งในประเทศและต่างประเทศจำนวนทั้งสิ้น 27 เรื่อง แบ่งเป็นงานวิจัยต่างประเทศจำนวน 17 เรื่องและงานวิจัยในประเทศจำนวน 10 เรื่อง มีรายละเอียดดังต่อไปนี้

งานวิจัยต่างประเทศ

Swaminathan and Rogers (1990) ได้เปรียบเทียบวิธีถดถอยโลจิสติก (Logistic Regression) กับวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) กลุ่มตัวอย่างขนาด 250 และ 500 คน และความยาวของแบบสอบ 3 ขนาด คือ 40 60 และ 80 ข้อ ผลการศึกษา พบว่า วิธีถดถอยโลจิสติก (Logistic Regression) และวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) ให้ผลการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป กล่าวคือ มีการตรวจสอบได้ถูกต้องร้อยละ 70 กรณีกลุ่มตัวอย่าง 250 คน และตรวจสอบได้ร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน ในทุกความยาวของแบบสอบ สำหรับการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป พบว่า วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) ตรวจสอบได้เล็กน้อย ส่วนวิธีถดถอยโลจิสติก (Logistic Regression) ตรวจสอบได้ถูกต้องร้อยละ 50 กรณีกลุ่มตัวอย่างน้อยและข้อสอบสั้นและถูกต้องร้อยละ 75 กรณีแบบสอบยาวและกลุ่มตัวอย่างขนาดใหญ่

Rogers and Swaminathan (1993) ได้เปรียบเทียบวิธีถดถอยโลจิสติก (Logistic Regression) กับวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการจำลองข้อมูล เพื่อศึกษาการกระจายของสถิติทดสอบและประสิทธิภาพการทำหน้าที่ต่างกันแบบเอกรูปและแบบอเนกรูป การศึกษาการกระจายของสถิติทดสอบ เป็นการศึกษาปัจจัยที่แปรเปลี่ยน คือ ขนาดของกลุ่มตัวอย่าง 250 และ 500 คน ความเหมาะสมของข้อมูลกับโมเดล ค่าความยากและอำนาจจำแนกของข้อสอบ ความยาวของแบบสอบ 40 ข้อ ผลการศึกษา พบว่า การกระจายของสถิติทดสอบทั้งสองวิธีน่าพอใจ ยกเว้นการกระจายของสถิติวิธีถดถอยโลจิสติก (Logistic Regression) ไม่เป็นไปตามที่คาดไว้ ในกรณีข้อสอบยากมากและอำนาจจำแนกสูง ด้านประสิทธิภาพการตรวจสอบ พบว่าทั้งสองวิธีมีประสิทธิภาพในการตรวจสอบเท่ากัน ในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป วิธี

ถดถอยโลจิสติก (Logistic Regression) ตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปได้ดีในกรณีที่ข้อสอบมีความยากปานกลางและอำนาจจำแนกสูง ส่วนวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) ตรวจสอบข้อสอบที่มีความยากปานกลางได้น้อยมาก แต่สามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ดีในกรณีที่ข้อสอบง่ายมากหรือข้อสอบที่ยากมาก

Mazor and al (1994) ศึกษาการใช้วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) ในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป โดยการจำลองข้อมูล กลุ่มตัวอย่างกลุ่มละ 1,000 คน แบบสอบ 25 ฉบับ แต่ละฉบับมี 75 ข้อ โดยในแต่ละฉบับมีข้อสอบที่ทำหน้าที่ไม่ต่างกัน 59 ข้อ และข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ กำหนดค่าอำนาจจำแนกไว้ 4 ระดับ (0.25, 0.6, 0.9, 1.25) กำหนดความแตกต่างของค่าอำนาจจำแนกระหว่างกลุ่มไว้ 5 ระดับ (0, 0.25, 0.50, 0.75, 1.0) กำหนดค่าความยากของข้อสอบกลุ่มอ้างอิงไว้ 5 ระดับ (-1.5, -1.0, 0, 1.0, 1.5) กำหนดความแตกต่างของค่าความยากระหว่างกลุ่มไว้ 4 ระดับ (0, 0.30, 0.60 1.0) ค่าโอกาสในการเดากำหนดไว้ที่ 0.2 และการกระจายความสามารถระหว่างกลุ่มผู้สอบกำหนดเป็น การกระจายความสามารถเท่ากันและไม่เท่ากัน ผลการศึกษา พบว่า เมื่อระดับความแตกต่างของค่าอำนาจจำแนกและค่าความยากระหว่างกลุ่มเพิ่มขึ้น จะทำให้อัตราการตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันมากขึ้น

Uttaro and Millsap (1994) ศึกษาปัจจัยที่มีอิทธิพลต่อวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการจำลองข้อมูล ปัจจัยที่ศึกษา คือ ความยาวของแบบสอบ 2 ขนาด คือ 20 และ 40 ข้อ ภายใต้เงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันเนื่องจากฟังก์ชันการตอบข้อสอบแปรเปลี่ยนไป โดยกำหนดฟังก์ชันการตอบของกลุ่มอ้างอิงคงที่ (a=1.0, b=0, c=0.2) ส่วนกลุ่มเปรียบเทียบกำหนดค่าอำนาจจำแนกไว้ 3 ระดับ ค่าความยากที่ต่างกัน 3 ระดับ และค่าโอกาสในการเดา 2 ระดับ กลุ่มตัวอย่างกลุ่มละ 500 คน ผลการศึกษา พบว่า ภายใต้เงื่อนไขข้อสอบที่ทำหน้าที่ต่างกัน ค่าความยาก ค่าอำนาจจำแนก ค่าโอกาสในการเดา การกระจายความสามารถและปฏิสัมพันธ์ระหว่างการกระจายความสามารถกับค่าพารามิเตอร์ของแบบสอบ มีผลต่อการประมาณค่า α_{MH} และข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่เป็นข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปมากกว่าแบบอเนกรูป

French and Miller (1996) ได้ศึกษาความเป็นไปได้ของการใช้วิธีการถดถอยโลจิสติก (Logistic Regression) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า โดยศึกษาจากแบบสอบที่จำลองขึ้น จำนวน 25 ข้อ แต่ละข้อมี 4 ลำดับขั้นคะแนน คือ ตั้งแต่ 0 ถึง 3 คะแนน ส่วนคะแนนสอบถูกจำลองขึ้นโดยใช้โมเดล GPCM ข้อมูลที่ได้ถูกนำมากำหนดรหัสใหม่ให้เป็น multiple dichotomies โดยใช้โมเดลตามแนวคิดของ Agresti (1990) 3 โมเดล คือโมเดลโลจิทของ

อัตราส่วนที่ต่อเนื่องกัน โมเดลโลจิสแบบสะสม และโมเดลโลจิสของลำดับขั้นที่ติดกัน เพื่อใช้กับวิธีการถดถอยโลจิสติก (Logistic Regression) ในการกำหนดรหัสให้ใช้กับรูปแบบการตอบแบบหลายค่า ผลการศึกษา พบว่า การเปลี่ยนแปลงขนาดของกลุ่มตัวอย่าง มีผลต่ออำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อกลุ่มตัวอย่างมีขนาดเล็กลง อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะลดลง วิธีโมเดลโลจิสของอัตราส่วนที่ต่อเนื่องกันและวิธีโมเดลโลจิสแบบสะสม มีอำนาจในการตรวจสอบสูงสุดในทุกๆ ครั้งของการถดถอย แต่ต่างกันที่วิธีโมเดลโลจิสของอัตราส่วนที่ต่อเนื่องกันมีการสูญเสียข้อมูลในการกำหนดรหัสครั้งที่ 2 และ 3 นอกจากนี้ยังพบว่า เมื่อค่าพารามิเตอร์อำนาจจำแนกของข้อสอบยิ่งแตกต่างกันมาก อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปจะยิ่งเพิ่มขึ้น

Narayanan and Swaminathan (1996) เปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) วิธีถดถอยโลจิสติก (Logistic Regression) และวิธีโครซิบท์ (CRO-SIB) ในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป โดยการจำลองข้อมูล ภายใต้เงื่อนไข คือ 1) กลุ่มตัวอย่างขนาด 500 และ 1,000 คน ในกลุ่มอ้างอิง และ 200, 500 คน ในกลุ่มเปรียบเทียบ 2) การกระจายความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเท่ากันและไม่เท่ากัน 3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ คือ 0%, 10% และ 20% 4) ขนาดของข้อสอบที่ทำหน้าที่ต่างกันหรือพื้นที่รวมแตกต่างระหว่างโค้งคุณลักษณะข้อสอบสองกลุ่ม 4 ระดับ คือ 0.4, 0.6, 0.8 และ 1.0 และ 5) ค่าโอกาสในการเดากำหนดเท่ากันที่ 0.2 และความยาวของแบบสอบเป็น 40 ข้อทุกเงื่อนไข ผลการศึกษาพบว่า วิธีถดถอยโลจิสติกและวิธีโครซิบท์ (CRO-SIB) ให้ผลใกล้เคียงกันในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูปและทั้งสองวิธีตรวจจับการทำหน้าที่ต่างกันได้ดีกว่าวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) ปัจจัยที่ส่งผลต่อการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป ประกอบด้วย ขนาดกลุ่มตัวอย่าง นั่นคือ เมื่อเพิ่มขนาดกลุ่มตัวอย่างทั้งสามวิธีสามารถตรวจสอบได้มากขึ้น การกระจายความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแบบเท่ากันทำให้ตรวจสอบได้มากขึ้น พื้นที่ความแตกต่างระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเพิ่มขึ้นจาก 0.4 เป็น 1.0 ทั้งสามวิธีสามารถตรวจสอบได้มากขึ้น ข้อสอบที่พบว่าทำหน้าที่ต่างกันด้วย วิธีถดถอยโลจิสติก (Logistic Regression) และวิธีโครซิบท์ (CRO-SIB) ส่วนใหญ่เป็นข้อสอบที่มีค่าความยากต่ำ ค่าอำนาจจำแนกสูง ส่วนวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) ตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ดีเฉพาะกรณีข้อสอบยากและข้อสอบง่ายซึ่งโค้งลักษณะข้อสอบ (ICC) ของผู้สอบสองกลุ่มตัดกันที่ระดับความสามารถสูงหรือความสามารถต่ำเท่านั้น

Roussos and Stout (1996) ศึกษาผลของกลุ่มตัวอย่างขนาดเล็กที่มีต่อความคลาดเคลื่อนชนิดที่ 1 ของวิธีซิบเทสท์ (SIBTEST) และวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) โดยศึกษาครั้งแรกใช้ขนาดของกลุ่มตัวอย่าง 100, 200, 300 และ 1,000 คน ความแตกต่างของค่าเฉลี่ยการกระจายความสามารถระหว่างกลุ่มเป็น 0, 0.5 และ 1.0 ข้อสอบจำนวน 25 ข้อ ผลการศึกษา พบว่า ค่าสถิติของวิธีซิบเทสท์ (SIBTEST) และวิธีวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) มีแนวโน้มที่จะมีความคลาดเคลื่อนชนิดที่ 1 เพิ่มขึ้น เมื่อขนาดกลุ่มตัวอย่างมีความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มเพิ่มขึ้น ส่วนการศึกษาครั้งที่สอง ใช้ขนาดของกลุ่มตัวอย่าง 500, 1,000 และ 3,000 คน ความแตกต่างของค่าเฉลี่ยการกระจายความสามารถระหว่างกลุ่มเป็น 0 และ 1.0 ค่าอำนาจจำแนก 3 ระดับ ค่าความยาก 5 ระดับ ค่าโอกาสในการเดา 3 ระดับ พบว่า เมื่อความแตกต่างของค่าเฉลี่ยการกระจายความสามารถเป็น 1.0 จะทำให้ความคลาดเคลื่อนชนิดที่ 1 เพิ่มขึ้นทั้งสองวิธี

Kim (2000) ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) วิธีแมนเทล(Mantel) และวิธีแมนเทล-แฮนส์เซลแบบทั่วไป (Generalize Mantel-Haenszel: GMH)โดยใช้ข้อมูลจากการประเมินโรงเรียนระดับอนุบาลของรัฐจอเจียร์ แบ่งขนาดของกลุ่มตัวอย่างที่ศึกษาออกเป็น 4 ขนาด คือ กลุ่มตัวอย่างขนาด 105,731 คน กลุ่มตัวอย่างขนาด10,000 คน กลุ่มตัวอย่างขนาด 1,000 คน และ กลุ่มตัวอย่างขนาด 100 คน ข้อคำถามที่นำมาตรวจสอบการทำหน้าที่ต่างกันเป็นข้อคำถามแบบให้คะแนนมากกว่า 2 ค่า การศึกษาในครั้งนี้มีวัตถุประสงค์ เพื่อศึกษาความสอดคล้องของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) วิธีแมนเทล (Mantel) และวิธีแมนเทล-แฮนส์เซลแบบทั่วไป (Generalized Mantel-Haenszel: GMH) ผลการศึกษา พบว่า ทั้ง 3 วิธี ให้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีเมื่อกลุ่มตัวอย่างมีขนาด 100 คน นอกจากนี้ยังมีข้อค้นพบที่สำคัญ คือ การใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินไปจะไม่มีประโยชน์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

Penfield (2001) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบในหลายกลุ่มด้วยวิธีแมนเทล-แฮนส์เซล 3 แบบ โดยมีวัตถุประสงค์ เพื่อศึกษาขนาดของความคลาดเคลื่อนชนิดที่ 1 เมื่อมีกลุ่มตัวอย่างที่ศึกษาพร้อมกันหลายกลุ่ม โดยมีเงื่อนไขที่ศึกษา คือ 1) ขนาดของกลุ่มตัวอย่าง นั่นคือ กลุ่มอ้างอิง 1 กลุ่มและกลุ่มเปรียบเทียบ 1 กลุ่ม กลุ่มอ้างอิง 1 กลุ่มและกลุ่มเปรียบเทียบ 2 กลุ่ม กลุ่มอ้างอิง 1 กลุ่มและกลุ่มเปรียบเทียบ 3 กลุ่ม กลุ่มอ้างอิง 1 กลุ่มและกลุ่มเปรียบเทียบ 4 กลุ่ม และ 2) การแจกแจงความสามารถของกลุ่มเปรียบเทียบ โดยเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลแบบไคสแควร์ที่ไม่ปรับระดับของ α วิธีแมนเทล-แฮนส์เซลแบบไคสแควร์ที่ปรับระดับของ α

ด้วย Bonferroni และวิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) ผลการศึกษา พบว่า วิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) ดีที่สุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในหลายกลุ่ม

Bolt (2002) ได้เปรียบเทียบวิธีการตรวจจับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบหลายค่าด้วยสถิติพาราเมตริกและนันพาราเมตริก การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลด้วยเทคนิคมอลติคาร์โล นั่นคือ การใช้วิธี GRM วิธี GRM-LR และวิธี GRM-DFIT ในการเปรียบเทียบวิธีการตรวจจับการทำหน้าที่ต่างกันของข้อสอบด้วยสถิตินันพาราเมตริกและใช้วิธีโพลีซิบเทสท์ (Poly-SIBTEST) ในการจำลองข้อมูล ผลการศึกษา พบว่า 1) วิธี GRM ให้ผลที่สอดคล้องกับข้อมูลเชิงประจักษ์มากที่สุด 2) วิธี GRM-LR ให้ค่าความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีอื่น และ 3) วิธี GRM-DFIT ให้ค่าความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีอื่น

Cohen and Bolt (2005) ได้วิเคราะห์โมเดลแบบผสมในการทำหน้าที่ต่างกันของข้อสอบ โดยวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามเพศและใช้แบบสอบแบบผสม (mixed format test) ตามแนวคิด IRT โดยมีวัตถุประสงค์เพื่อ 1) สำรวจมิติที่ทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ 2) ศึกษาคุณลักษณะของผู้สอบที่มีความเกี่ยวข้องกับมิติที่ทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ และ3) เปรียบเทียบคุณลักษณะทางวิชาการของผู้เรียนกับคุณลักษณะที่ปรากฏในชั้นเรียน แบบสอบที่ใช้เป็นแบบสอบจัดตำแหน่งวิชาคณิตศาสตร์ แบบเลือกตอบจำนวน 32 ข้อ กลุ่มตัวอย่างจำนวน 1,000 คน แบ่งเป็นเพศชาย 500 คน และเพศหญิง 500 คน วิเคราะห์ข้อมูลด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) ด้วยโปรแกรม MULTILOG ผลการศึกษา พบว่า มีข้อสอบที่ทำหน้าที่ต่างกันจำนวน 5 ข้อ โดยข้อสอบจำนวน 4 ข้อเข้าข้างเพศชายและข้อสอบอีก 1 ข้อ เข้าข้างเพศหญิง

Finch (2005) ได้เปรียบเทียบโมเดล MIMIC กับวิธีซิบเทสท์ (SIBTEST) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT Likelihood Ratio Test) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการจำลองข้อมูลจำนวนผู้สอบและจำนวนข้อสอบด้วยวิธีมอลติคาร์โล ผลการศึกษา พบว่า โมเดล MIMIC ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีในกรณีที่ข้อสอบมีจำนวน 50 ข้อ แบบ 2 พารามิเตอร์ และโมเดล MIMIC สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้สูงในกรณีที่ข้อสอบมีจำนวน 20 ข้อ แบบ 3 พารามิเตอร์โลจิสติก ส่วนความคลาดเคลื่อนชนิดที่ 1 มีค่าต่ำสุดในวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) นอกจากนี้ยังได้ข้อค้นพบว่า วิธีซิบเทสท์ (SIBTEST) ให้ผลคล้ายวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) แต่มีขนาดของความคลาดเคลื่อนชนิดที่ 1 สูงกว่า

Lei and al (2006) ได้ศึกษาการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ โดยการจำลองข้อมูลการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทางและไม่มีทิศทาง ภายใต้เงื่อนไขที่ศึกษา คือ กลุ่มตัวอย่างที่แตกต่างกัน นั่นคือ เงื่อนไขที่หนึ่ง จำนวนกลุ่มตัวอย่างทั้งสิ้น 1,000 คน แบ่งเป็นกลุ่มอ้างอิง 500 คน และกลุ่มเปรียบเทียบ 500 คน ส่วนเงื่อนไขที่สอง จำนวนกลุ่มตัวอย่าง 1,000 คนเท่ากัน แบ่งเป็นกลุ่มอ้างอิง 900 คน และกลุ่มเปรียบเทียบ 100 คน การแจกแจงความสามารถที่แตกต่างกัน พิจารณาจากการแจกแจงแบบปกติที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ส่วนเงื่อนไขข้อสอบแบ่งเป็น 3 แบบ คือ ไม่เกิดทำหน้าที่ต่างกันของข้อสอบ (No DIF) เกิดการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทาง (Unidirectional DIF) และเกิดการทำหน้าที่ต่างกันของข้อสอบแบบไม่มีทิศทาง (Nondirectional DIF) วิเคราะห์ข้อมูลด้วยวิธีถดถอยโลจิสติก (Logistic Regression) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT Likelihood Ratio Test) และ วิธีแคทซิบ (CATSIB) ผลการศึกษา พบว่า วิธีถดถอยโลจิสติก (Logistic Regression) และวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT Likelihood Ratio Test) ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทาง (Unidirectional DIF) และแบบไม่มีทิศทาง (Nondirectional DIF) ได้ดีเท่าๆ กัน และทั้งสองวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธีแคทซิบ (CATSIB) ในขณะที่วิธีแคทซิบ (CATSIB) ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทาง (Unidirectional DIF) ได้ดีกว่าแบบไม่มีทิศทาง (Nondirectional DIF)

Park (2006) ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านภาษาและเพศในการทดสอบการเขียนความเรียง MELAB ซึ่งเป็นการวัดความสามารถทางภาษาอังกฤษของรัฐมิชิแกนประเทศสหรัฐอเมริกา ประกอบด้วยการวัดทักษะการอ่าน การฟังและไวยากรณ์ กลุ่มตัวอย่างมีเวลาเขียนเรียงความ 30 นาที โดยเลือกเขียนเรียงความ 1 หัวข้อ จาก 2 หัวข้อที่กำหนดให้ กลุ่มตัวอย่างในการศึกษาครั้งนี้มีจำนวน 2,269 คน เป็นเพศชาย 686 คนและเป็นเพศหญิง 1,583 คน วิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิติก (Logistic Regression) แบบ 3 ขั้นตอนในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform DIF) และแบบอเนกรูป (Nonuniform DIF) ผลการศึกษา พบว่า ไม่เกิดการทำหน้าที่ต่างกันของแบบสอบ MELAB

Penfield (2006) ได้ศึกษาการประมาณค่าอิทธิพลของการทำหน้าที่ต่างกันในการวัดการทำหน้าที่ต่างกันของแบบสอบ (DTF) โดยไม่คิดเครื่องหมายในแบบสอบแบบผสม (mixed format test) การศึกษาครั้งนี้เป็นการจำลองข้อมูล ซึ่งประกอบด้วยข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) และข้อสอบที่ตรวจให้คะแนนแบบหลายค่า (polytomous) 4 ตัวเลือก การวิเคราะห์ข้อมูลแบ่งออกเป็น 2 กรณี คือ กรณีแรก แบบสอบที่ประกอบด้วยข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่าจำนวน 20

ข้อและข้อสอบที่ตรวจให้คะแนนแบบหลายค่าจำนวน 8 ข้อ กรณีที่สอง แบบสอบที่ประกอบด้วยข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่าจำนวน 8 ข้อและข้อสอบที่ตรวจให้คะแนนแบบ หลายค่าจำนวน 12 ข้อ โดยข้อสอบ 2 ค่าวิเคราะห์แบบ 3 พารามิเตอร์ ส่วนข้อสอบหลายค่าวิเคราะห์ด้วยวิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) กลุ่มตัวอย่างจำนวน 1,000 คน แบ่งเป็นกลุ่มอ้างอิงจำนวน 500 คน และกลุ่มเปรียบเทียบจำนวน 500 คน โดยพิจารณาจากการแจกแจงแบบปกติที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 1 และค่าเฉลี่ยขึ้นอยู่กับเงื่อนไขทั้งหมด 40 เงื่อนไขที่แตกต่างกัน (2 ระดับค่าเฉลี่ยการแจกแจงความสามารถ × 2 ชนิดของแบบสอบ × 2 พารามิเตอร์โอกาสในการเดา × 5 ขนาดอิทธิพลของการทำหน้าที่ต่างกัน) ผลการศึกษาพบว่า แบบสอบที่มีข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่าจำนวนมากจะส่งผลต่อความลำเอียงทางลบแต่แบบสอบที่มีข้อสอบที่ตรวจให้คะแนนแบบหลายค่าจำนวนมากจะส่งผลต่อความลำเอียงทางบวกเพียงเล็กน้อยเท่านั้น

Cohen and al (2007) ได้ศึกษาการทำหน้าที่ต่างกันและขนาดอิทธิพลของข้อสอบที่ตรวจให้คะแนนแบบหลายค่า (polytomous) โดยใช้กลุ่มตัวอย่างขนาดใหญ่ (N=105,731) เพื่อเปรียบเทียบถึงความสอดคล้องตามวิธีการตรวจสอบทั้ง 5 วิธี ได้แก่วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT Likelihood Ratio Test) วิธีการถดถอยโลจิสติก (Logistic Regression) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) วิธีแมนเทล (Mantel) และ วิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) โดยใช้โปรแกรม MULTILOG และโปรแกรม IRTLRDIF ในการวิเคราะห์ด้วยวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT Likelihood Ratio Test) โปรแกรม SAS ใช้ในการวิเคราะห์ด้วยวิธีการถดถอยโลจิสติก (Logistic Regression) และวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด (Likelihood Ratio Test) ส่วนวิธีแมนเทล (Mantel) และวิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) เขียนโปรแกรมด้วยภาษาฟอร์เทรน ผลการศึกษา พบว่า ตรวจพบการทำหน้าที่ต่างกันของข้อสอบทั้ง 10 ข้อจากทั้ง 5 วิธีและได้ข้อค้นพบที่สำคัญ คือ การใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินไปจะไม่มีประโยชน์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

Elosua and Jaúregui (2007) ได้ศึกษาแหล่งของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อการแปลแบบสอบ การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อหาแหล่งของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อการแปลแบบสอบ โดยจำแนกบนพื้นฐานของเกณฑ์ทางภาษาและวัฒนธรรม 4 แบบ คือ 1) ความเกี่ยวข้องทางวัฒนธรรม 2) ปัญหาการแปล 3) ไวยากรณ์ และ 4) การตีความหมายคำ การศึกษาในครั้งนี้ใช้ข้อคำถามที่ตรวจให้คะแนนแบบ 2 ค่าเกี่ยวกับภาษาจำนวน 53 ข้อ ซึ่งเป็นข้อคำถามที่สร้างเป็นภาษาสเปนจากนั้นแปลเป็นภาษาบาสก์ และมีการแปลย้อนกลับอีกครั้ง (back-translation) กลุ่มตัวอย่างที่ศึกษาเป็นนักเรียนอายุระหว่าง 9 – 11 ปี จำนวน 1,048 คน แบ่งเป็นกลุ่มอ้างอิง 498 คน และกลุ่ม

เปรียบเทียบ 550 คน วิเคราะห์ข้อมูลด้วยวิธีแมนเทล-แฮนเซล (Mantel-Haenszel) และจากความเห็นของผู้เชี่ยวชาญ (expert judgment) ผลการศึกษา พบว่า เกณฑ์ทั้ง 4 แบบ คือ 1) ความเกี่ยวข้องทางวัฒนธรรม (cultural relevance) 2) ปัญหาในการแปล (translation problems) 3) ไวยากรณ์ (grammar) และ 4) การตีความหมายคำ (semantic differences) ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบทั้งสิ้น และวิธีแมนเทล-แฮนเซลตรวจสอบพบว่าการทำหน้าที่ต่างกันของข้อสอบได้ทั้งสิ้น 32 ข้อ ผู้เชี่ยวชาญตรวจสอบพบว่าการทำหน้าที่ต่างกันของข้อสอบได้ 28 ข้อ และมีข้อคำถามที่ทั้งผู้เชี่ยวชาญและวิธีแมนเทล-แฮนเซลตรวจสอบพบว่าเกิดการทำหน้าที่ต่างกันของข้อสอบได้ตรงกันจำนวน 22 ข้อ และมีแหล่งของการทำหน้าที่ต่างกันของข้อสอบทั้งสิ้น 29 แหล่ง

งานวิจัยในประเทศ

กาญจนา วัธนสุนธร (2537) ได้พัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์ ด้วยวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และวิธีซิบเทสท์ (SIBTEST) โดยใช้ข้อมูลการตอบข้อสอบคัดเลือกศึกษาต่อในระดับอุดมศึกษา ปีการศึกษา 2535 ความยาวของแบบสอบ 20, 30 และ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 50, 60, 70 และ 80 ข้อสำหรับวิชาภาษาอังกฤษ และใช้ผู้สอบขนาด 100, 200, 400, 600, 800 และ 1,000 คน ผลการศึกษา พบว่า มีความไม่คงที่ข้ามขนาดผู้สอบและความยาวแบบสอบ ความสอดคล้องในการตรวจสอบข้อสอบลำเอียงภายในวิธีเดียวกันข้ามขนาดผู้สอบค่อนข้างต่ำ แต่จะสูงขึ้นเมื่อขนาดผู้สอบ 600 คนขึ้นไป ส่วนการวิเคราะห์ความลำเอียงของข้อสอบที่มีต่อเพศ พบว่า ข้อสอบวิชาภาษาอังกฤษลำเอียงเข้าข้างผู้หญิง ส่วนข้อสอบวิชาภาษาคณิตศาสตร์ลำเอียงเข้าข้างผู้ชาย

เกสร หว่างจิตร (2539) ศึกษาการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษาวิชาภาษาไทยและวิชาภาษาอังกฤษด้วยวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) โดยกลุ่มผู้สอบจำแนกตามเพศ ภูมิลำเนา ประสบการณ์ในการสอน และสังกัดสถานศึกษา ข้อมูลที่ใช้เป็นผลการตอบข้อสอบวิชาสอบร่วมในส่วนที่เป็นข้อสอบแบบเอกรูปของศูนย์ทดสอบทางภาษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย กลุ่มตัวอย่างเป็นผู้สอบวิชาภาษาไทยจำนวน 506 คน และวิชาภาษาอังกฤษจำนวน 501 คน ผลการศึกษา พบว่า ข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่เป็นแบบอเนกรูป (Nonuniform DIF) เมื่อวิเคราะห์ข้อสอบจำแนกตามเพศ พบว่า มีข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบเอกรูป (Uniform DIF)และแบบอเนกรูป (Nonuniform DIF) มากที่สุด รองลงมาคือ จำแนกตามภูมิลำเนา สังกัดสถานศึกษาและประสบการณ์ในการสอนตามลำดับ เมื่อพิจารณาข้อสอบที่ทำหน้าที่ต่างกัน พบว่า ส่วนใหญ่เป็นข้อสอบที่มีค่าอำนาจจำแนกค่อนข้างต่ำทั้งสองวิชา โดยข้อสอบที่ทำหน้าที่ต่างกันในวิชา

ภาษาไทยส่วนใหญ่เป็นข้อสอบที่ง่าย ส่วนข้อสอบที่ทำหน้าที่ต่างกันในวิชาภาษาอังกฤษส่วนใหญ่เป็นข้อสอบที่ยากมาก

จิติมา วรรณศรี (2539) ศึกษาการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel กับวิธีซิบเทสท์ (SIBTEST) เมื่อความยาวของแบบสอบ ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน โดยการจำลองข้อมูล ความยาวของแบบสอบ 3 ขนาด คือ 30, 60 และ 90 ข้อ ขนาดกลุ่ม 3 ขนาด คือ 200, 600 และ 1,000 คน และกำหนดอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 1:1, 1:0.9, 1:0.75 และ 1:0.5 ผลการศึกษา พบว่า วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) กับวิธีซิบเทสท์ (SIBTEST) มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในทุกขนาดกลุ่มตัวอย่างและทุกอัตราส่วนภายใต้แบบสอบความยาวเท่ากันและเมื่อใช้แบบสอบที่มีความยาว 60 ข้อนั้นทั้งสองวิธีสามารถตรวจสอบได้ดีที่สุดและวิธีซิบเทสท์ (SIBTEST) มีอัตราความคลาดเคลื่อนชนิดที่ 1 มากกว่าวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) เล็กน้อย

เสรี ชัดแช้ม (2539) ศึกษาการเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลแบบปกติกับวิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของแบบสอบ จากผลการตอบแบบวัดความสามารถในการอ่านภาษาไทยของนักเรียนระดับมัธยมศึกษาปีที่ 1 จังหวัดชลบุรี จำนวน 1,200 คน จำแนกตามเพศ ผลการศึกษา พบว่า วิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถผู้สอบและความยากของข้อสอบสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) ได้สอดคล้องกับวิธี IRT area และตรวจสอบได้เพิ่มขึ้นจากวิธีแมนเทล-แฮนส์เซลแบบปกติ ข้อสอบที่ตรวจพบส่วนใหญ่เป็นข้อสอบที่มีความยากระดับปานกลางและข้อสอบง่าย ซึ่งมีโค้งลักษณะข้อสอบ (ICC) ของผู้สอบทั้งสองกลุ่มตัดกันบริเวณใกล้ๆ จุดกลางของช่วงความสามารถ

ญาณภัทร สีหะมงคล (2540) ศึกษาการเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord' s $\chi 2$ วิธี Raju' s Area Measures และวิธี Closed Interval Area เมื่อขนาดของกลุ่มตัวอย่าง ความยาวของแบบสอบและสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบแตกต่างกัน กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 4 จำนวน 11,404 คน เครื่องมือที่ใช้เป็นแบบสอบจำนวน 80 ข้อ ผลการศึกษา พบว่า 1) ข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord' s $\chi 2$ วิธี Raju' s Area Measures และวิธี Closed Interval Area แตกต่างกันเมื่อขนาดกลุ่มตัวอย่างและความยาวของแบบสอบต่างกัน 2) ความสอดคล้องในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ ส่วนมากมีค่าปานกลางถึงต่ำในเกือบทุกเงื่อนไข 3) ความสัมพันธ์ของวิธีการตรวจสอบข้อสอบที่

ทำหน้าที่ต่างกันทั้ง 3 วิธี โดยภาพรวมมีสัมประสิทธิ์สหสัมพันธ์ค่อนข้างสูงมากและมีนัยสำคัญทางสถิติเกือบทุกเงื่อนไขของการศึกษา และ 4) อัตราความไม่สอดคล้องในการตรวจสอบ เมื่อแบบสอบยาวมากขึ้นและมีสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากขึ้น อัตราความไม่สอดคล้องชนิดที่ 1 และอัตราความไม่สอดคล้องชนิดที่ 2 ส่วนมากมีค่าใกล้เคียงกัน

พรรณี จิตมาศ (2540) ได้วิเคราะห์ความลำเอียงต่อเพศของแบบสอบวิชาคณิตศาสตร์ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และวิธีซิบเทสท์ (SIBTEST) ในแต่ละขนาดของกลุ่มผู้สอบ 500 และ 1,000 คน โดยเปรียบเทียบจำนวนข้อที่มีความลำเอียงและเปรียบเทียบค่าความเชื่อมั่นแบบครึ่งฉบับของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ผลการศึกษา พบว่า เมื่อวิเคราะห์จากกลุ่มตัวอย่างขนาด 500 คน วิธีซิบเทสท์ (SIBTEST) พบข้อสอบที่มีความลำเอียงมากที่สุดและวิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียงน้อยที่สุด และเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) พบข้อสอบมีความลำเอียงมากที่สุด วิธีแปลงค่าความยากไม่พบข้อสอบที่ลำเอียง โดยจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากกับวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และวิธีแปลงค่าความยากกับวิธี ซิบเทสท์ (SIBTEST) นั้น แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

รัชรินทร์ มุคดา (2540) ศึกษาการเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) กับวิธีการถดถอยโลจิสติก (Logistic Regression) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF) โดยการจำลองข้อมูลขึ้น ภายใต้เงื่อนไขอันได้แก่ 1) กลุ่มความสามารถ 3 ระดับ คือ กลุ่มผู้สอบที่มีความสามารถสูง ปานกลาง และต่ำ 2) ค่าความยากของแบบสอบ 3 ระดับ คือ ข้อสอบที่มีค่าความยากสูง ปานกลาง และต่ำ และ 3) ค่าอำนาจจำแนกของข้อสอบ 3 ระดับ คือ ข้อสอบที่มีค่าอำนาจจำแนกสูง ปานกลาง และต่ำ รวมเงื่อนที่ศึกษาทั้งสิ้น 27 เงื่อนไข ผลการศึกษา พบว่า วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) กับวิธีการถดถอยโลจิสติก (Logistic Regression) มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF) เท่ากันในกลุ่มผู้สอบที่มีความสามารถสูง ปานกลางและต่ำ โดยกลุ่มผู้สอบที่มีความสามารถสูงข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) มากที่สุดเป็นข้อสอบที่มีค่าความยากสูงและค่าอำนาจจำแนกสูงสำหรับกลุ่มผู้สอบที่มีความสามารถในระดับปานกลาง ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) มากที่สุดเป็นข้อสอบที่มีค่าความยากปานกลางและค่าอำนาจจำแนกสูงและกลุ่มผู้สอบที่มีความสามารถต่ำและข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป (Nonuniform DIF) มากที่สุดเป็นข้อสอบที่มีความยากต่ำและค่าอำนาจจำแนกสูง

นิคม กีรติวากูร (2542) เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และวิธีการตอบสนองข้อสอบ (IRT) แบบ 2 พารามิเตอร์ โดยเทียบกับเกณฑ์ที่กำหนด ศึกษาจากการจำลองข้อมูล ปัจจัยที่ศึกษาได้แก่ ขนาดกลุ่มตัวอย่าง 2 ขนาด คือ ขนาดเล็ก (300 คน) และขนาดใหญ่ (1,000 คน) ค่าความยาวแบบสอบ 2 ขนาด คือ แบบสอบสั้น (25 ข้อ) และแบบสอบยาว (75 ข้อ) ค่าความยากของข้อสอบแบ่งออกเป็น 3 ระดับคือ กลุ่มข้อสอบที่มีความยากสูง ปานกลาง และต่ำ ขนาดความลำเอียงของข้อสอบแบ่งออกเป็น 2 ขนาดคือ กลุ่มข้อสอบที่มีความลำเอียงสูงและต่ำ ผลการศึกษาพบว่า วิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงที่สุดโดย วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงภายใต้เงื่อนไขแบบสอบที่มีความยากต่ำ อำนาจจำแนกสูง ส่วนวิธี IRT แบบ 2 พารามิเตอร์ มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง ภายใต้เงื่อนไขแบบสอบที่มีความยากต่ำและวิธี IRT มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และวิธี RFA ตามลำดับ

วลีมาศ แซ่อึ้ง (2543) ศึกษาการเปรียบเทียบอำนาจการทดสอบและอัตราส่วนความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF) ระหว่างวิธีซิบเทสท์ปรับใหม่ (Modified SIBTEST) วิธีซิบเทสท์ (SIBTEST) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) วิธีการถดถอยโลจิสติก (Logistic Regression) ภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ โดยกำหนดค่าฏอกาสในการเดาคงที่ แล้วจัดกระทำข้อมูล 4 ตัวแปร คือ 1) ลักษณะข้อสอบที่มีค่าความยาก และอำนาจจำแนกระดับสูง ปานกลางและต่ำ จำนวน 9 ลักษณะ 2) ความยาวของแบบสอบ 2 ระดับ 3) สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกัน 3 ระดับ และ 4) ขนาดกลุ่มตัวอย่าง 6 ระดับ รวมข้อมูลที่ศึกษาทั้งสิ้น 324 เงื่อนไข ผลการศึกษา พบว่า อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป (Nonuniform DIF) ของวิธีซิบเทสท์ปรับใหม่ (Modified SIBTEST) และ วิธีการถดถอยโลจิสติก (Logistic Regression) มีค่าเท่าเทียมกันเกือบทุกเงื่อนไขและทั้งสองวิธีมีอำนาจการทดสอบสูงกว่าวิธีซิบเทสท์ (SIBTEST) และวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) เกือบทุกเงื่อนไข และอัตราส่วนความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของวิธีซิบเทสท์ปรับใหม่ (Modified SIBTEST) วิธีซิบเทสท์ (SIBTEST) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และวิธีการถดถอยโลจิสติก (Logistic Regression) มีค่าอยู่ภายในเกณฑ์ของความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10% เกือบทุกเงื่อนไข

สุมาลี แกล้วทนงค์ (2547) ศึกษาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนา และวัฒนธรรมต่างกัน กลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 1 การศึกษา 2546 สำนักงานเขตพื้นที่การศึกษาสงขลา พัทลุง ตรังและสตูล จำนวน 1,320 คน วิเคราะห์ข้อมูลด้วยวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) และวิธีซิบเทสท์ (SIBTEST) ผลการศึกษา พบว่า แบบสอบกลุ่มสาระการเรียนรู้ภาษาไทย 3 ฉบับ จำนวน 120 ข้อ มีข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ 9 ข้อ ด้านภาษาพูด 15 ข้อ และด้านเชื้อชาติ 28 ข้อ และกลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม มีข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ 22 ข้อ ด้านภาษาพูด 52 ข้อ และด้านเชื้อชาติ 20 ข้อ โดยข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ ส่วนใหญ่มีสาเหตุมาจากเนื้อหาและภาษาที่ใช้ในแบบสอบ ข้อสอบที่ทำหน้าที่ต่างกันด้านภาษาพูด ส่วนใหญ่มีสาเหตุมาจากการใช้คำศัพท์เฉพาะและบริบททางวัฒนธรรมและข้อสอบที่ทำหน้าที่ต่างกันด้านเชื้อชาติ ส่วนใหญ่มีสาเหตุมาจากบริบททางภาษาและบริบททางวัฒนธรรม

จากเอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ข้างต้น สรุปได้ดังตารางสังเคราะห์ที่ 2.5

ตารางที่ 2.5 การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF)

ชื่อผู้วิจัย (ปี)	วัตถุประสงค์	ผลการศึกษา
Swaminathan and Rogers (1990)	เปรียบเทียบวิธีถดถอยโลจิสติกกับวิธีแมนเทล-แฮนส์เซล	วิธีถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล ให้ผลการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป กล่าวคือ มีการตรวจสอบได้ถูกต้องร้อยละ 70 กรณีกลุ่มตัวอย่าง 250 คน และตรวจสอบได้ร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน ในทุกความยาวของแบบสอบ สำหรับการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป พบว่า วิธีแมนเทล-แฮนส์เซล ตรวจสอบได้เล็กน้อย ส่วนวิธีถดถอยโลจิสติกตรวจสอบได้ถูกต้องร้อยละ 50 กรณีกลุ่มตัวอย่างน้อยและข้อสอบสั้นและถูกต้องร้อยละ 75 กรณีแบบสอบยาวและกลุ่มตัวอย่างขนาดใหญ่
Rogers and Swaminathan (1993)	เปรียบเทียบวิธีถดถอยโลจิสติกกับวิธีแมนเทล-แฮนส์เซล	ทั้งสองวิธีมีประสิทธิภาพในการตรวจสอบเท่ากันในการตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป วิธีถดถอยโลจิสติกตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปได้ดีในกรณีที่ข้อสอบมีความยากปานกลางและอำนาจจำแนกสูง ส่วนวิธีแมนเทล-แฮนส์เซล ตรวจสอบข้อสอบที่มีความยากปานกลางได้น้อยมากแต่สามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้ดีในกรณีที่ข้อสอบง่ายมากหรือข้อสอบที่ยากมาก
Mazor and al (1994)	ศึกษาการใช้วิธีแมนเทล-แฮนส์เซล	เมื่อระดับความแตกต่างของค่าอำนาจจำแนกและในการตรวจสอบการทำหน้าที่ต่างกัน ค่าระหว่างกลุ่มเพิ่มขึ้น จะทำให้อัตราการตรวจสอบแบบเอกรูป พบข้อสอบที่ทำหน้าที่ต่างกันมากขึ้น

ตารางที่ 2.5 การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (ต่อ)

ชื่อผู้วิจัย (ปี)	วัตถุประสงค์	ผลการศึกษา
Uttaro and Millsap (1994)	ศึกษาปัจจัยที่มีอิทธิพลต่อวิธีแมนเทล-แฮนส์เซล	ภายใต้เงื่อนไขข้อสอบที่ทำหน้าที่ต่างกันค่าความยาก ค่าอำนาจจำแนก ค่าโอกาสในการเดา การกระจายความสามารถและปฏิสัมพันธ์ระหว่างการกระจายความสามารถกับค่าพารามิเตอร์ของแบบสอบ มีผลต่อการประมาณค่า α_{MH} และข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่เป็นข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปมากกว่าแบบอเนกรูป
French and Miller (1996)	ศึกษาความเป็นไปได้ของการใช้วิธีการถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่า	การเปลี่ยนแปลงขนาดของกลุ่มตัวอย่าง มีผลต่ออำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อกลุ่มตัวอย่างมีขนาดเล็กลง อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจะลดลง เมื่อค่าพารามิเตอร์อำนาจจำแนกของข้อสอบยิ่งแตกต่างกันมาก อำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปจะยิ่งเพิ่มขึ้น
Narayanan and Swaminathan (1996)	เปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซล วิธีถดถอยโลจิสติกและวิธีโครซิบท์ในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป	วิธีถดถอยโลจิสติกและวิธีโครซิบท์ ให้ผลใกล้เคียงกันในการตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป และทั้งสองวิธีตรวจสอบการทำหน้าที่ต่างกันได้ดีกว่าวิธีแมนเทล-แฮนส์เซล
Roussos and Stout (1996)	ศึกษาผลของกลุ่มตัวอย่างขนาดเล็กที่มีต่อความคลาดเคลื่อนชนิดที่ 1 ของวิธีซิบเทสท์และวิธีแมนเทล-แฮนส์เซล	ค่าสถิติของวิธีซิบเทสท์และวิธีวิธีแมนเทล-แฮนส์เซลมีแนวโน้มที่จะมีความคลาดเคลื่อนชนิดที่ 1 เพิ่มขึ้นเมื่อขนาดกลุ่มตัวอย่างมีความแตกต่างของค่าเฉลี่ยระหว่างกลุ่มเพิ่มขึ้น
Kim (2000)	ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3	ทั้ง 3 วิธี ให้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีเมื่อกลุ่มตัวอย่างมีขนาด 100 คน และการใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินไปจะไม่ส่งผลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
Penfield (2001)	ศึกษาขนาดของความคลาดเคลื่อนชนิดที่ 1 ด้วยวิธีแมนเทล-แฮนส์เซล 3 แบบ	วิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) ดีที่สุดในการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในหลายกลุ่ม

ตารางที่ 2.5 การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (ต่อ)

ชื่อผู้วิจัย (ปี)	วัตถุประสงค์	ผลการศึกษา
Bolt (2002)	เปรียบเทียบวิธีการตรวจจับการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพาราเมตริกและนันพาราเมตริก	วิธี GRM ให้ผลที่สอดคล้องกับข้อมูลเชิงประจักษ์มากที่สุด วิธีGRM-LR ให้ค่าความ หลายค่าด้วยสถิติประเภทที่ 1 สูงกว่าวิธีอื่นและวิธี GRM-DFIT ให้ค่าความคลาดเคลื่อนประเภทที่ 1 ต่ำที่สุด
Cohen and Bolt (2005)	ศึกษาการทำหน้าที่ต่างกันของข้อสอบตามตัวแปรเพศเมื่อใช้แบบสอบแบบผสม	มีข้อสอบที่ทำหน้าที่ต่างกันจำนวน 5 ข้อ โดยข้อสอบจำนวน 4 ข้อเข้าข้างเพศชายและข้อสอบอีก 1 ข้อเข้าข้างเพศหญิง
Finch (2005)	เปรียบเทียบโมเดล MIMIC กับวิธีซิบเทสท์ วิธีแมนเทล-แฮนส์เซลและวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT	โมเดล MIMIC ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีในกรณีที่ข้อสอบมีจำนวน 50 ข้อ แบบ 2 พารามิเตอร์และโมเดล MIMIC สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้สูงในกรณีที่ข้อสอบมีจำนวน 20 ข้อ แบบ 3 พารามิเตอร์โลจิสติก
Lei and al (2006)	ศึกษาการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์	วิธีถดถอยโลจิสติกและวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้งแบบมีทิศทางและแบบไม่มีทิศทางได้ดีเท่าๆ กันและทั้งสองวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธีแคทซิบ ในขณะที่วิธีแคทซิบตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทาง ได้ดีกว่าแบบไม่มีทิศทาง
Park (2006)	ศึกษาการทำหน้าที่ต่างกันของข้อสอบด้านภาษาและเพศในการทดสอบการเขียนความเรียง MELAB	ไม่เกิดการทำหน้าที่ต่างกันของแบบสอบ MELAB
Penfield (2006)	ศึกษาการประมาณค่าอิทธิพลของการทำหน้าที่ต่างกันในการวัดการทำหน้าที่ต่างกันของแบบสอบ (DTF) โดยไม่คิดเครื่องหมายในแบบสอบแบบผสม (mixed format test)	แบบสอบที่มีข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่าจำนวนมากจะส่งผลต่อความลำเอียงทางลบแต่แบบสอบที่มีข้อสอบที่ตรวจให้คะแนนแบบหลายค่าจำนวนมากจะส่งผลต่อความลำเอียงทางบวกเพียงเล็กน้อยเท่านั้น

ตารางที่ 2.5 การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (ต่อ)

ชื่อผู้วิจัย (ปี)	วัตถุประสงค์	ผลการศึกษา
Cohen and al (2007)	ศึกษาการทำหน้าที่ต่างกันและขนาดอิทธิพลของข้อสอบที่ตรวจให้คะแนนแบบหลายค่าโดยใช้กลุ่มตัวอย่างขนาดใหญ่	ตรวจพบการทำหน้าที่ต่างกันของข้อสอบทั้ง 10 ข้อ จากทั้ง 5 วิธีและได้ข้อค้นพบที่สำคัญ คือ การใช้กลุ่มตัวอย่างที่มีขนาดใหญ่เกินไปจะไม่มีประโยชน์ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
Elosua and Jaúregui (2007)	ศึกษาแหล่งของการทำหน้าที่ต่างกันของข้อสอบที่ส่งผลต่อการแปลแบบสอบ	เกณฑ์ทั้ง 4 แบบส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบทั้งสิ้น และวิธีแมนเทล-แฮนเซลตรวจสอบพบว่าการทำหน้าที่ต่างกันของข้อสอบได้ทั้งสิ้น 32 ข้อ ผู้เชี่ยวชาญตรวจสอบพบว่าการทำหน้าที่ต่างกันของข้อสอบได้ 28 ข้อ และมีข้อคำถามที่ทั้งผู้เชี่ยวชาญและวิธีแมนเทล-แฮนเซลตรวจสอบพบว่าเกิดการทำหน้าที่ต่างกันของข้อสอบได้ตรงกันจำนวน 22 ข้อ และมีแหล่งของการทำหน้าที่ต่างกันของข้อสอบทั้งสิ้น 29 แหล่ง
กาญจนา วัธนสุนธร (2537)	พัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศด้วยข้อมูลเชิงประจักษ์ ด้วยวิธีแมนเทล-แฮนส์เซลและวิธีซิบเทสท์	มีความไม่คงที่ข้ามขนาดผู้สอบและความยาวแบบสอบความสอดคล้องในการตรวจสอบข้อสอบลำเอียงภายในวิธีเดียวกันข้ามขนาดผู้สอบค่อนข้างต่ำ แต่จะสูงขึ้นเมื่อขนาดผู้สอบ 600 คนขึ้นไป เมื่อวิเคราะห์ความลำเอียงของข้อสอบที่มีต่อเพศ พบว่า ข้อสอบวิชาภาษาอังกฤษลำเอียงเข้าข้างผู้หญิง ส่วนข้อสอบวิชาภาษาคณิตศาสตร์ลำเอียงเข้าข้างผู้ชาย
เกสร หว่างจิตร (2539)	ศึกษาการทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษาวิชาภาษาไทยและวิชาภาษาอังกฤษด้วยวิธีแมนเทล-แฮนส์เซล	ข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่เป็นแบบอเนกรูปเมื่อวิเคราะห์ข้อสอบจำแนกตามเพศ พบว่า มีข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบเอกรูปและแบบอเนกรูป

ตารางที่ 2.5 การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (ต่อ)

ชื่อผู้วิจัย (ปี)	วัตถุประสงค์	ผลการศึกษา
จิติมา วรรณศรี (2539)	เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์	วิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในทุกขนาดกลุ่มตัวอย่างและทุกความยาวที่เท่ากันของแบบสอบและวิธีซิบเทสท์ มีอัตราความคลาดเคลื่อนชนิดที่ 1 มากกว่าวิธีแมนเทล-แฮนส์เซลเล็กน้อย
เสรี ชัดแช้ม (2539)	เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลแบบปกติกับวิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของแบบสอบ	วิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถผู้สอบและความยากของข้อสอบสามารถตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปได้สอดคล้องกับวิธี IRT area และตรวจสอบได้เพิ่มขึ้นจากวิธีแมนทล-แฮนส์เซลแบบปกติ ข้อสอบที่ตรวจพบส่วนใหญ่เป็นข้อสอบที่มีความยากระดับปานกลางถึงง่าย
ญาณภัทร สีหะมงคล (2540)	เปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord' s $\chi 2$ วิธี Raju' s Area Measures และวิธี Closed Interval Area	ข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord' s $\chi 2$ วิธี Raju' s Area Measures และวิธี Closed Interval Area แตกต่างกันเมื่อขนาดกลุ่มตัวอย่างและความยาวของแบบสอบต่างกัน
พรรณี จิตมาศ (2540)	วิเคราะห์ความลำเอียงต่อเพศของแบบสอบวิชาคณิตศาสตร์ ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธีแมนเทล-แฮนส์เซลและวิธีซิบเทสท์	เมื่อวิเคราะห์จากกลุ่มตัวอย่างขนาด 500 คน วิธีซิบเทสท์พบข้อสอบที่มีความลำเอียงมากที่สุด และวิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียงน้อยที่สุด และเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน วิธีแมนเทล-แฮนส์เซลพบข้อสอบมีความลำเอียงมากที่สุด วิธีแปลงค่าความยากไม่พบข้อสอบที่ลำเอียง

ตารางที่ 2.5 การสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) (ต่อ)

ชื่อผู้วิจัย (ปี)	วัตถุประสงค์	ผลการศึกษา
รัชรินทร์ มุคดา (2540)	เปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีการถดถอยโลจิสติก ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป	วิธีแมนเทล-แฮนส์เซลกับวิธีการถดถอยโลจิสติกมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปเท่ากันในกลุ่มผู้สอบที่มีความสามารถสูงปานกลางและต่ำ ในกลุ่มผู้สอบที่มีความสามารถสูง ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปมากที่สุดเป็นข้อสอบที่มีค่าความยากสูงและค่าอำนาจจำแนกสูง
นิคม กีรติวากูร (2542)	เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด วิธีแมนเทล-แฮนส์เซลและวิธีการตอบสนองข้อสอบแบบ 2 พารามิเตอร์	วิธี RFA มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงที่สุดโดย วิธีแมนเทล-แฮนส์เซลมีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูงภายใต้แบบสอบที่มีความยากต่ำ อำนาจจำแนกสูง
วลีมาศ แซ่อึ้ง (2543)	เปรียบเทียบอำนาจการทดสอบและอัตราส่วนความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิบเทสท์ปรับใหม่ วิธีซิบเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก	อำนาจการทดสอบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ของวิธีซิบเทสท์ปรับใหม่และวิธีการถดถอยโลจิสติกมีค่าเท่าเทียมกันเกือบทุกเงื่อนไขและทั้งสองวิธีมีอำนาจการทดสอบสูงกว่าวิธีซิบเทสท์และวิธีแมนเทล-แฮนส์เซลเกือบทุกเงื่อนไข และอัตราส่วนความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูปของวิธีซิบเทสท์ปรับใหม่ วิธีซิบเทสท์ วิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติกมีค่าอยู่ภายในเกณฑ์ของความคลาดเคลื่อนประเภทที่ 1 ที่ระดับ 10%
สุมาลี แกล้วทนงค์ (2547)	ศึกษาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรมต่างกันด้วยวิธีแมนเทล-แฮนส์เซลและวิธีซิบเทสท์	แบบสอบกลุ่มสาระการเรียนรู้ภาษาไทย 3 ฉบับ จำนวน120 ข้อ มีข้อสอบที่ทำหน้าที่ต่างกันด้านเพศ 9 ข้อด้านภาษาพูด 15 ข้อและด้านเชื้อชาติ 28 ข้อ

จากตารางที่ 2.5 สรุปประเด็นได้ ดังนี้ 1) วัตถุประสงค์ของการศึกษา สรุปได้ 2 ประเด็นที่ชัดเจน คือ ประเด็นแรกมุ่งศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) ส่วนประเด็นที่สอง เป็นการเปรียบเทียบประสิทธิภาพของแต่ละโปรแกรมหรือแต่ละเทคนิควิธีของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) 2) ลักษณะของข้อมูลที่นำมาศึกษา หากพิจารณาในประเด็นของการได้มาซึ่งข้อมูลในการวิเคราะห์นั้น กลุ่มแรกเป็นงานวิจัยที่ได้จากการเก็บรวบรวมข้อมูลจริงที่ประกอบไปด้วยข้อมูลปฐมภูมิ (primary data) หรือข้อมูลทุติยภูมิ (secondary data) และกลุ่มที่สองเป็นงานวิจัยที่อาศัยข้อมูลจำลอง (simulation data) 3) เครื่องมือที่ผู้วิจัยใช้ในการเก็บรวบรวมข้อมูลนั้น มีทั้งข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) และข้อสอบที่ตรวจให้คะแนนแบบมากกว่า 2 ค่า (polytomous) 4) ตัวแปรที่ศึกษาประกอบไปด้วย 1) ตัวแปรเพศ 2) ตัวแปรค่าความยากของแบบสอบ 3) ตัวแปรค่าอำนาจจำแนกของแบบสอบ 4) ตัวแปรความยาวของแบบสอบ 5) ตัวแปรขนาดของกลุ่มตัวอย่าง 6) ตัวแปรค่าโอกาสในการเดา 7) ตัวแปรการกระจายความสามารถ 8) ตัวแปรขนาดของข้อสอบที่ทำหน้าที่ต่างกัน 9) ตัวแปรอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบกลุ่มความสามารถ 10) ตัวแปรภาษาพูด 11) ตัวแปรสัดส่วนข้อสอบที่ DIF 12) ตัวแปรประสบการณ์ในการสอน 13) ตัวแปรเชื้อชาติ 14) ตัวแปรเกณฑ์การจับคู่ 15) ตัวแปรขนาดความลำเอียงของข้อสอบและ16) ตัวแปรวิธีในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ส่วน 5) วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) นั้น ประกอบไปด้วย 1) วิธีแมนเทล(Mantel) 2) วิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel) 3) วิธีแมนเทล-แฮนส์เซลแบบทั่วไป (GMH) 4) วิธีซิบเทสท์ (SIBTEST) 5) วิธีโพลีซิบเทสท์ (Poly-SIBTEST) 6) วิธีซิบเทสท์ปรับใหม่ (Modified SIBTEST) 7) วิธีดีเอฟไอที (DFIT) 8) วิธีโครซิบ (CRO SIB) 9) วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ด 10) วิธีทดสอบไค-สแควร์ของ Lord (Lord' 9s χ2) 11) วิธีการวัดพื้นที่ของ Raju (Raju' s Area Measures) 12) วิธีการวิเคราะห์องค์ประกอบจำกัด 13) วิธี Closed Interval Area (ญาณภัทร สีหะมงคล, 2540) 14) วิธีการถดถอยโลจิสติก (Logistic Regression) และ 15) วิธีขอความเห็นของผู้เชี่ยวชาญ (Expert Judgement)

5.2) งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM พบว่า การศึกษาในเรื่องนี้มีมาตั้งแต่ปี ค.ศ. 1995 (Kamata, 1998, 2001, 2002, 2003; Beretvas and Williams, 2004; Cheong, 2006) แต่โมเดล HGLM มาเป็นที่รู้จักกันมากขึ้นใน ปี ค.ศ. 1998 จากการศึกษาของ Kamata (1998) จนในปัจจุบันโมเดลนี้เป็นที่รู้จักกันอย่างแพร่หลายในนามของโมเดล Kamata (Kamata's Model) โดยจากการสังเคราะห์

เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM จำนวนทั้งสิ้น 14 เรื่อง มีรายละเอียดดังต่อไปนี้

Kamata (1998) ศึกษาโมเดล HGLM แบบ 1 พารามิเตอร์ โดยเป็นการประยุกต์โมเดล HGLM สู่ทฤษฎีการตอบสนองข้อสอบ (IRT) การศึกษาในครั้งนี้ถือว่าเป็นการบุกเบิกศาสตร์และพัฒนาศาสตร์ของโมเดล HGLM นั่นคือ การศึกษาในครั้งนี้มีจุดมุ่งหมายที่จะพิสูจน์ว่าโมเดลราสช์ (Rasch model) เป็นกรณีหนึ่งของโมเดล HGLM และมุ่งประมาณค่าพารามิเตอร์ข้อสอบ (item parameter) และค่าพารามิเตอร์ผู้สอบ (person parameter) ด้วยโปรแกรม HLM โดยศึกษาจากข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) ผลการศึกษา พบว่า โมเดลราสช์ (Rasch model) เป็นกรณีหนึ่ง (special case) ของโมเดลการตอบสนองข้อสอบ (IRT model)

Kamata (2001) ศึกษาการวิเคราะห์ข้อสอบด้วยโมเดล HGLM การศึกษาในครั้งนี้ใช้กรอบแนวคิดของ GLM (Generalized Linear Model) ในการวิเคราะห์ข้อสอบ (item analysis) แบบ 2 ระดับ จากนั้นใช้กรอบแนวคิดของ HLM (Hierarchical Linear Model) ในการวิเคราะห์โมเดลแบบ 2 ระดับ การศึกษาครั้งนี้เป็นการศึกษาแบบพหุระดับ โดย กำหนดระดับที่ 1 เป็นระดับข้อสอบ (item level) และระดับที่ 2 เป็นระดับนักเรียน (person level) เมื่อวิเคราะห์แบบ 2 ระดับได้แล้วจะเป็นการขยาย (extend) การวิเคราะห์พหุระดับสู่การวิเคราะห์แบบ 3 ระดับ โดยเพิ่มการวิเคราะห์ในระดับที่สาม คือ ระดับโรงเรียน (school level) เข้าไปในโมเดลการวิเคราะห์ ข้อมูลในการวิเคราะห์ในครั้งนี้เป็นข้อมูลทุติยภูมิ (secondary data) จาก Third International Mathematics and Science Studies (TIMSS) ของฐานข้อมูลของประเทศสหรัฐอเมริกา กลุ่มตัวอย่างเป็นนักเรียนจำนวน 1,130 คน จาก 68 โรงเรียน เครื่องมือที่ใช้เป็นข้อสอบแบบเลือกตอบ (multiple choice) จำนวน 17 ข้อ วิเคราะห์ข้อมูลด้วยโปรแกรม HLM ผลการศึกษา พบว่า ในการวิเคราะห์ข้อสอบ (item analysis) แบบ 2 ระดับด้วยโมเดล HGLM มีความทัดเทียมกัน (equivalent) กับการวิเคราะห์ด้วยโมเดลราสช์ (Rasch model) และสามารถประยุกต์โมเดล HGLM สู่การวิเคราะห์ในระดับที่ 3 ได้ โดยการวิเคราะห์ในระดับที่ 3 หรือการวิเคราะห์ระดับโรงเรียน (school level) นี้จะมีความสำคัญเป็นอย่างยิ่งหากต้องการศึกษาอิทธิพลของโรงเรียนหรือเมื่อต้องการวิเคราะห์มูลค่าเพิ่ม (value-added) ของโรงเรียน

Kamata (2002) เสนอผลการศึกษากระบวนการในการจัดกระทำกับการวิเคราะห์การตอบสนองข้อสอบด้วยโมเดล HGLM การศึกษาในครั้งนี้มีจุดมุ่งหมายเพื่อนำเสนอข้อค้นพบว่าการวิเคราะห์พหุระดับด้วยโมเดลราสช์ (Rasch model) แบบ 2 ระดับมีความทัดเทียมกัน (equivalent) กับการวิเคราะห์ด้วยโมเดล HGLM แบบ 2 ระดับเมื่อกำหนดตัวแปรระดับที่ 2 เป็นระดับนักเรียน (person level) และการวิเคราะห์พหุระดับด้วยโมเดลราสช์ (Rasch model) แบบ 3 ระดับก็มีความทัดเทียมกัน

(equivalent) กับการวิเคราะห์ด้วยโมเดล HGLM แบบ 3 ระดับด้วย เมื่อกำหนดตัวแปรระดับที่ 3 เป็นระดับกลุ่ม (group level) โดยใช้ข้อมูลทุติยภูมิ (secondary data) จาก Third International Mathematics and Science Studies (TIMSS) ของฐานข้อมูลของประเทศสหรัฐอเมริกาในการศึกษา ซึ่งประกอบไปด้วยกลุ่มตัวอย่างเป็นนักเรียนจำนวน 1,130 คน จาก 68 โรงเรียน เครื่องมือที่ใช้เป็นข้อสอบแบบเลือกตอบ (multiple choice) จำนวน 17 ข้อ วิเคราะห์ข้อมูลด้วยโปรแกรม HLM โดยเตรียมข้อมูลจากโปรแกรม SPSS โปรแกรม SAS และโปรแกรม SYSTAT

Chu and Kamata (2003) ศึกษาเรื่องการปรับเทียบแบบสอบและการศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยมีจุดประสงค์เพื่อทดสอบความถูกต้องและความคงที่ของพารามิเตอร์ความยากของข้อสอบและพารามิเตอร์ความสามารถของผู้สอบโดยใช้โมเดลระดับลดหลั่นแบบราสช์ (Hierarchical Rasch Model) การประเมินโมเดลระดับลดหลั่นแบบราสช์ (Hierarchical Rasch Model) ภายใต้การใช้ข้อสอบร่วมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ใน 2 ปัจจัย คือ จำนวนข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) และขนาดของการทำหน้าที่ต่างกันของข้อสอบ (amount of DIF) ผลการศึกษาพบว่า โมเดลระดับลดหลั่นแบบราสช์ (Hierarchical Rasch Model) สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในระหว่างที่ทำการปรับเทียบแบบสอบและการประมาณค่าความสามารถของผู้สอบได้ โดยโมเดลระดับลดหลั่นแบบราสช์ (Hierarchical Rasch Model) จะขจัดความลำเอียง (bias) ซึ่งเป็นผลมาจากการเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

Beretvas and Williams (2004) ศึกษาเรื่องการใช้โมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) สำหรับประเมินมิติของข้อสอบ โดยใช้โมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) เปรียบเทียบกับโมเดลทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional Item Response Theory Model: MIRT Model) โดยการจำลองข้อสอบจำนวน 17 ข้อ ซึ่งเป็นข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) แบ่งเป็นข้อสอบที่วัดความสามารถแบบ 1 มิติจำนวน 10 ข้อ โดยวัดในเรื่องของความรู้คณิตศาสตร์ ส่วนข้อสอบอีก 7 ข้อ เป็นข้อสอบที่วัดความสามารถแบบ 2 มิติ ให้น้ำหนักกับองค์ประกอบที่ 1 และองค์ประกอบที่ 2 วัดในเรื่องความสามารถคณิตศาสตร์ (maths ability) ประกอบด้วย 2 กลุ่มเนื้อหา (content/bundle) อันได้แก่ 1) มโนทัศน์เกี่ยวกับจำนวน (numerical concepts) และ 2) ความสามารถในการแก้ปัญหา (problem solving ability) และความสามารถในการอ่าน (reading ability) และจำลองข้อมูลนักเรียนจำนวน 5,652 คน วิเคราะห์ข้อมูลด้วยโปรแกรม NOHARM และ HLM 5 ผลการศึกษาพบว่า โมเดล HGLM และโมเดลทฤษฎีการตอบสนองข้อสอบแบบหลายมิติ (Multidimensional Item Response Theory: MIRT Model) ให้ผลไม่ต่างกันในการวัดมิติของข้อสอบ แต่โมเดล HGLM เหมาะกับข้อมูลแบบ 1 พารามิเตอร์มากกว่า (1 parameter Rasch model)

Chaimongkol (2005) ศึกษาโมเดลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยการวิเคราะห์โมเดลถดถอยโลจิสติกพหุระดับ ใช้การประมาณค่าแบบเบย์ การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาโมเดลทางสถิติเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) สำหรับข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) และศึกษาตัวแปรระดับโรงเรียนที่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยประยุกต์ใช้โมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) แบบ 3 ระดับ นั่นคือ ระดับที่ 1 เป็นระดับข้อสอบ (item level) ระดับที่ 2 เป็นระดับนักเรียน (student level) และระดับที่ 3 เป็นระดับโรงเรียน (school level) จำลองข้อมูลข้อสอบจำนวน 10 ข้อ ด้วยโปรแกรม S PLUS และวิเคราะห์ข้อมูลด้วยโมเดลถดถอยโลจิสติกพหุระดับแบบ 3 ระดับด้วยโปรแกรม WinBUGS 1.4 ผลการศึกษา พบว่า ค่าความลำเอียง (bias) ลดลง เมื่อจำนวนนักเรียนต่อโรงเรียนและจำนวนโรงเรียนที่นำเข้าสู่การวิเคราะห์เพิ่มขึ้น จากนั้นประยุกต์โมเดลที่ได้ไปสู่การวิเคราะห์ด้วยข้อมูลจริงที่เก็บรวมข้อมูลไว้ในปี 2003 จาก Administration of a mathematics assessment for 4^{th} graders in a statewide testing program จำนวน 3,312 คน จาก 30 โรงเรียนในประเทศสหรัฐอเมริกา วิชาคณิตศาสตร์เรื่องเรขาคณิต จำนวน 7 ข้อ ผลการศึกษา พบว่า ข้อสอบข้อ 5 เป็นข้อที่ยากสำหรับนักเรียนหญิงมากกว่านักเรียนชาย ส่วนข้อสอบข้อที่ 1 และข้อ 3 เป็นข้อที่ง่ายสำหรับนักเรียนหญิงมากกว่านักเรียนชายและเมื่อพิจารณาการเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) ระหว่างโรงเรียน พบว่า ข้อสอบข้อ 3 เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) สูงที่สุด รองลงมาคือข้อ 1, 2 และข้อ 6

Kamata and al (2005) ศึกษาอิทธิพลสุ่มของการทำหน้าที่ต่างกันของข้อสอบ (DIF) ระหว่างกลุ่มโดยประยุกต์ใช้โมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) ในการประมาณค่าความผันแปรของการทำหน้าที่ต่างกันของข้อสอบ (DIF) ระหว่างกลุ่มได้ โดยการประยุกต์โมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) แบบ 3 ระดับมาใช้ ข้อมูลที่นำมาวิเคราะห์เป็นข้อมูลทุติยภูมิ (secondary data) จาก National Assessment of Educational Progress: NAEP ในปี 2003 กลุ่มตัวอย่างเป็นนักเรียนที่มีความสามารถทางภาษาอังกฤษในระดับจำกัด (limited English proficiency: LEP) เกรด 4 จำนวน 2,243 คน จาก 236 โรงเรียน ในเนื้อหาวิชาคณิตศาสตร์ วิเคราะห์ข้อมูลโดยตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธีแมนเทลและแฮนเซล (Mantel and Hansel) จากนั้นนำข้อที่ตรวจสอบพบว่าเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) มาวิเคราะห์ด้วยโมเดล HGLM แบบ 2 ระดับอีกครั้ง ผลการศึกษาพบว่า จากการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยวิธีแมนเทลและแฮนเซล (Mantel and Hansel) พบข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) จำนวน 8 ข้อ จากนั้นเมื่อนำข้อสอบทั้ง 8 ข้อมาวิเคราะห์ซ้ำด้วยโมเดล HGLM แบบ 2 ระดับพบว่าข้อสอบเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) เพียง 6 ข้อเท่านั้น

Chaimongkol, Huffer and Kamata (2006) ศึกษาการทำหน้าที่ต่างกันข้อสอบ (DIF) โดยมุ่งเน้นการประเมินแหล่งของการเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยโมเดลโลจิสติกแบบพหุระดับ โดยใช้โมเดล HGLM เป็นกรอบแนวคิดในการศึกษาสำหรับข้อมูลแบบ 3 ระดับ โดยการจำลองข้อมูลด้วยเงื่อนไขที่แตกต่างกัน ประมาณค่าด้วยสถิติเบส์เซียนด้วยโปรแกรม WinBUGS 1.4 โดยกำหนดการทำซ้ำ 11,000 รอบ และตัด 1,000 รอบแรกออก ผลการศึกษา พบว่า ตัวแปรคุณลักษณะของนักเรียน (indivadual characteristics) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรคุณลักษณะของโรงเรียน (school or cluster characteristics) ซึ่งเป็นตัวแปรระดับที่ 3 ส่งผลต่อการทำหน้าที่ต่างกัน (DIF) ซึ่งข้อค้นพบที่ได้นี้ จะเป็นประโยชน์สำหรับการศึกษาในอนาคต นั่นคือ คุณลักษณะที่แตกต่างกันของตัวแปรระดับที่ 2 และระดับที่ 3 จะเป็นแหล่งที่อาจจะทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

Cheong (2006) ศึกษาการวิเคราะห์อิทธิพลของบริบทโรงเรียนที่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้โมเดล HGLM การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาบริบทโรงเรียนที่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบโดยใช้โมเดล HGLM เป็นกรอบแนวคิดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และระบุขนาดอิทธิพลของตัวแปรระดับโรงเรียนที่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ (DIF) การศึกษาในครั้งนี้เป็นการศึกษาโดยใช้ข้อมูลทุติยภูมิ (secondary data) ที่เก็บรวบรวมไว้แล้วตั้งแต่ปี ค.ศ. 2001 โดย Baldi และคณะ นั่นคือ กลุ่มตัวอย่างเป็นนักเรียนจำนวน 2,076 คน จาก 92 โรงเรียน และเครื่องมือที่ใช้เป็นแบบเลือกตอบ (multiple choice) จำนวน 13 ข้อ วิเคราะห์ข้อมูลแบบพหุระดับ โดยกำหนดระดับที่ 1 เป็นระดับข้อสอบ (item level) ระดับที่ 2 เป็นระดับนักเรียน (person level) และระดับที่ 3 เป็นระดับโรงเรียน (school level) ผลการศึกษา พบว่า มีข้อสอบจำนวน 3 ข้อที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) และตัวแปรระดับโรงเรียนส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ (DIF)

Vaughn (2006) ศึกษาโมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) ของการศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบสุ่มสำหรับข้อสอบที่ตรวจให้คะแนนมากกว่า 2 ค่า (polytomous) โดยการประมาณค่าแบบเบย์พหุระดับ การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบสุ่ม (random effect DIF) โดยประยุกต์ใช้โมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) แบบ 3 ระดับ นั่นคือ ระดับที่ 1 เป็นระดับข้อสอบ (item level) ระดับที่ 2 เป็นระดับนักเรียน (student level) และระดับที่ 3 เป็นระดับโรงเรียน (school level) จำลองข้อมูลข้อสอบจำนวน 17 ข้อ ด้วยโปรแกรม R และประมาณค่าพารามิเตอร์ด้วยเทคนิค Markov chain Monte Carlo (MCMC) ด้วยโมเดลถดถอยโลจิสติกพหุระดับแบบ 3 ระดับด้วยโปรแกรม OpenBUGS ผลการเปรียบเทียบการจำลองข้อมูลเพื่อตรวจสอบโมเดลการเกิดการทำหน้าที่ต่างกันของข้อสอบแบบกำหนด 3 ระดับ (Three level fixed DIF Model) และ

โมเดลการเกิดการทำหน้าที่ต่างกันของข้อสอบแบบสุ่ม 3 ระดับ (Three level random DIF Model) พบว่า ได้ผลการศึกษาคล้ายคลึงกัน จากนั้นประยุกต์โมเดลที่ได้ไปสู่การวิเคราะห์ด้วยข้อมูลจริง จาก National Center for Education Statistics (NCES): 88 High School Effects Study (HSES) กลุ่มตัวอย่างเป็นนักเรียนเกรด 10 – 12 จำนวน 2,190 คน จาก 109 โรงเรียนในประเทศสหรัฐอเมริกา จำนวน 4 ข้อ ผลการศึกษา พบว่าข้อสอบข้อ 2 -4 เป็นข้อสอบที่การเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) ระหว่างนักเรียนหญิงและนักเรียนชาย

Williams and Beretvas (2006) ศึกษาเรื่องการระบุการทำหน้าที่ต่างกันของข้อสอบโดยการใช้โมเดลโลจิสติกทั่วไประดับลดหลั่นสำหรับข้อสอบที่ตรวจให้คะแนนมากกว่า 2 ค่า (polytomous) โดยได้เสนอแนวทางในการประยุกต์ใช้โมเดลโลจิสติกทั่วไประดับลดหลั่น (Hierarchical Generalized Linear Logistic Model: HGLLM) ในการวิเคราะห์ข้อสอบที่ตรวจให้คะแนนมากกว่า 2 ค่า (polytomous) ทั้งยังประยุกต์ใช้โมเดลโลจิสติกทั่วไประดับลดหลั่น (Hierarchical Generalized Linear Logistic Model: HGLLM) ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) อีกด้วย ผลการศึกษา พบว่า โมเดลโลจิสติกทั่วไประดับลดหลั่น (Hierarchical Generalized Linear Logistic Model: HGLLM) สามารถประยุกต์ใช้ในการวิเคราะห์ข้อสอบที่ตรวจให้คะแนนมากกว่า 2 ค่า (polytomous) ได้ โดยเปรียบเทียบกับการวิเคราะห์ด้วยโมเดลทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Model: IRT Model) ที่ Mulaki (1990) ได้ปรับปรุงขึ้น นั่นคือ Modified Graded – Response Model (M-GRM) พบว่า ได้ค่าสถิติพื้นฐานที่ใกล้เคียงกัน คือ ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (SD) ค่า Bias ค่า RMSE และค่า Relative Bias นอกจากนี้ยังพบว่า โมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) สามารถใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ตรวจให้คะแนนมากกว่า 2 ค่า (polytomous) ได้ด้วย และเมื่อเปรียบเทียบกับการวิเคราะห์ด้วยวิธีแมนเทล-แฮนเซล (Mentel-Haenszel) พบว่า โมเดลเชิงเส้นทั่วไประดับลดหลั่น (HGLM) และวิธี Mentel-Haenszel สามารถประมาณค่าสถิติได้สมมูลกัน

Binici (2007) ศึกษาอิทธิพลแบบสุ่มของการทำหน้าที่ต่างกันของข้อสอบ (Random effect DIF) ภายใต้การเปรียบเทียบโมเดล HGLM และโมเดล GLLAMM การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของโมเดล HGLM ซึ่งวิเคราะห์จากโปรแกรม HLM แบบ 3 ระดับ นั่นคือ ระดับที่ 1 เป็นระดับข้อสอบ (item level) ระดับที่ 2 เป็นระดับนักเรียน (student level) และระดับที่ 3 เป็นระดับกลุ่ม (group level) และโมเดล GLLAMM ซึ่งวิเคราะห์จากโปรแกรม STATA โดยพิจารณาจากค่าอัตราความถูกต้อง (power) และอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 (type I error rate) จำลองข้อมูลข้อสอบจำนวน 10 ข้อ ตัวแปรที่ศึกษา ประกอบไปด้วย ขนาดของกลุ่ม (cluster size) จำนวนกลุ่ม (number of cluster) และอัตราการเกิดการทำหน้าที่ต่างกันของข้อสอบสูงสุด (magnitude of random

DIF) ผลการศึกษา พบว่า 1) ทั้งขนาดของกลุ่ม (cluster size) และจำนวนกลุ่ม (number of cluster) ส่งผลต่อการประมาณค่าพารามิเตอร์ ในขณะที่อัตราการเกิดการทำหน้าที่ต่างกันของข้อสอบสูงสุด (magnitude of random DIF) ไม่ส่งผลต่อการประมาณค่าพารามิเตอร์ 2) เมื่อขนาดของกลุ่ม (cluster size) และจำนวนกลุ่ม (number of cluster) เพิ่มขึ้น ค่าเฉลี่ยของ RMSE และค่า SE จะลดลงในขณะที่การเพิ่มขึ้นของขนาดของกลุ่ม (cluster size) และจำนวนกลุ่ม (number of cluster) ไม่ส่งผลต่อค่า mean squared bias และ 3) การประมาณค่าพารามิเตอร์ในระดับที่ 3 ไม่เสถียร (stable) เท่าการประมาณค่าพารามิเตอร์ในระดับที่ 2 จากนั้นประยุกต์โมเดลที่ได้ไปสู่การวิเคราะห์ด้วยข้อมูลจริงจำนวน 10 ข้อจาก Florida Comprehensive Assessment Test (FCAT) ของนักเรียนเกรด 4 เมื่อกลุ่มอ้างอิง (reference group) เป็นนักเรียนผิวขาว (White student) และกลุ่มอ้างอิง (focal group) เป็นนักเรียนผิวสี (African - American) ผลการศึกษา พบว่า ข้อสอบทั้ง 10 ข้อเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)

Kamata and Cheong (2007) ศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบสองทางด้วยโมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น (HGLM) โดยการศึกษาในครั้งนี้มีวัตถุประสงค์ในการประมาณค่าพารามิเตอร์ของโมเดลเชิงเส้นตรงทั่วไประดับลดหลั่น (HGLM) โดยใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์ในการคัดสรรโมเดล (Model selection criteria) โมเดลที่ใช้เป็นโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) ซึ่งเป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล ผลการศึกษา พบว่า ได้ค่าประมาณของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) แตกต่างจากค่าจริง

อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ศึกษาการวิเคราะห์ข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ: กาวิเคราะห์พหุระดับ มีวัตถุประสงค์ 3 ข้อ คือ 1) เพื่อวิเคราะห์พารามิเตอร์ข้อสอบ พารามิเตอร์ความสามารถของผู้สอบและผลของตัวแปรคุณลักษณะของผู้เรียนและตัวแปรคุณลักษณะของโรงเรียนต่อโอกาสในการตอบข้อสอบได้ถูกต้องโดยใช้วิธีการวิเคราะห์พหุระดับด้วยโปรแกรมโมเดลเชิงเส้นตรงระดับลดหลั่น 2) เพื่อเปรียบเทียบผลการประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ผู้สอบระหว่างการวิเคราะห์ข้อสอบแบบพหุระดับโดยการประยุกต์ใช้โปรแกรมโมเดลเชิงเส้นตรงระดับลดหลั่น และการวิเคราะห์ข้อสอบด้วยโปรแกรม BILOG-MG และ 3) เพื่อตรวจสอบและเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยโปรแกรม HLM กับโปรแกรม BILOG-MG กลุ่มตัวอย่างเป็นนักเรียนชั้น ม.3 จำนวน 1,588 คน ผู้บริหาร 32 คน จากโรงเรียน 32 แห่ง เครื่องมือที่ใช้ได้แก่ 1) แบบวัดความรู้ความวิชาคณิตศาสตร์ 2) แบบวัดเจตคติต่อวิชาคณิตศาสตร์ 3) แบบประเมินความเครียดของนักเรียน 4) แบบบันทึกข้อมูลนักเรียน 5) แบบประเมินภาวะผู้นำทางวิชาการของผู้บริหาร

และ 6) แบบบันทึกตัวแปรคุณลักษณะครูและโรงเรียน วิเคราะห์ข้อมูล 3 ขั้นตอน 1) ประมาณค่าพารามิเตอร์ข้อสอบและผู้สอบด้วยโมเดล HGLM-2L และ HGLM-3L และเปรียบเทียบผลการประมาณค่าพารามิเตอร์ข้อสอบและผู้สอบจากโปรแกรม HLM กับโปรแกรม BILOG-MG 2) ศึกษาผลของตัวแปรนักเรียนและโรงเรียนต่อโอกาสการตอบข้อสอบได้ถูก 3) เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ผลการศึกษาพบว่า 1) ผลการประมาณค่าพารามิเตอร์ข้อสอบด้วยโมเดล HGLM-2L และ HGLM-3L ด้วยสถิติ Empirical Bayesian มีความสัมพันธ์อย่างสัมบูรณ์กับผลการประมาณค่าด้วยโปรแกรม BILOG-MG ส่วนผลการประมาณค่าพารามิเตอร์ผู้สอบ HGLM-2L มีความสัมพันธ์อย่างสัมบูรณ์กับผลการประมาณค่าด้วยโปรแกรม BILOG-MG ส่วน HGLM-3L มีสัมประสิทธิ์ความสัมพันธ์ 0.793 2) ผลการวิเคราะห์ระดับนักเรียน พบว่าตัวแปรผลการเรียนวิชาคณิตศาสตร์ในภาคเรียนที่ผ่านมาส่งผลต่อค่าเฉลี่ยขอโอกาสในการตอบข้อสอบได้ถูกต้องในแต่ละโรงเรียนอย่างมีนัยสำคัญที่ระดับ 0.01 ส่วนผลการวิเคราะห์ระดับโรงเรียนว่าตัวแปรความเป็นผู้นำทางวิชาการของผู้บริหารส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบถูกในทุกโรงเรียนอย่างมีนัยสำคัญที่ระดับ 0.05 และ 3) การตรวจสอบ DIF ด้วยโมเดล HGLM สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันสอดคล้องตรงกับ BILOG-MG

จากเอกสารและงานวิจัยที่เกี่ยวข้องกับงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM ข้างต้น สรุปได้ดังตารางสังเคราะห์ที่ 2.6

ตาราง 2.6 งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM

ชื่อผู้วิจัย (ปี)	วัตถุประสงค์	ผลการศึกษา
Kamata (1998)	ศึกษาโมเดล HGLM แบบ 1 พารามิเตอร์ โดยประยุกต์โมเดล HGLM สู่ทฤษฎีการตอบสนองข้อสอบ (IRT)	โมเดลราสช์ (Rasch model) เป็นกรณีหนึ่งของโมเดล IRT
Kamata (2001)	ศึกษาการวิเคราะห์ข้อสอบด้วยโมเดล HGLM โดยขยายสู่การวิเคราะห์แบบ 3 ระดับ	โมเดล HGLM ให้ผลการวิเคราะห์ที่ทัดเทียมกับการวิเคราะห์ด้วยโมเดลราสช์ (Rasch model) และสามารถประยุกต์โมเดล HGLM สู่การวิเคราะห์ในระดับที่ 3 ได้
Kamata (2002)	เปรียบเทียบผลการวิเคราะห์พหุระดับแบบ 2 ระดับและ 3 ระดับ	การวิเคราะห์พหุระดับด้วยโมเดลราสช์ (Rasch model) แบบ 2 ระดับมีความทัดเทียมกันกับการวิเคราะห์โมเดล HGLM แบบ 2 ระดับและการวิเคราะห์พหุระดับด้วยโมเดลราสช์ (Rasch model) แบบ 3 ระดับก็มีความทัดเทียมกับการวิเคราะห์โมเดล HGLM แบบ 3 ระดับ
Chu and Kamata (2003)	ทดสอบความถูกต้องและความคงที่ของพารามิเตอร์ความยากของข้อสอบและพารามิเตอร์ความสามารถของผู้สอบ	โมเดลระดับลดหลั่นแบบราสช์ (Hierarchical Rasch Model) ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ในระหว่างที่ทำการปรับเทียบแบบสอบและการประมาณค่า
Beretvas and Williams (2004)	ประเมินมิติของข้อสอบโดยใช้โมเดล HGLM เปรียบเทียบกับโมเดล MIRT	โมเดล HGLM และโมเดล MIRT ให้ผลไม่ต่างกันในการวัดมิติของข้อสอบแต่โมเดล HGLM เหมาะกับข้อมูลแบบ 1 พารามิเตอร์มากกว่า
Chaimongkol (2005)	พัฒนาโมเดลทางสถิติเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และศึกษาตัวแปรระดับโรงเรียนที่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้โมเดล HGLM	ค่าความลำเอียง (bias) ลดลงเมื่อจำนวนนักเรียนต่อโรงเรียนและจำนวนโรงเรียนเพิ่มขึ้น เมื่อวิเคราะห์ด้วยข้อมูลจริง พบว่า ข้อสอบเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) ระหว่างนักเรียนชายและหญิง

ตาราง 2.6 งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM (ต่อ)

ชื่อผู้วิจัย (ปี)	วัตถุประสงค์	ผลการศึกษา
Kamata and al (2005)	ตรวจสอบการเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบสุ่ม โดยใช้โมเดล HGLM	จากการตรวจสอบด้วยวิธี MH พบข้อสอบที่เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) 8 ข้อ จากนั้นเมื่อนำข้อสอบทั้ง 8 ข้อมาวิเคราะห์ซ้ำด้วยโมเดล HGLM พบว่าข้อสอบเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) เพียง 6 ข้อ
Chaimongkol and al (2006)	ศึกษาการเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบ 3 ระดับด้วยโมเดลโลจิสติกแบบพหุระดับ	ตัวแปรคุณลักษณะของนักเรียนและตัวแปรคุณลักษณะของโรงเรียน ส่งผลต่อการทำหน้าที่ต่างกัน (DIF) นั่นคือ คุณลักษณะที่แตกต่างกันของตัวแปรระดับที่ 2 และระดับที่ 3 จะเป็นแหล่งที่อาจจะทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)
Cheong (2006)	ศึกษาบริบทโรงเรียนที่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบโดยใช้โมเดล HGLM	ข้อสอบ 3 ข้อเกิดการทำหน้าที่ต่างกันของข้อสอบและตัวแปรระดับโรงเรียนส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ (DIF)
Vaughn (2006)	ศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบสุ่ม โดยใช้ โมเดล HGLM	โมเดลการทำหน้าที่ต่างกันของข้อสอบแบบกำหนด 3 ระดับและโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบสุ่ม 3 ระดับ ให้ผลการศึกษาคล้ายคลึงกัน เมื่อวิเคราะห์ด้วยข้อมูลจริง พบว่าข้อสอบเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF) ระหว่างนักเรียนหญิงและนักเรียนชาย
Williams and Beretvas (2006)	ศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) โดยใช้ โมเดล HGLM	โมเดล HGLM ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้และเมื่อเปรียบเทียบกับผลการวิเคราะห์ด้วยวิธี MH พบว่าประมาณค่าสถิติได้สมมูลกัน

ตาราง 2.6 งานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM (ต่อ)

ชื่อผู้วิจัย (ปี)	วัตถุประสงค์	ผลการศึกษา
Binici (2007)	เปรียบเทียบประสิทธิภาพของโมเดล HGLM	ขนาดของกลุ่ม (cluster size) และโมเดล GLLAMM และจำนวนกลุ่ม (number of cluster) ส่งผลต่อ ในขณะที่อัตราการเกิดการทำหน้าที่ต่างกันของข้อสอบสูงสุด (magnitude of random DIF) ไม่ส่งผลต่อการประมาณค่าพารามิเตอร์ เมื่อวิเคราะห์ด้วยข้อมูลจริง พบว่า ข้อสอบทั้ง 10 ข้อเกิดการทำหน้าที่ต่างกันของข้อสอบ (DIF)
Kamata and Cheong (2007)	ประมาณค่าพารามิเตอร์ของโมเดล HGLM	ได้ค่าประมาณของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) แตกต่างจากค่าจริง
อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551)	เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกัน (DIF) ด้วยโปรแกรม BILOG-MG และโปรแกรม HLM	ผลที่ได้มีความสอดคล้องกันอย่างสัมบูรณ์ การตรวจสอบการทำหน้าที่ต่างกัน (DIF) จากทั้งสองโปรแกรมได้ผลที่สอดคล้องกัน

จากตารางที่ 2.6 สรุปได้ดังนี้ การศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ล้วนให้กรอบแนวคิดของโมเดล Kamata (Kamata's Model) หรือโมเดล HGLM เป็นกรอบในการศึกษาทั้งสิ้น โดยโมเดล HGLM นี้สามารถประยุกต์ใช้ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ (DIF) ได้ทั้งในการวิเคราะห์แบบ 2 ระดับและแบบ 3 ระดับ ซึ่งการศึกษาส่วนใหญ่จะเป็นการศึกษาโดยการจำลองข้อมูลขึ้นมาตามเงื่อนไขที่ต้องการศึกษา จากนั้นจะประยุกต์ผลที่ได้จากการศึกษาในเชิงทฤษฎีหรือการศึกษาจากการจำลองข้อมูลสู่การศึกษาด้วยข้อมูลจริง ซึ่งพบว่า โมเดล HGLM ให้ผลการประมาณค่าพารามิเตอร์ที่สอดคล้องกันทั้งข้อมูลที่จำลองขึ้นและข้อมูลเชิงประจักษ์

หากใช้เกณฑ์การวิเคราะห์ข้อมูลในการแบ่งการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) จะสามารถแบ่งออกเป็น 2 กลุ่มสำคัญ คือ 1) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยการวิเคราะห์แบบ 1 ระดับ (single level) หรือการทำหน้าที่ต่างกันของข้อสอบแบบ 1 ระดับ (single level DIF) และ 2) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยการวิเคราะห์แบบพหุระดับ (Multilevel) หรือการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ซึ่งการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) ด้วยการวิเคราะห์แบบ 1 ระดับ (single level) นั้น เป็นการศึกษาที่ได้รับ

ความนิยมอย่างแพร่หลายในอดีต ซึ่งจากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า ประกอบไปด้วยการศึกษาของ Swaminathan and Rogers (1990); Rogers and Swaminathan (1993); Mazor and al (1994); Uttaro and Millsap (1994); French and Miller (1996); Narayanan and Swaminathan (1996); Roussos and Stout (1996); Roussos and Stout (1996b); Oshima, Raju and Flowers (1997); Oshima, Raju, Flowers and Slinde (1998); Kim (2000); Penfield (2001); Walker and Beretvas (2001); Bolt (2002); Gierl, Bisanz, Bisanz and Boughton (2003); Cohen and Bolt (2005); Finch (2005); Lei and al (2006); Park (2006); Penfield (2006); Cohen and al (2007); Elosua and Jaúregui (2007); Walker Zhang and Surber (2008); กาญจนา วัธนสุนธร (2537); เกสร หว่างจิตร (2539); จิติมา วรรณศรี (2539); เสรี ชัดแช้ม (2539); ญาณภัทร สีหะมงคล (2540); พรรณี จิตมาศ (2540); รัชรินทร์ มุคดา (2540); นพมาศ พิพัฒนสุข (2541); นิคม กีรติวากูร (2542); วลีมาศ แซ่อึ้ง (2543); สิริรัตน์ วิภาสศิลป์ (2545); สุมาลี แกล้วทนงค์ (2547); ปิยะทิพย์ ตินวร (2549) และ อรินทร์ น่วมถนอม (2549) ส่วนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ประกอบไปด้วยการศึกษาของ Kamata (1998); Kamata (2001); Kamata (2002); Chu and Kamata (2003); Beretvas and Williams (2004); Chaimongkol (2005); Kamata and al (2005); Vaughn (2006); Williams and Beretvas (2006); Cheong (2006) และ Binici (2007) ซึ่งการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) จะให้สารสนเทศที่เหมาะสมกว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) แบบ 1 ระดับ (single level) ในกรณีที่ข้อมูลเป็นระดับชั้น (hierarchical data) (Bergh, Kuhlemeier and Wijnstra, 1995; Binici, 2007; Linjun, 1999; Luppescu, 2002)

ดังนั้นจึงอาจกล่าวได้ว่า เพื่อให้การศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) สอดคล้องกับสภาพจริงทางการศึกษามากยิ่งขึ้น การวิเคราะห์แบบพหุระดับ (multilevel) จึงเข้ามามีบทบาทแทนการวิเคราะห์แบบ 1 ระดับ (single level) ทั้งนี้ก็เนื่องมาจากตัวแปรทางการศึกษามีลักษณะเป็นระดับชั้น (hierarchical data) ซึ่งการศึกษาการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ผ่านมานั้น นักวิจัยส่วนหนึ่งละเลยต่อลักษณะของข้อมูลที่เป็นระดับชั้น (hierarchical data) โดยการจัดกระทำตัวแปรต่างระดับกันให้อยู่ในระดับเดียวกันแล้ววิเคราะห์เสมือนหนึ่งว่าข้อมูลเหล่านั้นอยู่ในระดับเดียวกัน ซึ่งผลการวิจัยที่ได้อาจจะบิดเบือนจากความเป็นจริง (Kamata, 2009; ศิริชัย กาญจนวาสี, 2550) ดังนั้นการวิเคราะห์ข้อมูลที่เป็นระดับชั้น (hierarchical data) ด้วยการวิเคราะห์แบบพหุระดับ (multilevel) จึงได้รับความนิยมในการศึกษากันอย่างแพร่หลายและกว้างขวาง ส่วนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) นั้นเป็นประเด็นทางการศึกษาที่ได้รับความนิยมในการศึกษามาระยะเวลาหนึ่งแล้ว

ซึ่งจากการสังเคราะห์เอกสารและงานวิจัยที่เกี่ยวข้องอย่างครอบคลุม พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับ (Multilevel DIF) ด้วยโมเดล HGLM นั้นเป็นประเด็นที่ศึกษาค้นคว้าโดย Chu and Kamata (2003), Chaimongkol (2005), Kamata and al (2005), Chaimongkol and al (2006), Cheong (2006), Vaughn (2006), Williams and Beretvas (2006), Binici (2007), Kamata and Cheong (2007) และอิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551)

ซึ่งเมื่อพิจารณาในส่วนของโมเดลการทำหน้าที่ต่างกันที่นักวิจัยเหล่านี้ใช้ พบว่า จากการศึกษาค้นคว้าที่ผ่านมา นักวิจัยเหล่านี้จะกำหนดให้ค่าพารามิเตอร์บางตัวในโมเดล HGLM เป็นศูนย์ ตัวอย่างเช่น การศึกษาของ Kamata and Cheong (2007) ที่ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับด้วยโมเดล HGLM แต่การศึกษานี้กำหนดให้พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์ นั่นคือ เป็นการศึกษาที่ไม่รวมเอาพารามิเตอร์นี้เข้าสู่โมเดลของการศึกษา ซึ่งสอดคล้องกับการศึกษาของ Chaimongkol, Huffer and Kamata (2006) และ Vaughn (2006) ซึ่ง Kamata (2009) กล่าวว่าเหตุผลสำคัญที่นักวิจัยมักกำหนดให้ค่าพารามิเตอร์บางตัวเป็นศูนย์เนื่องจากเพื่อลดความยุ่งยากในการแปลผล กอปรกับการที่ยังไม่มีการศึกษาค้นคว้ามาก่อนว่าพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) มีอิทธิพลต่อการผันแปรของการประมาณค่าการทำหน้าที่ต่างกันของข้อสอบหรือไม่เพียงใด รวมไปถึงยังไม่มีหลักฐานว่าควรจะกำหนดขนาดของพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นเท่าใดเพื่อให้ผลการประมาณค่าการทำหน้าที่ต่างกันของข้อสอบมีความน่าเชื่อถือมากที่สุด นอกจากนี้นักวิจัยบางคนจะใช้เกณฑ์การคัดเลือกโมเดล (model selection criteria) เป็นเกณฑ์ในการตัดสินว่าจะใช้โมเดลไหนในการวิเคราะห์ นั่นคือ นักวิจัยหรือนักสถิติส่วนใหญ่จะใช้ค่าสถิติเช่น ดัชนี AIC(Akaike's information criterion), ดัชนี BIC (Bayesian information criterion), หรือดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) ดัชนีใดดัชนีหนึ่งในการตัดสินใจว่าการศึกษานั้นๆ จะมีค่าพารามิเตอร์ใดอยู่ในโมเดลที่ศึกษาบ้าง หรือว่าโมเดลไหนเป็นโมเดลที่ดีที่สุด (best model) ที่ควรใช้ในการศึกษานั้นๆ

จากการศึกษาของ Patarapichayatham, Kanjanawasee and Kamata (2009) และจากการศึกษานำร่อง พบว่า โมเดลเต็มรูป (complete model) เป็นเพียงโมเดลเดียวเท่านั้นที่ได้ค่าพารามิเตอร์ที่ประมาณได้ใกล้เคียงกับค่าจริง (true value) ส่วนโมเดลไม่เต็มรูป (incomplete model) ทั้ง 3 โมเดลนั้นค่าพารามิเตอร์ที่ประมาณได้แตกต่างจากค่าจริง (true value) ซึ่งสามารถสรุปได้ว่าเกิดการทำหน้าที่ต่างกัน (DIF) ขึ้นในโมเดลไม่เต็มรูป (incomplete model) ทั้ง 3 โมเดล นั่นคือ การกำหนดให้ค่าพารามิเตอร์บางตัวเป็นศูนย์ทำให้เกิดการทำหน้าที่ต่างกัน (DIF) ขึ้น ทั้งนี้ก็เพราะว่า การละเลยต่อ

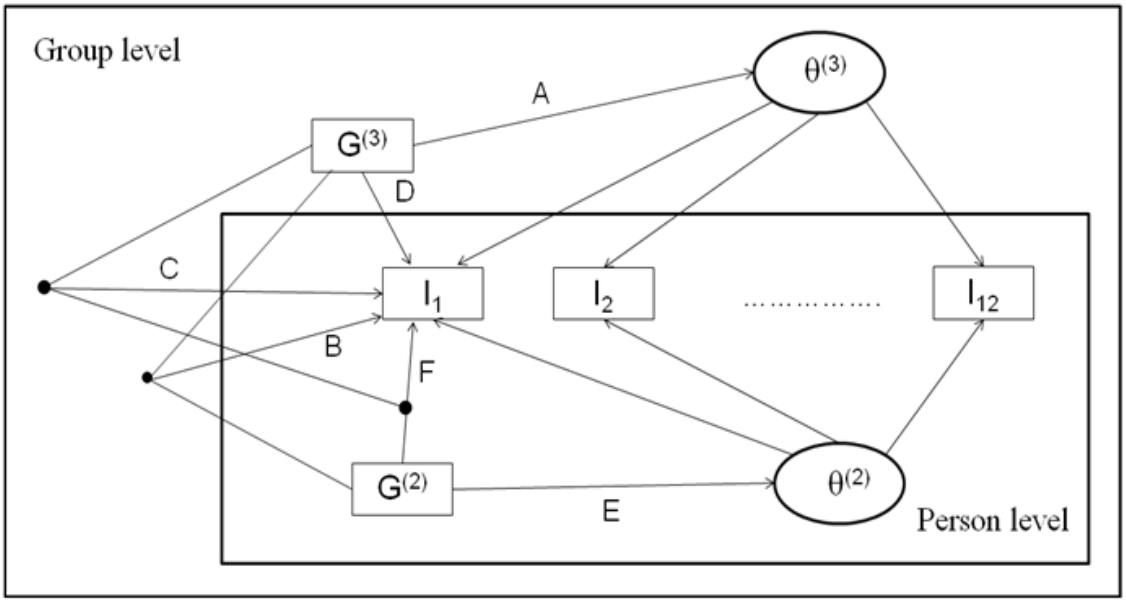
ขนาดของการทำหน้าที่ต่างกันของพารามิเตอร์เหล่านี้ส่งผลต่อการประมาณค่าพารามิเตอร์ของโมเดล HGLM

ดังนั้นการศึกษาในครั้งนี้จึงเป็นการศึกษาเพื่อเติมเต็มส่วนที่ขาดหายไปของโมเดล HGLM นั่นคือ เป็นการรวมพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (GLDIF) เข้าสู่โมเดลการวิเคราะห์ รวมไปถึงศึกษาขนาดอิทธิพลของจำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school or the number of individual in each cluster) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ดังนั้นผู้วิจัยมีวัตถุประสงค์ในการศึกษาในครั้งนี้ 2 ประการ คือ 1) เพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มและ 2) เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)

การศึกษาในครั้งนี้จึงเป็นการศึกษาโดยการจำลองข้อมูลด้วยเงื่อนไขที่แตกต่างกันจำนวน 18 เงื่อนไข (2 x 3 x 3) อันประกอบไปด้วย 1) จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school or the number of individual in each cluster) 2 ระดับ คือ โรงเรียนละ 50 คนและโรงเรียนละ 100 คน 2) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 3 ระดับ คือ 0.2, 0.4 และ 0.6 และ 3) ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ คือ 0.1, 0.2 และ 0.3 ส่วนจำนวนโรงเรียนนั้นกำหนดคงที่ทุกเงื่อนไขที่ 50 โรงเรียน โดยในแต่ละเงื่อนไขกำหนดการทำซ้ำเงื่อนไขละ 200 รอบ ผู้วิจัยมีโมเดลการตรวจสอบการทำหน้าที่ต่างกัน 4 โมเดลประกอบไปด้วย 1) โมเดลเต็มรูป (complete model) 2) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) 3) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF)และ 4) โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF)

การศึกษาในครั้งนี้มีพารามิเตอร์เป้าหมายในการประมาณค่า 6 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 2) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) 3) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 6) พารามิเตอร์ความสามารถระดับกลุ่ม (CA)

ผู้วิจัยมีเกณฑ์การคัดสรรโมเดล (Model selection criteria) 5 เกณฑ์ ได้แก่ 1) ดัชนี AIC (Akaike's information criterion, 2) ดัชนี BIC (Bayesian information criterion), 3) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), 4) ดัชนี 2 of 3 และ 5) ดัชนี p-value และมีเกณฑ์การประเมิน (evaluation criteria) ในการประมาณค่าพารามิเตอร์ 3 เกณฑ์ คือ 1) ค่าความลำเอียง (BIAS) 2) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และ 3) ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) โดยการศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลด้วยโปรแกรม R และประมาณค่าพารามิเตอร์ด้วยวิธี Maximum likelihood ด้วยโปรแกรม Mplus ภายใต้การเขียนคำสั่ง การนำข้อมูลไปวิเคราะห์และการแสดงผลในโปรแกรม R ดังนั้นจึงได้กรอบแนวคิดในการวิจัย ดังนี้



แผนภาพที่ 2.4 กรอบแนวคิดในการวิจัย

บทที่ 3

วิธีดำเนินการวิจัย

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มและเพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) สำหรับข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า จากการจำลองข้อมูลแบบ 3 ระดับ ในบทที่ 3 นี้ แบ่งออกเป็น 2 ส่วนสำคัญ คือ 3.1) การศึกษานำร่องและผลการศึกษานำร่องในส่วนของการทดสอบโมเดลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ประกอบไปด้วย 3.1.1) วัตถุประสงค์ของการศึกษา 3.1.2) โมเดลที่ใช้ในการศึกษา 3.1.3) กรอบแนวคิดในการศึกษา 3.1.4) การจำลองข้อมูลและการวิเคราะห์ข้อมูลและ 3.1.5) ผลการศึกษาและ 3.2) วิธีดำเนินการวิจัย ประกอบไปด้วย 3.2.1) เงื่อนไขในการจำลองข้อมูล 3.2.2) โมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) 3.2.3) เกณฑ์การคัดสรรโมเดล (Model selection criteria) 3.2.4) พารามิเตอร์เป้าหมายในการประมาณค่า 3.2.5) ขั้นตอนการจำลองข้อมูลและเกณฑ์การประเมินและ 3.2.6) การออกแบบการวิจัย ดังรายละเอียดต่อไปนี้

3.1) การศึกษานำร่องและผลการศึกษานำร่อง

3.1.1) วัตถุประสงค์ของการศึกษา

การวิจัยครั้งนี้ ผู้วิจัยมีวัตถุประสงค์ข้อแรกเพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม แต่เนื่องจากผลการศึกษาที่ผ่านมา ยังไม่มีการศึกษาค้นคว้าในประเด็นนี้แต่อย่างใด ดังที่ผู้วิจัยได้กล่าวไว้ในบทที่ 1 และบทที่ 2 ดังนั้น ผู้วิจัยจึงต้องศึกษานำร่องเพื่อทดสอบความเป็นไปได้และประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่ผู้วิจัยประยุกต์กรอบแนวคิดของโมเดล HGLM เป็นโมเดลในการศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) ในการศึกษาในครั้งนี้ ดังนั้นในการศึกษานำร่องนี้ ผู้วิจัยจึงมีวัตถุประสงค์เพื่อตรวจสอบคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม เมื่อระบุคุณลักษณะเฉพาะบางประการของโมเดลการทำหน้าที่ต่างกันของข้อสอบที่แตกต่างกัน 4 คุณลักษณะ

การศึกษาในครั้งนี้มีพารามิเตอร์เป้าหมายในการประมาณค่า 8 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 2) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) 3) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 6) พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 7) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 8) พารามิเตอร์ความสามารถระดับกลุ่ม (CA)

3.1.2) โมเดลที่ใช้ในการศึกษา

โมเดลในการศึกษาในครั้งนี้ ผู้วิจัยประยุกต์ใช้กรอบแนวคิดของโมเดล HGLM แบบ 3 ระดับ เมื่อระดับที่ 1 เป็นการวิเคราะห์ระดับข้อสอบ (item level) ระดับที่ 2 เป็นการวิเคราะห์ระดับบุคคล (individual level or person level or student level) และระดับที่ 3 เป็นการวิเคราะห์ระดับกลุ่มหรือระดับโรงเรียน (cluster level or group level or school level) มีตัวแบบหรือสมการดังต่อไปนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ

(item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 โดยในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

จากตัวแบบของสมการ HGLM ด้านซ้ายมือของสมการคือการ take log ของอัตราส่วนของความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องกับความน่าจะเป็นในการตอบข้อสอบข้อที่ i ไม่ถูกต้อง ซึ่งเป็นวิธีการ transformation ในส่วนของ logistic regression function ซึ่งเป็น nonlinear function และเป็นฐานคิดของโมเดล HGLM ให้เป็น linear function ดังด้านขวาของสมการ ซึ่งขั้นตอนนี้เรียกว่าเป็นแนวคิดของ link function ดังนั้น ผลที่ได้จาก link function คือสมการเชิงเส้น (linear function) จึงอาจกล่าวอีกนัยหนึ่งว่าโมเดล HGLM คือการแปลงโมเดลไม่เชิงเส้น (nonlinear modeling) ให้เป็นโมเดลเชิงเส้น (linear modeling) นั่นเอง

ผู้วิจัยมีโมเดลการทำหน้าที่ต่างกันของข้อสอบที่ต้องการศึกษาในครั้งนี้ 4 โมเดล โดยแต่ละโมเดลมีการระบุคุณลักษณะเฉพาะบางประการของโมเดลที่แตกต่างกัน รายละเอียดของแต่ละโมเดล มีดังต่อไปนี้

โมเดลที่ 1 เป็นโมเดลเต็มรูป (complete model) เป็นโมเดลที่ประกอบด้วยพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 2) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 3) ค่าความยากของข้อสอบข้อที่ *i* (item difficulty) 4) พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 5) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 6) พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 7) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ คือ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม (CA) มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1

X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA) ดังนั้นจึงอาจกล่าวได้ว่า โมเดลเต็มรูป (complete model) นี้เป็นโมเดลที่มีอิทธิพลครบทั้ง 9 อิทธิพลของโมเดล HGLM

โมเดลที่ 2 เป็นโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) เป็นโมเดลที่ประกอบด้วยพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 อิทธิพล โดยกำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) เป็นศูนย์ พารามิเตอร์ทั้ง 6 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 2) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 3) ค่าความยากของข้อสอบข้อที่ *i* (item difficulty) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 6) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ คือ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม (CA) มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA) ผู้วิจัยกำหนดค่าอิทธิพลของ γ_4 หรือพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ให้เป็นศูนย์ เพื่อต้องการทดสอบว่าการ drop อิทธิพลของพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) นั้นส่งผลต่อการประมาณค่าพารามิเตอร์เป้าหมายทั้ง 6 พารามิเตอร์หรือไม่

โมเดลที่ 3 เป็นโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) เป็นโมเดลที่ประกอบด้วยพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 พารามิเตอร์ โดยกำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์ พารามิเตอร์ทั้ง 6 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 2) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 3) ค่าความยากของข้อสอบข้อที่ *i* (item difficulty) 4) พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 5) พารามิเตอร์

การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 6) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ คือ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม (CA) มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA) ผู้วิจัยกำหนดค่าอิทธิพลของ γ_{6i} หรือพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) ให้เป็นศูนย์ เพื่อต้องการทดสอบว่าการ drop อิทธิพลของพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) นั้นส่งผลต่อการประมาณค่าพารามิเตอร์เป้าหมายทั้ง 6 พารามิเตอร์หรือไม่

โมเดลสุดท้ายเป็นโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ประกอบด้วยพารามิเตอร์อิทธิพลกำหนด (fixed effect) 5 พารามิเตอร์ นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์ พารามิเตอร์ทั้ง 5 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 2) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 3) ค่าความยากของข้อสอบข้อที่ *i* (item difficulty) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ คือ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม (CA) มีสมการ ดังนี้

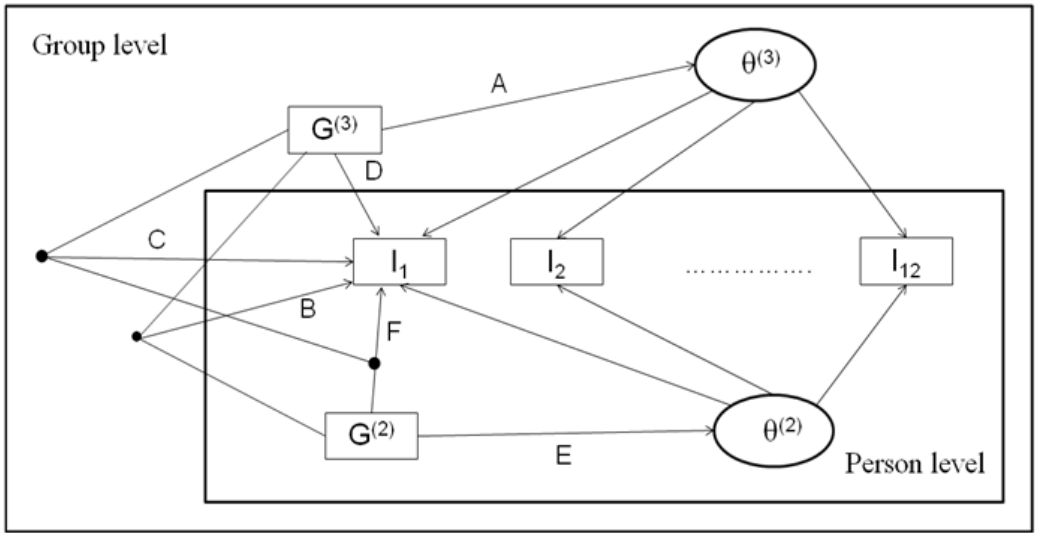
$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_{5i} X_j I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 5 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA) ผู้วิจัยกำหนดค่า

อิทธิพลของ γ_4 หรือพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ให้เป็นศูนย์และ γ_{6i} หรือพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) ให้เป็นศูนย์ เพื่อต้องการทดสอบว่าการ drop อิทธิพลของพารามิเตอร์ทั้งสองนั้น ส่งผลต่อการประมาณค่าพารามิเตอร์เป้าหมายทั้ง 6 พารามิเตอร์หรือไม่

3.1.3) กรอบแนวคิดในการศึกษา

ผู้วิจัยมีกรอบแนวคิดในการวิจัยในครั้งนี้ ดังนี้



แผนภาพที่ 3.1 กรอบแนวคิดในการวิจัย

จากแผนภาพจะสรุปได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) จำนวน 7 พารามิเตอร์ คือ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (A / CLDIF) 2) พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (B / 2WAYINT) 3) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (C / 3WAYINT) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (D / CLDIF) 5) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (E / ILADIFF) 6) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (F / ILDIF) และ 7) พารามิเตอร์ค่าความยากของข้อสอบ (item difficulty) นอกจากนี้ ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม ($\theta^{(2)}$ / IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม ($\theta^{(3)}$ / CA)

3.1.4) การจำลองข้อมูลและการวิเคราะห์ข้อมูล

การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลแบบ 3 ระดับ ผู้วิจัยขอนำเสนอวิธีการจำลองข้อมูลและการวิเคราะห์ข้อมูล ดังนี้

ผู้วิจัยเขียนคำสั่งและเงื่อนไขต่างๆ สำหรับจำลองข้อมูลด้วยโปรแกรม R โดยเขียนคำสั่งให้การทำงานทั้งหมดเสร็จสิ้นในขั้นตอนเดียว เพื่อป้องกันความผิดพลาดจากการทำงานที่ซ้ำซ้อนหลายขั้นตอนและเป็นการประหยัดเวลาในการศึกษา นั่นคือ ผู้วิจัยเขียนคำสั่งในโปรแกรม R เพื่อจำลองข้อมูลระดับโรงเรียนหรือระดับกลุ่ม (cluster level) จำนวน 30 โรงเรียนที่ได้มาอย่างสุ่ม โดย 15 โรงเรียนแรกถูกกำหนดให้เป็นกลุ่มอ้างอิง (reference group) ในพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF factor) ส่วนอีก 15 โรงเรียนที่เหลือถูกกำหนดให้เป็นกลุ่มเปรียบเทียบ (focal group) ในพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF factor) ในแต่ละโรงเรียนมีจำนวนนักเรียนโรงเรียนละ 100 คน ที่ได้มาอย่างสุ่มเช่นกัน โดยนักเรียน 50 คนแรกถูกกำหนดให้เป็นกลุ่มอ้างอิง (reference group) ในพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF factor) ส่วนนักเรียนอีก 50 คนที่เหลือถูกกำหนดให้เป็นกลุ่มเปรียบเทียบ (focal group) ในพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF factor)

ดังนั้นการศึกษานำร่องนี้มีกลุ่มตัวอย่างจำนวน 3,000 คน ซึ่ง Afshartous and Leeuw (2005), Binici (2007), Cohen (2005), Maas and Hox (2004, 2005) และ ศิริชัย กาญจนวาสี (2548) กล่าวสอดคล้องกันว่าในการวิเคราะห์ข้อมูลแบบพหุระดับที่ข้อมูลมีลักษณะแทรกสอดเป็นระดับขั้นนั้น ไม่มีการกำหนดขนาดของกลุ่มตัวอย่างที่เหมาะสมไว้แต่ควรให้ความสนใจต่อขนาดของกลุ่มตัวอย่างในระดับกลุ่ม หรือการวิเคราะห์ในระดับที่ 3 มากกว่าการวิเคราะห์ในระดับที่ 1 หรือ 2 เพราะเป็นการลดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์และเพิ่มความแม่นยำในการประมาณค่าพารามิเตอร์ ส่วนตัวแปรที่นิยมศึกษาในการวิเคราะห์พหุระดับนั้นเป็นตัวแปรที่น่าจะส่งผลต่อความถูกต้องในการประมาณค่าพารามิเตอร์ ตัวแปรการวิเคราะห์ในระดับที่ 2 ประกอบไปด้วย ตัวแปรจำนวนนักเรียนต่อห้อง (student per class) ตัวแปรจำนวนนักเรียนต่อกลุ่ม (student per group) ตัวแปรจำนวนห้องเรียนต่อโรงเรียน (classroom per school) และตัวแปรเพศ (gender) ส่วนตัวแปรการวิเคราะห์ในระดับที่ 3 ประกอบไปด้วย ตัวแปรจำนวนห้องเรียนต่อโรงเรียน (classroom per school) ตัวแปรจำนวนกลุ่ม (number of group) และตัวแปรจำนวนชั้นเรียนต่อโรงเรียน (school level per school) เป็นต้น โดยตัวแปรเหล่านี้ล้วนเป็นตัวแปรพื้นฐานที่สำคัญในการวิเคราะห์พหุระดับแต่ยังไม่มีข้อสรุปอย่างแน่ชัดว่าควรมีจำนวนเท่าไหร่จึงจะทำให้การประมาณค่าพารามิเตอร์มีความเชื่อถือได้มากที่สุด Kamata (2005, 2009)

นอกจากนี้จากการศึกษาของ Binici (2007) พบว่า ขนาดของกลุ่มตัวอย่างต่ำที่สุดสำหรับการวิเคราะห์ในระดับที่ 2 นั้นไม่ควรต่ำกว่า 20 คน และขนาดของกลุ่มตัวอย่างต่ำที่สุดสำหรับการวิเคราะห์ในระดับที่ 3 ไม่ควรต่ำกว่า 30 กลุ่ม สอดคล้องกับผลการศึกษาของ Maas and Hox (2004, 2005) ที่พบว่า ขนาดของกลุ่มตัวอย่างต่ำที่สุดสำหรับการวิเคราะห์ในระดับที่ 3 ก็ไม่ควรต่ำกว่า 30 กลุ่ม ในขณะที่ผลการศึกษาของ Snjders and Bosker (1999) พบว่าขนาดของกลุ่มตัวอย่างต่ำที่สุดสำหรับการวิเคราะห์ในระดับที่ 3 ไม่ควรต่ำกว่า 10 กลุ่ม

การศึกษาในครั้งนี้มีจำนวนข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า จำนวน 12 ข้อ โดยข้อสอบแต่ละข้อมีค่าความยาก ดังตารางที่ 3.1

ตารางที่ 3.1 ค่าความยากของข้อสอบแต่ละข้อ

ข้อสอบข้อที่	ค่าความยาก (item difficulty)
1	0
2	-2.0
3	-1.5
4	-1.0
5	0.5
6	0
7	0
8	0.5
9	1.0
10	1.5
11	2.0
12	2.5

ข้อสอบข้อที่ 1 เพียงข้อเดียวเท่านั้นที่ถูกกำหนดให้เป็นข้อสอบที่เกิดการทำหน้าที่ต่างกัน (DIF) โดยผู้วิจัยกำหนดให้ความแตกต่างของค่าเฉลี่ยความสามารถของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (the mean ability difference between the reference and focal groups) เท่ากับ 0.5 นั่นคือ กลุ่มเปรียบเทียบมีค่าเฉลี่ยความสามารถต่ำกว่ากลุ่มอ้างอิง 0.5 จากนั้นผู้วิจัยกำหนดค่าจริง (true value) ของพารามิเตอร์ต่างๆ ดังที่ผู้วิจัยได้กล่าวไว้ในตอนต้นว่าเนื่องจากการศึกษาเรื่องนี้เป็น

เรื่องใหม่และยังไม่มีการศึกษาค้นคว้ามาก่อน ผู้วิจัยจึงยึดผลการศึกษานำร่องเบื้องต้นก่อนหน้านี้เป็นเกณฑ์ในการกำหนดค่าพารามิเตอร์แต่ละค่า ซึ่งจากการจำลองข้อมูลและวิเคราะห์ครั้งแล้วครั้งเล่า จนได้ค่าจริง (true value) ที่มีนัยสำคัญและเป็นค่าที่เหมาะสมที่สุดออกมา 1 ชุด ดังตารางที่ 3.2

ตารางที่ 3.2 ค่าจริง (true value) ของค่าพารามิเตอร์แต่ละค่า

ค่าพารามิเตอร์	ค่าจริง
1) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF)	0.1
2) พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)	0.2
3) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT)	0.3
4) พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF)	0.5
5) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF)	0.2
6) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF)	-0.5
7) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC)	1.0
8) พารามิเตอร์ความสามารถระดับกลุ่ม (CA)	1.0

เนื่องจากการศึกษานำร่องในครั้งนี้ ผู้วิจัยมีโมเดลในการวิเคราะห์ 4 โมเดลดังที่กล่าวไว้ข้างต้นและโปรแกรม Mplus เป็นโปรแกรมที่มีศักยภาพสูงในการประมาณค่าของข้อมูลพหุระดับ มีความถูกต้องในการประมาณค่าพารามิเตอร์สูง มีขั้นตอนการเขียนคำสั่งที่ไม่ยุ่งยากและให้ผลการวิเคราะห์ข้อมูลที่เข้าใจง่ายและครบถ้วน นอกจากนี้โปรแกรม Mplus ใช้วิธี Maximum likelihood ในการประมาณค่าพารามิเตอร์ ซึ่งเป็นหลักการประมาณค่าที่ให้ความถูกต้องและแม่นยำสูง Muthen and Muthen (2007) และที่สำคัญผู้วิจัยสามารถวิเคราะห์ข้อมูลด้วยโปรแกรม Mplus ภายใต้การเขียนคำสั่ง การนำเข้าข้อมูลจากโปรแกรม R การแสดงผลทั้ง internal file และ external file ในโปรแกรม R การควบคุมสั่งการต่างๆ จากโปรแกรม R ดังนั้นผู้วิจัยจึงใช้โปรแกรม Mplus ในการประมาณค่าพารามิเตอร์ในครั้งนี้

สำหรับในส่วนของการทำซ้ำนั้น แม้ว่าการศึกษาที่ผ่านมา ยังไม่มีการวิจัยที่ชัดเจนว่าควรจำลองข้อมูลโดยการทำซ้ำกี่ครั้ง จึงจะให้ผลที่มีความเที่ยง (reliability) สูงสุด ซึ่งจากการศึกษาของ Harwell, Hsu and Kirisci (1996) พบว่า ในการศึกษาโดยใช้โมเดลทฤษฎีการตอบสนองข้อสอบ (IRT Model) เป็นฐานนั้นควรทำซ้ำอย่างน้อย 20 รอบ Binici (2007) ศึกษาโดยการทำซ้ำ 100 รอบ Zhou, Gierl and Tan (2006) ศึกษาโดยการทำซ้ำจำนวน 100 รอบ พัชรี จันทร์เพ็ง (2550) ศึกษาโดย

การทำซ้ำจำนวน 20 รอบ อรินทร์ น่วมถนอม (2549) ได้ศึกษาโดยการทำซ้ำ 50 รอบ และ Kamata (1999) ได้เสนอผลการศึกษานำร่องว่าควรทำซ้ำ 100 รอบ แต่เพื่อให้การศึกษาได้ค่าประมาณของพารามิเตอร์ที่ได้มีความเสถียรและเชื่อถือได้มากยิ่งขึ้นมากยิ่งขึ้น ผู้วิจัยจึงกำหนดการทำซ้ำจำนวน 120 รอบ ผู้วิจัยจำลองข้อมูลด้วยโปรแกรม R ประมาณค่าพารามิเตอร์ด้วยโปรแกรม Mplus

3.1.5) ผลการศึกษา

ผู้วิจัยได้ผลการศึกษาดังตารางที่ 3.3

ตารางที่ 3.3 ค่าพารามิเตอร์ที่ได้จากการประมาณค่าด้วยโมเดลการทำหน้าทำต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม

พารามิเตอร์	ค่าจริง	โมเดลที่ 1 Mean(SD)	โมเดลที่ 2 Mean(SD)	โมเดลที่ 3 Mean(SD)	โมเดลที่ 4 Mean(SD)
CLADIFF	0.1	0.063(0.196)	0.167(0.199)	0.113(0.199)	0.184(0.198)
2WAYINT	0.2	0.209(0.003)	-	0.161(0.002)	-
3WAYINT	0.3	0.292(0.002)	0.342(0.093)	0.806(0.038)	0.830(0.034)
CLDIF	0.5	0.512(0.015)	0.477(0.012)	-	-
ILADIFF	0.2	0.205(0.065)	0.046(0.066)	0.042(0.055)	0.046(0.066)
ILDIF	-0.5	-0.509(0.015)	0.087(0.010)	0.099(0.010)	0.101(0.010)
IAWC	1.0	1.095(0.028)	1.095(0.028)	1.094(0.028)	1.094(0.027)
CA	1.0	0.969(0.018)	0.974(0.019)	0.969(0.019)	0.972(0.019)

จากตาราง สรุปได้ว่า ค่าเฉลี่ย (mean) ของพารามิเตอร์ทุกพารามิเตอร์ของโมเดลเต็มรูป (complete model) หรือโมเดลที่ 1 มีค่าใกล้เคียงค่าจริง (true value) ในขณะที่โมเดลที่ 2 หรือโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) ได้ค่าเฉลี่ย (mean) ของพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง (true value) สำหรับโมเดลที่ 3 หรือโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) ได้ค่าเฉลี่ย (mean) ของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) พารามิเตอร์ความ

แตกต่างของความสามารถระดับบุคคล (ILADIFF) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง (true value) และโมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) ได้ค่าเฉลี่ย (mean) ของพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง

สำหรับค่าส่วนเบี่ยงเบนมาตรฐาน (SD) นั้น พบว่า ทุกพารามิเตอร์ได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกันทั้ง 4 โมเดล ยกเว้นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) ที่ได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ค่อนข้างแตกต่างกันในแต่ละโมเดล (0.002, 0.093, 0.038, 0.034) ตามลำดับ ซึ่งแสดงให้เห็นว่าค่าพารามิเตอร์ที่ประมาณได้ไม่เสถียรในแต่ละโมเดล

จากผลการศึกษานำร่องนี้ สรุปได้ว่า โมเดลเต็มรูป (complete model) เป็นเพียงโมเดลเดียวเท่านั้นที่ได้ค่าพารามิเตอร์ที่ประมาณได้ใกล้เคียงกับค่าจริง (true value) ส่วนโมเดลไม่เต็มรูป (incomplete model) ทั้ง 3 โมเดลได้ค่าพารามิเตอร์ที่ประมาณได้แตกต่างจากค่าจริง (true value) นั่นคือ สามารถสรุปได้ว่า การระบุคุณลักษณะเฉพาะบางประการของโมเดลที่แตกต่างกัน ส่งผลต่อคุณภาพของการประมาณค่าพารามิเตอร์ นั่นคือ เกิดการทำหน้าที่ต่างกันขึ้นในโมเดลไม่เต็มรูป (incomplete model) ทั้ง 3 โมเดล นั่นเอง

จากผลการศึกษาจะเห็นได้ว่าพารามิเตอร์อิทธิพลกำหนด (fixed effect) ทั้ง 4 พารามิเตอร์ อันได้แก่ พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ได้รับผลกระทบจากการระบุคุณลักษณะเฉพาะบางประการของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) ในขณะที่การระบุคุณลักษณะเฉพาะบางประการของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) นี้ ไม่ส่งผลต่อการประมาณค่าพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ ได้แก่ พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

3.2) วิธีดำเนินการวิจัย

3.2.1) เงื่อนไขในการจำลองข้อมูล

การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลด้วยโปรแกรม R และประมาณค่าพารามิเตอร์ด้วยโปรแกรม Mplus ภายใต้การเขียนคำสั่ง การประมวลผลและแสดงผลในโปรแกรม R โดยเป็นการทำงานร่วมกันของทุกขั้นตอนตั้งแต่การจำลองข้อมูลด้วยโปรแกรม R วิเคราะห์ข้อมูลด้วยโปรแกรม Mplus จากการเขียนคำสั่งและสั่งงานเพียงครั้งเดียว ดังต่อไปนี้

อย่างที่ผู้วิจัยได้กล่าวไว้ตั้งแต่ต้นว่า การศึกษาในครั้งนี้เป็นเรื่องใหม่ที่ยังไม่มีนักวิจัยท่านใดการศึกษาค้นคว้ามาก่อน ดังนั้นผู้วิจัยจึงยึดผลการศึกษานำร่องตามที่ได้นำเสนอไป ผลการศึกษาของ Patarapichayatham, Kanjanawasee and Kamata (2009) และจากการทดลองเบื้องต้นรวมทั้งศึกษาค้นคว้าเพิ่มเติม แล้วนำผลที่ได้มากำหนดเป็นเงื่อนไขในการศึกษาค้นคว้า ดังนี้ ผู้วิจัยกำหนดเงื่อนไขในการศึกษาที่แตกต่างกัน 18 เงื่อนไข (2 x 3 x 3) นั่นคือ 1) จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) 2 ระดับ คือ โรงเรียนละ 50 คนและโรงเรียนละ 100 คน โดยกลุ่มตัวอย่างจากโรงเรียนละ 50 คนนั้น ผู้วิจัยกำหนด 25 คนแรกเป็นกลุ่มเปรียบเทียบ (focal group) ส่วนอีก 25 คนหลังเป็นกลุ่มอ้างอิง (reference group) ส่วนกลุ่มตัวอย่างจากโรงเรียนละ 100 คนนั้น ผู้วิจัยกำหนด 50 คนแรกเป็นกลุ่มเปรียบเทียบ (focal group) ส่วนอีก 50 คนที่เหลือเป็นกลุ่มอ้างอิง (reference group) ซึ่งดำเนินการอย่างสุ่ม 2) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 3 ระดับ คือ 0.2, 0.4 และ 0.6 และ 3) ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ คือ 0.1, 0.2 และ 0.3 โดยในแต่ละเงื่อนไขกำหนดการทำซ้ำ 200 รอบ ส่วนจำนวนโรงเรียนนั้นเป็นเงื่อนไขที่ไม่ได้ศึกษา ผู้วิจัยจึงกำหนดคงที่ทุกเงื่อนไขที่ 50 โรงเรียน โดยกำหนด 25 โรงเรียนเป็นกลุ่มเปรียบเทียบ (focal group) ในพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF factor) ส่วนอีก 25 โรงเรียนที่เหลือเป็นกลุ่มอ้างอิง (reference group) ในพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF factor) และการทำงานเป็นไปอย่างสุ่มเช่นกัน ซึ่งจะเห็นได้ว่าจำนวนกลุ่มตัวอย่างในกลุ่มเปรียบเทียบ (focal group) และในกลุ่มอ้างอิง (reference group) นั้นเป็นไปตามเงื่อนไขของการวิเคราะห์พหุระดับ นั่นคือ มีจำนวนที่มากกว่าจำนวนกลุ่มตัวอย่างที่ Snjders and Bosker (1999) ได้กล่าวไว้ โดยเงื่อนไขการจำลองข้อมูลทั้งหมด มีดังต่อไปนี้

ตารางที่ 3.4 การจำลองข้อมูลตามเงื่อนไขจำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ที่แตกต่างกัน

ตัวแปรในการศึกษา			เงื่อนไขในการจำลองข้อมูล	เงื่อนไขที่
จำนวนนักเรียนในแต่ละโรงเรียน	ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF)	ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)		
50	0.2	0.1	(50,0.2,0.1)	1
		0.2	(50,0.2,0.2)	2
		0.3	(50,0.2,0.3)	3
	0.4	0.1	(50,0.4,0.1)	4
		0.2	(50,0.4,0.2)	5
		0.3	(50,0.4,0.3)	6
	0.6	0.1	(50,0.6,0.1)	7
		0.2	(50,0.6,0.2)	8
		0.3	(50,0.6,0.3)	9
100	0.2	0.1	(100,0.2,0.1)	10
		0.2	(100,0.2,0.2)	11
		0.3	(100,0.2,0.3)	12
	0.4	0.1	(100,0.4,0.1)	13
		0.2	(100,0.4,0.2)	14
		0.3	(100,0.4,0.3)	15
	0.6	0.1	(100,0.6,0.1)	16
		0.2	(100,0.6,0.2)	17
		0.3	(100,0.6,0.3)	18

ซึ่งจะเห็นได้ว่าขนาดอิทธิพลของพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 3 ระดับที่ผู้วิจัยศึกษา คือ 0.2, 0.4 และ 0.6 ซึ่งขนาดอิทธิพลตัวที่สองจะเป็นสองเท่าของขนาดอิทธิพลตัวที่

หนึ่งและขนาดอิทธิพลตัวที่สามจะเป็นสามเท่าของขนาดอิทธิพลตัวที่หนึ่งตามลำดับและผู้วิจัยมีแนวคิดในการกำหนดขนาดอิทธิพลของพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) เช่นเดียวกัน นั่นคือ ในการศึกษาในครั้งนี้ ผู้วิจัยกำหนดขนาดอิทธิพลของพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ คือ 0.1, 0.2 และ 0.3 ซึ่งจะเห็นได้ว่าขนาดอิทธิพลตัวที่สองจะเป็นสองเท่าของขนาดอิทธิพลตัวที่หนึ่งและขนาดอิทธิพลตัวที่สามจะเป็นสามเท่าของขนาดอิทธิพลตัวที่หนึ่งตามลำดับเช่นกัน ซึ่งผู้วิจัยมีฐานคิดในการกำหนดขนาดอิทธิพลว่าเมื่อขนาดอิทธิพลของแต่ละพารามิเตอร์เพิ่มขึ้นหนึ่งเท่าตัว จะส่งผลต่อการประมาณค่าพารามิเตอร์ทั้ง 6 พารามิเตอร์เป้าหมายของโมเดล HGLM หรือไม่

3.2.2) โมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF)

ผู้วิจัยใช้โมเดลทั้ง 4 โมเดลตามที่ได้ศึกษาไว้แล้วในการศึกษานำร่อง มาใช้ในการศึกษาในครั้งนี้ นั่นคือ ผู้วิจัยประยุกต์ใช้กรอบแนวคิดของโมเดล HGLM แบบ 3 ระดับ ซึ่งมีตัวแบบหรือสมการ ดังต่อไปนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรระดับข้อสอบ (item level) X_j เป็นตัวแปรระดับบุคคล (individual level) W_k เป็นตัวแปรระดับกลุ่ม (cluster level) ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) และตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) ตัวแปรระดับบุคคล (individual level DIF factor) และตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็น

พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

อย่างไรก็ตาม เนื่องจากการศึกษาในครั้งนี้ ผู้วิจัยมุ่งเน้นไปที่การศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม ดังนั้นผู้วิจัยจึงมีโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) ที่ศึกษาในครั้งนี้ 4 โมเดล ซึ่งแต่ละโมเดลมีคุณลักษณะเฉพาะบางประการของโมเดลที่แตกต่างกัน รายละเอียดของแต่ละโมเดล มีดังต่อไปนี้

โมเดลที่ 1 เป็นโมเดลเต็มรูป (complete model) ซึ่งเป็นโมเดลที่ประกอบด้วยพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 2) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 3) ค่าความยากของข้อสอบข้อที่ *i* (item difficulty) 4) พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 5) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 6) พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 7) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ คือ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม (CA) มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัว

แปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

โมเดลที่ 2 เป็นโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) เป็นโมเดลที่ประกอบด้วยพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 อิทธิพล โดยกำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) เป็นศูนย์ พารามิเตอร์ทั้ง 6 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 2) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 3) ค่าความยากของข้อสอบข้อที่ *i* (item difficulty) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 6) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ คือ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม (CA) มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับ

บุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

โมเดลที่ 3 เป็นโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) เป็นโมเดลที่ประกอบด้วยพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 พารามิเตอร์ โดยกำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์ พารามิเตอร์ทั้ง 6 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 2) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 3) ค่าความยากของข้อสอบข้อที่ *i* (item difficulty) 4) พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 5) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 6) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ คือ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม (CA) มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 6 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i}

เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

โมเดลสุดท้ายเป็นโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ประกอบด้วยพารามิเตอร์อิทธิพลกำหนด (fixed effect) 5 พารามิเตอร์ นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์ พารามิเตอร์ทั้ง 5 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 2) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 3) ค่าความยากของข้อสอบข้อที่ *i* (item difficulty) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ คือ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม (CA) มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_{5i} X_j I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม

(cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 5 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

3.2.3) เกณฑ์การคัดสรรโมเดล (model selection criteria)

จากการศึกษาค้นคว้าที่ผ่านมา การที่นักวิจัยจะตัดสินว่าจะใช้โมเดลไหนในการศึกษานั้น มีฐานคิดอยู่ 2 ฐานคิดสำคัญ ฐานคิดแรกนักวิจัยจะเลือกโมเดลต่างๆ ด้วยตนเองโดยไม่มีเหตุผลหรือหลักฐานทางสถิติใดๆ มารองรับ นั่นคือ นักวิจัยจะเลือกโมเดลโดยใช้วิจารณญาณส่วนตัว รวมไปถึงการที่จะละเลยหรือเพิกเฉยต่อพารามิเตอร์บางพารามิเตอร์ในโมเดลก็ขึ้นอยู่กับวิจารณญาณส่วนตัว เช่นกัน ส่วนอีกฐานคิดหนึ่งนักวิจัยจะใช้หลักฐานทางสถิติเป็นเกณฑ์ในการเลือกโมเดล นั่นคือการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) เป็นเกณฑ์ในการตัดสินว่าโมเดลไหนเป็นโมเดลที่ดีที่สุด (best model) หรือเป็นโมเดลที่เหมาะสมที่สุด (Kamata and Cheong, 2007; Kamata, 2009) โดยเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่เป็นที่นิยม ได้แก่ ดัชนี AIC (Akaike's information criterion) ดัชนี BIC (Bayesian information criterion) และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) ซึ่งเป็นดัชนีที่นิยมใช้กันอย่างแพร่หลายและกว้างขวาง

Shibata (2002) กล่าวว่า แม้ว่าดัชนี AIC (Akaike's information criterion) ดัชนี BIC (Bayesian information criterion) และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) จะเป็นดัชนีที่นิยมใช้กันอย่างแพร่หลาย แต่การที่จะพิจารณาว่าจะใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใดเป็นเกณฑ์ในการศึกษานั้น ยังไม่มีหลักเกณฑ์หรือข้อค้นพบที่ชัดเจน ดังนั้นการที่นักวิจัยจะเลือกใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใด

นั้น จึงขึ้นอยู่กับฐานคิดของนักวิจัยแต่ละคนว่าจะใช้เกณฑ์ใดถึงจะเหมาะสมที่สุด ซึ่งจากการศึกษาค้นคว้าที่ผ่านมา นักวิจัยส่วนใหญ่จะเลือกใช้เพียง ดัชนี AIC (Akaike's information criterion) หรือดัชนี BIC (Bayesian information criterion) หรือดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เพียงดัชนีใดดัชนีหนึ่งเท่านั้น ซึ่งสอดคล้องกับแนวคิดของ Muthen and Muthen (2007) และ Kamata (2009)

ด้วยเหตุผลดังกล่าวมาข้าวต้น นักวิจัยบางท่านจึงเลือกใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) เพียง 1 เกณฑ์เท่านั้น ในการคัดสรรโมเดลที่ดีที่สุด (best model) ตัวอย่างเช่น การศึกษาของ Kamata and Cheong (2007) ที่ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับด้วยโมเดล HGLM (multilevel DIF detection by Hierarchical Generalized Linear Modeling (HLGM)) ที่ใช้เพียงดัชนี AIC (Akaike's information criterion) เท่านั้น รวมทั้งการศึกษาในครั้งนั้น กำหนดให้ค่าพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (GLDIF) เป็นศูนย์ ซึ่งสอดคล้องกับการศึกษาของ Chaimongkol, Huffer and Kamata (2006) และ Vaughn (2006)

อย่างไรก็ตาม เพื่อให้ผลการศึกษาในครั้งนี้มีความน่าเชื่อถือมากที่สุด รวมไปถึงการศึกษาประสิทธิภาพของเกณฑ์การคัดสรรโมเดล (Model selection criteria) แต่ละเกณฑ์ ดังนั้นในการศึกษาครั้งนี้ ผู้วิจัยจึงกำหนดเกณฑ์การคัดสรรโมเดล (Model selection criteria) ถึง 5 เกณฑ์ ซึ่งทั้ง 5 เกณฑ์นี้เป็นเกณฑ์ที่นักวิจัยนิยมใช้ แต่การศึกษาที่ผ่านมายังไม่มีหลักฐานที่แน่ชัดว่าเกณฑ์ไหนดีที่สุด หรือผลของการศึกษาด้วยเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่แตกต่างกันได้ผลของโมเดลที่ดีที่สุด (best model) เหมือนหรือแตกต่างกันอย่างไร

ผู้วิจัยออกแบบการวิจัยโดยใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ในส่วนที่ 2 และส่วนที่ 3 ของการออกแบบการวิจัย โดยผู้วิจัยได้อธิบายรายละเอียดไว้ในหัวข้อ 3.2.6) การออกแบบการวิจัย อย่างไรก็ตาม ผู้วิจัยขอนำเสนอรายละเอียดของแต่ละเกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังนี้

เกณฑ์ที่ 1 คือ ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่พัฒนาขึ้นในปี 1971 โดย Hirotsugu Akaike เป็นที่รู้จักกันโดยทั่วไปในชื่อดัชนี AIC (Akaike's information criterion) ตั้งแต่ปี 1974 เป็นต้นมา โดยดัชนี AIC (Akaike's information criterion) นี้ มีสมการทั่วไป คือ $AIC = -2\log(L) + 2r$ เมื่อ r คือจำนวนพารามิเตอร์ในโมเดล (the number of model parameter) และ L คือ ค่าที่สูงที่สุดของ likelihood function ในโมเดลการประมาณค่า (estimated model) โดยค่าดัชนี AIC (Akaike's information criterion) ที่น้อยกว่าแสดงว่าโมเดลทางสถิติ (statitical model) และโมเดลที่ต้องการประมาณค่า

(estimate model) มีความสอดคล้องกันมากกว่า (smaller is better) นั่นคือ เป็นโมเดลที่ดีกว่านั่นเอง โดยค่าดัชนี AIC (Akaike's information criterion) นี้เป็นตัวบ่งชี้ถึงโมเดลที่ดีที่สุด (best model) ที่สามารถนำไปตีความหมายได้อย่างมีนัยสำคัญทางสถิติ (Shibata, 2002; Briggs and Wilson, 2003; Allen and Wilson, 2006; Muthen and Muthen, 2007; Kamata, 2009)

ในทุกเงื่อนไข ผู้วิจัยกำหนดการทำซ้ำ 200 รอบ โดยในแต่ละการทำซ้ำ ผู้วิจัยกำหนดดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่ 1 หลักการทำงานมีดังนี้ ผู้วิจัยเขียนคำสั่งให้โปรแกรม R นำค่าดัชนี AIC (Akaike's information criterion) จากทั้ง 4 โมเดลมาเปรียบเทียบกัน หากโมเดลไหนมีค่าน้อยที่สุด โมเดลนั้นจะเป็นโมเดลที่ดีที่สุด (best model) ตัวอย่างเช่นโมเดลที่ 1 มีค่าดัชนี AIC (Akaike's information criterion) น้อยที่สุด โมเดลที่ 1 จะเป็นโมเดลที่ดีที่สุด (best model) สำหรับดัชนี AIC (Akaike's information criterion)

เกณฑ์ที่ 2 คือ ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่เป็นที่รู้จักกันโดยทั่วไปในชื่อดัชนี BIC (Bayesian information criterion) ดัชนี SBC (Schwarz Bayesian Criterion) หรือ SBIC (Schwarz Bayesian information criterion) พัฒนาโดย Gideon E. Schwarz เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่ให้ค่าที่ใกล้เคียงกับดัชนี AIC (Akaike's information criterion) โดยดัชนี BIC (Bayesian information criterion)มีสมการทั่วไป คือ BIC = -2log(L) + rlog(n) เมื่อ n คือ จำนวนค่าสังเกตุ (the number of observations) หรือ กลุ่มตัวอย่าง (the sample size), และ L คือ ค่ามากที่สุดของ likelihood function ในโมเดลการประมาณค่า (estimated model) โดยค่าดัชนี BIC (Bayesian information criterion หรือ Schwarz Criterion) ที่น้อยกว่าแสดงว่าโมเดลทางสถิติ (statitical model) และโมเดลที่ต้องการประมาณค่า (estimate model) มีความสอดคล้องกันมากกว่า (smaller is better) นั่นคือ เป็นโมเดลที่ดีกว่านั่นเอง ซึ่งสามารถนำไปตีความหมายได้อย่างมีนัยสำคัญทางสถิติ (Shibata, 2002; Briggs and Wilson, 2003; Allen and Wilson, 2006; Muthen and Muthen, 2007; Kamata, 2009)

หลักการทำงานของดัชนี BIC (Bayesian information criterion) เหมือนกับหลักการทำงานของดัชนี AIC (Akaike's information criterion) นั่นคือ ในทุกเงื่อนไข ผู้วิจัยกำหนดการทำซ้ำ 200 รอบ โดยในแต่ละการทำซ้ำ ผู้วิจัยกำหนดดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่ 2 หลักการทำงานมีดังนี้ ผู้วิจัยเขียนคำสั่งให้โปรแกรม R นำค่าดัชนี BIC (Bayesian information criterion หรือ Schwarz Criterion) จากทั้ง 4 โมเดลมาเปรียบเทียบกัน หากโมเดลไหนมีค่าน้อยที่สุด โมเดลนั้นจะเป็นโมเดลที่ดีที่สุด (best model) ตัวอย่างเช่นโมเดลที่ 2 มีค่าดัชนี BIC (Bayesian information criterion) น้อยที่สุด โมเดลที่ 2 จะเป็นโมเดลที่ดีที่สุด (best model) สำหรับดัชนี BIC (Bayesian information criterion)

เกณฑ์ที่ 3 คือ ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นอีกเกณฑ์การคัดสรรโมเดล (Model selection criteria) หนึ่งที่ได้รับความนิยมอย่างแพร่หลาย โดยดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) มีสมการทั่วไป คือ ABIC = -2log(L) +rln n^* เมื่อ n^* = (n + 2) / 24 ซึ่งจะเห็นว่าขนาดของกลุ่มตัวอย่างจะส่งผลต่อค่าประมาณที่ได้จากการประมาณค่าด้วยดัชนีนี้ โดยค่าดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) ที่น้อยกว่าแสดงว่าโมเดลทางสถิติ (statitical model) และโมเดลที่ต้องการประมาณค่า (estimate model) มีความสอดคล้องกันมากกว่า (smaller is better) นั่นคือ เป็นโมเดลที่ดีกว่านั่นเอง โดยค่าดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) นี้เป็นตัวบ่งชี้ถึงโมเดลที่ดีที่สุด (best model) ที่สามารถนำไปตีความหมายได้อย่างมีนัยสำคัญทางสถิติ (Shibata, 2002; Briggs and Wilson, 2003; Allen and Wilson, 2006; Muthen and Muthen, 2007; Kamata, 2009)

หลักการทำงานของดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เหมือนกับหลักการทำงานของดัชนี AIC (Akaike's information criterion) และดัชนี BIC (Bayesian information criterion) นั่นคือ ในทุกเงื่อนไข ผู้วิจัยกำหนดการทำซ้ำ 200 รอบ โดยในแต่ละการทำซ้ำ ผู้วิจัยกำหนดดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่ 3 หลักการทำงานมีดังนี้ ผู้วิจัยเขียนคำสั่งให้โปรแกรม R นำค่าดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) จากทั้ง 4 โมเดลมาเปรียบเทียบกัน หากโมเดลไหนมีค่าน้อยที่สุด โมเดลนั้นจะเป็นโมเดลที่ดีที่สุด (best model) ตัวอย่างเช่นโมเดลที่ 3 มีค่าดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) น้อยที่สุด โมเดลที่ 3 จะเป็นโมเดลที่ดีที่สุด (best model) สำหรับดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)

เกณฑ์ที่ 4 คือ ดัชนี 2 of 3 เป็นดัชนีที่ผู้วิจัยพัฒนาขึ้น เพื่อยืนยันผลจากการใช้ดัชนี AIC (Akaike's information criterion) ดัชนี BIC (Bayesian information criterion) และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) โดยมีเกณฑ์ในการพิจารณา 3 กรณี ดังต่อไปนี้ กรณีที่ 1 สมมติจากการทำซ้ำรอบที่ 1 พบว่า เมื่อใช้ดัชนี AIC (Akaike's information criterion) ดัชนี BIC (Bayesian information criterion) และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 1 เป็นโมเดลที่ดีที่สุด (best model) ทั้ง 3 เกณฑ์ จะสรุปได้ว่าโมเดลที่ 1 เป็นโมเดลที่ดีที่สุด (best model) ของดัชนี 2 of 3 กรณีที่ 2 สมมติจากการทำซ้ำรอบที่ 2 พบว่าเมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 1 เป็นโมเดลที่ดีที่สุด (best model) เมื่อใช้ดัชนี BIC (Bayesian information criterion) และดัชนี

ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 2 เป็นโมเดลที่ดีที่สุด (best model) ทั้ง 2 เกณฑ์ จะสรุปได้ว่าโมเดลที่ 2 เป็นโมเดลที่ดีที่สุด (best model) ของดัชนี 2 of 3 กรณีที่ 3 สมมติจากการทำซ้ำรอบที่ 3 พบว่าเมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 1 เป็นโมเดลที่ดีที่สุด (best model) เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 2 เป็นโมเดลที่ดีที่สุด (best model) และเมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 3 เป็นโมเดลที่ดีที่สุด (best model) ดังนั้นจะสรุปได้ว่าไม่มีโมเดลไหนเป็นโมเดลที่ดีที่สุด (best model) ของดัชนี 2 of 3

เกณฑ์สุดท้าย คือ ดัชนี p – value ซึ่งเป็นอีกดัชนีหนึ่งที่นักวิจัยนิยมใช้อย่างแพร่หลาย โดยนักวิจัยจะพิจารณาดัชนี p – value ของตัวแปรที่ศึกษา แต่เนื่องจากการศึกษาในครั้งนี้ผู้วิจัยมุ่งเน้นไปที่การศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม ดังนั้นผู้วิจัยจึงพิจารณาดัชนี p – value ของพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) โดยผู้วิจัยกำหนดระดับนัยสำคัญทางสถิติไว้ที่ 0.05 โดยมีเกณฑ์ในการพิจารณา 3 กรณี ดังต่อไปนี้ กรณีที่ 1 สมมติจากการทำซ้ำรอบที่ 1 พบว่าทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) มีนัยสำคัญสำคัญทางสถิติที่ระดับ 0.05 ในโมเดลที่ 1 จะสรุปได้ว่าโมเดลที่ 1 เป็นโมเดลที่ดีที่สุด (best model) ของดัชนี p – value กรณีที่ 2 สมมติจากการทำซ้ำรอบที่ 2 พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) มีนัยสำคัญสำคัญทางสถิติที่ระดับ 0.05 ในโมเดลที่ 1 ในขณะที่พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) มีนัยสำคัญสำคัญทางสถิติที่ระดับ 0.05 ในโมเดลที่ 2 จะสรุปได้ว่าไม่มีโมเดลไหนเป็นโมเดลที่ดีที่สุด (best model) กรณีที่ 3 สมมติจากการทำซ้ำรอบที่ 3 พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) มีนัยสำคัญสำคัญทางสถิติที่ระดับ 0.05 ในโมเดลที่ 3 ในขณะที่พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) ไม่มีนัยสำคัญสำคัญทางสถิติจะสรุปได้ว่าไม่มีโมเดลไหนเป็นโมเดลที่ดีที่สุด (best model)

3.2.4) **พารามิเตอร์เป้าหมายในการประมาณค่า**

การศึกษาในครั้งนี้มีพารามิเตอร์เป้าหมายในการประมาณค่า 6 พารามิเตอร์ เป็นพารามิเตอร์อิทธิพลกำหนด (fixed effect) 4 พารามิเตอร์และพารามิเตอร์อิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ โดยพารามิเตอร์อิทธิพลกำหนด (fixed effect) 4 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 2) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) 3) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) และ 4)

พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ส่วนอิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 2) พารามิเตอร์ความสามารถระดับกลุ่ม (CA)

3.2.5) ขั้นตอนการจำลองข้อมูลและเกณฑ์การประเมิน

ผู้วิจัยเริ่มต้นจากการศึกษาโปรแกรม R และเขียนคำสั่งในการจำลองข้อมูล โดยเป็นคำสั่งในการจำลองข้อมูลแบบ 3 ระดับ ข้อมูลระดับที่ 1 เป็นข้อมูลระดับข้อสอบ (item level) ซึ่งเป็นข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่า (dichotomous) จำนวน 12 ข้อ โดยข้อสอบแต่ละข้อมีค่าความยากของข้อที่ 1 ไปถึงข้อ 12 ดังนี้ 0, -2.0, -1.5, -1.0, -0.5, 0, 0, 0.5, 1.0, 1.5, 2.0 และ 2.5 ข้อสอบข้อที่ 1 เพียงข้อเดียวเท่านั้นที่ถูกกำหนดให้เป็นข้อสอบที่เกิดการทำหน้าที่ต่างกัน (DIF) โดยผู้วิจัยกำหนดให้ความแตกต่างของค่าเฉลี่ยความสามารถของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (the mean ability difference between the reference and focal groups) เท่ากับ 0.5 นั่นคือ กลุ่มเปรียบเทียบ (focal group) มีค่าเฉลี่ยความสามารถต่ำกว่ากลุ่มอ้างอิง (reference group) 0.5

ผู้วิจัยกำหนดเงื่อนไขในการศึกษาที่แตกต่างกัน 18 เงื่อนไข (2 x 3 x 3) นั่นคือ 1) จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) 2 ระดับ คือ โรงเรียนละ 50 คนและโรงเรียนละ 100 คน โดยกลุ่มตัวอย่างจากโรงเรียนละ 50 คนนั้น ผู้วิจัยกำหนด 25 คนแรกเป็นกลุ่มเปรียบเทียบ (focal group) ส่วนอีก 25 คนหลังเป็นกลุ่มอ้างอิง (reference group) ส่วนกลุ่มตัวอย่างจากโรงเรียนละ 100 คนนั้น ผู้วิจัยกำหนด 50 คนแรกเป็นกลุ่มเปรียบเทียบ (focal group) ส่วนอีก 50 คนที่เหลือเป็นกลุ่มอ้างอิง (reference group) ซึ่งดำเนินการอย่างสุ่ม 2) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 3 ระดับ คือ 0.2, 0.4 และ 0.6 และ 3) ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ คือ 0.1, 0.2 และ 0.3 โดยในแต่ละเงื่อนไขกำหนดการทำซ้ำ 200 รอบ ส่วนจำนวนโรงเรียนนั้นเป็นเงื่อนไขที่ไม่ได้ศึกษา ผู้วิจัยจึงกำหนดคงที่ทุกเงื่อนไขที่ 50 โรงเรียน โดยกำหนด 25 โรงเรียนเป็นกลุ่มเปรียบเทียบ (focal group) ส่วนอีก 25 โรงเรียนที่เหลือเป็นกลุ่มอ้างอิง (reference group) และการทำงานเป็นไปอย่างสุ่มเช่นกัน

นอกจากนั้น ผู้วิจัยได้กำหนดค่าจริง (true value) ของอิทธิพลกำหนด (fixed effect) ทั้ง 4 อิทธิพลของโมเดล HGLM และอิทธิพลสุ่ม (random effect) 2 อิทธิพล โดยยึดผลการศึกษานำร่องเบื้องต้นก่อนหน้านี้เป็นเกณฑ์ในการกำหนดค่าพารามิเตอร์แต่ละค่า ดังต่อไปนี้ พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) กำหนดเป็น 0.1 พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) กำหนดเป็น 0.3 พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) กำหนดเป็น 0.2 พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) กำหนด

เป็น -0.5 ส่วนอิทธิพลสุ่ม (random effect) 2 พารามิเตอร์ ผู้วิจัยกำหนดค่าดังนี้ พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) กำหนดเป็น 1.0 และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) กำหนดเป็น 1.0

ขั้นตอนต่อไปเป็นการเขียนคำสั่งในโปรแกรม R เพื่อให้โปรแกรมนำข้อมูลจำลองที่จำลองได้ไปประมาณค่าพารามิเตอร์ด้วยโปรแกรม Mplus ภายใต้การทำงานในโปรแกรม R โดยผู้วิจัยเขียนคำสั่งให้การทำงานทั้งหมดเสร็จสิ้นในขั้นตอนเดียว นั่นคือ การจำลองข้อมูลด้วยโปรแกรม R จากนั้นโปรแกรม R ส่งข้อมูลไปประมาณค่าพารามิเตอร์ในโปรแกรม Mplus แล้วแสดงผลเป็น external file เก็บไว้ใน drive D โดยผู้วิจัยเขียนคำสั่งให้โปรแกรม R คำนวณค่าสถิติที่ต้องการทุกค่า รวมไปถึงการคำนวณค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ซึ่งทั้ง 3 ค่านี้เป็นเกณฑ์การประเมิน (evaluation criteria) ในการประมาณค่าพารามิเตอร์ในการศึกษาในครั้งนี้ นั่นคือ ค่าความลำเอียง (BIAS) เป็นความคลาดเคลื่อนอย่างเป็นระบบ (systematic error) ค่าความคลาดเคลื่อนมาตรฐาน (SE) เป็นความคลาดเคลื่อนอย่างสุ่ม (random error)และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) เป็นความคลาดเคลื่อนรวมทั้งหมด (total error) รายละเอียดของการคำนวณแต่ละสูตรมีดังต่อไปนี้ (Chu and Kamata, 2006; Vaughn, 2006; Binici, 2007; Kamata, 2009)

$$RMSE(\hat{\alpha}_j) = \sqrt{\frac{1}{s}\sum_{i=1}^{s}(\hat{\alpha}_{ij} - \alpha_j)^2}$$

$$Bias(\hat{\alpha}_j) = \frac{1}{s}\sum_{i=1}^{s}\hat{\alpha}_{ij} - \alpha_j$$

$$SE(\hat{\alpha}_j) = \sqrt{\frac{1}{s}\sum_{i=1}^{s}(\hat{\alpha}_{ij} - \frac{1}{s}\sum_{i=1}^{s}\hat{\alpha}_{ij})^2}$$

เมื่อ $\hat{\alpha}_j$ เป็นค่าความยากที่แท้จริงของข้อสอบข้อที่ j

$\hat{\alpha}_{ij}$ เป็นค่าความยากจากการประมาณค่าข้อสอบข้อที่ j จากการทำซ้ำ s ครั้ง

3.2.6) การออกแบบการวิจัย

ผู้วิจัยออกแบบการวิจัยในส่วนของการประมาณค่าพารามิเตอร์ โดยแบ่งออกเป็น 3 ส่วนสำคัญ ดังรายละเอียดต่อไปนี้ ส่วนแรกเป็นการประมาณค่าพารามิเตอร์ก่อนการใช้เกณฑ์การคัดสรรโมเดล (model selection criteria) ทั้ง 5 เกณฑ์ นั่นคือ ผู้วิจัยเขียนคำสั่งเพื่อให้โปรแกรม R และโปรแกรม Mplus ทำงานร่วมกันในการประมาณค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของแต่ละการทำซ้ำทั้ง 200 รอบ จากนั้นผู้วิจัยเขียนคำสั่งเพื่อให้โปรแกรม R คำนวณค่าเฉลี่ย (mean) ของค่าเฉลี่ย (mean) และค่าเฉลี่ย (mean) ของส่วนเบี่ยงเบนมาตรฐาน (SD) จากทั้ง 200 รอบ ซึ่งก็จะได้ค่าประมาณของพารามิเตอร์ทั้ง 6 พารามิเตอร์จากทั้ง โมเดล 4 การออกแบบการวิจัยในขั้นตอนแรกนี้ นอกจากจะตอบวัตถุประสงค์ในการศึกษาในครั้งนี้แล้ว ผู้วิจัยยังสามารถนำสารสนเทศที่ได้ไปเปรียบเทียบกับผลการประมาณค่าพารามิเตอร์หลังจากการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์แล้วได้อีกด้วย โดยผู้วิจัยนำเสนอผลที่ได้จากการออกแบบการวิจัยในส่วนแรกนี้ไว้ในตอนที่ 1 ของบทที่ 4

ค่าพารามิเตอร์จากการออกแบบการวิจัยในส่วนแรกนี้จะถูกนำมาใช้ในการออกแบบการวิจัยในส่วนที่ 2 นั่นคือ ค่าพารามิเตอร์ที่ประมาณได้จะถูกนำมาเปรียบเทียบกันเพื่อหาโมเดลที่ดีที่สุด (best model) ในแต่ละการทำซ้ำ โดยใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ เป็นเกณฑ์ในการตัดสิน ดังรายละเอียดที่ผู้วิจัยได้อธิบายไว้ในหัวข้อ 3.2.3) เกณฑ์การคัดสรรโมเดล (model selection criteria) โดยผู้วิจัยได้นำเสนอผลของการออกแบบการวิจัยในส่วนที่ 2 นี้ไว้ในตอนที่ 2 ของบทที่ 4 ซึ่งเป็นการนำเสนอโมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) รวมไปถึงการสรุปร้อยละของโมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ด้วย

การออกแบบการวิจัยในส่วนที่ 3 นี้ เป็นการสกัดค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของทั้ง 6 พารามิเตอร์จากโมเดลที่ดีที่สุด (best model) ที่สกัดได้จากการออกแบบการวิจัยในส่วนที่ 2 หลักการทำงาน มีดังนี้ ตัวอย่างเช่น การทำซ้ำที่ 1 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์ พบว่าโมเดลที่ 1 เป็นโมเดลที่ดีที่สุด (best model) โปรแกรม R จะสกัดค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของทั้ง 6 พารามิเตอร์จากโมเดลที่ 1 เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์ พบว่าโมเดลที่ 2 เป็นโมเดลที่ดีที่สุด (best model) โปรแกรม R จะสกัดค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของทั้ง 6 พารามิเตอร์จากโมเดลที่ 2 เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์ พบว่าโมเดลที่ 3 เป็นโมเดลที่ดีที่สุด (best model) โปรแกรม R จะสกัดค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของทั้ง 6 พารามิเตอร์จากโมเดลที่ 3 เมื่อใช้ดัชนี 2 ใน 3 เป็นเกณฑ์ พบว่าโมเดลที่ 3 เป็นโมเดลที่ดีที่สุด (best model) โปรแกรม R จะสกัดค่าเฉลี่ย (mean)

และส่วนเบี่ยงเบนมาตรฐาน (SD) ของทั้ง 6 พารามิเตอร์จากโมเดลที่ 3 และเมื่อใช้ดัชนี p-value เป็นเกณฑ์ พบว่าโมเดลที่ 4 เป็นโมเดลที่ดีที่สุด (best model) โปรแกรม R จะสกัดค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของทั้ง 6 พารามิเตอร์จากโมเดลที่ 4 ขั้นตอนการทำงานจะเป็นแบบนี้ไปเรื่อยๆ จนครบ 200 รอบในแต่ละเงื่อนไข ผู้วิจัยได้นำเสนอตัวอย่างของผลการสกัดค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์แต่ละพารามิเตอร์ในแต่ละรอบของการทำซ้ำในภาคผนวก ง

จากนั้น ผู้วิจัยเขียนคำสั่งให้โปรแกรม R หาค่าเฉลี่ย (mean) ของค่าเฉลี่ย (mean) และค่าเฉลี่ย (mean) ของส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์แต่ละพารามิเตอร์จากการทำซ้ำทั้ง 200 รอบในแต่ละเงื่อนไข จำแนกตามเกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์และได้นำเสนอผลไว้ในตอนที่ 3 ของบทที่ 4

นอกจากนี้ ผู้วิจัยยังเขียนคำสั่งให้โปรแกรม R คำนวณค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของแต่ละพารามิเตอร์ในแต่ละเงื่อนไข จำแนกตามเกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์และได้นำเสนอผลไว้ในตอนที่ 4 ของบทที่ 4

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม และเพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)

การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลด้วยเงื่อนไขที่แตกต่างกันจำนวน 18 เงื่อนไข (2 x 3 x 3) อันประกอบไปด้วย 1) จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school or the number of individual in each cluster) 2 ระดับ คือ โรงเรียนละ 50 คนและโรงเรียนละ 100 คน 2) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 3 ระดับ คือ 0.2, 0.4 และ 0.6 และ 3) ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ คือ 0.1, 0.2 และ 0.3 ส่วนจำนวนโรงเรียนนั้นกำหนดคงที่ทุกเงื่อนไขที่ 50 โรงเรียน โดยในแต่ละเงื่อนไขกำหนดการทำซ้ำเงื่อนไขละ 200 รอบ

ผู้วิจัยมีโมเดลการทำหน้าที่ต่างกันของข้อสอบที่ศึกษาในครั้งนี้ 4 โมเดล ได้แก่ 1) โมเดลเต็มรูป (complete model) ซึ่งเป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) ทั้ง 7 อิทธิพลและอิทธิพลสุ่ม (random effect) 2 อิทธิพล 2) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล 3) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพลและ 4) โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 5 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล

พารามิเตอร์เป้าหมายในการประมาณค่ามี 6 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 2) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) 3) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 6) พารามิเตอร์ความสามารถระดับกลุ่ม (CA)

ผู้วิจัยมีเกณฑ์การคัดสรรโมเดล (Model selection criteria) 5 เกณฑ์ ได้แก่ 1) ดัชนี AIC (Akaike's information criterion), 2) ดัชนี BIC (Bayesian information criterion), 3) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), 4) ดัชนี 2 of 3 และ 5) ดัชนี p-value และมีเกณฑ์การประเมิน (evaluation criteria) ในการประมาณค่าพารามิเตอร์ 3 เกณฑ์ คือ 1) ค่าความลำเอียง (BIAS) 2) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และ 3) ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลด้วยโปรแกรม R และประมาณค่าพารามิเตอร์ด้วยวิธี Maximum likelihood ด้วยโปรแกรม Mplus ภายใต้การเขียนคำสั่ง การนำข้อมูลไปวิเคราะห์และการแสดงผลในโปรแกรม R

เพื่อให้เกิดความเข้าใจที่ตรงกันในการแปลความหมายของผลการวิเคราะห์ข้อมูล ผู้วิจัยขอกำหนดสัญลักษณ์และอักษรย่อแทนชุดของข้อมูล เงื่อนไขในการศึกษาและค่าสถิติ ดังต่อไปนี้

Mean	คือ	ค่าเฉลี่ย
SD	ค่า	ค่าส่วนเบี่ยงเบนมาตรฐาน (SD)
BIAS	คือ	ค่าความลำเอียง (BIAS)
SE	คือ	ค่าความคลาดเคลื่อนมาตรฐาน (standard error)
RMSE	คือ	ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE)
T	คือ	ค่าจริง (true value)
f	คือ	ความถี่ (frequency)
M 1	คือ	โมเดลเต็มรูป (complete model)
M 2	คือ	โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction)
M 3	คือ	โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF)
M 4	คือ	โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF)

ILADIFF (γ_1)	คือ	พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF)
CLADIFF (γ_2)	คือ	พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF)
2WAYINT (γ_4)	คือ	พารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)
ILDIF (γ_{5i})	คือ	พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF)
CLDIF (γ_{6i})	คือ	พารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF)
3WAYINT (γ_{7i})	คือ	พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT)
IAWC ($\theta_j^{(2)}$)	คือ	พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC)
CA ($\theta_k^{(3)}$)	คือ	พารามิเตอร์ความสามารถระดับกลุ่ม (CA)
MSC	คือ	เกณฑ์การคัดสรรโมเดล (Model selection criteria)
1	คือ	ดัชนี AIC (Akaike's information criterion)
2	คือ	ดัชนี BIC (Bayesian information criterion)
3	คือ	ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)
4	คือ	ดัชนี 2 of 3
5	คือ	ดัชนี p-value
R	คือ	การทำซ้ำ (replication)
P	คือ	พารามิเตอร์ (parameter)
PEST	คือ	การประมาณค่าพารามิเตอร์ (parameter estimates)
Con	คือ	เงื่อนไขการจำลองข้อมูล (simulation condition)
C1	คือ	เงื่อนไขการจำลองข้อมูลที่ 1
C2	คือ	เงื่อนไขการจำลองข้อมูลที่ 2
C3	คือ	เงื่อนไขการจำลองข้อมูลที่ 3
C4	คือ	เงื่อนไขการจำลองข้อมูลที่ 4
C5	คือ	เงื่อนไขการจำลองข้อมูลที่ 5
C6	คือ	เงื่อนไขการจำลองข้อมูลที่ 6
C7	คือ	เงื่อนไขการจำลองข้อมูลที่ 7
C8	คือ	เงื่อนไขการจำลองข้อมูลที่ 8
C9	คือ	เงื่อนไขการจำลองข้อมูลที่ 9

C10 คือ เงื่อนไขการจำลองข้อมูลที่ 10
C11 คือ เงื่อนไขการจำลองข้อมูลที่ 11
C12 คือ เงื่อนไขการจำลองข้อมูลที่ 12
C13 คือ เงื่อนไขการจำลองข้อมูลที่ 13
C14 คือ เงื่อนไขการจำลองข้อมูลที่ 14
C15 คือ เงื่อนไขการจำลองข้อมูลที่ 15
C16 คือ เงื่อนไขการจำลองข้อมูลที่ 16
C17 คือ เงื่อนไขการจำลองข้อมูลที่ 17
C18 คือ เงื่อนไขการจำลองข้อมูลที่ 18

ในการนำเสนอผลการประมาณค่าพารามิเตอร์นั้น ผู้วิจัยแบ่งการนำเสนอผลการประมาณค่าพารามิเตอร์ออกเป็น 4 ตอน ตามขั้นตอนของการออกแบบการวิจัยในบทที่ 3 โดยแต่ละตอนผู้วิจัยนำเสนอผลด้วยตารางสังเคราะห์ บรรยายใต้ตารางและแผนภาพ โดยเริ่มต้นจากตอนที่ 1 เป็นการนำเสนอค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของการประมาณค่าพารามิเตอร์ทั้ง 6 พารามิเตอร์จากทั้ง 4 โมเดลก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ตอนที่ 2 เป็นการนำเสนอโมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำเมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ รวมไปถึงการสรุปร้อยละของโมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำจากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ตอนที่ 3 เป็นการสรุปค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของการประมาณค่าพารามิเตอร์จากการใช้ทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และตอนที่ 4 เป็นการสรุปค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของทั้ง 18 เงื่อนไข สำหรับผลการสกัดค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ผู้วิจัยสกัดออกมาจากแต่ละโมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำจากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) นั้น ผู้วิจัยนำเสนอตัวอย่างไว้ในภาคผนวก ง ผู้วิจัยขอนำเสนอผลการศึกษา ดังนี้

ตอนที่ 1 ค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของการประมาณค่าพารามิเตอร์จากทั้ง 4 โมเดล

ในตอนที่ 1 นี้ ผู้วิจัยนำเสนอค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของการประมาณค่าพารามิเตอร์ทั้ง 6 พารามิเตอร์จากทั้ง 4 โมเดล ซึ่งเป็นผลก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ ได้ผลดังตารางที่ 4.1

ตารางที่ 4.1 ค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของการประมาณค่าพารามิเตอร์ก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

Con	P	T	PEST			
			M1	M2	M3	M4
			Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
1	CLADIFF(γ_2)	0.1	0.109(0.174)	0.150(0.169)	0.125(0.179)	0.151(0.169)
	3WAYINT(γ_{7i})	0.3	0.296(0.096)	0.327(0.095)	0.492(0.038)	0.505(0.139)
	ILADIFF(γ_1)	0.2	0.213(0.072)	0.259(0.049)	0.222(0.072)	0.259(0.049)
	ILDIF(γ_{5i})	-0.5	-0.506(0.144)	-0.522(0.146)	-0.604(0.122)	-0.610(0.123)
	IAWC($\theta_j^{(2)}$)	1.0	1.095(0.058)	1.095(0.058)	1.094(0.058)	1.095(0.057)
	CA($\theta_k^{(3)}$)	1.0	0.969(0.198)	0.974(0.201)	0.969(0.199)	0.972(0.199)
2	CLADIFF(γ_2)	0.1	0.111(0.107)	0.161(0.105)	0.130 (0.106)	0.169(0.106)
	3WAYINT(γ_{7i})	0.3	0.275(0.001)	0.309(0.001)	0.496(0.001)	0.509(0.001)
	ILADIFF(γ_1)	0.2	0.208(0.046)	0.258(0.047)	0.219(0.045)	0.258(0.047)
	ILDIF(γ_{5i})	-0.5	-0.480(0.097)	-0.497(0.098)	-0.589(0.099)	-0.596(0.097)
	IAWC($\theta_j^{(2)}$)	1.0	1.094(0.044)	1.096 (0.046)	1.094 (0.045)	1.095(0.045)
	CA($\theta_k^{(3)}$)	1.0	0.986(0.100)	0.987(0.101)	0.986(0.100)	0.987(0.102)
3	CLADIFF(γ_2)	0.1	0.089(0.106)	0.146(0.104)	0.111(0.106)	0.150(0.105)
	3WAYINT(γ_{7i})	0.3	0.313(0.001)	0.348(0.002)	0.503(0.001)	0.516(0.001)
	ILADIFF(γ_1)	0.2	0.202(0.046)	0.256(0.047)	0.212(0.046)	0.256(0.047)
	ILDIF(γ_{5i})	-0.5	-0.503(0.097)	-0.520(0.097)	-0.597 (0.098)	-0.604 (0.099)
	IAWC($\theta_j^{(2)}$)	1.0	1.095 (0.045)	1.095(0.043)	1.094(0.044)	1.095(0.045)
	CA($\theta_k^{(3)}$)	1.0	1.000(0.104)	0.999(0.103)	0.998(0.103)	1.001(0.104)
4	CLADIFF(γ_2)	0.1	0.102(0.107)	0.161(0.106)	0.123(0.107)	0.169(0.109)
	3WAYINT(γ_{7i})	0.3	0.294(0.001)	0.331(0.002)	0.496(0.001)	0.511(0.001)
	ILADIFF(γ_1)	0.2	0.198(0.046)	0.253(0.047)	0.208(0.046)	0.253(0.047)
	ILDIF(γ_{5i})	-0.5	-0.511(0.096)	-0.531(0.097)	-0.613(0.098)	-0.621(0.099)
	IAWC($\theta_j^{(2)}$)	1.0	1.105 (0.046)	1.107(0.045)	1.105(0.044)	1.106(0.045)
	CA($\theta_k^{(3)}$)	1.0	0.953(0.098)	0.954(0.099)	0.953(0.098)	0.953(0.098)

ตารางที่ 4.1 ค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของการประมาณค่าพารามิเตอร์ก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) (ต่อ)

Con	P	T	PEST			
			M1	M2	M3	M4
			Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
5	CLADIFF(γ_2)	0.1	0.111(0.106)	0.161(0.107)	0.130(0.106)	0.169(0.106)
	3WAYINT(γ_{7i})	0.3	0.275(0.001)	0.309(0.002)	0.496(0.001)	0.508(0.001)
	ILADIFF(γ_1)	0.2	0.207(0.046)	0.257(0.047)	0.218(0.046)	0.257(0.047)
	ILDIF(γ_{5i})	-0.5	-0.479(0.098)	-0.497(0.098)	-0.589(0.099)	-0.595(0.099)
	IAWC($\theta_j^{(2)}$)	1.0	1.094 (0.044)	1.095 (0.044)	1.094 (0.045)	1.095 (0.044)
	CA($\theta_k^{(3)}$)	1.0	0.986(0.102)	0.986(0.101)	0.986(0.101)	0.987(0.101)
6	CLADIFF(γ_2)	0.1	0.089(0.106)	0.146(0.106)	0.111(0.105)	0.151(0.106)
	3WAYINT(γ_{7i})	0.3	0.314(0.001)	0.347(0.001)	0.503(0.001)	0.516 (0.001)
	ILADIFF(γ_1)	0.2	0.202(0.046)	0.256(0.045)	0.212(0.046)	0.256(0.045)
	ILDIF(γ_{5i})	-0.5	-0.502(0.097)	-0.520(0.097)	-0.597(0.098)	-0.603(0.098)
	IAWC($\theta_j^{(2)}$)	1.0	1.094(0.044)	1.095(0.044)	1.094(0.044)	1.095(0.044)
	CA($\theta_k^{(3)}$)	1.0	1.000(0.103)	0.998(0.103)	0.998(0.103)	1.001(0.104)
7	CLADIFF(γ_2)	0.1	0.103(0.107)	0.160(0.107)	0.123(0.108)	0.158(0.107)
	3WAYINT(γ_{7i})	0.3	0.292(0.001)	0.331(0.001)	0.496(0.001)	0.511(0.001)
	ILADIFF(γ_1)	0.2	0.198(0.046)	0.253(0.047)	0.208(0.046)	0.253(0.046)
	ILDIF(γ_{5i})	-0.5	-0.511(0.097)	-0.531(0.097)	-0.613 (0.098)	-0.621(0.099)
	IAWC($\theta_j^{(2)}$)	1.0	1.105 (0.045)	1.107 (0.045)	1.105 (0.045)	1.106 (0.045)
	CA($\theta_k^{(3)}$)	1.0	0.955(0.098)	0.954(0.099)	0.953(0.098)	0.953(0.099)
8	CLADIFF(γ_2)	0.1	0.077(0.106)	0.131(0.106)	0.100(0.106)	0.142(0.105)
	3WAYINT(γ_{7i})	0.3	0.290(0.001)	0.327(0.001)	0.495(0.001)	0.509(0.001)
	ILADIFF(γ_1)	0.2	0.195(0.046)	0.247(0.047)	0.205(0.046)	0.247(0.047)
	ILDIF(γ_{5i})	-0.5	-0.503(0.097)	-0.521(0.098)	-0.605(0.099)	-0.612(0.099)
	IAWC($\theta_j^{(2)}$)	1.0	1.095(0.045)	1.096(0.044)	1.094(0.044)	1.095(0.044)
	CA($\theta_k^{(3)}$)	1.0	1.013(0.102)	1.014(0.103)	1.013(0.102)	1.015 (0.103)

ตารางที่ 4.1 ค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของการประมาณค่าพารามิเตอร์ก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) (ต่อ)

Con	P	T	PEST			
			M1	M2	M3	M4
			Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
9	CLADIFF(γ_2)	0.1	0.097(0.106)	0.148(0.106)	0.116(0.107)	0.158(0.106)
	3WAYINT(γ_{7i})	0.3	0.329(0.002)	0.362(0.001)	0.505(0.001)	0.517(0.001)
	ILADIFF(γ_1)	0.2	0.207(0.046)	0.257(0.047)	0.215(0.046)	0.257(0.047)
	ILDIF(γ_{5i})	-0.5	-0.514(0.098)	-0.531(0.098)	-0.601(0.099)	-0.607(0.099)
	IAWC($\theta_j^{(2)}$)	1.0	1.103(0.043)	1.104(0.045)	1.103(0.044)	1.104(0.044)
	CA($\theta_k^{(3)}$)	1.0	1.009(0.104)	1.011(0.105)	1.010(0.104)	1.011(0.106)
10	CLADIFF(γ_2)	0.1	0.063(0.066)	0.118(0.066)	0.082(0.066)	0.128(0.067)
	3WAYINT(γ_{7i})	0.3	0.300(0.001)	0.338(0.001)	0.500(0.002)	0.517(0.001)
	ILADIFF(γ_1)	0.2	0.199(0.032)	0.253(0.033)	0.209(0.032)	0.253(0.033)
	ILDIF(γ_{5i})	-0.5	-0.508(0.068)	-0.528(0.068)	-0.608(0.070)	-0.617(0.070)
	IAWC($\theta_j^{(2)}$)	1.0	1.098(0.031)	1.099(0.032)	1.098(0.031)	1.099(0.031)
	CA($\theta_k^{(3)}$)	1.0	0.964(0.072)	0.963(0.071)	0.962(0.071)	0.968(0.071)
11	CLADIFF(γ_2)	0.1	0.065(0.067)	0.116(0.066)	0.084(0.066)	0.125(0.066)
	3WAYINT(γ_{7i})	0.3	0.294(0.001)	0.329(0.001)	0.502(0.001)	0.516(0.001)
	ILADIFF(γ_1)	0.2	0.203(0.032)	0.254(0.034)	0.212(0.033)	0.254(0.033)
	ILDIF(γ_{5i})	-0.5	-0.495(0.068)	-0.513(0.068)	-0.599(0.070)	-0.606(0.071)
	IAWC($\theta_j^{(2)}$)	1.0	1.102(0.031)	1.103(0.032)	1.102(0.033)	1.103(0.032)
	CA($\theta_k^{(3)}$)	1.0	0.973(0.071)	0.972(0.072)	0.973(0.072)	0.971(0.071)
12	CLADIFF(γ_2)	0.1	0.091(0.066)	0.139(0.066)	0.110(0.067)	0.145(0.067)
	3WAYINT(γ_{7i})	0.3	0.316(0.001)	0.349(0.001)	0.502(0.001)	0.515(0.001)
	ILADIFF(γ_1)	0.2	0.203(0.032)	0.251(0.033)	0.212(0.032)	0.250(0.033)
	ILDIF(γ_{5i})	-0.5	-0.501(0.069)	-0.518(0.069)	-0.594(0.070)	-0.601(0.071)
	IAWC($\theta_j^{(2)}$)	1.0	1.101(0.031)	1.102(0.031)	1.101(0.032)	1.101(0.031)
	CA($\theta_k^{(3)}$)	1.0	0.983(0.074)	0.983(0.073)	0.977(0.073)	0.982(0.074)

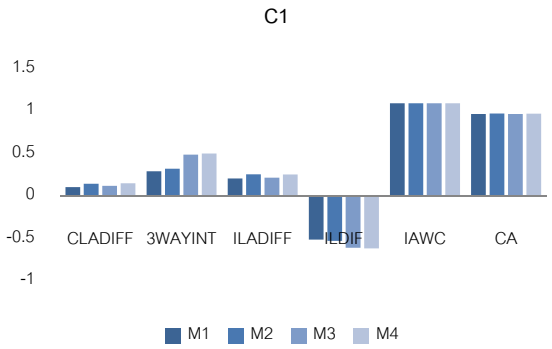
ตารางที่ 4.1 ค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของการประมาณค่าพารามิเตอร์ก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) (ต่อ)

			PEST			
		T	M1	M2	M3	M4
Con	P		Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
	CLADIFF(γ_2)	0.1	0.065(0.066)	0.116(0.067)	0.083(0.067)	0.125(0.066)
	3WAYINT(γ_{7i})	0.3	0.294(0.001)	0.329 (0.001)	0.502(0.001)	0.516(0.001)
13	ILADIFF(γ_1)	0.2	0.202(0.032)	0.253(0.033)	0.212(0.032)	0.253(0.033)
	ILDIF(γ_{5i})	-0.5	-0.495(0.068)	-0.512(0.068)	-0.598(0.070)	-0.606(0.070)
	IAWC($\theta_j^{(2)}$)	1.0	1.102(0.031)	1.103(0.031)	1.102(0.031)	1.102(0.031)
	CA($\theta_k^{(3)}$)	1.0	0.973(0.071)	0.972(0.072)	0.972(0.071)	0.971(0.072)
	CLADIFF(γ_2)	0.1	0.071(0.067)	0.126(0.066)	0.091(0.066)	0.132(0.066)
	3WAYINT(γ_{7i})	0.3	0.298(0.001)	0.333(0.001)	0.505(0.001)	0.521(0.001)
14	ILADIFF(γ_1)	0.2	0.195(0.032)	0.248(0.033)	0.204(0.032)	0.248(0.033)
	ILDIF(γ_{5i})	-0.5	-0.497(0.068)	-0.515(0.068)	-0.600(0.070)	-0.608(0.071)
	IAWC($\theta_j^{(2)}$)	1.0	1.105(0.031)	1.106(0.031)	1.105(0.032)	1.105(0.032)
	CA($\theta_k^{(3)}$)	1.0	0.958(0.071)	0.960(0.072)	0.958(0.070)	0.959(0.070)
	CLADIFF	0.1	0.074(0.066)	0.115(0.067)	0.087(0.066)	0.126(0.066)
	3WAYINT	0.3	0.311(0.001)	0.347(0.001)	0.509(0.001)	0.523(0.001)
15	ILADIFF	0.2	0.200(0.032)	0.250(0.033)	0.210(0.032)	0.250(0.033)
	ILDIF	-0.5	-0.502(0.068)	-0.520(0.068)	-0.600(0.069)	-0.607(0.070)
	IAWC	1.0	1.101(0.031)	1.102(0.031)	1.101(0.031)	1.102(0.031)
	CA	1.0	0.986(0.072)	0.984(0.073)	0.985(0.072)	0.987(0.072)
	CLADIFF(γ_2)	0.1	0.079(0.067)	0.131(0.066)	0.096(0.066)	0.135(0.066)
	3WAYINT(γ_{7i})	0.3	0.301(0.001)	0.337(0.001)	0.505(0.002)	0.518(0.001)
16	ILADIFF(γ_1)	0.2	0.201(0.032)	0.251(0.033)	0.210(0.032)	0.252(0.033)
	ILDIF(γ_{5i})	-0.5	-0.497(0.068)	-0.515 (0.068)	-0.599(0.070)	-0.605(0.070)
	IAWC($\theta_j^{(2)}$)	1.0	1.104(0.031)	1.105(0.032)	1.103(0.031)	1.104(0.031)
	CA($\theta_k^{(3)}$)	1.0	0.964(0.072)	0.967(0.073)	0.965(0.074)	0.970(0.073)

ตารางที่ 4.1 ค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของการประมาณค่าพารามิเตอร์ก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) (ต่อ)

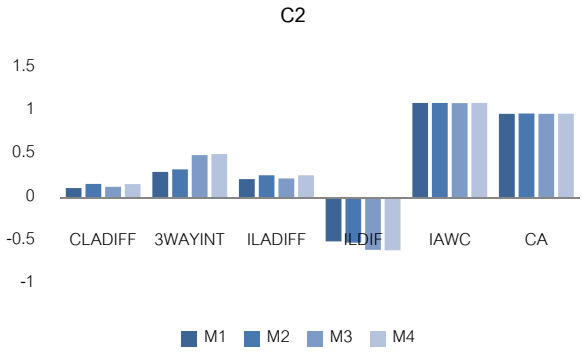
			PEST			
		T	M1	M2	M3	M4
Con	P		Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
	CLADIFF(γ_2)	0.1	0.096(0.067)	0.140(0.067)	0.113(0.066)	0.154(0.066)
	3WAYINT(γ_{7i})	0.3	0.310(0.001)	0.343(0.002)	0.513(0.001)	0.525(0.001)
17	ILADIFF(γ_1)	0.2	0.203(0.032)	0.253(0.033)	0.214(0.033)	0.254(0.033)
	ILDIF(γ_{5i})	-0.5	-0.510(0.069)	-0.526(0.069)	-0.611(0.071)	-0.617(0.071)
	IAWC($\theta_j^{(2)}$)	1.0	1.103(0.032)	1.104(0.031)	1.103(0.031)	1.104(0.031)
	CA($\theta_k^{(3)}$)	1.0	0.957(0.070)	0.955(0.070)	0.958(0.071)	0.962(0.071)
	CLADIFF(γ_2)	0.1	0.077(0.067)	0.125(0.066)	0.099(0.066)	0.140(0.066)
	3WAYINT(γ_{7i})	0.3	0.298(0.001)	0.330(0.001)	0.500(0.001)	0.514(0.001)
18	ILADIFF(γ_1)	0.2	0.208(0.032)	0.255(0.033)	0.218(0.032)	0.254(0.033)
	ILDIF(γ_{5i})	-0.5	-0.498(0.068)	-0.514(0.068)	-0.598(0.070)	-0.605(0.070)
	IAWC($\theta_j^{(2)}$)	1.0	1.104(0.031)	1.105(0.031)	1.104(0.030)	1.104(0.031)
	CA($\theta_k^{(3)}$)	1.0	0.953(0.071)	0.957(0.071)	0.955(0.071)	0.954(0.071)

จากตารางที่ 4.1 ผู้วิจัยขอสรุปผลที่ได้ทีละเงื่อนไข ดังนี้ เงื่อนไขที่ 1 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดล ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) แต่แตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อย แต่ได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกันทั้ง 4 โมเดล ดังแผนภาพ



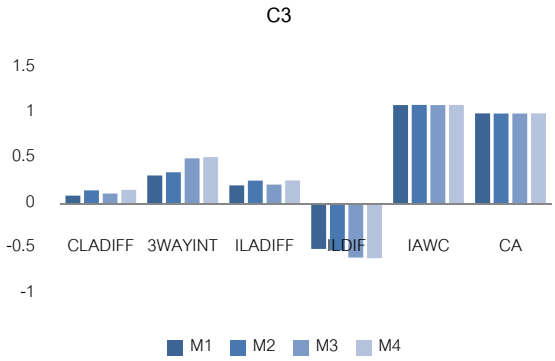
แผนภาพที่ 4.1 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 1 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 2 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกันทั้ง 4 โมเดล ดังแผนภาพ



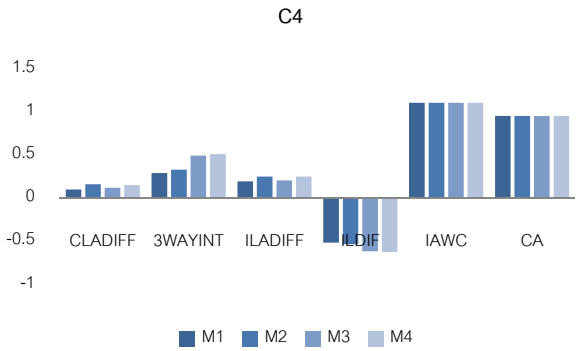
แผนภาพที่ 4.2 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 2 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 3 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกันทั้ง 4 โมเดล ดังแผนภาพ



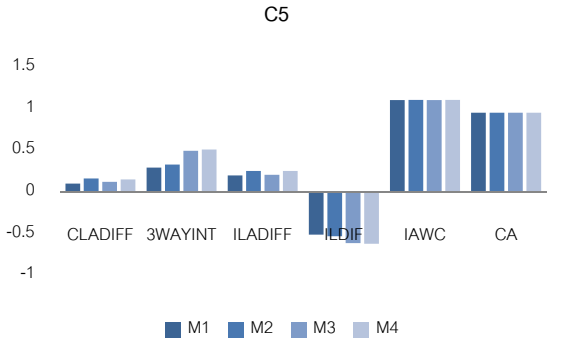
แผนภาพที่ 4.3 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 3 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 4 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกันทั้ง 4 โมเดล ดังแผนภาพ



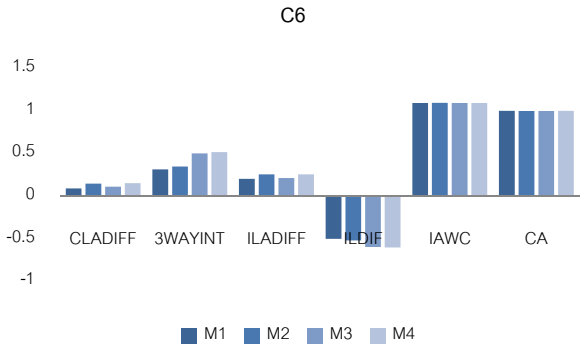
แผนภาพที่ 4.4 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 4 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 5 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) และได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกันทั้ง 4 โมเดล ดังแผนภาพ



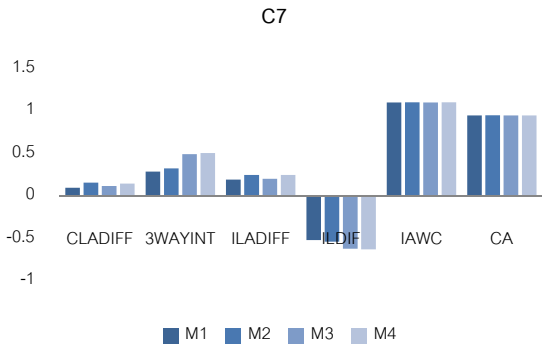
แผนภาพที่ 4.5 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 5 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 6 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



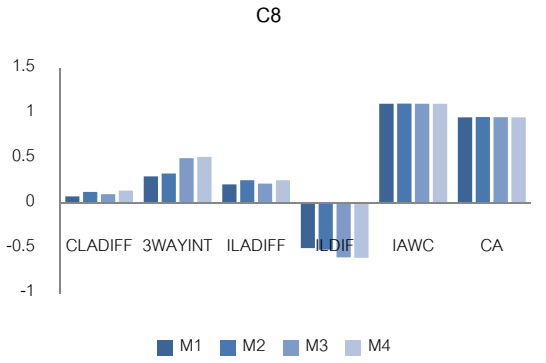
แผนภาพที่ 4.6 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 6 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 7 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



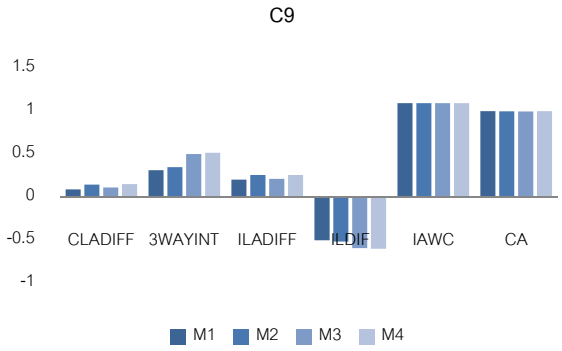
แผนภาพที่ 4.7 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 7 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 8 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่เท่ากันทั้ง 4 โมเดล ดังแผนภาพ



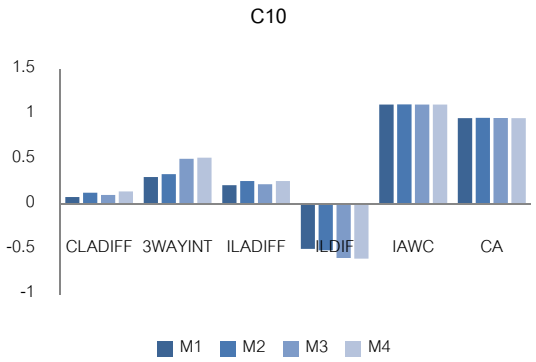
แผนภาพที่ 4.8 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 8 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 9 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 2 โมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อย และได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกันทั้ง 4 โมเดล ดังแผนภาพ



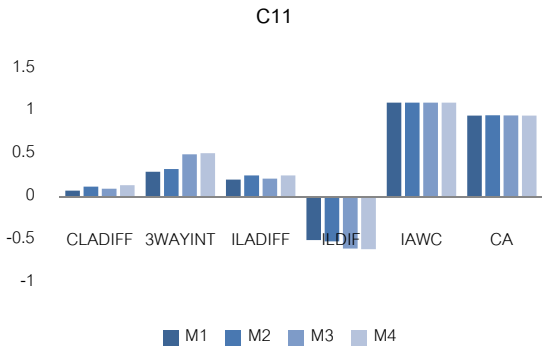
แผนภาพที่ 4.9 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 9 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 10 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) นั่นคือ ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) และได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



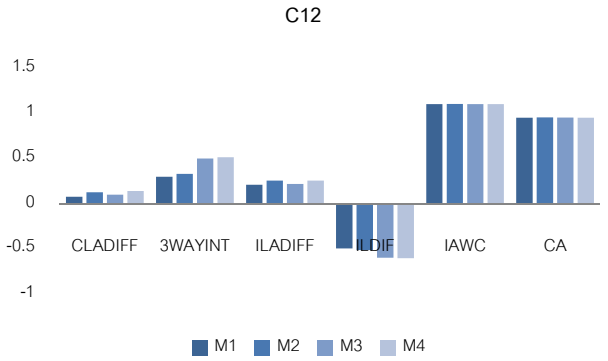
แผนภาพที่ 4.10 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 10 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 11 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) นั่นคือ ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) และได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



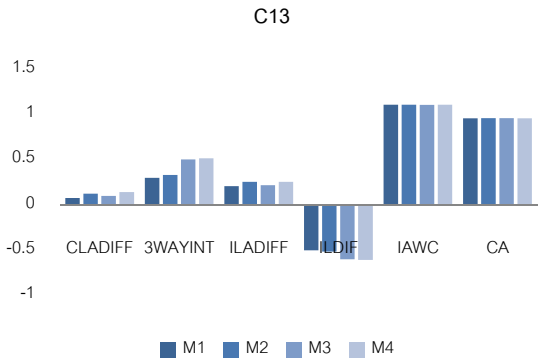
แผนภาพที่ 4.11 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 11 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 12 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



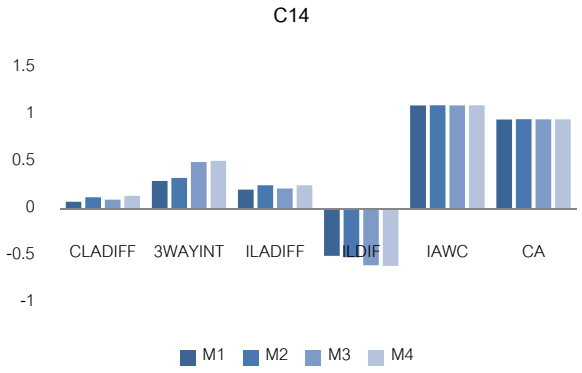
แผนภาพที่ 4.12 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 12 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 13 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) และได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



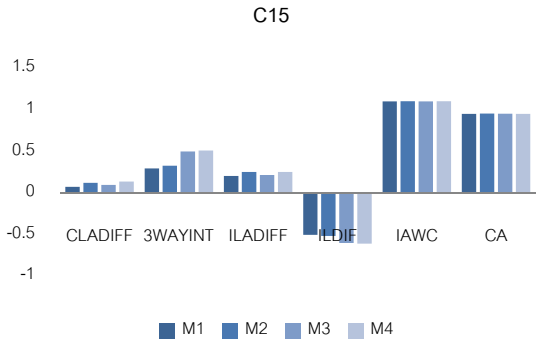
แผนภาพที่ 4.13 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 13 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 14 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



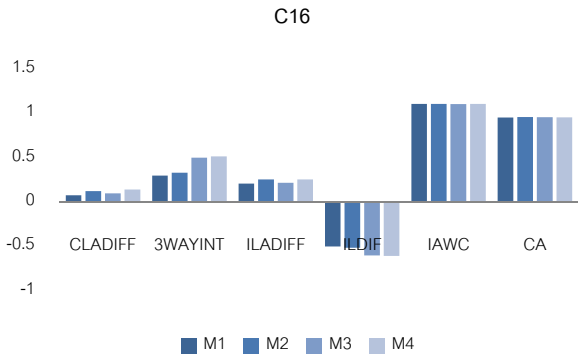
แผนภาพที่ 4.14 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 14 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 15 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



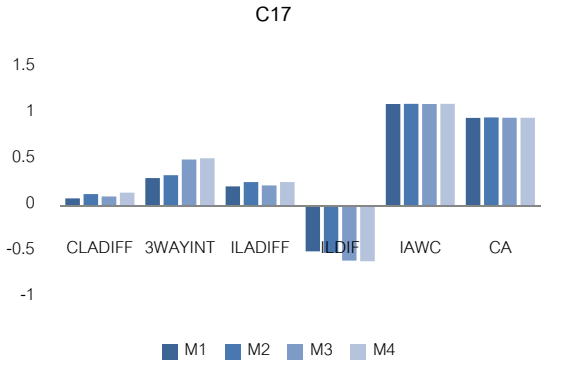
แผนภาพที่ 4.15 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 15 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 16 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



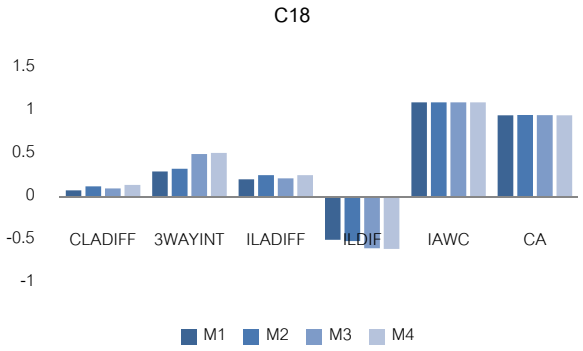
แผนภาพที่ 4.16 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 16 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 17 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



แผนภาพที่ 4.17 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 17 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 18 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) ซึ่งแตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกัน ดังแผนภาพ



แผนภาพที่ 4.18 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 18 ก่อนการใช้เกณฑ์การคัดสรรโมเดล

สรุปได้ว่า ทั้ง 18 เงื่อนไขได้ผลที่สอดคล้องกัน นั่นคือ ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 4 โมเดลใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกันในทั้ง 4 โมเดล ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 1 และโมเดลที่ 2 ได้ค่าประมาณที่ใกล้เคียงกับค่าจริง (true value) แต่แตกต่างจากค่าเฉลี่ย (mean) ที่ประมาณได้จากโมเดลที่ 3 และโมเดลที่ 4 ที่ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยและได้ค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ที่ใกล้เคียงกันทั้ง 4 โมเดล

การศึกษาในครั้งนี้มีวัตถุประสงค์ข้อแรกเพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม ซึ่งจากตารางที่ 4.1 จะสามารถตอบวัตถุประสงค์ข้อแรกได้ว่า โมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) นี้สามารถประมาณค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) ได้ใกล้เคียงกับค่าจริง (true value) ซึ่งอาจกล่าวได้ว่า โมเดลทั้ง 4 มีประสิทธิภาพในการประมาณค่าทั้ง 4 พารามิเตอร์ แต่ได้ค่าประมาณของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง (true value) ในทุกเงื่อนไข

การศึกษาในครั้งนี้มีวัตถุประสงค์ข้อที่ 2 เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ซึ่งจากผลการศึกษาทั้ง 18 เงื่อนไข สรุปได้ว่า โมเดลที่ 1 หรือโมเดลแบบเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) นั่นคือเป็นโมเดลที่มีประสิทธิภาพในการประมาณค่าพารามิเตอร์ทั้ง 6 พารามิเตอร์ได้ดีที่สุด ทั้งนี้ก็เพราะว่าได้ค่าประมาณพารามิเตอร์ทุกพารามิเตอร์ใกล้เคียงกับค่าจริง (true value) มากที่สุด รองลงมาคือโมเดลที่ 2 หรือโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) สำหรับโมเดลที่ 3 หรือ โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) และโมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) นั้นได้ค่าประมาณของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ที่แตกต่างจากค่าจริง (true value) ในทุกเงื่อนไข ส่วนพารามิเตอร์อื่นๆ มีค่าเฉลี่ย (mean) ไม่แตกต่างจากค่าจริง (true value) จึงอาจกล่าวได้ว่า การ drop ค่าพารามิเตอร์ γ_4 หรือพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ให้เป็นศูนย์และ γ_{6i} หรือพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) ให้เป็นศูนย์นั้น ส่งผลต่อการประมาณค่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF)

ตอนที่ 2 โมเดลที่ดีที่สุด (best model) **ของแต่ละการทำซ้ำเมื่อใช้เกณฑ์การคัดสรรโมเดล** (Model selection criteria) **ทั้ง 5 เกณฑ์**

ในตอนที่ 2 นี้ ผู้วิจัยนำเสนอผลการสกัดโมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำเมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ ได้ผลดังตารางที่ 4.2 และตารางที่ 4.3 เป็นค่าร้อยละของโมเดลที่ดีที่สุด (best model) ในแต่ละเกณฑ์การคัดสรรโมเดล ดังต่อไปนี้

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง

R	C1	C2	C3	C4	C5	C6
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1
2	2 4 4 4 1	4 4 4 4 3	1 4 3 0 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3
3	2 4 4 4 1	1 4 3 0 3	4 4 4 4 3	4 4 4 4 3	1 4 3 0 3	2 4 4 4 3
4	2 4 4 4 1	2 2 2 2 1	2 4 2 2 1	3 3 3 3 3	2 2 2 2 1	1 4 4 4 1
5	2 4 4 4 1	2 2 2 2 1	2 4 4 4 1	4 4 4 4 3	2 2 2 2 1	2 4 4 4 1
6	2 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1
7	4 4 4 4 3	3 4 4 4 3	2 4 4 4 1	1 4 2 0 1	3 4 4 4 3	4 4 4 4 3
8	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1
9	4 4 4 4 1	4 4 4 4 1	1 4 4 4 1	4 4 4 4 3	4 4 4 4 1	2 4 2 2 1
10	2 4 4 4 1	4 4 4 4 3	2 4 4 4 3	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3
11	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	1 4 3 0 1
12	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1
13	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3
14	4 4 4 4 3	2 4 4 4 3	2 4 4 4 1	2 4 4 4 1	2 4 4 4 3	4 4 4 4 3
15	1 4 4 4 1	1 4 4 4 3	3 4 4 4 3	4 4 4 4 3	1 4 4 4 3	2 4 4 4 1
16	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	2 2 2 2 1	2 4 4 4 1	1 3 3 3 1
17	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3	2 2 2 2 1	4 4 4 4 3	4 4 4 4 3
18	2 4 4 4 1	2 4 4 4 1	1 3 3 3 1	2 4 2 2 1	2 2 2 2 2	1 4 4 4 1
19	2 4 4 4 1	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	3 4 4 4 3
20	4 4 4 4 3	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1
21	2 4 4 4 1	2 4 4 4 1	2 2 2 2 1	2 4 4 4 1	4 4 4 4 3	2 2 2 2 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C1	C2	C3	C4	C5	C6
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
22	4 4 4 4 3	1 4 2 0 1	1 4 2 0 1	1 4 4 4 1	1 4 2 0 1	4 4 4 4 4
23	2 4 4 4 1	4 4 4 4 3	4 4 4 4 1	1 4 2 0 1	2 4 4 4 1	4 4 4 4 3
24	2 2 2 2 1	2 4 4 4 1	3 4 4 4 3	1 4 1 1 1	3 4 4 4 3	1 4 4 4 1
25	1 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 1 1 1	3 4 4 4 3	2 4 4 4 1
26	3 4 4 4 3	2 2 2 2 2	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1	4 4 4 4 3
27	1 4 4 4 1	1 2 2 2 1	2 4 4 4 1	1 4 4 4 1	1 2 2 2 1	2 4 4 4 1
28	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3	1 4 2 0 1
29	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	4 4 4 4 1	4 4 4 4 3	4 4 4 4 1
30	3 4 4 4 3	1 4 4 4 1	4 4 4 4 4	1 4 4 4 1	2 4 2 2 1	3 4 4 4 3
31	4 4 4 4 3	2 4 2 2 1	2 4 2 2 1	4 4 4 4 4	3 4 4 4 3	2 4 2 2 1
32	3 4 4 4 1	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3
33	2 2 2 2 1	1 3 1 1 1	3 4 4 4 3	2 4 2 2 1	2 4 2 2 1	4 4 4 4 3
34	4 4 4 4 3	3 3 3 3 3	4 4 4 4 3	2 2 2 2 1	3 3 3 3 3	3 3 3 3 3
35	2 4 2 2 1	2 4 2 2 1	4 4 4 4 3	2 4 2 2 3	1 3 1 1 1	4 4 4 4 3
36	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 3 0 3	3 4 4 4 3	1 4 4 4 1
37	4 4 4 4 3	3 4 4 4 3	2 4 4 4 1	2 4 2 2 1	2 4 2 2 1	2 4 4 4 1
38	2 4 4 4 1	2 4 2 2 1	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1
39	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	1 4 4 4 1	2 4 2 2 1	3 4 4 4 3
40	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	3 4 3 3 3	2 4 2 2 1	2 4 4 4 1
41	2 2 2 2 1	1 4 2 0 1	3 3 3 3 3	4 4 4 4 3	2 2 2 2 1	4 4 4 4 3
42	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3
43	4 4 4 4 3	2 4 2 2 1	3 4 4 4 3	2 4 2 2 1	2 4 2 2 1	3 4 4 4 3
44	4 4 4 4 3	1 2 2 2 1	1 4 4 4 1	3 3 3 3 3	1 4 2 0 1	1 2 2 2 1
45	2 4 2 2 1	2 2 2 2 1	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	3 4 4 4 3
46	4 4 4 4 3	4 4 4 4 1	4 4 4 4 1	3 4 4 4 3	3 4 4 4 3	3 4 3 3 3
47	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	3 4 4 4 3	2 4 2 2 1
48	3 4 3 3 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 1	2 4 2 2 1	4 4 4 4 3

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C1	C2	C3	C4	C5	C6
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
49	2 4 2 2 1	1 4 2 0 1	2 4 2 2 1	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3
50	2 4 4 4 1	4 4 4 4 3	3 4 3 3 3	4 4 4 4 3	1 4 2 0 1	4 4 4 4 1
51	2 4 4 4 1	3 4 4 4 3	3 4 4 4 3	3 4 4 4 3	2 4 4 4 1	1 4 4 4 1
52	4 4 4 4 3	3 4 4 4 3	1 2 2 2 1	2 4 2 2 1	2 4 2 2 1	1 4 4 4 1
53	1 4 4 4 1	2 4 2 2 1	1 4 4 4 1	4 4 4 4 3	1 2 2 2 1	1 4 4 4 1
54	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 2 2 2 1	4 4 4 4 3
55	1 3 1 1 1	2 2 2 2 1	2 4 2 2 1	2 4 4 4 2	4 4 4 4 1	2 4 2 2 1
56	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3
57	1 4 1 1 1	2 4 2 2 1	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	3 4 4 4 3
58	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3
59	4 4 4 4 3	2 4 2 2 1	2 2 2 2 1	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1
60	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 2 2 2 1
61	2 4 4 4 1	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1	4 4 4 4 3	2 4 2 2 1
62	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	1 4 3 0 1	2 4 4 4 1
63	3 3 3 3 1	4 4 4 4 3	3 4 4 4 3	1 4 4 4 1	4 4 4 4 3	3 4 4 4 3
64	1 4 4 4 1	1 2 1 1 1	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3	3 4 4 4 3
65	4 4 4 4 4	2 4 4 4 1	2 2 2 2 1	1 4 4 4 1	2 4 2 2 1	2 2 2 2 1
66	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1
67	4 4 4 4 3	1 4 2 0 1	2 2 2 2 1	4 4 4 4 3	2 4 4 4 1	2 2 2 2 1
68	1 4 3 0 1	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3	1 4 2 0 1	2 4 2 2 1
69	2 4 4 4 1	4 4 4 4 3	2 4 2 2 1	1 4 3 0 1	4 4 4 4 3	2 4 2 2 1
70	4 4 4 4 3	4 4 4 4 3	4 4 4 4 1	4 4 4 4 1	2 4 4 4 1	4 4 4 4 1
71	2 4 2 2 1	4 4 4 4 3	3 4 4 4 3	4 4 4 4 1	1 2 1 1 1	3 4 4 4 3
72	4 4 4 4 3	1 4 3 0 1	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1
73	2 4 2 2 1	4 4 4 4 3	3 4 3 3 3	1 2 1 1 1	1 4 4 4 1	3 4 3 3 3
74	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3
75	1 4 4 4 1	4 4 4 4 3	3 4 4 4 3	2 4 4 4 1	2 4 4 4 1	3 4 4 4 3

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C1	C2	C3	C4	C5	C6
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
76	2 4 2 2 1	1 4 4 4 1	1 4 1 1 1	4 4 4 4 3	1 4 4 4 1	1 4 1 1 1
77	2 2 2 2 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3
78	3 4 4 4 3	2 4 4 4 1	2 4 4 4 1	3 4 4 4 1	2 4 4 4 1	2 4 4 4 1
79	2 4 4 4 1	2 4 4 4 1	1 4 2 0 1	4 4 4 4 3	4 4 4 4 3	1 4 2 0 1
80	4 4 4 4 3	1 2 2 2 1	1 4 4 4 3	2 4 4 4 1	2 4 4 4 1	1 4 4 4 3
81	4 4 4 4 3	4 4 4 4 3	3 4 4 4 1	1 4 1 1 1	1 2 2 2 1	3 4 4 4 1
82	4 4 4 4 3	2 2 2 2 1	3 4 4 4 1	1 4 2 0 1	2 2 2 2 1	2 4 4 4 1
83	2 4 2 2 1	2 4 2 2 1	2 4 4 4 1	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3
84	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3
85	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1
86	2 2 2 2 2	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3
87	4 4 4 4 3	3 4 4 4 3	4 4 4 4 3	2 4 2 2 1	3 4 4 4 3	4 4 4 4 3
88	1 4 2 0 1	2 4 4 4 1	2 2 2 2 1	4 4 4 4 3	2 4 4 4 1	1 2 2 2 1
89	2 4 2 2 1	2 4 4 4 1	3 3 3 3 3	4 4 4 4 3	2 4 4 4 1	3 3 3 3 3
90	2 4 4 4 1	3 4 4 4 1	2 2 2 2 1	1 4 3 0 1	3 4 4 4 1	3 4 4 4 3
91	4 4 4 4 3	3 4 3 3 3	1 2 2 2 1	4 4 4 4 3	3 4 3 3 3	2 2 2 2 1
92	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 2 2 2 1	4 4 4 4 3	3 4 4 4 3
93	2 4 4 4 1	3 4 4 4 3	4 4 4 4 3	1 4 2 0 1	3 4 4 4 3	4 4 4 4 3
94	2 4 4 4 1	4 4 4 4 3	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 4 4 4 3
95	2 4 2 2 1	1 4 2 0 1	2 4 2 2 1	2 4 4 4 3	1 4 2 0 1	1 2 2 2 1
96	2 2 2 2 1	2 2 2 2 1	4 4 4 4 3	3 4 4 4 3	2 2 2 2 1	2 2 2 2 1
97	4 4 4 4 3	2 4 2 2 1	1 4 4 4 3	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3
98	4 4 4 4 3	4 4 4 4 1	1 2 2 2 1	2 4 4 4 1	4 4 4 4 1	4 4 4 4 3
99	4 4 4 4 3	2 4 4 4 1	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	3 4 4 4 3
100	3 3 3 3 3	4 4 4 4 3	1 4 2 0 1	1 4 4 4 1	4 4 4 4 3	1 3 3 3 1
101	4 4 4 4 3	4 4 4 4 3	3 3 3 3 3	4 4 4 4 3	4 4 4 4 3	3 4 3 3 3
102	1 4 4 4 1	3 4 4 4 3	1 4 4 4 1	3 4 3 3 3	3 4 4 4 3	3 4 3 3 3

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C1	C2	C3	C4	C5	C6
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
103	2 4 2 2 1	2 4 2 2 1	3 4 3 3 3	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1
104	3 4 4 4 3	4 4 4 4 3	3 4 3 3 3	2 4 2 2 1	3 4 4 4 3	3 3 3 3 3
105	3 4 4 4 3	4 4 4 4 3	1 3 3 3 1	4 4 4 4 3	1 2 2 2 1	1 4 2 0 1
106	2 4 4 4 1	4 4 4 4 1	3 4 4 4 3	2 4 4 4 1	2 4 2 2 1	2 4 4 4 1
107	2 4 2 2 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3
108	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1	1 4 1 1 1	4 4 4 4 3	2 4 4 4 1
109	2 4 4 4 1	1 2 2 2 1	2 4 4 4 1	2 4 4 4 1	4 4 4 4 1	1 4 4 4 1
110	2 2 2 2 1	3 4 4 4 3	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1	1 4 2 0 1
111	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1	4 4 4 4 3
112	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	2 4 4 4 1	2 4 2 2 1	3 4 4 4 1
113	2 2 2 2 1	4 4 4 4 3	3 4 4 4 3	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1
114	4 4 4 4 3	3 4 4 4 3	3 4 3 3 3	1 4 2 0 1	3 4 4 4 3	3 4 3 3 3
115	1 4 3 0 1	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3
116	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 1	4 4 4 4 3	2 4 4 4 1
117	4 4 4 4 3	2 2 2 2 1	4 4 4 4 3	4 4 4 4 3	2 2 2 2 1	4 4 4 4 3
118	2 4 2 2 1	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3
119	2 4 2 2 1	1 4 1 1 1	4 4 4 4 3	1 4 4 4 1	1 4 1 1 1	4 4 4 4 3
120	2 2 2 2 1	2 4 2 2 1	3 4 3 3 3	3 3 3 3 3	2 4 2 2 1	4 4 4 4 3
121	2 4 4 4 1	2 4 4 4 1	3 4 3 3 3	3 4 3 3 3	2 4 4 4 1	2 4 4 4 1
122	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3
123	1 2 2 2 1	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	2 4 4 4 3
124	4 4 4 4 3	2 2 2 2 1	4 4 4 4 3	2 2 2 2 1	3 4 4 4 3	3 4 3 3 3
125	4 4 4 4 3	2 4 2 2 1	3 4 3 3 3	1 4 4 4 1	2 4 4 4 1	1 2 2 2 1
126	1 4 1 1 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 3 0 1	4 4 4 4 3
127	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3	2 2 2 2 1	1 1 1 1 1	4 4 4 4 3
128	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 3 0 1	2 2 2 2 1
129	4 4 4 4 3	1 4 3 0 1	2 4 4 4 1	4 4 4 4 1	2 4 4 4 1	2 4 2 2 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C1	C2	C3	C4	C5	C6
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
130	3 4 3 3 3	1 1 1 1 1	3 4 4 4 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3
131	2 4 4 4 1	1 4 3 0 1	4 4 4 4 3	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3
132	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	1 4 1 1 1	2 2 2 2 1	4 4 4 4 3
133	4 4 4 4 3	3 4 4 4 3	2 4 4 4 3	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3
134	4 4 4 4 3	4 4 4 4 3	1 4 2 0 1	1 2 1 1 1	4 4 4 4 3	3 4 4 4 3
135	2 4 2 2 1	2 4 2 2 1	1 4 4 4 1	4 4 4 4 3	2 4 2 2 1	3 4 3 3 3
136	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	3 4 3 3 3
137	2 4 4 4 1	2 4 4 4 1	2 4 2 2 1	1 3 3 3 1	3 4 4 4 3	4 4 4 4 3
138	3 4 3 3 3	4 4 4 4 3	2 2 2 2 1	1 2 2 2 1	4 4 4 4 3	2 4 4 4 1
139	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 3 0 1	2 4 2 2 1	4 4 4 4 3
140	4 4 4 4 3	2 4 4 4 1	1 4 3 0 1	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3
141	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1
142	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1
143	3 4 4 4 3	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3
144	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	4 4 4 4 1	4 4 4 4 3	4 4 4 4 3
145	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	1 4 4 4 1	1 4 4 4 1	3 4 4 4 3
146	4 4 4 4 3	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1
147	3 4 3 3 3	2 4 2 2 1	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 1
148	2 4 4 4 1	4 4 4 4 3	4 4 4 4 1	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3
149	2 4 2 2 1	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	3 4 4 4 3
150	1 4 4 4 1	1 4 4 4 3	3 4 4 4 3	3 4 3 3 3	1 4 4 4 3	4 4 4 4 1
151	2 4 2 2 1	4 4 4 4 3	4 4 4 4 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3
152	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	2 4 4 4 1	1 4 1 1 1	2 4 4 4 1
153	4 4 4 4 3	1 4 3 0 1	2 4 2 2 1	4 4 4 4 1	2 4 4 4 1	4 4 4 4 3
154	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 3 0 1	2 4 4 4 1	2 4 2 2 1
155	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3
156	4 4 4 4 3	3 4 3 3 3	3 4 4 4 3	1 4 4 4 1	1 3 3 3 1	2 4 2 2 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C1	C2	C3	C4	C5	C6
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
157	1 3 3 3 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1
158	4 4 4 4 4	4 4 4 4 3	4 4 4 4 3	4 4 4 4 1	1 4 2 0 1	2 4 2 2 1
159	2 4 4 4 1	1 4 2 0 1	4 4 4 4 3	3 4 4 4 3	4 4 4 4 3	4 4 4 4 3
160	1 2 1 1 1	4 4 4 4 3	2 4 4 4 1	1 4 3 0 1	2 4 2 2 1	3 4 4 4 3
161	2 4 4 4 1	1 3 3 3 1	1 2 2 2 1	2 4 4 4 1	1 4 3 0 1	4 4 4 4 3
162	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3	2 2 2 2 1	4 4 4 4 3	2 4 4 4 1
163	2 4 2 2 1	2 4 4 4 1	3 4 4 4 3	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3
164	4 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 2 2 2 1	3 4 3 3 3	4 4 4 4 3
165	2 4 4 4 1	1 4 1 1 1	1 4 2 0 1	2 4 4 4 3	2 4 4 4 1	2 4 4 4 1
166	2 4 2 2 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	2 4 4 4 3
167	1 4 4 4 1	2 4 2 2 1	2 4 4 4 1	2 4 4 4 1	2 4 2 2 1	1 4 4 4 1
168	2 2 2 2 1	4 4 4 4 3	2 4 4 4 1	3 4 4 4 3	4 4 4 4 3	4 4 4 4 1
169	1 4 4 4 1	4 4 4 4 3	1 4 1 1 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1
170	2 4 4 4 1	2 4 4 4 2	1 4 4 4 1	3 4 3 3 3	4 4 4 4 3	4 4 4 4 1
171	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1
172	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1
173	2 2 2 2 3	2 4 2 2 1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	2 2 2 2 1
174	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 2 2 1	2 4 2 2 1	3 3 3 3 3
175	3 4 3 3 3	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1
176	1 4 3 0 1	4 4 4 4 3	3 4 4 4 3	2 4 4 4 1	2 4 4 4 2	1 4 4 4 1
177	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	3 3 3 3 1	2 4 2 2 1	4 4 4 4 3
178	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3	1 4 1 1 1	2 4 4 4 1	1 4 2 0 1
179	4 4 4 4 3	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3
180	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3	3 4 4 4 3
181	4 4 4 4 3	2 4 4 4 1	4 4 4 4 1	2 2 2 2 1	2 4 4 4 1	3 4 4 4 3
182	4 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C1	C2	C3	C4	C5	C6
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
183	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1	3 4 4 4 3	4 4 4 4 3
184	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3
185	2 4 4 4 1	2 4 4 4 1	1 4 3 0 1	4 4 4 4 3	2 4 4 4 1	1 4 3 0 1
186	4 4 4 4 1	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3
187	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	1 4 3 0 1
188	2 4 2 2 1	1 4 1 1 1	4 4 4 4 3	2 4 4 4 1	1 4 1 1 1	4 4 4 4 3
189	4 4 4 4 3	2 4 2 2 1	3 4 4 4 3	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3
190	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 1	4 4 4 4 3	2 4 4 4 1
191	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3
192	4 4 4 4 3	3 4 4 4 3	3 4 4 4 3	1 1 1 1 1	2 4 4 4 1	4 4 4 4 3
193	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3	3 4 3 3 3	1 4 4 4 1
194	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1	1 4 1 1 1
195	1 4 2 0 1	4 4 4 4 3	2 4 4 4 3	1 4 2 0 1	4 4 4 4 3	2 4 4 4 1
196	1 4 2 0 1	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3
197	4 4 4 4 3	4 4 4 4 3	4 4 4 4 1	2 4 2 2 1	4 4 4 4 3	2 4 4 4 1
198	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	3 4 4 4 3
199	3 4 4 4 3	3 4 3 3 3	3 3 3 3 3	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3
200	1 2 1 1 1	2 4 4 4 1	2 2 2 2 1	4 4 4 4 3	3 4 4 4 3	4 4 4 4 3

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C7	C8	C9	C10	C11	C12
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	2 4 4 4 3	3 4 4 4 3	1 4 2 0 1
2	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	3 4 4 4 3	2 4 4 4 1
3	4 4 4 4 3	3 4 3 3 3	2 2 2 2 1	1 4 1 1 1	3 4 4 4 3	1 4 4 4 1
4	3 3 3 3 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 1	1 2 2 2 1	4 4 4 4 3
5	4 4 4 4 3	1 4 4 4 1	3 4 4 4 3	3 4 4 4 3	3 4 4 4 1	4 4 4 4 1
6	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 2 2 2 1	3 4 4 4 3	4 4 4 4 1
7	1 4 2 0 1	4 4 4 4 3	3 4 3 3 3	2 2 2 2 1	4 4 4 4 3	4 4 4 4 4
8	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 2 0 1	2 4 4 4 1	2 4 4 4 1
9	4 4 4 4 3	3 4 4 4 3	2 4 2 2 1	1 4 4 4 1	4 4 4 4 3	2 4 4 4 1
10	3 4 4 4 3	2 4 2 2 1	4 4 4 4 3	4 4 4 4 1	1 4 4 4 1	1 4 1 1 1
11	4 4 4 4 3	3 4 4 4 1	4 4 4 4 3	2 2 2 2 1	1 4 2 0 1	4 4 4 4 3
12	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 1	2 2 2 2 1
13	1 4 4 4 1	2 4 4 4 1	3 4 4 4 3	1 4 3 0 1	2 2 2 2 1	2 4 4 4 1
14	2 4 4 4 1	4 4 4 4 4	3 4 4 4 3	2 2 2 2 1	4 4 4 4 3	2 2 2 2 1
15	4 4 4 4 3	1 4 4 4 3	2 4 2 2 1	1 4 2 0 1	2 4 4 4 1	4 4 4 4 3
16	2 2 2 2 1	2 4 4 4 1	2 4 4 4 1	1 4 1 1 1	4 4 4 4 1	1 4 1 1 1
17	2 2 2 2 1	2 2 2 2 1	4 4 4 4 3	4 4 4 4 1	2 4 4 4 1	2 4 2 2 1
18	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	1 4 2 0 1	1 4 2 0 1	2 4 4 4 1
19	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1
20	1 4 1 1 1	1 4 4 4 1	2 4 2 2 1	1 4 4 4 1	3 4 3 3 3	2 2 2 2 1
21	1 4 1 1 1	4 4 4 4 3	1 4 2 0 1	1 4 2 0 1	4 4 4 4 3	4 4 4 4 1
22	1 4 2 0 1	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1	1 4 4 4 1	1 4 4 4 1
23	1 4 4 4 1	1 3 3 3 1	1 2 2 2 1	4 4 4 4 3	1 2 1 1 1	3 4 4 4 3
24	2 4 4 4 1	1 2 2 2 1	4 4 4 4 3	2 4 4 4 1	1 4 2 0 1	3 4 4 4 1
25	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	2 4 2 2 1	2 2 2 2 1	1 4 1 1 1
26	4 4 4 4 3	4 4 4 4 3	1 2 1 1 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C7	C8	C9	C10	C11	C12
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
27	1 4 4 4 1	1 4 4 4 1	2 2 2 2 1	2 2 2 2 1	2 4 2 2 1	1 4 4 4 1
28	2 4 2 2 1	1 4 2 0 1	2 4 4 4 1	4 4 4 4 3	1 2 2 2 1	2 2 2 2 1
29	4 4 4 4 1	3 4 3 3 3	4 4 4 4 3	2 2 2 2 1	2 4 4 4 3	4 4 4 4 3
30	1 4 4 4 1	1 4 2 0 1	2 2 2 2 1	1 4 3 0 1	1 2 2 2 1	1 4 1 1 1
31	4 4 4 4 4	4 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1
32	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1	1 4 1 1 1	2 2 2 2 1	1 4 1 1 1
33	2 4 2 2 1	3 4 4 4 3	3 4 4 4 3	3 4 4 4 1	1 4 1 1 1	1 4 2 0 1
34	1 4 3 0 3	4 4 4 4 3	4 4 4 4 3	3 4 3 3 3	1 4 1 1 1	2 4 2 2 1
35	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1
36	2 2 2 2 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 4 2 0 1	1 4 4 4 1
37	2 4 2 2 1	2 4 4 4 1	3 4 4 4 3	1 4 4 4 1	1 4 4 4 1	1 4 2 0 1
38	4 4 4 4 3	3 4 4 4 3	4 4 4 4 3	2 4 2 2 1	3 4 4 4 3	4 4 4 4 1
39	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 1	1 4 4 4 1	4 4 4 4 3
40	3 4 3 3 3	2 4 2 2 1	4 4 4 4 3	1 4 4 4 1	1 4 4 4 3	1 4 4 4 1
41	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	2 4 2 2 1	2 2 2 2 1	2 4 4 4 1
42	3 4 4 4 3	3 4 4 4 3	2 4 4 4 1	1 4 4 4 1	1 4 4 4 1	2 2 2 2 1
43	4 4 4 4 3	3 4 3 3 3	2 4 4 4 1	4 4 4 4 3	1 2 2 2 1	3 4 4 4 3
44	3 3 3 3 3	3 4 4 4 3	4 4 4 4 3	1 4 4 4 1	1 2 2 2 1	4 4 4 4 1
45	2 4 2 2 1	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	2 4 2 2 1	1 4 1 1 1
46	1 4 4 4 1	1 4 1 1 1	4 4 4 4 3	1 4 3 0 1	4 4 4 4 3	1 4 2 0 1
47	4 4 4 4 3	1 2 1 1 1	3 3 3 3 3	1 4 4 4 1	2 4 4 4 1	2 2 2 2 1
48	4 4 4 4 1	3 4 4 4 3	1 4 3 0 1	1 4 2 0 1	1 2 2 2 1	2 4 2 2 1
49	2 4 4 4 2	3 3 3 3 3	1 4 4 4 1	1 4 4 4 1	1 4 4 4 1	1 4 3 0 1
50	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1
51	4 4 4 4 3	2 4 4 4 1	3 4 3 3 3	1 4 1 1 1	1 3 3 3 1	4 4 4 4 3
52	4 4 4 4 3	3 4 4 4 1	1 4 3 0 1	2 4 2 2 1	3 4 3 3 3	2 4 2 2 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C7	C8	C9	C10	C11	C12
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
53	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	1 3 1 1 1	1 2 1 1 1
54	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	2 2 2 2 1	1 4 2 0 1	1 2 2 2 1
55	2 4 2 2 1	2 2 2 2 1	1 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1
56	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	1 4 4 4 1	4 4 4 4 3
57	4 4 4 4 3	1 4 1 1 1	1 2 2 2 1	2 4 2 2 1	4 4 4 4 3	1 4 2 0 1
58	3 4 4 4 3	2 4 4 4 1	3 3 3 3 3	4 4 4 4 3	1 4 4 4 1	3 3 3 3 3
59	2 4 4 4 1	2 4 2 2 1	2 2 2 2 1	3 4 3 3 1	1 4 1 1 1	1 4 1 1 1
60	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 3 0 1	1 4 4 4 1
61	4 4 4 4 1	2 4 4 4 1	4 4 4 4 1	4 4 4 4 3	2 4 4 4 2	2 4 2 2 1
62	1 4 3 0 1	1 4 2 0 1	3 4 3 3 3	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1
63	4 4 4 4 3	2 4 4 4 1	1 4 4 4 3	4 4 4 4 3	3 4 4 4 3	4 4 4 4 1
64	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 2 0 1	2 4 4 4 1	3 4 4 4 3
65	4 4 4 4 3	4 4 4 4 1	1 4 1 1 1	2 4 2 2 1	1 4 4 4 1	2 4 2 2 1
66	1 4 4 4 1	3 4 4 4 3	3 4 3 3 3	1 4 3 0 1	4 4 4 4 3	2 4 2 2 1
67	4 4 4 4 3	4 4 4 4 3	4 4 4 4 1	1 4 4 4 1	1 4 3 0 3	2 4 2 2 1
68	1 4 4 4 1	4 4 4 4 3	1 4 4 4 1	1 4 3 0 1	2 4 2 2 1	2 4 4 4 1
69	4 4 4 4 3	4 4 4 4 3	3 4 4 4 3	1 4 2 0 1	2 4 4 4 1	2 4 2 2 1
70	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	3 4 4 4 3	2 4 4 4 1
71	4 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 2 2 2 1	3 4 3 3 3	4 4 4 4 3
72	1 4 2 0 1	3 4 4 4 3	3 4 4 4 3	3 3 3 3 3	2 4 4 4 1	3 4 4 4 3
73	2 4 4 4 1	1 4 2 0 1	4 4 4 4 3	1 4 1 1 1	2 4 4 4 1	3 3 3 3 3
74	2 4 4 4 1	1 4 4 4 1	3 4 4 4 3	2 4 2 2 1	3 4 3 3 3	2 4 2 2 1
75	4 4 4 4 3	2 4 2 2 1	1 4 2 0 1	1 4 2 0 1	2 4 4 4 1	4 4 4 4 3
76	1 4 1 1 1	4 4 4 4 3	2 4 4 4 3	1 4 4 4 1	2 4 4 4 1	2 4 4 4 1
77	2 4 4 4 1	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	3 4 4 4 3	3 4 4 4 1
78	4 4 4 4 3	1 3 1 1 1	1 4 4 4 1	1 4 1 1 1	1 2 2 2 1	1 4 4 4 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C7	C8	C9	C10	C11	C12
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
79	3 4 4 4 1	1 2 2 2 1	1 2 2 2 1	2 4 4 4 1	3 4 4 4 3	1 4 4 4 1
80	2 4 4 4 1	2 4 4 4 1	2 2 2 2 1	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3
81	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3
82	2 4 4 4 1	2 4 4 4 1	1 4 3 0 1	1 4 2 0 1	4 4 4 4 3	4 4 4 4 3
83	4 4 4 4 3	1 4 2 0 1	3 4 3 3 3	1 4 3 0 1	2 4 4 4 1	4 4 4 4 3
84	1 2 1 1 1	1 4 4 4 1	3 4 3 3 3	1 4 2 0 1	1 4 2 0 1	1 4 4 4 1
85	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 2 2 2	1 2 2 2 1	2 4 2 2 1
86	4 4 4 4 3	4 4 4 4 3	4 4 4 4 1	1 4 4 4 3	1 4 4 4 1	2 4 4 4 2
87	2 4 2 2 1	1 4 4 4 1	2 2 2 2 1	1 4 4 4 1	2 4 2 2 1	1 4 4 4 1
88	4 4 4 4 3	3 4 4 4 3	3 4 4 4 1	1 4 3 0 3	2 2 2 2 1	1 4 4 4 1
89	2 4 4 4 1	2 4 2 2 1	1 4 3 0 1	4 4 4 4 3	1 4 4 4 1	3 4 4 4 3
90	4 4 4 4 3	1 4 3 0 1	4 4 4 4 3	2 2 2 2 1	1 4 1 1 1	1 4 2 0 1
91	3 4 4 4 3	2 4 4 4 1	3 4 4 4 3	2 2 2 2 1	1 4 2 0 1	2 4 4 4 1
92	1 4 2 0 1	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 1	1 4 2 0 1
93	2 2 2 2 1	1 4 3 0 1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	1 4 1 1 1
94	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	1 4 4 4 1	1 4 4 4 1	1 4 4 4 3
95	1 4 3 0 1	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1
96	4 4 4 4 3	3 4 4 4 3	3 4 3 3 3	2 4 4 4 1	2 4 2 2 1	1 4 4 4 1
97	4 4 4 4 3	4 4 4 4 1	1 4 4 4 1	3 4 4 4 1	1 4 4 4 1	1 4 4 4 1
98	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3
99	2 4 2 2 1	2 2 2 2 1	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1
100	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	3 4 3 3 3	3 4 4 4 3	2 4 2 2 1
101	2 4 4 4 1	1 4 2 0 1	2 4 2 2 1	2 2 2 2 1	1 2 2 2 1	2 4 4 4 1
102	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 2 2 1	1 4 4 4 1	1 4 4 4 1
103	2 4 2 2 1	3 4 4 4 3	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	3 4 4 4 3
104	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 2 2 1	1 2 1 1 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C7	C8	C9	C10	C11	C12
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
105	3 4 3 3 3	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	3 4 3 3 3
106	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 2 1 1 1	2 4 4 4 1
107	1 4 4 4 1	2 2 2 2 1	4 4 4 4 1	1 2 1 1 1	1 4 1 1 1	1 4 1 1 1
108	1 4 2 0 1	2 2 2 2 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	1 4 4 4 1
109	3 3 3 3 3	4 4 4 4 1	1 4 2 0 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 1
110	4 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 1 1 1 1	3 4 4 4 3	4 4 4 4 3
111	4 4 4 4 3	3 4 4 4 3	4 4 4 4 3	1 4 4 4 1	1 2 2 2 1	1 4 3 0 1
112	3 4 4 4 3	2 4 2 2 1	2 4 4 4 1	1 4 3 0 1	1 4 3 0 1	2 2 2 2 1
113	2 4 4 4 1	4 4 4 4 3	3 4 4 4 3	2 4 4 4 1	1 4 4 4 1	1 2 1 1 1
114	2 2 2 2 1	1 4 4 4 1	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 3
115	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1
116	1 4 4 4 1	2 4 2 2 1	2 2 2 2 1	3 4 4 4 3	1 4 4 4 1	1 4 4 4 3
117	2 2 2 2 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3
118	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 2 2 2 1
119	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 1 1 1	4 4 4 4 3
120	3 4 3 3 3	4 4 4 4 3	2 4 4 4 1	3 4 4 4 3	3 4 4 4 3	1 2 1 1 1
121	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 2 0 1	2 4 4 4 1
122	2 4 4 4 1	2 2 2 2 1	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3
123	4 4 4 4 3	2 4 4 4 1	2 4 2 2 1	1 4 1 1 1	4 4 4 4 3	3 4 3 3 3
124	4 4 4 4 1	4 4 4 4 1	1 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 2 2 1
125	1 4 4 4 1	3 4 3 3 3	2 4 2 2 1	1 1 1 1 1	2 4 4 4 1	3 4 4 4 3
126	1 4 1 1 1	1 2 2 2 1	3 4 4 4 3	2 4 4 4 1	2 4 4 4 1	2 4 4 4 3
127	2 4 4 4 1	2 4 2 2 1	2 4 2 2 1	1 4 2 0 1	4 4 4 4 3	4 4 4 4 3
128	1 2 1 1 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3
129	4 4 4 4 3	4 4 4 4 3	3 3 3 3 3	1 2 2 2 1	4 4 4 4 3	2 4 4 4 1
130	1 4 4 4 1	4 4 4 4 3	3 4 4 4 3	1 4 4 4 1	3 4 4 4 1	1 4 4 4 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C7	C8	C9	C10	C11	C12
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
131	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	2 2 2 2 1	2 4 4 4 1	1 4 2 0 1
132	4 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3
133	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1	2 2 2 2 1	2 4 4 4 1	2 4 2 2 1
134	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1	1 1 1 1 1	1 2 1 1 1	2 4 4 4 1
135	4 4 4 4 3	2 2 2 2 1	4 4 4 4 3	2 2 2 2 1	3 4 4 4 3	4 4 4 4 3
136	2 4 4 4 3	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	2 4 2 2 1	1 4 4 4 1
137	2 4 4 4 1	2 4 2 2 1	2 2 2 2 1	4 4 4 4 3	1 4 1 1 1	2 4 4 4 1
138	2 4 2 2 1	3 3 3 3 3	3 4 3 3 3	2 4 4 4 1	1 1 1 1 1	4 4 4 4 3
139	2 4 4 4 1	3 4 4 4 3	4 4 4 4 3	1 2 2 2 1	4 4 4 4 3	2 4 4 4 1
140	3 4 4 4 3	2 4 4 4 1	1 4 4 4 1	2 2 2 2 1	1 4 4 4 1	1 2 2 2 1
141	1 4 1 1 1	4 4 4 4 1	2 4 4 4 1	4 4 4 4 1	2 2 2 2 1	1 4 4 4 1
142	3 4 3 3 3	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3
143	2 4 4 4 1	2 2 2 2 1	1 4 4 4 1	1 4 4 4 1	2 2 2 2 1	4 4 4 4 3
144	3 4 4 4 3	2 4 4 4 1	1 4 2 0 1	4 4 4 4 3	1 3 1 1 1	4 4 4 4 1
145	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 2 2 2 1	1 4 1 1 1	1 4 4 4 1
146	4 4 4 4 1	4 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	2 4 2 2 1
147	4 4 4 4 3	4 4 4 4 3	2 4 2 2 1	2 4 2 2 1	1 4 2 0 1	4 4 4 4 3
148	4 4 4 4 3	4 4 4 4 3	3 4 3 3 3	3 4 4 4 3	3 4 3 3 3	2 4 4 4 1
149	2 4 4 4 1	1 4 4 4 1	3 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1
150	3 4 4 4 3	4 4 4 4 3	4 4 4 4 1	1 4 4 4 1	3 4 4 4 1	2 4 4 4 1
151	1 4 3 0 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1	1 4 4 4 1
152	1 2 2 2 1	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	2 2 2 2 1	2 4 4 4 1
153	1 3 3 3 1	2 4 4 4 1	2 2 2 2 1	2 4 4 4 1	1 4 3 0 1	2 4 4 4 1
154	1 4 3 0 1	3 4 4 4 3	1 4 4 4 1	2 4 4 4 1	2 2 2 2 1	2 4 4 4 2
155	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 4 2 2 1	1 2 2 2 1
156	2 2 2 2 1	1 2 1 1 1	2 4 4 4 1	1 4 3 0 1	1 4 4 4 1	2 4 4 4 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C7	C8	C9	C10	C11	C12
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
157	2 4 4 4 3	1 4 1 1 1	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1
158	4 4 4 4 3	3 4 4 4 1	4 4 4 4 3	1 4 3 0 1	1 4 4 4 1	2 4 4 4 1
159	2 4 4 4 1	1 3 3 3 1	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1	1 4 4 4 3
160	3 4 4 4 3	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	1 2 2 2 1	1 4 4 4 1
161	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3	1 4 4 4 1	1 4 2 0 1	3 3 3 3 3
162	3 4 3 3 3	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3
163	2 2 2 2 1	4 4 4 4 3	1 2 2 2 1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1
164	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 2 2 2 1	2 4 4 4 1	2 4 2 2 1
165	1 4 3 0 1	3 4 4 4 3	4 4 4 4 3	1 4 1 1 1	1 1 1 1 1	2 4 2 2 1
166	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	4 4 4 4 1	1 1 1 1 1
167	4 4 4 4 1	4 4 4 4 3	2 2 2 2 1	3 4 4 4 1	2 4 4 4 2	4 4 4 4 1
168	4 4 4 4 3	3 4 3 3 3	1 4 4 4 1	2 2 2 2 1	2 4 2 2 1	1 4 4 4 1
169	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1
170	4 4 4 4 3	2 4 4 4 1	3 3 3 3 3	2 4 2 2 1	1 4 1 1 1	1 4 4 4 1
171	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 2 2 1	2 4 4 4 1
172	4 4 4 4 3	3 4 3 3 3	2 4 4 4 3	3 4 4 4 1	1 4 4 4 1	2 4 4 4 1
173	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 2 0 1	3 3 3 3 3
174	4 4 4 4 3	2 4 2 2 1	2 4 4 4 2	2 4 4 4 1	1 4 3 0 1	1 2 1 1 1
175	2 4 4 4 1	1 4 3 0 1	2 2 2 2 1	1 4 4 4 1	4 4 4 4 3	1 4 3 0 1
176	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 1	2 2 2 2 1	3 4 4 4 3
177	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1	1 4 3 0 1	2 4 4 4 1	2 4 4 4 1
178	2 4 4 4 1	3 4 3 3 1	3 4 4 4 1	3 4 4 4 3	1 4 3 0 1	2 4 4 4 1
179	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 4 2 0 1	1 4 4 4 1
180	2 2 2 2 1	2 4 4 4 1	2 4 4 4 1	1 4 1 1 1	2 4 4 4 1	2 4 4 4 1
181	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 1	3 4 4 4 3	1 1 1 1 1
182	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 2 2 2 1	2 4 4 4 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C7	C8	C9	C10	C11	C12
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
183	1 4 1 1 1	4 4 4 4 3	4 4 4 4 3	3 4 4 4 3	4 4 4 4 1	4 4 4 4 3
184	3 3 3 3 1	4 4 4 4 3	3 4 4 4 1	1 4 4 4 1	4 4 4 4 1	2 4 2 2 1
185	2 4 4 4 1	4 4 4 4 3	1 2 2 2 1	3 4 4 4 3	2 2 2 2 1	4 4 4 4 3
186	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1	4 4 4 4 3
187	2 4 2 2 1	1 3 1 1 1	4 4 4 4 3	1 4 2 0 1	1 4 2 0 1	4 4 4 4 3
188	1 4 4 4 1	4 4 4 4 3	22 4 4 4	1 4 4 4 1	1 4 4 4 1	2 2 2 2 1
189	4 4 4 4 1	3 4 3 3 1	4 4 4 4 3	3 4 4 4 3	2 2 2 2 1	1 1 1 1 1
190	4 4 4 4 3	1 4 4 4 1	3 4 4 4 3	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3
191	4 4 4 4 3	2 4 4 4 3	4 4 4 4 3	1 4 3 0 1	1 4 4 4 1	3 4 4 4 3
192	4 4 4 4 3	3 4 4 4 3	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1	1 4 4 4 1
193	4 4 4 4 3	2 4 4 4 1	3 4 3 3 3	1 4 2 0 1	4 4 4 4 3	2 4 2 2 1
194	2 4 2 2 1	2 4 2 2 1	3 4 4 4 3	3 3 3 3 3	2 4 4 4 1	1 4 4 4 1
195	2 4 4 4 1	1 4 4 4 1	4 4 4 4 4	1 2 2 2 1	4 4 4 4 1	1 4 1 1 1
196	1 4 2 0 1	1 2 1 1 1	4 4 4 4 1	2 4 2 2 1	1 2 2 2 1	4 4 4 4 3
197	1 4 4 4 1	2 4 4 4 1	2 2 2 2 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3
198	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1	2 2 2 2 1	2 4 2 2 1
199	1 1 1 1 1	3 4 4 4 3	1 4 3 0 1	1 4 1 1 1	2 2 2 2 1	1 4 4 4 1
200	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	1 4 4 4 1	2 4 4 4 1	1 4 1 1 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C13	C14	C15	C16	C17	C18
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
1	3 4 4 4 3	1 2 2 2 1	1 1 1 1 1	4 4 4 4 3	1 2 2 2 1	1 4 4 4 1
2	2 4 4 4 1	1 2 1 1 1	3 4 4 4 3	1 2 2 2 1	4 4 4 4 3	2 2 2 2 1
3	4 4 4 4 3	2 4 4 4 1	2 2 2 2 1	1 4 4 4 1	1 4 4 4 1	4 4 4 4 3
4	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	3 4 4 4 3	2 4 4 4 1
5	4 4 4 4 1	3 4 4 4 3	1 2 1 1 1	4 4 4 4 3	2 2 2 2 1	3 4 4 4 3
6	1 2 2 2 1	1 4 1 1 1	1 4 4 4 1	2 4 4 4 1	1 4 1 1 1	1 4 4 4 1
7	4 4 4 4 3	2 2 2 2 1	2 4 4 4 1	3 4 4 4 3	3 4 4 4 3	2 4 4 4 1
8	2 4 4 4 1	1 2 2 2 1	2 2 2 2 1	1 4 3 0 1	4 4 4 4 3	1 4 3 0 1
9	2 4 4 4 2	1 4 2 0 1	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	2 4 4 4 1
10	2 4 2 2 1	1 4 4 4 1	2 4 2 2 1	1 2 2 2 1	3 4 4 4 3	2 4 2 2 1
11	2 4 4 4 1	4 4 4 4 1	1 4 1 1 1	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1
12	1 4 4 4 1	1 2 2 2 1	1 4 3 0 1	4 4 4 4 3	2 4 4 4 1	1 2 2 2 1
13	1 4 1 1 1	2 4 4 4 1	2 2 2 2 1	3 4 4 4 3	1 1 1 1 1	4 4 4 4 3
14	2 4 2 2 1	1 4 4 4 1	4 4 4 4 1	2 4 2 2 1	4 4 4 4 1	2 4 4 4 1
15	1 4 3 0 1	1 4 2 0 1	2 4 4 4 1	1 4 4 4 1	4 4 4 4 1	1 4 4 4 1
16	4 4 4 4 3	3 4 3 3 3	2 4 4 4 1	1 2 1 1 1	4 4 4 4 3	2 2 2 2 1
17	2 2 2 2 1	4 4 4 4 4	2 2 2 2 1	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1
18	2 4 4 4 1	2 4 4 4 1	2 4 2 2 1	3 4 4 4 1	1 4 3 0 1	2 4 2 2 1
19	1 4 3 0 1	2 4 4 4 1	4 4 4 4 1	2 4 2 2 1	1 2 2 2 1	4 4 4 4 3
20	1 4 2 0 1	3 4 4 4 3	2 4 2 2 1	2 4 4 4 2	1 4 1 1 1	1 4 4 4 3
21	2 4 4 4 1	1 3 3 3 1	1 4 4 4 1	2 4 4 4 1	2 4 2 2 1	3 4 4 4 3
22	1 4 2 0 1	4 4 4 4 3	2 4 4 4 1	2 2 2 2 1	1 4 4 4 1	1 4 4 4 1
23	1 4 4 4 1	2 4 4 4 3	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1	3 4 4 4 3
24	2 2 2 2 1	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1	2 2 2 2 1
25	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 2 0 1	3 4 3 3 3
26	1 4 4 4 1	1 3 3 3 1	3 4 3 3 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C13	C14	C15	C16	C17	C18
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
27	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1	2 4 4 4 1
28	4 4 4 4 3	2 4 4 4 1	2 4 2 2 1	1 4 3 0 1	4 4 4 4 3	1 4 2 0 1
29	1 4 2 0 1	3 3 3 3 3	4 4 4 4 3	1 4 4 4 1	1 4 2 0 1	1 4 4 4 1
30	1 2 2 2 1	1 4 1 1 1	1 4 4 4 1	2 4 2 2 1	1 2 2 2 1	1 4 2 0 1
31	2 4 4 4 1	1 4 4 4 1	1 4 4 4 1	4 4 4 4 3	2 4 2 2 1	4 4 4 4 3
32	2 2 2 2 1	4 4 4 4 3	2 4 4 4 2	4 4 4 4 1	1 2 2 2 1	2 2 2 2 1
33	1 4 3 0 1	2 4 4 4 1	4 4 4 4 3	1 2 1 1 1	1 2 2 2 1	2 4 2 2 2
34	2 2 2 2 1	1 4 1 1 1	1 4 4 4 1	2 4 4 4 3	1 4 4 4 1	1 4 4 4 1
35	2 4 2 2 1	3 4 4 4 3	1 4 4 4 1	1 4 4 4 1	3 4 3 3 3	2 2 2 2 1
36	1 4 4 4 1	2 4 4 4 1	1 4 2 0 1	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1
37	4 4 4 4 3	1 3 3 3 1	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1
38	1 1 1 1 1	2 4 4 4 1	2 4 4 4 3	1 4 2 0 1	3 4 3 3 1	1 2 2 2 1
39	1 4 1 1 1	4 4 4 4 3	4 4 4 4 3	1 2 2 2 1	2 4 4 4 1	4 4 4 4 3
40	3 4 4 4 1	1 4 2 0 1	1 4 3 0 3	2 4 4 4 1	2 4 4 4 1	2 2 2 2 1
41	1 4 4 4 1	1 2 2 2 1	2 4 4 4 1	1 4 1 1 1	3 4 3 3 1	1 4 4 4 1
42	1 4 4 4 1	3 3 3 3 3	2 4 2 2 2	2 2 2 2 1	1 2 2 2 1	2 2 2 2 1
43	2 2 2 2 1	2 4 4 4 1	2 4 4 4 1	1 2 2 2 1	2 4 2 2 1	4 4 4 4 3
44	2 2 2 2 1	1 4 3 0 1	1 2 2 2 1	4 4 4 4 3	3 4 4 4 3	2 2 2 2 1
45	1 1 1 1 1	4 4 4 4 3	3 4 3 3 3	1 2 2 2 1	2 2 2 2 1	2 4 2 2 1
46	2 4 4 4 1	2 4 4 4 2	1 2 2 2 1	4 4 4 4 3	1 4 4 4 1	1 4 1 1 1
47	3 4 4 4 3	2 4 2 2 1	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3
48	4 4 4 4 1	2 2 2 2 1	2 2 2 2 1	1 4 2 0 1	2 4 4 4 1	1 4 4 4 1
49	1 4 3 0 1	1 2 2 2 1	1 4 4 4 1	2 4 2 2 1	2 2 2 2 1	2 2 2 2 1
50	1 4 2 0 1	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1	2 2 2 2 1	2 4 4 4 1
51	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3
52	2 2 2 2 1	2 4 4 4 3	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1	1 2 1 1 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C13	C14	C15	C16	C17	C18
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
53	4 4 4 4 1	1 4 1 1 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 3 3 3 1
54	4 4 4 4 1	4 4 4 4 3	1 4 3 0 1	2 4 4 4 1	1 4 4 4 1	1 3 3 3 1
55	3 4 4 4 3	3 4 3 3 3	4 4 4 4 3	2 4 2 2 1	2 4 2 2 1	2 4 4 4 1
56	2 4 2 2 1	2 4 2 2 1	4 4 4 4 3	4 4 4 4 1	4 4 4 4 3	3 3 3 3 3
57	4 4 4 4 3	2 4 4 4 1	1 4 2 0 1	1 4 4 4 1	2 4 2 2 1	2 4 4 4 1
58	2 2 2 2 1	2 4 4 4 1	1 4 4 4 1	3 4 4 4 3	2 4 2 2 1	1 4 4 4 1
59	1 4 2 0 1	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1	1 4 1 1 1	4 4 4 4 3
60	1 2 1 1 1	1 4 2 0 1	4 4 4 4 3	1 4 4 4 1	1 4 2 0 1	1 4 4 4 1
61	1 4 4 4 1	1 4 4 4 1	2 4 4 4 1	1 4 3 0 1	2 4 4 4 1	2 2 2 2 1
62	4 4 4 4 3	2 4 4 4 1	1 4 3 0 1	1 4 4 4 1	1 4 2 0 1	4 4 4 4 3
63	3 4 3 3 3	1 4 4 4 1	1 4 1 1 1	3 3 3 3 1	1 2 2 2 1	2 4 4 4 1
64	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	1 4 4 4 3	1 4 1 1 1
65	1 4 2 0 1	3 4 3 3 1	3 3 3 3 3	1 4 4 4 1	2 4 4 4 1	2 4 2 2 1
66	2 4 4 4 1	1 4 4 4 1	3 4 4 4 3	1 4 4 4 1	2 4 4 4 1	1 2 2 2 1
67	4 4 4 4 1	1 4 4 4 1	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	2 4 2 2 1
68	1 4 2 0 1	1 4 2 0 1	2 4 4 4 1	3 4 4 4 3	2 2 2 2 1	1 4 4 4 1
69	2 2 2 2 1	2 4 2 2 1	2 2 2 2 1	4 4 4 4 1	3 4 4 4 3	1 4 3 0 1
70	4 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 2 0 1	2 4 4 4 1	4 4 4 4 1
71	4 4 4 4 3	2 4 4 4 1	1 4 3 0 1	2 4 2 2 1	1 4 4 4 1	1 4 1 1 1
72	2 4 4 4 1	2 4 2 2 1	1 4 4 4 1	3 4 4 4 3	1 3 3 3 1	4 4 4 4 3
73	3 4 4 4 3	1 4 2 0 1	3 4 4 4 3	4 4 4 4 3	2 2 2 2 1	2 4 2 2 1
74	3 4 4 4 1	1 4 2 0 1	2 2 2 2 1	1 4 1 1 1	2 4 4 4 1	2 4 4 4 1
75	2 2 2 2 1	2 4 4 4 1	1 4 4 4 1	4 4 4 4 1	1 4 1 1 1	3 4 4 4 3
76	4 4 4 4 3	1 2 2 2 1	2 2 2 2 1	2 2 2 2 1	1 4 4 4 1	2 4 4 4 1
77	3 4 4 4 3	2 4 2 2 1	2 4 2 2 1	2 4 2 2 1	2 4 4 4 3	1 4 4 4 1
78	2 4 4 4 1	3 4 4 4 1	1 4 2 0 1	1 4 2 0 1	2 4 2 2 1	1 2 2 2 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C13	C14	C15	C16	C17	C18
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
79	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 2 2 2 1	4 4 4 4 1
80	1 4 4 4 1	4 4 4 4 3	1 4 4 4 1	1 2 2 2 1	4 4 4 4 1	4 4 4 4 3
81	1 2 2 2 1	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	2 4 4 4 1	1 2 2 2 1
82	1 4 4 4 1	2 4 4 4 1	2 4 4 4 1	1 4 1 1 1	4 4 4 4 3	1 4 4 4 1
83	4 4 4 4 3	3 3 3 3 3	1 4 3 0 1	1 4 4 4 1	2 4 4 4 1	2 4 2 2 1
84	1 4 1 1 1	1 4 1 1 1	2 2 2 2 2	1 2 2 2 1	2 4 4 4 1	1 4 4 4 1
85	3 4 4 4 3	2 4 4 4 1	1 2 2 2 1	4 4 4 4 1	1 4 3 0 1	2 4 4 4 1
86	2 4 4 4 1	1 4 4 4 1	3 4 3 3 3	3 4 4 4 3	1 4 2 0 1	3 4 3 3 3
87	2 4 4 4 1	1 4 3 0 1	2 4 4 4 1	1 4 4 4 1	1 2 2 2 1	4 4 4 4 1
88	3 4 4 4 3	1 2 2 2 1	2 4 2 2 1	3 4 4 4 1	1 1 1 1 1	4 4 4 4 3
89	1 4 4 4 1	2 4 2 2 1	1 2 2 2 1	4 4 4 4 3	1 2 2 2 1	1 2 2 2 1
90	1 2 1 1 1	2 4 2 2 1	1 4 4 4 1	2 4 4 4 1	4 4 4 4 4	1 4 4 4 1
91	1 4 4 4 1	1 2 1 1 1	2 2 2 2 1	2 4 2 2 1	2 4 2 2 1	1 2 1 1 1
92	2 2 2 2 1	3 4 4 4 3	4 4 4 4 1	2 4 4 4 1	1 4 1 1 1	3 4 3 3 1
93	2 4 2 2 1	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1
94	2 2 2 2 1	2 2 2 2 1	2 4 4 4 3	3 4 4 4 3	2 2 2 2 1	4 4 4 4 3
95	1 3 1 1 1	1 4 2 0 1	4 4 4 4 3	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3
96	4 4 4 4 3	2 4 4 4 1	2 2 2 2 1	4 4 4 4 3	4 4 4 4 3	1 1 1 1 1
97	3 4 3 3 3	1 4 2 0 1	1 4 4 4 1	4 4 4 4 3	4 4 4 4 1	1 4 1 1 1
98	2 4 4 4 1	1 2 2 2 1	1 1 1 1 1	4 4 4 4 3	2 2 2 2 1	1 3 3 3 1
99	3 4 4 4 1	2 4 4 4 1	2 4 4 4 1	1 4 2 0 1	3 4 4 4 3	2 4 2 2 1
100	4 4 4 4 3	1 2 1 1 1	4 4 4 4 1	1 4 3 0 1	2 2 2 2 1	2 4 2 2 1
101	1 4 4 4 1	2 4 4 4 1	1 4 3 0 1	1 4 4 4 1	4 4 4 4 3	4 4 4 4 1
102	4 4 4 4 3	2 4 4 4 1	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1	1 4 4 4 1
103	2 4 4 4 1	2 4 2 2 1	4 4 4 4 3	4 4 4 4 3	4 4 4 4 3	1 4 1 1 1
104	2 4 4 4 1	1 4 3 0 3	3 3 3 3 3	1 3 3 3 1	3 4 4 4 3	2 4 4 4 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C13	C14	C15	C16	C17	C18
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
105	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1	1 4 4 4 1	3 4 4 4 3	2 4 2 2 1
106	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	1 1 1 1 1	1 1 1 1 1
107	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 4 3 0 1	2 2 2 2 1	2 4 4 4 1
108	1 4 2 0 1	1 4 4 4 1	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3	3 4 4 4 3
109	2 4 2 2 1	2 2 2 2 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 2 2 2 1
110	1 4 1 1 1	1 4 4 4 1	2 4 4 4 1	2 4 4 4 1	2 2 2 2 1	2 2 2 2 1
111	1 4 4 4 1	1 1 1 1 1	2 4 4 4 1	4 4 4 4 1	2 4 4 4 1	4 4 4 4 3
112	1 4 2 0 1	2 2 2 2 1	4 4 4 4 1	1 2 2 2 1	2 4 4 4 1	4 4 4 4 1
113	1 4 4 4 1	2 4 2 2 1	4 4 4 4 3	2 2 2 2 1	3 4 4 4 3	2 4 4 4 1
114	1 4 4 4 1	2 2 2 2 1	2 2 2 2 1	2 2 2 2 1	2 4 4 4 1	3 4 3 3 3
115	1 4 4 4 1	1 4 2 0 1	2 2 2 2 1	2 4 4 4 1	4 4 4 4 1	1 2 2 2 1
116	2 4 2 2 1	1 4 4 4 1	1 2 2 2 1	2 4 4 4 1	1 2 1 1 1	2 4 2 2 1
117	2 4 4 4 1	2 4 4 4 1	1 4 1 1 1	1 3 3 3 3	4 4 4 4 3	2 4 2 2 1
118	1 4 4 4 1	2 4 4 4 1	2 4 2 2 1	1 2 2 2 1	4 4 4 4 3	2 4 2 2 1
119	2 4 2 2 1	4 4 4 4 3	3 4 4 4 3	1 4 4 4 1	1 4 2 0 1	4 4 4 4 1
120	4 4 4 4 1	2 4 2 2 1	3 4 4 4 3	2 2 2 2 1	2 4 4 4 1	4 4 4 4 1
121	1 4 2 0 1	2 2 2 2 1	3 4 3 3 3	1 4 1 1 1	1 4 4 4 1	2 2 2 2 1
122	1 4 1 1 1	2 4 4 4 1	2 2 2 2 1	1 4 2 0 1	4 4 4 4 1	4 4 4 4 3
123	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3
124	3 4 4 4 3	1 2 2 2 1	2 4 4 4 1	2 4 2 2 1	1 4 1 1 1	1 2 2 2 1
125	2 4 2 2 1	1 4 1 1 1	1 1 1 1 1	3 4 4 4 3	1 4 1 1 1	1 4 4 4 1
126	4 4 4 4 3	4 4 4 4 3	2 4 2 2 1	1 4 4 4 1	3 4 4 4 3	2 2 2 2 1
127	1 4 4 4 1	1 4 4 4 1	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1	2 4 4 4 1
128	1 2 2 2 1	1 4 2 0 1	2 4 4 4 1	1 1 1 1 1	2 4 4 4 1	2 4 4 4 1
129	2 2 2 2 1	2 4 2 2 1	1 4 1 1 1	3 4 3 3 3	1 2 1 1 1	2 4 4 4 1
130	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1	2 4 2 2 1	1 4 4 4 1	1 4 4 4 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C13	C14	C15	C16	C17	C18
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
131	2 4 4 4 1	4 4 4 4 1	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1
132	4 4 4 4 3	1 4 4 4 1	1 4 2 0 1	2 4 2 2 1	1 4 2 0 1	3 4 4 4 1
133	1 4 4 4 1	3 4 4 4 3	1 4 4 4 1	3 4 3 3 3	2 4 4 4 1	1 4 2 0 1
134	1 4 3 0 1	2 4 4 4 1	1 2 1 1 1	2 4 2 2 1	1 4 4 4 1	1 2 2 2 1
135	1 2 2 2 1	2 4 4 4 1	4 4 4 4 3	3 3 3 3 1	4 4 4 4 3	1 4 4 4 1
136	1 4 1 1 1	4 4 4 4 3	2 4 4 4 3	1 4 1 1 1	4 4 4 4 3	2 4 2 2 1
137	1 4 4 4 1	1 4 1 1 1	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3	4 4 4 4 3
138	4 4 4 4 3	1 2 2 2 1	1 4 4 4 1	2 4 4 4 1	1 3 3 3 1	4 4 4 4 1
139	1 2 1 1 1	1 4 4 4 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1
140	3 4 4 4 3	3 4 3 3 1	1 4 3 0 1	1 4 4 4 1	1 4 4 4 1	4 4 4 4 3
141	3 4 3 3 3	3 4 4 4 1	1 3 1 1 1	1 4 4 4 1	1 4 4 4 1	4 4 4 4 3
142	2 4 4 4 1	1 4 4 4 1	1 4 4 4 1	1 4 4 4 3	1 4 4 4 1	4 4 4 4 3
143	2 4 4 4 1	2 4 4 4 1	1 4 2 0 1	4 4 4 4 4	2 4 4 4 1	1 4 1 1 1
144	3 4 3 3 3	2 4 4 4 1	2 4 4 4 1	3 4 4 4 3	2 4 2 2 1	2 2 2 2 1
145	2 4 4 4 1	2 4 4 4 1	3 3 3 3 1	2 4 2 2 2	1 4 4 4 1	4 4 4 4 1
146	2 4 4 4 1	1 4 4 4 1	2 4 4 4 3	4 4 4 4 3	1 4 4 4 1	2 4 4 4 1
147	3 4 4 4 3	2 2 2 2 1	2 4 4 4 1	3 4 4 4 1	2 2 2 2 1	4 4 4 4 3
148	1 2 2 2 1	1 4 4 4 1	2 4 2 2 1	2 4 4 4 1	4 4 4 4 3	1 4 4 4 1
149	3 4 4 4 3	1 4 4 4 1	3 4 4 4 3	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1
150	1 2 2 2 1	1 4 2 0 1	3 4 4 4 1	1 4 4 4 1	2 4 2 2 1	1 4 1 1 1
151	1 4 4 4 1	4 4 4 4 3	4 4 4 4 3	1 2 2 2 1	1 4 4 4 1	2 4 4 4 1
152	2 4 4 4 1	1 4 4 4 1	1 3 3 3 1	2 4 2 2 1	2 4 4 4 1	2 2 2 2 1
153	1 3 3 3 1	1 4 1 1 1	3 4 4 4 3	1 4 4 4 1	1 3 3 3 1	1 2 2 2 1
154	3 4 4 4 3	3 4 4 4 3	4 4 4 4 3	1 4 2 0 1	2 4 4 4 1	1 4 2 0 1
155	2 4 2 2 1	3 4 3 3 3	2 4 4 4 1	1 4 2 0 1	2 4 4 4 1	2 4 4 4 3
156	1 4 4 4 1	2 4 2 2 1	3 4 4 4 3	2 2 2 2 1	3 4 4 4 1	2 4 4 4 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C13	C14	C15	C16	C17	C18
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
157	1 2 2 2 1	2 4 4 4 3	4 4 4 4 3	2 4 4 4 1	4 4 4 4 3	2 4 4 4 1
158	1 4 2 0 1	4 4 4 4 3	2 4 4 4 1	1 4 4 4 1	1 4 2 0 1	4 4 4 4 3
159	2 4 4 4 1	3 4 4 4 3	1 4 2 0 1	3 4 4 4 3	2 4 2 2 1	2 4 4 4 1
160	4 4 4 4 3	2 4 4 4 1	2 4 2 2 1	1 4 2 0 1	1 4 3 0 1	4 4 4 4 3
161	2 4 2 2 1	1 4 4 4 1	1 1 1 1 1	1 1 1 1 1	1 4 4 4 1	4 4 4 4 3
162	1 4 4 4 1	2 4 2 2 1	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1	4 4 4 4 3
163	4 4 4 4 3	1 4 1 1 1	2 4 2 2 1	2 4 2 2 1	1 2 2 2 1	1 1 1 1 1
164	2 4 4 4 1	2 4 4 4 1	4 4 4 4 1	2 4 4 4 1	3 4 4 4 1	4 4 4 4 3
165	2 4 4 4 1	1 4 4 4 1	2 4 4 4 1	2 4 2 2 1	2 4 4 4 1	2 2 2 2 1
166	1 4 2 0 1	2 4 2 2 1	1 4 2 0 1	1 2 2 2 1	1 4 1 1 1	1 3 3 3 1
167	1 3 1 1 1	2 2 2 2 1	2 2 2 2 1	4 4 4 4 3	2 4 2 2 1	2 4 4 4 1
168	3 4 3 3 3	1 4 4 4 1	1 4 4 4 1	1 4 4 4 1	2 4 4 4 1	1 2 2 2 1
169	2 4 4 4 3	4 4 4 4 3	3 4 3 3 3	4 4 4 4 3	3 4 3 3 1	1 4 4 4 1
170	1 2 2 2 1	1 2 1 1 1	1 2 1 1 1	1 3 1 1 1	4 4 4 4 1	3 4 4 4 3
171	2 4 4 4 1	2 2 2 2 1	1 2 1 1 1	1 3 1 1 1	2 4 4 4 1	3 4 3 3 3
172	2 2 2 2 1	3 4 4 4 3	1 4 2 0 1	1 4 4 4 1	1 4 3 0 1	1 4 4 4 1
173	1 4 1 1 1	3 4 4 4 3	2 4 4 4 1	1 4 2 0 1	3 4 4 4 3	4 4 4 4 1
174	1 4 1 1 1	1 4 3 0 1	4 4 4 4 3	4 4 4 4 3	1 2 2 2 1	4 4 4 4 3
175	2 4 4 4 1	1 4 1 1 1	1 4 3 0 1	2 4 4 4 1	4 4 4 4 3	2 4 2 2 1
176	1 4 2 0 1	2 4 2 2 1	1 1 1 1 1	2 2 2 2 1	2 4 4 4 1	4 4 4 4 3
177	1 4 4 4 1	2 4 2 2 1	1 4 4 4 1	2 4 4 4 1	3 4 4 4 3	4 4 4 4 3
178	3 4 4 4 3	1 4 3 0 1	2 4 4 4 1	1 4 2 0 1	1 4 3 0 1	2 4 4 4 1
179	1 4 4 4 1	1 3 3 3 1	3 4 4 4 3	1 4 4 4 1	2 4 2 2 1	1 4 2 0 1
180	1 2 2 2 1	2 4 2 2 1	2 4 4 4 3	1 4 4 4 1	3 3 3 3 1	2 4 4 4 1
181	1 2 2 2 1	1 4 4 4 1	1 4 1 1 1	3 4 4 4 3	1 4 2 0 1	1 4 3 0 1
182	1 4 4 4 1	3 4 3 3 3	1 4 4 4 1	1 4 4 4 3	1 4 2 0 1	2 4 2 2 1

ตารางที่ 4.2 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำแต่ละครั้ง (ต่อ)

R	C13	C14	C15	C16	C17	C18
	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5	1 2 3 4 5
183	2 2 2 2 1	1 4 4 4 1	4 4 4 4 1	2 4 4 4 1	1 4 2 0 1	1 4 2 0 1
184	1 4 4 4 3	1 2 1 1 1	4 4 4 4 3	2 4 4 4 1	4 4 4 4 1	4 4 4 4 3
185	2 4 2 2 1	2 4 2 2 1	2 4 4 4 1	1 4 4 4 1	2 2 2 2 1	2 2 2 2 1
186	1 4 1 1 1	3 4 4 4 3	1 4 2 0 1	4 4 4 4 3	1 2 2 2 1	1 4 3 0 1
187	1 4 4 4 1	2 4 2 2 1	2 4 2 2 1	1 4 1 1 1	2 4 4 4 1	2 4 4 4 1
188	4 4 4 4 3	2 2 2 2 1	1 4 4 4 1	1 4 4 4 1	2 4 2 2 1	4 4 4 4 1
189	1 4 4 4 1	2 4 4 4 3	4 4 4 4 3	1 4 2 0 1	1 3 3 3 1	4 4 4 4 3
190	1 2 2 2 1	2 2 2 2 1	1 4 4 4 1	2 4 4 4 1	2 4 4 4 1	1 4 4 4 1
191	2 4 4 4 1	2 4 2 2 1	4 4 4 4 4	1 4 4 4 1	1 4 1 1 1	4 4 4 4 3
192	2 4 2 2 1	2 2 2 2 1	1 4 4 4 1	2 4 2 2 1	1 4 4 4 1	2 4 4 4 1
193	1 4 3 0 3	2 4 4 4 1	2 4 4 4 1	4 4 4 4 1	2 4 4 4 1	2 4 4 4 1
194	4 4 4 4 3	1 4 4 4 1	3 4 4 4 3	2 4 4 4 1	3 4 3 3 3	1 4 4 4 1
195	1 4 4 4 1	1 4 4 4 1	1 2 2 2 1	2 4 4 4 1	1 4 4 4 1	1 2 2 2 1
196	2 4 4 4 1	2 4 4 4 1	3 4 4 4 3	1 4 2 0 1	1 4 1 1 1	2 4 4 4 1
197	3 4 4 4 3	4 4 4 4 1	3 4 3 3 3	1 4 1 1 1	3 4 4 4 3	4 4 4 4 1
198	2 4 2 2 1	1 4 4 4 1	1 4 2 0 1	1 4 1 1 1	2 4 4 4 1	1 4 4 4 1
199	2 4 4 4 2	1 4 4 4 1	2 4 2 2 1	2 4 4 4 1	1 4 4 4 1	4 4 4 4 3
200	1 4 3 0 1	4 4 4 4 3	1 4 4 4 3	1 2 1 1 1	1 4 2 0 1	3 4 4 4 3

จากตารางที่ 4.2 เป็นผลการคัดเลือกโมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำของทั้ง 18 เงื่อนไข เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์เป็นตัวตัดสิน ผู้วิจัยได้นำเสนอความถี่ (frequency) และร้อยละของผลการคัดสรรโมเดลที่ดีที่สุด (best model) ไว้ในตารางที่ 4.3

ตารางที่ 4.3 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำ เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์

Con	MSC	Model				
		M1	M2	M3	M4	none
		f (%)	f (%)	f (%)	f (%)	f (%)
1	AIC	26(13.0)	73(36.5)	23(11.5)	78(39.0)	0(0.0)
	BIC	0(0.0)	14(7.0)	4(2.0)	182(91.0)	0(0.0)
	ABIC	5(2.5)	35(17.5)	11(5.5)	149 (74.5)	0(0.0)
	2 of 3	5(2.5)	32(16.0)	8(4.0)	149(74.5)	6(3.0)
	p-value	103(51.5)	1(0.5)	94(47.0)	2 (1.0)	0(0.0)
2	AIC	36(18.0)	81(40.5)	21(10.5)	62(31.0)	0(0.0)
	BIC	1(0.5)	14(7.0)	3(1.5)	182(91.0)	0(0.0)
	ABIC	6(3.0)	48(24.0)	10(5.0)	136(68.0)	0(0.0)
	2 of 3	6(3.0)	42(21.0)	5(2.5)	136(68.0)	11(5.5)
	p-value	116(58.0)	2(1.0)	82(41.0)	0(0.0)	0(0.0)
3	AIC	33(16.5)	58(29.0)	34(17.0)	75(37.5)	0(0.0)
	BIC	0(0.0)	12(6.0)	6(3.0)	182(91.0)	0(0.0)
	ABIC	2(1.0)	29(14.5)	17(8.5)	152(76.0)	0(0.0)
	2 of 3	2(1.0)	24(12.0)	14(7.0)	152(76.0)	8(4.0)
	p-value	95(47.5)	0(0.0)	104(52.0)	1(0.5)	0(0.0)
4	AIC	42(21.0)	60(30.0)	21(10.5)	77(38.5)	0(0.0)
	BIC	1(0.5)	12(6.0)	5(2.5)	182(91.0)	0(0.0)
	ABIC	9(4.5)	30(15.0)	16(8.0)	145(72.5)	0(0.0)
	2 of 3	9(4.5)	24(12.0)	10(5.0)	145(72.5)	12(6.0)
	p-value	109(54.5)	1(0.5)	89(44.5)	1(0.5)	0(0.0)

ตารางที่ 4.3 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำ เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ (ต่อ)

Con	MSC	Model				
		M1	M2	M3	M4	none
		f (%)	f (%)	f (%)	f (%)	f (%)
5	AIC	36(18.0)	81(40.5)	21(10.5)	62(31.0)	0(0.0)
	BIC	1(0.5)	14(7.0)	3(1.5)	182(91.0)	0(0.0)
	ABIC	6(3.0)	48(24.0)	10(5.0)	136(68.0)	0(0.0)
	2 of 3	6(3.0)	42(21.0)	5(2.5)	136(68.0)	11(5.5)
	p-value	116(58.0)	2(1.0)	82(41.0)	0(0.0)	0(0.0)
6	AIC	33(16.5)	58(29.0)	34(17.0)	75(27.5)	0(0.0)
	BIC	0(0.0)	12(6.0)	6(3.0)	182(91.0)	0(0.0)
	ABIC	2(1.0)	29(14.5)	17(8.5)	152(76.0)	0(0.0)
	2 of 3	2(1.0)	24(12.0)	14(7.0)	152(76.0)	8(4.0)
	p-value	95(47.5)	0(0.0)	104(52.0)	1(0.5)	0(0.0)
7	AIC	40(20.0)	60(30.0)	21(10.5)	79(39.5)	0(0.0)
	BIC	1(0.5)	12(6.0)	5(2.5)	182(91.0)	0(0.0)
	ABIC	9(4.5)	30(15.0)	16(8.0)	145(72.5)	0(0.0)
	2 of 3	9(4.5)	24(12.0)	10(5.0)	147(73.5)	10(5.0)
	p-value	110(55.0)	1(0.5)	88(44.0)	1(0.5)	0(0.0)
8	AIC	42(21.0)	64(32.0)	32(16.0)	62(31.0)	0(0.0)
	BIC	0(0.0)	14(7.0)	6(3.0)	180(90.0)	0(0.0)
	ABIC	8(4.0)	34(17.0)	15(7.5)	143(71.5)	0(0.0)
	2 of 3	8(4.0)	28(14.0)	12(6.0)	143(71.5)	9(4.5)
	p-value	116(58.0)	0(0.0)	83(41.5)	1(0.5)	0(0.0)
9	AIC	30(15.0)	56(28.0)	33(16.5)	81(40.5)	0(0.0)
	BIC	0(0.0)	18(9.0)	4(2.0)	178(89.0)	0(0.0)
	ABIC	2(1.0)	33(16.5)	18(9.0)	147(73.5)	0(0.0)
	2 of 3	2(1.0)	29(14.5)	14(7.0)	147(73.5)	8(4.0)
	p-value	90(45.0)	1(0.5)	108(54.0)	1(0.5)	0(0.0)

ตารางที่ 4.3 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำ เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ (ต่อ)

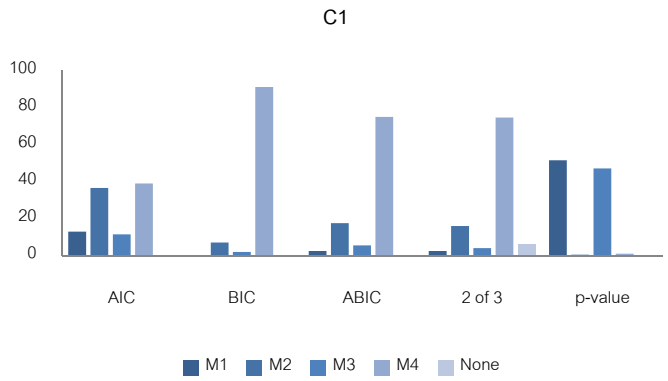
Con	MSC	Model				
		M1	M2	M3	M4	none
		f (%)	f (%)	f (%)	f (%)	f (%)
10	AIC	78(39.0)	65(32.5)	17(8.5)	40(20.0)	0(0.0)
	BIC	3(1.5)	22(11.0)	2(1.0)	173(86.5)	0(0.0)
	ABIC	14(7.0)	47(23.5)	17(8.5)	122(61.0)	0(0.0)
	2 of 3	14(7.0)	34(17.0)	5(2.5)	122(61.0)	25(12.5)
	p-value	151(75.5)	1(0.5)	48(24.0)	0(0.0)	0(0.0)
11	AIC	80(40.0)	67(33.5)	22(11.0)	31(15.5)	0(0.0)
	BIC	2(1.0)	30(15.0)	3(1.5)	165(82.5)	0(0.0)
	ABIC	16(8.0)	54(27.0)	13(6.5)	117(58.5)	0(0.0)
	2 of 3	16(8.0)	41(20.5)	6(3.0)	117(58.5)	20(10.0)
	p-value	152(76.0)	2(1.0)	46(23.0)	0(0.0)	0(0.0)
12	AIC	70(35.0)	69(34.5)	17(8.5)	44(22.0)	0(0.0)
	BIC	3(1.5)	17(8.5)	4(2.0)	176(88.0)	0(0.0)
	ABIC	19(9.5)	41(20.5)	9(4.5)	131(65.5)	0(0.0)
	2 of 3	19(9.5)	33(16.5)	6(3.0)	131(65.5)	11(5.5)
	p-value	143(71.5)	2(1.0)	54(27.0)	1(0.5)	0(0.0)
13	AIC	81(40.5)	67(33.5)	21(10.5)	31(15.5)	0(0.0)
	BIC	2(1.0)	30(15.0)	2(1.0)	166(83.0)	0(0.0)
	ABIC	16(8.0)	54(27.0)	12(6.0)	118(59.0)	0(0.0)
	2 of 3	15(7.5)	40(20.0)	6(3.0)	119(59.5)	0(0.0)
	p-value	154(77.0)	1(0.5)	45(22.5)	0(0.0)	0(0.0)
14	AIC	80(40.0)	75(37.5)	21(10.5)	24(12.0)	0(0.0)
	BIC	1(0.5)	28(14.0)	7(3.5)	164(82.0)	0(0.0)
	ABIC	16(8.0)	56(28.0)	18(9.0)	110(55.0)	0(0.0)
	2 of 3	16(8.0)	44(22.0)	13(6.5)	110(55.0)	17(8.5)
	p-value	156(78.0)	1(0.5)	42(21.0)	1(0.5)	0(0.0)

ตารางที่ 4.3 โมเดลที่ดีที่สุด (best model) ของแต่ละการทำซ้ำ เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ (ต่อ)

Con	MSC	Model				
		M1	M2	M3	M4	none
		f (%)	f (%)	f (%)	f (%)	f (%)
15	AIC	71(35.5)	68(34.0)	21(10.5)	40(20.0)	0(0.0)
	BIC	5(2.5)	25(12.5)	5(2.5)	165(82.5)	0(0.0)
	ABIC	15(7.5)	47(23.5)	19(9.5)	119(59.5)	0(0.0)
	2 of 3	15(7.5)	37(18.5)	10(5.0)	119(59.5)	19(9.5)
	p-value	139(69.5)	3(1.5)	57(28.5)	1(0.5)	0(0.0)
16	AIC	87(43.5)	62(31.0)	18(9.0)	33(16.5)	0(0.0)
	BIC	2(1.0)	22(11.0)	6(3.0)	170(85.0)	0(0.0)
	ABIC	15(7.5)	50(25.0)	11(5.5)	124(62.0)	0(0.0)
	2 of 3	15(7.5)	37(18.5)	6(3.0)	124(62.0)	18(9.0)
	p-value	155(77.5)	2(1.0)	42(21.0)	1(0.5)	0(0.0)
17	AIC	74(37.0)	69(34.5)	21(10.5)	36(18.0)	0(0.0)
	BIC	3(1.5)	28(14.0)	5(2.5)	164(82.0)	0(0.0)
	ABIC	15(7.5)	54(27.0)	15(7.5)	116(58.0)	0(0.0)
	2 of 3	15(7.5)	42(21.0)	10(5.0)	116(58.0)	17(8.5)
	p-value	155(77.5)	0(0.0)	44(22.0)	1(0.5)	0(0.0)
18	AIC	68 (34.0)	71(35.5)	14(7.0)	47(23.5)	0(0.0)
	BIC	3(1.5)	32(16.0)	5(2.5)	160(80.0)	0(0.0)
	ABIC	12(6.0)	53(26.5)	14(7.0)	121(60.5)	0(0.0)
	2 of 3	12(6.0)	47(23.5)	10(5.0)	121(60.5)	10(5.0)
	p-value	150(75.0)	1(0.5)	49(24.5)	0(0.0)	0(0.0)

จากตาราง 4.3 สรุปได้ว่าสำหรับเงื่อนไขที่ 1 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 39.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 36.5 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็นร้อยละ 13.0 และ 11.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information

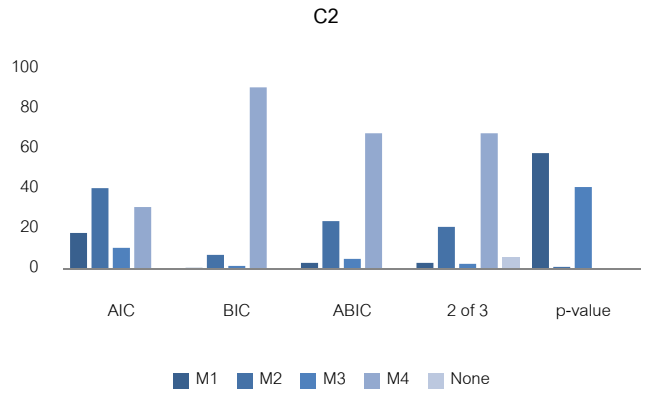
criterion) เป็นเกณฑ์พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 91.0 รองลงมาคือ โมเดลที่ 2 และโมเดลที่ 3 คิดเป็นร้อยละ 7.0 และ 2.0 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 74.5 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 17.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 5.5 และ 2.5 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 74.5 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 16.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 4.0 และ 2.5 ตามลำดับและเมื่อใช้ดัชนี p-value เป็นเกณฑ์ พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 51.5 รองลงมาคือ โมเดลที่ 3 คิดเป็นร้อยละ 47.0 ส่วนโมเดลที่ 4 และโมเดลที่ 2 คิดเป็นร้อยละ 1.0 และ 0.5 ตามลำดับ ดังแผนภาพ



แผนภาพที่ 4.19 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 1 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

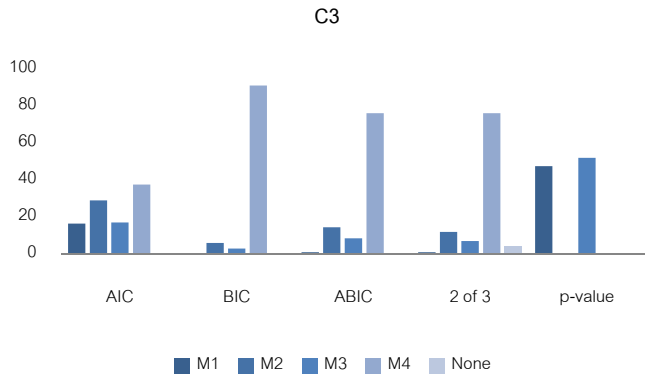
เงื่อนไขที่ 2 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 2 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 40.5 รองลงมาคือ โมเดลที่ 4 คิดเป็นร้อยละ 31.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็นร้อยละ 18.0 และ 10.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 91.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 7.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 1.5 และ 0.5 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 68.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 24.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 5.0 และ 3.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best

model) สูงสุด นั่นคือ ร้อยละ 68.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 21.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็นร้อยละ 3.0 และ 2.5 ตามลำดับ เมื่อใช้ ดัชนี p-value เป็นเกณฑ์ พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 58.0 รองลงมาคือ โมเดลที่ 3 คิดเป็น ร้อยละ 41.0 และโมเดลที่ 2 คิดเป็นร้อยละ 1.0 ตามลำดับ ดังแผนภาพ



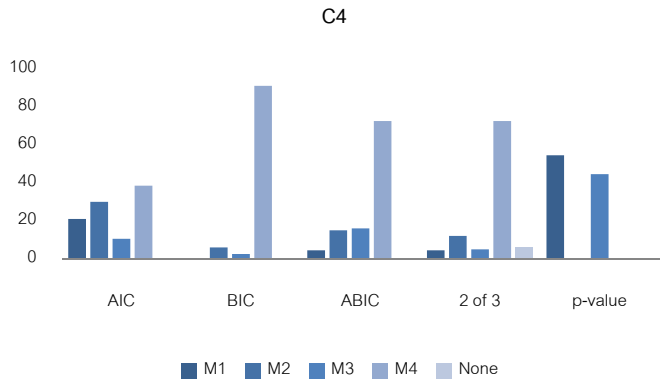
แผนภาพที่ 4.20 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 2 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 3 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 3 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 37.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 29.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 17.0 และ 16.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 91.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 6.0 และโมเดลที่ 3 คิดเป็น ร้อยละ 3.0 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 76.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 14.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 8.5 และ 1.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 76.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 12.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 7.0 และ 1.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์ พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 47.5 รองลงมาคือ โมเดลที่ 3 คิดเป็นร้อยละ 52.0 และโมเดลที่ 4 คิดเป็นร้อยละ 0.5 ตามลำดับ ดังแผนภาพ



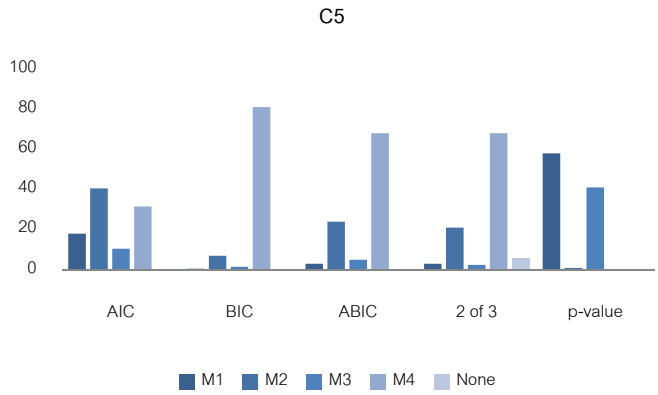
แผนภาพที่ 4.21 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 3 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 4 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 38.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 30.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็นร้อยละ 21.0 และ 10.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 91.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 6.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 2.5 และ 0.5 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 72.5 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 15.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 8.0 และ 4.5 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 72.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 12.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 5.0 และ 4.5 ตามลำดับ เมื่อใช้ ดัชนี p-value เป็นเกณฑ์ พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 54.5 รองลงมาคือ โมเดลที่ 3 คิดเป็นร้อยละ 44.5 โมเดลที่ 2 และโมเดลที่ 4 คิดเป็นร้อยละ 0.5 เท่ากันตามลำดับ ดังแผนภาพ



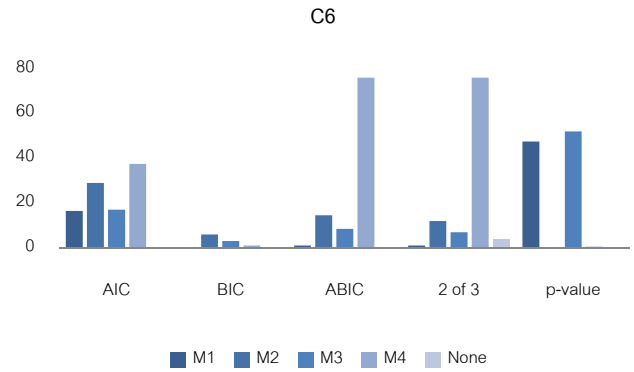
แผนภาพที่ 4.22 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 4 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 5 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 2 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 40.5 รองลงมาคือ โมเดลที่ 4 คิดเป็นร้อยละ 31.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็นร้อยละ 18.0 และ 10.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 91.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 7.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 1.5 และ 0.5 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 68.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 24.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 5.0 และ 3.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์ พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 68.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 21.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็นร้อยละ 3.0 และ 2.5 ตามลำดับ เมื่อใช้ ดัชนี p-value เป็นเกณฑ์ พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 58.0 รองลงมาคือ โมเดลที่ 3 คิดเป็นร้อยละ 41.0 และโมเดลที่ 2 คิดเป็นร้อยละ 1.0 ตามลำดับ ดังแผนภาพ



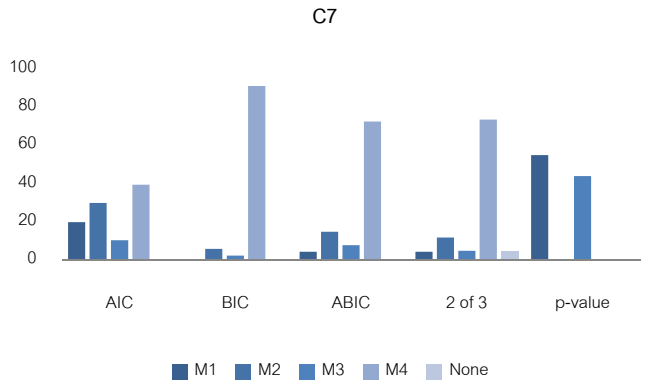
แผนภาพที่ 4.23 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 5 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 6 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 2 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 37.5 รองลงมาคือ โมเดลที่ 4 คิดเป็น ร้อยละ 29.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 17.0 และ 16.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 91.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 6.0 และโมเดลที่ 3 คิดเป็นร้อยละ 3.0 เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 76.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 14.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 8.5 และ 1.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือร้อยละ 76.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 12.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 7.0 และ 1.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 3 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 52.0 รองลงมาคือ โมเดลที่ 1 คิดเป็นร้อยละ 47.5 และโมเดลที่ 4 คิดเป็นร้อยละ 0.5 ตามลำดับ ดังแผนภาพ



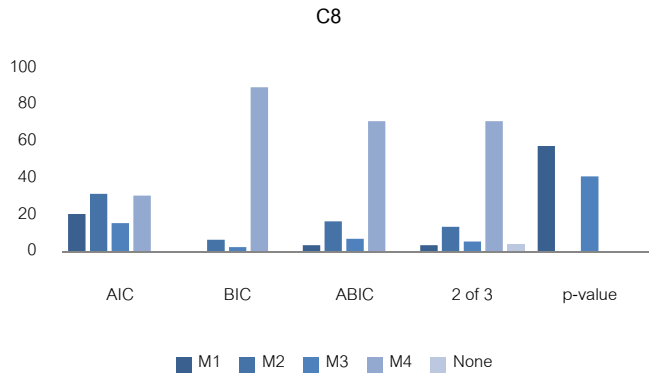
แผนภาพที่ 4.24 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 6 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 7 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 39.5 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 30.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็นร้อยละ 20.0 และ 10.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 91.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 6.0 โมเดลที่ 3 คิดเป็นร้อยละ 2.5 และโมเดล 1 คิดเป็นร้อยละ 0.5 เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 72.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 15.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 8.0 และ 4.5 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 73.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 12.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 5.0 และ 4.5 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 55.0รองลงมาคือ โมเดลที่ 3 คิดเป็นร้อยละ 44.0 ส่วนโมเดลที่ 1 และโมเดลที่ 4 คิดเป็นร้อยละ 0.5 เท่ากัน ดังแผนภาพ



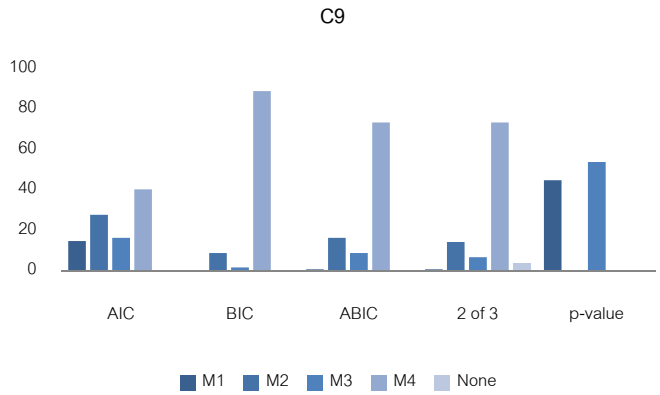
แผนภาพที่ 4.25 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 7 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 8 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 2 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 32.0 รองลงมาคือ โมเดลที่ 4 คิดเป็น ร้อยละ 31.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็นร้อยละ 21.0 และ 16.0 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 90.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 7.0 และโมเดลที่ 3 คิดเป็นร้อยละ 3.0 เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 71.5 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 17.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 7.5 และ 4.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 71.5 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 14.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 6.0 และ 4.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 58.0 รองลงมาคือ โมเดลที่ 3 คิดเป็นร้อยละ 41.5 และโมเดลที่ 4 คิดเป็นร้อยละ 0.5 ตามลำดับ ดังแผนภาพ



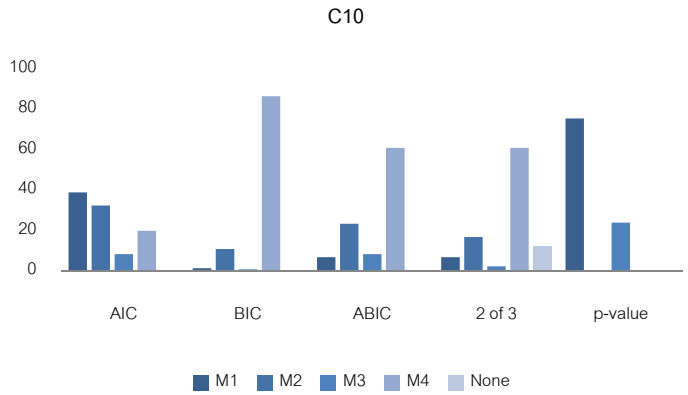
แผนภาพที่ 4.26 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 8 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 9 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 40.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 28.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น ร้อยละ 16.5 และ 15.0 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 89.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 9.0 และโมเดลที่ 3 คิดเป็น ร้อยละ 2.0 เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 73.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 16.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น ร้อยละ 9.0 และ 1.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 73.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 14.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น ร้อยละ 7.0 และ 1.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 3 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 54.0 รองลงมาคือ โมเดลที่ 1 คิดเป็นร้อยละ 45.0 ส่วนโมเดลที่ 2 และโมเดลที่ 4 คิดเป็นร้อยละ 0.5 เท่ากัน ดังแผนภาพ



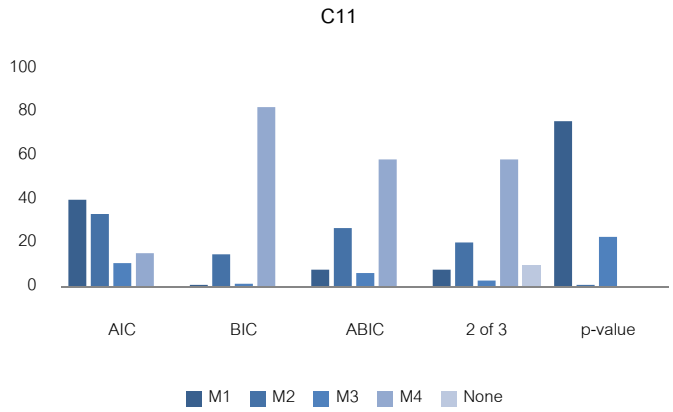
แผนภาพที่ 4.27 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 9 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 10 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 39.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 32.5 ส่วนโมเดลที่ 4 และโมเดลที่ 3 คิดเป็นร้อยละ 20.0 และ 8.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 86.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 11.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 1.5 และ 1.0 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 61.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 23.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น ร้อยละ 8.5 และ 7.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 61.0 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 17.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 7.0 และ 2.5 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 75.5 รองลงมาคือ โมเดลที่ 3 คิดเป็นร้อยละ 24.0 และโมเดลที่ 2 คิดเป็นร้อยละ 0.5 ตามลำดับ ดังแผนภาพ



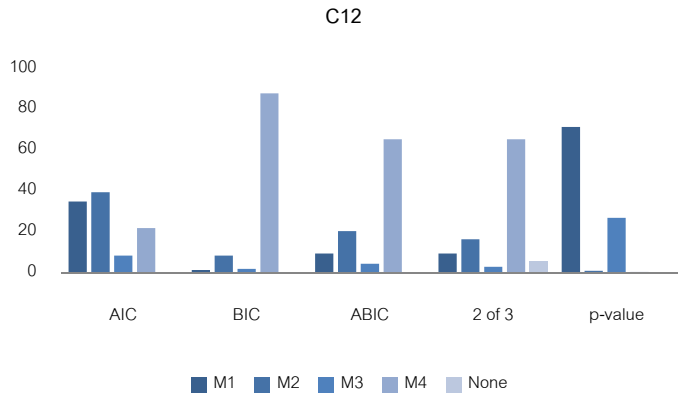
แผนภาพที่ 4.28 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 10 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 11 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 40.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 33.5 ส่วนโมเดลที่ 4 และโมเดลที่ 3 คิดเป็นร้อยละ 15.5 และ 11.0 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 82.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 15.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 1.5 และ 1.0 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 58.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 27.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 8.0 และ 6.5 ตามลำดับ เมื่อใช้ ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 58.5 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 20.5 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 8.0 และ 3.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 76.0 รองลงมาคือ โมเดลที่ 3 คิดเป็นร้อยละ 23.0 และโมเดลที่ 2 คิดเป็นร้อยละ 1.0 ตามลำดับ ดังแผนภาพ



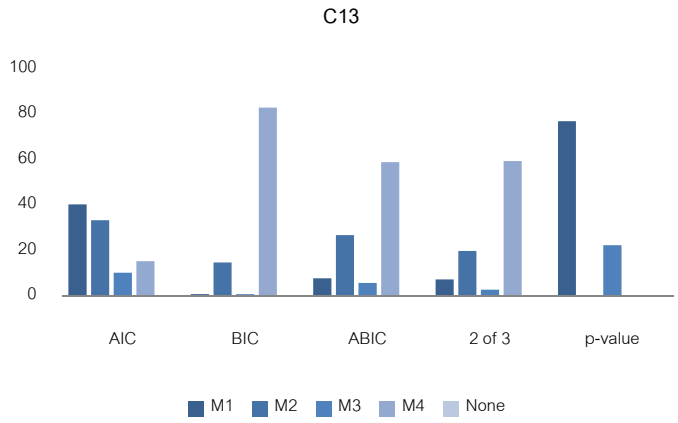
แผนภาพที่ 4.29 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 11 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 12 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 35.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 34.5 ส่วนโมเดลที่ 4 และโมเดลที่ 3 คิดเป็นร้อยละ 22.0 และ 8.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 88.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 8.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น ร้อยละ 2.0 และ 1.5 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 65.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 20.5 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 9.5 และ 4.5 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 65.5 รองลงมาคือ โมเดลที่ 2 คิดเป็นร้อยละ 16.5 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็นร้อยละ 9.5 และ 3.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 71.5 รองลงมาคือ โมเดลที่ 3 คิดเป็นร้อยละ 27.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 9.5 และ 3.0 ตามลำดับ ดังแผนภาพ



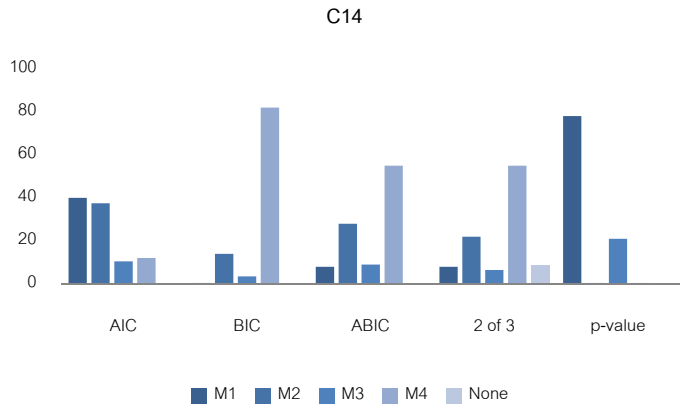
แผนภาพที่ 4.30 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 12 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 13 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 40.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 33.5 ส่วนโมเดลที่ 4 และโมเดลที่ 3 คิดเป็นร้อยละ 15.5 และ 10.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 83.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 15.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น 1.0 เท่ากัน เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 59.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 27.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 8.0 และ 6.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 59.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 20.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 7.5 และ 3.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 77.0 รองลงมาคือ โมเดลที่ 3 คิดเป็น ร้อยละ 22.5 และโมเดลที่ 2 คิดเป็นร้อยละ 0.5 ตามลำดับ ดังแผนภาพ



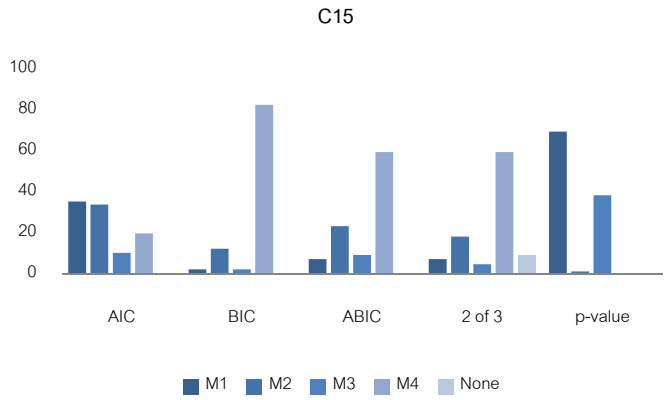
แผนภาพที่ 4.31 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 13 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 14 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 40.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 37.5 ส่วนโมเดลที่ 4 และโมเดลที่ 3 คิดเป็นร้อยละ 12.0 และ 10.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 82.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 14.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น 3.5 และ 0.5 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 55.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 28.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น ร้อยละ 9.0 และ 8.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 55.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 22.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 8.0 และ 6.5 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 78.0 รองลงมาคือ โมเดลที่ 3 คิดเป็น ร้อยละ 21.0 ส่วนโมเดลที่ 2 และโมเดลที่ 4 คิดเป็นร้อยละ 0.5 เท่ากัน ดังแผนภาพ



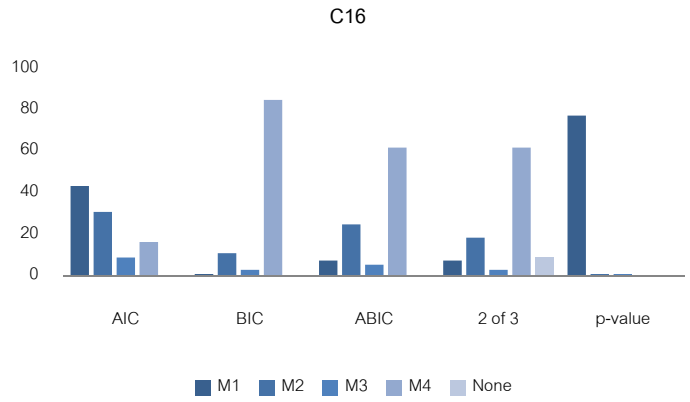
แผนภาพที่ 4.32 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 14 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 15 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 35.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 34.0 ส่วนโมเดลที่ 4 และโมเดลที่ 3 คิดเป็น ร้อยละ 20.0 และ 10.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 82.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 12.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น 2.5 เท่ากัน เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 59.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 23.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น ร้อยละ 9.5 และ 7.5 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 59.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 18.5 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 7.5 และ 5.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 69.5 รองลงมาคือ โมเดลที่ 3 คิดเป็น ร้อยละ 28.5 ส่วนโมเดลที่ 2 และโมเดลที่ 4 คิดเป็นร้อยละ 1.5 และ 0.5 ตามลำดับ ดังแผนภาพ



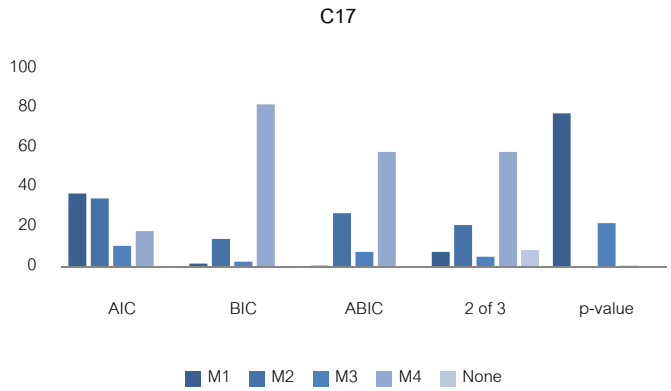
แผนภาพที่ 4.33 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 15 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 16 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 43.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 31.0 ส่วนโมเดลที่ 4 และโมเดลที่ 3 คิดเป็นร้อยละ 16.5 และ 9.0 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 85.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 11.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 3.0 และร้อยละ 1.0 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 62.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 25.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 7.5 และ 5.5 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 62.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 18.5 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 7.5 และ 3.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 77.5 รองลงมาคือ โมเดลที่ 3 คิดเป็น ร้อยละ 21.0 ส่วนโมเดลที่ 2 และโมเดลที่ 4 คิดเป็นร้อยละ 1.0 และ 0.5 ตามลำดับ ดังแผนภาพ



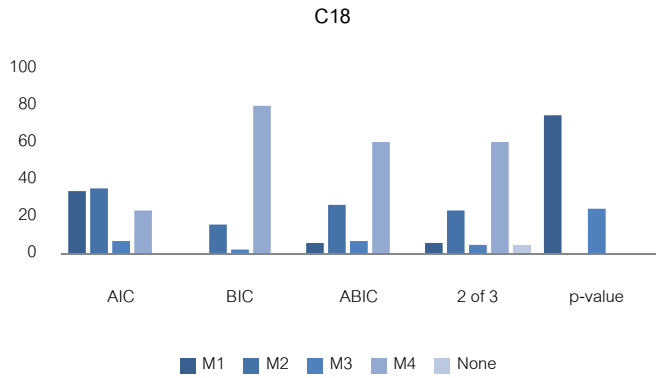
แผนภาพที่ 4.34 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 16 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 17 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 37.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 34.5 ส่วนโมเดลที่ 4 และโมเดลที่ 3 คิดเป็นร้อยละ 18.0 และ 10.5 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 82.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 14.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 2.5 และร้อยละ 1.5 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 58.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 27.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 7.5 เท่ากัน เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 58.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 21.0 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 7.5 และ 5.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 77.5 รองลงมาคือ โมเดลที่ 3 คิดเป็น ร้อยละ 22.0 และโมเดลที่ 4 คิดเป็นร้อยละ 0.5 ตามลำดับ ดังแผนภาพ



แผนภาพที่ 4.35 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 17 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

เงื่อนไขที่ 18 เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 2 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 35.5 รองลงมาคือ โมเดลที่ 1 คิดเป็น ร้อยละ 34.0 ส่วนโมเดลที่ 4 และโมเดลที่ 3 คิดเป็นร้อยละ 23.5 และ 7.0 ตามลำดับ เมื่อใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 80.0 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 16.0 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็นร้อยละ 2.5 และร้อยละ 1.5 ตามลำดับ เมื่อใช้ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 60.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 26.5 ส่วนโมเดลที่ 3 และโมเดลที่ 1 คิดเป็น ร้อยละ 7.0 และ 6.0 ตามลำดับ เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 4 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 60.5 รองลงมาคือ โมเดลที่ 2 คิดเป็น ร้อยละ 23.5 ส่วนโมเดลที่ 1 และโมเดลที่ 3 คิดเป็น ร้อยละ 6.0 และ 5.0 ตามลำดับ เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงสุด นั่นคือ ร้อยละ 75.0 รองลงมาคือ โมเดลที่ 3 คิดเป็น ร้อยละ 24.5 และโมเดลที่ 2 คิดเป็นร้อยละ 0.5 ตามลำดับ ดังแผนภาพ



แผนภาพที่ 4.36 ผลการสกัดโมเดลที่ดีที่สุด (best model) ของเงื่อนไขที่ 18 หลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria)

ดังนั้นจึงสรุปได้ว่า เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) ในเงื่อนไขที่ 10 - 18 ซึ่งเป็นเงื่อนไขการจำลองที่มีกลุ่มตัวอย่างขนาดใหญ่ นั่นคือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) มีจำนวน 100 คน โมเดลที่ 2 หรือโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) เป็นโมเดลที่ดีที่สุด (best model) ใน 4 เงื่อนไขคือเงื่อนไขที่ 2, 5, 6 และ 8 และโมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ดีที่สุด (best model) ใน 5 เงื่อนไขคือเงื่อนไขที่ 1, 3, 4, 7 และ 9

เมื่อใช้ดัชนี BIC (Bayesian information criterion) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ได้ผลสอดคล้องกันคือ โมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ดีที่สุด (best model) ในทุกเงื่อนไข

เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) ใน 15 เงื่อนไข คือ เงื่อนไขที่ 1, 2, 4, 5, 7, 8, 10, 11, 12, 13,14,15, 16,17,18 และโมเดลที่ 3 หรือโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) เป็นโมเดลที่ดีที่สุด (best model) ในเงื่อนไขที่ 3, 6, 9 เท่านั้น

การศึกษาในครั้งนี้มีวัตถุประสงค์ข้อที่ 2 คือเพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ซึ่งจากตาราง 4.3 และแผนภาพ สรุปได้ว่าเมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่แตกต่างกัน จะได้ผลการคัดสรรโมเดลที่ดีที่สุด (best model) แตกต่างกัน นั่นคือ เมื่อใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงที่สุด ในขณะที่เมื่อใช้ดัชนี BIC (Bayesian information criterion) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ได้ผลที่สอดคล้องกันคือ โมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) มีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงที่สุด

ดังนั้นจึงสรุปได้ว่า จากการศึกษาในครั้งนี้ พบโมเดลที่ดีที่สุด (best model) 2 โมเดล คือโมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) และโมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF)

ผู้วิจัยนำเสนอค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์ทั้ง 6 พารามิเตอร์ที่สกัดจากโมเดลที่ดีที่สุด (best model) เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์นี้ เพื่อตอบวัตถุประสงค์การศึกษาข้อแรก ที่ผู้วิจัยต้องการศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มไว้ในตอนที่ 3

ตอนที่ 3 ค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์ที่สกัดจากโมเดลที่ดีที่สุด (best model) เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์

ในตอนที่ 3 นี้ ผู้วิจัยนำเสนอผลการประมาณค่าพารามิเตอร์ที่สกัดจากโมเดลที่ดีที่สุด (best model) เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ ได้ผลดังตารางที่ 4.4

ตารางที่ 4.4 ค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์ที่สกัดจากโมเดลที่ดีที่สุด (best model) เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์

Con	MSC	PEST					
		CLADIFF(γ_2)	3WAYINT(γ_{7i})	ILDIF(γ_{5i})	ILADIFF(γ_1)	IAWC($\theta_j^{(2)}$)	CA($\theta_k^{(3)}$)
		Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
1	AIC	0.130(0.106)	0.351(0.001)	-0.534(0.099)	0.236(0.046)	1.094(0.044)	0.974(0.102)
	BIC	0.154(0.106)	0.473(0.001)	-0.595(0.099)	0.256(0.047)	1.095(0.044)	0.975(0.102)
	ABIC	0.145(0.106)	0.428(0.001)	-0.573(0.099)	0.248(0.047)	1.094(0.044)	0.975(0.102)
	2 of 3	0.142(0.103)	0.420(0.001)	-0.559(0.096)	0.243(0.045)	1.061(0.043)	0.948(0.099)
	p-value	0.110(0.107)	0.334(0.001)	-0.525(0.098)	0.215(0.046)	1.095(0.044)	0.970(0.102)
2	AIC	0.135(0.106)	0.322(0.001)	-0.503(0.098)	0.228(0.046)	1.095(0.045)	0.986(0.101)
	BIC	0.165(0.106)	0.473(0.001)	-0.578(0.099)	0.254(0.047)	1.095(0.045)	0.986(0.101)
	ABIC	0.155(0.106)	0.407(0.001)	-0.545(0.099)	0.247(0.047)	1.095(0.045)	0.987(0.101)
	2 of 3	0.145(0.100)	0.394(0.001)	-0.522(0.093)	0.236(0.044)	1.036(0.042)	0.932(0.096)
	p-value	0.114(0.106)	0.311(0.001)	-0.497(0.098)	0.210(0.046)	1.094(0.044)	0.985(0.101)
3	AIC	0.112(0.106)	0.365(0.001)	-0.528(0.097)	0.222(0.046)	1.094(0.044)	0.999(0.104)
	BIC	0.147(0.106)	0.487(0.001)	-0.589(0.098)	0.252(0.046)	1.095(0.044)	1.001(0.103)
	ABIC	0.138(0.106)	0.451(0.001)	-0.571(0.098)	0.244(0.046)	1.095(0.044)	1.001(0.104)
	2 of 3	0.135(0.101)	0.439(0.001)	-0.553(0.094)	0.236(0.044)	1.051(0.043)	0.965(0.099)
	p-value	0.096(0.106)	0.346(0.001)	-0.524(0.097)	0.205(0.046)	1.095(0.044)	0.998(0.103)
4	AIC	0.124(0.107)	0.343(0.001)	-0.537(0.097)	0.217(0.047)	1.106(0.045)	0.953(0.098)
	BIC	0.162(0.107)	0.480(0.001)	-0.605(0.098)	0.248(0.047)	1.106(0.045)	0.954(0.098)
	ABIC	0.148(0.107)	0.433(0.001)	-0.582(0.097)	0.236(0.047)	1.106(0.045)	0.954(0.098)
	2 of 3	0.138(0.101)	0.418(0.001)	-0.555(0.091)	0.225(0.044)	1.040(0.042)	0.903(0.093)
	p-value	0.107(0.107)	0.330(0.001)	-0.531(0.097)	0.201(0.046)	1.105(0.045)	0.953(0.098)

ตารางที่ 4.4 ค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์ที่สกัดจากโมเดลที่ดีที่สุด (best model) เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ (ต่อ)

Con	MSC	PEST					
		CLADIFF(γ_2)	3WAYINT(γ_{7i})	ILDIF(γ_{5i})	ILADIFF(γ_1)	IAWC($\theta_j^{(2)}$)	CA($\theta_k^{(3)}$)
		Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
5	AIC	0.134(0.106)	0.322(0.001)	-0.503(0.098)	0.228(0.046)	1.094(0.045)	0.986(0.101)
	BIC	0.164(0.106)	0.473(0.001)	-0.578(0.099)	0.254(0.047)	1.095(0.045)	0.986(0.101)
	ABIC	0.155(0.106)	0.406(0.001)	-0.545(0.098)	0.246(0.046)	1.095(0.044)	0.986(0.101)
	2 of 3	0.145(0.100)	0.394(0.001)	-0.522(0.093)	0.236(0.044)	1.036(0.042)	0.931(0.095)
	p-value	0.114(0.106)	0.311(0.001)	-0.497(0.098)	0.209(0.046)	1.094(0.044)	0.985(0.101)
6	AIC	0.111(0.106)	0.365(0.001)	-0.528(0.097)	0.222(0.046)	1.094(0.044)	0.999(0.103)
	BIC	0.146(0.105)	0.487(0.002)	-0.589(0.098)	0.251(0.046)	1.095(0.044)	1.001(0.103)
	ABIC	0.137(0.105)	0.451(0.001)	-0.571(0.097)	0.243(0.046)	1.094(0.044)	1.001(0.103)
	2 of 3	0.134(0.104)	0.439(0.001)	-0.552(0.093)	0.235(0.044)	1.050(0.042)	0.965(0.099)
	p-value	0.096(0.106)	0.355(0.001)	-0.524(0.097)	0.204(0.045)	1.094(0.044)	0.998(0.103)
7	AIC	0.125(0.105)	0.344(0.001)	-0.539(0.098)	0.216(0.046)	1.107(0.045)	0.954(0.098)
	BIC	0.163(0.105)	0.481(0.001)	-0.607(0.098)	0.248(0.047)	1.107(0.045)	0.955(0.098)
	ABIC	0.147(0.105)	0.434(0.001)	-0.582(0.097)	0.236(0.048)	1.107(0.045)	0.953(0.098)
	2 of 3	0.138(0.101)	0.418(0.001)	-0.553(0.091)	0.225(0.044)	1.040(0.042)	0.904(0.093)
	p-value	0.106(0.105)	0.327(0.001)	-0.527(0.097)	0.200(0.046)	1.107(0.045)	0.953(0.098)
8	AIC	0.100(0.106)	0.339(0.001)	-0.527(0.098)	0.211(0.046)	1.095(0.044)	1.015(0.103)
	BIC	0.137(0.105)	0.478(0.001)	-0.597(0.099)	0.243(0.046)	1.095(0.044)	1.015(0.103)
	ABIC	0.124(0.105)	0.429(0.001)	-0.572(0.098)	0.233(0.046)	1.095(0.044)	1.016(0.103)
	2 of 3	0.115(0.101)	0.416(0.001)	-0.551(0.094)	0.223(0.044)	1.045(0.042)	0.970(0.098)
	p-value	0.084(0.106)	0.321(0.001)	-0.518(0.098)	0.196(0.046)	1.095(0.044)	1.012(0.102)
9	AIC	0.118(0.106)	0.379(0.001)	-0.539(0.098)	0.223(0.046)	1.103(0.044)	1.010(0.104)
	BIC	0.152(0.106)	0.477(0.001)	-0.587(0.098)	0.253(0.047)	1.103(0.044)	1.010(0.104)
	ABIC	0.141(0.106)	0.446(0.001)	-0.572(0.098)	0.243(0.047)	1.103(0.044)	1.011(0.104)
	2 of 3	0.139(0.101)	0.434(0.001)	-0.551(0.094)	0.234(0.045)	1.059(0.042)	0.972(0.099)
	p-value	0.102(0.106)	0.368(0.001)	-0.533(0.098)	0.209(0.046)	1.103(0.044)	1.008(0.104)

ตารางที่ 4.4 ค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์ที่สกัดจากโมเดลที่ดีที่สุด (best model) เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ (ต่อ)

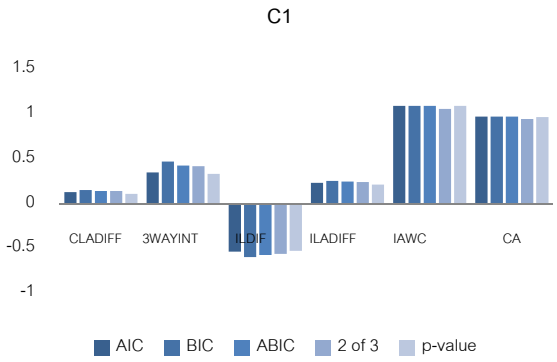
Con	MSC	PEST					
		CLADIFF(γ_2)	3WAYINT(γ_{7i})	ILDIF(γ_{5i})	ILADIFF(γ_1)	IAWC($\theta_j^{(2)}$)	CA($\theta_k^{(3)}$)
		Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
10	AIC	0.081(0.066)	0.332(0.001)	-0.524(0.068)	0.215(0.032)	1.098(0.031)	0.965(0.071)
	BIC	0.123(0.066)	0.473(0.001)	-0.594(0.070)	0.250(0.033)	1.099(0.031)	0.968(0.071)
	ABIC	0.107(0.066)	0.422(0.001)	-0.569(0.069)	0.237(0.033)	1.099(0.031)	0.965(0.071)
	2 of 3	0.101(0.058)	0.376(0.001)	-0.501(0.060)	0.211(0.029)	0.960(0.027)	0.844(0.062)
	p-value	0.066(0.066)	0.316(0.001)	-0.517(0.068)	0.200(0.032)	1.098(0.031)	0.963(0.071)
11	AIC	0.077(0.066)	0.325(0.001)	-0.510(0.068)	0.215(0.032)	1.102(0.031)	0.969(0.071)
	BIC	0.120(0.066)	0.462(0.001)	-0.579(0.069)	0.251(0.033)	1.103(0.031)	0.969(0.072)
	ABIC	0.104(0.066)	0.409(0.001)	-0.552(0.069)	0.239(0.033)	1.103(0.031)	0.971(0.072)
	2 of 3	0.104(0.059)	0.375(0.001)	-0.498(0.062)	0.217(0.029)	0.991(0.028)	0.873(0.064)
	p-value	0.063(0.066)	0.314(0.001)	-0.504(0.068)	0.203(0.032)	1.102(0.031)	0.973(0.072)
12	AIC	0.109(0.066)	0.346(0.001)	-0.516(0.069)	0.217(0.032)	1.101(0.031)	0.982(0.073)
	BIC	0.142(0.066)	0.481(0.001)	-0.584(0.070)	0.246(0.033)	1.101(0.031)	0.983(0.074)
	ABIC	0.126(0.066)	0.426(0.001)	-0.556(0.069)	0.236(0.033)	1.101(0.031)	0.982(0.073)
	2 of 3	0.120(0.062)	0.409(0.001)	-0.529(0.066)	0.225(0.031)	1.040(0.029)	0.923(0.069)
	p-value	0.095(0.066)	0.334(0.001)	-0.510(0.069)	0.203(0.032)	1.101(0.031)	0.982(0.073)
13	AIC	0.076(0.066)	0.325(0.001)	-0.510(0.068)	0.215(0.032)	1.102(0.031)	0.969(0.071)
	BIC	0.119(0.066)	0.462(0.001)	-0.578(0.069)	0.250(0.033)	1.103(0.031)	0.969(0.071)
	ABIC	0.104(0.066)	0.409(0.001)	-0.552(0.069)	0.238(0.033)	1.102(0.031)	0.971(0.071)
	2 of 3	0.103(0.059)	0.374(0.001)	-0.498(0.062)	0.217(0.029)	0.991(0.028)	0.873(0.064)
	p-value	0.063(0.066)	0.313(0.001)	-0.504(0.068)	0.203(0.032)	1.102(0.031)	0.973(0.071)
14	AIC	0.087(0.066)	0.323(0.001)	-0.510(0.068)	0.209(0.032)	1.105(0.031)	0.961(0.070)
	BIC	0.124(0.066)	0.474(0.001)	-0.585(0.070)	0.244(0.032)	1.105(0.031)	0.959(0.070)
	ABIC	0.111(0.066)	0.414(0.001)	-0.555(0.069)	0.232(0.032)	1.105(0.031)	0.960(0.070)
	2 of 3	0.102(0.061)	0.385(0.001)	-0.512(0.063)	0.214(0.029)	1.011(0.028)	0.877(0.065)
	p-value	0.072(0.066)	0.315(0.001)	-0.506(0.068)	0.195(0.032)	1.105(0.031)	0.961(0.070)

ตารางที่ 4.4 ค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของค่าพารามิเตอร์ที่สกัดจากโมเดลที่ดีที่สุด (best model) เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ (ต่อ)

Con	MSC	PEST					
		CLADIFF(γ_2)	3WAYINT(γ_{7i})	ILDIF(γ_{5i})	ILADIFF(γ_1)	IAWC($\theta_j^{(2)}$)	CA($\theta_k^{(3)}$)
		Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)	Mean(SD)
15	AIC	0.083(0.066)	0.342(0.001)	-0.517(0.068)	0.213(0.032)	1.101(0.031)	0.986(0.072)
	BIC	0.123(0.066)	0.468(0.001)	-0.580(0.069)	0.244(0.033)	1.102(0.031)	0.986(0.072)
	ABIC	0.111(0.066)	0.423(0.001)	-0.557(0.068)	0.233(0.033)	1.102(0.031)	0.988(0.072)
	2 of 3	0.108(0.059)	0.385(0.001)	-0.504(0.062)	0.212(0.029)	0.998(0.028)	0.898(0.065)
	p-value	0.074(0.066)	0.333(0.001)	-0.512(0.068)	0.201(0.032)	1.101(0.031)	0.986(0.072)
16	AIC	0.094(0.066)	0.331(0.001)	-0.512(0.068)	0.212(0.032)	1.104(0.031)	0.970(0.073)
	BIC	0.128(0.066)	0.476(0.001)	-0.584(0.069)	0.247(0.033)	1.104(0.031)	0.972(0.073)
	ABIC	0.118(0.066)	0.420(0.001)	-0.556(0.069)	0.238(0.032)	1.104(0.031)	0.972(0.073)
	2 of 3	0.106(0.060)	0.392(0.001)	-0.513(0.062)	0.218(0.029)	1.004(0.028)	0.889(0.066)
	p-value	0.080(0.066)	0.318(0.001)	-0.506(0.068)	0.202(0.032)	1.104(0.031)	0.965(0.073)
17	AIC	0.105(0.067)	0.339(0.001)	-0.525(0.069)	0.216(0.033)	1.103(0.031)	0.956(0.070)
	BIC	0.147(0.066)	0.472(0.001)	-0.591(0.070)	0.248(0.033)	1.104(0.031)	0.961(0.071)
	ABIC	0.134(0.066)	0.419(0.001)	-0.565(0.070)	0.238(0.033)	1.103(0.031)	0.961(0.071)
	2 of 3	0.118(0.060)	0.391(0.001)	-0.520(0.064)	0.217(0.030)	1.008(0.028)	0.881(0.065)
	p-value	0.095(0.067)	0.325(0.001)	-0.518(0.069)	0.204(0.032)	1.103(0.031)	0.957(0.071)
18	AIC	0.097(0.066)	0.334(0.001)	-0.516(0.068)	0.221(0.032)	1.104(0.031)	0.957(0.071)
	BIC	0.132(0.066)	0.454(0.001)	-0.575(0.069)	0.249(0.033)	1.104(0.031)	0.955(0.071)
	ABIC	0.121(0.066)	0.413(0.001)	-0.555(0.069)	0.241(0.032)	1.104(0.031)	0.956(0.071)
	2 of 3	0.111(0.063)	0.392(0.001)	-0.529(0.065)	0.229(0.031)	1.049(0.029)	0.908(0.067)
	p-value	0.080(0.067)	0.318(0.001)	-0.508(0.068)	0.209(0.032)	1.104(0.031)	0.954(0.071)

จากตารางที่ 4.4 ผู้วิจัยสรุปผลทีละเงื่อนไข ดังนี้ เงื่อนไขที่ 1 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์

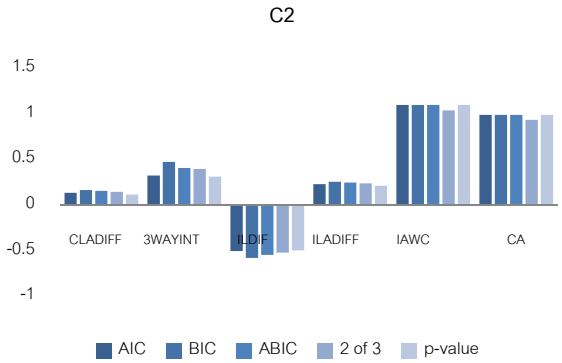
ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ ดัชนี p-value เป็นเกณฑ์เป็นเพียงค่าเดียวที่ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) มีค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value เป็นเกณฑ์ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่ต่ำกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



แผนภาพที่ 4.37 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 1 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 2 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion)และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size

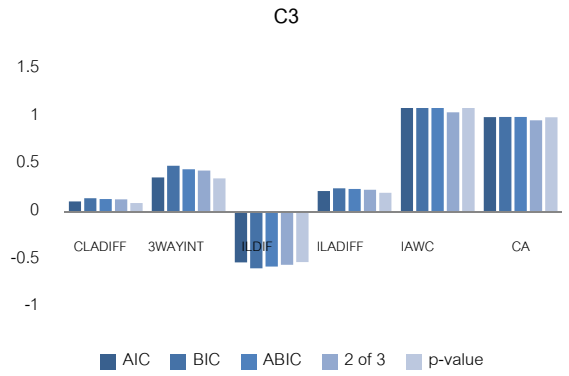
Adjusted Bayesian information criterion) และดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) มีค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3, ดัชนี p-value เป็นเกณฑ์ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์นั้น ได้ค่าประมาณที่ต่ำกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



แผนภาพที่ 4.38 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 2 หลังการใช้เกณฑ์การคัดสรรโมเดล

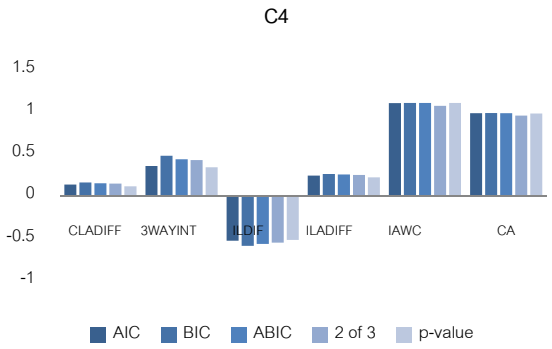
เงื่อนไขที่ 3 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และ ดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) มีค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value เป็นเกณฑ์ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean)

ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่ต่ำกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



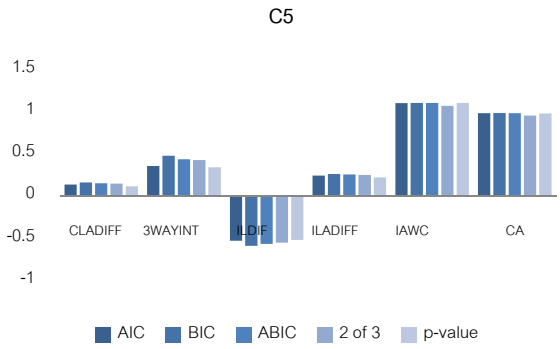
แผนภาพที่ 4.39 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 3 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 4 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



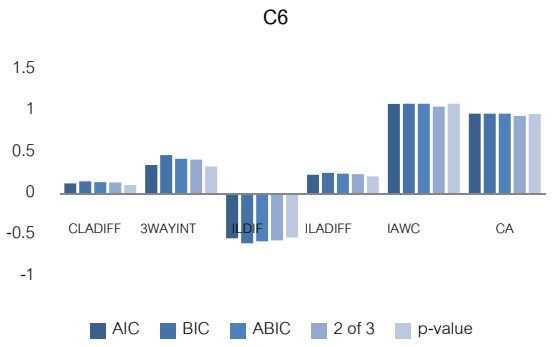
แผนภาพที่ 4.40 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 4 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 5 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันและพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) มีค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3, ดัชนี p-value เป็นเกณฑ์ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion) เป็นเกณฑ์นั้นต่ำกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



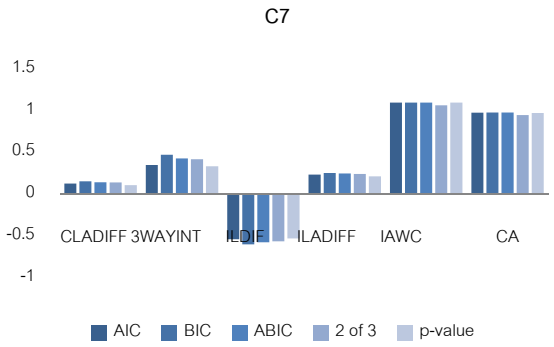
แผนภาพที่ 4.41 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 5 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 6 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) มีค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value เป็นเกณฑ์ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น ต่ำกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



แผนภาพที่ 4.42 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 6 หลังการใช้เกณฑ์การคัดสรรโมเดล

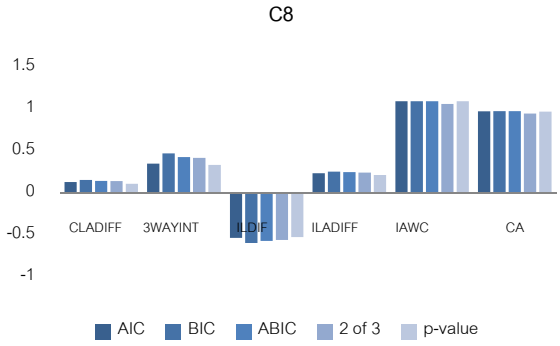
เงื่อนไขที่ 7 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3 สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) มีค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value เป็นเกณฑ์ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และดัชนี 2 of 3 เป็นเกณฑ์นั้น ต่ำกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



แผนภาพที่ 4.43 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 7 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 8 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์

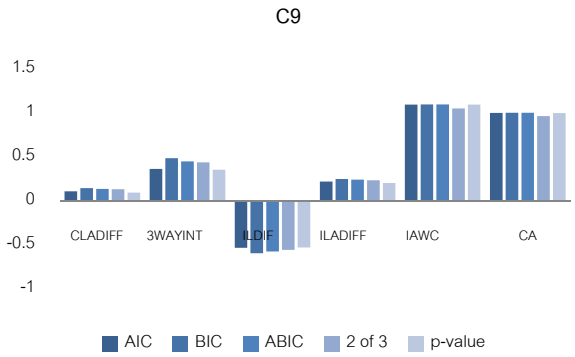
ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3 สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) มีค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value เป็นเกณฑ์ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น ต่ำกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



แผนภาพที่ 4.44 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 8 หลังการใช้เกณฑ์การคัดสรรโมเดล

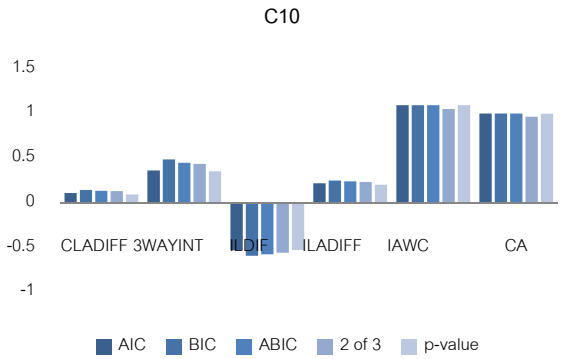
เงื่อนไขที่ 9 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันและพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) มีค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value เป็นเกณฑ์ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean)

ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น ต่ำกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล ดังแผนภาพ



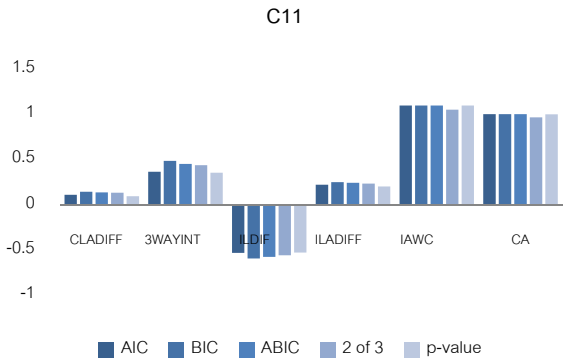
แผนภาพที่ 4.45 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 9 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 10 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี 2 of 3 และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion)และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)เป็นเกณฑ์นั้น สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



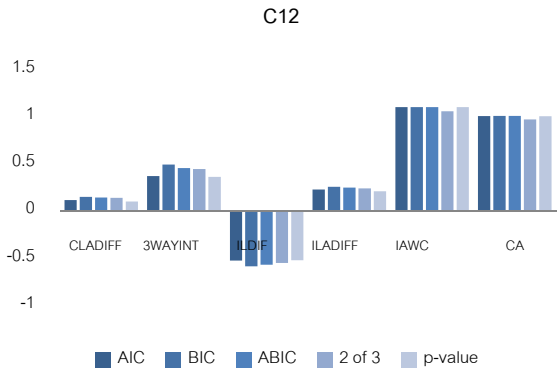
แผนภาพที่ 4.46 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 10 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 11 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี 2 of 3 และ ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion)และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)เป็นเกณฑ์นั้น สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล ดังแผนภาพ



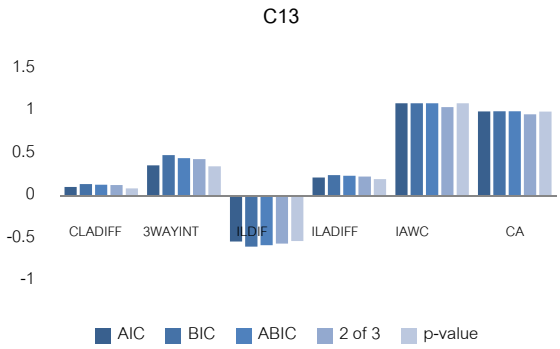
แผนภาพที่ 4.47 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 11 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 12 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)และดัชนี 2 of 3 เป็นเกณฑ์นั้น สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี 2 of 3 และ ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion)และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์นั้น สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล ดังแผนภาพ



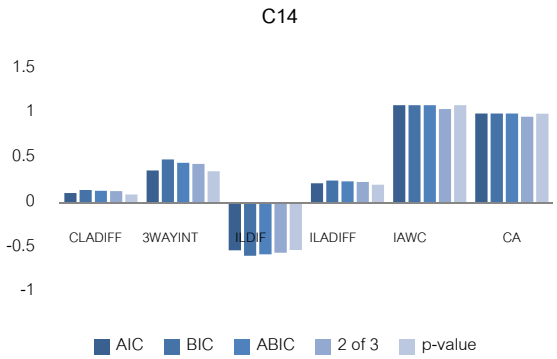
แผนภาพที่ 4.48 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 12 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 13 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี 2 of 3 และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion)และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



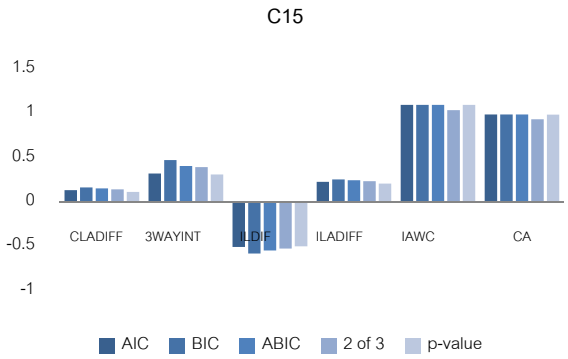
แผนภาพที่ 4.49 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 13 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 14 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี 2 of 3 และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion)และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



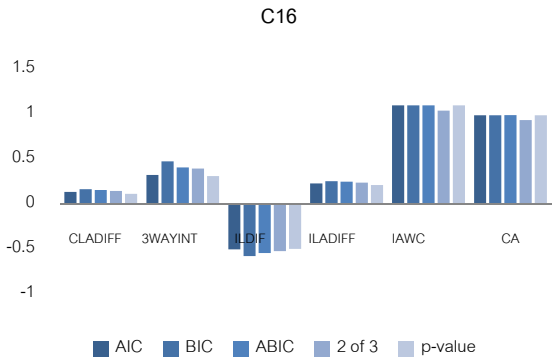
แผนภาพที่ 4.50 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 14 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 15 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3 เป็นเกณฑ์นั้น สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันและพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี 2 of 3 และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion)และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์นั้น สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกัน ดังแผนภาพ



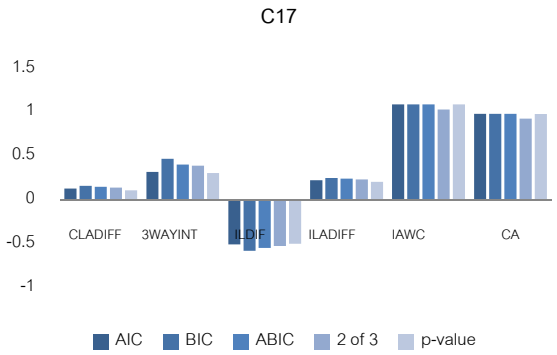
แผนภาพที่ 4.51 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 15 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 16 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี 2 of 3 และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion) และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



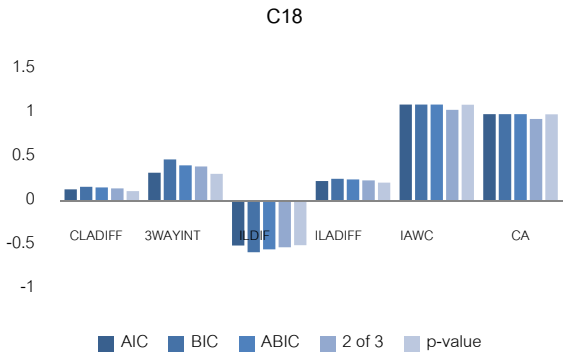
แผนภาพที่ 4.52 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 16 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 17 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี 2 of 3 และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion)และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



แผนภาพที่ 4.53 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 17 หลังการใช้เกณฑ์การคัดสรรโมเดล

เงื่อนไขที่ 18 ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3 เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion), ดัชนี 2 of 3 และดัชนี p-value ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion)และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion)เป็นเกณฑ์นั้น ได้ค่าประมาณที่สูงกว่าค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังแผนภาพ



แผนภาพที่ 4.54 ผลการประมาณค่าพารามิเตอร์ของเงื่อนไขที่ 18 หลังการใช้เกณฑ์การคัดสรรโมเดล

จากผลที่ได้สรุปได้ว่า ทั้ง 18 เงื่อนไขได้ผลที่สอดคล้องกัน คือ ค่าเฉลี่ย (mean) ของค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) ที่ประมาณค่าได้จากทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) ใกล้เคียงกับค่าจริง (true value) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) มีค่าใกล้เคียงกัน ซึ่งสอดคล้องกับค่าเฉลี่ย (mean) และค่าส่วนเบี่ยงเบนมาตรฐาน (SD) ของพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น ค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value เป็นเกณฑ์ได้ค่าเฉลี่ย (mean) ที่ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และดัชนี 2 of 3 ได้ค่าประมาณที่แตกต่างจากค่าจริง (true value) เล็กน้อยแต่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) มีค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี AIC (Akaike's information criterion) ดัชนี 2 of 3 และดัชนี p-value เป็นเกณฑ์ใกล้เคียงกับค่าจริง (true value) ส่วนค่าเฉลี่ย (mean) ที่ประมาณได้จากการใช้ดัชนี BIC (Bayesian information criterion) และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) นั้น แตกต่างจากค่าจริง (true value) เล็กน้อยในขณะที่ส่วนเบี่ยงเบนมาตรฐาน (SD) ใกล้เคียงกันในทั้ง 5 เกณฑ์การคัดสรรโมเดล (Model selection criteria)

สรุปได้ว่าจากทั้ง 18 เงื่อนไข ได้ข้อสรุปที่สอดคล้องกันในทุกเงื่อนไข นั่นคือ ค่าพารามิเตอร์ที่ได้ค่าเฉลี่ยแตกต่างจากค่าจริงมี 2 ค่า คือ พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ซึ่งสอดคล้องกับผลการวิเคราะห์ก่อนใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ที่พบว่า ทั้ง 18 เงื่อนไขได้ค่าเฉลี่ยของพารามิเตอร์

ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง (true value) ในทุกเงื่อนไข

เนื่องจากการศึกษาในครั้งนี้มีวัตถุประสงค์ข้อแรกเพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม ซึ่งจากตารางที่ 4.4 และแผนภาพจะสามารถตอบวัตถุประสงค์ข้อแรกได้ว่า โมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) นี้สามารถประมาณค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) ได้ใกล้เคียงกับค่าจริง (true value) แต่ได้ค่าประมาณของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง (true value) ในทุกเงื่อนไข

ตอนที่ 4 ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของแต่ละพารามิเตอร์

การศึกษาในครั้งนี้ ผู้วิจัยมีเกณฑ์การประเมิน (evaluation criteria) ในการประมาณค่าพารามิเตอร์ 3 เกณฑ์ คือ 1) ค่าความลำเอียง (BIAS) 2) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และ 3) ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ดังนั้นในตอนที่ 4 นี้ ผู้วิจัยขอนำเสนอค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของแต่ละพารามิเตอร์ทั้ง 18 เงื่อนไขแบ่งตามเกณฑ์การคัดสรรโมเดล (Model selection criteria) ดังตารางที่ 4.5

ตารางที่ 4.5 ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของแต่ละพารามิเตอร์ ตามเกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์

MSC	Con	PEST					
		CLADIFF(γ_2)	3WAYINT(γ_{7i})	ILDIF(γ_{5i})	ILADIFF(γ_1)	IAWC($\theta_j^{(2)}$)	CA($\theta_k^{(3)}$)
		BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE
AIC	1	0.030(0.106)0.012	0.051(0.001)0.003	-0.034(0.099)0.011	0.036(0.046)0.003	0.094(0.044)0.011	-0.026(0.102)0.011
	2	0.035(0.106)0.013	0.022(0.001)0.002	-0.004(0.098)0.010	0.028(0.046)0.003	0.095(0.045)0.011	-0.014(0.101)0.010
	3	0.012(0.106)0.011	0.065(0.001)0.004	-0.029(0.097)0.010	0.022(0.046)0.002	0.095(0.044)0.011	-0.001(0.104)0.001
	4	0.024(0.107)0.012	0.037(0.000)0.003	-0.037(0.097)0.012	0.017(0.046)0.002	0.106(0.045)0.013	-0.046(0.098)0.011
	5	0.034(0.106 0.012	0.022(0.001)0.001	-0.003(0.098)0.009	0.028(0.046)0.003	0.094(0.044)0.011	-0.014(0.101)0.010
	6	0.011(0.106)0.011	0.065(0.001)0.004	-0.028(0.097)0.010	0.022(0.046)0.002	0.094(0.044)0.011	-0.001(0.103)0.010
	7	0.025(0.107)0.012	0.043(0.001)0.002	-0.038(0.097)0.011	0.016(0.046)0.002	0.104(0.045)0.013	-0.045(0.098)0.011
	8	0.001(0.106)0.011	0.039(0.001)0.001	-0.027(0.098)0.010	0.011(0.046)0.002	0.095(0.044)0.011	0.015(0.103)0.010
	9	0.018(0.106)0.011	0.079(0.001)0.006	-0.039(0.098)0.011	0.023(0.046)0.002	0.103(0.044)0.012	0.010(0.104)0.011
	10	-0.018(0.066)0.004	0.032(0.001)0.002	-0.024(0.068)0.005	0.015(0.032)0.001	0.099(0.031)0.010	-0.034(0.071)0.006
	11	0.023(0.066)0.005	0.025(0.001)0.001	-0.010(0.068)0.004	0.015(0.032)0.001	0.102(0.031)0.011	-0.030(0.071)0.006
	12	0.009(0.066)0.004	0.046(0.001)0.002	-0.016(0.069)0.006	0.017(0.032)0.001	0.101(0.031)0.011	-0.017(0.073)0.005
	13	-0.023(0.066)0.004	0.025(0.001)0.001	-0.010(0.068)0.004	0.015(0.032)0.002	0.102(0.031)0.011	-0.030(0.071)0.006
	14	-0.012(0.066)0.004	0.023(0.001)0.002	-0.010(0.068)0.004	0.009(0.032)0.001	0.105(0.031)0.012	-0.038(0.070)0.006
	15	-0.016(0.066)0.004	0.042(0.001)0.001	-0.017(0.068)0.005	0.013(0.032)0.001	0.102(0.031)0.011	-0.013(0.072)0.005
	16	-0.005(0.066)0.003	0.031(0.001)0.002	-0.012(0.068)0.004	0.012(0.032)0.002	0.104(0.031)0.011	-0.029(0.073)0.006
	17	0.005(0.067)0.004	0.039(0.001)0.002	-0.025(0.069)0.005	0.016(0.033)0.001	0.103(0.031)0.011	-0.043(0.070)0.006
	18	-0.002(0.066)0.003	0.034(0.001)0.002	-0.016(0.068)0.005	0.021(0.032)0.001	0.104(0.031)0.012	-0.042(0.071)0.006
BIC	1	0.054(0.106)0.014	0.173(0.001)0.029	-0.095(0.099)0.019	0.056(0.047)0.005	0.095(0.044)0.011	-0.025(0.102)0.011
	2	0.065(0.106)0.015	0.173(0.001)0.029	-0.078(0.099)0.016	0.054(0.047)0.005	0.095(0.045)0.011	-0.014(0.101)0.010
	3	0.046(0.106)0.013	0.187(0.001)0.035	-0.089(0.098)0.017	0.052(0.046)0.005	0.095(0.044)0.011	-0.001(0.104)0.001
	4	0.062(0.107)0.015	0.181(0.001)0.032	-0.105(0.098)0.021	0.048(0.047)0.004	0.106(0.045)0.013	-0.046(0.098)0.011
	5	0.064(0.106)0.015	0.173(0.001)0.030	-0.078(0.099)0.016	0.054(0.046)0.005	0.095(0.044)0.011	-0.014(0.101)0.010
	6	0.046(0.105)0.013	0.187(0.001)0.035	-0.089(0.098)0.017	0.051(0.046)0.004	0.094(0.044)0.011	-0.001(0.103)0.010
	7	0.063(0.107)0.015	0.180(0.001)0.032	-0.106(0.098)0.020	0.048(0.047)0.004	0.104(0.045)0.013	-0.045(0.098)0.011
	8	0.037(0.105)0.012	0.178(0.001)0.032	-0.097(0.099)0.019	0.043(0.046)0.004	0.095(0.044)0.011	0.015(0.103)0.010
	9	0.052(0.106)0.014	0.177(0.001)0.031	-0.087(0.098)0.017	0.053(0.047)0.005	0.103(0.044)0.012	0.010(0.104)0.011
	10	0.023(0.066)0.004	0.173(0.001)0.029	-0.094(0.070)0.014	0.050(0.033)0.003	0.099(0.031)0.010	-0.032(0.071)0.006
	11	0.025(0.066)0.005	0.162(0.001)0.026	-0.079(0.069)0.011	0.050(0.033)0.003	0.103(0.031)0.011	-0.030(0.072)0.006
	12	0.042(0.066)0.006	0.181(0.001)0.033	-0.084(0.070)0.012	0.046(0.033)0.003	0.101(0.031)0.011	-0.017(0.074)0.005
	13	0.019(0.066)0.004	0.162(0.001)0.026	-0.078(0.069)0.011	0.050(0.033)0.003	0.102(0.031)0.011	-0.030(0.071)0.006
	14	0.024(0.066)0.005	0.174(0.001)0.030	-0.085(0.070)0.012	0.044(0.032)0.003	0.105(0.031)0.012	-0.040(0.070)0.006
	15	0.023(0.066)0.005	0.168(0.001)0.028	-0.080(0.069)0.011	0.044(0.033)0.003	0.102(0.031)0.011	-0.013(0.072)0.005
	16	0.028(0.066)0.005	0.176(0.001)0.031	-0.084(0.069)0.012	0.047(0.033)0.003	0.104(0.031)0.011	-0.027(0.073)0.006
	17	0.028(0.066)0.005	0.176(0.001)0.031	-0.084(0.069)0.012	0.047(0.033)0.003	0.104(0.031)0.011	-0.027(0.073)0.006
	18	0.032(0.066)0.005	0.154(0.001)0.023	-0.075(0.069)0.010	0.049(0.033)0.003	0.104(0.031)0.012	-0.044(0.071)0.007

ตารางที่ 4.5 ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของแต่ละพารามิเตอร์ ตามเกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ (ต่อ)

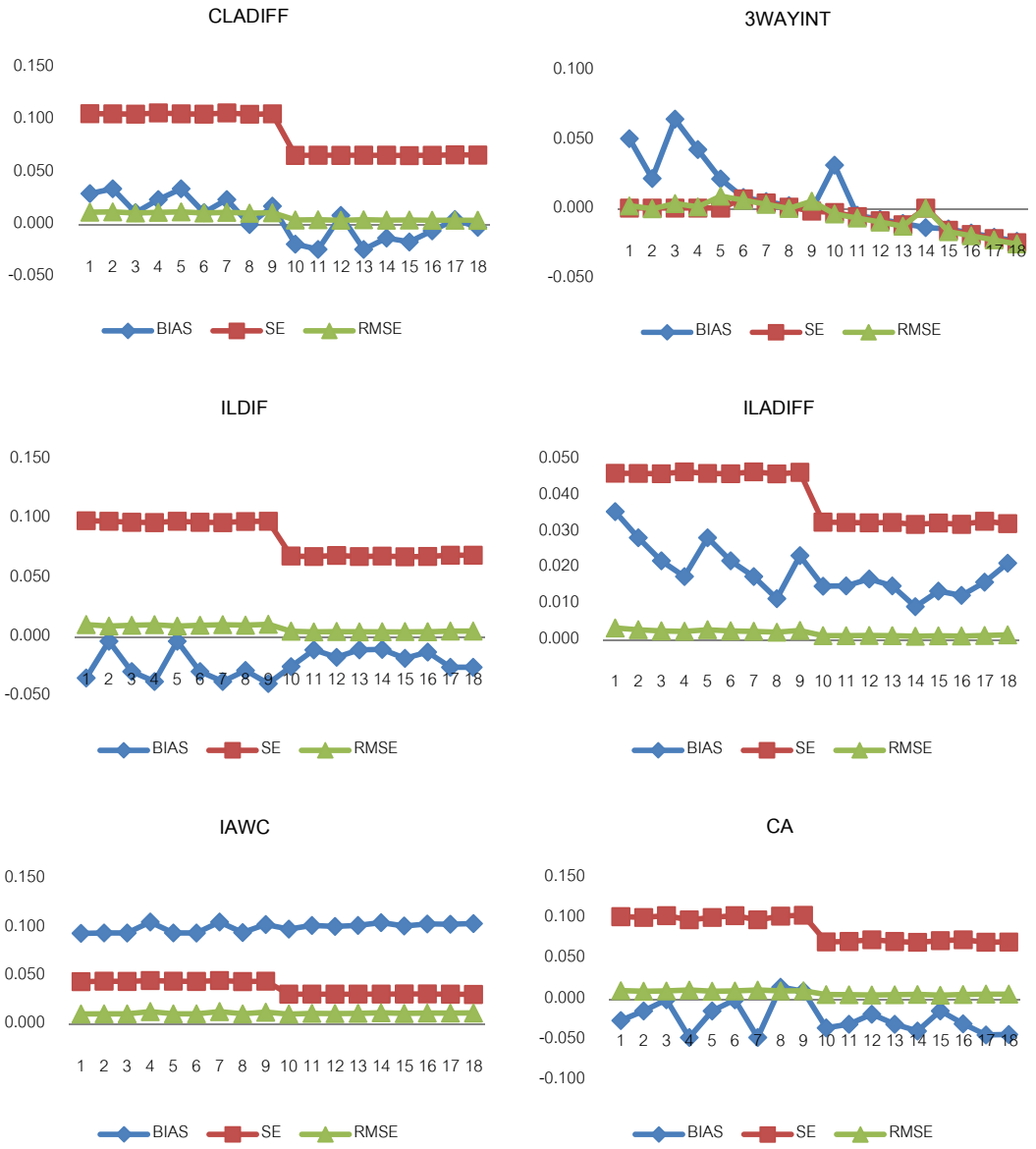
MSC	Con	PEST					
		CLADIFF(γ_2)	3WAYINT(γ_{7i})	ILDIF(γ_{5i})	ILADIFF(γ_1)	IAWC($\theta_j^{(2)}$)	CA($\theta_k^{(3)}$)
		BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE
ABIC	1	0.045(0.106)0.013	0.128(0.001)0.016	-0.073(0.099)0.015	0.048(0.046)0.004	0.094(0.044)0.011	-0.025(0.102)0.011
	2	0.055(0.106)0.014	0.107(0.001)0.011	-0.045(0.098)0.012	0.047(0.047)0.004	0.095(0.044)0.011	-0.013(0.101)0.010
	3	0.037(0.106)0.012	0.151(0.001)0.023	-0.071(0.098)0.015	0.043(0.046)0.004	0.095(0.045)0.011	-0.005(0.104)0.001
	4	0.047(0.107)0.013	0.133(0.001)0.018	-0.082(0.097)0.016	0.036(0.047)0.003	0.106(0.045)0.013	-0.046(0.098)0.012
	5	0.055(0.106)0.014	0.106(0.001)0.011	-0.045(0.098)0.012	0.046(0.046)0.004	0.094(0.045)0.011	-0.014(0.101)0.010
	6	0.037(0.105)0.012	0.151(0.001)0.022	-0.071(0.097)0.014	0.043(0.046)0.004	0.094(0.044)0.011	-0.001(0.103)0.010
	7	0.048(0.107)0.013	0.133(0.001)0.017	-0.084(0.097)0.016	0.036(0.047)0.003	0.104(0.045)0.013	-0.045(0.098)0.011
	8	0.024(0.105)0.011	0.129(0.001)0.016	-0.072(0.098)0.015	0.033(0.046)0.003	0.095(0.044)0.011	0.016(0.103)0.010
	9	0.041(0.106)0.013	0.146(0.001)0.021	-0.072(0.098)0.015	0.043(0.047)0.004	0.103(0.044)0.012	0.011(0.104)0.011
	10	0.007(0.066)0.004	0.122(0.001)0.014	-0.069(0.069)0.009	0.037(0.033)0.002	0.099(0.031)0.010	-0.034(0.071)0.006
	11	0.004(0.066)0.004	0.109(0.001)0.012	-0.052(0.069)0.007	0.038(0.033)0.002	0.103(0.031)0.011	-0.029(0.072)0.006
	12	0.026(0.066)0.005	0.126(0.001)0.015	-0.056(0.069)0.008	0.036(0.033)0.002	0.101(0.031)0.011	-0.017(0.073)0.005
	13	0.004(0.066)0.004	0.109(0.001)0.011	-0.052(0.069)0.007	0.038(0.033)0.002	0.102(0.031)0.011	-0.029(0.071)0.005
	14	0.011(0.066)0.004	0.114(0.001)0.013	-0.055(0.069)0.008	0.032(0.032)0.002	0.105(0.031)0.012	-0.039(0.070)0.006
	15	0.011(0.066)0.004	0.123(0.001)0.015	-0.057(0.068)0.008	0.033(0.033)0.002	0.102(0.031)0.011	-0.011(0.072)0.005
	16	0.018(0.066)0.004	0.120(0.001)0.014	-0.056(0.069)0.008	0.038(0.032)0.002	0.104(0.031)0.011	-0.028(0.073)0.006
	17	0.047(0.066)0.006	0.172(0.001)0.029	-0.091(0.070)0.013	0.048(0.033)0.003	0.103(0.031)0.011	-0.038(0.071)0.006
	18	0.021(0.066)0.004	0.113(0.001)0.012	-0.055(0.069)0.007	0.041(0.032)0.002	0.104(0.031)0.012	-0.043(0.071)0.007
2 of 3	1	0.042(0.103)0.012	0.119(0.001)0.014	-0.059(0.096)0.012	0.043(0.045)0.003	0.061(0.043)0.006	-0.052(0.099)0.013
	2	0.045(0.100)0.012	0.094(0.001)0.009	-0.022(0.093)0.009	0.036(0.044)0.003	0.036(0.042)0.003	-0.068(0.096)0.014
	3	0.035(0.102)0.012	0.139(0.001)0.019	-0.053(0.094)0.011	0.036(0.044)0.003	0.051(0.043)0.004	-0.035(0.099)0.011
	4	0.038(0.100)0.011	0.118(0.000)0.014	-0.055(0.091)0.013	0.025(0.044)0.002	0.040(0.042)0.003	-0.096(0.093)0.018
	5	0.044(0.100)0.012	0.094(0.001)0.009	-0.022(0.093)0.009	0.036(0.044)0.003	0.036(0.042)0.003	-0.068(0.095)0.014
	6	0.034(0.101)0.011	0.139(0.001)0.019	-0.052(0.093)0.011	0.035(0.044)0.003	0.050(0.042)0.004	-0.034(0.099)0.011
	7	0.038(0.100)0.011	0.118(0.001)0.014	-0.055(0.091)0.012	0.025(0.044)0.002	0.040(0.042)0.003	-0.096(0.093)0.018
	8	0.015(0.101)0.010	0.116(0.001)0.013	-0.051(0.094)0.011	0.023(0.044)0.002	0.045(0.042)0.003	0.029(0.098)0.011
	9	0.039(0.101)0.012	0.134(0.001)0.018	-0.051(0.094)0.011	0.034(0.045)0.003	0.059(0.042)0.005	-0.027(0.099)0.011
	10	0.001(0.058)0.003	0.076(0.001)0.005	-0.001(0.060)0.003	0.011(0.029)0.001	0.039(0.027)0.002	-0.015(0.062)0.028
	11	0.004(0.059)0.003	0.075(0.001)0.005	-0.002(0.062)0.004	0.017(0.029)0.001	0.008(0.028)0.001	-0.126(0.064)0.020
	12	0.020(0.062)0.005	0.109(0.001)0.012	-0.029(0.066)0.006	0.025(0.031)0.001	0.040(0.029)0.002	-0.076(0.069)0.010
	13	0.004(0.059)0.003	0.074(0.001)0.005	-0.001(0.062)0.003	0.017(0.029)0.002	0.008(0.028)0.001	-0.126(0.064)0.020
	14	0.002(0.061)0.003	0.085(0.001)0.007	-0.012(0.063)0.004	0.014(0.029)0.001	0.011(0.028)0.001	-0.122(0.064)0.019
	15	0.008(0.059)0.003	0.085(0.001)0.007	-0.004(0.062)0.003	0.012(0.029)0.001	0.002(0.028)0.001	-0.101(0.065)0.014
	16	0.006(0.060)0.003	0.092(0.001)0.008	-0.013(0.062)0.004	0.018(0.029)0.001	0.004(0.028)0.008	-0.110(0.066)0.016
	17	0.018(0.060)0.004	0.091(0.001)0.008	-0.020(0.064)0.004	0.017(0.030)0.001	0.008(0.028)0.001	-0.118(0.065)0.018
	18	0.011(0.063)0.004	0.092(0.001)0.008	-0.029(0.065)0.005	0.029(0.031)0.001	0.049(0.029)0.003	-0.091(0.067)0.012

ตารางที่ 4.5 ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของแต่ละพารามิเตอร์ ตามเกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ (ต่อ)

MSC	Con	PEST					
		CLADIFF(γ_2)	3WAYINT(γ_{7i})	ILDIF(γ_{5i})	ILADIFF(γ_1)	IAWC($\theta_j^{(2)}$)	CA($\theta_k^{(3)}$)
		BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE	BIAS(SE)RMSE
	1	0.010(0.107)0.011	0.034(0.001)0.001	-0.025(0.098)0.010	0.014(0.046)0.002	0.094(0.044)0.011	-0.029(0.102)0.011
	2	0.014(0.106)0.012	0.011(0.001)0.001	-0.003(0.098)0.009	0.010(0.046)0.002	0.095(0.045)0.011	-0.015(0.101)0.010
	3	0.004(0.106)0.009	0.056(0.001)0.003	-0.024(0.097)0.010	0.005(0.046)0.002	0.095(0.044)0.011	-0.002(0.103)0.001
	4	0.007(0.107)0.011	0.030(0.000)0.001	-0.030(0.097)0.010	0.001(0.046)0.001	0.106(0.045)0.013	-0.046(0.098)0.011
	5	0.014(0.106)0.011	0.010(0.001)0.001	-0.003(0.098)0.009	0.010(0.046)0.002	0.094(0.045)0.011	-0.014(0.101)0.010
	6	0.003(0.106)0.011	0.055(0.001)0.003	-0.024(0.097)0.009	0.004(0.045)0.001	0.094(0.044)0.011	-0.001(0.103)0.010
	7	0.007(0.107)0.010	0.027(0.001)0.001	-0.029(0.097)0.010	0.001(0.046)0.001	0.104(0.045)0.013	-0.045(0.098)0.011
	8	0.016(0.106)0.011	0.021(0.001)0.001	-0.018(0.098)0.009	0.003(0.046)0.001	0.095(0.044)0.011	0.012(0.102)0.010
p-value	9	0.002(0.106)0.010	0.068(0.001)0.004	-0.033(0.098)0.010	0.009(0.046)0.001	0.103(0.044)0.012	0.008(0.104)0.009
	10	-0.033(0.066)0.005	0.016(0.001)0.001	-0.017(0.068)0.005	0.001(0.032)0.001	0.099(0.031)0.010	-0.037(0.071)0.006
	11	0.036(0.066)0.006	0.013(0.001)0.001	-0.004(0.068)0.004	0.003(0.032)0.001	0.102(0.031)0.011	-0.026(0.072)0.006
	12	-0.005(0.066)0.004	0.034(0.001)0.001	-0.010(0.069)0.005	0.003(0.032)0.001	0.101(0.031)0.011	-0.017(0.073)0.005
	13	-0.036(0.066)0.005	0.013(0.001)0.001	-0.004(0.068)0.004	0.003(0.032)0.001	0.102(0.031)0.011	-0.026(0.071)0.005
	14	-0.027(0.066)0.005	0.015(0.001)0.001	-0.006(0.068)0.004	0.004(0.032)0.001	0.105(0.031)0.012	-0.037(0.070)0.006
	15	-0.025(0.066)0.005	0.033(0.001)0.001	-0.012(0.068)0.004	0.001(0.032)0.001	0.102(0.031)0.011	-0.013(0.072)0.005
	16	-0.019(0.066)0.004	0.018(0.001)0.001	-0.006(0.068)0.003	0.002(0.032)0.001	0.104(0.031)0.011	-0.034(0.073)0.006
	17	-0.004(0.067)0.004	0.025(0.001)0.001	-0.018(0.069)0.005	0.004(0.032)0.001	0.103(0.031)0.011	-0.043(0.071)0.006
	18	-0.019(0.067)0.004	0.018(0.001)0.001	-0.008(0.068)0.004	0.009(0.032)0.001	0.104(0.031)0.012	-0.045(0.071)0.007

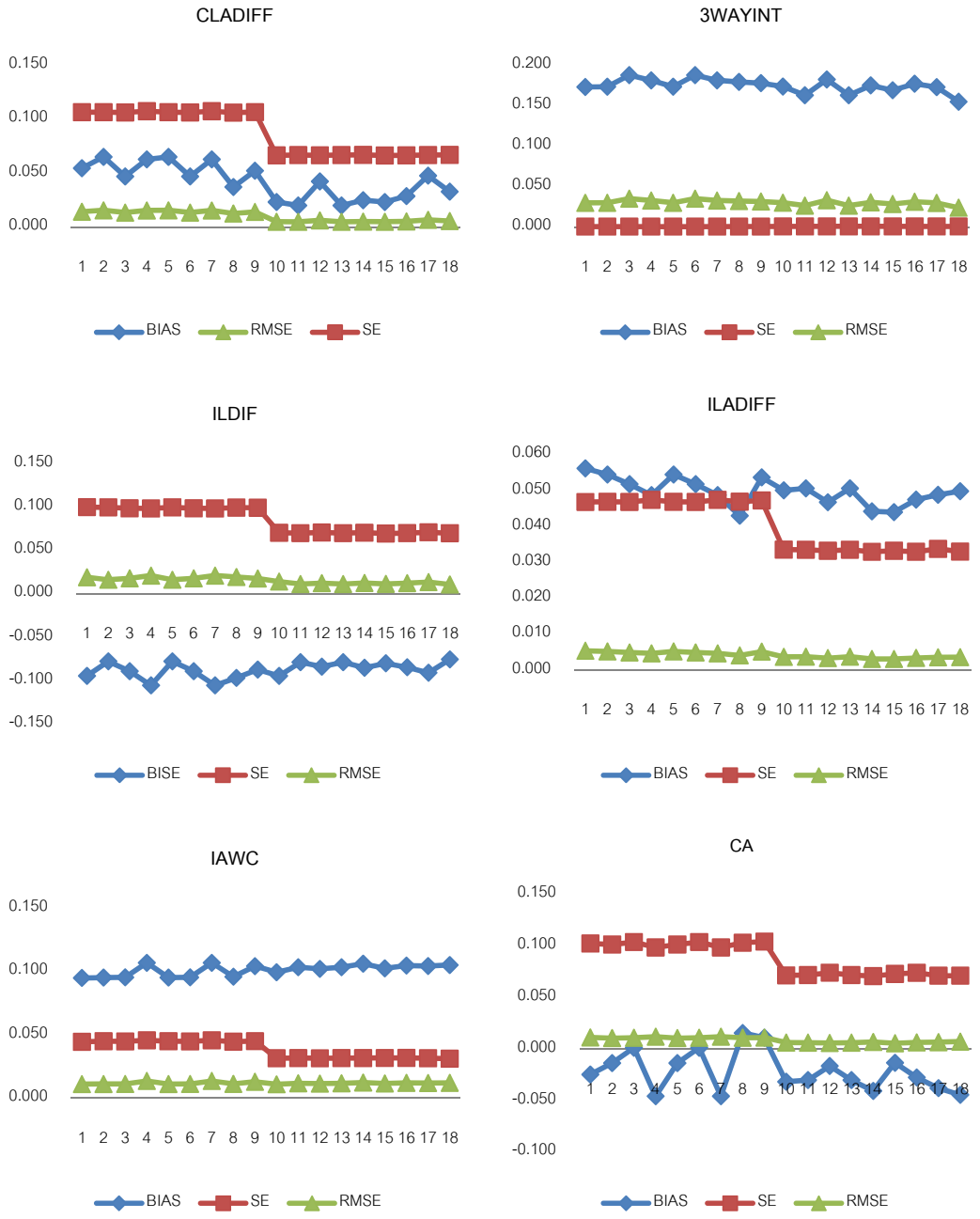
จากตารางที่ 4.5 สรุปได้ว่า เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (model selection criteria) สำหรับค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) พบว่า ค่าความลำเอียง (BIAS) มีค่าใกล้เคียงกันในทุกเงื่อนไข สำหรับค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) นั้น พบว่า เงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) จะลดลง ซึ่งสอดคล้องกับพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้น พบว่า ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) มีค่าใกล้เคียงกันในทุกเงื่อนไขการจำลองและพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) นั้น พบว่า ค่าความลำเอียง (BIAS) และค่ารากที่สองของความคลาดเคลื่อน

ยกกำลังสองเฉลี่ย (RMSE) ใกล้เคียงกันในทุกเงื่อนไข แต่ค่าความคลาดเคลื่อนมาตรฐาน (SE) ของเงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความคลาดเคลื่อนมาตรฐาน (SE) จะลดลง ดังแผนภาพ



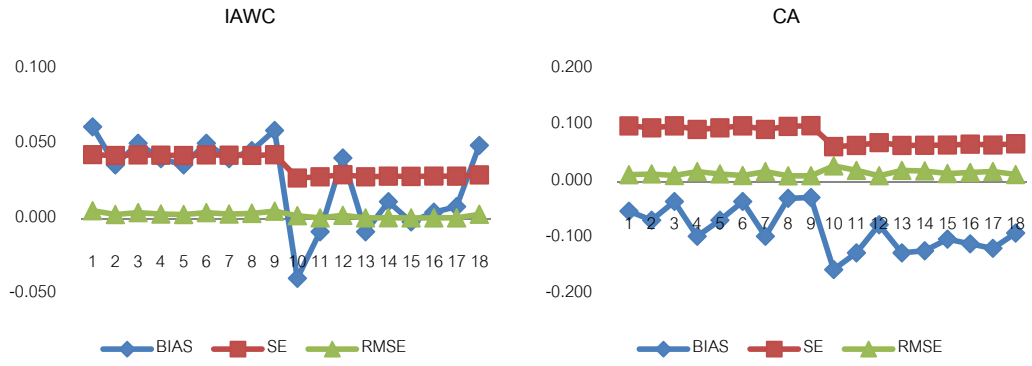
แผนภาพที่ 4.55 ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของทุกเงื่อนไข เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria)

เมื่อใช้ดัชนี BIC (Bayesian information criterion) และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า ได้ผลสอดคล้องกันในทุกเงื่อนไข นั่นคือ สำหรับค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) พบว่า ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) นั้น เงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) จะลดลง พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) พบว่า ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) มีค่าใกล้เคียงกันในทุกเงื่อนไข สำหรับพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) และพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) นั้นได้ผลที่สอดคล้องกัน นั่นคือ ค่าความลำเอียง (BIAS) ใกล้เคียงกันในทุกเงื่อนไข ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) นั้น เงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) จะลดลง สำหรับพารามิเตอร์ความสามารถระดับกลุ่ม (CA) พบว่า ค่าความลำเอียง (BIAS) ของเงื่อนไขที่ 3, 6, 9, 12, 15, 18 ได้ค่าที่ใกล้เคียงกันและมีค่าต่ำมาก ในขณะที่เงื่อนไขอื่นๆ ได้ค่าใกล้เคียงกัน ในขณะที่ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของเงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) จะลดลง ดังแผนภาพ



แผนภาพที่ 4.56 ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของทุกเงื่อนไข เมื่อใช้ดัชนี BIC (Bayesian information criterion) และดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria)

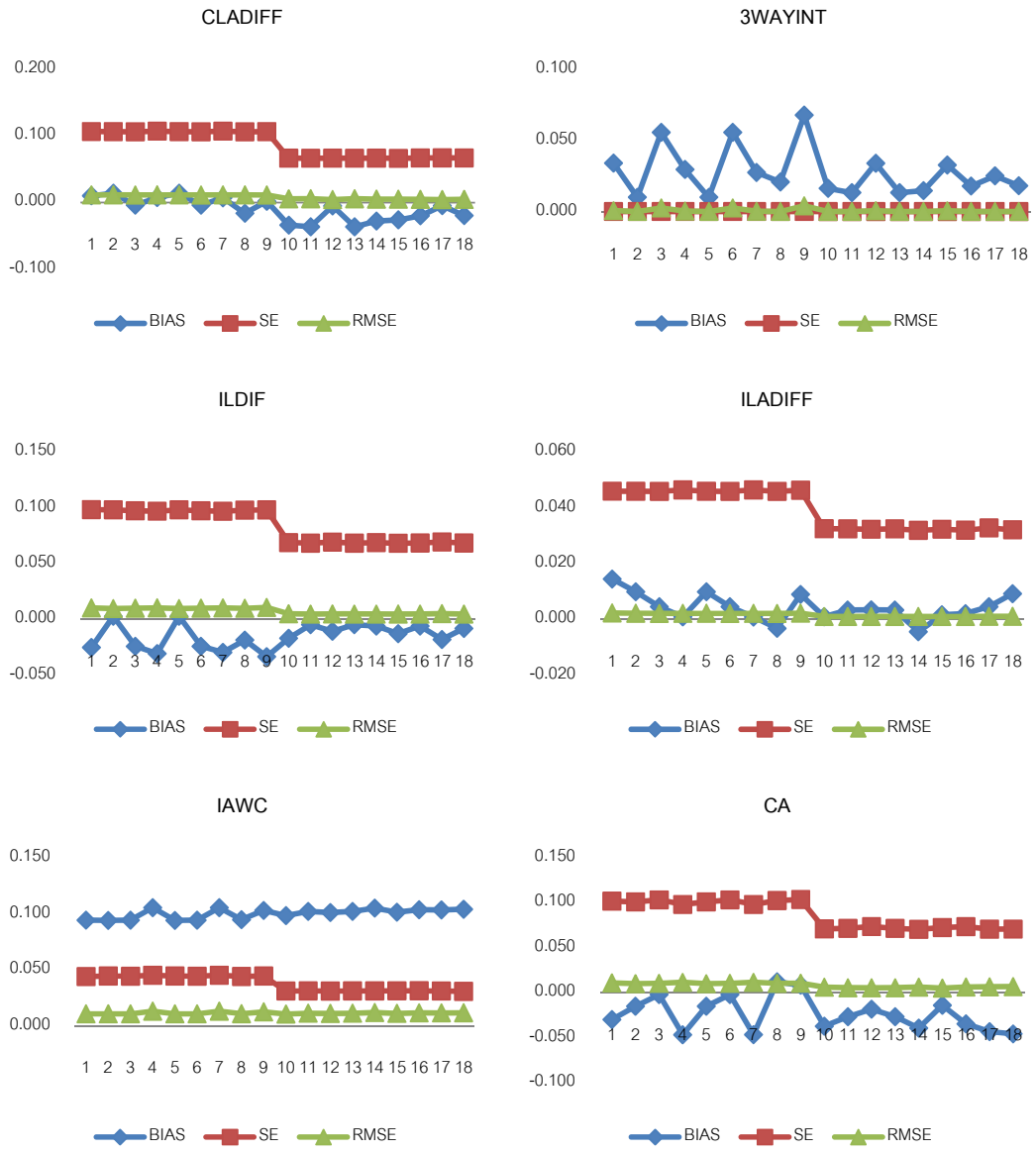
เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า สำหรับค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) พบว่า ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) นั้น เงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) จะลดลง พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) พบว่า ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) มีค่าใกล้เคียงกันในทุกเงื่อนไขการจำลอง สำหรับพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ พารามิเตอร์ความสามารถระดับกลุ่ม (CA) นั้นได้ผลที่สอดคล้องกัน นั่นคือ ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) นั้น เงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) จะลดลง ดังแผนภาพ



แผนภาพที่ 4.57 ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของทุกเงื่อนไข เมื่อใช้ดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria)

เมื่อใช้ดัชนี p - value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า สำหรับค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) พบว่า ค่าความลำเอียง (BIAS) มีค่าใกล้เคียงกันในทุกเงื่อนไขการจำลอง ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) นั้น เงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) จะลดลง พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) พบว่า ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) มีค่าใกล้เคียงกันในทุกเงื่อนไขการจำลอง สำหรับพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) นั้น ค่าความลำเอียง (BIAS) มีค่าใกล้เคียงกันในทุกเงื่อนไขการจำลอง ค่าความคลาดเคลื่อนมาตรฐาน (SE) ในเงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความคลาดเคลื่อนมาตรฐาน (SE) จะลดลง ในขณะที่ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) มีค่าใกล้เคียงกันในทุกเงื่อนไขการจำลอง พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) พบว่า ค่าความลำเอียง (BIAS) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ใกล้เคียงกันในทุกเงื่อนไขการจำลอง ในขณะที่ค่าความคลาดเคลื่อนมาตรฐาน (SE) ในเงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความคลาดเคลื่อนมาตรฐาน (SE) จะลดลงและ พารามิเตอร์ความสามารถระดับกลุ่ม (CA) นั้น ค่า

ความลำเอียง (BIAS) ใกล้เคียงกันในทุกเงื่อนไข แต่ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ในเงื่อนไขที่ 1 - 9 ได้ค่าใกล้เคียงกันและเงื่อนไขที่ 10 - 18 ก็ได้ค่าใกล้เคียงกันเช่นกัน นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าความคลาดเคลื่อนมาตรฐาน (SE) จะลดลง ผู้วิจัยขอนำเสนอด้วยแผนภาพ ดังนี้



แผนภาพที่ 4.58 ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ของทุกเงื่อนไข เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria)

บทที่ 5

สรุปผล อภิปรายผลและข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม และเพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)

การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลด้วยเงื่อนไขที่แตกต่างกันจำนวน 18 เงื่อนไข (2 x 3 x 3) อันประกอบไปด้วย 1) จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school or the number of individual in each cluster) 2 ระดับ คือ โรงเรียนละ 50 คนและโรงเรียนละ 100 คน 2) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 3 ระดับ คือ 0.2, 0.4 และ 0.6 และ 3) ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ คือ 0.1, 0.2 และ 0.3 ส่วนจำนวนโรงเรียนนั้นกำหนดคงที่ทุกเงื่อนไขที่ 50 โรงเรียน โดยในแต่ละเงื่อนไขกำหนดการทำซ้ำเงื่อนไขละ 200 รอบ

ผู้วิจัยมีโมเดลการทำหน้าที่ต่างกันของข้อสอบที่ศึกษาในครั้งนี้ 4 โมเดล ได้แก่ 1) โมเดลเต็มรูป (complete model) ซึ่งเป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) ทั้ง 7 อิทธิพลและอิทธิพลสุ่ม (random effect) 2 อิทธิพล 2) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล 3) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 6 อิทธิพล นั่นคือ กำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพลและ 4) โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ประกอบด้วยอิทธิพลกำหนด (fixed effect) 5 อิทธิพล นั่นคือ กำหนดพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์และมีอิทธิพลสุ่ม (random effect) 2 อิทธิพล

พารามิเตอร์เป้าหมายในการประมาณค่ามี 6 พารามิเตอร์ ได้แก่ 1) พารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) 2) พารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) 3) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) 4) พารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) 5) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ 6) พารามิเตอร์ความสามารถระดับกลุ่ม (CA)

ผู้วิจัยมีเกณฑ์การคัดสรรโมเดล (Model selection criteria) 5 เกณฑ์ ได้แก่ 1) ดัชนี AIC (Akaike's information criterion), 2) ดัชนี BIC (Bayesian information criterion), 3) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), 4) ดัชนี 2 of 3 และ 5) ดัชนี p-value และมีเกณฑ์การประเมิน (evaluation criteria) ในการประมาณค่าพารามิเตอร์ 3 เกณฑ์ คือ 1) ค่าความลำเอียง (BIAS) 2) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และ 3) ค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) การศึกษาในครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลด้วยโปรแกรม R และประมาณค่าพารามิเตอร์ด้วยวิธี Maximum likelihood ด้วยโปรแกรม Mplus ภายใต้การเขียนคำสั่ง การนำข้อมูลไปวิเคราะห์และการแสดงผลในโปรแกรม R

ในบทที่ 5 นี้ ผู้วิจัยขอนำเสนอ 5.1) สรุปผลการศึกษา 5.2) อภิปรายผลการศึกษาและ 5.3) ข้อเสนอแนะ 2 ประเด็น ได้แก่ 5.3.1) ข้อเสนอแนะในการนำผลการศึกษาไปใช้และ 5.3.2) ข้อเสนอแนะในการศึกษาวิจัยครั้งต่อไป ดังรายละเอียดต่อไปนี้

5.1 สรุปผลการศึกษา

การศึกษาในครั้งนี้มีวัตถุประสงค์ 2 ประการ คือ 1) เพื่อศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มและ 2) เพื่อเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ผู้วิจัยขอสรุปผลการศึกษาตามวัตถุประสงค์ ดังนี้

5.1.1) คุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม

จากผลการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม พบว่า ค่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง (true

value) ในทุกเงื่อนไข ในขณะที่ค่าพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) พารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) พารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และพารามิเตอร์ความสามารถระดับกลุ่ม (CA) ใกล้เคียงกับค่าจริง (true value) ในทุกเงื่อนไข ซึ่งผลที่ได้สอดคล้องกันทั้งก่อนและหลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์

โดยเงื่อนไขการจำลองที่แตกต่างกัน 3 ด้านที่ศึกษาในครั้งนี้ อันได้แก่ 1) ตัวแปรจำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) 3 ระดับ 2) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) 3 ระดับและ 3) ขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) 3 ระดับ ไม่ส่งผลต่อคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มมากนัก ทั้งนี้เนื่องจากผลที่ได้มีลักษณะที่สอดคล้องกันทุกเงื่อนไข

อย่างไรก็ตามจากผลการศึกษา พบว่า ปัจจัยที่ส่งผลต่อการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม คือ คุณลักษณะเฉพาะของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) แต่ละโมเดล

5.1.2) **การเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)**

จากผลการศึกษาสรุปได้ว่าประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดล ก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์นั้น พบว่าโมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) ในทุกเงื่อนไข แต่เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์เข้าไปพิจารณาเลือกโมเดลที่ดีที่สุด ผลปรากฏว่าการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่แตกต่างกัน จะได้ผลการคัดสรรโมเดลที่ดีที่สุด (best model) แตกต่างกัน นั่นคือ เมื่อใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) ในขณะที่เมื่อใช้ดัชนี BIC (Bayesian information criterion) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection

criteria) ได้ผลที่สอดคล้องกันคือ โมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ดีที่สุด (best model) ในทุกเงื่อนไข

นอกจากนี้ยังพบว่า การประมาณค่าพารามิเตอร์แต่ละพารามิเตอร์ด้วยเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่แตกต่างกันทั้ง 5 เกณฑ์ อันได้แก่ ดัชนี AIC (Akaike's information criterion), ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion), ดัชนี 2 of 3และดัชนี p-value ได้ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ที่ใกล้เคียงกันในทุกเงื่อนไขและเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) จะลดลง

5.2 อภิปรายผลการศึกษา

ผู้วิจัยมีประเด็นที่จะอภิปรายผลการศึกษา 2 ประเด็น ตามวัตถุประสงค์ของการศึกษา ประกอบไปด้วย 5.2.1) อภิปรายผลการศึกษาตามวัตถุประสงค์การศึกษาข้อที่ 1 นั่นคือ คุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มและ 5.2.2) อภิปรายผลการศึกษาตามวัตถุประสงค์การศึกษาข้อที่ 2 นั่นคือ การเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ดังนี้

5.2.1) อภิปรายผลการศึกษาตามวัตถุประสงค์การศึกษาข้อที่ 1 นั่นคือ คุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม

จากวัตถุประสงค์ข้อที่ 1 ที่ผู้วิจัยศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม โดยโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มนั้นประกอบไปด้วย 2 ตัวแปร คือ ตัวแปรการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) ซึ่งเป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และตัวแปรปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ซึ่งเป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level

DIF factor) ที่เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งผลการศึกษาที่ได้สอดคล้องกันทั้งก่อนและหลังการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์ นั่นคือ ค่าประมาณของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง (true value) ในทุกเงื่อนไข

จากผลการศึกษาที่ได้ จะสามารถกล่าวได้ว่าการระบุคุณลักษณะบางประการของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) ส่งผลต่อการประมาณค่าเฉลี่ยของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) หรืออาจกล่าวอีกนัยหนึ่งว่า ตัวแปรปฏิสัมพันธ์ข้ามระดับเมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มส่งผลต่อการประมาณค่าพารามิเตอร์ปฏิสัมพันธ์ข้ามระดับแบบ 2 ทางและปฏิสัมพันธ์ข้ามระดับแบบ 3 ทางในโมเดลการทำหน้าที่ต่างกันของข้อสอบ ดังนั้น ผู้วิจัยขออภิปรายถึงเหตุผลที่ทำให้ค่าประมาณของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง (true value) ด้วยสมการของโมเดล HGLM และแนวคิดของการวิเคราะห์แบบพหุระดับ ดังต่อไปนี้

สำหรับพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้น จากแนวคิดของการวิเคราะห์พหุระดับ (multilevel analysis) ซึ่งเป็นเทคนิคทางสถิติที่ใช้วิเคราะห์อิทธิพลของตัวแปรทำนายที่มีโครงสร้างเป็นระดับลดหลั่น (hierarchical) หรือเป็นระดับชั้นอย่างน้อย 2 ระดับ โดยตัวแปรทำนายและตัวแปรตามที่อยู่ระดับล่างกว่าจะมีความสัมพันธ์ซึ่งกันและกันและได้รับอิทธิพลร่วมกันจากตัวแปรทำนายที่อยู่ระดับสูงกว่า เช่น ผลการสอบของนักเรียนในชั้นเรียน ได้รับอิทธิพลจากผลการสอนของครูผู้สอน ผลการบริหารจัดการของผู้อำนวยการโรงเรียน รวมไปถึงผลของนโยบายการบริหารจัดการของผู้อำนวยการเขตพื้นที่การศึกษา เป็นต้น ซึ่งจากผลการศึกษาในครั้งนี้สอดคล้องกับหลักการและเหตุผลของการวิเคราะห์พหุระดับ นั่นคือ ตัวแปรทำนายที่อยู่ระดับสูงกว่าจะส่งผลต่อตัวแปรที่อยู่ระดับล่างกว่า โดยตัวแปรการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) ซึ่งเป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) ที่เป็นตัวแปรระดับที่ 1 ดังที่ผู้วิจัยได้กล่าวไว้ตั้งแต่แรกและตัวแปรปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ซึ่งเป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ผู้วิจัยได้กล่าวไว้แล้วเช่นกันน่าจะส่งผลต่อการประมาณค่าพารามิเตอร์ของพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ในทุกเงื่อนไข โดยพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นี้เป็นตัวแปรปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล

(individual level DIF factor) เป็นตัวแปรระดับที่ 2 กับตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ซึ่งเป็นข้อมูลระดับที่ต่ำกว่าตัวแปรระดับกลุ่มที่ผู้วิจัยต้องการศึกษาในการศึกษาครั้งนี้ นั่นคือ ผลการศึกษาที่ได้เป็นไปตามหลักการวิเคราะห์พหุระดับทุกประการ นั่นคือ ตัวแปรทำนายและตัวแปรตามที่อยู่ระดับล่างกว่าจะมีความสัมพันธ์ซึ่งกันและกันและได้รับอิทธิพลร่วมกันจากตัวแปรทำนายที่อยู่ระดับสูงกว่า (Heck & Thomas, 2000, Kamata,1998, 2000, 2005, 2009; ศิริชัย กาญจนวาสี, ก2548, 2550)

นอกจากนี้ ผลการประมาณค่าพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ยังสอดคล้องกับฐานคิดของโมเดล HGLM แบบ 3 ระดับ เมื่อระดับที่ 1 เป็นการวิเคราะห์ระดับข้อสอบ (item level) ระดับที่ 2 เป็นการวิเคราะห์ระดับบุคคล (individual or person level) และระดับที่ 3 เป็นการวิเคราะห์ระดับกลุ่มหรือระดับโรงเรียน (cluster level or group level or school level) มีสมการ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor)

ที่เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

นั่นคือ ตัวแปรต่างๆ ในโมเดล HGLM มีความสัมพันธ์ซึ่งกันและกัน รวมไปถึงการมีปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) ที่เป็นตัวแปรระดับที่ 3 ดังนั้นการละเลยตัวแปรการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) ซึ่งเป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) ตัวแปรระดับที่ 1 และตัวแปรปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ซึ่งเป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ที่เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) ที่เป็นตัวแปรระดับที่ 3 ย่อมส่งผลต่อการประมาณค่าพารามิเตอร์ของพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ที่เป็นตัวแปรปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ตัวแปรระดับที่ 2 กับตัวแปรค่าความยากของข้อสอบ (item difficulty) ตัวแปรระดับที่ 1 อย่างหลีกเลี่ยงไม่ได้ นั่นคือ ตัวแปรที่อยู่ระดับบนส่งผลต่อตัวแปรที่อยู่ระดับล่าง นั่นเอง

กล่าวโดยสรุป ผลการประมาณค่าพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) นั้นเป็นไปตามฐานคิดของการวิเคราะห์พหุระดับและแนวคิดของโมเดล HGLM ทุกประมาณนั้นคือ ตัวแปรทำนายและตัวแปรตามที่อยู่ระดับล่างกว่าจะมีความสัมพันธ์ซึ่งกันและกันและได้รับอิทธิพลร่วมกันจากตัวแปรทำนายที่อยู่ระดับสูงกว่า (Heck & Thomas, 2000; Kamata, 1998, 2000, 2005, 2009; ศิริชัย กาญจนวาสี, 2550)

สำหรับพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) ซึ่งเป็นตัวแปรเป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) ที่เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) ที่เป็นตัวแปรระดับที่ 3 นั้น หากพิจารณาในส่วนของตัวแบบของโมเดล HGLM แล้วจะเห็นได้อย่างชัดเจนว่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นี้ เกิดจากการรวมกันหรือการมีอิทธิพลร่วมกันของพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ซึ่งเป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) ตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) ตัวแปรระดับที่ 1และปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level

DIF factor) ตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 3 ตามลำดับเข้าด้วยกันนั่นเอง ด้วยเหตุนี้ การจัดกระทำกับพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) จึงส่งผลต่อการประมาณค่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) โดยตรงนั่นเอง เนื่องจากการศึกษาในครั้งนี้ประยุกต์แนวคิดของโมเดล HGLM สู่โมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) ที่ศึกษาในครั้งนี้ 4 โมเดล ได้แก่ 1) โมเดลเต็มรูป (complete model) 2) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) 3) โมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) และ 4) โมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) ดังนั้น ผลที่ได้จากการศึกษาในครั้งนี้ จึงเป็นไปตามหลักการของโมเดล HGLM นั่นเอง นอกจากนี้เมื่อพิจารณาจากค่าความคลาดเคลื่อนมาตรฐาน (SE) แล้วจะพบว่า ทั้ง 18 เงื่อนไข ได้ค่าความคลาดเคลื่อนมาตรฐาน (SE) ที่ต่ำมาก แสดงให้เห็นว่าการจำลองข้อมูลได้ค่าที่เสถียรมาก

กล่าวโดยสรุป ผลการประมาณพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นั้นเป็นไปตามฐานคิดของการวิเคราะห์พหุระดับสำหรับข้อมูลแบบระดับขั้นและแนวคิดของโมเดล HGLM ตามฐานคิดที่ว่าตัวแปรทำนายและตัวแปรตามที่อยู่ระดับล่างกว่าจะมีความสัมพันธ์ซึ่งกันและกันและได้รับอิทธิพลร่วมกันจากตัวแปรทำนายที่อยู่ระดับสูงกว่า (Heck & Thomas, 2000; Kamata, 1998, 2000, 2005, 2009; ศิริชัย กาญจนวาสี, 2550) นอกจากนี้ยังสรุปได้ว่าตัวแปรทั้ง 3 ระดับร่วมกันส่งผลต่อการประมาณพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT)

นอกจากนี้ ผลการประมาณค่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) ที่ได้จากการศึกษาในครั้งนี้ สอดคล้องกับผลการศึกษาของ Kamata and Cheong (2007) ที่ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับด้วยโมเดล HGLM (multilevel DIF detection by Hierarchical Generalized Linear Modeling (HLGM)) ด้วยโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) ซึ่งคือโมเดลที่ 3 ของการศึกษาในครั้งนี้เพียงโมเดลเดียว จากการศึกษาในครั้งนั้น ได้ค่าประมาณของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) แตกต่างจากค่าจริง

จากผลที่ได้นี้ สามารถนำไปประยุกต์ใช้ในทางปฏิบัติ (practical value) ได้ นั่นคือ จากการศึกษาในครั้งนี้ จะเห็นได้อย่างชัดเจนว่ามีปฏิสัมพันธ์ข้ามระดับเกิดขึ้น ทั้งในระดับที่ 3 กับระดับที่ 1 ซึ่งคือพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF)และระดับที่ 3 กับระดับที่ 2 ซึ่งก็คือพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) และปฏิสัมพันธ์ข้ามระดับเหล่านี้ส่งผลต่อ

ปฏิสัมพันธ์ข้ามระดับในระดับที่ 2 กับระดับที่ 1 ซึ่งคือพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF)และปฏิสัมพันธ์ข้ามระดับในระดับที่ 3 ระดับที่ 2 และระดับที่ 1 ซึ่งก็คือพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) ยกตัวอย่างเช่น ปฏิสัมพันธ์ระหว่างนโยบายการบริหารโรงเรียนซึ่งเป็นตัวแปรระดับที่ 3 กับค่าความยากของข้อสอบซึ่งเป็นตัวแปรระดับที่ 1 หรือปฏิสัมพันธ์ระหว่างนโยบายการบริหารโรงเรียนของผู้บริหารซึ่งเป็นตัวแปรระดับที่ 3 กับเพศของผู้เรียนซึ่งเป็นตัวแปรระดับที่ 2 ส่งผลต่อความสามารถในการทำข้อสอบได้ถูกต้องของนักเรียนแต่ละคนซึ่งเป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) หรือปฏิสัมพันธ์ระหว่างคุณภาพการสอนของครูในแต่ละโรงเรียนที่เป็นตัวแปรระดับที่ 3 กับค่าความยากของข้อสอบที่เป็นตัวแปรระดับที่ 1 หรือปฏิสัมพันธ์ระหว่างคุณภาพการสอนของครูในแต่ละโรงเรียนซึ่งเป็นตัวแปรระดับที่ 3 กับภาวะความเครียดของผู้เรียนแต่ละคนซึ่งเป็นตัวแปรระดับที่ 2 ส่งผลต่อความสามารถในการทำข้อสอบได้ถูกต้องของนักเรียนแต่ละคนที่ขึ้นอยู่กับโรงเรียนที่นักเรียนแต่ละคนสังกัดอยู่ ซึ่งคือพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) เป็นต้น

จากผลการศึกษาอาจกล่าวได้ว่าตัวแปรจำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ไม่ส่งผลต่อคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มมากนัก ทั้งนี้เนื่องจากผลที่ได้มีลักษณะที่สอดคล้องกันทุกเงื่อนไข ไม่ว่าจะกำหนดตัวแปรทั้ง 3 ขนาดใดก็ตาม รวมไปถึงจากการพิจารณาจากค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ที่พบว่าได้ค่าที่ใกล้เคียงกันในทุกเงื่อนไข แต่อย่างไรก็ตาม เมื่อจำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) มีขนาดใหญ่ขึ้นจะส่งผลให้ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ลดลง นั่นคือ ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ในเงื่อนไขที่ 10-18 จะมีซึ่งมีขนาดกลุ่มตัวอย่าง 5,000 คน มีค่าน้อยกว่าในเงื่อนไขที่ 1-9 ที่มีขนาดกลุ่มตัวอย่าง 2,500 คน ซึ่งสอดคล้องกับคำกล่าวของ Afshartous & Leeuw (2005),Cohen (2005), Maas & Hox (2004, 2005) ที่กล่าวว่าขนาดของกลุ่มตัวอย่างที่เพิ่มขึ้นจะช่วยลดความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์และเพิ่มความแม่นยำในการประมาณค่าพารามิเตอร์

5.2.2) **อภิปรายผลการศึกษาตามวัตถุประสงค์การศึกษาข้อที่ 2 นั่นคือ การเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT)**

จากวัตถุประสงค์ข้อที่ 2 ที่ผู้วิจัยเปรียบเทียบประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดลว่าโมเดลไหนมีประสิทธิภาพในการประมาณค่าพารามิเตอร์ได้ดีที่สุด เมื่อมีเงื่อนไขที่แตกต่างกัน 3 ด้าน คือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม)CLDIF) และขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) จากผลการศึกษา สรุปได้ว่า ประสิทธิภาพของโมเดลการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 โมเดล ก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์นั้น พบว่า โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) ในทุกเงื่อนไข แต่เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์เข้าไปพิจารณาเลือกโมเดลที่ดีที่สุด พบว่า เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) ในเงื่อนไขที่ 10 - 18 ซึ่งเป็นเงื่อนไขการจำลองที่มีกลุ่มตัวอย่างขนาดใหญ่ นั่นคือ จำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) มีจำนวน 100 คน โมเดลที่ 2 หรือโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (incomplete model without two-way interaction) เป็นโมเดลที่ดีที่สุด)best model (ใน 4 เงื่อนไขคือเงื่อนไขที่ 2, 5, 6 และ 8 และโมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม)incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ดีที่สุด (best model) ใน 5 เงื่อนไขคือเงื่อนไขที่ 1, 3, 4, 7 และ 9 ดังนั้น จึงอาจกล่าวได้ว่า เมื่อใช้ดัชนี AIC (Akaike's information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) นั้น โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model)

เมื่อใช้ดัชนี BIC (Bayesian information criterion) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ได้ผลสอดคล้องกัน นั่นคือ โมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ดีที่สุด (best model) ในทุกเงื่อนไข

เมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) ใน 15 เงื่อนไข คือ เงื่อนไขที่ 1, 2, 4, 5, 7, 8, 10, 11, 12, 13,14,15, 16,17,18 และโมเดลที่ 3 หรือโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) เป็นโมเดลที่ดีที่สุด (best model) ในเงื่อนไขที่ 3, 6 และ 9 เท่านั้น ดังนั้น จึงอาจกล่าวได้ว่าเมื่อใช้ดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model)

ดังนั้นจึงสรุปได้ว่าเมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่แตกต่างกัน จะได้ผลการคัดสรรโมเดลที่ดีที่สุด (best model) แตกต่างกัน ทั้งนี้น่าจะเนื่องมาจากแต่ละเกณฑ์การคัดสรรโมเดล (Model selection criteria) มีสูตรในการคำนวณที่แตกต่างกัน จึงทำให้ได้ผลการคัดสรรโมเดลที่ดีที่สุด (best model) ที่แตกต่างกัน ดังนี้ เมื่อใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) ในขณะที่เมื่อใช้ดัชนี BIC (Bayesian information criterion) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ได้ผลที่สอดคล้องกันคือ โมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ดีที่สุด (best model) ดังนั้นจึงอาจกล่าวได้ว่าการที่นักวิจัยจะเลือกใช้โมเดลไหนในการทำวิจัยในอนาคตนั้น ควรจะพิจารณาในส่วนของเกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่ใช้ด้วย ทั้งนี้ก็เพราะว่าเกณฑ์การคัดสรรโมเดล (Model selection criteria) เป็นปัจจัยหนึ่งที่ส่งผลต่อการได้มาซึ่งโมเดลที่ดีที่สุด (best model) ที่แตกต่างกัน นั่นเอง

จากผลที่ได้สรุปได้ว่าโมเดลที่ดีที่สุด (best model) มี 2 โมเดล นั่นคือ โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) ซึ่งผลที่ได้สอดคล้องกับผลการศึกษาก่อนการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้งนี้น่าจะเนื่องมาจากเป็นโมเดลที่มีพารามิเตอร์ทุกพารามิเตอร์ครบถ้วนตามแนวคิดของโมเดล HGLM และโมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) ทั้งนี้น่าจะเนื่องมาจากเป็นโมเดลที่ประหยัดที่สุด ซึ่งผลที่ได้นี้จะเป็นทางเลือกให้แก่นักวิจัยในอนาคตในการเลือกใช้โมเดลการทำหน้าที่ต่างกันของข้อสอบ (DIF) ที่เหมาะสมต่อไป นั่นคือ ในกรณีที่นักวิจัยต้องการศึกษาทุกพารามิเตอร์ในโมเดลก็ควรเลือกโมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) แต่หากนักวิจัยต้องการเลือกโมเดลที่ประหยัด

และง่ายต่อการตีความหมายก็ควรเลือกโมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) นอกจากนี้ เมื่อพิจารณาในส่วนของตัวแบบหรือโมเดลทั้ง 4 โมเดลจะพบว่าโมเดลที่ 4 เป็นโมเดลที่ง่ายและประหยัดที่สุด โดยแต่ละโมเดลมีสมการเรียงตามลำดับ ดังนี้

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_{5i} X_j I_i+\gamma_{6i} W_k I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_4 X_j W_k+\gamma_{5i} X_j I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

$$\log\left(\frac{p_{ijk}}{1-p_{ijk}}\right)=\gamma_1 X_j+\gamma_2 W_k+\gamma_{3i} I_i+\gamma_{5i} X_j I_i+\gamma_{7i} X_j W_k I_i+\theta_j^{(2)}+\theta_k^{(3)}$$

เมื่อ p_{ijk} เป็นความน่าจะเป็นของการตอบข้อสอบข้อที่ *i* ได้ถูกต้องของนักเรียนคนที่ *j* ซึ่งมาจากโรงเรียนหรือกลุ่มที่ *k* I_i เป็นตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 X_j เป็นตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 W_k เป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ดังนั้นจากโมเดล HGLM จะได้ว่ามีพารามิเตอร์อิทธิพลกำหนด (fixed effect) 7 พารามิเตอร์ คือ γ_1 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับบุคคล (ILADIFF) γ_2 เป็นพารามิเตอร์ความแตกต่างของความสามารถระดับกลุ่ม (CLADIFF) γ_{3i} เป็นค่าความยากของข้อสอบข้อที่ *i* (item difficulty) γ_4 เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) γ_{5i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) ซึ่งเป็นตัวแปรระดับที่ 2 และตัวแปรค่าความยากของข้อสอบ (item difficulty) ซึ่งเป็นตัวแปรระดับที่ 1 γ_{6i} เป็นพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) สำหรับข้อสอบข้อที่ *i* เป็นปฏิสัมพันธ์แบบ 2 ทาง

ระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 และตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 และ γ_{7i} เป็นปฏิสัมพันธ์แบบ 3 ทางระหว่างตัวแปรค่าความยากของข้อสอบ (item difficulty) เป็นตัวแปรระดับที่ 1 ตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) เป็นตัวแปรระดับที่ 3 ซึ่งในการศึกษาในครั้งนี้เรียกว่าเป็นพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) นอกจากนี้ยังมีพารามิเตอร์อิทธิพลสุ่ม (random effect) อีก 2 พารามิเตอร์ คือ $\theta_j^{(2)}$ เป็นพารามิเตอร์ความสามารถระดับบุคคลภายในกลุ่ม (IAWC) และ $\theta_k^{(3)}$ เป็นพารามิเตอร์ความสามารถระดับกลุ่ม (CA)

จากผลที่ได้ในการศึกษาในครั้งนี้ สามารถสรุปได้ว่าคุณลักษณะบางประการของโมเดลการทำหน้าที่ต่างกันของข้อสอบส่งผลต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบ นั่นคือ การละเลยต่อค่าพารามิเตอร์บางพารามิเตอร์ในโมเดล HGLM ซึ่งเป็นโมเดลที่ใช้เป็นกรอบแนวคิดในการศึกษาในครั้งนี้นั้น จะทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบขึ้นในโมเดลการทำหน้าที่ต่างกันของข้อสอบและส่งผลต่อประสิทธิภาพในการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบ ตัวอย่างเช่น การศึกษาของ Kamata and Cheong (2007) ที่ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบพหุระดับด้วยโมเดล HGLM (multilevel DIF detection by Hierarchical Generalized Linear Modeling (HLGM)) แต่กำหนดค่าพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์ รวมไปถึงจากการศึกษาของ Chaimongkol, Huffer & Kamata (2006) และ Vaughn (2006) ก็ศึกษาโมเดลการทำหน้าที่ต่างกันของข้อสอบโดยกำหนดพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) เป็นศูนย์เช่นกัน นั่นคือเป็นการศึกษาด้วยโมเดลไม่เต็มรูปที่ไม่มีพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without cluster-level DIF) หรือโมเดลที่ 3 ของการศึกษาในครั้งนี้เพียงโมเดลเดียวเท่านั้น จากการศึกษาในครั้งนี้ ทำให้ได้สารสนเทศใหม่ว่าโมเดลที่ Kamata and Cheong (2007), Chaimongkol, Huffer and Kamata (2006) และ Vaughn (2006) ใช้เป็นโมเดลในการศึกษานั้น ไม่ใช่โมเดลที่ดีที่สุด (best model) สำหรับการประมาณค่าพารามิเตอร์ นั่นคือ เป็นโมเดลที่ทำให้ได้ค่าประมาณของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง นั่นคือ เกิดการทำหน้าที่ต่างกัน นั่นเอง รวมไปถึงผลที่ได้อาจจะไม่มีความน่าเชื่อถือเพียงพอ แต่อย่างไรก็ตามเนื่องจากการศึกษาในครั้งนั้น ยังไม่สารสนเทศที่เพียงพอว่าโมเดลไหนเป็นโมเดลที่ดีที่สุด (best model) หรือนักวิจัยละเลยในประเด็นของโมเดลที่ใช้ในการวิเคราะห์ว่าเป็นโมเดลที่ดีที่สุด (best model) แล้วหรือยัง

อย่างไรก็ตาม เนื่องจากการศึกษาในครั้งนี้เป็นการศึกษาในเชิงทฤษฎีเท่านั้น หากพิจารณาบริบทจริงทางการศึกษาจะพบว่ามีบางปัจจัยหรือบางตัวแปรที่เป็นตัวแปรที่ส่งต่อผลสัมฤทธิ์

ทางการศึกษาของนักเรียนหรือเป็นตัวแปรที่ส่งผลต่อความน่าจะเป็นในการตอบข้อสอบข้อที่ i ได้ถูกต้องของนักเรียน เช่น ตัวแปรการกวดวิชา แต่อาจจะต้องมีการตีความหมายหรือต้องศึกษาค้นคว้าเพิ่มเติมว่าตัวแปรดังกล่าวนี้ควรจัดไว้ในพารามิเตอร์ใดของโมเดล HGLM

นอกจากนี้ยังพบว่า การประมาณค่าพารามิเตอร์เป้าหมายแต่ละตัว ด้วยเกณฑ์การประเมินที่แตกต่างกันนั้น ได้ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ที่ใกล้เคียงกัน ดังนั้น จากการศึกษาค้นคว้าในครั้งนี้ เป็นการศึกษาที่ได้สารสนเทศที่เป็นประโยชน์สำหรับการศึกษาค้นคว้าต่อไปในอนาคต นั่นคือ นักวิจัยสามารถเลือกใช้เพียงเกณฑ์ใดเกณฑ์หนึ่งก็ได้ในการประเมินประสิทธิภาพของโมเดล โดยไม่จำเป็นต้องใช้เกณฑ์การประเมินถึง 5 เกณฑ์ในการศึกษา ซึ่งจะทำให้นักวิจัยประหยัดเวลาในการดำเนินการวิจัยมากยิ่งขึ้น จากการศึกษาในครั้งนี้ ยังพบอีกว่าค่าพารามิเตอร์ที่ประมาณได้จากใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) นั้น จะสูงกว่าค่าพารามิเตอร์ที่ประมาณด้วยดัชนี AIC (Akaike's information criterion) เล็กน้อย ซึ่งสอดคล้องกับผลการศึกษาของ Shibata (2002) ที่ได้ข้อสรุปว่าเมื่อใช้ดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) นั้นจะได้ค่าประมาณของพารามิเตอร์ต่างๆ สูงกว่าดัชนี AIC (Akaike's information criterion) ทั้งนี้เนื่องจากทั้งดัชนี BIC (Bayesian information criterion), ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) มีความอ่อนไหวกว่าดัชนี AIC (Akaike's information criterion) นั่นเอง ซึ่งสารสนเทศที่ได้จากการศึกษาในครั้งนี้จะเป็นการเสนอทางเลือกให้แก่นักวิจัยในอนาคตในส่วนของการคัดเลือกโมเดลที่เหมาะสมสำหรับการวิเคราะห์ข้อมูลต่อไป

5.3 ข้อเสนอแนะ

ผู้วิจัยมีข้อเสนอแนะจากการศึกษาวิจัยในครั้งนี้ 2 ประการ คือ 6.3.1) ข้อเสนอแนะในการนำผลการศึกษาไปใช้และ 6.3.2) ข้อเสนอแนะในการทำวิจัยครั้งต่อไป ดังต่อไปนี้

5.3.1) ข้อเสนอแนะในการนำผลการศึกษาไปใช้

ผู้วิจัยมีข้อเสนอแนะในการนำผลการศึกษาไปใช้ โดยยึดตามวัตถุประสงค์ของการศึกษาและผลที่ได้เป็นสำคัญ ดังต่อไปนี้

1) จากการศึกษาในครั้งนี้ พบว่า ค่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริง (true value) ในทุกเงื่อนไข ดังนั้น ผลที่ได้จากการศึกษาในครั้งนี้จึงให้สารสนเทศในระดับนโยบาย นั่นคือ ผู้ที่

เกี่ยวข้องกับการจัดการศึกษาในระดับนโยบายควรตระหนักและเห็นความสำคัญของปฏิสัมพันธ์แบบ 2 ทางข้ามระดับ นั่นคือ ตัวแปรการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) ซึ่งเป็นปฏิสัมพันธ์แบบ 2 ทางข้ามระดับระหว่างตัวแปรระดับกลุ่ม (cluster level DIF factor) ที่เป็นตัวแปรระดับที่ 3และตัวแปรค่าความยากของข้อสอบ (item difficulty) ที่เป็นตัวแปรระดับที่ 1 โดยอาจจะเป็นปฏิสัมพันธ์ร่วมของข้อสอบกับตัวแปรระดับโรงเรียนหรือระดับกลุ่ม อาจจะส่งผลต่อการประมาณค่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ตัวอย่างเช่น การที่นักเรียนที่ศึกษาอยู่ในโรงเรียนในเมืองคุ้นเคยกับข้อสอบแบบเขียนตอบมากกว่านักเรียนที่ศึกษาอยู่ในโรงเรียนในชนบท ทั้งนี้เนื่องมาจากโรงเรียนในเมืองมีนโยบายมุ่งเน้นการทำข้อสอบแบบเขียนตอบมากกว่า รวมไปถึงการที่ครูผู้สอนให้ความสำคัญกับการทำข้อสอบแบบเขียนตอบมากกว่าโรงเรียนในชนบท หากข้อสอบสอบเข้ามหาลัยมุ่งเน้นการเขียนตอบเป็นส่วนใหญ่ก็จะส่งผลให้เกิดการได้เปรียบเสียเปรียบกันได้ อีกตัวอย่างหนึ่งที่น่าสนใจ คือ หากโรงเรียนในเมืองสอนทักษะการคิดวิเคราะห์ในขณะที่โรงเรียนในชนบทละเลยในการสอนเรื่องการคิดวิเคราะห์ แล้วข้อสอบ O-NET วัดการคิดวิเคราะห์เป็นส่วนหนึ่งของข้อสอบ ก็จะทำให้เกิดการได้เปรียบเสียเปรียบระหว่างนักเรียนหรือผู้เข้าสอบได้ นั่นคือ ข้อสอบขาดความยุติธรรมสำหรับผู้เข้าสอบนั่นเอง นอกจากนี้ ตัวแปรของปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ซึ่งเป็นปฏิสัมพันธ์ 2 ทางระหว่างตัวแปรระดับบุคคล (individual level DIF factor) เป็นตัวแปรระดับที่ 2 และตัวแปรระดับกลุ่ม (cluster level DIF factor) ที่เป็นตัวแปรระดับที่ 3 นั่นคือ เป็นปฏิสัมพันธ์ร่วมของตัวแปรระดับนักเรียนกับตัวแปรระดับโรงเรียนหรือระดับกลุ่ม ซึ่งจากการศึกษาในครั้งนี้อาจกล่าวได้ว่าตัวแปรนี้อาจผลต่อการประมาณค่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ด้วยเช่นกัน ตัวอย่างเช่น การที่นักเรียนชายซึ่งเป็นตัวแปรระดับบุคคล (individual level DIF factor) ที่ศึกษาอยู่ในโรงเรียนกีฬาที่มีนโยบายมุ่งเน้นด้านกีฬาซึ่งเป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) คุ้นเคยกับการเล่นกีฬามากกว่านักเรียนหญิงซึ่งเป็นตัวแปรระดับบุคคล (individual level DIF factor) ที่ศึกษาอยู่ในโรงเรียนที่มุ่งเน้นด้านวิชาการซึ่งเป็นตัวแปรระดับกลุ่ม (cluster level DIF factor) หากข้อสอบเข้ามหาลัยบางข้อถามรายละเอียดของกีฬาบางประเภทก็จะส่งผลให้เกิดการได้เปรียบเสียเปรียบกันได้ นั่นคือ ข้อสอบขาดความยุติธรรมสำหรับผู้เข้าสอบทุกคนนั่นเอง เนื่องจากนักเรียนแต่ละคนมีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน ซึ่งปัญหาเรื่องความยุติธรรมของข้อสอบกำลังเป็นปัญหาที่หลายฝ่ายให้ความสำคัญอยู่ในขณะนี้ แม้กระทั่งประเทศไทยเองก็เกิดปัญหานี้เช่นกัน ดังนั้น หากผู้ที่เกี่ยวข้องกับการวางนโยบายการศึกษา การออกข้อสอบเข้ามหาวิทยาลัย ผู้อำนวยการเขตพื้นที่การศึกษา ครูอาจารย์รวมถึงทุกภาคส่วนนำผลที่ได้จากการศึกษาในครั้งนี้ไป

ประกอบการตัดสินใจในการออกข้อสอบหรือการวางระบบทางการศึกษา ก็จะเป็นการพัฒนาการศึกษาของชาติได้อีกทางหนึ่ง

จากผลการศึกษา พบว่าปฏิสัมพันธ์ข้ามระดับ เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มส่งผลต่อพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) ในทุกเงื่อนไข นั่นคือ ไม่ว่าผู้วิจัยจะกำหนดจำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) เป็นเท่าใด ขนาดการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) ระดับไหนหรือขนาดปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) เพียงใดก็ตาม ดังนั้น ผู้ที่เกี่ยวข้องกับการจัดการศึกษา ไม่ว่าจะเป็นกระทรวงศึกษาธิการ สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ) สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) (สทศ) เขตพื้นที่การศึกษา รวมไปถึงผู้บริหารสถานศึกษาและผู้ปกครองจึงควรตระหนักและเห็นความสำคัญของปฏิสัมพันธ์ข้ามระดับนี้ รวมไปถึงการหาแนวคิดหรือนโยบายที่จะป้องกันหรือแก้ไขปัญหาของปฏิสัมพันธ์ข้ามระดับดังกล่าวนี้ ทั้งนี้ก็เพื่อให้การจัดการศึกษามีความเท่าเทียม ลดการได้เปรียบเสียเปรียบกันทางการศึกษา รวมไปถึงเพื่อให้การจัดการสอบระดับชาติ (national testing or high stage testing) เช่น การสอบ O-NET หรือการสอบตรงเข้ามหาลัยของสถาบันการศึกษาต่างๆ เป็นการจัดการทดสอบที่ยุติธรรมสำหรับผู้เข้าสอบทุกคนอย่างแท้จริง

2) จากการศึกษาในครั้งนี้ พบว่า ตัวแปรจำนวนนักเรียนในแต่ละโรงเรียน (the number of student in each school) ส่งผลต่อการประมาณค่าพารามิเตอร์ด้วยเช่นกัน นั่นคือ หากจำนวนนักเรียนในแต่ละโรงเรียนมีขนาดใหญ่หรือกล่าวอีกนัยหนึ่งว่ากลุ่มตัวอย่างมีขนาดใหญ่จะส่งผลให้ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) ลดลง ดังนั้น ผลการศึกษาที่ได้จึงเป็นสารสนเทศที่เป็นประโยชน์แก่ผู้ที่ต้องการศึกษาค้นคว้าหรือนำผลไปใช้ในการประมาณค่าพารามิเตอร์ต่อไปในอนาคต นั่นคือ เมื่อกลุ่มตัวอย่างเพิ่มมากขึ้น ค่าพารามิเตอร์ที่ได้จะเสถียรมากยิ่งขึ้น ซึ่งการศึกษาในครั้งนี้ศึกษาเพียง 2 ขนาดกลุ่มตัวอย่าง คือ 2,500 คนและ 5,000 คน ซึ่งผลที่ได้มีความน่าเชื่อถือทั้งคู่และได้ค่าที่ใกล้เคียงกัน ทั้งนี้ก็เพราะว่าการศึกษาในครั้งนี้ศึกษาจากกลุ่มตัวอย่างขนาดใหญ่ โดยมีจำนวนมากกว่าขนาดกลุ่มตัวอย่างที่ Afshartous and Leeuw (2005), Cohen (2005), Maas and Hox (2004, 2005), Binici (2007) และ ศิริชัย กาญจนวาสี (2548) ได้กล่าวไว้ นอกจากนี้ยังพบว่าเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น ค่าความลำเอียง (BIAS) ค่าความคลาดเคลื่อนมาตรฐาน (SE) และค่ารากที่สองของความคลาดเคลื่อนยกกำลังสองเฉลี่ย (RMSE) จะลดลง อย่างไรก็ตามในทางกลับกันการใช้กลุ่มตัวอย่างขนาดใหญ่จะใช้เวลาในการประมาณค่าพารามิเตอร์มากกว่ากลุ่มตัวอย่างที่มีขนาดเล็ก ดังนั้น นักวิจัยจึงควรพิจารณาถึงปัจจัยด้านเวลาและปัจจัยด้านสมรรถนะของเครื่อง

คอมพิวเตอร์ที่จะใช้ในการจำลองข้อมูลและประมาณค่าพารามิเตอร์ด้วย ซึ่งจากการศึกษาครั้งนี้ ผู้วิจัยใช้คอมพิวเตอร์ 2 เครื่องที่มี Processor ของเครื่อง 4.0 GHz และ Memory (Ram) 12 GB ระบบปฏิบัติการ Windows7 ซึ่งถือว่ามีสมรรถนะและความเร็วสูงมากในปัจจุบันในการจำลองข้อมูลและประมาณค่าพารามิเตอร์ ซึ่งใช้เวลา 48 ชั่วโมงหรือ 2 วันต่อ 1 เงื่อนไข ซึ่งหากนักวิจัยใช้คอมพิวเตอร์ที่มีสมรรถนะที่ต่ำกว่านี้หรือมีสมรรถนะเท่ากับคอมพิวเตอร์ตามท้องตลาดทั่วไปก็อาจจะใช้เวลา 4-5 วันต่อ 1 เงื่อนไข

3) จากการศึกษาในครั้งนี้ พบว่า เมื่อใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ทั้ง 5 เกณฑ์เข้าไปพิจารณาเลือกโมเดลที่ดีที่สุด (best model) ผลปรากฏว่าการใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่แตกต่างกัน จะได้ผลการคัดสรรโมเดลที่ดีที่สุด (best model) แตกต่างกัน นั่นคือ เมื่อใช้ดัชนี AIC (Akaike's information criterion) และดัชนี p-value เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) พบว่า โมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เป็นโมเดลที่ดีที่สุด (best model) ในขณะที่เมื่อใช้ดัชนี BIC (Bayesian information criterion) ดัชนี ABIC (Sample-Size Adjusted Bayesian information criterion) และดัชนี 2 of 3 เป็นเกณฑ์การคัดสรรโมเดล (Model selection criteria) ได้ผลที่สอดคล้องกันคือ โมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เป็นโมเดลที่ดีที่สุด (best model) ซึ่งผลการศึกษาที่ได้นี้ จะให้สารสนเทศแก่นักวิจัยใน 2 ประเด็นคือ ประเด็นแรก การเลือกใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) เนื่องจากผลการศึกษาในครั้งนี้ชัดเจนว่าการเลือกใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่แตกต่างกัน จะได้ผลการคัดสรรโมเดลที่ดีที่สุด (best model) ที่แตกต่างกัน ดังนั้นนักวิจัยจะต้องใช้วิจารณญาณและฐานคิดที่อิงตามหลักวิชาในการเลือกใช้เกณฑ์การคัดสรรโมเดล (Model selection criteria) ที่เหมาะสมกับงานวิจัยที่จะทำให้มากที่สุด ประเด็นที่ 2 หากนักวิจัยต้องการเลือกโมเดลที่ดีที่สุดเพียง 1 โมเดลเป็นโมเดลในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) ในอนาคต ผลการศึกษาที่ได้จากการศึกษาในครั้งนี้ จะเป็นสารสนเทศที่สำคัญที่นักวิจัยจะสามารถนำไปสรุปอ้างอิงหรือนำไปเป็นหลักเกณฑ์ในการเลือก นั่นคือ นักวิจัยสามารถเลือกโมเดลที่ 4 หรือโมเดลไม่เต็มรูปที่ไม่มีทั้งพารามิเตอร์ปฏิสัมพันธ์แบบ 2 ทางและพารามิเตอร์การทำหน้าที่ต่างกันระดับกลุ่ม (incomplete model without both two-way interaction and cluster-level DIF) เนื่องจากเป็นโมเดลที่ประหยัดและมีร้อยละของการเป็นโมเดลที่ดีที่สุด (best model) สูงที่สุดหรือเลือกโมเดลที่ 1 หรือโมเดลเต็มรูป (complete model) เมื่อนักวิจัยไม่ต้องการละเลยค่าพารามิเตอร์บางพารามิเตอร์ในโมเดล เป็นต้น

5.3.2) ข้อเสนอแนะในการศึกษาวิจัยครั้งต่อไป

ผู้วิจัยมีข้อเสนอแนะในการศึกษาวิจัยครั้งต่อไป ดังนี้

1) จากการศึกษาในครั้งนี้เป็นการศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่มด้วย 4 โมเดลที่มีการระบุคุณลักษณะเฉพาะที่แตกต่างกัน ผลการศึกษาพบว่า การประมาณค่าพารามิเตอร์ของพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) แตกต่างจากค่าจริงในทุกเงื่อนไข ดังนั้น เพื่อให้การศึกษาค้นคว้าการทำหน้าที่ต่างกันของข้อสอบกว้างขวางและมีคุณค่ามากยิ่งขึ้น การศึกษาในอนาคตจึงควรศึกษาเพิ่มเติมว่าตัวแปรใดส่งผลต่อคุณภาพของการประมาณค่าพารามิเตอร์ดังกล่าว รวมไปถึงการศึกษาปฏิสัมพันธ์ระหว่างตัวแปรการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และตัวแปรปฏิสัมพันธ์แบบ 2 ทาง (2WAYINT) ว่าทั้งสองตัวแปรร่วมกันส่งผลต่อคุณภาพของการประมาณค่าพารามิเตอร์ปฏิสัมพันธ์แบบ 3 ทาง (3WAYINT) และพารามิเตอร์การทำหน้าที่ต่างกันระดับบุคคล (ILDIF) หรือไม่ อย่างไร

2) จากการศึกษาในครั้งนี้เป็นการศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) เมื่อตัวแปรหนึ่งเป็นตัวแปรระดับกลุ่ม นั่นคือ ผู้วิจัยศึกษาโดยกำหนดขนาดของตัวแปรการทำหน้าที่ต่างกันระดับกลุ่ม (CLDIF) และตัวแปรปฏิสัมพันธ์แบบ 2 ทาง(2WAYINT) ที่แตกต่างกัน โดยกำหนดขนาดของกลุ่มตัวอย่างเพียง 2 ขนาด คือ 2,500 คน และ 5,000 คน ซึ่งเป็นกลุ่มตัวอย่างขนาดใหญ่ และกำหนดขนาดของกลุ่มอ้างอิง (reference) และกลุ่มเปรียบเทียบ (focal group) ในอัตราส่วน 1:1 ซึ่งการศึกษาค้นคว้าในอนาคตควรศึกษาขนาดของกลุ่มตัวอย่างที่หลากหลายขนาด รวมไปถึงการกำหนดอัตราส่วนของกลุ่มอ้างอิง (reference) และกลุ่มเปรียบเทียบ (focal group) ที่แตกต่างกันไป การศึกษาความยาวของข้อสอบที่หลากหลาย รวมไปถึงจำนวนข้อสอบที่เกิดการทำหน้าที่ต่างกันในแบบสอบ 1 ฉบับ ซึ่งผลที่ได้น่าจะเป็นประโยชน์และเป็นการขยายศาสตร์ด้านการวัดผลให้กว้างขวางมากยิ่งขึ้น

3) จากการศึกษาในครั้งนี้เป็นการศึกษาคุณภาพของการประมาณค่าพารามิเตอร์ของโมเดลการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) สำหรับข้อสอบที่ตรวจให้คะแนนแบบ 2 ค่าแบบเอกมิติ (unidimensional) เท่านั้น ซึ่งเนื่องด้วยกระแสของการศึกษาข้อสอบที่ตรวจให้คะแนนแบบหลายค่า (polytomous) รวมไปถึงแบบสอบแบบพหุมิติ (multidimensional) ที่นักวิจัยส่วนหนึ่งให้ความสนใจและกำลังศึกษาค้นคว้าอย่างต่อเนื่อง ดังนั้นเพื่อเป็นการขยายแนวคิดของการศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) ให้กว้างขวางและมีคุณค่าต่อศาสตร์แห่งการวัดผลมากยิ่งขึ้น การศึกษาในอนาคตจึงควรศึกษาเพิ่มเติม

ในส่วนของการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) สำหรับข้อสอบที่ตรวจให้คะแนนแบบหลายค่า (polytomous) และ/หรือจากแบบวัดแบบพหุมิติ (multidimensional) ซึ่งอาจจะศึกษาทั้งในส่วนของความเป็นพหุมิติภายในแบบสอบหรือความเป็นพหุมิติระดับข้อสอบ ซึ่งผลการศึกษาที่ได้นั้น น่าจะเป็นการศึกษาที่มีประโยชน์และเป็นการขยายศาสตร์ด้านการวัดผลให้มีความก้าวหน้าและทัดเทียมศาสตร์อื่นๆ มากยิ่งขึ้น

4) การศึกษาในครั้งนี้เป็นการศึกษาในเชิงทฤษฎี เพื่อบุกเบิกองค์ความรู้ใหม่ที่ยังไม่มีการศึกษาค้นคว้ามาก่อน โดยไม่มีการศึกษาจากข้อมูลจริงแต่อย่างใด รวมไปถึงไม่มีการทดสอบ Type I error rate, type II error rate และ power ดังนั้น หากการศึกษาค้นคว้าในอนาคตศึกษาเพิ่มเติมในส่วนของการศึกษาการทำหน้าที่ต่างกันของข้อสอบแบบ 2 ทาง (two-way DIF) ด้วยข้อมูลจริง รวมไปถึงทดสอบ Type I error rate, type II error rate และ power ก็น่าจะเป็นการศึกษาที่มีประโยชน์และมีคุณค่าในทางปฏิบัติมากยิ่งขึ้น

รายการอ้างอิง

ภาษาไทย

กัลยา วานิชย์บัญชา. (2550). **หลักสถิติ**. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

กาญจนา วัธนสุนธร. (2537). **การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ**. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาวิจัยการศึกษา ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

เกสร หว่างจิตร. (2539). **การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับแบบสอบคัดเลือกระดับบัณฑิตศึกษาวิชาภาษาไทยและวิชาภาษาอังกฤษด้วยวิธีแมนเทล-แฮนส์เซล**. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาวิจัยการศึกษา ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

จิติมา วรรณศรี. (2539). **การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีซิบเทสท์ เมื่อความยาวของแบบสอบ ขนาดกลุ่มตัวอย่างและอัตราส่วนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

จิตติมา ผสมญาติ. (2546). **การเปรียบเทียบวิธีการคัดเลือกสมการถดถอยที่ดีที่สุดเชิงเบส์เมื่อใช้การแจกแจงแบบคู่สังยุคปกติ**. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาสถิติการศึกษา ภาควิชาสถิติ คณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย.

ญาณภัทร สีหะมงคล. (2540). **การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี** Lord' s χ2 **วิธี** Raju' s Area Measures **และวิธี** Closed Interval Area. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

ธีระพร วีระถาวร. (2536). **การอนุมานเชิงสถิติขั้นกลาง: โครงสร้างและความหมาย**. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

นพมาศ พิพัฒนสุข. (2541). **การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-เฮลส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้เกณฑ์จับคู่เปรียบเทียบแตกต่างกันในแบบสอบชนิดพหุมิติ**. วิทยานิพนธ์ปริญญามหาบัณฑิต. สาขาวิชาวิจัยการศึกษา ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

นิคม กีรติวรางกูร. (2542). **การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัด แมนเทล-แฮนส์เซลและการตอบสนองข้อสอบ**. วิทยานิพนธ์ปริญญามหาบัณฑิต. สาขาวิชาวิจัยการศึกษา ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ปิยะทิพย์ ตินวร. (2549). **การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ : การเปรียบเทียบประสิทธิภาพระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก**. วิทยานิพนธ์ปริญญามหาบัณฑิต. ภาควิชาวิจัยและวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยบูรพา.

พรรณี จินตมาศ. (2540). **การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ขนาดกลุ่ม ผู้สอบและวิธีวิเคราะห์ต่างกัน**. วิทยานพินธ์ปริญญามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

พัชรี จันทร์เพ็ง. (2550). **การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติภายใต้การหมุนแกน โครงสร้างเชิงมิติและระดับความสัมพันธ์ที่แตกต่างกัน**. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

พร้อมพรรณ อุดมสิน. (2544). **การวัดและการประเมินผลการเรียนการสอนคณิตศาสตร์**. กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

มานพ วราภักดิ์. (2548). **ทฤษฎีความน่าจะเป็น**. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ยุพิน พิพิธกุล. (2546). **การเรียนการสอนคณิตศาสตร์: ยุคปฏิรูปการศึกษา**. กรุงเทพมหานคร: บพิธการพิมพ์.

รัชรินทร์ มุคดา. (2540). **การเปรียบเทียบประสิทธิภาพระหว่างวิธีแมนเทล-แฮนส์เซลกับวิธีถดถอยโลจิสติกในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ในกรณีการจัดกลุ่มความสามารถ ค่าความยากของแบบสอบ และค่าอำนาจจำแนกของข้อสอบต่างกัน**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วลีมาศ แซ่อึ้ง. (2543). **การเปรียบเทียบอำนาจการทดสอบและอัตราส่วนความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูประหว่างวิธีซิบเทสท์ปรับใหม่ วิธีซิบเทสท์ วิธีแมนเทล-แฮนส์เซล วิธีการถดถอย**. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

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

วินัย วงศ์ฤทัยวัฒนา .(2533) .**การเปรียบเทียบผลการประมาณค่าพารามิเตอร์ของโมเดลโลจิสติกแบบสองพารามิเตอร์ระหว่างวิธีของเบส์กับวิธีแมกซิมัมไลดค์ลิฮู้ด**วิทยานิพนธ์ . ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

หัชชา ศรีปลั่ง. (2550). **การวิเคราะห์ทางสถิติพื้นฐานด้วย** R-ICE. สงขลา: ชานเมืองการพิมพ์.

ศิริชัย กาญจนวาสี. (2550). **ทฤษฎีการทดสอบแนวใหม่**. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ศึกษาธิการ, กระทรวง. (2545). **แผนพัฒนาคุณภาพการเรียนการสอนภาษาไทยและการใช้ภาษาไทย**. กรุงเทพมหานคร: โรงพิมพ์คุรุสภาลาดพร้าว.

ศึกษาธิการ, กระทรวง. (2545). **เอกสารประกอบหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544คู่มือการจัดการเรียนรู้กลุ่มสาระการเรียนรู้ภาษาไทย**. กรุงเทพมหานคร: โรงพิมพ์องค์การรับส่งสินค้าและพัสดุภัณฑ์.

สิริรัตน์ วิภาสศิลป์. (2545). **การเปรียบเทียบวิธีซิปเทสท์และดีเอฟไอทีในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ**. วิทยานิพนธ์ปริญญาการศึกษาดุษฎีบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

สำราญ มีแจ้ง.(2544) . **สถิติขั้นสูงสำหรับการวิจัย**. กรุงเทพ: โรงพิมพ์นิชินแอดเวอร์ไทซิ่งกรุ๊ป.

สุมาลี แกล้วทนงค์. (2547) .**สาเหตุของการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรมต่างกัน**. วิทยานิพนธ์ปริญญามหาบัณฑิต สาขาวิชาวิจัยการศึกษา ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

สุวิมล ติรการนันท์. (2548). **การตรวจสอบความเป็นเอกมิติ. (ปรับปรุงใหม่**). กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

เสรี ชัดแช้ม. (2539) .**การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทล-แฮนส์เซลแบบปกติกับวิธีแมนเทล-แฮนส์เซลแบบแบ่งกลุ่มความสามารถของผู้สอบและความยากของแบบสอบ**. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

อรินทร์ น่วมถนอม . (2549). **การเปรียบเทียบวิธีโพลี-ซิปเทสท์ วิธีการถดถอยโลจิสติกแบบจัดอันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า**. วิทยานิพนธ์ปริญญาการศึกษาดุษฎีบัณฑิต บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.

อิทธิฤทธิ์ พงษ์ปิยะรัตน์. (2551). **การวิเคราะห์ข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ: การวิเคราะห์พหุระดับ**. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต ภาควิชาวิจัยและจิตวิทยาการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

อุทุมพร จามรมาน. (2537). **ทฤษฎีการวัดทางจิตวิทยา**. กรุงเทพมหานคร: ฟันนี่พับบลิชชิ่ง.

ภาษาอังกฤษ

Ackerman, T. A. (1992). A Didactic Explanation of Item Bias, Item Impact, and Item Validity from a Multidimensional Perspective. **Journal of Educational Measurement**, 29(1): 67-91.

Adams, R. J., Wilson, M., and Wang, W. (1997). The multidimensional random coefficients multinomial logit model. **Applied Psychological Measurement**, 21(1): 1-23.

Afshartous, D. and Leeuw, J. (2005). Prediction in multilevel models. **Journal of Educational and Behavioral Statistics** 30(2): 133 – 152.

Beretvas, S. N., and Williams, N. J. (2004). The use of hierarchical generalized linear model for item dimensionality assessment. **Journal of Educational Measurement**, 41(4): 379-395.

Binici, S. (2007). **Random-effect differential item functioning via hierarchical generalized linear model and generalized linear latent mixed model: a comparison of estimation methods**. Doctoral Dissertation. Department of Educational Psychology and Learning Systems, Faculty of Education, Florida State University. (Unpublished)

Briggs, D.C., and Wilson, M. (2003). An introduction to multidimensional measurement Using rasch model. **Journal of Applied Measurement**, 4(1): 87-100.

Bock, R. D. (1989). **Multilevel analysis of educational data**. San Diego: Academic Press.

Bolt, D. M. (2002). A Monte Carlo comparison of parametric and nonparametric polytomous DIF detection methods. **Applied Measurement in Educational**, 15(2): 113-141.

Bryk, A. S. and Raudenbush, S. W. (1992). **Hierarchical Linear Models: Applications and data analysis methods**. Newbury Park, California: Sage Publication.

Camilli, G., and Shepard, L. A. (1994). **Methods for identifying biased test Items**. California: Sage Publications, Inc.

Camilli, G. (1993). **The case against DIF techniques based on internal criteria: Do item bias procedures obscure test fairness issues?** Hillsdale, N. J.: Lawrence Erlbaum.

Chaimongkol, S. (2005). **Modeling differential item functioning (DIF) using multilevel logistic regression models: A Bayesian Perspective**. Doctoral Dissertation. Department of Educational Psychology and Learning Systems, Faculty of Education, Florida State University. (Unpublished)

Chaimongkol, S., Huffer, W. F., and Kamata A. (2006). An explanatory different item functioning (DIF) model by the WinBUGS 1.4. **Songklanakarin Journal of Science and Technology**, 29(1): 449-458.

Cheong, Y. F. (2006). Analysis of school context effects on differential item functioning using hierarchical generalized linear models. **International Journal of Testing**, 6(1): 57-79.

Cheong, Y. F., and Raudenbush, S. W. (2000). Measurement and structural models for children's problem behaviors. **Psychological methods**, 5(4): 477-495.

Clauser, R. E., and Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. **Educational Measurement: Issues and Practice**, 17(1): 31-44.

Cohen, A.S. and Bolt, D. M. (2005). A mixture model analysis of differential item functioning. **Journal of Educational Measurement**, 42(2): 13-148.

Donoghue, J., Holland, P., and Thayer, D. (1993). A Monte Carlo study of factors that affect the Mantel-Haenszel and standardization measures of differential item functioning. In P.W. Holland and H. Wainer (Eds.), **Differential item functioning** pp. 137-166. Hillsdale, NJ: Erlbaum.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P.W. Holland and H. Wainer (Eds.), **Differential item functioning** pp. 46-66. Hillsdale, NJ: Erlbaum.

Elosua, P. and Jaúregui, A. L. (2007). Potential Sources of differential item functioning in the adaptation of tests. **International Journal of Testing**, *7*(1): 39–52.

Embretson, S.E. and Reise, S.P. (2000). **Item response theory for psychologists**. Mahwah: Lawrence Erlbaum Associates.

Fidalgo, A. M., Hashimoto, K., Bartram, D., and Muzis, J. (2007). Empirical Bayes versus standard Mantel-Haenzel statistics for detecting differential item functioning under small sample conditions. **Journal of Experimental Education**, 75(4): 293-314.

Finch, H. (2005). The MIMIC model as a method for detecting DIF: comparison with Mantel-Haenszel, SIBTEST, and the IRT Likelihood Ratio. **Applied Psychological Measurement**, 29(4): 278-295.

French, A. W. and Miller T. R. (1996). Logistic Regression and its use in detecting Differential item functioning in polytomous items. **Journal of Education Measurement**, 33(3): 315–332.

Guilford, J.P. and B. Fruchter. (1978). **Fundamental Statistics in Psychology and Education**. Singapore : McGraw-Hill.

Guo, G., and Zhao, H. (2000). **Multilevel modeling for binary data**. Annual Reviews. Sociological, 26(1): 441-462.

Hambleton, R. K., Clauser, B. E., Mazor, K. M., and Jones, R. W. (1993). **Advanced in the detection of differentially functioning test items. Laboratory of psychometric and evaluation research report No. 237**. Amherst, MA: University of Massachusetts, School of Education.

Hambleton, R. K., Swaminathan, H. and Rogers, H. J. (1991). **Fundamentals of item response theory**. Newbury Park, C. A.: Sage Publications.

Hambleton, R. K. and Swaminathan, H. Rogers, H. J. (1985). **Item Response Theory: Principles and applications**. Boston: Kluwer Nijhoff Publishing.

Heck, R. H. (2000). **An introduction to multilevel modeling techniques**. London. Lawrence Erlbaum associates.

Heck, R. H., and Thomas, S. L. (2000). **An introduction to multilevel modeling techniques**. Lawrence Erlbaum Associates.

Leeuw, J. D. and Meijer, E. (2008). **Handbook of Multilevel Analysis**. London. Springer.

Kamata, A. (1998). **Some generalizations of the Rasch model: an application of the hierarchical generalized linear model**. Doctoral Dissertation. Department of Educational Psychology, Faculty of Education, Michigan State University. (Unpublished)

Kamata, A. (1998). One-parameter hierarchical generalized linear logistic model: an application of HGLM to IRT. **Applied Psychological Measurement**, 29(1): 128-135.

Kamata, A. (2001). Item analysis by the hierarchical generalized linear model. **Journal of Educational Measurement**, 38(1), 79-93.

Kamata, A. (2002). Procedure to perform item response analysis by hierarchical generalized linear model. **Applied Psychological Measurement**, 31(2): 178-196.

Kamata, A. (2009). The Global Issues and Trends in Educational Statistics: Merging Psychometric and Statistical Models. **International Journal of Testing**, 7(3): 149–162.

Kamata, A., and Binici, S. (2003). Random Effect DIF Analysis via Hierarchical **Generalized** Linear Modeling. **Applied Measurement in Education**, 7(4): 259-274.

Kamata, A., Chaimongkol, S., Genc, E., and Bilir, K. (2005). Random-Effect Differential Item Functioning Across Group Unites by the Hierarchical Generalized Linear Model. **Applied Measurement in Education**, 8(3): 245-255.

Kamata, A. and Cheong, F. (2007). Multilevel Rasch model. In M. von Davier and C. H. Carstensen (Eds). **Multivariate and mixture distribution Rasch models: Extensions and applications** pp. 217-232. New York: Springer.

Kamata, A., and Vaughn, B. K. (2004). An Introduction to Differential Item Functioning Analysis. Learning Disabilities: **A Contemproary Journal**, 2(2): 49-64.

Kim, S. H., and Cohen, A. S. (1991). A comparison of two area measures for detecting differential item functioning. **Applied Psychological Measurement**, 15(1): 221-235.

Kim, S. H., and Cohen, A. S. (1995). A comparison of Lord's chi-square, Raju's measures, and the likelihood ratio test on detection of differential item functioning. **Applied Measurement in Education**, 8(2): 213-234.

Kim, S. H., (2000). An investigation of the Likelihood Ratio Test, the Mantel Test and the Generalized Mantel-Haenzel Test of DIF. **Applied Measurement in Education**, 13(1): 155-169.

Kim, S. H., Cohen, A. S., Alagoz, C. and Kim, S. W. (2007). **DIF** detection and effect size measures for polytomously scored items. **Journal of Educational Measurement**, 4(2): 93-116.

Lee, V. E. (2000). Using hierarchical linear modeling to study social context: The case of school effects. **Educational Psychologist**, 35(2): 125-141.

Leeuw, J. D. (2007). **Handbook of Multilevel Analysis**. New York: Springer.

Lei, P. W., Chen, S. Y., and Yu, L. (2006). Comparing methods of assessing differential item functioning in a computerized adaptive testing environment. **Journal of Educational Measurement**, 43(3): 245-264.

Liu,O.L., Wilson,M., and Paek, I. (2008). A multidimensional rasch analysis of gender differences in PISA mathematics. **Journal of Applied Measurement**, 9(1): 18-35.

Maas. C. and Hox. J. (2004). Robustness issues in multilevel regression analysis. **Statistica Neerlandica**, 58(3): 307 – 330.

Maas. C. and Hox. J. (2004). Sufficient sample size for multilevel modeling. **Methodology**, 58(3): 307 – 330.

Mazor, K. M., Clauser, B. E., and Hambleton, R. K. (1994). Identification of nonuniform differential item Functioning using a variation of the Mantel-Haenszel procedure. **Educational and Psychological Measurement**, 16(2): 211-245.

McNamara and Roever. (2004). **Psychometric approaches to fairness: bias and DIF**. New York: Springer.

Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. **Journal of Educational Statistics**, 7(2): 105-107.

Muthén, L.K. and Muthén, B.O. (1998-2007). **Mplus User's Guide**. Fifth Edition. Los Angeles, CA: Muthén & Muthén.

Narayanan, P., and Swaminathan, H. (1996). Identification of items that show nonuniform DIF. **Applied Psychological Measurement**, 20(2): 134-156.

Narayanan, P., and Swaminathan, H. (1994). Performance of the Identification of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. **Applied Psychological Measurement**, 18(4): 235-256.

Park, T. (2006). Detecting DIF across different language and gender groups in the MELAB essay test using the logistic regression method. **Journal of Educational Measurement**, 43(3): 188-212.

Patarapichayatham, C., Kanjanawasee, S., and Kamata, A. (2009). Cross-level Two-way Differential Item Functioning Analysis Model by the Multilevel Rasch Modeling: Investigation of Parameter Estimates. **Journal of Educational Measurement**, 123(4): 288-302.

Patz, R. J., and Junker, B. W. (1999). A Straightforward approach to markov chain monte carlo methods for item response models. **Journal of Educational and Behavioral Statistics**, 24(2): 146-178.

Penfield, R. D. (2005). DIFAS: differential item functioning analysis system. **Applied Psychological Measurement**, 29(2): 150-165.

Penfield, R. D. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. **Journal of Educational Measurement**, 43(4), 295-312.

Penfield, R. D. (2001). Assessing differential item functioning across multiple groups: A comparison of three Mantel-Haenszel procedures. **Applied Measurement in Education**, 14(1): 235-259.

Penfield, R. D., and Lam, T. C. M. (2000). Assessing differential item functioning assessment. Review and recommendations. **Educational Measurement: Issues and Practice**, 19(3), 5-15.

Penfield, R. D., and Camilli, G. (2007). Differential item functioning and item bias. In S. Sinharay & C. R. Rao (Eds.) **Handbook of Statistics**, 26, 125-167.

Potenza, M.T., and Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. **Applied Psychological Measurement**. 19(2): 211-237.

Raudenbush, S. W. and Bryk, A. S. (2002). **Hierarchical Linear Model: An applications and data analysis methods**. California: Sage Publication.

Rogers, H. J. and Swaminathan, H. (1993). A comparison of Logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. **Applied Measurement in Education**, 19(2): 254-275.

Shealy, R. and Stout, W. F. (1993a). **An item response theory model for test bias**. Hillsdale, N.J.: Lawrence Erlbaum.

Shibata, R. (2002). Information Criteria for Statistical Model Selection. **Electronics and Communications in Japan**, 85(4): 32-38.

Snijders, T. A. and Bosker, R. J. (1999). **Multilevel analysis: An introduction to basic and advanced multilevel modeling**. Thousand Oaks, CA: Sage Publication.

Su, Y. H., and Wang, W. C. (2005). Efficiency of the Mantel, Generalized Mantel–Haenszel, and Logistic Discriminant function analysis methods in detecting differential item functioning for polytomous items. **Applied Measurement in Education**, 18(4): 313–350.

Swaminathan, H. and Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. **Journal of Educational Measurement**, 27(2): 244-267.

Thissen, D., Steinberg, L., and Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In **Differential item functioning**. W. P. Holland; & H. Wainer. Hillsdale, N.J.: Lawrence Erlbaum.

Uttaro, T., and Millsap, R. E. (1994). Factors influencing the Mantel-Haenszel procedure in the detection of differential item functioning. **Applied Psychological Measurement**, 18(3): 325-341.

Wang, W. C., and Su, Y. H. (2004). Factors influencing the Mantel and Generalized Mantel-Haenszel methods for the assessment of differential item functioning in polytomous items. **Applied Psychological Measurement**, 28(4): 345-371.

Wang, W.-C., Yao, G., Tsai, Y.-J., Wang, J.-D., and Hsieh, C.-L. (2006). Validating, Improving reliability, and estimating correlation of the four subacales in the WHOQOL-BREF using multidimensional rasch analysis. **Quality of Life Research**, 607-620.

Wiberg, M. (2007). **Measuring and detecting differential item functioning in criterion-referenced licensing test: A theoretic comparison of methods**. Thousand Oaks, CA: Sage Publication.

Wiersma, W., and Jurs, S. G. (1990). **Test bias: educational measurement and testing**. Boston: Sage Publication.

Williams, N. J. and Beretvas, S. N. (2006). DIF identification using HGLM for polytomous items. **Applied Psychological Measurement**, 30(1): 22-42.

ภาคผนวก

ภาคผนวก ก
คำสั่งในการจำลองข้อมูลและการวิเคราะห์ข้อมูล

```
function(g=50,p,r101,r011,k=12,r=200)
{
st<-matrix(rep(0,r*46+r*44+r*44+r*42),nrow=r)
cat(c("Start iterations",date(),"\n"),file="d:/RMplus/progress.txt",append=T)
for (i in 1:r){
datagen(g,p,r011,r101,k)
system("d:/RMplus/fullmodel.bat", show.output.on.consol=F, invisible=F)
A<-scan("d:/RMplus/fullmodel.dat")
out1<-c(A[1:46])
system("d:/RMplus/without2w.bat", show.output.on.consol=F, invisible=F)
B<-scan("d:/RMplus/without2w.dat")
out2<-c(B[1:44])
system("d:/RMplus/withoutgld.bat", show.output.on.consol=F, invisible=F)
E<-scan("d:/RMplus/withoutgld.dat")
out3<-c(E[1:44])
system("d:/RMplus/without2wgld.bat", show.output.on.consol=F, invisible=F)
F<-scan("d:/RMplus/without2wgld.dat")
out4<-c(F[1:42])
out<-c(out1, out2, out3, out4)
st[i,]<-out
cat(c("Finished iteration", i, date(),"\n"),file="d:/RMplus/progress.txt",append=T)
cat(c(out,"\n"),file="d://RMplus/resultstep1.out",append=T)
data<-matrix(st, nrow=r, ncol=176)

## Find the mean of the mean of 4 model in step1##
Mean.of.gp.ab.dif.m1=mean(data[,4])
Mean.of.gp.ab.dif.m2=mean(data[,50])
Mean.of.gp.ab.dif.m3=mean(data[,94])
Mean.of.gp.ab.dif.m4=mean(data[,138])
Mean.of.2way.m1=mean(data[,5])
```

```
Mean.of.2way.m3=mean(data[,95])
Mean.of.3way.m1=mean(data[,6])
Mean.of.3way.m2=mean(data[,51])
Mean.of.3way.m3=mean(data[,96])
Mean.of.3way.m4=mean(data[,139])
Mean.of.gp.dif.m1=mean(data[,7])
Mean.of.gp.dif.m2=mean(data[,52])
Mean.of.in.ab.dif.m1=mean(data[,2])
Mean.of.in.ab.dif.m2=mean(data[,48])
Mean.of.in.ab.dif.m3=mean(data[,92])
Mean.of.in.ab.dif.m4=mean(data[,136])
Mean.of.in.dif.m1=mean(data[,3])
Mean.of.in.dif.m2=mean(data[,49])
Mean.of.in.dif.m3=mean(data[,93])
Mean.of.in.dif.m4=mean(data[,137])
Mean.of.r0.m1=mean(data[,1])
Mean.of.r0.m2=mean(data[,47])
Mean.of.r0.m3=mean(data[,91])
Mean.of.r0.m4=mean(data[,135])
Mean.of.u00.m1=mean(data[,8])
Mean.of.u00.m2=mean(data[,53])
Mean.of.u00.m3=mean(data[,97])
Mean.of.u00.m4=mean(data[,140])
allmeanstep1=cbind(Mean.of.gp.ab.dif.m1,Mean.of.gp.ab.dif.m2,Mean.of.gp.ab.dif.m3,Mea
n.of.gp.ab.dif.m4,Mean.of.2way.m1,Mean.of.2way.m3,Mean.of.3way.m1,Mean.of.3way.m2,
Mean.of.3way.m3,Mean.of.3way.m4,Mean.of.gp.dif.m1,Mean.of.gp.dif.m2,Mean.of.in.ab.dif.
m1,Mean.of.in.ab.dif.m2,Mean.of.in.ab.dif.m3,Mean.of.in.ab.dif.m4,Mean.of.in.dif.m1,Mean.
of.in.dif.m2,Mean.of.in.dif.m3,Mean.of.in.dif.m4,Mean.of.r0.m1,Mean.of.r0.m2,Mean.of.r0.m3
,Mean.of.r0.m4,Mean.of.u00.m1,Mean.of.u00.m2,Mean.of.u00.m3,Mean.of.u00.m4)
Meanstep1SHOW=matrix(allmeanstep1,ncol=1)
```

```
## Find the best model by using the AIC method ##
 AIC1=data[,44]
 AIC2=data[,88]
 AIC3=data[,132]
 AIC4=data[,174]
 show.AIC=matrix(c(AIC1,AIC2,AIC3,AIC4),ncol=4)
    r1=double(r)
   for( i in 1:r)
   {
   r1[i]<-min(show.AIC[i,1],show.AIC[i,2],show.AIC[i,3],show.AIC[i,4])
   }
   Model.aic=double(r)
   for (i in 1:r)
   {
   if (show.AIC[i,1]==r1[i]) Model.aic[i]<-1
   else if (show.AIC[i,2]==r1[i]) Model.aic[i]<-2
   else if (show.AIC[i,3]==r1[i]) Model.aic[i]<-3
   else Model.aic[i]<-4
   }
best.AIC=matrix(c(Model.aic),ncol=1)

  ## Count the best model of AIC##
ma1<-0
ma2<-0
ma3<-0
ma4<-0
for (i in 1:r)
{ if (best.AIC[i,1]==1) ma1<-ma1+1
  else if (best.AIC[i,1]==2) ma2<-ma2+1
  else if (best.AIC[i,1]==3) ma3<-ma3+1
```

```
  else ma4<-ma4+1
}
countAIC=c(ma1,ma2,ma3,ma4)
countAIC

  ## Find the best model by using the BIC method ##
 BIC1=data[,45]
 BIC2=data[,89]
 BIC3=data[,133]
 BIC4=data[,175]
 show.BIC=matrix(c(BIC1,BIC2,BIC3,BIC4),ncol=4)
    r2=double(r)
   for( i in 1:r)
   {
   r2[i]<-min(show.BIC[i,1],show.BIC[i,2],show.BIC[i,3],show.BIC[i,4])
   }
   Model.bic=double(r)
   for (i in 1:r)
   {
   if (show.BIC[i,1]==r2[i]) Model.bic[i]<-1
   else if (show.BIC[i,2]==r2[i]) Model.bic[i]<-2
   else if (show.BIC[i,3]==r2[i]) Model.bic[i]<-3
   else Model.bic[i]<-4
   }
best.BIC=matrix(c(Model.bic),ncol=1)

  ## Count the best model of BIC##
mb1<-0
mb2<-0
mb3<-0
```

```
mb4<-0
for (i in 1:r)
{ if (best.BIC[i,1]==1) mb1<-mb1+1
  else if (best.BIC[i,1]==2) mb2<-mb2+1
  else if (best.BIC[i,1]==3) mb3<-mb3+1
  else mb4<-mb4+1
}
countBIC=c(mb1,mb2,mb3,mb4)

## Find the best model by using the ABIC method ##
 ABIC1=data[,46]
 ABIC2=data[,90]
 ABIC3=data[,134]
 ABIC4=data[,176]
 show.ABIC=matrix(c(ABIC1,ABIC2,ABIC3,ABIC4),ncol=4)
    r3=double(r)
   for( i in 1:r)
   {
   r3[i]<-min(show.ABIC[i,1],show.ABIC[i,2],show.ABIC[i,3],show.ABIC[i,4])
   }
   Model.abic=double(r)
   for (i in 1:r)
   {
   if (show.ABIC[i,1]==r3[i]) Model.abic[i]<-1
   else if (show.ABIC[i,2]==r3[i]) Model.abic[i]<-2
   else if (show.ABIC[i,3]==r3[i]) Model.abic[i]<-3
   else Model.abic[i]<-4
   }
best.ABIC=matrix(c(Model.abic),ncol=1)
```

```
## Count the best model of ABIC##
mab1<-0
mab2<-0
mab3<-0
mab4<-0
for (i in 1:r)
{ if (best.ABIC[i,1]==1) mab1<-mab1+1
  else if (best.ABIC[i,1]==2) mab2<-mab2+1
  else if (best.ABIC[i,1]==3) mab3<-mab3+1
  else mab4<-mab4+1
}
countABIC=c(mab1,mab2,mab3,mab4)
## Results of the best model of using three model selections :AIC BIC ABIC ##
best=cbind(best.AIC,best.BIC,best.ABIC)

## Find the best model by using 2of3 ##
   twoOf3=double(r)
   for( i in 1:r)
{
   if      (best[i,1]==best[i,2] & best[i,2]==best[i,3]) {twoOf3[i]<-best[i,1]}
   else if (best[i,1]==best[i,2] & best[i,2]!=best[i,3]) {twoOf3[i]<-best[i,1]}
   else if (best[i,1]==best[i,3] & best[i,2]!=best[i,3]) {twoOf3[i]<-best[i,1]}
   else if (best[i,1]!=best[i,2] & best[i,2]==best[i,3]) {twoOf3[i]<-best[i,2]}
}
best.2of3=matrix(c(twoOf3),ncol=1)
## Count the best model of 2of3##
m2of3.1<-0
m2of3.2<-0
m2of3.3<-0
m2of3.4<-0
```

```
for (i in 1:r)
{ if (best.2of3[i,1]==1)      m2of3.1<-m2of3.1+1
  else if (best.2of3[i,1]==2) m2of3.2<-m2of3.2+1
  else if (best.2of3[i,1]==3) m2of3.3<-m2of3.3+1
  else if (best.2of3[i,1]==4) m2of3.4<-m2of3.4+1
}
count2of3=c(m2of3.1,m2of3.2,m2of3.3,m2of3.4)

##### P-value Model Selection #####
## two-way interaction ##
  est2way<-data[,5]
  se2way<-data[,25]
  est<-matrix(c(est2way),ncol=1)
  se<-matrix(c(se2way),ncol=1)
  pLower=est-1.96*se
  pUpper=est+1.96*se
  sig=double(r)
  for( i in 1:r)
{ if (pLower[i,1] < 0 & pUpper[i,1] > 0) sig[i]=0
  else sig[i] = 1
}
sig2way=matrix(c(sig),ncol=1)

## Group level Dif ##
  estGdif<-data[,7]
  se2Gdif<-data[,27]
  estGdif<-matrix(c(estGdif),ncol=1)
  seGdif<-matrix(c(se2Gdif),ncol=1)
  pLowerGdif=estGdif-1.96*se2Gdif
  pUpperGdif=estGdif+1.96*se2Gdif
```

```
  sigGdif=double(r)
  for( i in 1:r)
{ if (pLowerGdif[i,1] < 0 & pUpperGdif[i,1] > 0) sigGdif[i]=0
  else sigGdif[i] = 1
}
 sigG=matrix(c(sigGdif),ncol=1)
## Find the best model by P-Value Method ##
SigPv<-matrix(c(sig2way,sigG),ncol=2)
mp=double(r)
for ( i in 1:r) {
    if (SigPv[i,1]==1 & SigPv[i,2]==1) mp[i]<-1
    else if(SigPv[i,1]==0 & SigPv[i,2]==1) mp[i]<-2
    else if(SigPv[i,1]==1 & SigPv[i,2]==0) mp[i]<-3
    else mp[i]<-4
}
best.pvalue<-matrix(c(mp),ncol=1)
 ## Count the best model of P-Value Method##
p1<-0
p2<-0
p3<-0
p4<-0
for (i in 1:r)
{ if (best.pvalue[i,1]==1)      p1<-p1+1
  else if (best.pvalue[i,1]==2) p2<-p2+1
  else if (best.pvalue[i,1]==3) p3<-p3+1
  else if (best.pvalue[i,1]==4) p4<-p4+1
}
countPval=c(p1,p2,p3,p4)
```

```
### Find estimate and s.e. value of 4 fixed effects: group-level ability difference, 3-way
interaction, individual-level ability difference, individual-level dif and 2 random effects:
individual ability within group, and group ability###
est.gr.ab.dif=matrix(double(r),nrow=r,ncol=5)
se.gr.ab.dif=matrix(double(r),nrow=r,ncol=5)
est.3way=matrix(double(r),nrow=r,ncol=5)
se.3way=matrix(double(r),nrow=r,ncol=5)
est.in.dif=matrix(double(r),nrow=r,ncol=5)
se.in.dif=matrix(double(r),nrow=r,ncol=5)
est.in.ab.dif=matrix(double(r),nrow=r,ncol=5)
se.in.ab.dif=matrix(double(r),nrow=r,ncol=5)
est.r0=matrix(double(r),nrow=r,ncol=5)
se.r0=matrix(double(r),nrow=r,ncol=5)
est.u00=matrix(double(r),nrow=r,ncol=5)
se.u00=matrix(double(r),nrow=r,ncol=5)
for( j in 1:5) {
  for (i in 1:r) {
     if (bestmodel[i,j]==1)
        { est.gr.ab.dif[i,j]<-data[i,4];se.gr.ab.dif[i,j]<-data[i,24];
          est.3way[i,j]<-data[i,6]; se.3way[i,j]<-data[i,26];
          est.in.dif[i,j]<-data[i,3];se.in.dif[i,j]<-data[i,23];
          est.in.ab.dif[i,j]<-data[i,2];se.in.ab.dif[i,j]<-data[i,22];
          est.r0[i,j]<-data[i,1];se.r0[i,j]<-data[i,21];
          est.u00[i,j]<-data[i,8];se.u00[i,j]<-data[i,28]}
    else if (bestmodel[i,j]==2)
        { est.gr.ab.dif[i,j]<-data[i,50];se.gr.ab.dif[i,j]<-data[i,69];
          est.3way[i,j]<-data[i,51];se.3way[i,j]<-data[i,70];
          est.in.dif[i,j]<-data[i,49];se.in.dif[i,j]<-data[i,68];
          est.in.ab.dif[i,j]<-data[i,48];se.in.ab.dif[i,j]<-data[i,67];
          est.r0[i,j]<-data[i,47];se.r0[i,j]<-data[i,66];
```

```
        est.u00[i,j]<-data[i,53];se.u00[i,j]<-data[i,72]}
    else if (bestmodel[i,j]==3)
       { est.gr.ab.dif[i,j]<-data[i,94];se.gr.ab.dif[i,j]<-data[i,113];
        est.3way[i,j]<-data[i,96];se.3way[i,j]<-data[i,115];
        est.in.dif[i,j]<-data[i,93];se.in.dif[i,j]<-data[i,112];
        est.in.ab.dif[i,j]<-data[i,92];se.in.ab.dif[i,j]<-data[i,111];
        est.r0[i,j]<-data[i,91];se.r0[i,j]<-data[i,110];
        est.u00[i,j]<-data[i,97];se.u00[i,j]<-data[i,116]}
    else if (bestmodel[i,j]==4)
      { est.gr.ab.dif[i,j]<-data[i,138];se.gr.ab.dif[i,j]<-data[i,156];
       est.3way[i,j]<-data[i,139];se.3way[i,j]<-data[i,157];
       est.in.dif[i,j]<-data[i,137];se.in.dif[i,j]<-data[i,155];
       est.in.ab.dif[i,j]<-data[i,136];se.in.ab.dif[i,j]<-data[i,154];
       est.r0[i,j]<-data[i,135];se.r0[i,j]<-data[i,153];
       est.u00[i,j]<-data[i,140];se.u00[i,j]<-data[i,158]}
    else
      { est.gr.ab.dif[i,j]<-0;se.gr.ab.dif[i,j]<-0;
       est.3way[i,j]<-0;se.3way[i,j]<-0;
       est.in.dif[i,j]<-0;se.in.dif[i,j]<-0;
       est.in.ab.dif[i,j]<-0;se.in.ab.dif[i,j]<-0;
       est.r0[i,j]<-0;se.r0[i,j]<-0;
       est.u00[i,j]<-0;se.u00[i,j]<-0}
}
}
MeanSe.aic=cbind(est.gr.ab.dif[,1],se.gr.ab.dif[,1],est.3way[,1],se.3way[,1],est.in.dif[,1],se.i
n.dif[,1],est.in.ab.dif[,1],se.in.ab.dif[,1],est.r0[,1],se.r0[,1],est.u00[,1],se.u00[,1])
MeanSe.bic=cbind(est.gr.ab.dif[,2],se.gr.ab.dif[,2],est.3way[,2],se.3way[,2],est.in.dif[,2],se.i
n.dif[,2],est.in.ab.dif[,2],se.in.ab.dif[,2],est.r0[,2],se.r0[,2],est.u00[,2],se.u00[,2])
MeanSe.abic=cbind(est.gr.ab.dif[,3],se.gr.ab.dif[,3],est.3way[,3],se.3way[,3],est.in.dif[,3],se
.in.dif[,3],est.in.ab.dif[,3],se.in.ab.dif[,3],est.r0[,3],se.r0[,3],est.u00[,3],se.u00[,3])
```

```
MeanSe.2of3=cbind(est.gr.ab.dif[,4],se.gr.ab.dif[,4],est.3way[,4],se.3way[,4],est.in.dif[,4],se
.in.dif[,4],est.in.ab.dif[,4],se.in.ab.dif[,4],est.r0[,4],se.r0[,4],est.u00[,4],se.u00[,4])
MeanSe.Pval=cbind(est.gr.ab.dif[,5],se.gr.ab.dif[,5],est.3way[,5],se.3way[,5],est.in.dif[,5],se
.in.dif[,5],est.in.ab.dif[,5],se.in.ab.dif[,5],est.r0[,5],se.r0[,5],est.u00[,5],se.u00[,5])

### 1.Find mean of mean and se value : AIC method###
MeanOfEst.aic.gr.ab.dif=mean(est.gr.ab.dif[,1])
MeanOfSe.aic.gr.ab.dif=mean(se.gr.ab.dif[,1])
MeanOfEst.aic.3way=mean(est.3way[,1])
MeanOfSe.aic.3way=mean(se.3way[,1])
MeanOfEst.aic.in.dif=mean(est.in.dif[,1])
MeanOfSe.aic.in.dif=mean(se.in.dif[,1])
MeanOfEst.aic.in.ab.dif=mean(est.in.ab.dif[,1])
MeanOfSe.aic.in.ab.dif=mean(se.in.ab.dif[,1])
MeanOfEst.aic.r0=mean(est.r0[,1])
MeanOfSe.aic.r0=mean(se.r0[,1])
MeanOfEst.aic.u00=mean(est.u00[,1])
MeanOfSe.aic.u00=mean(se.u00[,1])
MeanSe.mean.aic=c
(MeanOfEst.aic.gr.ab.dif,MeanOfSe.aic.gr.ab.dif,MeanOfEst.aic.3way,MeanOfSe.aic.3way,
MeanOfEst.aic.in.dif,MeanOfSe.aic.in.dif,MeanOfEst.aic.in.ab.dif,MeanOfSe.aic.in.ab.dif,Me
anOfEst.aic.r0,MeanOfSe.aic.r0,MeanOfEst.aic.u00,MeanOfSe.aic.u00)

### 2.Find mean of mean and se value from BIC method###
MeanOfEst.bic.gr.ab.dif=mean(est.gr.ab.dif[,2])
MeanOfSe.bic.gr.ab.dif=mean(se.gr.ab.dif[,2])
MeanOfEst.bic.3way=mean(est.3way[,2])
MeanOfSe.bic.3way=mean(se.3way[,2])
MeanOfEst.bic.in.dif=mean(est.in.dif[,2])
MeanOfSe.bic.in.dif=mean(se.in.dif[,2])
```

```
MeanOfEst.bic.in.ab.dif=mean(est.in.ab.dif[,2])
MeanOfSe.bic.in.ab.dif=mean(se.in.ab.dif[,2])
MeanOfEst.bic.r0=mean(est.r0[,2])
MeanOfSe.bic.r0=mean(se.r0[,2])
MeanOfEst.bic.u00=mean(est.u00[,2])
MeanOfSe.bic.u00=mean(se.u00[,2])
MeanSe.mean.bic=c
(MeanOfEst.bic.gr.ab.dif,MeanOfSe.bic.gr.ab.dif,MeanOfEst.bic.3way,MeanOfSe.bic.3way,
MeanOfEst.bic.in.dif,MeanOfSe.bic.in.dif,MeanOfEst.bic.in.ab.dif,MeanOfSe.bic.in.ab.dif,Me
anOfEst.bic.r0,MeanOfSe.bic.r0,MeanOfEst.bic.u00,MeanOfSe.bic.u00)

### 3.Find mean of mean and se value from ABIC method###
MeanOfEst.abic.gr.ab.dif=mean(est.gr.ab.dif[,3])
MeanOfSe.abic.gr.ab.dif=mean(se.gr.ab.dif[,3])
MeanOfEst.abic.3way=mean(est.3way[,3])
MeanOfSe.abic.3way=mean(se.3way[,3])
MeanOfEst.abic.in.dif=mean(est.in.dif[,3])
MeanOfSe.abic.in.dif=mean(se.in.dif[,3])
MeanOfEst.abic.in.ab.dif=mean(est.in.ab.dif[,3])
MeanOfSe.abic.in.ab.dif=mean(se.in.ab.dif[,3])
MeanOfEst.abic.r0=mean(est.r0[,3])
MeanOfSe.abic.r0=mean(se.r0[,3])
MeanOfEst.abic.u00=mean(est.u00[,3])
MeanOfSe.abic.u00=mean(se.u00[,3])
MeanSe.mean.abic=c
(MeanOfEst.abic.gr.ab.dif,MeanOfSe.abic.gr.ab.dif,MeanOfEst.abic.3way,MeanOfSe.abic.3
way,MeanOfEst.abic.in.dif,MeanOfSe.abic.in.dif,MeanOfEst.abic.in.ab.dif,MeanOfSe.abic.in.
ab.dif,MeanOfEst.abic.r0,MeanOfSe.abic.r0,MeanOfEst.abic.u00,MeanOfSe.abic.u00)
```

```
### 4.Find mean of mean and se value from 2Of3 method###
MeanOfEst.2Of3.gr.ab.dif=mean(est.gr.ab.dif[,4])
MeanOfSe.2Of3.gr.ab.dif=mean(se.gr.ab.dif[,4])
MeanOfEst.2Of3.3way=mean(est.3way[,4])
MeanOfSe.2Of3.3way=mean(se.3way[,4])
MeanOfEst.2Of3.in.dif=mean(est.in.dif[,4])
MeanOfSe.2Of3.in.dif=mean(se.in.dif[,4])
MeanOfEst.2Of3.in.ab.dif=mean(est.in.ab.dif[,4])
MeanOfSe.2Of3.in.ab.dif=mean(se.in.ab.dif[,4])
MeanOfEst.2Of3.r0=mean(est.r0[,4])
MeanOfSe.2Of3.r0=mean(se.r0[,4])
MeanOfEst.2Of3.u00=mean(est.u00[,4])
MeanOfSe.2Of3.u00=mean(se.u00[,4])
MeanSe.mean.2of3=c(MeanOfEst.2Of3.gr.ab.dif,MeanOfSe.2Of3.gr.ab.dif,MeanOfEst.2Of3.
3way,MeanOfSe.2Of3.3way,MeanOfEst.2Of3.in.dif,MeanOfSe.2Of3.in.dif,MeanOfEst.2Of3.in
.ab.dif,MeanOfSe.2Of3.in.ab.dif,MeanOfEst.2Of3.r0,MeanOfSe.2Of3.r0,MeanOfEst.2Of3.u00
,MeanOfSe.2Of3.u00)

### 5.Find mean of mean and se value from P-Value method###
MeanOfEst.pval.gr.ab.dif=mean(est.gr.ab.dif[,5])
MeanOfSe.pval.gr.ab.dif=mean(se.gr.ab.dif[,5])
MeanOfEst.pval.3way=mean(est.3way[,5])
MeanOfSe.pval.3way=mean(se.3way[,5])
MeanOfEst.pval.in.dif=mean(est.in.dif[,5])
MeanOfSe.pval.in.dif=mean(se.in.dif[,5])
MeanOfEst.pval.in.ab.dif=mean( est.in.ab.dif[,5])
MeanOfSe.pval.in.ab.dif=mean(se.in.ab.dif[,5])
MeanOfEst.pval.r0=mean(est.r0[,5])
MeanOfSe.pval.r0=mean(se.r0[,5])
MeanOfEst.pval.u00=mean(est.u00[,5])
```

```
MeanOfSe.pval.u00=mean(se.u00[,5])
MeanSe.mean.pval=c
(MeanOfEst.pval.gr.ab.dif,MeanOfSe.pval.gr.ab.dif,MeanOfEst.pval.3way,MeanOfSe.pval.3
way,MeanOfEst.pval.in.dif,MeanOfSe.pval.in.dif,MeanOfEst.pval.in.ab.dif,MeanOfSe.pval.in.
ab.dif,MeanOfEst.pval.r0,MeanOfSe.pval.r0,MeanOfEst.pval.u00,MeanOfSe.pval.u00)
SumMeanSe=matrix(c(MeanSe.mean.aic,MeanSe.mean.bic,MeanSe.mean.abic,MeanSe.me
an.2of3,MeanSe.mean.pval),ncol=1)

## Calculate BIAS=Mean-True Value ##
## Extract Mean values ##
mean_aic=c(SumMeanSe[1,1],SumMeanSe[3,1],SumMeanSe[5,1],SumMeanSe[7,1],SumMe
anSe[9,1],SumMeanSe[11,1])
mean_bic=c(SumMeanSe[13,1],SumMeanSe[15,1],SumMeanSe[17,1],SumMeanSe[19,1],S
umMeanSe[21,1],SumMeanSe[23,1])
mean_abic=c(SumMeanSe[25,1],SumMeanSe[27,1],SumMeanSe[29,1],SumMeanSe[31,1],
SumMeanSe[33,1],SumMeanSe[35,1])
mean_2of3=c(SumMeanSe[37,1],SumMeanSe[39,1],SumMeanSe[41,1],SumMeanSe[43,1],
SumMeanSe[45,1],SumMeanSe[47,1])
mean_pval=c(SumMeanSe[49,1],SumMeanSe[51,1],SumMeanSe[53,1],SumMeanSe[55,1],S
umMeanSe[57,1],SumMeanSe[59,1])

## Calculate BIAS ##
All.Mean=c(mean_aic,mean_bic,mean_abic,mean_2of3,mean_pval)
BIAS=All.Mean-true.value
BIASShow=matrix(BIAS,ncol=1)

## Extract SE values ##
se_aic=c(SumMeanSe[2,1],SumMeanSe[4,1],SumMeanSe[6,1],SumMeanSe[8,1],SumMean
Se[10,1],SumMeanSe[12,1])
```

```
se_bic=c(SumMeanSe[14,1],SumMeanSe[16,1],SumMeanSe[18,1],SumMeanSe[20,1],Sum
MeanSe[22,1],SumMeanSe[24,1])
se_abic=c(SumMeanSe[26,1],SumMeanSe[28,1],SumMeanSe[30,1],SumMeanSe[32,1],Sum
MeanSe[34,1],SumMeanSe[36,1])
se_2of3=c(SumMeanSe[38,1],SumMeanSe[40,1],SumMeanSe[42,1],SumMeanSe[44,1],Sum
MeanSe[46,1],SumMeanSe[48,1])
se_pval=c(SumMeanSe[50,1],SumMeanSe[52,1],SumMeanSe[54,1],SumMeanSe[56,1],Sum
MeanSe[58,1],SumMeanSe[60,1])

## SE ##
All.SE=c(se_aic,se_bic,se_abic,se_2of3,se_pval)
SE=matrix(All.SE,ncol=1)

## Calculate RMSE ##
RMSE=(BIAS^2)+(All.SE^2)
RMSEShow=matrix(RMSE,ncol=1)
write(format(t(Meanstep1SHOW)),file="d:/RMplus/SumMeanStep1.out",ncol=1)
write(format(t(bestmodel)),file="d:/RMplus/bestmodel.out",ncol=5)
write(format(t(countbestmodel)),file="d:/RMplus/Countbestmodel.out",ncol=5)
write(format(t(allEstAndSe)),file="d:/RMplus/allMeanSeStep3.out",ncol=60)
write(format(t(SumMeanSe)),file="d:/RMplus/SumMeanSeStep3.out",ncol=1)
write(format(t(BIASShow)),file="d:/RMplus/BIAS.out",ncol=1)
write(format(t(SE)),file="d:/RMplus/SE.out",ncol=1)
write(format(t(RMSEShow)),file="d:/RMplus/RMSE.out",ncol=1)
}
```

ภาคผนวก ข
คำสั่งในการประมาณค่าพารามิเตอร์ด้วยโปรแกรม Mplus

```
TITLE: Complete model

DATA:  FILE IS 3level.dat;
VARIABLE:
        NAMES ARE sch x1 w1 u1-u12;
        CATEGORICAL = u1-u12;
        WITHIN = x1;
        BETWEEN = w1;
        CLUSTER = sch;
        usevariable = sch x1 w1 u1-u12;
        MISSING = ALL (999);
ANALYSIS:
        TYPE = random twolevel;
        integration = 4;
        processors = 4;
MODEL:
        %WITHIN%
                fw BY u1-u12@1;
                s0 | fw ON x1;
                s1 | u1 ON x1;
        %BETWEEN%
                fb BY u1-u12@1;
                fb ON w1;
                u1 ON w1;
                s0 ON w1;
                 s1 ON w1;
                s0@0;
                s1@0;
!OUTPUT:    TECH1 TECH8;
SAVEDATA:  results are fullmodel.dat;
```

```
TITLE: Incomplete model without two-way interaction

DATA:  FILE IS 3level.dat;
VARIABLE:
        NAMES ARE sch x1 w1 u1-u12;
        CATEGORICAL = u1-u12;
         WITHIN = x1;
        BETWEEN = w1;
        CLUSTER = sch;
        usevariable = sch x1 w1 u1-u12;
        MISSING = ALL (999);
ANALYSIS:
        TYPE = random twolevel;
       integration = 4;
       processors = 4;
MODEL:
        %WITHIN%
                fw BY u1-u12@1;
                s0 | fw ON x1;
                s1 | u1 ON x1;
        %BETWEEN%
                fb BY u1-u12@1;
                fb ON w1;
                u1 ON w1;
                s1 ON w1;
                s0@0;
                s1@0;
!OUTPUT:    TECH1 TECH8;
SAVEDATA:  results are without2w.dat;
```

```
TITLE: Incomplete model without cluster-level DIF

DATA:  FILE IS 3level.dat;
VARIABLE:
        NAMES ARE sch x1 w1 u1-u12;
        CATEGORICAL = u1-u12;
        WITHIN = x1;
        BETWEEN = w1;
        CLUSTER = sch;
        usevariable = sch x1 w1 u1-u12;
        MISSING = ALL (999);
ANALYSIS:
        TYPE = random twolevel;
        integration = 4;
        processors = 4;
MODEL:
        %WITHIN%
                fw BY u1-u12@1;
                s0 | fw ON x1;
                s1 | u1 ON x1;
        %BETWEEN%
                fb BY u1-u12@1;
                fb ON w1;
                s0 ON w1;
                s1 ON w1;
                s0@0;
                s1@0;
!OUTPUT:    TECH1 TECH8;
SAVEDATA:  results are withoutgld.dat;
```

```
TITLE: Incomplete model without both two-way interaction and cluster-level DIF

DATA:  FILE IS 3level.dat;
VARIABLE:
        NAMES ARE sch x1 w1 u1-u12;
        CATEGORICAL = u1-u12;
        WITHIN = x1;
        BETWEEN = w1;
        CLUSTER = sch;
        usevariable = sch x1 w1 u1-u12;
        MISSING = ALL (999);
ANALYSIS:
        TYPE = random twolevel;
        integration = 4;
        processors = 4;
MODEL:
        %WITHIN%
                fw BY u1-u12@1;
                s0 | fw ON x1;
                s1 | u1 ON x1;
        %BETWEEN%
                fb BY u1-u12@1;
                fb ON w1;
                s1 ON w1;
                s0@0;
                s1@0;
! OUTPUT:    TECH1 TECH8;
SAVEDATA:  results are without2wgld.dat;
```

ภาคผนวก ค
ตัวอย่างผลการสกัดค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (SD) จากโมเดลที่ดีที่สุด (best model) ของแต่ละเกณฑ์การประเมิน (model selection criteria) สำหรับ 1 เงื่อนไขการจำลอง

PEST: AIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	-0.076	0.111	0.250	0.000	-0.466	0.107	0.335	0.050	1.204	0.047	0.926	0.103
2	0.222	0.102	0.395	0.001	-0.422	0.100	0.223	0.050	1.012	0.050	0.987	0.080
3	-0.039	0.105	0.396	0.000	-0.622	0.088	0.175	0.046	1.115	0.035	0.947	0.090
4	-0.497	0.097	0.384	0.001	-0.566	0.106	0.177	0.048	1.066	0.048	1.390	0.113
5	-0.231	0.110	0.614	0.001	-0.707	0.112	0.315	0.047	1.142	0.052	0.895	0.109
6	0.100	0.110	0.380	0.001	-0.569	0.092	0.263	0.045	1.164	0.049	0.888	0.074
7	0.079	0.105	0.389	0.000	-0.435	0.078	0.245	0.048	1.102	0.041	1.102	0.122
8	-0.161	0.109	0.215	0.001	-0.441	0.095	0.137	0.043	1.141	0.047	0.992	0.130
9	0.107	0.104	0.650	0.001	-0.695	0.094	0.331	0.051	1.035	0.038	1.108	0.116
10	-0.080	0.108	0.331	0.001	-0.453	0.078	0.229	0.045	1.115	0.045	0.896	0.098
11	0.442	0.102	0.525	0.000	-0.671	0.079	0.220	0.039	1.045	0.040	1.250	0.132
12	0.595	0.105	0.313	0.000	-0.272	0.100	0.149	0.043	1.133	0.051	1.256	0.110
13	0.293	0.106	0.094	0.000	-0.500	0.095	0.196	0.040	1.048	0.049	1.047	0.105
14	0.176	0.111	0.135	0.000	-0.495	0.111	0.309	0.042	1.196	0.040	0.819	0.082
15	0.175	0.108	0.285	0.000	-0.502	0.102	0.169	0.052	1.101	0.046	0.750	0.095
16	-0.365	0.105	0.428	0.001	-0.793	0.116	0.152	0.046	1.101	0.046	1.009	0.092
17	0.158	0.101	0.389	0.001	-0.785	0.105	0.237	0.051	1.068	0.044	1.384	0.124
18	0.118	0.106	0.112	0.000	-0.589	0.092	0.360	0.046	1.076	0.044	0.997	0.129
19	-0.261	0.108	0.205	0.000	-0.504	0.120	0.253	0.043	1.129	0.051	1.022	0.111
20	0.587	0.106	0.017	0.000	-0.406	0.093	0.259	0.046	1.062	0.046	0.895	0.088
21	0.314	0.109	0.365	0.001	-0.675	0.071	0.313	0.052	1.126	0.056	0.802	0.075
22	0.039	0.100	0.612	0.001	-0.690	0.097	0.250	0.045	1.068	0.045	1.376	0.104
23	0.377	0.109	0.520	0.000	-0.603	0.100	0.378	0.049	1.169	0.045	1.090	0.079
24	0.092	0.105	0.095	0.000	-0.406	0.101	0.141	0.040	1.105	0.048	1.106	0.157
25	-0.279	0.103	0.491	0.001	-0.420	0.105	0.291	0.032	1.040	0.049	1.084	0.085
26	0.289	0.108	0.255	0.000	-0.394	0.106	-0.005	0.044	1.086	0.047	0.937	0.085
27	-0.283	0.100	0.387	0.001	-0.463	0.098	0.225	0.036	1.058	0.040	1.357	0.115
28	-0.145	0.108	0.181	0.000	-0.370	0.094	0.094	0.037	1.110	0.046	0.808	0.083
29	0.265	0.109	0.571	0.001	-0.564	0.109	0.304	0.040	1.173	0.052	0.992	0.083
30	-0.119	0.109	0.544	0.000	-0.503	0.098	0.289	0.050	1.107	0.052	0.712	0.089

PEST: AIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
31	-0.059	0.106	0.380	0.000	-0.534	0.094	0.169	0.044	1.050	0.046	0.828	0.086
32	0.623	0.109	0.245	0.000	-0.472	0.109	0.174	0.045	1.112	0.050	0.670	0.080
33	0.527	0.110	0.495	0.001	-0.758	0.083	0.230	0.040	1.146	0.047	0.682	0.064
34	0.542	0.105	0.222	0.000	-0.510	0.092	0.239	0.048	1.115	0.046	1.145	0.105
35	-0.337	0.103	0.236	0.000	-0.360	0.093	0.231	0.041	1.042	0.044	0.877	0.073
36	0.079	0.105	0.176	0.001	-0.423	0.103	0.320	0.040	1.111	0.044	1.253	0.135
37	-0.071	0.110	0.675	0.001	-0.679	0.100	0.228	0.046	1.110	0.043	0.805	0.113
38	0.116	0.106	0.667	0.001	-0.633	0.095	0.239	0.044	1.100	0.037	1.229	0.110
39	-0.196	0.103	0.117	0.000	-0.426	0.100	0.257	0.046	1.049	0.037	1.194	0.141
40	0.056	0.109	0.243	0.000	-0.450	0.098	0.177	0.049	1.098	0.048	0.903	0.122
41	0.338	0.106	0.366	0.000	-0.416	0.121	0.133	0.049	1.132	0.048	1.136	0.119
42	0.000	0.104	0.152	0.000	-0.573	0.102	0.133	0.050	1.009	0.045	0.946	0.095
43	-0.101	0.102	0.353	0.001	-0.700	0.102	0.130	0.042	1.055	0.046	1.155	0.129
44	-0.003	0.107	0.212	0.000	-0.363	0.115	0.154	0.044	1.193	0.051	1.527	0.158
45	0.385	0.105	0.148	0.000	-0.320	0.095	0.053	0.045	1.000	0.040	0.610	0.106
46	0.465	0.105	0.649	0.001	-0.723	0.102	0.158	0.038	1.123	0.050	1.250	0.108
47	0.014	0.111	0.196	0.000	-0.327	0.082	0.230	0.041	1.160	0.050	1.034	0.114
48	0.333	0.106	0.568	0.000	-0.706	0.110	0.218	0.047	1.045	0.038	0.701	0.087
49	0.582	0.106	0.674	0.001	-0.682	0.111	0.205	0.050	1.077	0.041	1.032	0.111
50	0.414	0.106	0.542	0.001	-0.816	0.101	0.293	0.052	1.119	0.042	0.854	0.116
51	-0.186	0.109	0.159	0.000	-0.512	0.102	0.255	0.048	1.142	0.046	0.904	0.087
52	0.145	0.106	0.377	0.000	-0.361	0.097	0.069	0.037	1.071	0.045	0.934	0.093
53	0.043	0.105	0.697	0.001	-0.822	0.098	0.276	0.059	1.081	0.038	1.091	0.120
54	-0.398	0.108	0.263	0.000	-0.360	0.098	0.128	0.046	1.023	0.035	0.744	0.092
55	-0.211	0.106	0.341	0.001	-0.407	0.126	0.233	0.043	1.127	0.044	0.992	0.109
56	-0.363	0.105	0.201	0.000	-0.331	0.085	0.171	0.046	1.078	0.046	0.938	0.086
57	0.179	0.107	0.336	0.000	-0.513	0.087	0.162	0.051	1.140	0.054	1.003	0.102
58	-0.128	0.108	0.343	0.000	-0.459	0.073	0.235	0.045	1.058	0.048	0.831	0.083
59	0.083	0.103	0.168	0.000	-0.464	0.089	0.122	0.049	1.050	0.037	1.106	0.093
60	-0.414	0.105	0.008	0.000	-0.474	0.105	0.287	0.054	1.056	0.045	0.952	0.089

PEST: AIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
61	0.247	0.112	0.248	0.000	-0.532	0.092	0.063	0.045	1.121	0.049	0.581	0.073
62	-0.074	0.104	0.508	0.000	-0.634	0.098	0.334	0.046	1.115	0.042	1.085	0.104
63	-0.521	0.106	0.734	0.001	-0.681	0.099	0.282	0.051	1.127	0.047	0.949	0.081
64	-0.456	0.105	0.176	0.000	-0.395	0.092	0.061	0.045	1.047	0.039	0.837	0.076
65	0.062	0.107	0.600	0.001	-0.704	0.093	0.222	0.044	1.065	0.046	0.674	0.085
66	0.136	0.107	0.545	0.001	-0.680	0.111	0.131	0.045	1.110	0.048	1.093	0.098
67	0.028	0.108	0.482	0.001	-0.645	0.100	0.222	0.045	1.166	0.052	1.083	0.112
68	0.445	0.110	0.448	0.001	-0.508	0.097	0.275	0.045	1.167	0.042	1.117	0.112
69	0.302	0.104	0.489	0.001	-0.582	0.102	0.245	0.041	1.048	0.037	0.993	0.088
70	-0.231	0.099	0.069	0.000	-0.439	0.119	0.220	0.050	1.071	0.037	1.294	0.121
71	0.314	0.105	0.561	0.001	-0.846	0.085	0.209	0.044	1.083	0.040	1.077	0.117
72	0.118	0.109	0.143	0.000	-0.569	0.102	0.314	0.046	1.037	0.039	0.707	0.064
73	0.521	0.104	0.474	0.000	-0.535	0.107	0.210	0.051	1.009	0.039	0.917	0.120
74	0.069	0.107	0.481	0.001	-0.577	0.094	0.253	0.043	1.057	0.043	0.520	0.060
75	-0.378	0.106	0.646	0.001	-0.769	0.092	0.310	0.045	1.154	0.042	1.084	0.102
76	0.205	0.105	0.193	0.000	-0.340	0.086	0.241	0.049	1.115	0.042	1.067	0.091
77	0.193	0.107	0.544	0.000	-0.601	0.080	0.097	0.054	1.133	0.052	1.088	0.160
78	0.133	0.112	-4.450	0.000	-0.352	0.112	0.281	0.045	1.170	0.051	0.839	0.125
79	0.021	0.105	0.250	0.000	-0.564	0.100	0.273	0.041	1.049	0.043	1.001	0.116
80	0.111	0.105	0.235	0.000	-0.491	0.082	0.245	0.039	1.071	0.044	1.194	0.118
81	0.377	0.098	0.227	0.000	-0.537	0.128	0.271	0.047	1.039	0.039	1.513	0.112
82	0.312	0.107	0.592	0.001	-0.605	0.110	0.165	0.048	1.196	0.053	1.348	0.099
83	-0.042	0.109	0.728	0.001	-0.756	0.090	0.099	0.041	1.049	0.048	0.737	0.127
84	0.102	0.107	0.211	0.000	-0.429	0.093	0.379	0.050	1.097	0.040	1.135	0.111
85	-0.444	0.104	0.525	0.001	-0.619	0.104	0.130	0.046	1.083	0.041	1.226	0.183
86	-0.131	0.109	0.425	0.001	-0.499	0.073	0.237	0.046	1.052	0.036	0.748	0.070
87	-0.014	0.111	0.366	0.000	-0.610	0.104	0.269	0.044	1.091	0.047	0.837	0.089
88	0.102	0.099	0.067	0.000	-0.225	0.083	0.161	0.046	0.970	0.045	1.164	0.117
89	-0.199	0.109	0.371	0.000	-0.486	0.102	0.152	0.047	1.077	0.047	0.819	0.096
90	-0.001	0.108	0.235	0.000	-0.515	0.112	0.303	0.043	1.093	0.034	1.001	0.101

PEST: AIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
91	0.335	0.107	0.567	0.001	-0.637	0.093	0.195	0.034	1.168	0.046	1.143	0.119
92	-0.109	0.105	0.709	0.001	-0.765	0.101	0.315	0.043	1.069	0.039	0.918	0.103
93	0.298	0.110	0.273	0.000	-0.475	0.091	0.235	0.046	1.128	0.036	0.959	0.092
94	-0.088	0.107	0.443	0.001	-0.549	0.096	0.258	0.045	1.059	0.038	0.832	0.089
95	0.209	0.104	0.734	0.001	-0.564	0.104	0.229	0.053	1.056	0.036	0.717	0.088
96	-0.222	0.105	0.342	0.001	-0.628	0.089	0.210	0.050	1.037	0.030	0.925	0.100
97	-0.299	0.114	0.105	0.000	-0.592	0.095	0.188	0.043	1.214	0.051	0.456	0.055
98	-0.079	0.109	0.499	0.000	-0.556	0.102	0.229	0.042	1.111	0.050	0.978	0.092
99	0.127	0.110	0.532	0.001	-0.724	0.096	0.175	0.053	1.163	0.048	0.970	0.093
100	0.414	0.107	0.252	0.000	-0.502	0.099	0.133	0.046	1.126	0.037	0.957	0.097
101	-0.398	0.106	0.552	0.000	-0.675	0.085	0.230	0.037	1.014	0.038	0.766	0.086
102	0.491	0.109	0.550	0.001	-0.791	0.098	0.384	0.054	1.100	0.044	0.786	0.085
103	0.718	0.103	0.267	0.000	-0.477	0.099	0.167	0.041	1.033	0.045	1.072	0.092
104	-0.308	0.111	0.537	0.001	-0.588	0.103	0.193	0.044	1.166	0.041	0.898	0.091
105	0.452	0.112	0.355	0.000	-0.601	0.107	0.160	0.049	1.286	0.046	1.170	0.130
106	0.070	0.099	0.319	0.000	-0.535	0.090	0.303	0.050	1.064	0.043	1.075	0.074
107	0.240	0.108	0.180	0.000	-0.339	0.092	0.229	0.047	1.147	0.046	0.984	0.105
108	0.466	0.109	0.637	0.001	-0.634	0.103	0.280	0.039	1.080	0.042	0.859	0.103
109	0.161	0.110	0.146	0.000	-0.684	0.102	0.275	0.041	1.114	0.047	0.721	0.080
110	-0.250	0.108	0.337	0.001	-0.472	0.079	0.164	0.040	1.098	0.043	0.916	0.096
111	0.205	0.107	0.771	0.001	-0.732	0.085	0.351	0.044	0.968	0.036	0.597	0.080
112	-0.574	0.110	0.425	0.000	-0.602	0.088	0.280	0.047	1.121	0.044	0.927	0.108
113	0.070	0.107	0.530	0.001	-0.631	0.121	0.094	0.048	1.148	0.049	1.007	0.094
114	-0.472	0.106	0.730	0.001	-0.832	0.090	0.256	0.041	1.107	0.050	1.046	0.116
115	0.631	0.105	0.104	0.000	-0.420	0.090	0.240	0.046	1.084	0.037	1.150	0.155
116	0.460	0.109	0.363	0.000	-0.698	0.083	0.180	0.047	1.108	0.044	0.652	0.085
117	0.631	0.104	0.883	0.001	-0.859	0.119	0.260	0.037	1.123	0.050	1.222	0.148
118	0.288	0.103	-6.070	0.000	-0.428	0.100	0.045	0.040	1.025	0.046	1.110	0.088
119	0.387	0.106	0.678	0.001	-0.597	0.098	0.338	0.055	1.136	0.041	1.054	0.123
120	-0.179	0.105	0.589	0.001	-0.540	0.097	0.250	0.050	1.098	0.041	1.074	0.088

PEST: AIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
121	-0.097	0.095	0.455	0.001	-0.403	0.108	0.252	0.039	1.010	0.041	1.545	0.122
122	-0.043	0.108	0.596	0.001	-0.692	0.110	0.265	0.050	1.102	0.051	0.923	0.097
123	0.557	0.104	0.289	0.000	-0.287	0.096	0.239	0.048	1.170	0.046	1.337	0.115
124	-0.273	0.110	0.125	0.000	-0.649	0.095	0.121	0.054	1.159	0.048	0.929	0.116
125	0.254	0.108	0.484	0.001	-0.544	0.084	0.255	0.042	1.105	0.047	0.862	0.097
126	0.575	0.109	0.508	0.001	-0.507	0.100	0.300	0.045	1.227	0.049	1.081	0.085
127	-0.022	0.108	0.123	0.000	-0.689	0.092	0.204	0.042	1.131	0.050	0.997	0.094
128	0.086	0.099	0.281	0.000	-0.705	0.098	0.217	0.044	1.020	0.045	1.271	0.127
129	0.056	0.101	0.296	0.000	-0.559	0.118	0.115	0.045	1.130	0.040	1.513	0.172
130	0.457	0.108	0.326	0.001	-0.561	0.103	0.123	0.052	1.217	0.057	1.219	0.093
131	0.242	0.105	0.667	0.001	-0.594	0.096	0.268	0.048	1.086	0.042	0.919	0.129
132	0.007	0.106	0.003	0.000	-0.307	0.093	0.270	0.043	1.098	0.037	1.039	0.082
133	0.352	0.108	0.740	0.000	-0.738	0.091	0.269	0.054	1.148	0.053	1.030	0.116
134	0.386	0.103	0.267	0.000	-0.472	0.117	0.248	0.041	1.078	0.046	0.908	0.103
135	-0.086	0.110	0.283	0.001	-0.304	0.083	0.257	0.045	1.193	0.045	1.049	0.116
136	-0.297	0.106	0.260	0.000	-0.421	0.088	0.207	0.047	1.078	0.041	0.972	0.117
137	0.223	0.107	0.003	0.000	-0.337	0.083	0.104	0.048	1.119	0.040	0.907	0.098
138	0.375	0.104	-2.187	0.000	-0.404	0.109	0.292	0.044	1.101	0.048	1.098	0.094
139	0.222	0.109	0.089	0.000	-0.190	0.093	0.243	0.050	1.110	0.050	1.094	0.111
140	0.288	0.110	0.332	0.001	-0.456	0.102	0.244	0.049	1.149	0.043	0.594	0.056
141	-0.246	0.112	0.063	0.000	-0.265	0.101	0.086	0.041	1.074	0.046	0.728	0.088
142	-0.280	0.105	0.226	0.000	-0.503	0.091	0.215	0.040	1.067	0.056	0.978	0.091
143	0.118	0.108	-9.172	0.000	-0.414	0.095	0.243	0.053	1.115	0.037	0.871	0.097
144	0.040	0.105	0.581	0.000	-0.708	0.081	0.232	0.048	1.103	0.042	1.129	0.108
145	-0.057	0.105	0.618	0.001	-0.520	0.104	0.278	0.043	1.037	0.047	0.931	0.105
146	0.214	0.106	0.671	0.001	-0.602	0.097	0.212	0.045	1.126	0.042	1.135	0.111
147	0.118	0.108	0.351	0.000	-0.548	0.094	0.339	0.050	1.138	0.042	1.008	0.154
148	0.475	0.112	0.424	0.001	-0.722	0.105	0.206	0.040	1.162	0.042	0.757	0.071
149	0.613	0.112	0.432	0.001	-0.688	0.115	0.217	0.042	1.207	0.049	0.725	0.071
150	0.731	0.104	0.781	0.001	-0.697	0.098	0.159	0.043	1.031	0.047	0.998	0.089

PEST: AIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
151	0.237	0.104	0.367	0.000	-0.598	0.111	0.289	0.036	1.084	0.043	0.790	0.087
152	0.093	0.109	0.520	0.001	-0.631	0.093	0.354	0.050	1.120	0.043	0.645	0.075
153	0.605	0.106	0.024	0.000	-0.521	0.093	0.181	0.037	1.067	0.042	1.122	0.103
154	0.173	0.105	0.522	0.001	-0.622	0.115	0.260	0.052	1.053	0.045	0.911	0.087
155	0.124	0.105	0.409	0.001	-0.494	0.093	0.224	0.042	1.045	0.047	0.696	0.070
156	0.318	0.110	0.408	0.001	-0.666	0.103	0.276	0.045	1.078	0.049	0.668	0.081
157	0.413	0.110	0.166	0.000	-0.599	0.101	0.234	0.044	1.028	0.041	0.869	0.105
158	0.063	0.108	0.349	0.001	-0.457	0.109	0.239	0.049	1.139	0.040	1.068	0.095
159	0.198	0.103	-1.576	0.000	-0.311	0.094	0.037	0.042	1.058	0.043	1.409	0.171
160	0.042	0.111	0.467	0.000	-0.689	0.098	0.267	0.044	1.131	0.045	0.668	0.074
161	0.282	0.109	0.591	0.001	-0.723	0.083	0.255	0.042	1.232	0.046	0.876	0.082
162	-0.069	0.104	0.395	0.000	-0.478	0.094	0.209	0.050	1.091	0.047	1.163	0.102
163	0.203	0.104	0.686	0.000	-0.882	0.103	0.281	0.036	1.046	0.046	1.082	0.114
164	0.482	0.106	0.653	0.000	-0.842	0.092	0.335	0.046	1.127	0.046	1.337	0.141
165	0.451	0.111	0.627	0.000	-0.596	0.095	0.191	0.047	1.151	0.051	0.773	0.087
166	0.445	0.103	0.581	0.001	-0.577	0.115	0.287	0.047	1.031	0.047	0.954	0.093
167	0.031	0.104	-1.073	0.000	-0.296	0.088	0.190	0.040	1.064	0.044	1.022	0.095
168	-0.012	0.102	0.187	0.000	-0.446	0.097	0.213	0.040	1.078	0.044	1.154	0.112
169	0.306	0.102	0.261	0.000	-0.325	0.082	0.295	0.045	1.042	0.045	1.187	0.081
170	0.327	0.104	0.182	0.000	-0.446	0.098	0.155	0.048	1.059	0.040	0.957	0.139
171	0.206	0.109	0.395	0.001	-0.555	0.092	0.301	0.043	1.162	0.057	0.956	0.139
172	-0.450	0.102	0.524	0.001	-0.789	0.116	0.254	0.043	1.042	0.035	1.192	0.195
173	0.345	0.108	0.460	0.001	-0.560	0.104	0.210	0.051	1.142	0.038	0.938	0.111
174	0.049	0.103	0.030	0.000	-0.459	0.101	0.265	0.055	1.104	0.041	1.346	0.130
175	-0.123	0.110	0.636	0.001	-0.892	0.110	0.264	0.041	1.172	0.050	0.820	0.074
176	-0.002	0.103	0.710	0.001	-0.917	0.088	0.186	0.047	1.098	0.056	1.136	0.128
177	0.040	0.105	0.380	0.000	-0.665	0.099	0.254	0.051	1.049	0.044	1.073	0.113
178	0.062	0.106	0.438	0.000	-0.356	0.101	0.206	0.046	1.047	0.046	0.971	0.102
179	-0.385	0.103	0.522	0.000	-0.581	0.107	0.298	0.049	1.114	0.044	1.254	0.170
180	-0.024	0.110	0.602	0.000	-0.803	0.091	0.275	0.052	1.168	0.038	0.901	0.089

PEST: AIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
181	0.434	0.102	0.391	0.000	-0.466	0.101	0.255	0.041	0.976	0.042	0.970	0.099
182	0.051	0.106	0.552	0.001	-0.670	0.118	0.229	0.048	1.099	0.042	1.082	0.101
183	-0.062	0.107	0.466	0.001	-0.602	0.099	0.195	0.047	1.123	0.055	1.045	0.098
184	0.739	0.104	0.353	0.000	-0.370	0.107	0.214	0.048	1.173	0.040	1.363	0.140
185	0.326	0.109	0.241	0.000	-0.627	0.097	0.314	0.051	1.114	0.044	0.810	0.098
186	0.276	0.103	0.203	0.000	-0.538	0.104	0.349	0.053	0.994	0.045	0.988	0.099
187	0.304	0.110	-2.261	0.000	-0.209	0.100	0.260	0.051	1.186	0.042	0.808	0.087
188	0.187	0.107	0.517	0.000	-0.707	0.086	0.253	0.039	1.076	0.044	0.852	0.076
189	-0.012	0.105	-9.203	0.000	-0.249	0.114	0.199	0.047	1.065	0.046	0.829	0.092
190	0.288	0.110	0.600	0.001	-0.569	0.104	0.122	0.052	1.055	0.044	0.696	0.081
191	-0.018	0.106	0.286	0.000	-0.384	0.105	0.269	0.043	1.135	0.052	1.101	0.096
192	0.107	0.102	0.615	0.001	-0.653	0.088	0.268	0.040	1.013	0.035	0.952	0.115
193	-0.009	0.106	0.347	0.000	-0.347	0.102	0.089	0.041	1.203	0.048	0.764	0.117
194	0.213	0.106	0.749	0.001	-0.643	0.105	0.161	0.051	1.028	0.042	0.813	0.101
195	0.164	0.105	0.468	0.000	-0.629	0.092	0.346	0.046	1.032	0.048	1.069	0.124
196	0.469	0.107	0.250	0.000	-0.354	0.100	0.098	0.043	1.046	0.041	0.910	0.112
197	0.102	0.109	0.496	0.001	-0.536	0.094	0.141	0.043	1.050	0.040	0.676	0.069
198	-0.121	0.111	0.318	0.001	-0.567	0.113	0.215	0.051	1.173	0.052	0.832	0.105
199	0.167	0.104	0.464	0.000	-0.535	0.106	0.245	0.045	1.030	0.047	1.056	0.123
200	-0.325	0.111	0.158	0.000	-0.678	0.099	0.272	0.050	1.129	0.043	0.603	0.061

PEST: BIC

R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	-0.109	0.111	0.632	0.001	-0.653	0.110	0.330	0.050	1.205	0.047	0.913	0.102
2	0.222	0.102	0.395	0.001	-0.422	0.100	0.223	0.050	1.012	0.050	0.987	0.080
3	-0.016	0.106	0.697	0.000	-0.771	0.090	0.176	0.046	1.114	0.035	0.946	0.090
4	-0.497	0.097	0.384	0.001	-0.566	0.106	0.177	0.048	1.066	0.048	1.390	0.113
5	-0.231	0.110	0.614	0.001	-0.707	0.112	0.315	0.047	1.142	0.052	0.895	0.109
6	0.100	0.110	0.380	0.001	-0.569	0.092	0.263	0.045	1.164	0.049	0.888	0.074
7	0.038	0.105	0.626	0.000	-0.552	0.080	0.250	0.048	1.105	0.041	1.124	0.122
8	-0.027	0.109	0.253	0.001	-0.460	0.095	0.253	0.046	1.147	0.047	0.995	0.130
9	0.107	0.104	0.650	0.001	-0.695	0.094	0.331	0.051	1.035	0.038	1.108	0.116
10	-0.080	0.108	0.331	0.001	-0.453	0.078	0.229	0.045	1.115	0.045	0.896	0.098
11	0.364	0.101	0.456	0.000	-0.636	0.079	0.133	0.041	1.049	0.040	1.279	0.134
12	0.546	0.104	0.550	0.001	-0.390	0.101	0.151	0.043	1.132	0.051	1.260	0.109
13	0.304	0.103	0.291	0.001	-0.598	0.096	0.196	0.040	1.046	0.049	1.049	0.105
14	0.168	0.112	0.374	0.001	-0.613	0.112	0.313	0.042	1.199	0.040	0.820	0.082
15	0.310	0.110	0.556	0.001	-0.638	0.104	0.282	0.054	1.104	0.046	0.736	0.094
16	-0.258	0.105	0.471	0.001	-0.814	0.116	0.249	0.048	1.102	0.046	1.012	0.092
17	0.158	0.101	0.389	0.001	-0.785	0.105	0.237	0.051	1.068	0.044	1.384	0.124
18	0.133	0.106	0.399	0.001	-0.731	0.095	0.360	0.046	1.075	0.044	0.999	0.129
19	-0.255	0.109	0.468	0.001	-0.641	0.122	0.252	0.043	1.130	0.051	1.003	0.110
20	0.587	0.106	0.017	0.000	-0.406	0.093	0.259	0.046	1.062	0.046	0.895	0.088
21	0.314	0.109	0.365	0.001	-0.675	0.071	0.313	0.052	1.126	0.056	0.802	0.075
22	0.039	0.100	0.612	0.001	-0.690	0.097	0.250	0.045	1.068	0.045	1.376	0.104
23	0.377	0.109	0.520	0.000	-0.603	0.100	0.378	0.049	1.169	0.045	1.090	0.079
24	0.241	0.107	0.453	0.001	-0.589	0.104	0.243	0.042	1.103	0.048	1.090	0.155
25	-0.279	0.103	0.491	0.001	-0.420	0.105	0.291	0.032	1.040	0.049	1.084	0.085
26	0.393	0.107	0.514	0.001	-0.524	0.109	0.030	0.044	1.086	0.047	0.941	0.085
27	-0.283	0.100	0.387	0.001	-0.463	0.098	0.225	0.036	1.058	0.040	1.357	0.115
28	-0.026	0.108	0.552	0.000	-0.560	0.097	0.205	0.040	1.111	0.045	0.821	0.084
29	0.265	0.109	0.571	0.001	-0.564	0.109	0.304	0.040	1.173	0.052	0.992	0.083
30	-0.119	0.109	0.827	0.001	-0.641	0.101	0.291	0.050	1.106	0.052	0.713	0.089

PEST: BIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
31	-0.051	0.106	0.691	0.001	-0.687	0.097	0.170	0.044	1.051	0.046	0.840	0.087
32	0.634	0.108	0.531	0.000	-0.610	0.111	0.174	0.045	1.112	0.049	0.671	0.080
33	0.527	0.110	0.495	0.001	-0.758	0.083	0.230	0.040	1.146	0.047	0.682	0.064
34	0.557	0.105	0.462	0.001	-0.631	0.095	0.239	0.048	1.115	0.045	1.153	0.106
35	-0.180	0.104	0.552	0.000	-0.514	0.097	0.319	0.042	1.043	0.044	0.868	0.072
36	0.079	0.105	0.176	0.001	-0.423	0.103	0.320	0.040	1.111	0.044	1.253	0.135
37	-0.071	0.110	0.675	0.001	-0.679	0.100	0.228	0.046	1.110	0.043	0.805	0.113
38	0.116	0.106	0.667	0.001	-0.633	0.095	0.239	0.044	1.100	0.037	1.229	0.110
39	-0.085	0.102	0.372	0.001	-0.553	0.101	0.347	0.047	1.051	0.037	1.202	0.142
40	0.058	0.107	0.493	0.000	-0.580	0.099	0.177	0.049	1.097	0.048	0.916	0.123
41	0.471	0.104	0.619	0.001	-0.538	0.123	0.233	0.050	1.135	0.048	1.148	0.121
42	0.130	0.101	0.489	0.001	-0.741	0.106	0.255	0.053	1.010	0.045	0.985	0.097
43	-0.029	0.102	0.389	0.001	-0.718	0.102	0.221	0.044	1.055	0.046	1.167	0.129
44	0.011	0.106	0.473	0.000	-0.485	0.117	0.155	0.044	1.191	0.051	1.495	0.156
45	0.444	0.107	0.491	0.001	-0.489	0.098	0.076	0.045	0.996	0.040	0.608	0.105
46	0.524	0.106	0.687	0.001	-0.742	0.102	0.251	0.040	1.126	0.050	1.233	0.108
47	0.041	0.110	0.558	0.000	-0.503	0.086	0.231	0.041	1.159	0.050	1.006	0.112
48	0.363	0.105	0.816	0.001	-0.828	0.112	0.221	0.047	1.046	0.038	0.724	0.089
49	0.582	0.106	0.674	0.001	-0.682	0.111	0.205	0.050	1.077	0.041	1.032	0.111
50	0.414	0.106	0.542	0.001	-0.816	0.101	0.293	0.052	1.119	0.042	0.854	0.116
51	-0.209	0.110	0.431	0.001	-0.647	0.104	0.255	0.048	1.144	0.046	0.871	0.085
52	0.298	0.107	0.654	0.001	-0.500	0.099	0.179	0.039	1.074	0.045	0.925	0.093
53	0.095	0.105	0.730	0.001	-0.838	0.098	0.354	0.060	1.083	0.038	1.099	0.120
54	-0.291	0.108	0.574	0.001	-0.518	0.101	0.237	0.048	1.026	0.035	0.744	0.092
55	-0.211	0.106	0.341	0.001	-0.407	0.126	0.233	0.043	1.127	0.044	0.992	0.109
56	-0.289	0.106	0.571	0.000	-0.515	0.090	0.206	0.046	1.075	0.045	0.921	0.085
57	0.297	0.107	0.602	0.000	-0.647	0.089	0.253	0.052	1.143	0.054	0.987	0.101
58	-0.101	0.108	0.541	0.000	-0.561	0.075	0.235	0.045	1.058	0.048	0.826	0.083
59	0.117	0.101	0.523	0.001	-0.643	0.092	0.193	0.050	1.051	0.037	1.146	0.094
60	-0.414	0.105	0.008	0.000	-0.474	0.105	0.287	0.054	1.056	0.045	0.952	0.089

PEST: BIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
61	0.364	0.110	0.513	0.001	-0.665	0.094	0.159	0.047	1.120	0.049	0.587	0.073
62	0.002	0.106	0.623	0.001	-0.692	0.099	0.337	0.046	1.119	0.042	1.087	0.104
63	-0.521	0.106	0.734	0.001	-0.681	0.099	0.282	0.051	1.127	0.047	0.949	0.081
64	-0.304	0.105	0.438	0.000	-0.527	0.095	0.180	0.048	1.049	0.039	0.843	0.076
65	0.062	0.107	0.600	0.001	-0.704	0.093	0.222	0.044	1.065	0.046	0.674	0.085
66	0.270	0.106	0.606	0.001	-0.710	0.111	0.270	0.049	1.113	0.048	1.103	0.098
67	0.028	0.108	0.482	0.001	-0.645	0.100	0.222	0.045	1.166	0.052	1.083	0.112
68	0.445	0.110	0.448	0.001	-0.508	0.097	0.275	0.045	1.167	0.042	1.117	0.112
69	0.302	0.104	0.489	0.001	-0.582	0.102	0.245	0.041	1.048	0.037	0.993	0.088
70	-0.136	0.102	0.296	0.001	-0.559	0.120	0.216	0.050	1.069	0.037	1.250	0.117
71	0.370	0.105	0.595	0.001	-0.863	0.085	0.293	0.045	1.085	0.039	1.085	0.118
72	0.143	0.107	0.407	0.001	-0.702	0.104	0.314	0.046	1.037	0.039	0.713	0.064
73	0.521	0.104	0.474	0.000	-0.535	0.107	0.210	0.051	1.009	0.039	0.917	0.120
74	0.069	0.107	0.481	0.001	-0.577	0.094	0.253	0.043	1.057	0.043	0.520	0.060
75	-0.378	0.106	0.646	0.001	-0.769	0.092	0.310	0.045	1.154	0.042	1.084	0.102
76	0.230	0.107	0.511	0.001	-0.498	0.089	0.241	0.049	1.116	0.042	1.035	0.090
77	0.248	0.109	0.579	0.000	-0.619	0.080	0.191	0.055	1.138	0.052	1.054	0.157
78	0.229	0.111	0.275	0.001	-0.487	0.113	0.368	0.047	1.171	0.051	0.846	0.126
79	0.111	0.106	0.487	0.001	-0.679	0.101	0.348	0.042	1.052	0.043	0.978	0.114
80	0.078	0.102	0.443	0.001	-0.594	0.084	0.245	0.039	1.070	0.044	1.216	0.120
81	0.395	0.098	0.502	0.001	-0.671	0.130	0.271	0.047	1.041	0.039	1.517	0.113
82	0.312	0.107	0.592	0.001	-0.605	0.110	0.165	0.048	1.196	0.053	1.348	0.099
83	0.056	0.109	0.755	0.001	-0.769	0.090	0.186	0.043	1.051	0.048	0.736	0.127
84	0.143	0.105	0.332	0.001	-0.489	0.093	0.379	0.051	1.096	0.040	1.134	0.111
85	-0.365	0.105	0.560	0.001	-0.636	0.104	0.231	0.048	1.085	0.041	1.187	0.180
86	-0.131	0.109	0.425	0.001	-0.499	0.073	0.237	0.046	1.052	0.036	0.748	0.070
87	0.016	0.109	0.566	0.001	-0.707	0.106	0.269	0.044	1.091	0.047	0.836	0.089
88	0.215	0.097	0.508	0.000	-0.442	0.089	0.278	0.049	0.975	0.045	1.201	0.119
89	-0.134	0.108	0.336	0.000	-0.468	0.101	0.245	0.049	1.079	0.047	0.837	0.098
90	0.017	0.106	0.541	0.001	-0.669	0.115	0.304	0.043	1.093	0.034	1.006	0.101

PEST: BIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
91	0.335	0.107	0.567	0.001	-0.637	0.093	0.195	0.034	1.168	0.046	1.143	0.119
92	-0.109	0.105	0.709	0.001	-0.765	0.101	0.315	0.043	1.069	0.039	0.918	0.103
93	0.264	0.109	0.624	0.000	-0.649	0.094	0.227	0.046	1.126	0.036	0.955	0.092
94	-0.088	0.107	0.443	0.001	-0.549	0.096	0.258	0.045	1.059	0.038	0.832	0.089
95	0.209	0.104	0.734	0.001	-0.564	0.104	0.229	0.053	1.056	0.036	0.717	0.088
96	-0.222	0.105	0.342	0.001	-0.628	0.089	0.210	0.050	1.037	0.030	0.925	0.100
97	-0.214	0.113	0.413	0.000	-0.747	0.097	0.285	0.045	1.215	0.051	0.465	0.056
98	-0.058	0.108	0.736	0.001	-0.669	0.104	0.228	0.042	1.110	0.050	0.977	0.092
99	0.127	0.110	0.532	0.001	-0.724	0.096	0.175	0.053	1.163	0.048	0.970	0.093
100	0.537	0.107	0.507	0.001	-0.627	0.100	0.234	0.048	1.127	0.037	0.962	0.097
101	-0.421	0.105	0.821	0.001	-0.811	0.087	0.229	0.037	1.010	0.038	0.771	0.087
102	0.491	0.109	0.550	0.001	-0.791	0.098	0.384	0.054	1.100	0.044	0.786	0.085
103	0.861	0.104	0.566	0.001	-0.627	0.101	0.246	0.042	1.035	0.045	1.081	0.093
104	-0.234	0.109	0.573	0.001	-0.606	0.103	0.281	0.046	1.168	0.041	0.929	0.093
105	0.466	0.113	0.678	0.001	-0.757	0.109	0.159	0.049	1.285	0.046	1.167	0.130
106	0.133	0.105	0.613	0.000	-0.680	0.093	0.297	0.050	1.063	0.043	1.027	0.071
107	0.286	0.108	0.522	0.001	-0.505	0.095	0.227	0.047	1.146	0.046	0.974	0.104
108	0.594	0.109	0.736	0.001	-0.681	0.105	0.370	0.042	1.083	0.042	0.859	0.103
109	0.207	0.110	0.473	0.001	-0.840	0.106	0.275	0.041	1.112	0.047	0.729	0.082
110	-0.250	0.108	0.337	0.001	-0.472	0.079	0.164	0.040	1.098	0.043	0.916	0.096
111	0.205	0.107	0.771	0.001	-0.732	0.085	0.351	0.044	0.968	0.036	0.597	0.080
112	-0.574	0.110	0.425	0.000	-0.602	0.088	0.280	0.047	1.121	0.044	0.927	0.108
113	0.070	0.107	0.530	0.001	-0.631	0.121	0.094	0.048	1.148	0.049	1.007	0.094
114	-0.472	0.106	0.730	0.001	-0.832	0.090	0.256	0.041	1.107	0.050	1.046	0.116
115	0.667	0.107	0.433	0.001	-0.594	0.092	0.242	0.046	1.084	0.037	1.128	0.153
116	0.460	0.109	0.363	0.000	-0.698	0.083	0.180	0.047	1.108	0.044	0.652	0.085
117	0.631	0.104	0.883	0.001	-0.859	0.119	0.260	0.037	1.123	0.050	1.222	0.148
118	0.320	0.103	0.412	0.001	-0.665	0.105	0.140	0.042	1.024	0.046	1.069	0.086
119	0.387	0.106	0.678	0.001	-0.597	0.098	0.338	0.055	1.136	0.041	1.054	0.123
120	-0.179	0.105	0.589	0.001	-0.540	0.097	0.250	0.050	1.098	0.041	1.074	0.088

PEST: BIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
121	-0.041	0.097	0.487	0.001	-0.419	0.108	0.334	0.041	1.011	0.041	1.517	0.121
122	-0.043	0.108	0.596	0.001	-0.692	0.110	0.265	0.050	1.102	0.051	0.923	0.097
123	0.528	0.105	0.452	0.001	-0.369	0.096	0.242	0.048	1.168	0.046	1.317	0.114
124	-0.137	0.109	0.196	0.000	-0.684	0.096	0.253	0.056	1.160	0.046	0.942	0.117
125	0.401	0.105	0.543	0.001	-0.573	0.084	0.364	0.045	1.107	0.047	0.885	0.099
126	0.575	0.109	0.508	0.001	-0.507	0.100	0.300	0.045	1.227	0.049	1.081	0.085
127	0.003	0.107	0.416	0.000	-0.833	0.094	0.204	0.042	1.131	0.050	0.997	0.094
128	0.258	0.098	0.352	0.000	-0.741	0.098	0.295	0.046	1.020	0.045	1.271	0.127
129	0.123	0.102	0.558	0.001	-0.688	0.120	0.225	0.047	1.129	0.040	1.531	0.173
130	0.452	0.112	0.360	0.001	-0.578	0.103	0.260	0.056	1.224	0.057	1.175	0.091
131	0.242	0.105	0.667	0.001	-0.594	0.096	0.268	0.048	1.086	0.042	0.919	0.129
132	0.007	0.106	0.003	0.000	-0.307	0.093	0.270	0.043	1.098	0.037	1.039	0.082
133	0.299	0.105	0.487	0.001	-0.613	0.092	0.275	0.054	1.145	0.053	1.057	0.119
134	0.377	0.106	0.565	0.001	-0.619	0.119	0.248	0.041	1.076	0.046	0.896	0.102
135	-0.086	0.110	0.283	0.001	-0.304	0.083	0.257	0.045	1.193	0.045	1.049	0.116
136	-0.331	0.104	0.525	0.001	-0.552	0.090	0.207	0.047	1.076	0.041	1.006	0.119
137	0.377	0.108	0.351	0.000	-0.516	0.086	0.218	0.050	1.121	0.040	0.898	0.098
138	0.375	0.104	0.312	0.001	-0.572	0.111	0.293	0.044	1.100	0.048	1.093	0.093
139	0.240	0.104	0.365	0.000	-0.329	0.095	0.243	0.050	1.110	0.050	1.113	0.112
140	0.288	0.110	0.332	0.001	-0.456	0.102	0.244	0.049	1.149	0.043	0.594	0.056
141	-0.218	0.106	0.236	0.001	-0.348	0.103	0.095	0.041	1.069	0.046	0.764	0.091
142	-0.229	0.104	0.563	0.001	-0.673	0.095	0.316	0.042	1.068	0.055	0.989	0.091
143	0.118	0.108	-9.172	0.000	-0.414	0.095	0.243	0.053	1.115	0.037	0.871	0.097
144	0.040	0.105	0.581	0.000	-0.708	0.081	0.232	0.048	1.103	0.042	1.129	0.108
145	-0.057	0.105	0.618	0.001	-0.520	0.104	0.278	0.043	1.037	0.047	0.931	0.105
146	0.214	0.106	0.671	0.001	-0.602	0.097	0.212	0.045	1.126	0.042	1.135	0.111
147	0.125	0.108	0.545	0.001	-0.648	0.094	0.340	0.050	1.140	0.042	0.986	0.152
148	0.475	0.112	0.424	0.001	-0.722	0.105	0.206	0.040	1.162	0.042	0.757	0.071
149	0.613	0.112	0.432	0.001	-0.688	0.115	0.217	0.042	1.207	0.049	0.725	0.071
150	0.731	0.104	0.781	0.001	-0.697	0.098	0.159	0.043	1.031	0.047	0.998	0.089

PEST: BIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
151	0.256	0.105	0.574	0.001	-0.704	0.112	0.291	0.036	1.085	0.043	0.807	0.088
152	0.093	0.109	0.520	0.001	-0.631	0.093	0.354	0.050	1.120	0.043	0.645	0.075
153	0.734	0.106	0.498	0.001	-0.761	0.099	0.272	0.039	1.066	0.041	1.133	0.103
154	0.173	0.105	0.522	0.001	-0.622	0.115	0.260	0.052	1.053	0.045	0.911	0.087
155	0.124	0.105	0.409	0.001	-0.494	0.093	0.224	0.042	1.045	0.047	0.696	0.070
156	0.318	0.110	0.408	0.001	-0.666	0.103	0.276	0.045	1.078	0.049	0.668	0.081
157	0.513	0.105	0.411	0.001	-0.723	0.102	0.334	0.046	1.029	0.041	0.882	0.106
158	0.063	0.108	0.349	0.001	-0.457	0.109	0.239	0.049	1.139	0.040	1.068	0.095
159	0.301	0.102	0.179	0.000	-0.486	0.097	0.144	0.045	1.059	0.042	1.409	0.171
160	0.042	0.111	0.467	0.000	-0.689	0.098	0.267	0.044	1.131	0.045	0.668	0.074
161	0.282	0.109	0.591	0.001	-0.723	0.083	0.255	0.042	1.232	0.046	0.876	0.082
162	0.066	0.105	0.435	0.000	-0.497	0.095	0.315	0.052	1.093	0.048	1.142	0.101
163	0.294	0.104	0.901	0.001	-0.993	0.104	0.275	0.036	1.047	0.046	1.080	0.114
164	0.482	0.106	0.653	0.000	-0.842	0.092	0.335	0.046	1.127	0.046	1.337	0.141
165	0.451	0.111	0.627	0.000	-0.596	0.095	0.191	0.047	1.151	0.051	0.773	0.087
166	0.445	0.103	0.581	0.001	-0.577	0.115	0.287	0.047	1.031	0.047	0.954	0.093
167	0.150	0.105	0.191	0.000	-0.446	0.090	0.283	0.041	1.065	0.044	1.014	0.095
168	0.065	0.102	0.569	0.001	-0.638	0.098	0.313	0.042	1.079	0.043	1.164	0.112
169	0.270	0.100	0.524	0.000	-0.455	0.085	0.295	0.045	1.042	0.044	1.198	0.081
170	0.437	0.105	0.431	0.001	-0.567	0.100	0.254	0.049	1.060	0.041	0.951	0.137
171	0.206	0.109	0.395	0.001	-0.555	0.092	0.301	0.043	1.162	0.057	0.956	0.139
172	-0.450	0.102	0.524	0.001	-0.789	0.116	0.254	0.043	1.042	0.035	1.192	0.195
173	0.345	0.108	0.460	0.001	-0.560	0.104	0.210	0.051	1.142	0.038	0.938	0.111
174	0.049	0.103	0.030	0.000	-0.459	0.101	0.265	0.055	1.104	0.041	1.346	0.130
175	-0.123	0.110	0.636	0.001	-0.892	0.110	0.264	0.041	1.172	0.050	0.820	0.074
176	0.068	0.105	0.734	0.001	-0.929	0.088	0.266	0.048	1.101	0.057	1.107	0.126
177	0.040	0.105	0.380	0.000	-0.665	0.099	0.254	0.051	1.049	0.044	1.073	0.113
178	0.076	0.105	0.650	0.001	-0.462	0.102	0.207	0.046	1.046	0.046	0.974	0.102
179	-0.351	0.104	0.745	0.001	-0.695	0.109	0.298	0.049	1.114	0.044	1.254	0.170
180	-0.024	0.110	0.602	0.000	-0.803	0.091	0.275	0.052	1.168	0.038	0.901	0.089

PEST: BIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
181	0.434	0.102	0.391	0.000	-0.466	0.101	0.255	0.041	0.976	0.042	0.970	0.099
182	0.051	0.106	0.552	0.001	-0.670	0.118	0.229	0.048	1.099	0.042	1.082	0.101
183	0.028	0.106	0.489	0.001	-0.614	0.099	0.284	0.049	1.125	0.055	1.058	0.099
184	0.739	0.104	0.353	0.000	-0.370	0.107	0.214	0.048	1.173	0.040	1.363	0.140
185	0.354	0.108	0.601	0.001	-0.799	0.100	0.313	0.051	1.113	0.044	0.825	0.099
186	0.380	0.100	0.495	0.001	-0.684	0.106	0.351	0.053	0.996	0.045	1.016	0.101
187	0.304	0.110	-2.261	0.000	-0.209	0.100	0.260	0.051	1.186	0.042	0.808	0.087
188	0.190	0.109	0.677	0.001	-0.785	0.087	0.253	0.039	1.075	0.044	0.853	0.076
189	0.101	0.106	-4.001	0.000	-0.293	0.114	0.308	0.049	1.068	0.046	0.814	0.091
190	0.358	0.108	0.637	0.001	-0.588	0.105	0.207	0.053	1.056	0.044	0.709	0.082
191	0.006	0.104	0.642	0.000	-0.564	0.108	0.271	0.043	1.134	0.052	1.143	0.098
192	0.107	0.102	0.615	0.001	-0.653	0.088	0.268	0.040	1.013	0.035	0.952	0.115
193	0.072	0.112	0.614	0.001	-0.478	0.104	0.220	0.045	1.206	0.046	0.755	0.116
194	0.213	0.106	0.749	0.001	-0.643	0.105	0.161	0.051	1.028	0.042	0.813	0.101
195	0.211	0.102	0.684	0.001	-0.733	0.094	0.346	0.046	1.031	0.048	1.074	0.125
196	0.633	0.106	0.588	0.001	-0.528	0.102	0.234	0.047	1.049	0.041	0.902	0.111
197	0.102	0.109	0.496	0.001	-0.536	0.094	0.141	0.043	1.050	0.040	0.676	0.069
198	-0.121	0.111	0.318	0.001	-0.567	0.113	0.215	0.051	1.173	0.052	0.832	0.105
199	0.226	0.103	0.733	0.001	-0.671	0.108	0.248	0.045	1.028	0.046	1.047	0.123
200	-0.351	0.110	0.528	0.001	-0.862	0.102	0.272	0.050	1.128	0.043	0.596	0.061

PEST: ABIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	-0.109	0.111	0.632	0.001	-0.653	0.110	0.330	0.050	1.205	0.047	0.913	0.102
2	0.222	0.102	0.395	0.001	-0.422	0.100	0.223	0.050	1.012	0.050	0.987	0.080
3	-0.016	0.106	0.697	0.000	-0.771	0.090	0.176	0.046	1.114	0.035	0.946	0.090
4	-0.497	0.097	0.384	0.001	-0.566	0.106	0.177	0.048	1.066	0.048	1.390	0.113
5	-0.231	0.110	0.614	0.001	-0.707	0.112	0.315	0.047	1.142	0.052	0.895	0.109
6	0.100	0.110	0.380	0.001	-0.569	0.092	0.263	0.045	1.164	0.049	0.888	0.074
7	0.038	0.105	0.626	0.000	-0.552	0.080	0.250	0.048	1.105	0.041	1.124	0.122
8	-0.161	0.109	0.215	0.001	-0.441	0.095	0.137	0.043	1.141	0.047	0.992	0.130
9	0.107	0.104	0.650	0.001	-0.695	0.094	0.331	0.051	1.035	0.038	1.108	0.116
10	-0.080	0.108	0.331	0.001	-0.453	0.078	0.229	0.045	1.115	0.045	0.896	0.098
11	0.364	0.101	0.456	0.000	-0.636	0.079	0.133	0.041	1.049	0.040	1.279	0.134
12	0.546	0.104	0.550	0.001	-0.390	0.101	0.151	0.043	1.132	0.051	1.260	0.109
13	0.304	0.103	0.291	0.001	-0.598	0.096	0.196	0.040	1.046	0.049	1.049	0.105
14	0.168	0.112	0.374	0.001	-0.613	0.112	0.313	0.042	1.199	0.040	0.820	0.082
15	0.310	0.110	0.556	0.001	-0.638	0.104	0.282	0.054	1.104	0.046	0.736	0.094
16	-0.258	0.105	0.471	0.001	-0.814	0.116	0.249	0.048	1.102	0.046	1.012	0.092
17	0.158	0.101	0.389	0.001	-0.785	0.105	0.237	0.051	1.068	0.044	1.384	0.124
18	0.133	0.106	0.399	0.001	-0.731	0.095	0.360	0.046	1.075	0.044	0.999	0.129
19	-0.255	0.109	0.468	0.001	-0.641	0.122	0.252	0.043	1.130	0.051	1.003	0.110
20	0.587	0.106	0.017	0.000	-0.406	0.093	0.259	0.046	1.062	0.046	0.895	0.088
21	0.314	0.109	0.365	0.001	-0.675	0.071	0.313	0.052	1.126	0.056	0.802	0.075
22	0.039	0.100	0.612	0.001	-0.690	0.097	0.250	0.045	1.068	0.045	1.376	0.104
23	0.377	0.109	0.520	0.000	-0.603	0.100	0.378	0.049	1.169	0.045	1.090	0.079
24	0.241	0.107	0.453	0.001	-0.589	0.104	0.243	0.042	1.103	0.048	1.090	0.155
25	-0.279	0.103	0.491	0.001	-0.420	0.105	0.291	0.032	1.040	0.049	1.084	0.085
26	0.289	0.108	0.255	0.000	-0.394	0.106	-0.005	0.044	1.086	0.047	0.937	0.085
27	-0.283	0.100	0.387	0.001	-0.463	0.098	0.225	0.036	1.058	0.040	1.357	0.115
28	-0.056	0.107	0.251	0.000	-0.405	0.094	0.206	0.040	1.114	0.045	0.815	0.083
29	0.265	0.109	0.571	0.001	-0.564	0.109	0.304	0.040	1.173	0.052	0.992	0.083
30	-0.119	0.109	0.827	0.001	-0.641	0.101	0.291	0.050	1.106	0.052	0.713	0.089

PEST: ABIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
31	-0.059	0.106	0.380	0.000	-0.534	0.094	0.169	0.044	1.050	0.046	0.828	0.086
32	0.634	0.108	0.531	0.000	-0.610	0.111	0.174	0.045	1.112	0.049	0.671	0.080
33	0.527	0.110	0.495	0.001	-0.758	0.083	0.230	0.040	1.146	0.047	0.682	0.064
34	0.557	0.105	0.462	0.001	-0.631	0.095	0.239	0.048	1.115	0.045	1.153	0.106
35	-0.244	0.103	0.301	0.000	-0.392	0.093	0.318	0.042	1.043	0.044	0.879	0.073
36	0.079	0.105	0.176	0.001	-0.423	0.103	0.320	0.040	1.111	0.044	1.253	0.135
37	-0.071	0.110	0.675	0.001	-0.679	0.100	0.228	0.046	1.110	0.043	0.805	0.113
38	0.116	0.106	0.667	0.001	-0.633	0.095	0.239	0.044	1.100	0.037	1.229	0.110
39	-0.085	0.102	0.372	0.001	-0.553	0.101	0.347	0.047	1.051	0.037	1.202	0.142
40	0.058	0.107	0.493	0.000	-0.580	0.099	0.177	0.049	1.097	0.048	0.916	0.123
41	0.471	0.104	0.619	0.001	-0.538	0.123	0.233	0.050	1.135	0.048	1.148	0.121
42	0.000	0.104	0.152	0.000	-0.573	0.102	0.133	0.050	1.009	0.045	0.946	0.095
43	-0.029	0.102	0.389	0.001	-0.718	0.102	0.221	0.044	1.055	0.046	1.167	0.129
44	0.011	0.106	0.473	0.000	-0.485	0.117	0.155	0.044	1.191	0.051	1.495	0.156
45	0.385	0.105	0.148	0.000	-0.320	0.095	0.053	0.045	1.000	0.040	0.610	0.106
46	0.524	0.106	0.687	0.001	-0.742	0.102	0.251	0.040	1.126	0.050	1.233	0.108
47	0.014	0.111	0.196	0.000	-0.327	0.082	0.230	0.041	1.160	0.050	1.034	0.114
48	0.333	0.106	0.568	0.000	-0.706	0.110	0.218	0.047	1.045	0.038	0.701	0.087
49	0.582	0.106	0.674	0.001	-0.682	0.111	0.205	0.050	1.077	0.041	1.032	0.111
50	0.414	0.106	0.542	0.001	-0.816	0.101	0.293	0.052	1.119	0.042	0.854	0.116
51	-0.209	0.110	0.431	0.001	-0.647	0.104	0.255	0.048	1.144	0.046	0.871	0.085
52	0.298	0.107	0.654	0.001	-0.500	0.099	0.179	0.039	1.074	0.045	0.925	0.093
53	0.095	0.105	0.730	0.001	-0.838	0.098	0.354	0.060	1.083	0.038	1.099	0.120
54	-0.291	0.108	0.574	0.001	-0.518	0.101	0.237	0.048	1.026	0.035	0.744	0.092
55	-0.211	0.106	0.341	0.001	-0.407	0.126	0.233	0.043	1.127	0.044	0.992	0.109
56	-0.363	0.105	0.201	0.000	-0.331	0.085	0.171	0.046	1.078	0.046	0.938	0.086
57	0.297	0.107	0.602	0.000	-0.647	0.089	0.253	0.052	1.143	0.054	0.987	0.101
58	-0.101	0.108	0.541	0.000	-0.561	0.075	0.235	0.045	1.058	0.048	0.826	0.083
59	0.127	0.103	0.230	0.000	-0.496	0.089	0.193	0.050	1.053	0.037	1.115	0.093
60	-0.414	0.105	0.008	0.000	-0.474	0.105	0.287	0.054	1.056	0.045	0.952	0.089

PEST: ABIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
61	0.364	0.110	0.513	0.001	-0.665	0.094	0.159	0.047	1.120	0.049	0.587	0.073
62	0.002	0.106	0.623	0.001	-0.692	0.099	0.337	0.046	1.119	0.042	1.087	0.104
63	-0.521	0.106	0.734	0.001	-0.681	0.099	0.282	0.051	1.127	0.047	0.949	0.081
64	-0.304	0.105	0.438	0.000	-0.527	0.095	0.180	0.048	1.049	0.039	0.843	0.076
65	0.062	0.107	0.600	0.001	-0.704	0.093	0.222	0.044	1.065	0.046	0.674	0.085
66	0.136	0.107	0.545	0.001	-0.680	0.111	0.131	0.045	1.110	0.048	1.093	0.098
67	0.028	0.108	0.482	0.001	-0.645	0.100	0.222	0.045	1.166	0.052	1.083	0.112
68	0.445	0.110	0.448	0.001	-0.508	0.097	0.275	0.045	1.167	0.042	1.117	0.112
69	0.302	0.104	0.489	0.001	-0.582	0.102	0.245	0.041	1.048	0.037	0.993	0.088
70	-0.136	0.102	0.296	0.001	-0.559	0.120	0.216	0.050	1.069	0.037	1.250	0.117
71	0.370	0.105	0.595	0.001	-0.863	0.085	0.293	0.045	1.085	0.039	1.085	0.118
72	0.143	0.107	0.407	0.001	-0.702	0.104	0.314	0.046	1.037	0.039	0.713	0.064
73	0.521	0.104	0.474	0.000	-0.535	0.107	0.210	0.051	1.009	0.039	0.917	0.120
74	0.069	0.107	0.481	0.001	-0.577	0.094	0.253	0.043	1.057	0.043	0.520	0.060
75	-0.378	0.106	0.646	0.001	-0.769	0.092	0.310	0.045	1.154	0.042	1.084	0.102
76	0.205	0.105	0.193	0.000	-0.340	0.086	0.241	0.049	1.115	0.042	1.067	0.091
77	0.248	0.109	0.579	0.000	-0.619	0.080	0.191	0.055	1.138	0.052	1.054	0.157
78	0.229	0.111	0.275	0.001	-0.487	0.113	0.368	0.047	1.171	0.051	0.846	0.126
79	0.111	0.106	0.487	0.001	-0.679	0.101	0.348	0.042	1.052	0.043	0.978	0.114
80	0.078	0.102	0.443	0.001	-0.594	0.084	0.245	0.039	1.070	0.044	1.216	0.120
81	0.395	0.098	0.502	0.001	-0.671	0.130	0.271	0.047	1.041	0.039	1.517	0.113
82	0.312	0.107	0.592	0.001	-0.605	0.110	0.165	0.048	1.196	0.053	1.348	0.099
83	0.056	0.109	0.755	0.001	-0.769	0.090	0.186	0.043	1.051	0.048	0.736	0.127
84	0.143	0.105	0.332	0.001	-0.489	0.093	0.379	0.051	1.096	0.040	1.134	0.111
85	-0.365	0.105	0.560	0.001	-0.636	0.104	0.231	0.048	1.085	0.041	1.187	0.180
86	-0.131	0.109	0.425	0.001	-0.499	0.073	0.237	0.046	1.052	0.036	0.748	0.070
87	0.016	0.109	0.566	0.001	-0.707	0.106	0.269	0.044	1.091	0.047	0.836	0.089
88	0.102	0.099	0.067	0.000	-0.225	0.083	0.161	0.046	0.970	0.045	1.164	0.117
89	-0.134	0.108	0.336	0.000	-0.468	0.101	0.245	0.049	1.079	0.047	0.837	0.098
90	-0.001	0.108	0.235	0.000	-0.515	0.112	0.303	0.043	1.093	0.034	1.001	0.101

PEST: ABIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
91	0.335	0.107	0.567	0.001	-0.637	0.093	0.195	0.034	1.168	0.046	1.143	0.119
92	-0.109	0.105	0.709	0.001	-0.765	0.101	0.315	0.043	1.069	0.039	0.918	0.103
93	0.298	0.110	0.273	0.000	-0.475	0.091	0.235	0.046	1.128	0.036	0.959	0.092
94	-0.088	0.107	0.443	0.001	-0.549	0.096	0.258	0.045	1.059	0.038	0.832	0.089
95	0.209	0.104	0.734	0.001	-0.564	0.104	0.229	0.053	1.056	0.036	0.717	0.088
96	-0.222	0.105	0.342	0.001	-0.628	0.089	0.210	0.050	1.037	0.030	0.925	0.100
97	-0.214	0.113	0.413	0.000	-0.747	0.097	0.285	0.045	1.215	0.051	0.465	0.056
98	-0.058	0.108	0.736	0.001	-0.669	0.104	0.228	0.042	1.110	0.050	0.977	0.092
99	0.127	0.110	0.532	0.001	-0.724	0.096	0.175	0.053	1.163	0.048	0.970	0.093
100	0.537	0.107	0.507	0.001	-0.627	0.100	0.234	0.048	1.127	0.037	0.962	0.097
101	-0.421	0.105	0.821	0.001	-0.811	0.087	0.229	0.037	1.010	0.038	0.771	0.087
102	0.491	0.109	0.550	0.001	-0.791	0.098	0.384	0.054	1.100	0.044	0.786	0.085
103	0.861	0.104	0.566	0.001	-0.627	0.101	0.246	0.042	1.035	0.045	1.081	0.093
104	-0.234	0.109	0.573	0.001	-0.606	0.103	0.281	0.046	1.168	0.041	0.929	0.093
105	0.452	0.112	0.355	0.000	-0.601	0.107	0.160	0.049	1.286	0.046	1.170	0.130
106	0.133	0.105	0.613	0.000	-0.680	0.093	0.297	0.050	1.063	0.043	1.027	0.071
107	0.240	0.108	0.180	0.000	-0.339	0.092	0.229	0.047	1.147	0.046	0.984	0.105
108	0.594	0.109	0.736	0.001	-0.681	0.105	0.370	0.042	1.083	0.042	0.859	0.103
109	0.161	0.110	0.146	0.000	-0.684	0.102	0.275	0.041	1.114	0.047	0.721	0.080
110	-0.250	0.108	0.337	0.001	-0.472	0.079	0.164	0.040	1.098	0.043	0.916	0.096
111	0.205	0.107	0.771	0.001	-0.732	0.085	0.351	0.044	0.968	0.036	0.597	0.080
112	-0.574	0.110	0.425	0.000	-0.602	0.088	0.280	0.047	1.121	0.044	0.927	0.108
113	0.070	0.107	0.530	0.001	-0.631	0.121	0.094	0.048	1.148	0.049	1.007	0.094
114	-0.472	0.106	0.730	0.001	-0.832	0.090	0.256	0.041	1.107	0.050	1.046	0.116
115	0.631	0.105	0.104	0.000	-0.420	0.090	0.240	0.046	1.084	0.037	1.150	0.155
116	0.460	0.109	0.363	0.000	-0.698	0.083	0.180	0.047	1.108	0.044	0.652	0.085
117	0.631	0.104	0.883	0.001	-0.859	0.119	0.260	0.037	1.123	0.050	1.222	0.148
118	0.212	0.103	-9.430	0.000	-0.411	0.099	0.139	0.042	1.026	0.046	1.100	0.087
119	0.387	0.106	0.678	0.001	-0.597	0.098	0.338	0.055	1.136	0.041	1.054	0.123
120	-0.179	0.105	0.589	0.001	-0.540	0.097	0.250	0.050	1.098	0.041	1.074	0.088

PEST: ABIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
121	-0.041	0.097	0.487	0.001	-0.419	0.108	0.334	0.041	1.011	0.041	1.517	0.121
122	-0.043	0.108	0.596	0.001	-0.692	0.110	0.265	0.050	1.102	0.051	0.923	0.097
123	0.528	0.105	0.452	0.001	-0.369	0.096	0.242	0.048	1.168	0.046	1.317	0.114
124	-0.273	0.110	0.125	0.000	-0.649	0.095	0.121	0.054	1.159	0.048	0.929	0.116
125	0.254	0.108	0.484	0.001	-0.544	0.084	0.255	0.042	1.105	0.047	0.862	0.097
126	0.575	0.109	0.508	0.001	-0.507	0.100	0.300	0.045	1.227	0.049	1.081	0.085
127	-0.022	0.108	0.123	0.000	-0.689	0.092	0.204	0.042	1.131	0.050	0.997	0.094
128	0.258	0.098	0.352	0.000	-0.741	0.098	0.295	0.046	1.020	0.045	1.271	0.127
129	0.123	0.102	0.558	0.001	-0.688	0.120	0.225	0.047	1.129	0.040	1.531	0.173
130	0.457	0.108	0.326	0.001	-0.561	0.103	0.123	0.052	1.217	0.057	1.219	0.093
131	0.242	0.105	0.667	0.001	-0.594	0.096	0.268	0.048	1.086	0.042	0.919	0.129
132	0.007	0.106	0.003	0.000	-0.307	0.093	0.270	0.043	1.098	0.037	1.039	0.082
133	0.299	0.105	0.487	0.001	-0.613	0.092	0.275	0.054	1.145	0.053	1.057	0.119
134	0.377	0.106	0.565	0.001	-0.619	0.119	0.248	0.041	1.076	0.046	0.896	0.102
135	-0.086	0.110	0.283	0.001	-0.304	0.083	0.257	0.045	1.193	0.045	1.049	0.116
136	-0.331	0.104	0.525	0.001	-0.552	0.090	0.207	0.047	1.076	0.041	1.006	0.119
137	0.377	0.108	0.351	0.000	-0.516	0.086	0.218	0.050	1.121	0.040	0.898	0.098
138	0.375	0.104	-2.187	0.000	-0.404	0.109	0.292	0.044	1.101	0.048	1.098	0.094
139	0.240	0.104	0.365	0.000	-0.329	0.095	0.243	0.050	1.110	0.050	1.113	0.112
140	0.288	0.110	0.332	0.001	-0.456	0.102	0.244	0.049	1.149	0.043	0.594	0.056
141	-0.218	0.106	0.236	0.001	-0.348	0.103	0.095	0.041	1.069	0.046	0.764	0.091
142	-0.229	0.104	0.563	0.001	-0.673	0.095	0.316	0.042	1.068	0.055	0.989	0.091
143	0.118	0.108	-9.172	0.000	-0.414	0.095	0.243	0.053	1.115	0.037	0.871	0.097
144	0.040	0.105	0.581	0.000	-0.708	0.081	0.232	0.048	1.103	0.042	1.129	0.108
145	-0.057	0.105	0.618	0.001	-0.520	0.104	0.278	0.043	1.037	0.047	0.931	0.105
146	0.214	0.106	0.671	0.001	-0.602	0.097	0.212	0.045	1.126	0.042	1.135	0.111
147	0.125	0.108	0.545	0.001	-0.648	0.094	0.340	0.050	1.140	0.042	0.986	0.152
148	0.475	0.112	0.424	0.001	-0.722	0.105	0.206	0.040	1.162	0.042	0.757	0.071
149	0.613	0.112	0.432	0.001	-0.688	0.115	0.217	0.042	1.207	0.049	0.725	0.071
150	0.731	0.104	0.781	0.001	-0.697	0.098	0.159	0.043	1.031	0.047	0.998	0.089

PEST: ABIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
151	0.256	0.105	0.574	0.001	-0.704	0.112	0.291	0.036	1.085	0.043	0.807	0.088
152	0.093	0.109	0.520	0.001	-0.631	0.093	0.354	0.050	1.120	0.043	0.645	0.075
153	0.687	0.106	0.096	0.000	-0.557	0.094	0.272	0.039	1.070	0.042	1.122	0.103
154	0.173	0.105	0.522	0.001	-0.622	0.115	0.260	0.052	1.053	0.045	0.911	0.087
155	0.124	0.105	0.409	0.001	-0.494	0.093	0.224	0.042	1.045	0.047	0.696	0.070
156	0.318	0.110	0.408	0.001	-0.666	0.103	0.276	0.045	1.078	0.049	0.668	0.081
157	0.513	0.105	0.411	0.001	-0.723	0.102	0.334	0.046	1.029	0.041	0.882	0.106
158	0.063	0.108	0.349	0.001	-0.457	0.109	0.239	0.049	1.139	0.040	1.068	0.095
159	0.301	0.102	0.179	0.000	-0.486	0.097	0.144	0.045	1.059	0.042	1.409	0.171
160	0.042	0.111	0.467	0.000	-0.689	0.098	0.267	0.044	1.131	0.045	0.668	0.074
161	0.282	0.109	0.591	0.001	-0.723	0.083	0.255	0.042	1.232	0.046	0.876	0.082
162	0.066	0.105	0.435	0.000	-0.497	0.095	0.315	0.052	1.093	0.048	1.142	0.101
163	0.294	0.104	0.901	0.001	-0.993	0.104	0.275	0.036	1.047	0.046	1.080	0.114
164	0.482	0.106	0.653	0.000	-0.842	0.092	0.335	0.046	1.127	0.046	1.337	0.141
165	0.451	0.111	0.627	0.000	-0.596	0.095	0.191	0.047	1.151	0.051	0.773	0.087
166	0.445	0.103	0.581	0.001	-0.577	0.115	0.287	0.047	1.031	0.047	0.954	0.093
167	0.150	0.105	0.191	0.000	-0.446	0.090	0.283	0.041	1.065	0.044	1.014	0.095
168	0.065	0.102	0.569	0.001	-0.638	0.098	0.313	0.042	1.079	0.043	1.164	0.112
169	0.306	0.102	0.261	0.000	-0.325	0.082	0.295	0.045	1.042	0.045	1.187	0.081
170	0.437	0.105	0.431	0.001	-0.567	0.100	0.254	0.049	1.060	0.041	0.951	0.137
171	0.206	0.109	0.395	0.001	-0.555	0.092	0.301	0.043	1.162	0.057	0.956	0.139
172	-0.450	0.102	0.524	0.001	-0.789	0.116	0.254	0.043	1.042	0.035	1.192	0.195
173	0.345	0.108	0.460	0.001	-0.560	0.104	0.210	0.051	1.142	0.038	0.938	0.111
174	0.049	0.103	0.030	0.000	-0.459	0.101	0.265	0.055	1.104	0.041	1.346	0.130
175	-0.123	0.110	0.636	0.001	-0.892	0.110	0.264	0.041	1.172	0.050	0.820	0.074
176	0.068	0.105	0.734	0.001	-0.929	0.088	0.266	0.048	1.101	0.057	1.107	0.126
177	0.040	0.105	0.380	0.000	-0.665	0.099	0.254	0.051	1.049	0.044	1.073	0.113
178	0.076	0.105	0.650	0.001	-0.462	0.102	0.207	0.046	1.046	0.046	0.974	0.102
179	-0.351	0.104	0.745	0.001	-0.695	0.109	0.298	0.049	1.114	0.044	1.254	0.170
180	-0.024	0.110	0.602	0.000	-0.803	0.091	0.275	0.052	1.168	0.038	0.901	0.089

PEST: ABIC												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
181	0.434	0.102	0.391	0.000	-0.466	0.101	0.255	0.041	0.976	0.042	0.970	0.099
182	0.051	0.106	0.552	0.001	-0.670	0.118	0.229	0.048	1.099	0.042	1.082	0.101
183	0.028	0.106	0.489	0.001	-0.614	0.099	0.284	0.049	1.125	0.055	1.058	0.099
184	0.739	0.104	0.353	0.000	-0.370	0.107	0.214	0.048	1.173	0.040	1.363	0.140
185	0.326	0.109	0.241	0.000	-0.627	0.097	0.314	0.051	1.114	0.044	0.810	0.098
186	0.380	0.100	0.495	0.001	-0.684	0.106	0.351	0.053	0.996	0.045	1.016	0.101
187	0.304	0.110	-2.261	0.000	-0.209	0.100	0.260	0.051	1.186	0.042	0.808	0.087
188	0.190	0.109	0.677	0.001	-0.785	0.087	0.253	0.039	1.075	0.044	0.853	0.076
189	0.101	0.106	-4.001	0.000	-0.293	0.114	0.308	0.049	1.068	0.046	0.814	0.091
190	0.358	0.108	0.637	0.001	-0.588	0.105	0.207	0.053	1.056	0.044	0.709	0.082
191	-0.018	0.106	0.286	0.000	-0.384	0.105	0.269	0.043	1.135	0.052	1.101	0.096
192	0.107	0.102	0.615	0.001	-0.653	0.088	0.268	0.040	1.013	0.035	0.952	0.115
193	-0.007	0.109	0.575	0.001	-0.459	0.103	0.101	0.041	1.203	0.047	0.767	0.117
194	0.213	0.106	0.749	0.001	-0.643	0.105	0.161	0.051	1.028	0.042	0.813	0.101
195	0.211	0.102	0.684	0.001	-0.733	0.094	0.346	0.046	1.031	0.048	1.074	0.125
196	0.469	0.107	0.250	0.000	-0.354	0.100	0.098	0.043	1.046	0.041	0.910	0.112
197	0.102	0.109	0.496	0.001	-0.536	0.094	0.141	0.043	1.050	0.040	0.676	0.069
198	-0.121	0.111	0.318	0.001	-0.567	0.113	0.215	0.051	1.173	0.052	0.832	0.105
199	0.226	0.103	0.733	0.001	-0.671	0.108	0.248	0.045	1.028	0.046	1.047	0.123
200	-0.325	0.111	0.158	0.000	-0.678	0.099	0.272	0.050	1.129	0.043	0.603	0.061

PEST: 2 of 3

R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	-0.109	0.111	0.632	0.001	-0.653	0.110	0.300	0.050	1.205	0.047	0.913	0.102
2	0.222	0.102	0.395	0.001	-0.422	0.100	0.200	0.050	1.012	0.050	0.987	0.080
3	-0.016	0.106	0.697	0.000	-0.771	0.090	0.200	0.046	1.114	0.035	0.946	0.090
4	-0.497	0.097	0.384	0.001	-0.566	0.106	0.200	0.048	1.066	0.048	1.390	0.113
5	-0.231	0.110	0.614	0.001	-0.707	0.112	0.300	0.047	1.142	0.052	0.895	0.109
6	0.100	0.110	0.380	0.001	-0.569	0.092	0.300	0.045	1.164	0.049	0.888	0.074
7	0.038	0.105	0.626	0.000	-0.552	0.080	0.200	0.048	1.105	0.041	1.124	0.122
8	-0.161	0.109	0.215	0.001	-0.441	0.095	0.100	0.043	1.141	0.047	0.992	0.130
9	0.107	0.104	0.650	0.001	-0.695	0.094	0.300	0.051	1.035	0.038	1.108	0.116
10	-0.080	0.108	0.331	0.001	-0.453	0.078	0.200	0.045	1.115	0.045	0.896	0.098
11	0.364	0.101	0.456	0.000	-0.636	0.079	0.100	0.041	1.049	0.040	1.279	0.134
12	0.546	0.104	0.550	0.001	-0.390	0.101	0.200	0.043	1.132	0.051	1.260	0.109
13	0.304	0.103	0.291	0.001	-0.598	0.096	0.200	0.040	1.046	0.049	1.049	0.105
14	0.168	0.112	0.374	0.001	-0.613	0.112	0.300	0.042	1.199	0.040	0.820	0.082
15	0.310	0.110	0.556	0.001	-0.638	0.104	0.300	0.054	1.104	0.046	0.736	0.094
16	-0.258	0.105	0.471	0.001	-0.814	0.116	0.200	0.048	1.102	0.046	1.012	0.092
17	0.158	0.101	0.389	0.001	-0.785	0.105	0.200	0.051	1.068	0.044	1.384	0.124
18	0.133	0.106	0.399	0.001	-0.731	0.095	0.400	0.046	1.075	0.044	0.999	0.129
19	-0.255	0.109	0.468	0.001	-0.641	0.122	0.300	0.043	1.130	0.051	1.003	0.110
20	0.587	0.106	0.017	0.000	-0.406	0.093	0.300	0.046	1.062	0.046	0.895	0.088
21	0.314	0.109	0.365	0.001	-0.675	0.071	0.300	0.052	1.126	0.056	0.802	0.075
22	0.039	0.100	0.612	0.001	-0.690	0.097	0.300	0.045	1.068	0.045	1.376	0.104
23	0.377	0.109	0.520	0.000	-0.603	0.100	0.400	0.049	1.169	0.045	1.090	0.079
24	0.241	0.107	0.453	0.001	-0.589	0.104	0.200	0.042	1.103	0.048	1.090	0.155
25	-0.279	0.103	0.491	0.001	-0.420	0.105	0.300	0.032	1.040	0.049	1.084	0.085
26	0.289	0.108	0.255	0.000	-0.394	0.106	0.000	0.044	1.086	0.047	0.937	0.085
27	-0.283	0.100	0.387	0.001	-0.463	0.098	0.200	0.036	1.058	0.040	1.357	0.115
28	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29	0.265	0.109	0.571	0.001	-0.564	0.109	0.300	0.040	1.173	0.052	0.992	0.083
30	-0.119	0.109	0.827	0.001	-0.641	0.101	0.300	0.050	1.106	0.052	0.713	0.089

PEST: 2 of 3

R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
31	-0.059	0.106	0.380	0.000	-0.534	0.094	0.200	0.044	1.050	0.046	0.828	0.086
32	0.634	0.108	0.531	0.000	-0.610	0.111	0.200	0.045	1.112	0.049	0.671	0.080
33	0.527	0.110	0.495	0.001	-0.758	0.083	0.200	0.040	1.146	0.047	0.682	0.064
34	0.557	0.105	0.462	0.001	-0.631	0.095	0.200	0.048	1.115	0.045	1.153	0.106
35	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
36	0.079	0.105	0.176	0.001	-0.423	0.103	0.300	0.040	1.111	0.044	1.253	0.135
37	-0.071	0.110	0.675	0.001	-0.679	0.100	0.200	0.046	1.110	0.043	0.805	0.113
38	0.116	0.106	0.667	0.001	-0.633	0.095	0.200	0.044	1.100	0.037	1.229	0.110
39	-0.085	0.102	0.372	0.001	-0.553	0.101	0.300	0.047	1.051	0.037	1.202	0.142
40	0.058	0.107	0.493	0.000	-0.580	0.099	0.200	0.049	1.097	0.048	0.916	0.123
41	0.471	0.104	0.619	0.001	-0.538	0.123	0.200	0.050	1.135	0.048	1.148	0.121
42	0.000	0.104	0.152	0.000	-0.573	0.102	0.100	0.050	1.009	0.045	0.946	0.095
43	-0.029	0.102	0.389	0.001	-0.718	0.102	0.200	0.044	1.055	0.046	1.167	0.129
44	0.011	0.106	0.473	0.000	-0.485	0.117	0.200	0.044	1.191	0.051	1.495	0.156
45	0.385	0.105	0.148	0.000	-0.320	0.095	0.100	0.045	1.000	0.040	0.610	0.106
46	0.524	0.106	0.687	0.001	-0.742	0.102	0.300	0.040	1.126	0.050	1.233	0.108
47	0.014	0.111	0.196	0.000	-0.327	0.082	0.200	0.041	1.160	0.050	1.034	0.114
48	0.333	0.106	0.568	0.000	-0.706	0.110	0.200	0.047	1.045	0.038	0.701	0.087
49	0.582	0.106	0.674	0.001	-0.682	0.111	0.200	0.050	1.077	0.041	1.032	0.111
50	0.414	0.106	0.542	0.001	-0.816	0.101	0.300	0.052	1.119	0.042	0.854	0.116
51	-0.209	0.110	0.431	0.001	-0.647	0.104	0.300	0.048	1.144	0.046	0.871	0.085
52	0.298	0.107	0.654	0.001	-0.500	0.099	0.200	0.039	1.074	0.045	0.925	0.093
53	0.095	0.105	0.730	0.001	-0.838	0.098	0.400	0.060	1.083	0.038	1.099	0.120
54	-0.291	0.108	0.574	0.001	-0.518	0.101	0.200	0.048	1.026	0.035	0.744	0.092
55	-0.211	0.106	0.341	0.001	-0.407	0.126	0.200	0.043	1.127	0.044	0.992	0.109
56	-0.363	0.105	0.201	0.000	-0.331	0.085	0.200	0.046	1.078	0.046	0.938	0.086
57	0.297	0.107	0.602	0.000	-0.647	0.089	0.300	0.052	1.143	0.054	0.987	0.101
58	-0.101	0.108	0.541	0.000	-0.561	0.075	0.200	0.045	1.058	0.048	0.826	0.083
59	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
60	-0.414	0.105	0.008	0.000	-0.474	0.105	0.300	0.054	1.056	0.045	0.952	0.089

PEST: 2 of 3

R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
61	0.364	0.110	0.513	0.001	-0.665	0.094	0.200	0.047	1.120	0.049	0.587	0.073
62	0.002	0.106	0.623	0.001	-0.692	0.099	0.300	0.046	1.119	0.042	1.087	0.104
63	-0.521	0.106	0.734	0.001	-0.681	0.099	0.300	0.051	1.127	0.047	0.949	0.081
64	-0.304	0.105	0.438	0.000	-0.527	0.095	0.200	0.048	1.049	0.039	0.843	0.076
65	0.062	0.107	0.600	0.001	-0.704	0.093	0.200	0.044	1.065	0.046	0.674	0.085
66	0.136	0.107	0.545	0.001	-0.680	0.111	0.100	0.045	1.110	0.048	1.093	0.098
67	0.028	0.108	0.482	0.001	-0.645	0.100	0.200	0.045	1.166	0.052	1.083	0.112
68	0.445	0.110	0.448	0.001	-0.508	0.097	0.300	0.045	1.167	0.042	1.117	0.112
69	0.302	0.104	0.489	0.001	-0.582	0.102	0.200	0.041	1.048	0.037	0.993	0.088
70	-0.136	0.102	0.296	0.001	-0.559	0.120	0.200	0.050	1.069	0.037	1.250	0.117
71	0.370	0.105	0.595	0.001	-0.863	0.085	0.300	0.045	1.085	0.039	1.085	0.118
72	0.143	0.107	0.407	0.001	-0.702	0.104	0.300	0.046	1.037	0.039	0.713	0.064
73	0.521	0.104	0.474	0.000	-0.535	0.107	0.200	0.051	1.009	0.039	0.917	0.120
74	0.069	0.107	0.481	0.001	-0.577	0.094	0.300	0.043	1.057	0.043	0.520	0.060
75	-0.378	0.106	0.646	0.001	-0.769	0.092	0.300	0.045	1.154	0.042	1.084	0.102
76	0.205	0.105	0.193	0.000	-0.340	0.086	0.200	0.049	1.115	0.042	1.067	0.091
77	0.248	0.109	0.579	0.000	-0.619	0.080	0.200	0.055	1.138	0.052	1.054	0.157
78	0.229	0.111	0.275	0.001	-0.487	0.113	0.400	0.047	1.171	0.051	0.846	0.126
79	0.111	0.106	0.487	0.001	-0.679	0.101	0.300	0.042	1.052	0.043	0.978	0.114
80	0.078	0.102	0.443	0.001	-0.594	0.084	0.200	0.039	1.070	0.044	1.216	0.120
81	0.395	0.098	0.502	0.001	-0.671	0.130	0.300	0.047	1.041	0.039	1.517	0.113
82	0.312	0.107	0.592	0.001	-0.605	0.110	0.200	0.048	1.196	0.053	1.348	0.099
83	0.056	0.109	0.755	0.001	-0.769	0.090	0.200	0.043	1.051	0.048	0.736	0.127
84	0.143	0.105	0.332	0.001	-0.489	0.093	0.400	0.051	1.096	0.040	1.134	0.111
85	-0.365	0.105	0.560	0.001	-0.636	0.104	0.200	0.048	1.085	0.041	1.187	0.180
86	-0.131	0.109	0.425	0.001	-0.499	0.073	0.200	0.046	1.052	0.036	0.748	0.070
87	0.016	0.109	0.566	0.001	-0.707	0.106	0.300	0.044	1.091	0.047	0.836	0.089
88	0.102	0.099	0.067	0.000	-0.225	0.083	0.200	0.046	0.970	0.045	1.164	0.117
89	-0.134	0.108	0.336	0.000	-0.468	0.101	0.200	0.049	1.079	0.047	0.837	0.098
90	-0.001	0.108	0.235	0.000	-0.515	0.112	0.300	0.043	1.093	0.034	1.001	0.101

PEST: 2 of 3

R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
91	0.335	0.107	0.567	0.001	-0.637	0.093	0.200	0.034	1.168	0.046	1.143	0.119
92	-0.109	0.105	0.709	0.001	-0.765	0.101	0.300	0.043	1.069	0.039	0.918	0.103
93	0.298	0.110	0.273	0.000	-0.475	0.091	0.200	0.046	1.128	0.036	0.959	0.092
94	-0.088	0.107	0.443	0.001	-0.549	0.096	0.300	0.045	1.059	0.038	0.832	0.089
95	0.209	0.104	0.734	0.001	-0.564	0.104	0.200	0.053	1.056	0.036	0.717	0.088
96	-0.222	0.105	0.342	0.001	-0.628	0.089	0.200	0.050	1.037	0.030	0.925	0.100
97	-0.214	0.113	0.413	0.000	-0.747	0.097	0.300	0.045	1.215	0.051	0.465	0.056
98	-0.058	0.108	0.736	0.001	-0.669	0.104	0.200	0.042	1.110	0.050	0.977	0.092
99	0.127	0.110	0.532	0.001	-0.724	0.096	0.200	0.053	1.163	0.048	0.970	0.093
100	0.537	0.107	0.507	0.001	-0.627	0.100	0.200	0.048	1.127	0.037	0.962	0.097
101	-0.421	0.105	0.821	0.001	-0.811	0.087	0.200	0.037	1.010	0.038	0.771	0.087
102	0.491	0.109	0.550	0.001	-0.791	0.098	0.400	0.054	1.100	0.044	0.786	0.085
103	0.861	0.104	0.566	0.001	-0.627	0.101	0.200	0.042	1.035	0.045	1.081	0.093
104	-0.234	0.109	0.573	0.001	-0.606	0.103	0.300	0.046	1.168	0.041	0.929	0.093
105	0.452	0.112	0.355	0.000	-0.601	0.107	0.200	0.049	1.286	0.046	1.170	0.130
106	0.133	0.105	0.613	0.000	-0.680	0.093	0.300	0.050	1.063	0.043	1.027	0.071
107	0.240	0.108	0.180	0.000	-0.339	0.092	0.200	0.047	1.147	0.046	0.984	0.105
108	0.594	0.109	0.736	0.001	-0.681	0.105	0.400	0.042	1.083	0.042	0.859	0.103
109	0.161	0.110	0.146	0.000	-0.684	0.102	0.300	0.041	1.114	0.047	0.721	0.080
110	-0.250	0.108	0.337	0.001	-0.472	0.079	0.200	0.040	1.098	0.043	0.916	0.096
111	0.205	0.107	0.771	0.001	-0.732	0.085	0.400	0.044	0.968	0.036	0.597	0.080
112	-0.574	0.110	0.425	0.000	-0.602	0.088	0.300	0.047	1.121	0.044	0.927	0.108
113	0.070	0.107	0.530	0.001	-0.631	0.121	0.100	0.048	1.148	0.049	1.007	0.094
114	-0.472	0.106	0.730	0.001	-0.832	0.090	0.300	0.041	1.107	0.050	1.046	0.116
115	0.631	0.105	0.104	0.000	-0.420	0.090	0.200	0.046	1.084	0.037	1.150	0.155
116	0.460	0.109	0.363	0.000	-0.698	0.083	0.200	0.047	1.108	0.044	0.652	0.085
117	0.631	0.104	0.883	0.001	-0.859	0.119	0.300	0.037	1.123	0.050	1.222	0.148
118	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
119	0.387	0.106	0.678	0.001	-0.597	0.098	0.300	0.055	1.136	0.041	1.054	0.123
120	-0.179	0.105	0.589	0.001	-0.540	0.097	0.200	0.050	1.098	0.041	1.074	0.088

PEST: 2 of 3

R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
121	-0.041	0.097	0.487	0.001	-0.419	0.108	0.300	0.041	1.011	0.041	1.517	0.121
122	-0.043	0.108	0.596	0.001	-0.692	0.110	0.300	0.050	1.102	0.051	0.923	0.097
123	0.528	0.105	0.452	0.001	-0.369	0.096	0.200	0.048	1.168	0.046	1.317	0.114
124	-0.273	0.110	0.125	0.000	-0.649	0.095	0.100	0.054	1.159	0.048	0.929	0.116
125	0.254	0.108	0.484	0.001	-0.544	0.084	0.300	0.042	1.105	0.047	0.862	0.097
126	0.575	0.109	0.508	0.001	-0.507	0.100	0.300	0.045	1.227	0.049	1.081	0.085
127	-0.022	0.108	0.123	0.000	-0.689	0.092	0.200	0.042	1.131	0.050	0.997	0.094
128	0.258	0.098	0.352	0.000	-0.741	0.098	0.300	0.046	1.020	0.045	1.271	0.127
129	0.123	0.102	0.558	0.001	-0.688	0.120	0.200	0.047	1.129	0.040	1.531	0.173
130	0.457	0.108	0.326	0.001	-0.561	0.103	0.100	0.052	1.217	0.057	1.219	0.093
131	0.242	0.105	0.667	0.001	-0.594	0.096	0.300	0.048	1.086	0.042	0.919	0.129
132	0.007	0.106	0.003	0.000	-0.307	0.093	0.300	0.043	1.098	0.037	1.039	0.082
133	0.299	0.105	0.487	0.001	-0.613	0.092	0.300	0.054	1.145	0.053	1.057	0.119
134	0.377	0.106	0.565	0.001	-0.619	0.119	0.200	0.041	1.076	0.046	0.896	0.102
135	-0.086	0.110	0.283	0.001	-0.304	0.083	0.300	0.045	1.193	0.045	1.049	0.116
136	-0.331	0.104	0.525	0.001	-0.552	0.090	0.200	0.047	1.076	0.041	1.006	0.119
137	0.377	0.108	0.351	0.000	-0.516	0.086	0.200	0.050	1.121	0.040	0.898	0.098
138	0.375	0.104	-2.187	0.000	-0.404	0.109	0.300	0.044	1.101	0.048	1.098	0.094
139	0.240	0.104	0.365	0.000	-0.329	0.095	0.200	0.050	1.110	0.050	1.113	0.112
140	0.288	0.110	0.332	0.001	-0.456	0.102	0.200	0.049	1.149	0.043	0.594	0.056
141	-0.218	0.106	0.236	0.001	-0.348	0.103	0.100	0.041	1.069	0.046	0.764	0.091
142	-0.229	0.104	0.563	0.001	-0.673	0.095	0.300	0.042	1.068	0.055	0.989	0.091
143	0.118	0.108	-9.172	0.000	-0.414	0.095	0.200	0.053	1.115	0.037	0.871	0.097
144	0.040	0.105	0.581	0.000	-0.708	0.081	0.200	0.048	1.103	0.042	1.129	0.108
145	-0.057	0.105	0.618	0.001	-0.520	0.104	0.300	0.043	1.037	0.047	0.931	0.105
146	0.214	0.106	0.671	0.001	-0.602	0.097	0.200	0.045	1.126	0.042	1.135	0.111
147	0.125	0.108	0.545	0.001	-0.648	0.094	0.300	0.050	1.140	0.042	0.986	0.152
148	0.475	0.112	0.424	0.001	-0.722	0.105	0.200	0.040	1.162	0.042	0.757	0.071
149	0.613	0.112	0.432	0.001	-0.688	0.115	0.200	0.042	1.207	0.049	0.725	0.071
150	0.731	0.104	0.781	0.001	-0.697	0.098	0.200	0.043	1.031	0.047	0.998	0.089

PEST: 2 of 3

R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
151	0.256	0.105	0.574	0.001	-0.704	0.112	0.300	0.036	1.085	0.043	0.807	0.088
152	0.093	0.109	0.520	0.001	-0.631	0.093	0.400	0.050	1.120	0.043	0.645	0.075
153	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
154	0.173	0.105	0.522	0.001	-0.622	0.115	0.300	0.052	1.053	0.045	0.911	0.087
155	0.124	0.105	0.409	0.001	-0.494	0.093	0.200	0.042	1.045	0.047	0.696	0.070
156	0.318	0.110	0.408	0.001	-0.666	0.103	0.300	0.045	1.078	0.049	0.668	0.081
157	0.513	0.105	0.411	0.001	-0.723	0.102	0.300	0.046	1.029	0.041	0.882	0.106
158	0.063	0.108	0.349	0.001	-0.457	0.109	0.200	0.049	1.139	0.040	1.068	0.095
159	0.301	0.102	0.179	0.000	-0.486	0.097	0.100	0.045	1.059	0.042	1.409	0.171
160	0.042	0.111	0.467	0.000	-0.689	0.098	0.300	0.044	1.131	0.045	0.668	0.074
161	0.282	0.109	0.591	0.001	-0.723	0.083	0.300	0.042	1.232	0.046	0.876	0.082
162	0.066	0.105	0.435	0.000	-0.497	0.095	0.300	0.052	1.093	0.048	1.142	0.101
163	0.294	0.104	0.901	0.001	-0.993	0.104	0.300	0.036	1.047	0.046	1.080	0.114
164	0.482	0.106	0.653	0.000	-0.842	0.092	0.300	0.046	1.127	0.046	1.337	0.141
165	0.451	0.111	0.627	0.000	-0.596	0.095	0.200	0.047	1.151	0.051	0.773	0.087
166	0.445	0.103	0.581	0.001	-0.577	0.115	0.300	0.047	1.031	0.047	0.954	0.093
167	0.150	0.105	0.191	0.000	-0.446	0.090	0.300	0.041	1.065	0.044	1.014	0.095
168	0.065	0.102	0.569	0.001	-0.638	0.098	0.300	0.042	1.079	0.043	1.164	0.112
169	0.306	0.102	0.261	0.000	-0.325	0.082	0.300	0.045	1.042	0.045	1.187	0.081
170	0.437	0.105	0.431	0.001	-0.567	0.100	0.300	0.049	1.060	0.041	0.951	0.137
171	0.206	0.109	0.395	0.001	-0.555	0.092	0.300	0.043	1.162	0.057	0.956	0.139
172	-0.450	0.102	0.524	0.001	-0.789	0.116	0.300	0.043	1.042	0.035	1.192	0.195
173	0.345	0.108	0.460	0.001	-0.560	0.104	0.200	0.051	1.142	0.038	0.938	0.111
174	0.049	0.103	0.030	0.000	-0.459	0.101	0.300	0.055	1.104	0.041	1.346	0.130
175	-0.123	0.110	0.636	0.001	-0.892	0.110	0.300	0.041	1.172	0.050	0.820	0.074
176	0.068	0.105	0.734	0.001	-0.929	0.088	0.300	0.048	1.101	0.057	1.107	0.126
177	0.040	0.105	0.380	0.000	-0.665	0.099	0.300	0.051	1.049	0.044	1.073	0.113
178	0.076	0.105	0.650	0.001	-0.462	0.102	0.200	0.046	1.046	0.046	0.974	0.102
179	-0.351	0.104	0.745	0.001	-0.695	0.109	0.300	0.049	1.114	0.044	1.254	0.170
180	-0.024	0.110	0.602	0.000	-0.803	0.091	0.300	0.052	1.168	0.038	0.901	0.089

PEST: 2 of 3

R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
181	0.434	0.102	0.391	0.000	-0.466	0.101	0.300	0.041	0.976	0.042	0.970	0.099
182	0.051	0.106	0.552	0.001	-0.670	0.118	0.200	0.048	1.099	0.042	1.082	0.101
183	0.028	0.106	0.489	0.001	-0.614	0.099	0.300	0.049	1.125	0.055	1.058	0.099
184	0.739	0.104	0.353	0.000	-0.370	0.107	0.200	0.048	1.173	0.040	1.363	0.140
185	0.326	0.109	0.241	0.000	-0.627	0.097	0.300	0.051	1.114	0.044	0.810	0.098
186	0.380	0.100	0.495	0.001	-0.684	0.106	0.400	0.053	0.996	0.045	1.016	0.101
187	0.304	0.110	-2.261	0.000	-0.209	0.100	0.300	0.051	1.186	0.042	0.808	0.087
188	0.190	0.109	0.677	0.001	-0.785	0.087	0.300	0.039	1.075	0.044	0.853	0.076
189	0.101	0.106	-4.001	0.000	-0.293	0.114	0.300	0.049	1.068	0.046	0.814	0.091
190	0.358	0.108	0.637	0.001	-0.588	0.105	0.200	0.053	1.056	0.044	0.709	0.082
191	-0.018	0.106	0.286	0.000	-0.384	0.105	0.300	0.043	1.135	0.052	1.101	0.096
192	0.107	0.102	0.615	0.001	-0.653	0.088	0.300	0.040	1.013	0.035	0.952	0.115
193	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
194	0.213	0.106	0.749	0.001	-0.643	0.105	0.200	0.051	1.028	0.042	0.813	0.101
195	0.211	0.102	0.684	0.001	-0.733	0.094	0.300	0.046	1.031	0.048	1.074	0.125
196	0.469	0.107	0.250	0.000	-0.354	0.100	0.100	0.043	1.046	0.041	0.910	0.112
197	0.102	0.109	0.496	0.001	-0.536	0.094	0.100	0.043	1.050	0.040	0.676	0.069
198	-0.121	0.111	0.318	0.001	-0.567	0.113	0.200	0.051	1.173	0.052	0.832	0.105
199	0.226	0.103	0.733	0.001	-0.671	0.108	0.200	0.045	1.028	0.046	1.047	0.123
200	-0.325	0.111	0.158	0.000	-0.678	0.099	0.300	0.050	1.129	0.043	0.603	0.061

PEST: p-value												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	-0.123	0.114	0.361	0.000	-0.521	0.108	0.331	0.050	1.204	0.047	0.893	0.101
2	0.151	0.103	0.376	0.001	-0.412	0.100	0.181	0.050	1.011	0.050	0.978	0.080
3	-0.062	0.107	0.396	0.000	-0.622	0.088	0.138	0.046	1.116	0.035	0.920	0.089
4	-0.455	0.100	0.381	0.001	-0.564	0.106	0.138	0.048	1.068	0.049	1.358	0.111
5	-0.298	0.110	0.583	0.001	-0.692	0.112	0.245	0.046	1.140	0.052	0.894	0.109
6	0.101	0.109	0.349	0.001	-0.554	0.092	0.239	0.045	1.163	0.049	0.900	0.075
7	-0.046	0.105	0.395	0.000	-0.438	0.078	0.192	0.047	1.103	0.040	1.101	0.121
8	-0.161	0.109	0.215	0.001	-0.441	0.095	0.137	0.043	1.141	0.047	0.992	0.130
9	0.165	0.103	0.674	0.001	-0.707	0.094	0.374	0.050	1.034	0.038	1.122	0.117
10	-0.162	0.110	0.093	0.000	-0.330	0.076	0.170	0.045	1.114	0.045	0.875	0.096
11	0.442	0.102	0.525	0.000	-0.671	0.079	0.220	0.039	1.045	0.040	1.250	0.132
12	0.503	0.108	0.291	0.000	-0.260	0.100	0.127	0.043	1.134	0.051	1.224	0.107
13	0.264	0.106	0.091	0.000	-0.498	0.095	0.173	0.040	1.048	0.049	1.054	0.105
14	0.243	0.112	0.147	0.000	-0.501	0.111	0.333	0.042	1.198	0.041	0.816	0.081
15	0.175	0.108	0.285	0.000	-0.502	0.102	0.169	0.052	1.101	0.046	0.750	0.095
16	-0.365	0.105	0.428	0.001	-0.793	0.116	0.152	0.046	1.101	0.046	1.009	0.092
17	0.051	0.100	0.165	0.000	-0.675	0.104	0.177	0.050	1.066	0.044	1.424	0.127
18	0.061	0.107	0.074	0.000	-0.570	0.092	0.309	0.045	1.076	0.044	0.990	0.129
19	-0.272	0.109	0.179	0.000	-0.490	0.120	0.201	0.042	1.129	0.051	1.006	0.110
20	0.544	0.107	0.027	0.000	-0.411	0.092	0.267	0.046	1.062	0.046	0.896	0.088
21	0.242	0.109	0.337	0.001	-0.660	0.071	0.237	0.051	1.124	0.057	0.803	0.076
22	-0.023	0.101	0.402	0.000	-0.586	0.096	0.221	0.045	1.070	0.045	1.378	0.104
23	0.337	0.109	0.510	0.000	-0.598	0.100	0.336	0.049	1.169	0.045	1.094	0.079
24	0.092	0.105	0.095	0.000	-0.406	0.101	0.141	0.040	1.105	0.048	1.106	0.157
25	-0.252	0.103	0.510	0.001	-0.429	0.106	0.270	0.031	1.041	0.049	1.086	0.084
26	0.289	0.108	0.255	0.000	-0.394	0.106	-0.005	0.044	1.086	0.047	0.937	0.085
27	-0.401	0.101	0.379	0.001	-0.459	0.098	0.172	0.035	1.059	0.040	1.354	0.115
28	-0.145	0.108	0.181	0.000	-0.370	0.094	0.094	0.037	1.110	0.046	0.808	0.083
29	0.230	0.111	0.567	0.001	-0.562	0.109	0.325	0.040	1.173	0.052	0.960	0.082
30	-0.187	0.108	0.498	0.000	-0.480	0.098	0.231	0.050	1.106	0.052	0.716	0.089

PEST: p-value												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
31	-0.092	0.107	0.393	0.000	-0.540	0.094	0.136	0.044	1.052	0.046	0.823	0.085
32	0.644	0.111	0.234	0.000	-0.466	0.109	0.199	0.045	1.112	0.050	0.648	0.078
33	0.382	0.112	0.489	0.001	-0.755	0.083	0.174	0.039	1.145	0.047	0.665	0.063
34	0.466	0.104	0.152	0.000	-0.474	0.092	0.166	0.047	1.115	0.046	1.160	0.106
35	-0.337	0.103	0.236	0.000	-0.360	0.093	0.231	0.041	1.042	0.044	0.877	0.073
36	0.016	0.104	0.191	0.001	-0.431	0.103	0.287	0.040	1.110	0.044	1.267	0.136
37	-0.069	0.109	0.688	0.001	-0.686	0.100	0.227	0.046	1.110	0.043	0.823	0.115
38	0.135	0.106	0.685	0.001	-0.643	0.095	0.247	0.044	1.100	0.037	1.229	0.110
39	-0.196	0.103	0.117	0.000	-0.426	0.100	0.257	0.046	1.049	0.037	1.194	0.141
40	0.056	0.109	0.243	0.000	-0.450	0.098	0.177	0.049	1.098	0.048	0.903	0.122
41	0.351	0.107	0.588	0.001	-0.522	0.123	0.141	0.049	1.134	0.048	1.119	0.118
42	0.000	0.104	0.152	0.000	-0.573	0.102	0.133	0.050	1.009	0.045	0.946	0.095
43	-0.101	0.102	0.353	0.001	-0.700	0.102	0.130	0.042	1.055	0.046	1.155	0.129
44	-0.055	0.108	0.169	0.000	-0.342	0.115	0.078	0.043	1.192	0.051	1.495	0.156
45	0.385	0.105	0.148	0.000	-0.320	0.095	0.053	0.045	1.000	0.040	0.610	0.106
46	0.465	0.105	0.649	0.001	-0.723	0.102	0.158	0.038	1.123	0.050	1.250	0.108
47	0.018	0.110	0.203	0.000	-0.330	0.082	0.217	0.041	1.158	0.050	1.034	0.114
48	0.274	0.106	0.495	0.000	-0.670	0.110	0.187	0.047	1.047	0.038	0.711	0.087
49	0.540	0.104	0.685	0.001	-0.688	0.112	0.198	0.050	1.076	0.041	1.058	0.112
50	0.325	0.109	0.511	0.001	-0.801	0.101	0.224	0.051	1.120	0.042	0.830	0.113
51	-0.228	0.108	0.137	0.000	-0.501	0.102	0.217	0.048	1.142	0.045	0.908	0.088
52	0.145	0.106	0.377	0.000	-0.361	0.097	0.069	0.037	1.071	0.045	0.934	0.093
53	0.043	0.105	0.697	0.001	-0.822	0.098	0.276	0.059	1.081	0.038	1.091	0.120
54	-0.398	0.108	0.263	0.000	-0.360	0.098	0.128	0.046	1.023	0.035	0.744	0.092
55	-0.274	0.107	0.107	0.000	-0.287	0.124	0.178	0.042	1.126	0.044	0.976	0.108
56	-0.363	0.105	0.201	0.000	-0.331	0.085	0.171	0.046	1.078	0.046	0.938	0.086
57	0.179	0.107	0.336	0.000	-0.513	0.087	0.162	0.051	1.140	0.054	1.003	0.102
58	-0.137	0.108	0.327	0.000	-0.451	0.073	0.213	0.045	1.058	0.048	0.834	0.083
59	0.083	0.103	0.168	0.000	-0.464	0.089	0.122	0.049	1.050	0.037	1.106	0.093
60	-0.441	0.106	-3.219	0.000	-0.469	0.105	0.271	0.053	1.057	0.045	0.951	0.089

PEST: p-value												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
61	0.247	0.112	0.248	0.000	-0.532	0.092	0.063	0.045	1.121	0.049	0.581	0.073
62	-0.001	0.106	0.599	0.001	-0.680	0.098	0.363	0.046	1.115	0.042	1.081	0.104
63	-0.533	0.106	0.729	0.001	-0.679	0.099	0.266	0.051	1.127	0.047	0.949	0.081
64	-0.456	0.105	0.176	0.000	-0.395	0.092	0.061	0.045	1.047	0.039	0.837	0.076
65	0.022	0.110	0.386	0.000	-0.600	0.092	0.196	0.044	1.065	0.045	0.659	0.083
66	0.136	0.107	0.545	0.001	-0.680	0.111	0.131	0.045	1.110	0.048	1.093	0.098
67	0.010	0.106	0.456	0.001	-0.631	0.100	0.193	0.044	1.163	0.053	1.113	0.114
68	0.367	0.110	0.428	0.001	-0.498	0.097	0.188	0.044	1.165	0.042	1.117	0.112
69	0.364	0.103	0.518	0.001	-0.597	0.102	0.273	0.041	1.048	0.037	1.002	0.088
70	-0.158	0.101	-2.188	0.000	-0.390	0.118	0.256	0.049	1.070	0.037	1.269	0.118
71	0.314	0.105	0.561	0.001	-0.846	0.085	0.209	0.044	1.083	0.040	1.077	0.117
72	0.089	0.109	0.091	0.000	-0.543	0.102	0.261	0.045	1.036	0.039	0.703	0.064
73	0.432	0.104	0.242	0.000	-0.419	0.106	0.145	0.050	1.009	0.039	0.919	0.120
74	0.088	0.104	0.346	0.000	-0.510	0.093	0.293	0.043	1.056	0.043	0.525	0.060
75	-0.358	0.105	0.641	0.001	-0.767	0.092	0.310	0.046	1.154	0.042	1.104	0.103
76	0.259	0.106	0.166	0.000	-0.327	0.086	0.175	0.049	1.117	0.042	1.058	0.090
77	0.193	0.107	0.544	0.000	-0.601	0.080	0.097	0.054	1.133	0.052	1.088	0.160
78	0.133	0.112	-4.450	0.000	-0.352	0.112	0.281	0.045	1.170	0.051	0.839	0.125
79	0.021	0.105	0.250	0.000	-0.564	0.100	0.273	0.041	1.049	0.043	1.001	0.116
80	0.129	0.102	0.448	0.001	-0.596	0.084	0.252	0.039	1.070	0.044	1.216	0.119
81	0.266	0.096	0.224	0.000	-0.535	0.129	0.217	0.046	1.039	0.039	1.563	0.114
82	0.346	0.105	0.597	0.001	-0.608	0.110	0.183	0.048	1.196	0.053	1.380	0.101
83	-0.056	0.109	0.535	0.000	-0.660	0.089	0.088	0.041	1.049	0.048	0.737	0.127
84	0.137	0.106	0.120	0.000	-0.383	0.093	0.356	0.050	1.097	0.040	1.134	0.111
85	-0.444	0.104	0.525	0.001	-0.619	0.104	0.130	0.046	1.083	0.041	1.226	0.183
86	-0.282	0.107	0.405	0.001	-0.489	0.073	0.190	0.045	1.049	0.036	0.762	0.070
87	0.069	0.110	0.372	0.000	-0.613	0.104	0.313	0.044	1.091	0.047	0.848	0.090
88	0.102	0.099	0.067	0.000	-0.225	0.083	0.161	0.046	0.970	0.045	1.164	0.117
89	-0.199	0.109	0.371	0.000	-0.486	0.102	0.152	0.047	1.077	0.047	0.819	0.096
90	-0.079	0.107	0.205	0.000	-0.500	0.112	0.230	0.042	1.092	0.034	1.013	0.102

PEST: p-value												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
91	0.286	0.109	0.572	0.001	-0.639	0.093	0.188	0.034	1.169	0.046	1.130	0.118
92	-0.104	0.106	0.706	0.001	-0.763	0.101	0.314	0.043	1.069	0.039	0.905	0.102
93	0.378	0.110	0.280	0.000	-0.478	0.091	0.256	0.046	1.128	0.036	0.953	0.091
94	-0.179	0.106	0.472	0.001	-0.563	0.096	0.213	0.044	1.060	0.037	0.849	0.090
95	0.133	0.106	0.696	0.001	-0.546	0.104	0.149	0.053	1.055	0.036	0.700	0.087
96	-0.276	0.105	0.103	0.000	-0.507	0.087	0.148	0.050	1.037	0.030	0.929	0.101
97	-0.299	0.114	0.105	0.000	-0.592	0.095	0.188	0.043	1.214	0.051	0.456	0.055
98	-0.158	0.109	0.457	0.000	-0.535	0.102	0.157	0.041	1.110	0.050	0.979	0.092
99	0.163	0.109	0.534	0.001	-0.725	0.096	0.188	0.053	1.164	0.048	0.997	0.094
100	0.414	0.107	0.252	0.000	-0.502	0.099	0.133	0.046	1.126	0.037	0.957	0.097
101	-0.401	0.106	0.775	0.001	-0.789	0.086	0.215	0.037	1.011	0.038	0.769	0.087
102	0.494	0.109	0.541	0.001	-0.787	0.098	0.384	0.054	1.100	0.044	0.787	0.085
103	0.718	0.103	0.267	0.000	-0.477	0.099	0.167	0.041	1.033	0.045	1.072	0.092
104	-0.308	0.111	0.537	0.001	-0.588	0.103	0.193	0.044	1.166	0.041	0.898	0.091
105	0.399	0.114	0.337	0.000	-0.591	0.107	0.136	0.049	1.290	0.046	1.147	0.129
106	0.054	0.103	0.259	0.000	-0.506	0.090	0.231	0.049	1.064	0.043	1.031	0.071
107	0.021	0.106	0.043	0.000	-0.271	0.092	0.171	0.046	1.148	0.046	0.982	0.107
108	0.451	0.110	0.471	0.000	-0.553	0.103	0.274	0.039	1.081	0.042	0.859	0.103
109	0.110	0.111	0.048	0.000	-0.637	0.101	0.208	0.041	1.111	0.047	0.718	0.081
110	-0.179	0.107	0.349	0.001	-0.478	0.079	0.212	0.040	1.098	0.043	0.922	0.097
111	0.154	0.105	0.740	0.001	-0.717	0.085	0.317	0.043	0.968	0.036	0.612	0.081
112	-0.587	0.108	0.406	0.000	-0.592	0.088	0.245	0.047	1.121	0.044	0.951	0.110
113	0.070	0.107	0.530	0.001	-0.631	0.121	0.094	0.048	1.148	0.049	1.007	0.094
114	-0.513	0.108	0.502	0.000	-0.714	0.089	0.183	0.039	1.107	0.050	1.037	0.115
115	0.594	0.107	0.088	0.000	-0.411	0.090	0.211	0.045	1.085	0.037	1.125	0.153
116	0.397	0.108	0.348	0.000	-0.690	0.083	0.140	0.047	1.107	0.044	0.663	0.086
117	0.511	0.106	0.673	0.000	-0.752	0.118	0.212	0.037	1.120	0.050	1.169	0.144
118	0.288	0.103	-6.070	0.000	-0.428	0.100	0.045	0.040	1.025	0.046	1.110	0.088
119	0.430	0.105	0.661	0.001	-0.589	0.098	0.314	0.055	1.135	0.041	1.046	0.124
120	-0.207	0.105	0.578	0.001	-0.535	0.097	0.222	0.050	1.098	0.041	1.076	0.089

PEST: p-value												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
121	-0.097	0.095	0.455	0.001	-0.403	0.108	0.252	0.039	1.010	0.041	1.545	0.122
122	0.052	0.107	0.600	0.001	-0.694	0.109	0.302	0.049	1.101	0.051	0.941	0.098
123	0.656	0.106	0.329	0.000	-0.307	0.096	0.249	0.048	1.171	0.046	1.303	0.113
124	-0.273	0.110	0.125	0.000	-0.649	0.095	0.121	0.054	1.159	0.048	0.929	0.116
125	0.254	0.108	0.484	0.001	-0.544	0.084	0.255	0.042	1.105	0.047	0.862	0.097
126	0.534	0.109	0.493	0.001	-0.500	0.100	0.256	0.045	1.227	0.049	1.076	0.084
127	0.031	0.108	0.159	0.000	-0.707	0.092	0.231	0.041	1.129	0.050	1.003	0.094
128	0.086	0.099	0.281	0.000	-0.705	0.098	0.217	0.044	1.020	0.045	1.271	0.127
129	0.056	0.101	0.296	0.000	-0.559	0.118	0.115	0.045	1.130	0.040	1.513	0.172
130	0.457	0.108	0.326	0.001	-0.561	0.103	0.123	0.052	1.217	0.057	1.219	0.093
131	0.300	0.100	0.628	0.001	-0.574	0.096	0.215	0.047	1.085	0.042	0.945	0.133
132	-0.054	0.108	-7.752	0.000	-0.301	0.093	0.245	0.043	1.100	0.037	1.005	0.080
133	0.317	0.109	0.705	0.000	-0.721	0.091	0.205	0.054	1.147	0.053	1.028	0.115
134	0.341	0.103	0.213	0.000	-0.446	0.117	0.194	0.041	1.077	0.046	0.913	0.104
135	-0.089	0.109	0.259	0.001	-0.292	0.083	0.224	0.045	1.192	0.045	1.049	0.116
136	-0.388	0.104	0.251	0.000	-0.416	0.088	0.168	0.047	1.079	0.041	1.012	0.120
137	0.223	0.107	0.003	0.000	-0.337	0.083	0.104	0.048	1.119	0.040	0.907	0.098
138	0.056	0.104	0.034	0.000	-0.432	0.109	0.275	0.044	1.097	0.048	1.081	0.094
139	0.134	0.108	0.053	0.000	-0.172	0.093	0.187	0.050	1.111	0.050	1.103	0.112
140	0.277	0.111	0.132	0.000	-0.358	0.101	0.237	0.049	1.149	0.043	0.595	0.056
141	-0.246	0.112	0.063	0.000	-0.265	0.101	0.086	0.041	1.074	0.046	0.728	0.088
142	-0.280	0.105	0.226	0.000	-0.503	0.091	0.215	0.040	1.067	0.056	0.978	0.091
143	0.043	0.109	-1.178	0.000	-0.401	0.095	0.210	0.053	1.114	0.037	0.869	0.096
144	0.022	0.105	0.604	0.000	-0.719	0.081	0.221	0.047	1.103	0.042	1.118	0.108
145	-0.056	0.107	0.424	0.000	-0.426	0.103	0.277	0.043	1.037	0.047	0.944	0.106
146	0.142	0.108	0.717	0.001	-0.626	0.097	0.167	0.044	1.131	0.042	1.116	0.111
147	0.162	0.108	0.370	0.000	-0.558	0.094	0.385	0.050	1.137	0.042	1.008	0.153
148	0.489	0.112	0.428	0.001	-0.724	0.105	0.217	0.040	1.162	0.042	0.759	0.071
149	0.642	0.111	0.438	0.001	-0.691	0.115	0.237	0.042	1.207	0.050	0.728	0.072
150	0.654	0.105	0.763	0.001	-0.688	0.098	0.114	0.043	1.030	0.047	0.988	0.088

PEST: p-value												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
151	0.116	0.104	0.336	0.000	-0.583	0.110	0.221	0.035	1.082	0.043	0.778	0.086
152	0.067	0.109	0.513	0.001	-0.628	0.093	0.330	0.050	1.120	0.043	0.644	0.075
153	0.605	0.106	0.024	0.000	-0.521	0.093	0.181	0.037	1.067	0.042	1.122	0.103
154	0.154	0.107	0.521	0.001	-0.622	0.115	0.272	0.052	1.055	0.045	0.895	0.086
155	0.014	0.108	0.357	0.001	-0.468	0.092	0.173	0.041	1.048	0.048	0.681	0.069
156	0.261	0.110	0.392	0.001	-0.658	0.103	0.246	0.044	1.079	0.049	0.664	0.081
157	0.466	0.108	0.362	0.001	-0.698	0.102	0.249	0.044	1.029	0.041	0.852	0.104
158	0.020	0.108	0.364	0.001	-0.464	0.109	0.205	0.049	1.139	0.040	1.064	0.095
159	0.198	0.103	-1.576	0.000	-0.311	0.094	0.037	0.042	1.058	0.043	1.409	0.171
160	0.019	0.112	0.465	0.000	-0.688	0.098	0.239	0.044	1.131	0.045	0.665	0.073
161	0.256	0.109	0.581	0.001	-0.717	0.083	0.223	0.041	1.231	0.046	0.878	0.082
162	-0.124	0.102	0.212	0.000	-0.386	0.093	0.196	0.050	1.091	0.048	1.163	0.102
163	0.211	0.103	0.616	0.000	-0.846	0.103	0.223	0.035	1.045	0.047	1.084	0.114
164	0.464	0.104	0.661	0.000	-0.846	0.092	0.353	0.046	1.127	0.046	1.341	0.142
165	0.422	0.111	0.600	0.000	-0.582	0.094	0.139	0.047	1.150	0.051	0.773	0.087
166	0.362	0.106	0.370	0.000	-0.475	0.114	0.231	0.047	1.032	0.046	0.952	0.093
167	0.031	0.104	-1.073	0.000	-0.296	0.088	0.190	0.040	1.064	0.044	1.022	0.095
168	-0.012	0.102	0.187	0.000	-0.446	0.097	0.213	0.040	1.078	0.044	1.154	0.112
169	0.250	0.102	0.223	0.000	-0.306	0.082	0.245	0.045	1.041	0.045	1.188	0.081
170	0.327	0.104	0.182	0.000	-0.446	0.098	0.155	0.048	1.059	0.040	0.957	0.139
171	0.162	0.108	0.350	0.001	-0.533	0.093	0.250	0.043	1.161	0.057	0.963	0.139
172	-0.506	0.102	0.502	0.001	-0.778	0.116	0.202	0.042	1.041	0.036	1.189	0.195
173	0.426	0.109	0.492	0.001	-0.576	0.104	0.253	0.050	1.142	0.038	0.920	0.110
174	-0.026	0.104	-1.321	0.000	-0.437	0.101	0.198	0.054	1.104	0.041	1.325	0.129
175	-0.137	0.110	0.612	0.001	-0.880	0.110	0.244	0.041	1.171	0.050	0.812	0.073
176	-0.002	0.103	0.710	0.001	-0.917	0.088	0.186	0.047	1.098	0.056	1.136	0.128
177	0.012	0.105	0.356	0.000	-0.652	0.099	0.224	0.051	1.049	0.044	1.074	0.114
178	-0.016	0.106	0.414	0.000	-0.343	0.101	0.161	0.046	1.047	0.046	0.971	0.103
179	-0.400	0.104	0.512	0.000	-0.576	0.107	0.250	0.049	1.113	0.044	1.259	0.170
180	0.000	0.109	0.623	0.000	-0.814	0.092	0.275	0.052	1.169	0.038	0.913	0.089

PEST: p-value												
R	CLADIFF		3WAYINT		ILDIF		ILADIFF		IAWC		CA	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
181	0.503	0.100	0.398	0.000	-0.469	0.101	0.274	0.041	0.974	0.042	0.987	0.101
182	0.077	0.107	0.568	0.001	-0.679	0.118	0.272	0.048	1.101	0.042	1.067	0.100
183	-0.062	0.107	0.466	0.001	-0.602	0.099	0.195	0.047	1.123	0.055	1.045	0.098
184	0.690	0.105	0.342	0.000	-0.365	0.107	0.171	0.047	1.173	0.040	1.349	0.140
185	0.305	0.109	0.195	0.000	-0.604	0.097	0.309	0.050	1.114	0.044	0.810	0.098
186	0.302	0.102	0.184	0.000	-0.529	0.104	0.300	0.053	0.996	0.045	1.011	0.100
187	0.264	0.110	-2.329	0.000	-0.205	0.100	0.235	0.050	1.185	0.042	0.815	0.088
188	0.236	0.105	0.465	0.000	-0.681	0.086	0.203	0.038	1.073	0.044	0.870	0.077
189	-0.012	0.105	-9.203	0.000	-0.249	0.114	0.199	0.047	1.065	0.046	0.829	0.092
190	0.288	0.110	0.600	0.001	-0.569	0.104	0.122	0.052	1.055	0.044	0.696	0.081
191	-0.050	0.105	0.252	0.000	-0.367	0.105	0.217	0.042	1.135	0.052	1.129	0.098
192	-0.078	0.102	0.521	0.001	-0.608	0.088	0.140	0.039	1.010	0.034	0.956	0.115
193	-0.009	0.106	0.347	0.000	-0.347	0.102	0.089	0.041	1.203	0.048	0.764	0.117
194	0.195	0.105	0.743	0.001	-0.640	0.105	0.152	0.051	1.028	0.042	0.821	0.101
195	0.141	0.104	0.449	0.000	-0.620	0.092	0.323	0.046	1.032	0.048	1.084	0.125
196	0.469	0.107	0.250	0.000	-0.354	0.100	0.098	0.043	1.046	0.041	0.910	0.112
197	0.102	0.109	0.496	0.001	-0.536	0.094	0.141	0.043	1.050	0.040	0.676	0.069
198	-0.162	0.113	0.133	0.000	-0.477	0.112	0.169	0.050	1.174	0.052	0.825	0.104
199	0.153	0.103	0.429	0.000	-0.518	0.106	0.182	0.044	1.030	0.047	1.065	0.124
200	-0.324	0.111	0.167	0.000	-0.682	0.099	0.274	0.050	1.129	0.043	0.604	0.061

ประวัติผู้เขียนวิทยานิพนธ์

นางสาวชลี ภัทรพิชญธรรม เกิดวันที่ 17 พฤษภาคม พ.ศ. 2522 สำเร็จการศึกษาครุศาสตรบัณฑิต (เกียรตินิยมอันดับ 1) สาขามัธยมศึกษาวิทยาศาสตร์ วิชาเอกคณิตศาสตร์ ในปีการศึกษา 2544 และสำเร็จการศึกษาครุศาสตรมหาบัณฑิต สาขาวิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2546 จากนั้นเข้าทำงาน ณ ศูนย์การศึกษาต่อเนื่องแห่งจุฬาลงกรณ์มหาวิทยาลัย ตำแหน่งเจ้าหน้าที่พัฒนาหลักสูตร โครงการ Math Square และเข้าศึกษาต่อในหลักสูตรครุศาสตรดุษฎีบัณฑิต สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2550 โดยได้รับทุนโครงการเครือข่ายเชิงกลยุทธ์เพื่อการผลิตและพัฒนาอาจารย์ในสถาบันอุดมศึกษาแบบร่วมในประเทศและต่างประเทศ ประจำปี 2550 ของสำนักงานคณะกรรมการการอุดมศึกษา โดยศึกษาค้นคว้า ณ จุฬาลงกรณ์มหาวิทยาลัย 2 ปีและศึกษาค้นคว้าในบางรายวิชาที่เกี่ยวข้องกับสาขาวิชาวัดผลเพิ่มเติมและทำวิทยานิพนธ์ร่วมกับ Professor Akihito Kamata (Ph.D) ณ Department of Educational Methodology, Policy, and Leadership (EMPL). College of Education, University of Oregon ประเทศสหรัฐอเมริกา ในฐานะ International Research Scholar เป็นระยะเวลา 1 ปี