

การเปรียบเทียบประสิทธิภาพของกระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนอง  
ข้อสอบแบบพหุมิติ สำหรับแบบสอบรูปแบบผสม ภายใต้ผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม



นางสาวพนิดา พานิชวัฒน์นะ

บทคัดย่อและแฟ้มข้อมูลฉบับเต็มของวิทยานิพนธ์ตั้งแต่ปีการศึกษา 2554 ที่ให้บริการในคลังปัญญาจุฬาฯ (CUIR)  
เป็นแฟ้มข้อมูลของนิสิตเจ้าของวิทยานิพนธ์ ที่ส่งผ่านทางบัณฑิตวิทยาลัย

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

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต

สาขาวิชาการศึกษาและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2560

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

A COMPARISON OF EFFICIENCY AMONG FULL MIRT OBSERVED-SCORE  
EQUATING PROCEDURES FOR MIXED-FORMAT TESTS UNDER  
NON-EQUIVALENT GROUPS WITH ANCHOR TEST DESIGN

Miss Panida Panidvadtana



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 2017

Copyright of Chulalongkorn University



จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**

หัวข้อวิทยานิพนธ์

การเปรียบเทียบประสิทธิภาพของกระบวนการปรับเทียบ  
คะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุ  
มิติ สำหรับแบบสอบรูปแบบผสม ภายใต้ผู้สอบกลุ่มไม่เท่า  
เทียมกันโดยใช้ข้อสอบร่วม

โดย

นางสาวพนิดา พานิชวัฒนะ

สาขาวิชา

การวัดและประเมินผลการศึกษา

อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ

อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

อาจารย์ ดร.สิวะโชติ ศรีสุทธิยากร

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้บัณฑิตวิทยาลัย  
ของการศึกษาตามหลักสูตรปริญญาครุศึกษบัณฑิต

..... คณะบดีคณะครุศาสตร์

(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ

(ศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก

(รองศาสตราจารย์ ดร.ศิริเดช สุชีวะ)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม

(อาจารย์ ดร.สิวะโชติ ศรีสุทธิยากร)

..... กรรมการ

(รองศาสตราจารย์ ดร.โชติกา ภาษีผล)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ณัฐภรณ์ หลาวทอง)

..... กรรมการภายนอกมหาวิทยาลัย

(ผู้ช่วยศาสตราจารย์ ดร.สังวรณ์ ังดกระโทก)

พนิดา พานิชวัฒนะ : การเปรียบเทียบประสิทธิภาพของกระบวนการปรับเทียบคะแนนสังเกตได้ตาม  
ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ สำหรับแบบสอบรูปแบบผสม ภายใต้ผู้สอบกลุ่มไม่เท่าเทียมกัน  
โดยใช้ข้อสอบร่วม (A COMPARISON OF EFFICIENCY AMONG FULL MIRT OBSERVED-SCORE  
EQUATING PROCEDURES FOR MIXED-FORMAT TESTS UNDER NON-EQUIVALENT GROUPS  
WITH ANCHOR TEST DESIGN) อ.ที่ปรึกษาวิทยานิพนธ์หลัก: รศ. ดร.ศิริเดช สุขชีวะ, อ.ที่ปรึกษา  
วิทยานิพนธ์ร่วม: อ. ดร.สฤษดิ์ โชติ ศรีสวัสดิ์, หน้า.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของกระบวนการปรับเทียบคะแนนสังเกตได้  
ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูปแบบ (MOSE) 2 แบบ ได้แก่ กระบวนการ MOSE ที่ประมาณ  
ค่าพารามิเตอร์พร้อมกัน (CMOSE) และกระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แยกกันและใช้การเชื่อมโยง  
สเกลด้วยวิธี TCF (SMOSE) ด้วยการจำลองแบบมอนติคาร์โล สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบผู้สอบ  
กลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม (NEAT) เมื่อโครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบร่วม และ  
อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า แตกต่างกัน ประสิทธิภาพของกระบวนการปรับเทียบคะแนน  
วัดจากตัวบ่งชี้ 3 ตัว ได้แก่ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ค่าความลำเอียงสัมพัทธ์  
(RB) และสัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (CVSE)

ผลการวิจัยด้านประสิทธิภาพโดยรวม (RMSE) พบว่า มีปฏิสัมพันธ์อิทธิพลของสัดส่วนคะแนนข้อสอบ  
ร่วมกับอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าต่อค่า RMSE อย่างมีนัยสำคัญทางสถิติที่ระดับ .05  
เมื่อวิเคราะห์อิทธิพลอย่างง่ายในเงื่อนไขสัดส่วนคะแนนข้อสอบร่วม (1) กรณีอัตราส่วนฯ คือ 60:40 และ 50:50  
พบว่า กระบวนการ MOSE มีประสิทธิภาพโดยรวมมากขึ้น เมื่อสัดส่วนคะแนนข้อสอบร่วมเพิ่มขึ้นอย่างมีนัยสำคัญ  
ทางสถิติที่ระดับ .05 และ (2) กรณีอัตราส่วนฯ คือ 70:30 และ 40:60 พบว่า ประสิทธิภาพโดยรวมมากที่สุด  
เมื่อสัดส่วนคะแนนข้อสอบร่วม คือ 20%, 30% และ 10% ตามลำดับ เมื่อวิเคราะห์อิทธิพลอย่างง่ายในเงื่อนไข  
อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า พบว่า (1) สัดส่วนคะแนนข้อสอบร่วม 30% พบว่า  
มีประสิทธิภาพโดยรวมมากที่สุด เมื่ออัตราส่วนฯ คือ 60:40, 50:50, 40:60 และ 70:30 ตามลำดับ (2) สัดส่วน  
คะแนนข้อสอบร่วม 20% พบว่า มีประสิทธิภาพโดยรวมมากที่สุด เมื่ออัตราส่วนฯ คือ 40:60, 70:30, 50:50 และ  
60:40 ตามลำดับ และ (3) สัดส่วนคะแนนข้อสอบร่วม คือ 10% พบว่า มีประสิทธิภาพโดยรวมมากที่สุด  
เมื่ออัตราส่วนฯ คือ 40:60, 70:30 และ 60:40 ตามลำดับ ผลด้านความแม่นยำ (RB) พบว่า มีปฏิสัมพันธ์อิทธิพล  
ระหว่างตัวแปรอิสระทั้ง 4 ตัวแปรต่อค่า RB อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนผลด้านความถูกต้อง (CVSE)  
พบว่า ไม่มีปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อค่า CVSE ในเชิงปฏิบัติ เมื่อพิจารณาอิทธิพลหลักของ  
ตัวแปรอิสระ พบว่า ตัวแปรอิสระที่มีผลต่อค่า CVSE ระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ได้แก่  
กระบวนการ MOSE โครงสร้างมิติความสามารถ และสัดส่วนคะแนนข้อสอบร่วม โดยมีขนาดอิทธิพลร้อยละ 6.90,  
.30 และ .90 ตามลำดับ ขณะที่อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าไม่มีผลต่อค่า CVSE

ภาควิชา	วิจัยและจิตวิทยาการศึกษา	ลายมือชื่อนิสิต .....
สาขาวิชา	การวัดและประเมินผลการศึกษา	ลายมือชื่อ อ.ที่ปรึกษาหลัก .....
ปีการศึกษา	2560	ลายมือชื่อ อ.ที่ปรึกษาร่วม .....



## กิตติกรรมประกาศ

การทำวิทยานิพนธ์ฉบับนี้ ผู้วิจัยได้รับความรู้และคำแนะนำต่าง ๆ ด้วยความกรุณาอย่างยิ่งจากอาจารย์ที่ปรึกษาวิทยานิพนธ์ คือ รองศาสตราจารย์ ดร. ศิริเดช สุชีวะ ที่ได้ให้คำแนะนำในการตัดสินใจเลือกหัวข้อวิทยานิพนธ์ รวมถึงคำแนะนำในการทำวิทยานิพนธ์ฉบับนี้ และอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม คือ อาจารย์ ดร. ลีระโชติ ศรีสุทธิยากร ที่ให้ความกรุณาแนะนำการเขียนคำสั่งในโปรแกรม R และให้คำปรึกษาเกี่ยวกับสมการและสถิติวิเคราะห์ที่ผู้วิจัยไม่เข้าใจหรือไม่แน่ใจ ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ด้วยความละเอียดถี่ถ้วนจนงานวิจัยฉบับนี้เสร็จสมบูรณ์

ขอกราบขอบพระคุณ ศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี ประธานกรรมการสอบวิทยานิพนธ์ รองศาสตราจารย์ ดร. โชติกา ภาชีผล ผู้ช่วยศาสตราจารย์ณัฐภรณ์ หลาวทอง และผู้ช่วยศาสตราจารย์สังวรณ์ ังตกระโทก กรรมการสอบวิทยานิพนธ์ ที่ได้สละเวลาอันมีค่ามาร่วมสอบวิทยานิพนธ์ และให้ข้อเสนอแนะที่เป็นประโยชน์ในการทำวิทยานิพนธ์ฉบับนี้

ขอกราบขอบพระคุณอาจารย์ ดร. ชยุตม์ ภิรมย์สมบัติ ที่ได้สละเวลาให้คำปรึกษาเกี่ยวกับการปรับเทียบคะแนน รวมถึงคำสั่งที่ใช้ในการจำลองข้อมูล ซึ่งทำให้ผู้วิจัยเกิดความมั่นใจในการดำเนินการและแก้ไขข้อบกพร่องต่าง ๆ ให้ถูกต้อง และอาจารย์ ดร.กนิษฐ์ ศรีเคลือบ ที่ให้ความกรุณาในการพิจารณาบทความเพื่อตีพิมพ์เผยแพร่ในวารสารวิธีวิทยาการวิจัย

ขอกราบขอบพระคุณคณาจารย์ภาควิชาวิจัยและจิตวิทยาการศึกษา ที่ได้ให้ความรู้ความห่วงใย แก่นิสิตด้วยความเมตตา

ขอขอบคุณเพื่อน ๆ สาขาวิชาการวัดและประเมินผลการศึกษาทุกคน ที่ให้กำลังใจตลอดช่วงเวลาของการเรียนและการทำวิทยานิพนธ์จนสำเร็จลุล่วงไปด้วยดี

ท้ายที่สุด ขอกราบขอบพระคุณ บิดา มารดา สำหรับความรัก ความห่วงใย ความเอาใจใส่ และกำลังใจที่มีผู้ให้วิจัยเสมอมา วิทยานิพนธ์ครั้งนี้สำเร็จลุล่วงได้เพราะมีบิดา มารดาเป็นแรงจูงใจ และแรงผลักดันสำคัญในการก้าวผ่านอุปสรรคที่เกิดขึ้นระหว่างการทำวิทยานิพนธ์ คุณค่าและประโยชน์ที่มีต่อวิทยานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบให้แก่ท่านทั้งสอง



## สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ง
บทคัดย่อภาษาอังกฤษ.....	จ
กิตติกรรมประกาศ.....	ฉ
สารบัญ.....	ช
สารบัญตาราง.....	ฐ
สารบัญภาพ .....	ฌ
บทที่ 1 บทนำ .....	1
ความเป็นมาและความสำคัญของปัญหา .....	1
คำถามการวิจัย .....	7
วัตถุประสงค์การวิจัย .....	7
สมมติฐานการวิจัย .....	8
ข้อตกลงเบื้องต้น.....	12
ขอบเขตของการวิจัย.....	12
คำจำกัดความที่ใช้ในการวิจัย.....	14
ประโยชน์ที่คาดว่าจะได้รับ.....	17
บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง .....	18
ตอนที่ 1 ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ .....	19
1.1 มโนทัศน์และประเภทของโมเดลการตอบสนองข้อสอบแบบพหุมิติ .....	19
1.2 ค่าพารามิเตอร์และฟังก์ชันสารสนเทศของข้อสอบและแบบสอบ .....	28
1.3 การประมาณค่าพารามิเตอร์ของข้อสอบ-ผู้สอบ และการแปลงค่าพารามิเตอร์ไประบบพิกัด.....	31
1.4 โครงสร้างของแบบสอบพหุมิติ .....	42
1.5 โปรแกรมที่ใช้จำลองข้อมูลพหุมิติและการกำหนดค่าพารามิเตอร์ของข้อสอบ .....	48

ตอนที่ 2 การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ .....	54
2.1 มโนทัศน์เกี่ยวกับการปรับเทียบคะแนน .....	54
2.2 การออกแบบการเก็บรวบรวมข้อมูลสำหรับการปรับเทียบคะแนน .....	55
2.3 กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ .....	61
2.4 กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองแบบพหุมิติ.....	63
2.5 การประเมินประสิทธิภาพของการปรับเทียบคะแนน.....	76
2.6 โปรแกรมที่ใช้ปรับเทียบคะแนนภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ ข้อสอบร่วม .....	81
2.7 งานวิจัยเกี่ยวกับกระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบ พหุมิติ .....	82
ตอนที่ 3 การเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ .....	90
3.1 มโนทัศน์เกี่ยวกับการเชื่อมโยงสเกล .....	90
3.2 แนวทางการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบ .....	91
3.3 งานวิจัยเกี่ยวกับแนวทางการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบ .....	92
3.4 การเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ .....	97
3.5 งานวิจัยที่เกี่ยวข้องกับการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองแบบพหุมิติ .....	104
ตอนที่ 4 สังเคราะห์งานวิจัยเกี่ยวกับการปรับเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล MIRT และกรอบแนวคิดการวิจัย .....	118
4.1 สังเคราะห์งานวิจัยเกี่ยวกับการปรับเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล MIRT.....	118
4.2 กรอบแนวคิดการวิจัย .....	124
บทที่ 3 วิธีดำเนินการวิจัย.....	125
ตอนที่ 1 การจำลองข้อมูล .....	126
1.1 การจำลองค่าความสามารถของผู้สอบ .....	126

1.2 การจำลองค่าพารามิเตอร์ของข้อสอบ .....	127
1.3 การจำลองความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง.....	134
1.4 การจำลองผลการตอบข้อสอบ .....	134
ตอนที่ 2 การตรวจสอบความถูกต้องของการจำลองข้อมูล.....	135
2.1 การตรวจสอบความถูกต้องของค่ามุมที่จำลองขึ้น .....	135
2.2 การตรวจสอบความถูกต้องของค่าความสามารถของผู้สอบที่จำลองขึ้น .....	138
ตอนที่ 3 การวัดประสิทธิภาพของการปรับเทียบคะแนน .....	142
3.1 ค่าความลำเอียงสัมพัทธ์ .....	142
3.2 สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน .....	143
3.3 รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง.....	143
3.4 ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ .....	143
ตอนที่ 4 การวิเคราะห์ประสิทธิภาพของการปรับเทียบคะแนน .....	144
ตอนที่ 5 ขั้นตอนการดำเนินงานวิจัย.....	144
5.1 การจำลองค่าพารามิเตอร์ของข้อสอบและแบบแผนการตอบข้อสอบ .....	144
5.2 การประมาณค่าพารามิเตอร์ข้อสอบ .....	145
5.3 การเชื่อมโยงสเกลระหว่างแบบสอบ.....	145
5.4 การปรับเทียบคะแนนระหว่างแบบสอบ.....	146
บทที่ 4 ผลการวิเคราะห์ข้อมูล.....	148
ตอนที่ 1 ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพของการปรับเทียบคะแนน .....	149
1.1 ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพโดยรวม (RMSE).....	149
1.2 ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อความแม่นยำในการปรับเทียบคะแนน (RB) และความถูกต้องในการปรับเทียบคะแนน (CVSE).....	150
ตอนที่ 2 ประสิทธิภาพระหว่างกระบวนการ CMOSE กับกระบวนการ SMOSE.....	156

2.1 ผลด้านประสิทธิภาพโดยรวม (RMSE).....	156
2.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB) .....	159
2.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE).....	162
ตอนที่ 3 ประสิทธิภาพของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบร่วมแตกต่างกัน ....	164
3.1 ผลด้านประสิทธิภาพโดยรวม (RMSE).....	164
3.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB) .....	167
3.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE).....	169
ตอนที่ 4 ประสิทธิภาพของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้ คะแนนสองค่าแตกต่างกัน .....	170
4.1 ผลด้านประสิทธิภาพโดยรวม (RMSE).....	170
4.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB) .....	173
4.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE).....	175
ตอนที่ 5 ประสิทธิภาพของกระบวนการ MOSE เมื่อโครงสร้างมิติความสามารถแตกต่างกัน ....	177
5.1 ผลด้านประสิทธิภาพโดยรวม (RMSE).....	177
5.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB) .....	178
5.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE).....	180
ตอนที่ 6 สรุปประสิทธิภาพของการปรับเทียบคะแนนแยกตามวัตถุประสงค์.....	181
ตอนที่ 7 ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ .....	182
บทที่ 5 สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ .....	184
สรุปผลการวิจัย.....	185
1. ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพของการปรับเทียบคะแนน	185
2. ประสิทธิภาพระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE .....	185
2.1 ผลด้านประสิทธิภาพโดยรวมของการปรับเทียบคะแนน (RMSE).....	185

2.2	ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB).....	185
2.3	ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE).....	185
3.	ประสิทธิภาพของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบรวมแตกต่างกัน.....	186
3.1	ผลด้านประสิทธิภาพโดยรวมของการปรับเทียบคะแนน (RMSE).....	186
3.2	ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB).....	186
3.3	ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE).....	186
4.	ประสิทธิภาพของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้ คะแนนสองค่าแตกต่างกัน.....	187
4.1	ผลด้านประสิทธิภาพโดยรวมของการปรับเทียบคะแนน.....	187
4.2	ผลด้านความแม่นยำในการปรับเทียบคะแนน.....	187
4.3	ผลด้านความถูกต้องในการปรับเทียบคะแนน.....	187
5.	ประสิทธิภาพของกระบวนการ MOSE เมื่อโครงสร้างมิติความสามารถแตกต่างกัน.....	188
5.1	ผลด้านประสิทธิภาพโดยรวมของการปรับเทียบคะแนน.....	188
5.2	ผลด้านความแม่นยำในการปรับเทียบคะแนน.....	188
5.3	ผลด้านความถูกต้องในการปรับเทียบคะแนน.....	188
อภิปรายผล.....		189
1.	ปฏิสัมพันธ์ระหว่างตัวแปรอิสระที่ส่งผลต่อประสิทธิภาพในการปรับเทียบคะแนน ...	189
2.	ประสิทธิภาพระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE.....	190
3.	ประสิทธิภาพของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบรวมแตกต่างกัน .	192
4.	ประสิทธิภาพของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้ คะแนน สองค่าแตกต่างกัน.....	195
5.	ประสิทธิภาพของกระบวนการ MOSE เมื่อโครงสร้างมิติความสามารถแตกต่างกัน .	197
6.	อภิปรายเพิ่มเติมในประเด็นที่น่าสนใจ.....	199
ข้อเสนอแนะ.....		201

1. ข้อเสนอแนะในการนำไปปฏิบัติ .....	201
2. ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป .....	202
3. ข้อเสนอแนะเชิงทฤษฎี.....	203
รายการอ้างอิง .....	204
ภาคผนวก ก ค่าความยากรวม (MDIFF) ค่าอำนาจจำแนกรวม (MDISC) และค่ามุมระหว่าง เวกเตอร์ในแต่ละเงื่อนไข .....	211
ภาคผนวก ข ผลการตรวจสอบความถูกต้องของค่ามุมที่จำลองขึ้น .....	234
ภาคผนวก ค ผลการตรวจสอบความถูกต้องของค่าความสามารถที่จำลองขึ้น .....	243
ภาคผนวก ง ผลการประเมินประสิทธิภาพโดยรวมของการเชื่อมโยงสเกล.....	246
ภาคผนวก จ ผลของประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบในภาพรวม .....	250
ภาคผนวก ฉ ผลของประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบในแต่ละเงื่อนไข .....	254
ประวัติผู้เขียนวิทยานิพนธ์ .....	266

## สารบัญตาราง

ตาราง 1.1 ตัวแปรตามและตัวแปรอิสระที่ใช้ในการวิจัย .....	12
ตาราง 2.1 ความแตกต่างระหว่าง Compensatory Models และ Partially Compensatory Models .....	25
ตาราง 2.2 สรุปค่าพารามิเตอร์ในการจำลองข้อมูลแบบพหุมิติที่ใช้ในงานวิจัย .....	51
ตาราง 2.3 การกำหนดค่าพารามิเตอร์ในการจำลองข้อมูลแบบพหุมิติสำหรับงานวิจัยเกี่ยวกับการเปรียบเทียบคะแนนและการเชื่อมโยงสเกลสำหรับแบบสอบพหุมิติ .....	52
ตาราง 2.4 การเปรียบเทียบการออกแบบการเก็บรวบรวมข้อมูลแต่ละรูปแบบในปัจจุบันต่าง ๆ .....	61
ตาราง 2.5 เปรียบเทียบสมการที่ใช้ในกระบวนการ MOSE ของข้อสอบทั้ง 3 รูปแบบ .....	68
ตาราง 2.6 สรุปขั้นตอนของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติและพหุมิติ .....	73
ตาราง 2.7 เปรียบเทียบกระบวนการปรับเทียบคะแนนสังเกตได้ด้วยโมเดล UIRT และโมเดล MIRT .....	75
ตาราง 2.8 สรุปประเด็นของงานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT .....	88
ตาราง 2.9 สรุปประเด็นสำคัญของงานวิจัยเกี่ยวกับประสิทธิภาพของวิธีการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ ภายใต้รูปแบบ NEAT .....	96
ตาราง 2.10 สรุปเงื่อนไขของวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT หลังจากการประมาณค่าพารามิเตอร์ .....	103
ตาราง 2.11 สรุปประเด็นสำคัญของงานวิจัยการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT .....	109
ตาราง 2.12 สรุปประเด็นของงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า .....	116
ตาราง 2.13 สรุปตัวแปรที่ศึกษาของงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า .....	117
ตาราง 2.14 จำนวนการทำซ้ำในงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลและปรับเทียบคะแนนด้วยโมเดล MIRT	123

ตาราง 3.1 การกระจายค่าความสามารถและความสัมพันธ์ระหว่างมิติสำหรับผู้สอบแต่ละกลุ่ม .....	127
ตาราง 3.2 อัตราส่วนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแยกตามสัดส่วนคะแนนข้อสอบร่วม.....	128
ตาราง 3.3 สัดส่วนคะแนนข้อสอบร่วมและข้อสอบเฉพาะในแต่ละเงื่อนไข สำหรับโครงสร้างที่ไม่ซับซ้อน.....	129
ตาราง 3.4 สัดส่วนคะแนนข้อสอบร่วมและข้อสอบเฉพาะในแต่ละเงื่อนไข สำหรับโครงสร้างที่ซับซ้อน.	131
ตาราง 3.5 การกระจายค่าความยากรวมและค่าอำนาจจำแนกรวมใน 5 ระดับ .....	133
ตาราง 3.6 ขนาดของมุมเมื่อวัดจากมิติที่ 1 และมิติที่ 2 สำหรับโครงสร้าง APSS และ CS .....	133
ตาราง 3.7 ผลการทดสอบความถูกต้องของค่าความสามารถสำหรับผู้สอบทั้งสองกลุ่ม.....	141
ตาราง 4.1 ผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ .....	150
ตาราง 4.2 ผลการทดสอบความแปรปรวนพหุนาม.....	152
ตาราง 4.3 ผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ .....	155
ตาราง 4.4 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของประสิทธิภาพโดยรวมของการปรับเทียบคะแนน	158
ตาราง 4.5 ค่าสถิติบรรยายของความแม่นยำและความถูกต้องในการปรับเทียบคะแนน .....	161
ตาราง 4.6 ผลวิเคราะห์อิทธิพลอย่างง่ายของตัวแปรสัดส่วนข้อสอบร่วม .....	166
ตาราง 4.7 ค่าสถิติบรรยายของประสิทธิภาพโดยรวมของการปรับเทียบคะแนน .....	167
ตาราง 4.8 ผลวิเคราะห์อิทธิพลอย่างง่ายของอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า ..	173
ตาราง 4.9 ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบในภาพรวม.....	183
ตาราง 5.1 ความแตกต่างระหว่างค่ามุมในแต่ละกลุ่ม สำหรับสัดส่วนคะแนนข้อสอบร่วม.....	195



## สารบัญภาพ

ภาพ 1.1 โครงสร้างที่ซับซ้อน (CS).....	13
ภาพ 1.2 โครงสร้างที่ไม่ซับซ้อน (APSS).....	13
ภาพ 2.1 ตัวอย่างพื้นผิวของการตอบสนองข้อสอบสำหรับโมเดล MGPC (Reckase, 2009).....	25
ภาพ 2.2 พื้นผิวคะแนนที่คาดหวังสำหรับข้อสอบในภาพที่ 2.1 (Reckase, 2009).....	26
ภาพ 2.3 การนำเสนอค่าพารามิเตอร์ข้อสอบจำนวน 45 ข้อในพื้นที่ 3 มิติ (ค่า $\theta_1$ , $\theta_2$ และ $\theta_3$ ).....	29
ภาพ 2.4 การนำเสนอค่าพารามิเตอร์ข้อสอบจำนวน 2 ข้อในพื้นที่ 2 มิติ (ค่า $\theta_1$ และ $\theta_2$ ).....	29
ภาพ 2.5 พื้นผิวสารสนเทศสำหรับโมเดล M2PL.....	30
ภาพ 2.6 พื้นผิวโค้งคุณลักษณะของแบบสอบ.....	31
ภาพ 2.7 โครงสร้างอย่างง่ายของแบบสอบ 2 มิติ.....	43
ภาพ 2.8 โครงสร้างที่ไม่ซับซ้อนของแบบสอบ 2 มิติ.....	44
ภาพ 2.9 เวกเตอร์ข้อสอบ (item vectors) สำหรับโครงสร้างที่ไม่ซับซ้อน.....	44
ภาพ 2.10 โครงสร้างที่ซับซ้อนของแบบสอบ 5 มิติ.....	45
ภาพ 2.11 เวกเตอร์ข้อสอบ (item vectors) สำหรับโครงสร้างซับซ้อน.....	45
ภาพ 2.12 การออกแบบการเก็บรวบรวมข้อมูลสำหรับการปรับเทียบคะแนน.....	57
ภาพ 2.13 กรอบแนวคิดในการวิจัย.....	124
ภาพ 3.1 ขั้นตอนการจำลองข้อมูล.....	126
ภาพ 3.2 ตัวอย่างกราฟแสดงเวกเตอร์ของแบบสอบที่มีโครงสร้างที่ไม่ซับซ้อน.....	136
ภาพ 3.3 ตัวอย่างกราฟแสดงเวกเตอร์ของแบบสอบที่มีโครงสร้างที่ซับซ้อน.....	138
ภาพ 3.4 ฮิสโทแกรมแสดงความสามารถ มิติที่ 1 ระหว่างผู้สอบทั้งสองกลุ่ม.....	141
ภาพ 3.5 เส้นโค้งปกติแสดงความสามารถ มิติที่ 1 ระหว่างผู้สอบทั้งสองกลุ่ม.....	141
ภาพ 3.6 เส้นโค้งปกติแสดงความสามารถ มิติที่ 2 ระหว่างผู้สอบทั้งสองกลุ่ม.....	141
ภาพ 3.7 ฮิสโทแกรมแสดงความสามารถมิติที่ 2 ระหว่างผู้สอบทั้งสองกลุ่ม.....	141

ภาพ 4.1 ค่า RMSE ของกระบวนการ MOSE สำหรับโครงสร้าง APSS และ CS.....	157
ภาพ 4.2 แผนภูมิ Box plot ค่า RMSE ของกระบวนการ CMOSE โครงสร้าง APSS และ CS.....	157
ภาพ 4.3 แผนภูมิ Box plot ค่า RMSE ของกระบวนการ SMOSE โครงสร้าง APSS และ CS.....	157
ภาพ 4.4 ค่า RB ของกระบวนการ MOSE สำหรับโครงสร้าง APSS และ CS .....	160
ภาพ 4.5 แผนภูมิ Box plot ค่า RB ของกระบวนการ CMOSE โครงสร้าง APSS และ CS .....	160
ภาพ 4.6 แผนภูมิ Box plot ค่า RB ของกระบวนการ SMOSE โครงสร้าง APSS และ CS .....	160
ภาพ 4.7 ค่า CVSE ของกระบวนการ MOSE สำหรับโครงสร้าง APSS และ CS.....	163
ภาพ 4.8 แผนภูมิ Box plot ค่า CVSE ของกระบวนการ CMOSE โครงสร้าง APSS และ CS.....	163
ภาพ 4.9 แผนภูมิ Box plot ค่า CVSE ของกระบวนการ SMOSE โครงสร้าง APSS และ CS.....	164
ภาพ 4.10 แผนภูมิแท่งเปรียบเทียบค่า RMSE สำหรับสัดส่วนคะแนนข้อสอบรวม .....	165
ภาพ 4.11 ปฏิสัมพันธ์อิทธิพลระหว่าง ANCHOR X DICHO ต่อค่า RMSE.....	166
ภาพ 4.12 แผนภูมิแท่งเปรียบเทียบค่า RB สำหรับสัดส่วนคะแนนข้อสอบรวม .....	168
ภาพ 4.13 แผนภูมิแท่งเปรียบเทียบค่า CVSE สำหรับสัดส่วนคะแนนข้อสอบรวม .....	170
ภาพ 4.14 แผนภูมิแท่งเปรียบเทียบค่า RMSE สำหรับสัดส่วนคะแนนข้อสอบรวม .....	171
ภาพ 4.15 ปฏิสัมพันธ์อิทธิพลของ ANCHOR X DICHO ต่อค่า RMSE.....	172
ภาพ 4.16 แผนภูมิแท่งเปรียบเทียบค่า RB สำหรับอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้ คะแนนสองค่า.....	174
ภาพ 4.17 แผนภูมิแท่งเปรียบเทียบค่า CVSE สำหรับอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้ คะแนนสองค่า.....	176
ภาพ 4.18 กราฟแท่งเปรียบเทียบค่า RMSE สำหรับโครงสร้างมิติความสามารถ .....	178
ภาพ 4.19 แผนภูมิแท่งเปรียบเทียบค่า RB สำหรับโครงสร้างมิติความสามารถ .....	179
ภาพ 4.20 แผนภูมิแท่งเปรียบเทียบค่า CVSE สำหรับโครงสร้างมิติความสามารถ.....	181
ภาพ 4.21 ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ.....	182

## บทที่ 1

### บทนำ

#### ความเป็นมาและความสำคัญของปัญหา

สถานการณ์การทดสอบปัจจุบันมีแนวโน้มใช้แบบสอบหลายฉบับในการวัดครั้งเดียวหรือใช้แบบสอบหลายฉบับประเมินการเรียนรู้ในเวลาที่แตกต่างกัน เนื่องด้วยเหตุผลของความปลอดภัยในการทดสอบ แบบสอบหลายฉบับเหล่านั้นถูกพัฒนาให้มีลักษณะเนื้อหาหรือโครงสร้างเป็นคู่ขนานกัน แต่ก็ไม่สามารถรับประกันได้ว่าความยากของแบบสอบแต่ละฉบับเท่ากัน ดังนั้นหากต้องการเปรียบเทียบคะแนนระหว่างแบบสอบเหล่านั้นต้องอาศัยการปรับเทียบ (equating) ซึ่งเป็นกระบวนการทางสถิติที่ใช้ในการปรับเทียบคะแนนระหว่างแบบสอบที่มุ่งวัดในเนื้อหาหรือโครงสร้างเดียวกัน โดยแปลงคะแนนจากแบบสอบฉบับหนึ่งให้อยู่ในสเกลของแบบสอบอีกฉบับหนึ่ง เพื่อให้คะแนนที่ปรับเทียบจากแบบสอบเหล่านั้นสามารถใช้เปรียบเทียบหรือใช้แทนกันได้ ซึ่งการปรับเทียบคะแนนมีข้อตกลงที่เข้มงวด เช่น แบบสอบที่นำมาปรับเทียบคะแนนต้องมีโครงสร้างและความเที่ยงเท่าเทียมกัน และผลการปรับเทียบคะแนนระหว่างแบบสอบทั้งสองฉบับมีความสมมาตร เป็นต้น โดยการปรับเทียบคะแนนระหว่างแบบสอบที่มีประสิทธิภาพอาศัยการออกแบบการปรับเทียบคะแนน 2 กระบวนการสำคัญ ได้แก่ การออกแบบวิธีการเก็บรวบรวมข้อมูล และการออกแบบวิธีการปรับเทียบคะแนน

การออกแบบวิธีการเก็บรวบรวมข้อมูลมี 3 รูปแบบหลัก ได้แก่ (1) รูปแบบผู้สอบกลุ่มสุ่ม (random group) (2) รูปแบบผู้สอบกลุ่มเดียว (single group) และ (3) รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม (non-equivalent group with anchor test: NEAT) โดยการปรับเทียบคะแนนนิยมใช้รูปแบบ NEAT เนื่องจากผู้สอบแต่ละกลุ่มทำแบบสอบเพียงฉบับเดียว ซึ่งแบบสอบแต่ละฉบับมีข้อสอบร่วมกันจำนวนหนึ่ง ทำให้ผู้สอบแต่ละกลุ่มไม่จำเป็นต้องทำแบบสอบในช่วงเวลาเดียวกัน นอกจากนี้รูปแบบ NEAT ยังใช้ได้กับผู้สอบที่อยู่คนละระดับชั้นเพื่อพิจารณาพัฒนาการผลสัมฤทธิ์ทางการเรียนได้อีกด้วย (Kolen & Brennan, 2004; ศิริชัย กาญจนวาสี, 2555) โดยรายละเอียดของวิธีการเก็บรวบรวมข้อมูลทั้ง 3 แบบ ระบุไว้ในบทที่ 2 (ตอนที่ 2 หัวข้อ 2.2)

วิธีการปรับเทียบคะแนนมีหลายวิธี ไม่ว่าจะเป็น วิธีปรับเทียบค่าเฉลี่ย วิธีปรับเทียบเชิงเส้น วิธีปรับเทียบอิกวิเปอร์เซ็นไทล์ และวิธีการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ โดยการปรับเทียบคะแนนมักกระทำภายใต้กรอบแนวคิดของทฤษฎีการตอบสนองข้อสอบ เนื่องจากผลการปรับเทียบคะแนนมีความแม่นยำตลอดสเกลคะแนน และความยากของแบบสอบแต่ละฉบับสามารถแตกต่างกันได้มากกว่าวิธีการปรับเทียบเชิงเส้น (Reckase, 1985) การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบ่งได้เป็น 2 ประเภทหลักๆ ได้แก่ (1) การปรับเทียบคะแนน

ตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (UIRT equating) และ (2) การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT equating) การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบสำหรับแบบสอบเอกมิติ ใช้แปลงคะแนนระหว่างแบบสอบซึ่งมุ่งวัดคุณลักษณะเดียวโดยประสิทธิภาพในการปรับเทียบคะแนนด้วยโมเดล IRT ขึ้นอยู่กับการเป็นไปตามข้อตกลงเบื้องต้นที่ระบุว่าแบบสอบต้องวัดคุณลักษณะ/ความสามารถในมิติเดียว ดังนั้นหากปรับเทียบคะแนนแบบสอบพหุมิติด้วยด้วยโมเดล IRT ย่อมทำให้ผลการปรับเทียบคะแนนมีความคลาดเคลื่อนสูง จึงมีการพัฒนาการปรับเทียบคะแนนด้วยโมเดล MIRT สำหรับแบบสอบพหุมิติ ซึ่งเมื่อเปรียบเทียบประสิทธิภาพของการปรับเทียบคะแนนแบบสอบพหุมิติด้วยโมเดล UIRT และ MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่าภายใต้รูปแบบผู้สอบกลุ่มเท่าเทียม และแบบสอบรูปแบบผสมภายใต้รูปแบบผู้สอบกลุ่มสุ่ม พบว่า การปรับเทียบคะแนนด้วยโมเดล MIRT มีประสิทธิภาพมากกว่าการปรับเทียบคะแนนด้วยโมเดล UIRT (Bradley Grant Brossman, 2010; Bradley G. Brossman & Lee, 2013)

กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (Full MIRT score equating procedure: FMIRT) ประกอบด้วย 3 กระบวนการ ได้แก่ (1) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (Full MIRT observed score equating procedure: MOSE) (2) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูปโดยใช้การประมาณเอกมิติ (unidimensional approximation to MIRT observed score equating procedure: AOSE) และ (3) กระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูปโดยใช้การประมาณเอกมิติ (unidimensional approximation to MIRT true score equating procedure: ATSE) โดยงานวิจัยของ Brossman (2010) พบว่า กระบวนการ FMIRT ทั้งหมดมีประสิทธิภาพมากกว่าการปรับเทียบคะแนนด้วยโมเดล UIRT ส่วนงานวิจัยที่เปรียบเทียบประสิทธิภาพระหว่างกระบวนการ FMIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT เมื่อมีเชื่อมโยงสเกล 5 วิธี ได้แก่ วิธี OD วิธี TCF วิธี ICF วิธี M และวิธี NOP พบว่า กระบวนการ ATSE มีประสิทธิภาพมากที่สุดในทุกเงื่อนไขความสามารถของผู้สอบและวิธีการเชื่อมโยงสเกล และกระบวนการปรับเทียบคะแนน FMIRT ทั้ง 3 กระบวนการภายใต้การเชื่อมโยงสเกลด้วยวิธี TCF วิธี ICF และวิธี OD มีประสิทธิภาพดีกว่า วิธี M และ NOP (O. Zhang, 2012)

อย่างไรก็ตามงานวิจัยที่เปรียบเทียบประสิทธิภาพระหว่าง (1) กระบวนการ MOSE (2) กระบวนการ UIRT (3) วิธีอีควิเปอร์เซ็นไทล์ และ (4) วิธี Bi-factor equating สำหรับแบบสอบรูปแบบผสมที่ประกอบด้วยข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่า ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม พบว่า กระบวนการ MOSE มีประสิทธิภาพมากที่สุดสำหรับข้อสอบพหุมิติ (Peterson, 2014) สอดคล้องกับงานวิจัยที่ศึกษากับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม ซึ่งพบว่า การปรับเทียบคะแนนด้วยกระบวนการ MOSE มีความแม่นยำมากกว่ากระบวนการ AOSE,

ATSE และวิธีอีควิเปอร์เซ็นไทล์ โดยเฉพาะอย่างยิ่ง เมื่อความสัมพันธ์ระหว่างมิติมีค่าต่ำ (E. Lee, 2013) ด้วยเหตุนี้กระบวนการ MOSE จึงเหมาะสมสำหรับการปรับเทียบคะแนนแบบสอบพหุมิติ โดย งานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT มีข้อสังเกตสำคัญ คือ *งานวิจัยทั้งหมดมีการประมาณค่าพารามิเตอร์แบบแยกกัน* ซึ่งต้องอาศัยการเชื่อมโยงสเกลก่อนการปรับเทียบคะแนน อีกทั้งยังไม่มีศึกษาการปรับเทียบคะแนนด้วยโมเดล MIRT สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT

การเชื่อมโยงสเกล (scale linking) เป็นการแปลงสเกลจากแบบสอบฉบับหนึ่งไปยังแบบสอบอีกฉบับหนึ่ง โดยแบบสอบอาจมีเนื้อหา/โครงสร้างแตกต่างกัน ทำให้คะแนนที่ได้จากการเชื่อมโยงสเกลระหว่างแบบสอบไม่สามารถแลกเปลี่ยนกันได้ (Kolen & Brennan, 2004) การเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบ่งเป็น 2 ประเภท ได้แก่ การเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (UIRT scale linking) และการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT scale linking) แนวทางการเชื่อมโยงสเกลทั้งแบบเอกมิติและพหุมิติแบ่งเป็น 3 แนวทางหลัก ได้แก่ (1) การประมาณค่าพารามิเตอร์แบบพร้อมกัน (concurrent calibration) (2) การประมาณค่าพารามิเตอร์เมื่อกำหนดค่าพารามิเตอร์ของข้อสอบร่วม (fixed common item parameter calibration: FCIP) และ (3) การเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แบบแยกกัน (scale linking after separate calibration)

*การประมาณค่าพารามิเตอร์แบบพร้อมกัน* ข้อมูลการตอบของผู้สอบจากการทดสอบที่ต่างกันถูกนำมารวมกันและประมาณค่าด้วยการ run ในครั้งเดียว ค่าพารามิเตอร์ข้อสอบและค่าความสามารถย่อมอยู่ในเมทริกซ์เดียวกัน จึงไม่จำเป็นต้องใช้การเชื่อมโยงสเกล วิธี FCIP นำไปใช้ได้เมื่อทราบค่าพารามิเตอร์ของข้อสอบร่วมหรือค่าความสามารถของผู้สอบร่วม โดยวิธีนี้จะทำให้ค่าพารามิเตอร์จากการทดสอบต่าง ๆ ให้อยู่ในเมทริกซ์เดียวกัน จึงไม่จำเป็นต้องใช้การเชื่อมโยงสเกล (O. Zhang, 2012) ส่วนการเชื่อมโยงสเกลหลังจากการประมาณค่าพารามิเตอร์แบบแยกกัน ใช้เมื่อข้อมูลการตอบจากผู้สอบแต่ละกลุ่มหรือแบบสอบแต่ละฉบับ ถูกนำไปประมาณค่าพารามิเตอร์ของข้อสอบด้วยการ run แบบแยกกัน จึงต้องใช้การเชื่อมโยงสเกลเพื่อแปลงค่าพารามิเตอร์จากแบบสอบที่แตกต่างกันให้อยู่บนเมทริกซ์เดียวกัน

การพิจารณาว่าต้องใช้การเชื่อมโยงสเกลหรือไม่อาจพิจารณาความเท่าเทียมของผู้สอบ กล่าวคือ *ถ้าผู้สอบแต่ละกลุ่มมีความสามารถเท่าเทียมกัน* เมื่อทำการประมาณค่าพารามิเตอร์สำหรับผู้สอบแต่ละกลุ่มแยกกัน ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานสำหรับผู้สอบแต่ละกลุ่มถูกกำหนดให้เท่ากัน จึงไม่จำเป็นต้องใช้การเชื่อมโยงสเกล ขณะที่การประมาณค่าพารามิเตอร์แบบแยกกันสำหรับรูปแบบ NEAT จำเป็นต้องใช้การเชื่อมโยงสเกล เพื่อให้แน่ใจว่าค่าพารามิเตอร์ข้อสอบและค่าความสามารถของผู้สอบแต่ละกลุ่มที่อาจอยู่คนละระดับชั้นหรือทำแบบสอบในช่วงเวลาต่างกันอยู่

บนสเกลเดียวกัน แต่ถ้ารูปแบบ NEAT ที่มีการประมาณค่าพารามิเตอร์แบบพร้อมกัน ไม่จำเป็นต้องใช้การเชื่อมโยงสเกลก่อนการปรับเทียบคะแนน (Peterson, 2014; O. Zhang, 2012)

วิธีการเชื่อมโยงสเกลแบ่งตามมิติของแบบสอบได้ 2 แบบ ได้แก่ การเชื่อมโยงสเกลแบบเอกมิติ และแบบพหุมิติ โดยการเชื่อมโยงสเกลด้วยโมเดล UIRT มีหลายวิธี ได้แก่ วิธีค่าเฉลี่ย/ซิกมา วิธีค่าเฉลี่ย/ค่าเฉลี่ย วิธีของ Haebara และวิธีของ Stocking-Lord ส่วนการเชื่อมโยงสเกลด้วยโมเดล MIRT เป็นการแปลงเชิงเส้นเช่นเดียวกับการเชื่อมโยงสเกลด้วยโมเดล UIRT แต่การแปลงของการเชื่อมโยงสเกลด้วยโมเดล MIRT กระทำบนหลายมิติ การเชื่อมโยงสเกลด้วยโมเดล MIRT มีหลายวิธี ได้แก่ วิธีของ Hirsch วิธีของ Thompson, Nearing และ Davey วิธีของ Li และ Lisztz วิธีของ Min (M) วิธีของ Reckase และ Martineau (NOP method) วิธีของ Oshima, Davey และ Lee (ODL) โดยวิธี ODL ประกอบด้วย วิธี Direct (OD) วิธี Equated function procedure วิธี Test characteristic function (TCF) procedure และวิธี Item characteristic function (ICF) procedure งานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT พบว่า กระบวนการ FMIRT ภายใต้การเชื่อมโยงสเกลด้วยวิธี TCF วิธี ICF มีประสิทธิภาพดีกว่าวิธี M และ วิธี NOP (O. Zhang, 2012) สอดคล้องกับงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT ที่พบว่า การเชื่อมโยงสเกลสำหรับแบบสอบพหุมิติด้วยวิธี TCF สำหรับรูปแบบ NEAT ให้ผลการเชื่อมโยงสเกลที่ดี (Yao & Boughton, 2009; ศิวะทัศน์ สุขสุวรรณ, พัชรี จันทร์เพ็ญ, & สังวรรณ รัตตะโทก, 2555) ดังนั้นวิธี TCF จึงมีความเหมาะสมในการเชื่อมโยงสเกลระหว่างแบบสอบ เมื่อมีการประมาณค่าพารามิเตอร์แบบแยกกัน

อย่างไรก็ตามการประมาณค่าพารามิเตอร์แบบพร้อมกันมีประสิทธิภาพมากกว่าแบบแยกกัน เนื่องจากการประมาณค่าพารามิเตอร์แบบสอบสองฉบับพร้อมกัน ทำให้ตัวอย่างในการประมาณค่ามีขนาดใหญ่ ส่งผลให้คลาดเคลื่อนในการประมาณค่าพารามิเตอร์ต่ำลง อีกทั้งยังช่วยลดความคลาดเคลื่อนในการปรับเทียบคะแนนอันเนื่องจากการเชื่อมโยงสเกล (S.-H. Kim & Cohen, 1997; Seonghoon Kim & Kolen, 2006) สอดคล้องกับงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT ภายใต้รูปแบบ NEAT ที่พบว่า การประมาณค่าพารามิเตอร์พร้อมกันมีประสิทธิภาพมากกว่าการประมาณค่าแบบแยกกัน (Lin, 2008; Simon, 2008) นอกจากนี้ยังพบว่า การประมาณค่าพารามิเตอร์ของข้อมูลจริงมีความสอดคล้องกับโมเดลและปรากฏความเป็นพหุมิติ เมื่อมีการสเกลข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่าพร้อมกัน และมีการสเกลข้อมูลการตอบจากกลุ่มผู้สอบในระดับต่าง ๆ พร้อมกัน (Patz & Yao, 2007) และสอดคล้องกับงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล IRT สำหรับแบบสอบรูปแบบผสมที่พบว่า วิธีการปรับค่าพารามิเตอร์แบบพร้อมกันมีประสิทธิภาพดีกว่าการประมาณค่าพารามิเตอร์แบบแยกกัน เนื่องจากมีความคลาดเคลื่อนน้อยกว่า และมีความประสิทธิภาพในการกู้คืนค่าพารามิเตอร์มากกว่าวิธีการเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แบบแยกกัน (Seonghoon Kim & Kolen, 2006; Tian, 2011; อัญชลี ศรีกลชาญ, 2552) และเนื่องจาก

งานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT ยังไม่มีศึกษากับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT และใช้การประมาณค่าพารามิเตอร์แบบแยกกันเท่านั้น ด้วยเหตุนี้ผู้วิจัยจึงสนใจเปรียบเทียบประสิทธิภาพของกระบวนการ Full MIRT observed-score equating procedure: MOSE 2 รูปแบบ ได้แก่ กระบวนการ MOSE ที่มีการประมาณค่าพารามิเตอร์แบบพร้อมกัน และกระบวนการ MOSE ที่มีการประมาณค่าพารามิเตอร์แบบแยกกัน และใช้การเชื่อมโยงสเกลด้วยวิธี TCF สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบ (NEAT)

การปรับเทียบคะแนนแบบสอบพหุมิติรูปแบบผสมจำเป็นต้องพิจารณาโครงสร้างมิติความสามารถและความสัมพันธ์ระหว่างมิติ โครงสร้างมิติความสามารถมี 3 รูปแบบ ได้แก่ โครงสร้างอย่างง่าย (SS) โครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างซับซ้อน (CS) แต่โครงสร้างอย่างง่ายไม่สอดคล้องกับสภาพจริงของการทดสอบเพราะเป็นไปได้ยากที่ข้อสอบข้อหนึ่งจะอยู่ในมิติใดมิติหนึ่งเพียงมิติเดียว (Zhang & Stout, 1999 อ้างอิงใน พัชร จันทร์เพ็ง, 2550) การปรับเทียบคะแนนด้วยโมเดล MIRT ส่วนใหญ่ศึกษากับโครงสร้างที่ไม่ซับซ้อนและโครงสร้างซับซ้อนแบบ 2 มิติ ผู้วิจัยจึงสนใจที่จะศึกษาแบบสอบที่มีโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อน (CS) แบบ 2 มิติ เมื่อพิจารณาความสัมพันธ์ระหว่างมิติความสามารถ พบว่า งานวิจัยที่ผ่านมาศึกษาความสัมพันธ์ระหว่างมิติในระดับต่าง ๆ ได้แก่ ไม่มีความสัมพันธ์ มีความสัมพันธ์ในระดับปานกลาง และระดับสูง โดยงานวิจัยเกี่ยวกับการปรับเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล MIRT มีข้อค้นพบในทิศทางเดียวกัน คือ เมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้น การปรับเทียบคะแนนมีประสิทธิภาพลดลง เนื่องจากมีความลำเอียงและความคลาดเคลื่อนมาตรฐานเพิ่มขึ้น (E. Lee, 2013; W.-C. Lee & Brossman, 2012; Lin, 2008; Simon, 2008) จึงไม่จำเป็นต้องศึกษาในประเด็นดังกล่าวอีก (รายละเอียดของผลวิจัยดังกล่าวอยู่ในบทที่ 2 หัวข้อ 3.6) อย่างไรก็ตามเมื่อพิจารณางานวิจัยที่ใช้ข้อมูลจากการทดสอบจริง พบว่า แบบสอบส่วนใหญ่มีความสัมพันธ์ระหว่างมิติความสามารถตั้งแต่ระดับปานกลางไปจนถึงระดับสูง (Bradley Grant Brossman, 2010) และเนื่องจากผู้วิจัยไม่ได้สนใจศึกษาตัวแปรนี้จึงกำหนดให้ความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับปานกลาง คือ 0.5

การปรับเทียบคะแนนภายใต้รูปแบบ NEAT จำเป็นต้องพิจารณาปัจจัยด้านแบบสอบร่วม/ข้อสอบร่วม และเนื่องจากการสร้างแบบสอบแต่ละฉบับต้องใช้เวลาและค่าใช้จ่าย จึงต้องกำหนดความยาวข้อสอบร่วมที่เหมาะสม โดยสัดส่วนข้อสอบร่วมที่มีประสิทธิภาพในการปรับเทียบคะแนนสำหรับกลุ่มตัวอย่าง 400 คน คือ 10% และ 20% และถ้าข้อสอบร่วมมีคุณภาพสูง สัดส่วนข้อสอบร่วม 10% ก็มีความเพียงพอ (McKinley & Reckase, 1981; Wright 1974 cited in Gao, 2004) ส่วนการเชื่อมโยงสเกลด้วย MIRT พบว่า สัดส่วนข้อสอบร่วมที่มีความเหมาะสมสำหรับแบบสอบรูปแบบผสม ที่มีโครงสร้างมิติความสามารถแบบซับซ้อนด้วยวิธี TCF คือ 25% (ศิวัช ทัศน สุสุวรรณ

และคณะ 2555) ขณะที่สัดส่วนข้อสอบรวมน้อยที่สุดที่มีประสิทธิภาพสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า เมื่อความยาวของแบบสอบเป็น 160, 120, 100 และ 80 ข้อ คือ 20% และเมื่อความยาวแบบสอบเป็น 60 และ 40 ข้อ สัดส่วนข้อสอบรวมนที่มีประสิทธิภาพ คือ 25% และ 33% ตามลำดับ (Xiong, Ye, Ding, & Luo, 2012) อย่างไรก็ตามการปรับเทียบคะแนนด้วยโมเดล MIRT ภายใต้รูปแบบ NEAT เลือกใช้สัดส่วนคะแนนข้อสอบรวม 50% ซึ่งเป็นสัดส่วนจำนวนมาก นอกจากนี้การปรับเทียบคะแนนด้วยโมเดล MIRT ยังไม่มีศึกษาเกี่ยวกับสัดส่วนข้อสอบรวม ดังนั้นผู้วิจัยจึงสนใจศึกษาความยาวข้อสอบรวมที่สั้นลง เพื่อเพิ่มทางเลือกและประหยัดทรัพยากรในการสร้าง แบบสอบพหุมิติ สำหรับรูปแบบ NEAT ด้วยสัดส่วนคะแนนข้อสอบรวม 3 แบบ คือ 10%, 20% และ 30%

การปรับเทียบคะแนนแบบสอบรูปแบบผสมควรใช้แบบสอบรวมรูปแบบผสมเพราะการใช้ข้อสอบรวมที่ตรวจให้คะแนนสองค่าอย่างเดียวทำให้เกิดความคลาดเคลื่อนสูง ส่วนการใช้ข้อสอบรวมที่ตรวจให้คะแนนหลายค่ามักประสบปัญหาความเป็นปรนัยในการตรวจให้คะแนน (Sooyeon Kim, Walker, & McHale, 2010) และเมื่อใช้แบบสอบรวมรูปแบบผสมจึงต้องกำหนดสัดส่วนข้อสอบรวมแต่ละประเภท โดยงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลแบบสอบพหุมิติรูปแบบผสมมีการศึกษาอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40, 70:30 และ 80:20 (Patz & Yao, 2007; Yao & Boughton, 2009; ศิวะทัศน์ สุขสุวรรณ, 2555) ซึ่งจะสังเกตเห็นว่า ยังไม่มีการศึกษากรณีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนหลายค่ามากกว่าตรวจให้คะแนนสองค่า นอกจากนี้การปรับเทียบคะแนนแบบสอบพหุมิติรูปแบบผสม ภายใต้รูปแบบ NEAT ยังไม่มีการศึกษาในประเด็นดังกล่าว ผู้วิจัยจึงสนใจอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 4 แบบ คือ 40:60, 50:50, 60:40 และ 70:30

ผลจากการเปรียบเทียบประสิทธิภาพกระบวนการ MOSE 2 แบบ คือ กระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบพร้อมกัน (concurrent calibrated MOSE: CMOSE) และกระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบแยกกัน และใช้การเชื่อมโยงสเกลด้วยวิธี TCF (separated calibrated MOSE with Test characteristic function procedure: SMOSE) สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบรวม (NEAT) เมื่อโครงสร้างมิติตความสามารถ สัดส่วนคะแนนข้อสอบรวม และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าแตกต่างกัน ด้วยการจำลองข้อมูลแบบมอนติคาร์โล จะได้สารสนเทศอันเป็นประโยชน์ต่อการตัดสินใจเลือกกระบวนการปรับเทียบคะแนนแบบสอบพหุมิติที่มีประสิทธิภาพและสอดคล้องกับสถานการณ์ที่ต้องการศึกษามากที่สุด อีกทั้งยังเป็นการขยายองค์ความรู้เกี่ยวกับการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติอีกด้วย



## คำถามการวิจัย

การเปรียบเทียบประสิทธิภาพระหว่างกระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์พร้อมกัน (CMOSE) และกระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แยกกัน และใช้การเชื่อมโยงสเกลด้วยวิธี TCF (SMOSE) สำหรับแบบสอปรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วมในเงื่อนไขต่าง ๆ โดยพิจารณาจากค่าความลำเอียงสัมพัทธ์ (RB) สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (CVSE) และ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) มีคำถามการวิจัย ดังนี้

1. ประสิทธิภาพในการปรับเทียบคะแนนระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE กระบวนการใดมีประสิทธิภาพมากกว่า
2. ประสิทธิภาพของกระบวนการ MOSE ทั้ง 2 แบบจะเป็นอย่างไร เมื่อสัดส่วนคะแนนข้อสอบร่วมเพิ่มขึ้น
3. ประสิทธิภาพของกระบวนการ MOSE ทั้ง 2 แบบจะเป็นอย่างไร เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าเพิ่มขึ้น
4. ประสิทธิภาพของกระบวนการ MOSE ทั้ง 2 แบบ ภายใต้โครงสร้างมิตិความสามารถแบบไม่ซับซ้อน (APSS) และโครงสร้างมิติความสามารถแบบซับซ้อน (CS) แตกต่างกันหรือไม่

## วัตถุประสงค์การวิจัย

การวิจัยครั้งนี้ใช้ข้อมูลจำลองด้วยเทคนิคมอนติคาร์โล สำหรับแบบสอปรูปแบบผสม ภายใต้รูปแบบ NEAT โดยมีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE เมื่อโครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบร่วม และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า แตกต่างกัน ด้วยการจำลองข้อมูลแบบมอนติคาร์โล มีวัตถุประสงค์การวิจัยเพื่อ

1. เปรียบเทียบประสิทธิภาพระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE
2. ศึกษาประสิทธิภาพของกระบวนการ MOSE ทั้ง 2 แบบ เมื่อสัดส่วนคะแนนข้อสอบร่วมแตกต่างกัน
3. ศึกษาประสิทธิภาพของกระบวนการ MOSE ทั้ง 2 แบบ เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแตกต่างกัน
4. ศึกษาประสิทธิภาพของกระบวนการ MOSE ทั้ง 2 แบบ ภายใต้โครงสร้างมิติความสามารถที่ไม่ซับซ้อน (APSS) และโครงสร้างมิติความสามารถที่ซับซ้อน (CS)

## สมมติฐานการวิจัย

การประมาณค่าพารามิเตอร์จากแบบสอป 2 ฉบับพร้อมกัน ทำให้ตัวอย่างในการประมาณค่าพารามิเตอร์มีขนาดใหญ่ ส่งผลให้ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ต่ำลง อีกทั้งยังช่วยลดความคลาดเคลื่อนในการปรับเทียบคะแนนอันเนื่องจากการเชื่อมโยงสเกล ขณะที่การประมาณค่าพารามิเตอร์แบบแยกกันต้องใช้การเชื่อมโยงสเกล เพื่อนำค่าพารามิเตอร์ที่ประมาณได้ไว้บนสเกลของข้อสอบร่วม ทำให้เกิดความคลาดเคลื่อนอย่างสุ่ม (random error) เนื่องจากการสุ่มตัวอย่างและการประมาณค่าสัมประสิทธิ์ของการเชื่อมโยงสเกล ส่งผลให้ค่าพารามิเตอร์หลังจากการเชื่อมโยงมีความแปรปรวน (S.-H. Kim & Cohen, 1997; Seonghoon Kim & Kolen, 2006; Meng, 2007)

งานวิจัยที่เปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลด้วยโมเดล IRT ระหว่างวิธีประมาณค่าพารามิเตอร์แบบพร้อมกันและวิธีประมาณค่าพารามิเตอร์แบบแยกกัน (ได้แก่ วิธีค่าเฉลี่ย/ค่าเฉลี่ย วิธีค่าเฉลี่ย/ซิกมา วิธีของ Heabara และวิธีของ Stocking-Lord) สำหรับแบบสอปรูปแบบผสมภายใต้รูปแบบ NEAT พบว่า วิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันมีความคลาดเคลื่อนต่ำกว่าวิธีประมาณค่าพารามิเตอร์แบบแยกกัน (Seonghoon Kim & Kolen, 2006; อัญชลี ศรีภักขาน, 2552) เช่นเดียวกับงานวิจัยที่เปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลด้วยโมเดล IRT ระหว่างวิธีประมาณค่าพารามิเตอร์แบบพร้อมกัน และวิธีประมาณค่าพารามิเตอร์แบบแยกกันและเชื่อมโยงสเกลด้วยวิธีของ Stocking-Lord ที่พบว่า การประมาณค่าพารามิเตอร์แบบพร้อมกันมีค่า RMSD ต่ำกว่าการประมาณค่าพารามิเตอร์แบบแยกกันและเชื่อมโยงสเกลด้วยวิธีของ Stocking-Lord เนื่องจากค่าความสามารถของผู้สอบจากการประมาณค่าพารามิเตอร์แบบพร้อมกันมีความใกล้เคียงกับความสามารถที่แท้จริงของผู้สอบมากกว่าการประมาณค่าพารามิเตอร์แบบแยกกัน (Tian, 2011) สอดคล้องกับผลการวิจัยสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT ที่พบว่า ผลการเชื่อมโยงสเกลของวิธีประมาณค่าพารามิเตอร์แบบพร้อมกันมีความเสถียรมากกว่าวิธีของ Stocking-Lord และการประมาณค่าพารามิเตอร์แบบพร้อมกันมีความคลาดเคลื่อนต่ำกว่าการประมาณค่าพารามิเตอร์แบบแยกกัน (Bradley A. Hanson & Anton A. Béguin, 2002; Keller & Keller, 2015) นอกจากนี้ งานวิจัยที่ศึกษาการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า พบว่า ค่าอำนาจจำแนกจากการประมาณค่าพารามิเตอร์พร้อมกันมีค่า RMSE และค่า Bias ต่ำกว่าการประมาณค่าพารามิเตอร์แบบแยกกัน แต่ค่าความยากจากวิธีการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธี ICF มีค่า RMSE และค่า Bias ต่ำกว่าการประมาณค่าพารามิเตอร์แบบพร้อมกันและการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธีการเชื่อมโยงสเกลอื่น ๆ (วิธี TCF, วิธีของ Min และ วิธี Direct) สำหรับรูปแบบผู้สอบกลุ่มเท่าเทียมและรูปแบบ NEAT (Simon, 2008) ดังนั้นผู้วิจัยจึงกำหนดสมมติฐานได้ดังนี้

## สมมติฐานที่ 1 กระบวนการ CMOSE น่าจะมีประสิทธิภาพมากกว่ากระบวนการ SMOSE เมื่อพิจารณาจากค่าความลำเอียงสัมพัทธ์ (RB) และค่า RMSE

Angoff และคณะ (1971) แนะนำว่า ถ้าแบบสอบมีข้อสอบ 40 ข้อ ควรใช้จำนวนข้อสอบรวมอย่างต่ำร้อยละ 20 ของข้อสอบทั้งหมด และถ้าแบบสอบมีข้อสอบมากกว่า 40 ข้อ ควรใช้ข้อสอบรวมอย่างต่ำร้อยละ 30 ของข้อสอบทั้งหมด ขณะที่การปรับเทียบคะแนนส่วนใหญ่ใช้จำนวนข้อสอบรวมประมาณ 20-25% ของแบบสอบทั้งฉบับ (Kolen & Brennan, 2004) การเชื่อมโยงสเกลด้วยวิธี NOP สำหรับแบบสอบรูปแบบผสมที่มีโครงสร้าง APSS พบว่า การเชื่อมโยงสเกลด้วยวิธี NOP โดยใช้สัดส่วนน้ำหนักคะแนนข้อสอบรวม 25% มีค่า Bias และค่า RMSE น้อยที่สุดเมื่อเทียบกับสัดส่วนคะแนน 20% และ 30% (Sukswana, Junpengb, Ngudgratokec, & Guayjarempanishk, 2012) และการเชื่อมโยงสเกลด้วยวิธี TCF สำหรับแบบสอบรูปแบบผสมที่มีโครงสร้าง APSS และ CS เมื่อสัดส่วนน้ำหนักคะแนนข้อสอบรวมแตกต่างกัน คือ 20% 25% และ 30% พบว่า สัดส่วนน้ำหนักคะแนนข้อสอบรวมและโครงสร้างมิตีความสามารถไม่ร่วมกันส่งผลต่อประสิทธิภาพของการเชื่อมโยงสเกล แต่สัดส่วนน้ำหนักคะแนนข้อสอบรวมและโครงสร้างมิตีความสามารถต่างส่งผลต่อประสิทธิภาพของการเชื่อมโยงสเกล เมื่อพิจารณาอิทธิพลหลักในส่วนของสัดส่วนน้ำหนักคะแนนข้อสอบรวม พบว่า สัดส่วนคะแนนข้อสอบรวม 25% มีค่า Bias และค่า RMSE ต่ำที่สุด นั่นคือ การเชื่อมโยงสเกลในโครงสร้าง APSS และสัดส่วนน้ำหนักคะแนนข้อสอบรวม 25% มีประสิทธิภาพมากกว่าสัดส่วนอื่น ๆ (ศิวัชทัศนีย์ สุขสุวรรณ และคณะ, 2555) นอกจากนี้ยังมีงานวิจัยที่ระบุสัดส่วนข้อสอบรวมที่น้อยที่สุดที่ทำให้การปรับเทียบคะแนนข้อสอบที่ตรวจให้คะแนนสองค่ามีประสิทธิภาพ เมื่อความยาวของแบบสอบเป็น 160, 120, 100 และ 80 ข้อ คือ 20% และเมื่อความยาวแบบสอบเป็น 60 และ 40 ข้อ สัดส่วนข้อสอบรวมที่มีประสิทธิภาพ คือ 25% และ 33% ตามลำดับ (Xiong et al., 2012)

อย่างไรก็ตามการเชื่อมโยงสเกลด้วยโมเดล UIRT โดยการใช้การประมาณค่าพารามิเตอร์แบบแยกกันและวิธีประมาณค่าพารามิเตอร์แบบพร้อมกัน สำหรับแบบสอบรูปแบบผสมที่มีจำนวนข้อสอบรวมแตกต่างกัน 3 รูปแบบ ได้แก่ 5, 10 และ 15 ข้อ คิดเป็น 10%, 20% และ 30% ตามลำดับ พบว่าเมื่อข้อสอบรวมเพิ่มขึ้นจะมีค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (MSE) ลดลง เนื่องจากการเพิ่มจำนวนข้อสอบรวมทำให้ค่าความสามารถของผู้สอบที่ประมาณได้ใกล้เคียงกับความสามารถที่แท้จริงมากขึ้น (Tian, 2011) สอดคล้องกับการเชื่อมโยงสเกลด้วยโมเดล UIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า โดยแบบสอบแต่ละฉบับมีข้อสอบทั้งหมด 50 ข้อ มีข้อสอบรวมแตกต่างกัน 2 รูปแบบ คือ 20% (10 ข้อ) และ 40% (20 ข้อ) ที่พบว่า การเชื่อมโยงสเกลด้วยจำนวนข้อสอบรวม 40% มีค่าเฉลี่ยความคลาดเคลื่อนยกกำลังสอง (MSE) น้อยกว่าจำนวนข้อสอบรวม 20% (Bradley A. Hanson & Anton A. Béguin, 2002) จะเห็นว่าสัดส่วนคะแนนข้อสอบรวมและจำนวนข้อสอบรวมส่งผลต่อประสิทธิภาพของ

การปรับเทียบคะแนนและการเชื่อมโยงสเกล แต่ผลการวิจัยที่ผ่านมายังมีความขัดแย้งกันอาจเป็นเพราะบริบทที่ศึกษามีความแตกต่างกัน เช่น วิธีการเชื่อมโยงสเกล ขนาดกลุ่มตัวอย่าง และความยาวของแบบสอบ เป็นต้น แต่อย่างไรก็ตามการปรับเทียบคะแนนด้วยโมเดล MIRT ภายใต้รูปแบบ NEAT เลือกใช้สัดส่วนคะแนนข้อสอบร่วม 50% แต่ยังไม่มีการศึกษาเกี่ยวกับสัดส่วนข้อสอบร่วมดังนั้นในเบื้องต้นผู้วิจัยจึงตั้งสมมติฐานได้ดังนี้

**สมมติฐานที่ 2 กระบวนการ MOSE ทั้ง 2 แบบ น่าจะมีค่า CVSE และค่า RMSE ลดลงเมื่อสัดส่วนคะแนนข้อสอบร่วมเพิ่มขึ้น**

การปรับเทียบคะแนนสำหรับแบบสอบเอกมิติด้วยวิธีเชิงเส้น ได้แก่ วิธี Chained linear วิธีของ Tucker และวิธีของ Levine ภายใต้รูปแบบ NEAT พบว่า แบบสอบร่วมซึ่งประกอบด้วยข้อสอบที่ตรวจให้คะแนนสองค่าอย่างเดียวยังมีความลำเอียงและความคลาดเคลื่อนมากกว่าแบบสอบร่วมรูปแบบผสม (Sooyeon Kim et al., 2010) การเชื่อมโยงสเกลด้วยโมเดล IRT ได้แก่ วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันและวิธีโค้งคุณลักษณะ สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มเท่าเทียมและรูปแบบ NEAT โดยอัตราส่วนคะแนนข้อสอบที่ตรวจให้คะแนนสองค่า คือ 50:50, 60:40, 70:30: และ 80:20 และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 3:1, 1:1, และ 1:3 พบว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความคลาดเคลื่อนลดลง เมื่อแบบสอบร่วมประกอบด้วยข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่า และจำนวนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าลดลง (อัญชลี ศรีกลชาญ, 2552) และการเชื่อมโยงสเกลด้วยโมเดล IRT ได้แก่ วิธีของ Stocking-Lord และวิธีประมาณค่าพารามิเตอร์พร้อมกัน สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT ที่พบว่า ค่า MSE ลดลงเมื่อจำนวนข้อสอบร่วมแบบสร้างคำตอบเพิ่มขึ้น (Kim, 2004 อ้างอิงในอัญชลี ศรีกลชาญ, 2552) ซึ่ง Kim (2004 cited in Tian, 2011) อธิบายความสัมพันธ์ระหว่างผลของการปรับเทียบคะแนนและประเภทของข้อสอบข้อสอบร่วมไว้ว่า จำนวนของชั้นการตอบ (response categories) ส่งผลต่อกระบวนการเชื่อมโยงสเกล กล่าวคือ ถ้าจำนวนชั้นของการตอบเพิ่มขึ้น ผลของการเชื่อมโยงสเกลจะมีความเสถียร (stable) และถูกต้องแม่นยำ (accurate) มากขึ้น ดังนั้นผู้วิจัยจึงตั้งสมมติฐานได้ดังนี้

**สมมติฐานที่ 3 กระบวนการ MOSE ทั้ง 2 แบบ น่าจะมีค่าความลำเอียงสัมพัทธ์ (RB) และ ค่า CVSE เพิ่มขึ้น เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าเพิ่มขึ้น**

การเปรียบเทียบประสิทธิภาพระหว่างวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT 4 วิธี ได้แก่ (1) Direct (2) Equated Function (3) TCF และ (4) ICF สำหรับรูปแบบ NEAT พบว่า การเชื่อมโยงสเกลทั้ง 4 วิธีสำหรับโครงสร้าง APSS ให้ผลการเชื่อมโยงสเกลที่ถูกต้องแม่นยำและมีความเสถียรมากกว่าโครงสร้าง CS เนื่องจากการประมาณโครงสร้างแบบสอบและค่าพารามิเตอร์ของข้อสอบสำหรับ

โครงสร้าง APSS มีความถูกต้องมากกว่าโครงสร้าง CS ซึ่งพิจารณาจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความแตกต่างของค่าพารามิเตอร์ข้อสอบที่ประมาณได้ระหว่างกลุ่มฐาน (base group) และกลุ่มที่ถูกปรับเทียบ (equated group) ในกรณีโครงสร้าง APSS วิธีการเชื่อมโยงสเกลทั้ง 4 วิธีให้ผลการเชื่อมโยงสเกลที่ถูกต้องแม่นยำและมีความเสถียรใกล้เคียงกัน แต่กรณีโครงสร้าง CS พบว่า (1) วิธี Equated Function ไม่มีประสิทธิภาพในการเชื่อมโยงสเกล (2) ประสิทธิภาพของวิธีการเชื่อมโยงสเกลทั้ง 3 วิธีที่เหลือขึ้นอยู่กับเงื่อนไขของการทดสอบ (3) วิธี Direct method มีประสิทธิภาพในการเชื่อมโยงสเกลมากที่สุด แต่เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ คือ 2,000 คน ประสิทธิภาพของวิธีการเชื่อมโยงสเกลทุกวิธียกเว้นวิธี Equated Function สำหรับโครงสร้าง APSS และ CS ใกล้เคียงกัน เนื่องจากกลุ่มตัวอย่างขนาดใหญ่ช่วยลดความคลาดเคลื่อนการประมาณค่าพารามิเตอร์ของข้อสอบ (Wei, 2008) สอดคล้องกับผลการวิจัยของการเชื่อมโยงสเกลด้วยวิธี TCF สำหรับแบบสอบรูปแบบผสมที่มีโครงสร้าง APSS และ CS เมื่อสัดส่วนน้ำหนักคะแนนข้อสอบรวมแตกต่างกัน ซึ่งพบว่า สัดส่วนน้ำหนักคะแนนข้อสอบรวมและโครงสร้างมิติความสามารถไม่ร่วมกันส่งผลต่อค่า Bias และ ค่า RMSE ของการเชื่อมโยงสเกล แต่สัดส่วนน้ำหนักคะแนนข้อสอบรวมและโครงสร้างมิติความสามารถต่างส่งผลต่อค่า Bias และค่า RMSE เมื่อพิจารณาคูณภาพของการเชื่อมโยงสเกลตามอิทธิพลหลักในส่วนของโครงสร้างมิติความสามารถ พบว่า โครงสร้าง APSS มีค่า Bias และค่า RMSE ต่ำกว่าโครงสร้าง CS (ศิวัชรัตน์ สุขสุวรรณ และคณะ, 2555) และสอดคล้องกับผลการวิจัยของการเปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลในแนวตั้ง 2 วิธี ได้แก่ วิธี TCF และวิธี NOP สำหรับรูปแบบ NEAT ซึ่งใช้ข้อมูลจริงและข้อมูลจำลอง โดยศึกษาทั้งโครงสร้างมิติความสามารถ 2 ลักษณะ คือ APSS และ CS พบว่า การเชื่อมโยงสเกลในแนวตั้งสำหรับโครงสร้าง CS มีความถูกต้องต่ำกว่าโครงสร้าง APSS แต่ถ้าต้องการเชื่อมโยงสเกลสำหรับโครงสร้าง APSS ควรเลือกใช้วิธี NOP (Yon, 2006 อ้างอิงในพัชรี จันทร์เพ็ง, 2547) ดังนั้นผู้วิจัยจึงตั้งสมมติฐานได้ดังนี้

**สมมติฐานที่ 4** กระบวนการ MOSE ทั้ง 2 แบบ สำหรับโครงสร้างที่ซับซ้อน (CS) น่าจะมีค่าความลำเอียงสัมพัทธ์ (RB) และค่า RMSE มากกว่าโครงสร้างที่ไม่ซับซ้อน (APSS)

## ข้อตกลงเบื้องต้น

วิธีการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์ในการวิจัยครั้งนี้ คือ วิธีอีควิเปอร์เซ็นไทล์ที่ปรับโค้งให้ราบเรียบ (pre-smoothed equipercentile equating method) ซึ่งปรับโค้งการแจกแจงด้วย Polynomial log-linear model โดยมีข้อตกลงเบื้องต้นว่าวิธีนี้เป็นเกณฑ์ที่เหมาะสมสำหรับงานวิจัยในครั้งนี้อย่างไรก็ตาม รายละเอียดของวิธีการดังกล่าวอยู่ในที่บทที่ 2 หัวข้อ 2.5

## ขอบเขตของการวิจัย

ขอบเขตของการวิจัย ประกอบด้วย ตัวแปรที่ใช้ในการวิจัย โมเดลสำหรับการจำลองข้อมูล ค่ามุมที่ใช้จำลองข้อมูล จำนวนการทำซ้ำและขนาดตัวอย่าง และวิธีการปรับเทียบคะแนนเกณฑ์

1. **ตัวแปรที่ใช้ในการวิจัย** ประกอบด้วย ตัวแปรตาม 1 ตัว คือ ประสิทธิภาพในการปรับเทียบคะแนน และตัวแปรอิสระ 4 ตัว คือ (1) รูปแบบของกระบวนการ MOSE (2) โครงสร้างมิติความสามารถ (3) สัดส่วนคะแนนข้อสอบรวม และ (4) อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า รายละเอียดดังตาราง 1.1

ตาราง 1.1 ตัวแปรตามและตัวแปรอิสระที่ใช้ในการวิจัย

ตัวแปรตาม	ตัวแปรอิสระ
<b>ประสิทธิภาพในการปรับเทียบคะแนน วัดจากตัวบ่งชี้ 3 ตัว</b> 1) ค่าความลำเอียงสัมพัทธ์ (RB) 2) สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (CVSE) 3) รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE)	<b>1) รูปแบบของกระบวนการ MOSE 2 แบบ</b> - กระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบพร้อมกัน (CMOSE) - กระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบแยกกัน และใช้การเชื่อมโยงสเกลด้วยวิธี TCF (SMOSE)
	<b>2) โครงสร้างมิติความสามารถ 2 แบบ</b> - โครงสร้างที่ไม่ซับซ้อน (APSS) - โครงสร้างที่ซับซ้อน (CS)
	<b>3) สัดส่วนคะแนนข้อสอบรวม 3 แบบ</b> - 10%, 20% และ 30%
	<b>4) อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 4 แบบ</b> - 40:60, 50:50, 60:40 และ 70:30

## 2. โมเดลสำหรับการจำลองข้อมูล

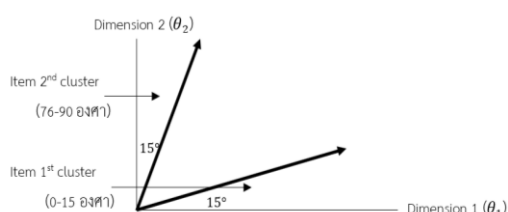
โมเดลในการจำลองข้อมูลเป็นโมเดล MIRT แบบทดแทนได้ (compensatory model) ประกอบด้วย โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (M3PL) สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า และโมเดล Multidimensional generalized partial credit แบบ 2 พารามิเตอร์ (MGPC) สำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า ที่มีการกำหนดค่าพารามิเตอร์ในโมเดลจากค่าอำนาจจำแนกรวม (MDISC) และค่าความยากรวม (MDIFF) ตามแนวคิดของ Min (2003 อ้างอิงใน Zhang, 2012) แบ่งชุดข้อสอบออกเป็น 5 ระดับ โดยค่าอำนาจจำแนกรวม 5 ระดับ คือ 0.4, 0.8, 1.2, 1.6 และ 2.0 ตามลำดับ และค่าความยากรวม 5 ระดับ คือ -1.5, 1.0, 0, -1.0 และ 1.5 ตามลำดับ

## 3. ค่ามุมที่ใช้จำลองข้อมูล

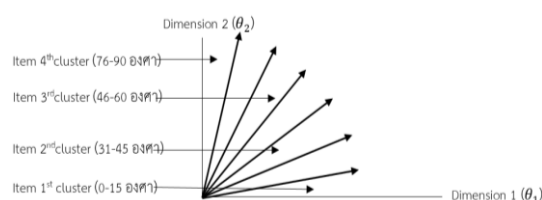
การวิจัยครั้งนี้ศึกษาโครงสร้างมิติความสามารถ 2 แบบ คือ โครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อน (CS) ซึ่งมีการกำหนดมุมระหว่างเวกเตอร์แตกต่างกัน โดยความสัมพันธ์ระหว่างมิติความสามารถที่ 1 และ 2 เป็น 0.5

**3.1 โครงสร้างที่ไม่ซับซ้อน (APSS)** ข้อสอบในแต่ละสัดส่วนแบ่งออกเป็น 2 กลุ่ม ข้อสอบกลุ่มแรกมีลักษณะอยู่ในมิติที่ 1 มากกว่ามิติที่ 2 ข้อสอบกลุ่มสองมีลักษณะอยู่ในมิติที่ 2 มากกว่ามิติที่ 1 กลุ่ม ซึ่งลักษณะเวกเตอร์ของข้อสอบในมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ( $\theta_1$ ) ตั้งแต่ 0 - 15 องศา ส่วนเวกเตอร์ของข้อสอบในมิติที่ 2 ( $\theta_2$ ) ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 76-90 องศา ดังแผนภาพ 1.1

**3.2 โครงสร้างที่ซับซ้อน (CS)** ข้อสอบแต่ละข้อของชุดข้อสอบมีความสัมพันธ์กันทั้งมิติที่หนึ่งและมิติที่สองซึ่งมีทั้งการวัดในแต่ละมิติและมีการวัดมิติโดยภาพรวม (composite dimension) โดยแบ่งแบบสอบออกเป็น 4 ชุดย่อย แบบสอบย่อย 2 ชุดแรก แยกเป็นมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 0-15 องศา และมิติที่ 2 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 76-90 องศา ส่วนแบบสอบย่อยอีก 2 ชุดวัดในลักษณะมิติรวม โดยชุดแรกมีความใกล้เคียงกับมิติที่ 1 มากกว่ามิติที่ 2 ทำมุมกับแกนที่ 1 ประมาณ 31-45 องศา ส่วนชุดสอง มีความใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 46-60 องศา ดังแผนภาพ 1.2 ซึ่งความสัมพันธ์ระหว่างมิติความสามารถที่ 1 และ 2 เป็น 0.5



ภาพ 1.2 โครงสร้างที่ไม่ซับซ้อน (APSS)



ภาพ 1.1 โครงสร้างที่ซับซ้อน (CS)

#### 4. จำนวนการทำซ้ำและขนาดตัวอย่าง

การวิจัยครั้งนี้ใช้การจำลองข้อมูลแบบมอนติคาร์โล ภายใต้อายุ 22 เดือน ซึ่งเวลาที่ใช้ในการดำเนินการ 1 รอบ ประมาณ 1 ชั่วโมงเป็นอย่างต่ำ ผู้วิจัยจึงสามารถกระทำซ้ำข้อมูลได้ 40 ครั้งในแต่ละเดือน โดยผู้วิจัยได้นำเสนอจำนวนการทำซ้ำของงานวิจัยที่ผ่านมาไว้ในบทที่ 2 หัวข้อ 4.1.5 และเนื่องจากการเปรียบเทียบคะแนนด้วยโมเดล MIRT ส่วนใหญ่ใช้กลุ่มตัวอย่างมากกว่า 2,000 คน ซึ่งขนาดกลุ่มตัวอย่างแต่ละกลุ่มที่น้อยที่สุด คือ 2,000 คน (O. Zhang, 2012) โดยงานวิจัยที่ใช้ข้อมูลจากการทดสอบจริง (การทดสอบ Iowa Tests of Educational Development: ITED) มีขนาดกลุ่มตัวอย่างกลุ่มละ 2,500 คน (Bradley Grant Brossman, 2010) นอกจากนี้ Kolen และ Brennan แนะนำว่า การเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติควรใช้กลุ่มตัวอย่างขนาดใหญ่ ด้วยเหตุนี้ผู้วิจัยจึงกำหนดให้ตัวอย่างแต่ละกลุ่มสำหรับการวิจัยครั้งนี้ คือ 3,000 คน

#### 5. วิธีการปรับเทียบคะแนนเกณฑ์

วิธีการปรับเทียบคะแนนเกณฑ์ คือ วิธีอีควิเปอร์เซ็นไทล์ที่ปรับโค้งให้ราบเรียบ (presmoothed equipercentile equating method) ซึ่งใช้ Polynomial log-linear model ปรับโค้งการแจกแจงให้ราบเรียบด้วยดีกรี (degree) ต่าง ๆ เพื่อหาโมเดลที่เหมาะสมกับข้อมูลมากที่สุด โดยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ได้จากการปรับโค้งแล้วต้องมีค่าเท่ากับข้อมูลเดิมที่ยังไม่ได้ทำการปรับ โดยรายละเอียดของการดำเนินการสำหรับวิธีดังกล่าวอยู่ในบทที่ 3 ตอนที่ 3

#### คำจำกัดความที่ใช้ในการวิจัย

**กระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์พร้อมกัน (CMOSE)** หมายถึง กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (MOSE) ที่นำไปใช้กับแบบสอบสองฉบับที่มีการประมาณพร้อมกัน (concurrent calibration)

**กระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบแยกกัน และใช้การเชื่อมโยงสเกลด้วยวิธี TCF (SMOSE)** หมายถึง กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (MOSE) ที่นำไปใช้กับแบบสอบสองฉบับที่มีการประมาณแยกกัน (separate calibration) ซึ่งต้องเชื่อมโยงสเกลระหว่างแบบสอบทั้งสองฉบับให้อยู่บนสเกลเดียวกันด้วยวิธี TCF หลังจากนั้นจึงทำการปรับเทียบคะแนนระหว่างแบบสอบ

**แบบสอบพหุมิติรูปแบบผสม** หมายถึง แบบสอบ 2 ฉบับที่มีโครงสร้างภายในเป็นองค์ประกอบเชิงซ้อน ซึ่งแบบสอบมุ่งวัดคุณลักษณะแฝงหรือความสามารถมากกว่า 1 มิติ โดยแบบสอบแต่ละฉบับประกอบด้วยข้อสอบที่ตรวจให้คะแนนสองค่า (0 และ 1 คะแนน) และตรวจให้คะแนนหลายค่า (0, 1 และ 2 คะแนน) ในอัตราส่วนคะแนน 70:30 และมีข้อสอบร่วมรูปแบบผสมในสัดส่วนต่าง ๆ



**โครงสร้างมิติความสามารถ** หมายถึง การรวมกันของค่าพารามิเตอร์ของข้อสอบในแต่ละมิติ ซึ่งสะท้อนความสัมพันธ์ระหว่างเวกเตอร์ข้อสอบแต่ละข้อกับมิติที่ศึกษา

**โครงสร้างที่ไม่ซับซ้อน (approximation simple structure: APSS)** หมายถึง ข้อสอบแต่ละข้อของชุดข้อสอบมีความสัมพันธ์กันสูงในมิติที่หนึ่งมากกว่ามิติที่สอง โดยมีค่าอำนาจจำแนกสูงในมิติเดียวกัน แต่มีค่าอำนาจจำแนกต่ำในอีกมิติหนึ่ง ซึ่งข้อสอบทั้งหมดแบ่งออกเป็น 2 กลุ่มย่อย ข้อสอบกลุ่มที่ 1 มีน้ำหนักในมิติที่ 1 มากกว่ามิติที่ 2 ข้อสอบกลุ่มที่สองมีน้ำหนักในมิติที่ 2 มากกว่ามิติที่ 2 (CS) ลักษณะเวกเตอร์ของข้อสอบในมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 0 - 15 องศา ส่วนเวกเตอร์ของข้อสอบในมิติที่ 2 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 76-90 องศา โดยกลุ่มข้อสอบรวมในแต่ละมิติมีคะแนนเต็มและจำนวนข้อสอบรวมในแต่ละมิติแปรเปลี่ยนไปตามเงื่อนไขของสัดส่วนคะแนนข้อสอบรวมที่กำหนดไว้

**โครงสร้างที่ซับซ้อน (complex structure: CS)** หมายถึง ข้อสอบแต่ละข้อของชุดข้อสอบมีความสัมพันธ์กันทั้งในมิติที่หนึ่งและมิติที่สองซึ่งมีทั้งการวัดในแต่ละมิติและการวัดมิติโดยภาพรวม (composite dimension) แบ่งแบบสอบออกเป็น 4 ชุดย่อย แบบสอบย่อย 2 ชุดแรก แยกเป็นมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 0-15 องศา และมิติที่ 2 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 76-90 องศา ส่วนแบบสอบย่อยอีก 2 ชุดวัดในลักษณะมิติรวม ชุดแรกมีความใกล้เคียงกับมิติที่ 1 มากกว่ามิติที่ 2 ทำมุมกับแกนที่ 1 ประมาณ 31-45 องศา ส่วนชุดที่ 2 มีความใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 45-60 องศา

**ความสัมพันธ์ระหว่างมิติความสามารถ** หมายถึง ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าความสามารถของผู้สอบในมิติที่ 1 และค่าความสามารถของผู้สอบในมิติที่ 2 ซึ่งมีค่าเป็น 0.5

**สัดส่วนคะแนนข้อสอบรวม** หมายถึง คะแนนเต็มของข้อสอบรวมที่อยู่ในแบบสอบแต่ละฉบับ เมื่อเทียบกับคะแนนเต็มทั้งฉบับ การวิจัยครั้งนี้ศึกษาสัดส่วนคะแนนข้อสอบรวม 3 แบบ คือ 10%, 20% และ 30%

**อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า** หมายถึง อัตราส่วนของคะแนนเต็มของข้อสอบรวมที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าเมื่อเทียบกับคะแนนเต็มของข้อสอบรวมทั้งหมด การวิจัยครั้งนี้ศึกษาอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 4 แบบ คือ 40:60, 50:50, 60:40 และ 70:30

**ประสิทธิภาพของการปรับเทียบคะแนน** หมายถึง ความแม่นยำถูกต้องของการปรับเทียบคะแนนระหว่างแบบสอบพหุมิติรูปแบบผสมทั้ง 2 ฉบับ พิจารณาจากค่าความลำเอียงสัมพัทธ์ (RB) และสัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (CVSE) และในภาพรวมพิจารณาจากค่า RMSE

**ค่าความลำเอียงสัมพัทธ์ (relative bias: RB)** หมายถึง ค่าสัมบูรณ์ของผลต่างระหว่างค่าเฉลี่ยของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา ( $\overline{\hat{e}_{base}(x_i)}$ ) และค่าเฉลี่ยของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนเกณฑ์ ( $\overline{e_{base}(x_i)}$ ) เมื่อเทียบกับส่วนเบี่ยงเบนมาตรฐานของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนเกณฑ์ ซึ่งสะท้อนความแม่นยำในการปรับเทียบคะแนน ดังสมการ

$$RB = \frac{|\overline{\hat{e}_{base}(x_i)} - \overline{e_{base}(x_i)}|}{\sigma_{base}(x_i)} \times 100$$

**สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (coefficient of variance of standard error of equating: CVSE)** หมายถึง การกระจายของค่าความคลาดเคลื่อนมาตรฐานในการปรับเทียบคะแนน ซึ่งสะท้อนความถูกต้องในการปรับเทียบคะแนน ดังสมการ

$$CVSE = \frac{SEE_i}{\overline{e_{base}(x_i)}} \times 100$$

โดยที่ความคลาดเคลื่อนมาตรฐาน (SE) เป็นขนาดของความคลาดเคลื่อนอย่างสุ่มในการปรับเทียบคะแนน ดังสมการ

$$SE_i = \sqrt{\frac{1}{N} \sum_{k=1}^N [\hat{e}_{base_k}(x_i) - \overline{\hat{e}_{base}(x_i)}]^2}$$

**รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (root mean square error: RMSE)** หมายถึง รากที่สองของผลรวมของความแปรปรวนความคลาดเคลื่อนมาตรฐานและความลำเอียงกำลังสอง ซึ่งสะท้อนประสิทธิภาพของการปรับเทียบคะแนนโดยรวมทั้งด้านความแม่นยำและความถูกต้อง ดังสมการ

$$RMSE_i = \sqrt{SE_i^2 + Bias_i^2}$$

**ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ (RMSE<sub>k</sub>)** หมายถึง ความสามารถของตัวประมาณในการประมาณค่าความคลาดเคลื่อนมาตรฐาน (SE) ของคะแนนปรับเทียบแต่ละคะแนนได้อย่างแม่นยำและถูกต้อง

$$RMSE_k = \sqrt{\frac{\sum_{i=1}^{rep} (SE_{ik} - MCSE_k)^2}{rep-1}}$$

## ประโยชน์ที่คาดว่าจะได้รับ

### 1. ประโยชน์ทางด้านวิชาการ

ผลการวิจัยเป็นการขยายองค์ความรู้เกี่ยวกับการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติสำหรับแบบสอบรูปแบบผสม ในสถานการณ์การทดสอบที่แตกต่างกัน เนื่องจากการวิจัยได้ศึกษาในประเด็นที่ยังไม่ได้ทำการศึกษา คือ การปรับเทียบคะแนนในรูปแบบผู้สอบรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม และในแง่ของข้อสอบที่เกี่ยวกับสัดส่วนคะแนนข้อสอบร่วม และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า

### 2. ประโยชน์ทางการนำไปใช้

- ในบริบทของการทดสอบขนาดใหญ่หรือการทดสอบมาตรฐานที่มีการวัดหลายคุณลักษณะ และต้องการปรับเทียบคะแนนระหว่างแบบสอบในแต่ละปีหรือแต่ละครั้ง สามารถใช้ผลการวิจัยนี้เป็นแนวทางสำหรับกำหนดโครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบร่วม ซึ่งกรณีที่ใช้แบบสอบพหุมิติรูปแบบผสม สามารถใช้ผลการวิจัยครั้งนี้เป็นแนวทางสำหรับกำหนดอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าได้อีกด้วย

- ผลการวิจัยครั้งนี้สามารถใช้ในการตัดสินใจเลือกวิธีการประมาณค่าพารามิเตอร์ระหว่างแบบสอบที่เหมาะสมได้ ซึ่งในเชิงทฤษฎีวิธีการประมาณค่าแบบพร้อมกันมีประสิทธิภาพดีกว่าการประมาณค่าพารามิเตอร์แยกกัน แต่กรณีที่ต้องใช้การประมาณค่าพารามิเตอร์แบบแยกกันก็สามารถใช้ผลการวิจัยครั้งนี้เป็นแนวทางในการกำหนดโครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบร่วม และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าได้เช่นกัน เนื่องจากผู้วิจัยได้ศึกษาประสิทธิภาพในการปรับเทียบคะแนนของทั้งกระบวนการ CMOSE และกระบวนการ SMOSE ครอบคลุมประเด็นดังกล่าว

## บทที่ 2

### เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการนำเสนอเอกสารและงานวิจัยที่เกี่ยวข้องกับการวิจัยเรื่อง “การเปรียบเทียบประสิทธิภาพของการเชื่อมโยงสเกล สำหรับแบบสอบถามแบบผสม ภายใต้ผู้สอบกลุ่มไม่เท่าเทียมกัน โดยใช้ข้อสอบร่วม: การศึกษากระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนอง ข้อสอบแบบพหุมิติภายใต้เงื่อนไขที่แตกต่างกัน” แบ่งเป็น 4 ตอนดังนี้

#### ตอนที่ 1 ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

- 1.1 มโนทัศน์และประเภทโมเดลการตอบสนองข้อสอบแบบพหุมิติ
- 1.2 ค่าพารามิเตอร์ฟังก์ชันสารสนเทศของข้อสอบและแบบสอบ
- 1.3 การประมาณค่าพารามิเตอร์และการแปลงค่าประมาณพารามิเตอร์ของข้อสอบและผู้สอบ1.4) โครงสร้างของแบบสอบพหุมิติ
- 1.5 โปรแกรมที่ใช้ในการจำลองข้อมูลแบบพหุมิติและการกำหนดช่วงของค่าพารามิเตอร์ข้อสอบ

#### ตอนที่ 2 การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

- 2.1 มโนทัศน์เกี่ยวกับการปรับเทียบคะแนน
- 2.2 การออกแบบการเก็บรวบรวมข้อมูลสำหรับการปรับเทียบคะแนน
- 2.3 กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ
- 2.4 กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองแบบพหุมิติ
- 2.5 การประเมินประสิทธิภาพของการปรับเทียบคะแนน
- 2.6 โปรแกรมการปรับเทียบคะแนนภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม
- 2.7 งานวิจัยเกี่ยวกับกระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

#### ตอนที่ 3 การเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

- 3.1 มโนทัศน์เกี่ยวกับการเชื่อมโยงสเกล
- 3.2 แนวทางการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบ
- 3.3 งานวิจัยเกี่ยวกับประสิทธิภาพของแนวทางการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบ
- 3.4 การเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ
- 3.5 งานวิจัยเกี่ยวกับการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองแบบพหุมิติ

#### ตอนที่ 4 สังเคราะห์งานวิจัยเกี่ยวกับการปรับเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล

##### MIRT และกรอบแนวคิดในการวิจัย

- 4.1 สังเคราะห์ประเด็นวิจัยเกี่ยวกับการปรับเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล MIRT
- 4.2 กรอบแนวคิดในการวิจัย

## ตอนที่ 1 ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (multidimensional item response theory: MIRT) พัฒนาจากแนวคิดทางจิตวิทยา การศึกษา การพัฒนาแบบสอบ และสถิติศาสตร์ โดยทฤษฎีดังกล่าวพัฒนาขึ้นเพื่อตอบสนองสถานการณ์การทดสอบจริง ซึ่งผู้สอบอาจใช้ทักษะหรือความสามารถมากกว่า 1 มิติเพื่อทำข้อสอบข้อหนึ่ง ความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องจึงขึ้นอยู่กับความสามารถหรือทักษะหลายด้านๆ อย่างเช่น ปัญหาทางวิทยาศาสตร์ที่ผู้สอบต้องใช้ทั้งทักษะทางวิทยาศาสตร์ในแก้ปัญหาและใช้ความสามารถทางภาษาในการทำความเข้าใจคำถาม ดังนั้นทฤษฎี การตอบสนองข้อสอบแบบพหุมิติจึงได้รับการพัฒนาเพื่อใช้อธิบายความสัมพันธ์ระหว่างมิติความสามารถของผู้สอบ 1 มิติหรือมากกว่า 1 มิติกับข้อสอบ โดยลักษณะพหุมิติเกิดจากหลายปัจจัย เช่น บริบทของข้อสอบ เนื้อหาของแบบสอบ และคุณลักษณะหลายทักษะ (trait) ที่ใช้ในการทำข้อสอบข้อหนึ่ง (Reckase, 2009)

### 1.1 มโนทัศน์และประเภทของโมเดลการตอบสนองข้อสอบแบบพหุมิติ

#### 1.1.1) มโนทัศน์ของโมเดลการตอบสนองข้อสอบแบบพหุมิติ

โมเดล MIRT ที่สอดคล้องกับสัดส่วนของการตอบข้อสอบได้ถูกต้องที่สังเกตได้ย่อมให้ค่าประมาณของความน่าจะเป็นอย่างมีเงื่อนไขของการตอบข้อสอบได้ถูกต้อง ซึ่งให้ฟังก์ชัน  $\theta = (\theta_1, \theta_2, \dots, \theta_m)$  สำหรับพื้นที่มิติ  $m$  เมื่อ  $m$  คือ จำนวนมิติ โมเดล MIRT เป็นฟังก์ชันความน่าจะเป็นอย่างต่อเนื่องที่เชื่อมโยงกับพื้นที่ซึ่งระบุโดย  $\theta$  ไปยังความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องสำหรับข้อสอบข้อที่  $i$  ด้วยพารามิเตอร์เชิงโครงสร้าง (structural parameters) ที่ระบุไว้ โมเดลนี้สามารถใช้กับข้อสอบที่มีการให้คะแนนแบบ 0 และ 1

$$P_i(\theta_1, \dots, \theta_m) = \Pr(u_i = 1 | \theta_1, \dots, \theta_m) \equiv P_i(\theta) \quad (2.1)$$

เมื่อ  $i = 1, 2, 3, \dots, i$  ซึ่งก็คือ ข้อสอบ อาจพิจารณาว่าเป็นชุดของค่าพารามิเตอร์เชิงโครงสร้าง (structural parameters)

โมเดลการตอบสนองข้อสอบแบบพหุมิติรูปแบบทั่วไปเป็นดังนี้

$$P_i(U = u | \theta) = f(\theta, \eta_i, u) \quad (2.2)$$

เมื่อ  $\eta$  คือ เวกเตอร์ของพารามิเตอร์เชิงโครงสร้าง ซึ่งอธิบายคุณลักษณะของข้อสอบ

$U$  คือ คะแนนของข้อสอบ

$u$  คือ คะแนนที่เป็นไปได้

$f$  คือ ฟังก์ชันอธิบายความสัมพันธ์ระหว่างตำแหน่งของผู้สอบที่ระบุโดย  $\theta$  กับความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง

คะแนนของข้อสอบอาจปรากฏทั้งสองข้างของสมการถ้าคะแนนของข้อสอบมีทั้งค่า 0 และ 1 แต่สมการดังกล่าวไม่เหมาะสมกับข้อสอบที่ตรวจให้คะแนนหลายค่า

โมเดล MIRT ส่วนใหญ่ถือว่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ( $P_i(\theta)$ ) เป็นฟังก์ชันเพิ่มทางเดียวของ  $\theta$  ข้อตกลงเบื้องต้นนี้เรียกว่า Monotonicity นอกจากนี้ยังถือว่าผู้สอบเป็นอิสระในการตอบข้อสอบแต่ละข้อ โดยการตอบข้อสอบของผู้สอบคนหนึ่งสำหรับข้อสอบข้อหนึ่งไม่ได้เปลี่ยนแปลงตามแนวโน้มการตอบข้อสอบข้ออื่น ๆ และไม่มีการลอกข้อสอบระหว่างผู้สอบ จึงถือว่าการตอบข้อสอบของผู้สอบขึ้นอยู่กับเวกเตอร์  $\theta$  และเวกเตอร์  $\eta$  เท่านั้น โดยข้อตกลงเบื้องต้นเกี่ยวกับการอิสระในการตอบข้อสอบทุกข้อสำหรับผู้สอบทุกคน เรียกว่า local independence

### ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

1. การเพิ่มขึ้นทางเดียว (monotonicity assumption) กล่าวคือ ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องเพิ่มขึ้นเมื่อโครงสร้างตามสมมติฐานที่ต้องการวัดเพิ่มขึ้น แสดงว่า ถ้าสมาชิกของเวกเตอร์  $\theta$  เพิ่มขึ้น ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องย่อมเพิ่มขึ้นด้วย

2. ความเป็นอิสระในการตอบข้อสอบ (local independence assumption) ระบุว่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องหลายข้อพิจารณาจากผลคูณของความน่าจะเป็นในการตอบข้อสอบถูกหลายข้อ เมื่อการตอบข้อสอบในแต่ละบุคคลเป็นอิสระจากกัน ดังสมการ

$$P(U_1 = u_1, \dots, U_i = u_i | \theta) = \prod_{i=1}^I P(U_i = u_i | \theta) \quad (2.3)$$

McDonald (1967, 1981 cited in Reckase, 2009) ระบุว่าข้อตกลงเบื้องต้นที่อ่อนกว่า คือ ความแปรปรวนร่วมอย่างมีเงื่อนไขระหว่างข้อสอบแต่ละคู่เป็นศูนย์ นั่นคือ

$$E(\text{cov}(U_i, U_j) | \theta) = 0, i \neq j \quad (2.4)$$

สำหรับค่าเวกเตอร์  $\theta$  ทุกค่า

#### 1.1.2) ประเภทของโมเดลการตอบสนองข้อสอบแบบพหุมิติ

โมเดล MIRT แสดงถึงปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบในรูปแบบของสมการทางคณิตศาสตร์ แนวคิดสำคัญของโมเดล คือ คุณสมบัติเฉพาะของผู้สอบมากกว่า 1 องค์ประกอบส่งผลต่อการตอบข้อสอบ วิธีการวิเคราะห์ข้อสอบแบบพหุมิติประกอบด้วย (1) โมเดลเชิงสำรวจ สำหรับประมาณค่าพารามิเตอร์ความสามารถของผู้สอบที่มีมากกว่า 1 มิติและค่าพารามิเตอร์ของข้อสอบ เพื่อให้โมเดลกับข้อมูลเชิงประจักษ์ มีความสอดคล้องกัน ซึ่งเป็นการหาจำนวนองค์ประกอบเฉพาะของผู้สอบที่มีการตอบข้อสอบ และ (2) โมเดลเชิงยืนยัน เพื่อประมาณค่าพารามิเตอร์ความสามารถของผู้สอบที่มีมากกว่า 1 มิติและค่าพารามิเตอร์ของข้อสอบตามจำนวนองค์ประกอบของคุณลักษณะแฝงที่ระบุไว้ Reckase (2009) กำหนดประเภทของโมเดลการตอบสนองข้อสอบไว้ 2 ประเภท ได้แก่ (1) โมเดลที่สามารถทดแทนได้ (compensatory model) กล่าวคือ ถ้าผู้สอบมีความสามารถน้อยในมิติที่หนึ่งสามารถทดแทนจุดอ่อนนี้ได้ด้วยจุดแข็งในอีกมิติ และโมเดลที่ไม่สามารถทดแทนได้ (noncompensatory model) กล่าวคือ ถ้าผู้สอบมีความสามารถน้อยในมิติที่ไม่สามารถทดแทนจุดอ่อนนี้ได้ด้วยจุดแข็งจากอีกมิติหนึ่งได้ ในปัจจุบันโปรแกรมสำหรับการประมาณค่า MIRT ใช้ได้เฉพาะกับโมเดลที่สามารถทดแทนได้

ดังนั้นโมเดล MIRT ที่จะอธิบายในส่วนต่อไปจึงเป็นโมเดลที่สามารถทดแทนได้ (compensatory model) เท่านั้น โดยแบ่งการนำเสนอเป็น 2 ส่วนตามประเภทของการตรวจให้คะแนน 2 ประเภท คือ การตรวจให้คะแนนสองค่า และการตรวจให้คะแนนหลายค่า

### 1) โมเดลการตอบสนองข้อสอบแบบพหุมิติสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า

โมเดลการตอบสนองข้อสอบแบบพหุมิติสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า แบ่งเป็น 2 ประเภท คือ Compensatory Extensions of the UIRT Models และ Partially Compensatory Extensions of UIRT Models (Reckase, 2009)

#### 1.1) Compensatory Extensions of the UIRT Models

โมเดล Compensatory Extensions of the UIRT Model ประกอบด้วย (1) โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (M2PL) (2) โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (M3PL) (3) โมเดล Multidimensional extension of Rash model และ (4) โมเดล Multidimensional extension of the normal ogive model มีรายละเอียดดังนี้

##### 1. Multidimensional extension of the two-parameter logistic model (M2PL)

โมเดลโลจิสติกแบบ 2 มิเตอร์ มีเลขชี้กำลังเป็น  $a(\theta - b) = a\theta - ab$  ถ้าแทน  $-ab$  เป็น  $d$  ( $d$  เป็นความชัน/จุดตัด) ได้เป็น  $a\theta + d$  ซึ่งขยายโมเดลโลจิสติก 2 พารามิเตอร์ไปยังเวกเตอร์  $\theta$  ที่มีองค์ประกอบหลากหลายด้วยการแทนที่ความชัน/จุดตัดด้วยเลขชี้กำลัง  $\mathbf{a}_i\theta_j + d_i$  จึงได้สมการแบบพหุมิติของโมเดลโลจิสติกแบบ 2 พารามิเตอร์ ดังนี้

$$P(U_{ij} = 1 | \theta_j, \mathbf{a}_i, d_i) = \frac{e^{\mathbf{a}_i\theta_j + d_i}}{1 + e^{\mathbf{a}_i\theta_j + d_i}} \quad (2.5)$$

โดย

$$\mathbf{a}_i\theta_j + d_i = \sum_{l=1}^m a_{il}\theta_{jl} + d_i \quad (2.6)$$

เมื่อ  $P(U_{ij} = 1 | \theta_j, \mathbf{a}_i, d_i)$  คือ ความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้อง ( $U_{ij} = 1$ ) สำหรับผู้สอบ  $j$  ของข้อสอบข้อที่  $i$

$\mathbf{a}_i$  คือ  $m \times 1$  เวกเตอร์ของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

$d_i$  คือ ค่าพารามิเตอร์ที่เป็นสเกลาร์ซึ่งสัมพันธ์กับความยากของข้อสอบ

$\theta_j$  คือ  $m \times 1$  เวกเตอร์ของค่าพารามิเตอร์ความสามารถสำหรับผู้สอบ

$m$  คือ จำนวนของมิติความสามารถ

เลขชี้กำลัง  $e$  เป็นฟังก์ชันเชิงเส้นของส่วนประกอบของค่า  $\theta$  กับพารามิเตอร์  $d$  เป็นจุดตัดและส่วนประกอบของเวกเตอร์  $\mathbf{a}$  เป็นพารามิเตอร์ความชัน โดยนิพจน์ในเลขชี้กำลังเป็นเส้นตรงในพื้นที่มิติ  $m$  ถ้ากำหนดเลขชี้กำลังเป็นค่าคงที่ใดๆ เช่น  $k$  โดย  $k = \mathbf{a}_i\theta_j + d_i$  นิพจน์ดังกล่าวอยู่บนเส้นตรง ซึ่งแสดงให้เห็นว่ามีความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องเท่ากัน

ค่าพารามิเตอร์  $a$  แสดงแนวโน้มของ Equiprobable contour และอัตราการเปลี่ยนแปลงของ  $P(U_{ij} = 1|\theta_j, \mathbf{a}_i, d_i)$  เมื่อเทียบกับ  $\theta$  ดังนั้นอนุพันธ์ในส่วนของนิพจน์ที่เกี่ยวข้องมีตร่วม  $\theta_i$  โดยผลของ  $P(U_{ij} = 1|\theta_j, \mathbf{a}_i, d_i) = P$  และ  $Q = (1 - P)$  เป็นดังสมการ

$$\frac{\partial P}{\partial \theta_i} = a_i P(1 - p) = a_i P Q \quad (2.7)$$

ค่าพารามิเตอร์  $a$  มีความสัมพันธ์กับความชันของพื้นผิวและอัตราการเปลี่ยนแปลงของค่าความน่าจะเป็นบนแกนพิกัด ดังนั้นค่าพารามิเตอร์  $a$  จึงเรียกว่าความชัน หรือค่าพารามิเตอร์อำนาจจำแนก ส่วนพารามิเตอร์  $d$  ไม่ได้เป็นค่าพารามิเตอร์ความยากในความหมายเดียวกับโมเดล UIRT เพราะค่าดังกล่าวไม่ได้เป็นตัวชี้วัดเฉพาะ (unique indicator) ของค่าความยากของข้อสอบ โดยความยากเชิงสัมพัทธ์ของข้อสอบที่สัมพันธ์กับมิติร่วมที่สอดคล้องกันหาได้จากจุดตัดที่เป็นค่าลบหารด้วยส่วนประกอบเวกเตอร์ของพารามิเตอร์อำนาจจำแนกของข้อสอบ

## 2. Multidimensional extension of the tree-parameter logistic model (M3PL)

โมเดล M2PL กำหนดค่าโอกาสในการเดาสำหรับผู้สอบที่มีความสามารถต่ำจะตอบข้อสอบได้ถูกต้องเป็นศูนย์ ส่วนโมเดลโลจิสติก 3 พารามิเตอร์ ค่าโอกาสในการเดาไม่เท่ากับศูนย์ (Lord, 1980 อ้างอิงใน Reckase, 2009) ค่าโอกาสในการเดาคำตอบได้ถูกต้องมีเพียงค่าเดียว และสมการแบบพหุมิติของโมเดลโลจิสติกแบบ 3 พารามิเตอร์ คือ

$$P(U_{ij} = 1|\theta_j, \mathbf{a}_i, c_i, d_i) = c_i + (1 - c_i) \frac{e^{a_i \theta_j + d_i}}{1 + e^{a_i \theta_j + d_i}} \quad (2.8)$$

เมื่อค่าพารามิเตอร์  $c$  คือ พารามิเตอร์ของการเดาข้อสอบข้อที่  $i$  ซึ่งก็คือ ความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องสำหรับผู้สอบที่มีค่าความสามารถ  $\theta$  ต่ำมาก

## 3. Multidimensional extension of Rash model

โมเดล M2PL กำหนดค่าพารามิเตอร์  $a$  เป็น 1.0 ทำให้โมเดลโลจิสติก Rash 1 พารามิเตอร์ เป็นโมเดลย่อยของโมเดลโลจิสติก Rash 2 พารามิเตอร์ สมการทั่วไปของเลขชี้กำลังของโมเดล M2PL คือ

$$a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + \dots + a_{im}\theta_{jm} + d_i \quad (2.9)$$

กำหนดค่าพารามิเตอร์  $a$  เท่ากันทั้งหมด เรียกว่า  $a_{i*}$  จะได้ว่า

$$a_{i*}(\theta_{j1} + a_{i2}\theta_{j2} + \dots + a_{im}\theta_{jm}) + d_i \quad (2.10)$$

ผู้สอบ  $j$  พิกัด  $\theta$  ถูกกำหนดเป็นค่าหนึ่ง (fix) และรวมค่าความสามารถเหล่านั้นเป็น  $\theta_*$  จะได้ว่า

$$a_{i*}(\theta_{j*}) + d_i \quad (2.11)$$

กำหนดให้  $b_i = -d_i/a_{i*}$  จะได้เลขชี้กำลังเป็น;  $a_{i*}(\theta_{j*} - b_i)$  (2.12)

ถ้ากำหนดให้  $a_{i*}$  เป็น 1 เลขชี้กำลังที่ได้เหมือนกับ UIRT Rash model นั่นคือ ค่าพารามิเตอร์ความสามารถเพียงค่าเดียวที่มีความแตกต่าง



โมเดล Rash แบบพหุมิติเต็มรูปมีความซับซ้อนมากกว่าดังสมการ (2.13) โมเดลนี้ใช้สำหรับข้อสอบที่มีคะแนนสูงสุดสำหรับข้อสอบข้อที่  $i$  ไปจนถึงข้อที่  $K_i$  ถ้าระดับคะแนนต่ำสุดเท่ากับ 0 จำนวนของระดับคะแนนจึงเท่ากับ  $K_i + 1$  และกรณีของข้อสอบที่ตรวจให้คะแนนสองค่า  $K_i$  เท่ากับ 1 และระดับคะแนน คือ 0 และ 1 ซึ่งแทนด้วย  $k$  ส่วนตัวแปรสุ่ม  $X_{ik}$  เป็นตัวชี้วัดที่ระบุว่าคำตอบของข้อสอบข้อที่  $i$  เท่ากับ  $k$  ถ้าคะแนนเท่ากับ  $k$  ตัวชี้วัดกำหนดเป็น 1 หรือ 0 ดังนั้นข้อสอบที่ตรวจให้คะแนนสองค่าแล้วกำหนดให้  $X_{ik} = 1$  คะแนนของคำตอบที่ถูกต้องเท่ากับ 1

$$P(X_{ik} = 1 | \mathbf{A}, \mathbf{B}, \xi, \boldsymbol{\theta}) = \frac{e^{\mathbf{b}_{ik}\boldsymbol{\theta} + a_{ik}\xi}}{\sum_{k=0}^{K_i} e^{\mathbf{b}_{ik}\boldsymbol{\theta} + a_{ik}\xi}} \quad (2.13)$$

เมื่อ  $\mathbf{A}$  คือ เมทริกซ์การออกแบบ (design matrix) ประกอบด้วยเวกเตอร์  $\mathbf{a}_{ik}$  ซึ่งเลือกค่าพารามิเตอร์ข้อสอบที่เหมาะสมสำหรับการให้คะแนนข้อสอบ

$\mathbf{B}$  คือ เมทริกซ์การให้คะแนน (scoring matrix) ประกอบด้วยเวกเตอร์  $\mathbf{b}_{ik}$  ซึ่งระบุมิติเพียงมิติเดียวหรือหลายมิติ ได้คะแนน  $k$  สำหรับข้อสอบชิ้นนั้นๆ

$\xi$  คือ เวกเตอร์ของค่าพารามิเตอร์ความยากของข้อสอบ

$\boldsymbol{\theta}$  คือ เวกเตอร์ของพิกัดสำหรับตำแหน่งของบุคคลในพื้นที่โครงสร้าง

กรณีข้อสอบที่ตรวจให้คะแนนสองค่า *Rash model* แบบหลายมิติ เป็นดังสมการ

$$P(U_{ij} = 1 | \mathbf{a}_i, d_i, \boldsymbol{\theta}_j) = \frac{e^{\mathbf{a}_i\boldsymbol{\theta}_j + d_i}}{1 + e^{\mathbf{a}_i\boldsymbol{\theta}_j + d_i}} \quad (2.14)$$

เมื่อ  $\mathbf{a}_i$  เป็นปริมาณเวกเตอร์ ซึ่ง  $\mathbf{a}_i = \mathbf{b}_{ik}$  และ  $d_i$  เป็นปริมาณสเกลาร์ ซึ่ง  $d_i = a_{ik}\xi$

#### 4. Multidimensional extension of the normal ogive model

สมการแบบพหุมิติของ Normal ogive model เป็นดังนี้

$$P(U_{ij} = 1 | \boldsymbol{\theta}_j, \mathbf{a}_i, c_i, d_i) = c_i + (1 - c_i) \frac{1}{\sqrt{2\pi}} \int e^{\frac{-t^2}{2}} d \quad (2.15)$$

เมื่อ  $Z_{ij}(\boldsymbol{\theta}_j) = \mathbf{a}_i\boldsymbol{\theta}_j + d_i$  และสัญลักษณ์อื่นยังคงเหมือนเดิม ถ้า  $c_i = 0$  ผลลัพธ์คือ

Normal ogive ของโมเดลโลจิสติก 2 มิติแบบหลายมิติ โดยรูปแบบของโมเดลดังกล่าวจะระบุความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องเป็นพื้นที่ภายใต้การแจกแจงแบบปกติมาตรฐานตั้งแต่  $-Z_{ij}(\boldsymbol{\theta}_j)$  ไปจนถึงอนันต์ และเนื่องจากการแจกแจงแบบปกติมีลักษณะสมมาตร พื้นที่ภายใต้ดังกล่าวจึงเท่ากับพื้นที่ที่อยู่ภายใต้  $Z_{ij}(\boldsymbol{\theta}_j)$  (Bock and Schilling 2003; McDonald 1999; Samejima 1974 อ้างอิงใน Reakase, 2009)

#### 1.2) Partially Compensatory Extensions of UIRT Models

โมเดล Partially Compensatory Extensions of UIRT Model ประกอบด้วย 2 โมเดล ได้แก่ โมเดล Multidimensional extension of the one-parameter logistic และ โมเดล Multidimensional extension of the three-parameter logistic model มีรายละเอียดดังนี้

### 1. Multidimensional extension of the one-parameter logistic model

Maris (1995) เสนอ Rasch model ที่เชื่อมโยงกับสมการ (2.14) ซึ่งเป็นการขยายโมเดลโลจิสติกแบบ 1 พารามิเตอร์ดังสมการ (2.16) โดยเพิ่มผลลัพธ์เหมือนกับโมเดล Rasch UIRT model แต่โมเดลแบบพหุมิติไม่สอดคล้องกับข้อตกลงของ Rasch model เพราะไม่มีค่าสถิติที่เหมาะสมกับเวกเตอร์ความสามารถของผู้สอบ ด้วยเหตุนี้ Adams และคณะ (1997 อ้างอิงใน Reakase, 2009) จึงระบุว่าสมการ (2.16) เป็น Rasch model ได้ถ้าข้อสอบนั้นมีผลการตอบ 2 ระดับ

$$P(U_{ij} = 1 | \theta_j, \mathbf{b}_i) = \left( \prod_{k=1}^m \frac{e^{(\theta_{jk} - b_{ik})}}{1 + e^{(\theta_{jk} - b_{ik})}} \right) \quad (2.16)$$

กรณีโมเดล 2 พารามิเตอร์ เทอมของผลลัพธ์ แทนด้วย  $p_l$  เมื่อ  $l$  คือ มิติของความสนใจ โมเดลอย่างง่ายเป็นดังสมการ

$$k = \prod_{l=1}^m p_l \quad (2.17)$$

กรณี 2 มิติ  $k = p_1 p_2$  ซึ่งเป็นสมการไฮเพอร์โบล่าที่มีค่าเป็นเมตริกซ์ความน่าจะเป็น (probability metric)

### 2. Multidimensional extension of the three-parameter logistic model

Sympson (1978 อ้างอิงใน Reakase, 2009; Zhang, 2012) เสนอโมเดล Partially compensatory extensions of UIRT แสดงปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ ดังสมการ

$$P(U_{ij} = 1 | \theta_j, \mathbf{a}_i, c_i, d_i) = c_i + (1 - c_i) \left( \prod_{l=1}^m \frac{e^{1.7a_{il}(\theta_{jl} - b_{il})}}{1 + e^{1.7a_{il}(\theta_{jl} - b_{il})}} \right) \quad (2.18)$$

เมื่อ  $P(U_{ij} = 1 | \theta_j, \mathbf{a}_i, c_i, d_i)$  คือ ความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องของผู้สอบคนที่  $j$  สำหรับข้อสอบข้อที่  $i$

$U_{ij}$  คือ ผลการตอบข้อสอบ (0 = ผิด, 1 = ถูก) ของผู้สอบคนที่  $j$  สำหรับข้อสอบข้อที่  $i$

$\mathbf{a}_i$  คือ  $m \times 1$  เวกเตอร์ของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

$d_i$  คือ ค่าพารามิเตอร์ที่เป็นสเกลาร์ซึ่งสัมพันธ์กับความยากของข้อสอบ

$\theta_j$  คือ  $m \times 1$  เวกเตอร์ของค่าพารามิเตอร์ความสามารถสำหรับผู้สอบ

$m$  คือ จำนวนของมิติความสามารถ

$D$  คือ ค่าสัมประสิทธิ์ เท่ากับ 1.702 ดังนั้นค่าพารามิเตอร์ของโมเดลโลจิสติกจึงมีความหมายเหมือนกับค่าพารามิเตอร์ของโมเดล normal ogive model

การศึกษารูปแบบของพื้นผิวจากสมการ (2.18) พิจารณาที่  $c_i = 0$  และค่าความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องเป็นค่าที่ใดๆ  $k$

### ความแตกต่างระหว่าง Compensatory Models และ Partially Compensatory Models

Reakase (2005) ระบุความแตกต่างระหว่าง Compensatory Models และ Partially Compensatory Models ไว้ดังนี้

ตาราง 2.1 ความแตกต่างระหว่าง Compensatory Models และ Partially Compensatory Models

ประเด็น	Compensatory Models	Partially Compensatory Models
ความน่าจะเป็นในการตอบข้อสอบได้	เป็นผลรวมของ $\theta_s$ หลายๆค่าโดยใช้ฟังก์ชันเชิงเส้น (linear function) ในเลขชี้กำลังของ Compensatory Models	เป็นผลรวมของ $\theta_s$ หลายๆค่าโดยใช้เทอมผลลัพธ์ของ Partially Compensatory Models
เงื่อนไขความสอดคล้อง	โมเดลมีความสอดคล้องกับสมมติฐานมากกว่า เนื่องจากพิจารณาปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบแบบองค์รวมมากกว่า	โมเดลมีความสอดคล้องกับสมมติฐานเมื่อข้อสอบแต่ละข้อมีส่วนที่แตกต่างกันซึ่งสัมพันธ์กับทักษะและความรู้แตกต่างกัน ซึ่งเป็นความสำเร็จโดยรวมในการตอบข้อสอบแต่ละข้อได้ถูกต้อง

## 2) โมเดลการตอบสนองข้อสอบแบบพหุมิติสำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า

โมเดลการตอบสนองข้อสอบแบบพหุมิติสำหรับข้อสอบที่ตรวจให้คะแนนหลายค่าแบ่งเป็น 3 ประเภท ได้แก่ Multidimensional Generalized Partial Credit Model dimensional, Multidimensional Partial Credit Model และ Multidimensional Graded Response Model

### 2.1) Multidimensional Generalized Partial Credit Model

Multidimensional extension of the generalized partial credit (MGPC) อธิบายปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบที่มีการตรวจให้คะแนนมากกว่าสองค่า ดังสมการ

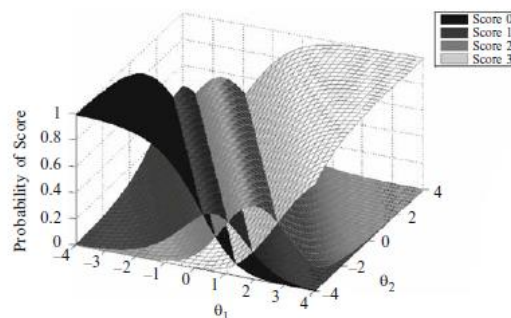
$$P(u_{ij} = k | \theta_i) = \frac{e^{ka_i\theta_j - \sum_{u=0}^k \beta_{iu}}}{\sum_{v=0}^{K_i} e^{va_i\theta_j - \sum_{u=0}^v \beta_{iu}}} \quad (2.19)$$

เมื่อ  $K_i$  คือ คะแนนสูงสุดสำหรับข้อสอบข้อที่  $i$  โดยคะแนนต่ำสุดกำหนดให้เป็น 0

$K_i + 1$  คือ ระดับคะแนนทั้งหมด  $k$  คือ คะแนนของผู้สอบสำหรับข้อสอบข้อที่  $i$

$\beta_{iu}$  คือ ค่าพารามิเตอร์เทอร์สโฮลด์ (threshold) สำหรับระดับคะแนน  $u$  ขณะที่  $\beta_{i0}$  เป็น 0

พื้นผิวของการตอบสนองข้อสอบ (item response surface) สำหรับโมเดล MGPC ในกรณีที่มีปฏิสัมพันธ์ระหว่างข้อสอบและผู้สอบสามารถแสดงอยู่ในรูปของพื้นที่ในสองมิติร่วมกันได้ ดังภาพ 2.1



ภาพ 2.1 ตัวอย่างพื้นผิวของการตอบสนองข้อสอบสำหรับโมเดล MGPC (Reckase, 2009)

เมื่อ  $\beta_{iu} = 0, -2.5, -1.5, .5$  และ  $\mathbf{a}_i = [1.2 \ .7]$

อินเตอร์เซกกระหว่างพื้นผิวสำหรับระดับคะแนนที่อยู่ติดกันซึ่งเป็นเส้นตรงในแกนนอน  $\theta$  โดยเส้น คือชุดของจุดในแกนนอน  $\theta$  ที่มีความน่าจะเป็นในการได้คะแนนที่ติดกันเท่ากัน โดยสมการสำหรับชุดของจุดดังกล่าวหาได้จากสมการ (2.20)

$$k\mathbf{a}_i\theta_j - \sum_{u=0}^k \beta_{iu} = (k+1)\mathbf{a}_i\theta_j - \sum_{u=0}^{k+1} \beta_{iu} \quad (2.20)$$

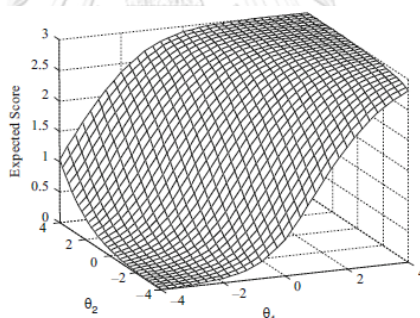
สมการนี้เป็นสมการสำหรับเส้นในพื้นที่มิติ  $m$  ที่เป็นตัวแทนของข้อสอบข้อหนึ่ง ๆ เทอมของจุดตัด  $\beta$  อาจเปลี่ยนแปลงเนื่องจากความแตกต่างของระดับคะแนนที่อยู่ติดกัน และเช่นเดียวกับโมเดล GPCM ค่าพารามิเตอร์เป็นกำหนดตำแหน่งของ Threshold ระหว่างคะแนนแต่ละระดับ

$$0 = \mathbf{a}_i\theta_j - \beta_{i,k+1} \quad \text{เมื่อ } k = 0, \dots, k-1 \quad (2.21)$$

สมการ (2.19) แทนความน่าจะเป็นที่จะได้แต่ละคะแนนสำหรับข้อสอบข้อหนึ่ง พื้นผิวที่แสดงถึงคะแนนที่คาดหวัง (expected score) ของข้อสอบข้อหนึ่งสำหรับผู้สอบคนหนึ่งที่มีเวกเตอร์  $\theta$  ดังสมการ

$$E(u_{ij}|\theta_j) = \sum_{k=0}^{K_i} kP(u_{ij} = k|\theta_j) \quad (2.22)$$

พื้นผิวถูกกำหนดด้วยสมการ (2.22) มีลักษณะดังภาพ 2.1 ส่วนพื้นผิวคะแนนที่คาดหวังเป็นดังภาพ 2.2 โดยพื้นผิวในภาพ 2.2 แสดงพื้นผิวการตอบสนองข้อสอบสำหรับข้อสอบที่ตรวจให้คะแนนหลายค่าสำหรับโมเดลแบบทดแทนได้ โดยคะแนนสูงสุด (expected score) คือ 3



ภาพ 2.2 พื้นผิวคะแนนที่คาดหวังสำหรับข้อสอบในภาพที่ 2.1 (Reckase, 2009)

## 2.2) Multidimensional Partial Credit Model

Kelderman และ Rijkens (1994 อ้างอิงใน Reckase, 2009) เสนอส่วนขยายแบบพหุมิติของ Rasch model สำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า ดังสมการ

$$P(u_{ij} = k|\theta_j) = \frac{e^{\sum_{l=1}^m (\theta_{jl} - b_{ilk}) W_{ilk}}}{\sum_{r=0}^{K_i} e^{\sum_{l=1}^m (\theta_{jl} - b_{ilr}) W_{ilr}}} \quad (2.23)$$

เมื่อ  $b_{ilk}$  คือ พารามิเตอร์ความยากของข้อสอบข้อที่  $i$  ในมิติ  $l$  สำหรับระดับคะแนน  $k$

$W_{ilk}$  คือ น้ำหนักการให้คะแนนที่กำหนดไว้ก่อนของข้อสอบข้อที่  $i$  ซึ่งสอดคล้องกับมิติ  $l$  และระดับคะแนน  $k$

ข้อสังเกต คือ ทักษะหรือความรู้ที่จำเป็นต้องใช้เพื่อให้ได้คะแนนสูงอาจมีแตกต่างจากทักษะและความรู้ที่จำเป็นต้องใช้เพื่อให้ได้คะแนนต่ำ ดังนั้นผู้ใช้โมเดลนี้จึงอาจประสบปัญหาการระบุทักษะและความรู้เฉพาะที่จำเป็นเฉพาะสำหรับข้อสอบข้อหนึ่ง แต่ก็ไม่ใช่อุปสรรคสำคัญนัก

### 2.3) Multidimensional Graded Response Model

ค่าพารามิเตอร์ของโมเดลสำหรับคะแนนต่ำสุดของข้อสอบข้อที่  $i$  เท่ากับ 0 และคะแนนสูงสุดเท่ากับ  $m_i$  โมเดลความน่าจะเป็นของผลสัมฤทธิ์ (accomplish)  $k$  หรือมากกว่าเป็นโมเดล normal ogive แบบ 2 พารามิเตอร์ที่มีพารามิเตอร์ผู้สอบเป็นผลรวมของเวกเตอร์  $\theta$  ซึ่งถูกถ่วงน้ำหนักด้วยค่าพารามิเตอร์อำนาจจำแนก ความน่าจะเป็นที่จะได้รับคะแนน  $k$  คือความแตกต่างระหว่างความน่าจะเป็นที่จะได้รับคะแนน  $k$  หรือมากกว่า และความน่าจะเป็นที่จะได้รับคะแนน  $k + 1$  หรือมากกว่า ถ้าความน่าจะเป็นของการได้คะแนน  $k$  หรือมากกว่าที่ระดับ  $\theta$  เฉพาะ คือ  $P^*(u_{ij} = k|\theta_j)$  ความน่าจะเป็นที่ผู้สอบจะได้รับคะแนน  $k$  ดังสมการ

$$P(u_{ij} = k|\theta_j) = P^*(u_{ij} = k|\theta_j) - P^*(u_{ij} = k + 1|\theta_j) \quad (2.24)$$

Samejima (1969 อ้างอิงใน Reakase, 2009) กำหนดเทอมขวาของสมการเป็นฟังก์ชันของการตอบแบบสะสม (cumulative category response functions) และเทอมซ้ายของสมการเป็นฟังก์ชันการตอบ (category response functions) โมเดล Normal ogive สำหรับโมเดล GRM คือ

$$P(u_{ij} = k|\theta_j) = \frac{1}{\sqrt{2\pi}} \int_{a_j\theta_j + d_{ik+1}}^{a_j\theta_j + d_{ik}} e^{-\frac{t^2}{2}} dt \quad (2.25)$$

เมื่อ  $k$  คือ คะแนนของข้อสอบ 0, 1, ..., m

$a_j$  คือ เวกเตอร์ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ

$d_{ik}$  คือ พารามิเตอร์ที่สอดคล้องกับผู้สอบที่อยู่ในขั้น  $k$  ของข้อสอบข้อนั้น

โดย  $d_{ik}$  มีความสัมพันธ์ผกผันกับคะแนนของข้อสอบ

ความน่าจะเป็นของระดับการตอบ  $k$  คำนวณจากความแตกต่างระหว่างสมการอินทิเกรตสองสมการ ดังสมการ

$$P(u_{ij} = k|\theta_j) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_j\theta_j + d_{ik}} e^{-\frac{t^2}{2}} dt - P(u_{ij} = k|\theta_j) - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a_j\theta_j + d_{ik+1}} e^{-\frac{t^2}{2}} dt \quad (2.26)$$

คะแนนที่คาดหวังของข้อสอบคำนวณจากการคูณคะแนนด้วยความน่าจะเป็นของคะแนนนั้น คะแนนที่คาดหวังเข้าใกล้ 0 เมื่อส่วนประกอบของเวกเตอร์  $\theta$  เข้าใกล้ -4 และคะแนนที่คาดหวังเข้าใกล้ 3 เมื่อส่วนประกอบของเวกเตอร์  $\theta$  เข้าใกล้ 4 พื้นผิวความน่าจะเป็นของการตอบของโมเดล MGRM จึงเหมือนกับคอนทัวร์ความน่าจะเป็น ดังนั้นพื้นผิวของคะแนนที่คาดหวังจึงเหมือนกับคอนทัวร์ของคะแนน โดยความชันของพื้นผิวขึ้นอยู่กับขนาดของส่วนประกอบของเวกเตอร์พารามิเตอร์  $a$  และความแปรปรวนของพารามิเตอร์  $d$  นอกจากนี้โมเดล MGRM ยังมีฟังก์ชันการ

ตอบในแต่ละระดับคล้ายคลึงกับ โมเดล MGPCM แต่อย่างไรก็ตามพบว่าชุดข้อมูลที่ได้จากการจำลอง ด้วยโมเดลทั้งสองมีความแตกต่างกัน

## 1.2 ค่าพารามิเตอร์และฟังก์ชันสารสนเทศของข้อสอบและแบบสอบ

ในส่วนนี้เป็นการอธิบายค่าพารามิเตอร์ของข้อสอบ สารสนเทศของข้อสอบ และฟังก์ชันของแบบสอบ มีรายละเอียดดังนี้

### 1.2.1) ค่าพารามิเตอร์ของข้อสอบ ประกอบด้วยค่าความยากและค่าอำนาจจำแนก

#### 1. ค่าความยาก

ค่าความยากสำหรับโมเดล MIRT แทนด้วยสัญลักษณ์  $B_i$  หรือ  $MDIFF$  ซึ่งเทียบได้กับพารามิเตอร์  $b$  สำหรับโมเดลการตอบสนองข้อสอบแบบเอกมิติ ค่า  $B$  มีค่าเป็นบวกสูง แสดงว่าข้อสอบยาก (ต้องมีค่าของส่วนประกอบของ  $\theta$  สูงจึงจะมีโอกาสตอบข้อสอบได้ถูกต้อง) แต่ถ้าค่า  $B$  น้อยแสดงว่ามีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องสูงสำหรับความสามารถ  $\theta$

การแปลความหมายค่า  $B$  พิจารณาเพียงทิศทางของเวกเตอร์  $\alpha$  ซึ่งอธิบายได้ 2 ลักษณะ ได้แก่ การระบุความยากของข้อสอบ (เช่น  $B$ ) และการรวมกันของแกนพิกัด (coordinate axes) ซึ่งเป็นค่าความยากส่วนใหญ่ของข้อสอบ ซึ่งการรวมกันของแกนร่วมระบุโดยทิศทางของความชันที่สูงชันที่สุดจากจุดกำเนิดของพื้นที่  $\theta$

$$B_i = \frac{-d_i}{\sqrt{\sum_{k=1}^m a_{ik}^2}} \quad (2.27)$$

Muraki และ Carlson (1993) นำเสนอค่าสถิติสำหรับโมเดล Multidimensional graded response model โดยสมการค่าอำนาจจำแนกแบบพหุมิติ ( $A_i$ ) และเวกเตอร์ของมุม  $\alpha$  เหมือนเดิม แต่ค่าความยากเป็นดังสมการ (2.28)

$$B_{ik} = \frac{-d_{ik}}{\sqrt{\sum_{i=1}^m a_{ik}^2}} \quad (2.28)$$

เมื่อ  $B_{ik}$  คือ ค่าความยาก (step difficulty) สำหรับขั้นที่  $k$  ของข้อสอบที่มีการให้คะแนนหลายค่า  $d_{ik}$  คือ ค่าพารามิเตอร์ (step parameters) ซึ่งถูกระบุในโมเดล

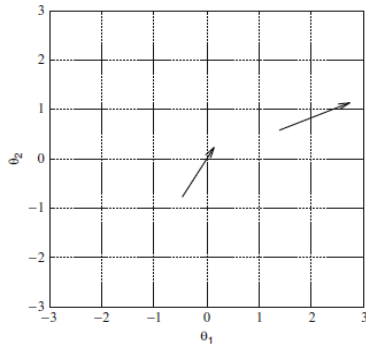
#### 2. ค่าอำนาจจำแนก

ค่าอำนาจจำแนกสำหรับโมเดล MIRT แทนด้วยสัญลักษณ์  $A_i$  หรือ  $MDISC_i$  ซึ่งก็คือความชันของพื้นผิวการตอบข้อสอบ ณ จุดของความชันที่มีค่ามากที่สุดในทิศทางจากจุดกำเนิดของพื้นที่  $\theta$

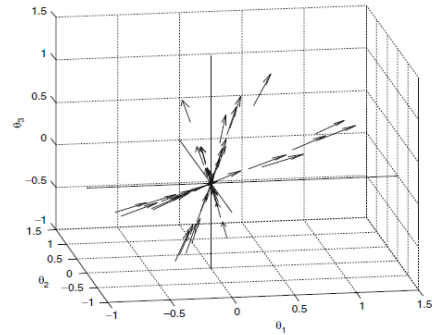
$$A_i = \sqrt{\sum_{k=1}^m a_{ik}^2} \quad (2.29)$$

การนำเสนอค่าความยากและค่าอำนาจจำแนกอาจนำเสนอ ดังภาพ 2.3 ความยาวของลูกศรแทนค่าอำนาจการจำแนกของข้อสอบ  $A_i$  ระยะห่างจากจุดกำเนิดไปยังฐานของลูกศรแทนค่าความ

ยาก  $B_i$  และทิศทาง  $\alpha_i$  ของลูกศรแทนทิศทางการเปลี่ยนแปลงเชิงบวกในความชันของข้อสอบ ส่วนภาพ 2.4 เป็นการนำเสนอค่าพารามิเตอร์จำนวน 45 ข้อในพื้นที่ 3 มิติ โดยภาพนี้แสดงให้เห็นว่าความชันที่จุดสูงสุดมีทิศทางเดียวกัน



ภาพ 2.4 การนำเสนอค่าพารามิเตอร์ข้อสอบจำนวน 2 ข้อในพื้นที่ 2 มิติ (ค่า  $\theta_1$  และ  $\theta_2$ )



ภาพ 2.3 การนำเสนอค่าพารามิเตอร์ข้อสอบจำนวน 45 ข้อในพื้นที่ 3 มิติ (ค่า  $\theta_1$ ,  $\theta_2$  และ  $\theta_3$ )

### 1.2.2) สารสนเทศของข้อสอบ (item information)

นิยามของสารสนเทศใน MIRT เหมือนกับ UIRT คือ ความชันยกกำลังสองของการถดถอยของคะแนนข้อสอบบนความสามารถ  $\theta$  หาด้วยความแปรปรวนของคะแนนที่ระดับความสามารถ  $\theta$  อย่างไรก็ตามความชันของพื้นผิวการตอบสนองข้อสอบแบบพหุมิติ (multidimensional item response surface) ของแต่ละจุดในพื้นที่  $\theta$  แตกต่างกันขึ้นอยู่กับทิศทางการเคลื่อนที่จากจุดหนึ่งไปยังอีกจุดหนึ่ง นิยามของสารสนเทศข้อสอบที่ตอบได้ถูกมีลักษณะทั่วไป คือ

$$I_{\alpha}(\theta) = \frac{[\nabla_{\alpha} P(\theta)]^2}{P(\theta)Q(\theta)} \quad (2.30)$$

เมื่อ  $\alpha$  คือ เวกเตอร์ของมุมกับแกนพิกัด ซึ่งระบุทิศทางที่ออกจากจุด  $\theta$

$\nabla_{\alpha}$  คือ อนุพันธ์ (directional derivative) หรือเกรเดียนต์ (gradient) ในทิศทาง  $\alpha$  สมการ (2.30) แสดงสารสนเทศของข้อสอบข้อหนึ่ง ณ ตำแหน่งหนึ่งในพื้นที่  $\theta$  ดังนั้นจึงไม่มีตัวห้อยของข้อสอบและผู้สอบในสมการ

อนุพันธ์ (directional derivative) สำหรับพื้นผิวการตอบสนองข้อสอบเป็นดังสมการ

$$\nabla_{\alpha} P(\theta) = \frac{\partial P(\theta)}{\partial \theta_1} \cos \alpha_1 + \frac{\partial P(\theta)}{\partial \theta_2} \cos \alpha_2 + \dots + \frac{\partial P(\theta)}{\partial \theta_m} \cos \alpha_m \quad (2.31)$$

ถ้าโมเดล MIRT คือ โมเดล multidimensional extension of the two-parameter logistic model อนุพันธ์เป็นดังสมการ

$$\nabla_{\alpha} P(\theta) = a_1 P(\theta) Q(\theta) \cos \alpha_1 + a_2 P(\theta) Q(\theta) \cos \alpha_2 + \dots + a_m P(\theta) Q(\theta) \cos \alpha_m \quad (2.32)$$

สมการข้างต้นสามารถเขียนได้อีกแบบ คือ

$$\nabla_{\alpha} P(\theta) = P(\theta) Q(\theta) \sum_{v=1}^m a_v \cos \alpha_v \quad (2.33)$$

แทนสมการ (2.32) ลงในสมการ (2.29) จะได้ว่า

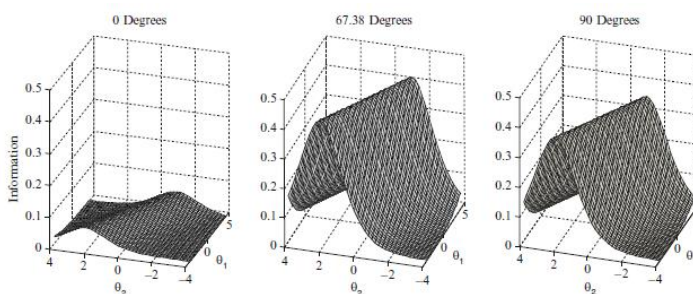
$$I_\alpha(\theta) = \frac{[P(\theta)Q(\theta)\sum_{v=1}^m a_v \cos\alpha_v]^2}{P(\theta)Q(\theta)} = P(\theta)Q(\theta)(\sum_{v=1}^m a_v \cos\alpha_v)^2 \quad (2.34)$$

เมื่อโมเดล MIRT มีเพียง 2 มิติ สารสนเทศในทิศทางระบุด้วยเวกเตอร์  $\alpha$  อธิบายได้ด้วยพื้นผิวสารสนเทศ (information surface) ส่วนความสูงของพื้นผิวเหนือระดับ  $\theta$  แทนขนาดของสารสนเทศ แต่ละตำแหน่งในระดับ  $\theta$  พื้นผิวสารสนเทศสำหรับข้อสอบอธิบายสำหรับโมเดล M2PL ดังภาพ 2.3 ซึ่งพื้นผิวดังกล่าวแสดงสารสนเทศใน 3 ทิศทาง ได้แก่  $0^\circ$ ,  $67.38^\circ$  และ  $90^\circ$  กับแกน  $\theta_1$

ถ้าแทนที่  $\cos\alpha_k$  ในสมการ (2.33) ได้สารสนเทศของข้อสอบในทิศทางของความชันสูงสุดดังสมการ

$$I_{\alpha max}(\theta) = P(\theta)Q(\theta)\sum_{v=1}^m a_k^2 = P(\theta)Q(\theta)A^2 \quad (2.35)$$

เมื่อ  $A$  คือ ค่าอำนาจจำแนกแบบพหุมิติสำหรับข้อสอบ



ภาพ 2.5 พื้นผิวสารสนเทศสำหรับโมเดล M2PL

เมื่อค่า  $a_1 = 1.2$ ,  $a_2 = 1.2$  และ  $d = -0.6$  ใน 3 มิติ

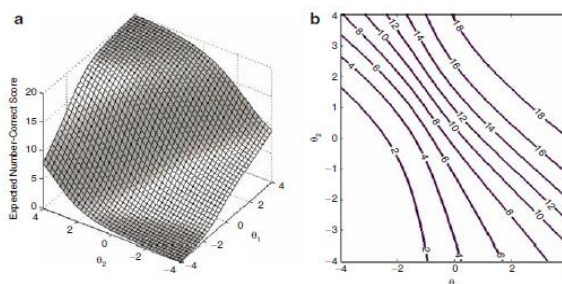
### 1.2.3) ฟังก์ชันของแบบสอบ (test functioning)

คุณลักษณะของแบบสอบใน MIRT อธิบายได้ด้วยโค้งคุณลักษณะของแบบสอบ (test characteristic curve: TCC) โดยโค้งคุณลักษณะของแบบสอบ คือ สมการถดถอยของผลรวมของคะแนนสอบบนความสามารถ  $\theta$  และสมการการถดถอยดังกล่าวสามารถใช้กับแบบสอบพหุมิติ ดังนั้นพื้นผิวคุณลักษณะของแบบสอบ (test characteristic surface: TCS) คือ สมการการถดถอยของผลรวมของคะแนนสอบบนเวกเตอร์ความสามารถ  $\theta$  สมการสำหรับพื้นผิวคุณลักษณะของแบบสอบสำหรับข้อสอบที่ตรวจให้คะแนนสองค่าเป็นดังสมการ

$$E(y_i|\theta_j) = E(\sum_{i=1}^n u_{ij}|\theta_j) = \sum_{i=1}^n E(u_{ij}|\theta_j) = \sum_{i=1}^n P(u_{ij}|\theta_j) \quad (2.36)$$

พื้นผิวคุณลักษณะของแบบสอบ (TCS) คือ ผลรวมของพื้นผิวคุณลักษณะของข้อสอบสำหรับแบบสอบที่ตรวจให้คะแนนหลายค่าหรือแบบสอบรูปแบบผสม สมการพื้นผิวคุณลักษณะของแบบสอบยังเหมือนกับสมการ (2.36) ยกเว้นสมการเทอมขวาสุด โดย TCS สำหรับแบบสอบที่ตรวจให้คะแนนหลายค่าหรือแบบสอบรูปแบบผสม คือ ผลรวมของคะแนนที่คาดหวังในข้อสอบบนเวกเตอร์  $\theta$  ซึ่ง TCS ของแบบสอบสามารถอธิบายได้ดังภาพ 2.6





ภาพ 2.6 พื้นผิวโค้งคุณลักษณะของแบบสอบ

a แสดงพื้นผิวและพล็อต b แสดง Equal score contours สำหรับพื้นผิว

## สรุป

คุณลักษณะของข้อสอบและแบบสอบสำหรับโมเดล MIRT สามารถอธิบายด้วยปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบ โมเดล MIRT พิจารณาฟังก์ชันของแบบสอบในพื้นที่หลายมิติจึงต้องระบุทิศทางในพื้นที่ว่าแบบสอบใดมีอำนาจจำแนกดีที่สุดในค่าอำนาจจำแนกแทนด้วยลูกศรหรือเวกเตอร์ในพื้นที่พหุมิติ ส่วนสารสนเทศของข้อสอบและแบบสอบใช้ประมาณตำแหน่งของผู้สอบในพื้นที่  $\theta$  โดยสารสนเทศเกี่ยวกับตำแหน่งของผู้สอบในพื้นที่  $\theta$  มีความแตกต่างกันขึ้นอยู่กับทิศทางที่ออกจากตำแหน่ง

### 1.3 การประมาณค่าพารามิเตอร์ของข้อสอบ-ผู้สอบ และการแปลงค่าพารามิเตอร์ไประบบพิกัด

ในส่วนนี้เป็นการอธิบายการประมาณเวกเตอร์  $\theta$  ด้วยค่าพารามิเตอร์ของข้อสอบที่ทราบค่า และการแปลงค่าพารามิเตอร์ไปยังระบบพิกัด

#### 1.3.1) การประมาณเวกเตอร์ $\theta$ ด้วยค่าพารามิเตอร์ของข้อสอบที่ทราบค่า

การประมาณค่าพารามิเตอร์ในโมเดล MIRT มีความท้าทายอยู่หลายประการ (1) โมเดลประกอบด้วยพารามิเตอร์ผู้สอบและข้อสอบ และโดยทั่วไปไม่สามารถประมาณค่าพารามิเตอร์สองชุดอย่างเป็นอิสระจากกันได้ จึงต้องประมาณพารามิเตอร์ข้อสอบและผู้สอบร่วมกัน (2) ต้องประมาณพารามิเตอร์จำนวนมาก อย่างเช่น โมเดลที่สามารถทดแทนได้ ถ้า  $m$  คือ จำนวนของแกนร่วม  $n$  คือ จำนวนข้อสอบ และ  $N$  คือ จำนวนผู้สอบ ดังนั้นพารามิเตอร์ที่ต้องประมาณทั้งหมด คือ  $n(m + 1) + m \times n$  ประการที่สาม คือ ลักษณะ Indeterminacies ในโมเดล อย่างเช่น ตำแหน่งของจุดกำเนิดในพื้นที่ หน่วยของการวัดสำหรับแต่ละแกนร่วมและแนวโน้มของแกนร่วมซึ่งสัมพันธ์กับตำแหน่งของผู้สอบ ปัญหาเหล่านี้สามารถแก้ไขด้วยโปรแกรมสำหรับการประมาณค่าพารามิเตอร์ (Reckase, 2009)

การประมาณค่าพารามิเตอร์ข้อสอบต้องวิเคราะห์ด้วยโปรแกรมที่มีคุณภาพสูงและใช้กลุ่มตัวอย่างขนาดใหญ่และเหมาะสม มีข้อตกลงเบื้องต้น ได้แก่ ค่าพารามิเตอร์ของข้อสอบไม่มีความคลาดเคลื่อนและไม่มีปัญหาในการประมาณตำแหน่งของผู้สอบ ผู้สอบถูกเลือกและแบบสอบนำไปใช้กับผู้สอบกลุ่มนั้นเพื่อประมาณตำแหน่งของผู้สอบในพื้นที่  $\theta$  ผลจากการทดสอบ คือ ผลการตอบใน

แต่ละข้อ และสารสนเทศสำหรับการประมาณตำแหน่งของผู้สอบ คือ String ของคะแนนข้อสอบและค่าพารามิเตอร์ของข้อสอบ เมื่อคำนวณค่าประมาณตำแหน่งได้แล้วต้องพิจารณาว่าเป็นค่าประมาณที่ดีหรือไม่ เกณฑ์ที่ใช้ในการพิจารณา ได้แก่ Maximum likelihood, Maximum a posteriori Bayesian และ Least squares การประมาณตำแหน่งของผู้สอบเริ่มด้วยการใช้เกณฑ์ Maximum likelihood หลังจากนั้นใช้เกณฑ์ Maximum a posteriori Bayesian และ Least squares และเปรียบเทียบกับ Maximum likelihood

เกณฑ์ Maximum likelihood เป็นเกณฑ์เบื้องต้น การประมาณตำแหน่งของผู้สอบคือเวกเตอร์  $\theta$  ซึ่งเป็นผลของความน่าจะเป็นสูงสุดสำหรับ String ที่สังเกตได้ของคะแนนสำหรับเวกเตอร์  $\theta$  ที่เป็นไปได้ เทอมของ Likelihood ใช้สำหรับความน่าจะเป็นของ String ของคะแนนสำหรับเวกเตอร์  $\theta$  ที่เป็นไปได้ เวกเตอร์  $\theta$  เฉพาะที่มีค่า Likelihood สูงกว่าเวกเตอร์  $\theta$  อื่นๆ คือค่าประมาณ Maximum likelihood สำหรับชุดข้อสอบและ String ของคะแนนสอบนั่นเอง และโดยทั่วไป Likelihood สำหรับ String ของตำแหน่งเฉพาะในพื้นที่  $\theta$  เป็นดังสมการ

$$L(\mathbf{U}_j|\theta_j) = \prod_{i=1}^n P(u_{ij}|\theta_j)^{u_{ij}} Q(u_{ij}|\theta_j)^{1-u_{ij}} \quad (2.37)$$

เมื่อ  $L(\mathbf{U}_j|\theta_j)$  คือ Likelihood ของ String คำตอบ  $\mathbf{U}_j$  สำหรับผู้สอบ  $j$  ตำแหน่งที่  $\theta_j$

$u_{ij}$  คือ คะแนนของข้อสอบข้อที่  $i$  สำหรับผู้สอบ  $j$  บนข้อสอบ  $n$  ข้อ

เกณฑ์ Bayesian อาศัยทฤษฎีความสัมพันธ์ระหว่างความน่าจะเป็นอย่างมีเงื่อนไขของ Bayes ปี 1963 โดยทฤษฎีของ Bayes เป็นดังสมการ

$$P(A_i|E) = \frac{P(E|A_i)P(A_i)}{\sum_j P(E|A_j)P(A_j)} \quad (2.38)$$

เมื่อ  $E$  คือ เหตุการณ์ที่สังเกตได้

$A_i$  คือ สาเหตุที่เป็นไปได้ทั้งหมดสำหรับเหตุการณ์นั้น

$\Sigma$  คือ ผลรวมของสาเหตุทั้งหมดสำหรับเหตุการณ์นั้น

เกณฑ์ Least squares เป็นความแตกต่างระหว่างค่าสังเกตได้และค่าที่คาดหวังยกกำลังสอง โดยค่า Least squares ใช้พิจารณาเมื่อมีการประมาณ  $\theta$  เพื่อหาเวกเตอร์  $\theta$  ที่มีค่าความแตกต่างดังกล่าวน้อยที่สุด ดังสมการ

$$SS_\theta = \sum_{i=1}^n (u_i - P(u_i = 1|\theta, \mathbf{a}_i, d_i))^2 \quad (2.39)$$

เมื่อ  $SS_\theta$  คือ ผลรวมของความแตกต่างยกกำลังสองสำหรับค่าเฉพาะของ  $\theta$

$u_i$  คือ คะแนน 0 หรือ 1 สำหรับข้อสอบข้อที่  $i$

ผลรวมของความแตกต่างยกกำลังสองสามารถคำนวณได้ในแต่ละจุด  $\theta$  โดยจุด  $\theta$  ที่มีค่า  $SS_\theta$  น้อยที่สุด คือ ตัวประมาณยกกำลังที่น้อยที่สุดสำหรับ  $\theta$

## โปรแกรมคอมพิวเตอร์สำหรับประมาณค่าพารามิเตอร์ในโมเดล MIRT สำหรับแบบสอบพหุมิติ

### 1) BMIRT

BMIRT (Bayesian Multivariate Item Response Theory) เป็นโปรแกรมที่อยู่ในชุดเครื่องมือ BMIRT (BMIRT toolkit) สำหรับประมาณค่าพารามิเตอร์สำหรับโมเดล MIRT โดยใช้วิธี Markov chain Monte Carlo (MCMC) โดยวิธี MCMC เป็นการจำลองกลุ่มตัวอย่างจากการแจกแจง (posterior distribution) ของค่าพารามิเตอร์ข้อสอบและผู้สอบซึ่งได้จากเมทริกซ์การตอบข้อสอบที่สังเกตได้ ซึ่งใช้ Metropolis–Hastings sampling algorithm ประมาณค่าพารามิเตอร์ข้อสอบ ผู้สอบ และการแจกแจงของประชากรแบบเอกมิตีและพหุมิติ ซึ่งได้ทั้งผู้สอบกลุ่มเดียวหรือผู้สอบหลายกลุ่ม โปรแกรมนี้สามารถประมาณค่าพารามิเตอร์แบบพร้อมกัน (concurrent calibration) สำหรับผู้สอบหลายกลุ่มได้ (ไม่ว่าจะเป็นแต่ละกลุ่มมีขนาดเท่ากัน แต่ละกลุ่มมีขนาดแตกต่างกัน แต่ละกลุ่มมีขนาดต่างกันและมีการแจกแจงแตกต่างกัน) ซึ่งมีงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลที่ใช้โปรแกรม BMIRT ประมาณค่าพารามิเตอร์แบบพร้อมกันสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า และข้อสอบรูปแบบผสม ซึ่งผู้วิจัยจะกล่าวในส่วนต่อไป

โปรแกรมนี้มีความสอดคล้องกับโมเดลสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ตรวจให้คะแนนหลายค่า และแบบสอบรูปแบบผสม ไม่ว่าจะเป็น โมเดลโลจิสติกแบบ 3 พารามิเตอร์ และ 2 พารามิเตอร์, โมเดล Rasch, โมเดล Generalized two-parameter partial credit, โมเดล Grade response และโมเดล Higher-order IRT model โดยโปรแกรมนี้สามารถกำหนดค่าพารามิเตอร์สำหรับข้อสอบร่วมและประมาณค่าพารามิเตอร์ข้อสอบที่ไม่ใช่ข้อสอบร่วม และสามารถคำนวณได้ทั้งคะแนนแต่ละโดเมนและคะแนน (Yao, 2006)

### 2) flexMIRT

โปรแกรม flexMIRT เวอร์ชัน 3 เป็นซอฟต์แวร์ล่าสุดสำหรับการวิเคราะห์ข้อสอบสำหรับโมเดล IRT แบบหลายระดับและหลายกลุ่ม โปรแกรม flexMIRT มีความเหมาะสม (fit) กับโมเดล UIRT และ MIRT กับข้อมูลระดับเดียวและข้อมูลหลายระดับโดยใช้ Maximum likelihood หรือ modal bayes ด้วย Bock-Atikin EM (ด้วย Generalized dimension reduction) หรือ MH-RM estimation algorithms และเนื่องจาก Generalized dimension reduction EM algorithms สามารถใช้กับเงื่อนไขบังคับของค่าพารามิเตอร์ที่กำหนดโดยผู้ใช้โปรแกรม โปรแกรมนี้ จึงมีความยืดหยุ่นและ MH-RM algorithm ยังทำให้โปรแกรมประมาณโมเดลแบบหลายมิติได้อย่างมีประสิทธิภาพ โปรแกรม flexMIRT สามารถให้คะแนนรวมไปจนถึงตารางคะแนนสเกล: EAP conversion table สำหรับโมเดล UIRT และ MIRT เมื่อพิจารณาการประมาณค่าพารามิเตอร์ข้อสอบ โปรแกรมนี้สามารถใช้กับแบบสอบรูปแบบผสมได้ เนื่องจากสามารถประมาณโมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PL) โมเดล Logistic graded response model (รวมถึง 2PL และ 1PL) และโมเดล Nominal categories model (อย่างเช่น

Generalized partial credit model, Partial credit model และ Rating scale model) ผสมกันได้ทั้ง ข้อมูลระดับเดียวและหลายระดับ (Houts & Cai, 2015)

โปรแกรม flexMIRT พัฒนาให้ใช้กับโมเดล MIRT และโมเดล Multilevel IRT ได้ดีขึ้น เนื่องจากสามารถประมาณความคลาดเคลื่อนมาตรฐานสำหรับค่าพารามิเตอร์ข้อสอบด้วยวิธีการที่หลากหลาย นอกจากนี้ยังสามารถวิเคราะห์ค่าสถิติความสอดคล้องของโมเดล (model fit statistics) สำหรับการวิเคราะห์มิติของข้อมูล การทดสอบความสอดคล้องของข้อสอบ และการวิเคราะห์ความปกติของตัวแปรแฝงได้ด้วย อีกทั้งยังสามารถใช้กับงานวิจัยเกี่ยวกับฟังก์ชันข้อสอบที่แตกต่างกัน (DIF) และการเชื่อมโยงสเกล (รวมถึงการสเกลในแนวตั้ง) โปรแกรมดังกล่าวยังให้ Monte Carlo simulation module เพื่อการจำลองข้อมูลจากโมเดลทั้งหมดใน flexMIRT นอกจากนี้โปรแกรม flexMIRT สามารถประมาณค่าพารามิเตอร์ได้แล้ว ยังสามารถจำลองข้อมูลจากโมเดล IRT รวมถึงโมเดลพหุมิติ โมเดลสององค์ประกอบ (bifactor models) และโมเดลหลายระดับ (multilevel models)

### 3) Package “mirt” ในโปรแกรม R

Package “mirt” เวอร์ชัน 1.19 ใช้สำหรับประมาณค่าพารามิเตอร์ของข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่า โดยใช้โมเดลคุณลักษณะแฝงแบบเอกมิตีและพหุมิติ โดยการประมาณโมเดลเชิงสำรวจและโมเดลเชิงยืนยันใช้วิธี Quadrature (EM) หรือ Stochastic (MHRM) แพคเกจนี้ยังสามารถวิเคราะห์องค์ประกอบสองตัวแปรและวิเคราะห์แบบสองระดับ (two-tier analysis) เพื่อสร้าง item testlets อีกทั้งยังสามารถวิเคราะห์แบบหลายกลุ่ม (multiple-group analysis) และออกแบบผลกระทบ (mixed effect designs) อย่างหลากหลายสำหรับตรวจสอบความแตกต่างระหว่างฟังก์ชันของข้อสอบและแบบสอบ รวมถึงการจำลองข้อมูลการตอบของผู้สอบได้อีกด้วย นอกจากนี้ยังสามารถวิเคราะห์โมเดล Latent class ต่างได้ เช่น DINA, DINO และ multidimensional latent class (Chalmers, 2012)

ฟังก์ชันที่ใช้ในการประมาณโมเดล MIRT มี 4 ฟังก์ชัน ได้แก่ mirt(), bfactor(), polymirt() และ confmirt() โดยฟังก์ชัน polymirt() และ confmirt() ใช้อัลกอริทึม MH-RM ฟังก์ชัน **mirt()** เหมาะกับโมเดลการวิเคราะห์องค์ประกอบสำหรับข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่า โดยใช้อัลกอริทึม MH-RM หรืออัลกอริทึม EM ในการใช้งานต้องระบุค่ารูปแบบการตอบข้อสอบ ซึ่งอาจอยู่ในรูปตาราง พร้อมทั้งระบุจำนวนองค์ประกอบของโมเดล ผลลัพธ์ที่ได้จากฟังก์ชัน คือ ค่าความสอดคล้องของโมเดล (เช่นค่า Log-likelihood, AIC, BIC, RMSEA และค่า p-value) ค่าพารามิเตอร์ข้อสอบ (ค่าความชัน: a, ค่าจุดตัด: d และโอกาสในการเดา) ค่าน้ำหนักองค์ประกอบ อีกทั้งยังมีการตรวจสอบความเหมาะสมของโมเดลด้วยคำสั่ง residual() ซึ่งคำนวณค่า Local dependence (LD) pairwise ระหว่างข้อสอบแต่ละคู่ โดยค่า LD มีความคล้ายคลึงกับค่าไค-สแควร์ ซึ่งค่าสถิติ LD เป็นค่า

ที่อยู่เหนือแนวทแยง โดยจะใช้ในการตีความเมื่อข้อสอบมีรูปแบบการตอบสนองมากกว่า 2 แบบและมีองค์อิสระจำนวนมาก

**ฟังก์ชัน bfactor()** ใช้ศึกษาโมเดลสององค์ประกอบ (bifactor model) จึงสามารถวิเคราะห์ได้สูงสุด 2 มิติ สำหรับข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่า และถ้าจำนวนขององค์ประกอบมากกว่าหนึ่ง จะสามารถหาค่าอำนาจจำแนกและจุดตัดแบบพหุตัวแปรได้ **ฟังก์ชัน polymirt()** ใช้ประมาณค่าพารามิเตอร์ของโมเดลด้วยอัลกอริทึม MH-RM โดยใช้กับการวิเคราะห์ข้อสอบเชิงสำรวจสำหรับข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่า ซึ่งฟังก์ชันนี้ใช้ได้กับข้อมูลพหุมิติระดับสูง และสามารถประมาณค่าโอกาสในการเดา (lower asymptotes) **ฟังก์ชัน confirmirt()** สามารถกำหนดเงื่อนไขบังคับของข้อสอบได้ โดยฟังก์ชันนี้ใช้เวลาในการประมาณค่ามากกว่าฟังก์ชัน mirt() และ bfactor() ในภาพรวม

แพ็คเกจนี้สามารถประมาณค่าต่าง ๆ ได้ ได้แก่ ค่าความยากแบบพหุมิติ (MDIFF) ค่าอำนาจจำแนกแบบพหุมิติ (MDIFF) ความคลาดเคลื่อนมาตรฐาน (bootstrapped standard error) สำหรับโมเดลที่ถูกประมาณ ค่าความสอดคล้องของข้อสอบ และยังสามารถแสดงภาพพื้นผิวของข้อสอบได้อีกด้วยกรณีที่มีค่าพารามิเตอร์สำหรับแบบสอบหลายฉบับพร้อมกันต้องใช้ฟังก์ชัน **multipleGroup ()** ซึ่งจะต้องกำหนดข้อสอบที่อยู่ในแต่ละกลุ่มก่อนการวิเคราะห์ข้อมูล

#### 1.4.2) การแปลงค่าพารามิเตอร์ไประบบพิกัด

Indeterminacy หมายถึง โครงสร้างอย่างยากที่ไม่สามารถคำนวณหาคำตอบโดยใช้สมการพื้นฐานทั่วไป ซึ่งโมเดล MIRT ก็มีปัญหาดังกล่าว อีกทั้งโมเดล MIRT ไม่มีระบบพิกัดมาตรฐาน ผู้วิจัยจึงพัฒนาโปรแกรมในการประมาณค่า ซึ่งแต่ละโปรแกรมมีการกำหนดระบบพิกัดเป็นของตัวเอง วิธีการปกติ คือ กำหนดจุดกำเนิดของพื้นที่ให้มีค่าเฉลี่ยของเวกเตอร์  $\theta$  เป็นเวกเตอร์ 0 การพัฒนาโปรแกรม MIRT อาจใช้ส่วนเบี่ยงเบนมาตรฐานของพิกัดร่วมกับแกนพิกัดเป็นหน่วยของการวัด ซึ่งวิธีการนี้มักพบบ่อยในโปรแกรมประมาณค่าพารามิเตอร์ (calibration) ของโมเดล UIRT โดยวิธีนี้เก็บหน่วยของการวัดที่มีค่า -3 ถึง 3 ไว้ นอกจากนี้ในการใช้โปรแกรมการประมาณค่าอาจกำหนดสหสัมพันธ์ระหว่างพิกัดเป็นศูนย์เพื่อให้สะดวกต่อการประมาณค่าทางสถิติ การระบุระบบพิกัดตั้งต้นอาจใช้การบังคับค่าสหสัมพันธ์แล้วทำการหมุนแกนเพื่อให้ได้โครงสร้างที่เหมาะสมสำหรับการวัด โดยการหมุนแกนกล่าวไม่ได้ทำให้ตำแหน่งของผู้สอบในพื้นที่เปลี่ยนไป

ระบบพิกัด (coordinate system) ใช้นำเสนอตำแหน่งของผู้สอบและคุณลักษณะของข้อสอบภายใต้โมเดล MIRT ซึ่งตำแหน่งของผู้สอบและคุณลักษณะของข้อสอบคำนวณจากพารามิเตอร์ของโมเดล MIRT ในส่วนนี้ผู้วิจัยเริ่มจากอธิบายการเปลี่ยนแปลงค่าพารามิเตอร์ของโมเดล MIRT ด้วยการเปลี่ยนระบบพิกัด ตามด้วยการอธิบายวิธีสำหรับการแปลงค่าประมาณพารามิเตอร์ไป

ยังระบบพิกัดที่ระบุไว้ โดยวิธีการเหล่านี้มีความสำคัญต่อการเทียบมาตรฐาน (calibration) จากแบบสอบต่างฉบับกันให้อยู่ในระบบพิกัดแบบพหุมิติเดียวกัน นอกจากนี้ยังอธิบายถึงการวิธีการเลือกระบบพิกัดที่เหมาะสม

การเลือกระบบพิกัดที่แตกต่างกันบนค่าพารามิเตอร์จากโมเดล MIRT ตำแหน่งของจุดกำเนิดการเลือกหน่วยของหน่วยการวัดร่วมกับแกน และแนวโน้มของแกนจะถูกนำเสนอเชื่อมโยงกับแต่ละประเภทของ indeterminacy ซึ่งเสนอไว้ในโมเดล MIRT เนื่องจากตำแหน่งของผู้สอบและคุณลักษณะของข้อสอบเป็นค่าคงที่ในทุกกรณี ดังนั้นความสามารถและคุณลักษณะอื่น ๆ ของผู้สอบจึงคงที่ และความไวของข้อสอบต่อคุณลักษณะของผู้สอบที่แตกต่างกันยังคงเหมือนเดิม โดยปฏิสัมพันธ์ระหว่างข้อสอบ-ผู้สอบคงที่ ณ ความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องเดียวกัน ซึ่งความไม่แปรเปลี่ยนของโมเดล MIRT แสดงให้เห็นว่าความน่าจะเป็นของการเลือกตอบไม่ได้เปลี่ยนแปลงไปตามการเปลี่ยนแปลงของระบบพิกัด และโครงสร้างอย่างยาก (indeterminacy) ของโมเดล MIRT แสดงให้เห็นว่าระบบพิกัดที่แตกต่างกันเป็นผลดีอย่างเท่าเทียม ซึ่งระบบพิกัดขึ้นอยู่กับการกำหนดจุดตั้งแต่ หน่วย และแนวโน้มของแกนของผู้ใช้งาน

### 1. การเลื่อนขนานของจุดกำเนิดของพื้นที่ความสามารถ $\theta$ (translation of the origin of the $\theta$ space)

การแปลง (conversion) ค่าพารามิเตอร์อย่างง่ายที่ความน่าจะเป็นของการตอบข้อสอบไม่แปรเปลี่ยน คือ การเลื่อนขนานของจุดกำเนิดของพื้นที่  $\theta$  แต่การเลื่อนขนานไม่เกี่ยวข้องกับภาระงานพิกัด การเลื่อนขนานของจุดกำเนิดของพื้นที่  $\theta$  มักเกิดขึ้นในขั้นตอนการประมาณค่าพารามิเตอร์ (calibration) ชุดของข้อสอบและการประมาณค่าพารามิเตอร์ของผู้สอบ ถ้าวิเคราะห์ข้อมูลด้วยโปรแกรม TESTFACT ค่าเฉลี่ยของการเทียบมาตรฐานถูกกำหนดเป็นเวกเตอร์ 0 ซึ่งช่วยจัดการกับปัญหา Indeterminacy ในการเลื่อนขนาน การแปลงแกนของตำแหน่งผู้สอบจากแกนชุดแรกไปยังแกนชุดที่สอง ดังสมการ

$$v_j = \theta_j + \delta \quad (2.43)$$

เมื่อ  $v_j$  คือ เวกเตอร์ของพิกัดสำหรับผู้สอบ  $j$  ในระบบพิกัดใหม่

$\theta_j$  คือ เวกเตอร์ของพิกัดสำหรับผู้สอบ  $j$  ในระบบพิกัดเก่า

$\delta$  คือ เวกเตอร์ของตำแหน่งจุดกำเนิดใหม่โดยใช้ระบบพิกัดเดิม

แม้ว่าค่าเฉลี่ยความสามารถสำหรับแต่ละพิกัดมีการเปลี่ยนแปลง แต่ส่วนเบี่ยงเบนมาตรฐาน และค่าสหสัมพันธ์ไม่เปลี่ยนแปลง เนื่องจากเป็นการเลื่อน (shift) อย่างคงที่ในแต่ละแกน โดยค่าเฉลี่ยที่เปลี่ยนแปลง คือ  $\delta$  ดังนั้นการกำหนดจำนวนของการเลื่อนในจุดกำเนิดระบบแกน ดังสมการ

$$\bar{\theta} - \bar{v} = \bar{\theta} - (\bar{\theta} - \bar{\delta}) = \bar{\delta} = \delta \quad (2.44)$$

ความแตกต่างของค่าเฉลี่ยความสามารถเท่ากับเวกเตอร์ของการเปลี่ยนพิกัดของจุดกำเนิด โดยค่าเฉลี่ยของ  $\delta$  คือ  $\delta$  เนื่องจากเป็นค่าคงที่

เมื่อวิเคราะห์ชุดข้อมูลที่แตกต่างกันด้วยโปรแกรม TESTFACT ขนาดของการเลื่อนจุดกำเนิดยังไม่รู้ค่าเพราะโปรแกรมจะตั้งจุดกำเนิดของพื้นที่เป็นเวกเตอร์ 0 และเมื่อพล็อตตำแหน่งของผู้สอบจากการประมาณค่าพารามิเตอร์ทั้งสองครั้งให้อยู่บนแกนเดียวกัน พบว่า ตำแหน่งของผู้สอบมีการเลื่อน (shift) แทนที่จุดกำเนิดจะมีการเลื่อน แต่ถ้าถือว่าตำแหน่งของผู้สอบไม่เปลี่ยนแปลง สามารถใช้สมการ (2.43) ในการกำหนดการเลื่อนตำแหน่งผู้สอบในจุดกำเนิดได้ การแปลงค่าพารามิเตอร์ข้อสอบไปยังระบบพิกัดด้วยจุดกำเนิดใหม่มีความยุ่งยากมากกว่าขั้นตอนการแปลงพารามิเตอร์ของผู้สอบเล็กน้อย หลังจากการแปลงไปยังระบบพิกัดด้วยจุดกำเนิดใหม่ในระบบพิกัดเริ่มต้น (initial coordinate system) พบว่าจุดในพื้นที่พิกัดควรมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องเท่ากัน คุณสมบัตินี้เรียกว่าความไม่แปรเปลี่ยนของโมเดล MIRT สมการระบบพิกัดเริ่มต้น เป็นดังสมการ (กรณีที่มี 2 มิติ)

$$a_{11}\theta_1 + a_{12}\theta_2 + d = 0 \quad (2.45)$$

แทนค่า  $\theta_v = v_v + \delta_v$  ลงในสมการ (2.45) เมื่อ  $v$  คือ ดัชนีสำหรับมิติ จะได้ว่า

$$a_{11}(v_1 + \delta_1) + a_{12}(v_2 + \delta_2) + d = a_{11}v_1 + a_{12}v_2 + d + a_{11}\delta_1 + a_{12}\delta_2 \quad (2.46)$$

สมการ (2.46) แสดงการเปลี่ยนจุดกำเนิดซึ่งมีการเปลี่ยนแปลงของค่าพารามิเตอร์  $d$  เพื่อคงไว้ซึ่งคุณสมบัติความไม่แปรเปลี่ยนของโมเดล MIRT โดยการเปลี่ยนแปลงของค่าพารามิเตอร์  $d$  เป็นการเลื่อนจุดกำเนิดที่ถ่วงน้ำหนักด้วยค่าพารามิเตอร์  $a$  ซึ่งสอดคล้องกับแกนพิกัด ในเทอมของเมทริกซ์เวกเตอร์ของพารามิเตอร์  $d$  ซึ่งก็คือ  $\mathbf{d}$  แปลงไปยังชุดแกนใหม่ โดย  $\mathbf{d}$  เป็นดังสมการ

$$\mathbf{\tilde{d}} = \mathbf{d} + \mathbf{a}\delta \quad (2.47)$$

คุณสมบัติความไม่แปรเปลี่ยนของโมเดล MIRT เกิดขึ้น ค่าพารามิเตอร์ของโมเดลมาจากระบบพิกัดเดียวกัน ความน่าจะเป็นของการตอบข้อสอบยังคงเหมือนเดิม หลังจากการเลื่อนในจุดกำเนิด ค่าพารามิเตอร์  $\mathbf{a}$  ไม่เปลี่ยนแปลง แต่ต้องแปลงค่าพารามิเตอร์ของผู้สอบและค่าพารามิเตอร์  $\mathbf{a}$  ไปยังระบบพิกัดใหม่ สมการในเอ็กซ์โปเนนเชียลของโมเดล  $\mathbf{a}\theta + 1\mathbf{d}$  และ  $\mathbf{a}v + 1\mathbf{d}$  ยังมีความน่าจะเป็นในการตอบข้อสอบเหมือนเดิมสำหรับทุกคนและทุกข้อ

การแปลงค่าพารามิเตอร์ที่ประมาณได้จากระบบพิกัดหนึ่งไปยังอีกพิกัดหนึ่งมีความสำคัญต่อการนำโมเดล MIRT ไปใช้ในการแก้ปัญหาทางการทดสอบ ไม่ว่าจะเป็นการปรับเทียบคะแนน การพัฒนาคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ถ้าทราบค่าพารามิเตอร์สำหรับผู้สอบคนเดียวกันจากระบบพิกัดทั้งสองระบบ การเลื่อนในจุดกำเนิดระหว่างระบบพิกัดทั้งสองสามารถคำนวณจากสมการ (2.48) อย่างไรก็ตามมักหาตำแหน่งของผู้สอบจากการประมาณค่าพารามิเตอร์ (calibration) ทั้งสองครั้งไม่ได้ แต่สามารถประมาณค่าพารามิเตอร์ข้อสอบสำหรับข้อสอบ

เดียวกันจากการประมาณค่าพารามิเตอร์ ทั้งสองครั้งจากระบบพิกัดที่แตกต่างกันทั้งสองระบบได้ จึงสามารถกำหนดการเลื่อนในจุดกำเนิดของระบบพิกัดจากชุดค่าพารามิเตอร์ข้อสอบที่ประมาณทั้งสองชุดได้ ถ้าค่าพารามิเตอร์  $\mathbf{a}$  สำหรับการประมาณค่าพารามิเตอร์ข้อสอบจากกลุ่มตัวอย่างทั้งสองกลุ่มเหมือนกันสามารถใช้ ConQuest วิเคราะห์ข้อมูลการตอบข้อสอบด้วยค่าพารามิเตอร์  $\mathbf{a}$  ที่ระบุไว้ หลังจากนั้นจึงใช้สมการ (2.47) เพื่อหา  $\delta$  แล้วนำผลที่ได้มาใช้ในสมการ (2.48)

$$(\mathbf{a}'\mathbf{a})^{-1}\mathbf{a}'(\mathbf{d} - \mathbf{d}') = \delta \quad (2.48)$$

ค่าจากสมการดังกล่าวสามารถใช้แปลงค่าพารามิเตอร์ข้อสอบสำหรับข้อสอบที่ไม่ได้อยู่ในชุดข้อสอบรวมและค่าพารามิเตอร์ของผู้สอบไปยังค่าในระบบพิกัดรวมได้

## 2) การหมุนแกนพิกัดของพื้นที่ความสามารถ $\theta$ (rotating the coordinate axes of $\theta$ -space)

ปัญหา Interdimancy ประเภทที่สองเกิดขึ้นในโมเดล MIRT แบบทดแทนได้ ซึ่งก็คือแนวโน้มของแกนพิกัด โดยการแปลงพิกัดในพื้นที่ไปยังชุดแกนพิกัดซึ่งถูกหมุนคำนวณจาก การคูณพิกัดตั้งต้นด้วยเมทริกซ์การหมุน (rotation matrix) เมทริกซ์การหมุนสำหรับข้อมูลสองมิติ คือ

$$\mathbf{Rot} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{bmatrix}$$

เมื่อ  $\alpha$  คือ จำนวนของดีกรีการหมุนในทิศตามเข็มนาฬิกา โดยการหมุนแบบตามเข็มนาฬิกา ค่า  $\alpha$  เป็นลบ เมทริกซ์ของพิกัด  $\theta \ n \times m$  และเมทริกซ์ของพารามิเตอร์  $\mathbf{a} \ i \times m$  ถูกคูณด้วย เมทริกซ์การหมุน เมทริกซ์การแปลงแบบตั้งฉาก (orthogonal transformation matrix) คือ Transpose ของเมทริกซ์นั้น ๆ ซึ่งเป็นส่วนกลับของตัวเอง จะได้ว่า  $\mathbf{Rot}\mathbf{Rot}' = \mathbf{Rot}'\mathbf{Rot} = \mathbf{I}$  เรียกว่าเมทริกซ์เอกลักษณ์ (identity matrix) จากผลดังกล่าวเอกโพเนนเชียลของโมเดลก่อนและหลังการหมุนมีค่าเอกลักษณ์ หลังจากนั้นทำการหมุนแกนอำนาจจำแนกและความสามารถของผู้สอบ การหมุนแกนตามเข็มนาฬิกาทำให้ตำแหน่งของผู้สอบและเวกเตอร์ข้อสอบถูกหมุนตามเข็มนาฬิกาสัมพันธ์กับแกน

การหมุนร่วมกับค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานและค่าสหสัมพันธ์ของพิกัด พบว่า ค่าเฉลี่ยของพิกัดมีการเปลี่ยนแปลง ส่วนค่าสหสัมพันธ์และส่วนเบี่ยงเบนมาตรฐานไม่มีการเปลี่ยนแปลง ส่วนการหมุนรอบจุดกำเนิดเดิม พบว่า ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานและค่าสหสัมพันธ์ มีการเปลี่ยนแปลง แม้ว่าระยะระหว่างตำแหน่งของผู้สอบไม่ได้เปลี่ยนไป ซึ่งผลจากการหมุนดังกล่าวสามารถกำหนดจากสหสัมพันธ์ระหว่างเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมก่อนและหลังการหมุน เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมหลังการหมุน คือ  $\mathbf{Rot} \Sigma \mathbf{Rot}'$  ซึ่งเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมมีค่าความแปรปรวนเท่ากันในแนวทแยง การหมุนจึงเป็นผลในเมทริกซ์



ความแปรปรวน-ความแปรปรวนร่วมเหมือนกัน และเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของประชากรจึงไม่เปลี่ยนแปลง เนื่องจากข้อมูลสุ่มมาจากการแจกแจงมาตรฐานแบบสองตัวที่มี  $\rho = 0$

การประมาณค่าพารามิเตอร์ข้อสอบและผู้สอบโดยใช้โปรแกรมการประมาณค่าพารามิเตอร์แบบพหุมิติ (MIRT calibration) อย่างเช่น TESTFACT โดยแนวโน้มของแกนถูกกำหนดด้วยการบังคับค่าและกระบวนการหาค่าที่เหมาะสมที่สุด แสดงให้เห็นว่าโปรแกรมประมาณค่าพารามิเตอร์ต่างกันให้ผลในระบบพิกัดไม่เหมือนกัน และแม้ว่าจะใช้โปรแกรมเดียวแต่มาจากกลุ่มตัวอย่างคนละกลุ่ม (แบบสอบฉบับเดียวกัน) ก็อาจทำให้แนวโน้มของแกนพิกัดแตกต่างกันจึงจำเป็นต้องพิจารณาวิธีการหมุนแกน โดยวิธีการหมุนที่นำมาใช้ได้ คือ วิธี Procrustes โดยวิธีการนี้ต้องทราบเมทริกซ์เป้าหมาย (target matrix) หลังจากนั้นจึงกำหนดการหมุนที่มีระยะห่างระหว่างจุดการหมุนน้อยที่สุด ซึ่งการหมุนด้วยวิธี Procrustes แบบตั้งฉากจากเมทริกซ์หนึ่งไปยังเมทริกซ์เป้าหมายได้จากการแยกค่าเอกฐาน (singular) ของ matrix product ของเมทริกซ์ทั้งสอง ถ้าเมทริกซ์เป้าหมาย คือ เมทริกซ์ของพิกัดตำแหน่งของผู้สอบ  $\theta$  และเมทริกซ์ทางเลือก (alternate matrix) คือ ชุดของพิกัดที่ถูกหมุน  $\nu$  และการแยกค่าเอกฐานของผลลัพธ์ เป็นดังสมการ

$$\text{svd}(\theta\nu) = \mathbf{USV} \quad (2.49)$$

เมื่อ  $\mathbf{S}$  คือ เมทริกซ์แนวทแยงที่มีค่าบวกในแนวทแยง  $m \times m$

$\mathbf{V}$  คือ เมทริกซ์ตั้งฉาก  $m \times m$

การหมุนด้วยวิธี Procrustes  $\mathbf{R}_p$  เพื่อแปลงค่าใน  $\nu$  ไปยังค่าของ  $\theta$  เป็นดังสมการ

$$\mathbf{R}_p = \mathbf{VU} \quad (2.50)$$

เนื่องจากเมทริกซ์การหมุนเดียวกันถูกใช้เพื่อแปลงทั้งเมทริกซ์  $\theta$  ของตำแหน่งผู้สอบและเมทริกซ์ของค่าพารามิเตอร์อำนาจจำแนก  $\mathbf{a}$  จึงสามารถใช้การแยกค่าเอกฐานของ Product ของเมทริกซ์  $\mathbf{a}$  เป้าหมาย และเมทริกซ์  $\mathbf{a}$  ซึ่งถูกหมุนเพื่อคำนวณเมทริกซ์การหมุนเดียวกัน ถ้าเมทริกซ์เป้าหมายของค่าพารามิเตอร์อำนาจจำแนก คือ  $\mathbf{a}$  และเมทริกซ์ของค่าพารามิเตอร์อำนาจจำแนกอื่น ๆ คือ  $\hat{\mathbf{a}}$  หลังจากนั้นแทนลงในสมการ (2.49) และ (2.50) จะได้ว่า

$$\text{svd}(\mathbf{a}\hat{\mathbf{a}}) = \mathbf{USV} \quad \text{และ} \quad \mathbf{R}_p = \mathbf{VU} \quad (2.51)$$

### 3) การเปลี่ยนหน่วยของแกนพิกัด (changing the units of the coordinate axes)

ประเภทที่ 3 ของ Indeterminacy ในโมเดล MIRT แบบทดแทนได้ คือ หน่วยของการวัดซึ่งใช้ร่วมกับแกนพิกัดแต่ละแกน โดยปกติหน่วยของการวัดถูกกำหนดด้วยการระบุขนาดของค่าพารามิเตอร์  $\mathbf{a}$  ในโปรแกรม อย่างเช่น ConQuest หรือกำหนดค่าส่วนเบี่ยงเบนมาตรฐานของพิกัดตำแหน่ง ซึ่งถ้าใช้ในโปรแกรม TESTFACT กำหนดเป็น 1.0 แสดงให้เห็นว่า ผลของการประมาณค่าพารามิเตอร์ที่แตกต่างกันของชุดข้อมูลการตอบเดียวกันอาจทำให้ขนาดของหน่วยแกนพิกัดมีความแตกต่างกัน

การแปลงพิกัดจากชุดของหน่วยหนึ่งไปยังอีกหน่วยหนึ่งต้องคูณด้วยค่าคงที่ กล่าวคือ นำเมทริกซ์  $\theta$  คูณด้วยเมทริกซ์แนวทแยง (diagonal matrix) ที่มีค่าคงที่อยู่ในแนวทแยง โดยสมการเมทริกซ์สำหรับการแปลงหน่วย ดังสมการ

$$v = \theta C \quad (2.52)$$

เมื่อ  $v$  คือ เมทริกซ์ของพิกัดหลังการแปลงหน่วยของพิกัด

$\theta$  คือ เมทริกซ์ของพิกัดตั้งต้น

$C$  คือ เมทริกซ์แนวทแยงของค่าคงที่ในการสร้างสเกล (scaling constants)

การคูณตัวแปรด้วยค่าคงที่ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าใหม่ เท่ากับค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานดั้งเดิมคูณด้วยค่าคงที่ วิธีการนี้เป็นการระบุค่าของการสเกลสำหรับแกนพิกัดแต่ละแกน โดยสัดส่วนของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าใหม่ต่อค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานดั้งเดิมควรกู้คืน (recover) ค่าคงที่ของการสเกล ดังสมการ

$$\frac{\bar{v}_w}{\theta_w} = \frac{s_{v_w}}{s_{\theta_w}} = c_w \quad (2.53)$$

เมื่อ  $c_w$  คือ ส่วนประกอบแนวทแยงของเมทริกซ์การสเกลที่สอดคล้องกับมิติ  $w$

ค่าพารามิเตอร์ของข้อสอบต้องถูกแปลงหน่วย (conversion) หลังจากมีการเปลี่ยนหน่วยสำหรับแกนพิกัดแล้ว เพื่อให้โมเดล MIRT ยังคงมีคุณสมบัติด้านความไม่แปรเปลี่ยน และมีการแปลงหน่วยของค่าพารามิเตอร์ที่เหมาะสมต้องพิจารณาเอกโพเนนเชียลเชิงเส้นของโมเดล MIRT แบบทดแทนได้ โดยค่าเอกโพเนนเชียลต้องเหมือนเดิมหลังจากการเปลี่ยนสเกลของความน่าจะเป็นในการตอบข้อสอบจากโมเดล กรณีของการเปลี่ยนหน่วยของสเกลพิกัด (coordinate scales) ค่าสเกลใหม่ คือ  $v_{jw} = c_w \theta_{jw}$  เมื่อ  $j$  คือ ดัชนีของผู้สอบ และ  $w$  คือ ดัชนีของมิติ จะได้ว่า  $\theta_{jw} = \frac{v_{jw}}{c_w}$  หลังจากนั้นแทนที่สมการด้วยตัวแปร  $\theta$  ในสมการเอกโพเนนเชียลของมิติ สำหรับกรณีข้อมูล 2 มิติ จะได้ว่า

$$a_1 \theta_1 + a_2 \theta_2 + d = a_1 \frac{v_1}{c_1} + a_2 \frac{v_2}{c_2} + d = \frac{a_1}{c_1} v_1 + \frac{a_2}{c_2} v_2 + d \quad (2.54)$$

จากสมการข้างต้นแสดงให้เห็นว่าค่าเอกโพเนนเชียลยังเหมือนเดิม ส่วนค่าพารามิเตอร์อำนาจจำแนก  $a_w$  ถูกหารด้วยค่าคงที่ของการสเกลที่สอดคล้องกัน โดยค่าพารามิเตอร์  $d$  ไม่เปลี่ยน ในเทอมของ matrix algebra ค่าพารามิเตอร์  $\mathbf{a}$  ในพื้นที่  $\theta$  ซึ่งมีหน่วยใหม่ของแกนพิกัดคำนวณจากการคูณค่าพารามิเตอร์  $\mathbf{a}$  ตั้งต้นด้วยส่วนกลับของเมทริกซ์ค่าคงที่ของการสเกล  $\mathbf{aC}^{-1}$

ค่าคงที่ของการสเกลประมาณได้จากชุดของค่าพารามิเตอร์ตั้งต้นและค่าพารามิเตอร์  $\mathbf{a}$  ที่แปลงหน่วยแล้ว ซึ่งมาจากพิกัด  $\theta$  จำนวน 2 ชุด ในทางทฤษฎี ค่าคงที่ของการสเกลคำนวณจากค่าพารามิเตอร์สองชุดสำหรับข้อสอบหนึ่งข้อ แต่การประมาณค่าพารามิเตอร์ของข้อสอบและการประมาณค่าคงที่ของการสเกลจะต้องใช้ค่าเฉลี่ยของค่าพารามิเตอร์เพื่อค่ามีความเสถียร ดังสมการ

$$c_w = \frac{\bar{a}_{ow}}{\bar{a}_{nw}} \quad (2.55)$$

เมื่อ  $c_w$  คือ ค่าคงที่ของสเกลสำหรับแกนพิกัด  $w$

$a_{ow}$  คือ ชุดของค่าพารามิเตอร์  $a$  ตั้งต้นสำหรับมิติ

$a_{nw}$  คือ ชุดของค่าพารามิเตอร์  $a$  สำหรับมิติ หลังจากเปลี่ยนสเกลแล้ว

**การแปลงค่าพารามิเตอร์โดยใช้การเลื่อนขนาน (translation) การหมุน (rotation) และการเปลี่ยนหน่วย (change of the units)**

Indeterminacy มี 3 ประเภท ได้แก่ การเลื่อนขนานของจุดกำเนิดของพื้นที่ความสามารถ  $\theta$  (translation of the origin of the  $\theta$  space) การหมุนแกนพิกัดของพื้นที่ความสามารถ  $\theta$  (rotating the coordinate axes of  $\theta$ -space) และการเปลี่ยนหน่วยของแกนพิกัด (changing the units of the coordinate axes) สำหรับโมเดล MIRT แบบทดแทนเชิงเส้น (linear compensatory model) ในส่วนนี้เป็นการอธิบายการเปลี่ยนแปลงของระบบพิกัดเมื่อใช้การแปลงทางเรขาคณิต 3 ประเภท ได้แก่ การเลื่อนจุดกำเนิด การหมุนแกน และการเปลี่ยนหน่วยบนแกนพิกัด มีข้อตกลงเบื้องต้น 4 ประการ ได้แก่ (1) ตำแหน่งของผู้สอบเชิงสัมพัทธ์ในพื้นที่ระบุด้วยโมเดล MIRT ไม่มีการเปลี่ยนแปลง มีเพียงระบบพิกัดซึ่งใช้อธิบายตำแหน่งของผู้สอบเท่านั้นที่เปลี่ยนแปลง (2) ประสิทธิภาพของข้อสอบสำหรับการแยกแยะระหว่างตำแหน่งของผู้สอบไม่เปลี่ยนแปลงการอธิบายเชิงสถิติของข้อสอบอาจเปลี่ยนแปลงเพราะว่าระบบพิกัดมีการเปลี่ยนแปลง แต่เป็นเพียงการอธิบายคุณลักษณะเดียวกันที่แตกต่างกัน (3) ผลจากข้อ 1 และ 2 ความน่าจะเป็นของการตอบข้อสอบที่สังเกตได้ของผู้สอบแต่ละคนในข้อสอบแต่ละข้อเชิงทำนายยังคงไม่เปลี่ยนแปลง ซึ่งก็คือ คุณสมบัติความไม่แปรเปลี่ยนของโมเดล MIRT และ (4) แกนพิกัดแบบตั้งฉากในทุกกรณี มีเพียงการหมุนแบบตั้งฉากที่มีระบบพิกัดเปลี่ยนแปลง

การเปลี่ยนระบบพิกัดต้องใช้การแปลงหน่วย (conversion) ที่หลากหลาย เริ่มจากการเปลี่ยนหน่วยของแกนพิกัดก่อน หลังจากนั้นจึงเลื่อนจุดกำเนิดในหน่วยใหม่ การเลื่อนพิกัดของจุดกำเนิดเชิงตัวเลขครั้งแรกใช้หน่วยตั้งต้นของแกนพิกัดและเปลี่ยนหน่วยในภายหลัง แม้ว่าใช้ค่าคงที่เป็นตัวคูณเดียวกันก็จะให้ผลของแกนพิกัดของจุดต่าง ๆ ในพื้นที่  $\theta$  แตกต่างกัน การหมุนของแกนพิกัดมักใช้จุดกำเนิดของพื้นที่  $\theta$  เป็นศูนย์กลางของการหมุน ดังนั้นการหมุนก่อนการเลื่อนจุดกำเนิดจะให้ผลลัพธ์แตกต่างกันมากกว่าการหมุนหลังจากการเลื่อนจุดกำเนิด ถ้ากำหนด (fixed) เมทริกซ์สำหรับการหมุนของผลลัพธ์ และการสเกลหน่วยของแกนพิกัด และกำหนดเวกเตอร์ของระยะห่างสำหรับการเลื่อนจุดกำเนิดในพื้นที่ สมการที่เป็นไปได้สำหรับการแปลงทางเรขาคณิตของพื้นที่  $\theta$  มี 6 สมการ แต่ละสมการให้ชุดของค่าพิกัดเชิงตัวเลขที่แตกต่างกันซึ่งแสดงถึงตำแหน่งของผู้สอบคนเดียวกัน และตำแหน่งของผู้สอบดังกล่าวไม่ได้เปลี่ยนแปลง มีเพียงแนวโน้มของแกนพิกัดและหน่วยของแกนที่เปลี่ยนไปเท่านั้น

สมการของการแปลงทางเรขาคณิตทั้ง 6 สมการ เป็นดังสมการ (2.56) ถึง (2.61) เมื่อ  $\mathbf{v}$  คือ เมทริกซ์ของพิกัดในระบบการแปลงทางเรขาคณิต และ  $\boldsymbol{\theta}$  คือ เมทริกซ์ของพิกัดตั้งต้น เมทริกซ์ **Rot** และ **C** และเวกเตอร์- $\delta$  มีนิยามเหมือนกันเมื่ออธิบายการเลื่อนในแกนพิกัด โดยที่ 1 ในสมการ คือ  $n \times 1$  เวกเตอร์ ของ 1s ใช้สร้างเมทริกซ์ซึ่งค่าการเลื่อนของจุดกำเนิดในระบบพิกัดมีค่าเอกลักษณ์ในแนวแถว

$$\mathbf{v} = (\boldsymbol{\theta} - \mathbf{1}\delta)\mathbf{RotC} \quad (2.56)$$

$$\mathbf{v} = (\boldsymbol{\theta} - \mathbf{1}\delta)\mathbf{CRot} \quad (2.57)$$

$$\mathbf{v} = (\boldsymbol{\theta}\mathbf{Rot} - \mathbf{1}\delta)\mathbf{C} \quad (2.58)$$

$$\mathbf{v} = (\boldsymbol{\theta}\mathbf{C} - \mathbf{1}\delta)\mathbf{Rot} \quad (2.59)$$

$$\mathbf{v} = \boldsymbol{\theta}\mathbf{RotC} - \mathbf{1}\delta \quad (2.60)$$

$$\mathbf{v} = \boldsymbol{\theta}\mathbf{CRot} - \mathbf{1}\delta \quad (2.61)$$

สมการของการแปลงทางเรขาคณิต (2.56), (2.58) และ (2.59) มีค่าสหสัมพันธ์เท่ากับ และสมการ (2.57), (2.59) และ (2.61) มีค่าสหสัมพันธ์เท่ากับ โดย 3 สมการแรกใช้ **RotC** เป็นตัวคูณกับ  $\boldsymbol{\theta}$  ขณะที่ 3 สมการชุดหลังใช้ **CRot** และเนื่องจากใช้การหมุนกับสเกลต่างกัน ผลที่ได้จึงแตกต่างกัน การแปลงทางเรขาคณิต **RotC** เป็นการหมุนบนพิกัด  $\boldsymbol{\theta}$  ตั้งต้นก่อนแล้วจึงแปลงหน่วยของแกนพิกัดของผลลัพธ์ ขณะที่ **CRot** แปลงหน่วยของแกนพิกัดก่อนแล้วจึงหมุนบนพิกัดที่ถูกแปลง โดยการแปลงทางเรขาคณิตของพิกัด  $\boldsymbol{\theta}$  ต้องทำร่วมกับการแปลงทางเรขาคณิตของค่าพารามิเตอร์ข้อสอบ เพื่อคงไว้ซึ่งความไม่แปรเปลี่ยนของโมเดล MIRT

#### 1.4 โครงสร้างของแบบสอบพหุมิติ

การใช้ทฤษฎีการตอบสนองข้อสอบต้องวิเคราะห์โครงสร้างของข้อมูลการตอบ โดยใช้การวิเคราะห์เชิงสำรวจหรือการวิเคราะห์เชิงยืนยัน การวิเคราะห์ข้อมูลการตอบข้อสอบเชิงสำรวจใช้เมื่อยังไม่มีสมมติฐานเกี่ยวกับโครงสร้างของข้อมูลการตอบที่แน่นอน ส่วนการวิเคราะห์เชิงยืนยันใช้เมื่อมีสมมติฐานเกี่ยวกับโครงสร้างดังกล่าวแล้ว โดยสมมติฐานกล่าวถึงจำนวนของมิติพิกัด (coordinate dimensions) ที่จำเป็นต่อการจำลองข้อมูลและความสัมพันธ์ของพื้นผิวคุณลักษณะของข้อสอบ (item characteristic surface) กับแกนพิกัด การวิเคราะห์เชิงยืนยันเป็นการกระบุความสัมพันธ์ของทิศทางการวัดที่ดีที่สุดของข้อสอบ (ทิศทางของอำนาจจำแนกสูงสุด) ด้วยแกนพิกัด จึงเป็นไปได้ที่จะมีสมมติฐานเกี่ยวกับความยากของแบบสอบ แต่สมมติฐานเหล่านี้ต้องได้รับการตรวจสอบ เนื่องจากค่าพารามิเตอร์ความยากเป็นพารามิเตอร์ที่ต้องประมาณค่าแบบอิสระ

### 1.4.1) โครงสร้างของแบบสอบพหุมิติ

โครงสร้าง (structure) คือ การรวมกันของโครงสร้างของแบบสอบและแกนพิกัด โดยแกนพิกัดระบุ ระนาบพิกัด ถ้าเวกเตอร์ของแบบสอบแต่ละฉบับ คือ ระนาบพิกัด 1 ระนาบหรือมากกว่า การรวมกันของ โครงสร้างของแบบสอบและแกนพิกัด เรียกว่า โครงสร้างอย่างง่าย (simple structure) รูปแบบของ องค์ประกอบที่สอดคล้องจะมีการกรอกข้อมูลมากกว่าศูนย์ในแต่ละแถว (Thurstone, 1947 cited in Reckase, 2009) โดยโครงสร้างอย่างง่าย (simple structure) ใช้กับแบบสอบที่มีค่าพารามิเตอร์  $\alpha$  ไม่ เป็นศูนย์เพียงมิติเดียวสำหรับข้อสอบแต่ละคน ส่วนโครงสร้างแบบไม่ซับซ้อน (approximate simple structure) ใช้เมื่อข้อสอบมีค่าพารามิเตอร์  $\alpha$  สูงเพียงมิติเดียวและมิติอื่น ๆ มีค่าใกล้เคียงศูนย์ ซึ่งหัวลูกศร ของข้อสอบขนานกับแกนพิกัด

#### 1) โครงสร้างอย่างง่าย (simple structure: SS)

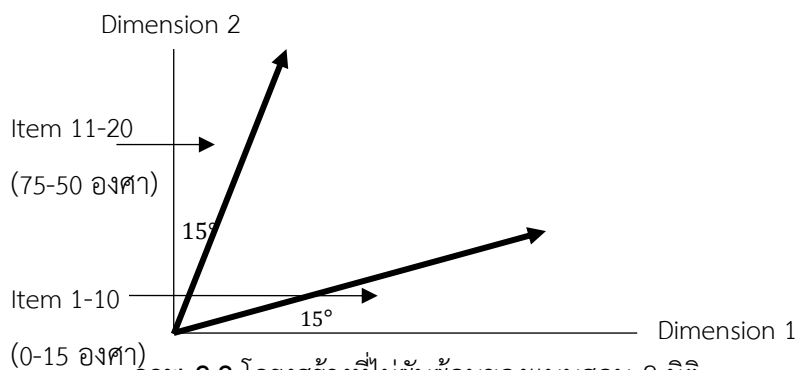
ลักษณะของโครงสร้างอย่างง่าย คือ ชุดข้อสอบสามารถแบ่งความแตกต่างของ ข้อสอบย่อยตามมิติได้เพียงมิติเดียว โดยมุมระหว่างเวกเตอร์ข้อสอบและมิติที่ศึกษาทำมุม 0 องศา หรือ 90 องศา แต่โครงสร้างนี้ไม่สอดคล้องกับสภาพจริงของการทดสอบ เพราะเป็นไปได้ยากที่ ข้อสอบข้อหนึ่งจะอยู่ในมิติใดมิติหนึ่งเพียงมิติเดียว มีลักษณะดังแผนภาพ 2.7 (Zhang & Stout, 1999 อ้างอิงใน พัทธี จันทรพิง, 2550)



ภาพ 2.7 โครงสร้างอย่างง่ายของแบบสอบ 2 มิติ

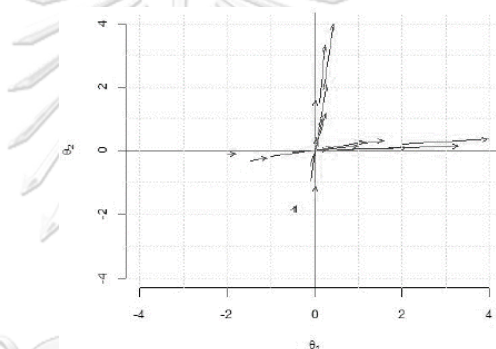
#### 2) โครงสร้างที่ไม่ซับซ้อน (approximate simple structure: APSS)

ลักษณะโครงสร้างที่ไม่ซับซ้อน คือ ข้อสอบแต่ละข้อของชุดข้อสอบมีความสัมพันธ์กัน สูงในมิติที่หนึ่งมากกว่ามิติที่สอง โดยค่าอำนาจจำแนกที่สูงในมิติเดียวกันแต่มีค่าอำนาจจำแนกต่ำในอีก มิติหนึ่ง อย่างเช่น งานวิจัยของพัทธิ จันทรพิง (2550) ซึ่งศึกษากับแบบสอบที่มีลักษณะ 2 มิติและมี ข้อสอบรวมในแต่ละระดับทั้งหมดมี 20 ข้อ โดยแต่ละมิติมีข้อสอบเท่ากัน 10 ข้อ ลักษณะเวกเตอร์ของ ข้อสอบในมิติที่ 1 ( $\theta_1$ ) ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 0 - 15 องศา นั่นคือ  $0^\circ \leq \alpha_k \leq 15^\circ$  เมื่อ  $\alpha_k$  คือ มุมระหว่างเวกเตอร์ในมิติที่ 1 กับแกนของมิติที่ 1 ส่วนลักษณะเวกเตอร์ของข้อสอบในมิติที่ 2 ( $\theta_2$ ) ทำมุม กับแกนมิติที่ 1 ตั้งแต่ 75 - 90 องศา นั่นคือ  $90^\circ - \alpha_k$  มีลักษณะดังแผนภาพ 2.8



ภาพ 2.8 โครงสร้างที่ไม่ซับซ้อนของแบบสอบ 2 มิติ

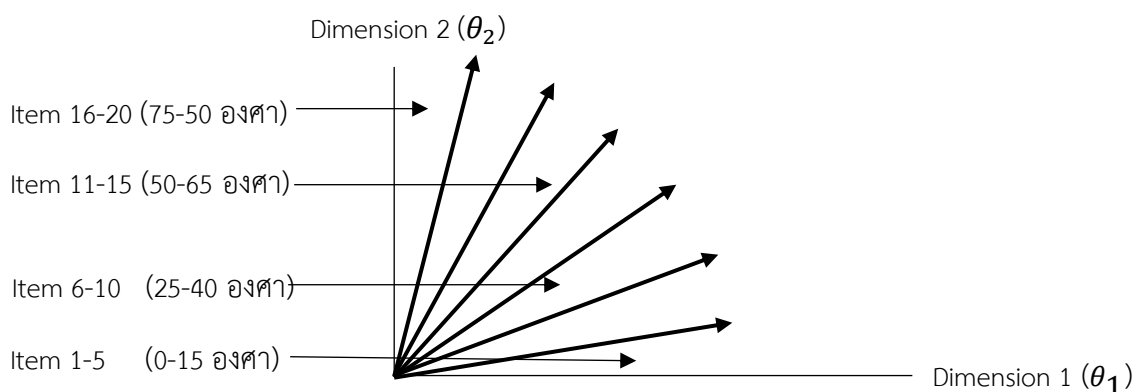
ทิศทางของเวกเตอร์ข้อสอบ (item vectors) มีลักษณะโครงสร้างที่ไม่ซับซ้อนเป็นดังภาพ 2.9 โดยแบบสอบชุดแรกมีลักษณะอยู่ในมิติที่ 1 มากกว่ามิติที่ 2 ส่วนข้อสอบชุดที่ 2 ใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 (พัชรี จันทร์เพ็ง, 2550)



ภาพ 2.9 เวกเตอร์ข้อสอบ (item vectors) สำหรับโครงสร้างที่ไม่ซับซ้อน

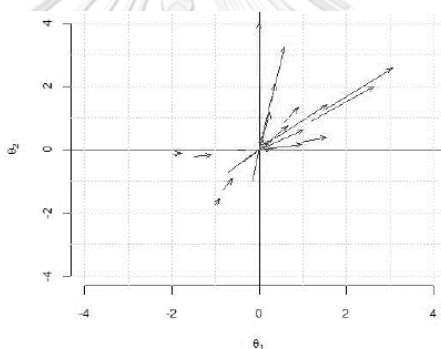
### 3) โครงสร้างที่ซับซ้อน (complex structure: CS)

ลักษณะของโครงสร้างที่ซับซ้อนมากขึ้น คือ โครงสร้างประกอบด้วยหลายมิติ โครงสร้างดังกล่าวมีลักษณะคล้ายคลึงกับโครงสร้างที่ไม่ซับซ้อน คือ ลักษณะโครงสร้างมีการวัดในแต่ละมิติและมีการวัดมิติโดยภาพรวม แต่ขนาดของมุมระหว่างข้อสอบกับแกนมิติต่างกันไป อย่างเช่น งานวิจัยของพัชรี จันทร์เพ็ง (2550) แบบสอบย่อย 2 ชุดแรก ชุดละ 5 ข้อ (ข้อที่ 1-5) แยกออกเป็นมิติที่ 1 มุม 0-15 องศา และมิติที่ 2 (ข้อที่ 6-10) มุม 75-90 องศา ส่วนแบบสอบย่อยอีก 2 ชุดหลัง คือ ข้อที่ 6-10 และข้อที่ 11-15 วัดในมิติรวม โดยข้อที่ 6-10 มีลักษณะใกล้เคียงกับมิติที่ 1 มากกว่ามิติที่ 2 ส่วนข้อที่ 11-15 มีลักษณะใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 มีลักษณะดังแผนภาพ 2.10



ภาพ 2.10 โครงสร้างที่ซับซ้อนของแบบสอบ 5 มิติ

ทิศทางของเวกเตอร์ข้อสอบ (item vectors) มีลักษณะโครงสร้างที่ซับซ้อนเป็นดังภาพ 2.11 โดยความยาวของเวกเตอร์แสดงถึงระดับของค่าอำนาจจำแนก (MDISC) ส่วนระยะทางระหว่างจุดกำเนิดและจุดเริ่มต้นของเวกเตอร์แสดงถึงค่าความยาก (MDIFF) จากภาพ พบว่า เวกเตอร์ทั้งหมดมุ่งสู่จุดกำเนิดซึ่งอยู่ในควอดแดนท์ที่ 1 และ 3 ดังนั้นค่าพารามิเตอร์อำนาจจำแนกมีค่าบวก



ภาพ 2.11 เวกเตอร์ข้อสอบ (item vectors) สำหรับโครงสร้างซับซ้อน

#### 1.4.2) วิธีการกำหนดจำนวนมิติที่ต้องการ

วิธีการกำหนดจำนวนมิติใช้เพื่อยืนยันความถูกต้องของความสัมพันธ์ในเมทริกซ์ของคะแนนสอบ มีรายละเอียดดังนี้

##### 1. DIMTEST

วิธีการทั่วไปสำหรับการกำหนดมิติของข้อมูลจำนวน 1 มิติ หรือมากกว่า จำเป็นต้องยืนยันความถูกต้องของเมทริกซ์คะแนนสอบของโมเดล ซึ่งสามารถใช้โปรแกรม DIMTEST จัดการได้ ขั้นตอนของ DIMTEST ถือว่าปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบสามารถอธิบายได้ด้วยรูปแบบทั่วไปของโมเดล MIRT ซึ่งฟังก์ชันความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง คือ  $\mathbf{a}\theta + d$  และรูปแบบของโมเดล MIRT ถือว่าความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องเพิ่มขึ้นทางเดียวเมื่อส่วนประกอบของเวกเตอร์  $\theta$  เพิ่มขึ้น

วิธีการใน DIMTEST กำหนดทิศทางของค่าอำนาจจำแนกที่มากที่สุดสำหรับแบบสอบ สถิติเบื้องต้นที่ใช้ในวิธีการนี้ คือ ค่าที่คาดหวังของความแปรปรวนร่วมระหว่างข้อสอบบนคอมพิวเตอร์ของ ส่วนประกอบ  $\theta$  ซึ่งวัดได้ดีที่สุดในทิศทางของค่าอำนาจจำแนกที่มากที่สุด  $\theta_Y = \mathbf{a}_Y \theta$  เมื่อ  $\mathbf{a}_Y$  คือ เวกเตอร์ของพารามิเตอร์  $\mathbf{a}$  โดยทิศทางของการวัดที่ดีที่สุดเป็นดังสมการ (Reckase, 2009)

$$E[\text{cov}(U_i, U_j | \theta_Y)], \quad i \neq j \quad (2.62)$$

เมื่อ  $i$  และ  $j$  คือ ดัชนีสำหรับข้อสอบแต่ละคู่ในแบบสอบ

DIMTEST ทดสอบสมมติฐานหลัก คือ ความแปรปรวนร่วมเท่ากับศูนย์ ถ้าปฏิเสธ สมมติฐานศูนย์แสดงว่า ความแปรปรวนร่วมเท่ากับศูนย์ นั่นคือ ข้อมูลไม่ได้เป็นของโมเดลการ ตอบสนองข้อสอบแบบเอกมิติ โปรแกรมนี้สามารถวิเคราะห์ข้อสอบที่ตรวจให้คะแนนสองค่า (0, 1) มากที่สุด 120 ข้อ กับผู้สอบไม่เกิน 6,000 คน ซึ่งการวิเคราะห์ไม่ซับซ้อนมากนัก แต่ผู้ใช้ต้องใช้ ข้อสอบย่อยในแบบสอบเพื่อวัดมิติแฝงที่มีคุณลักษณะเด่นเดียวกันพร้อมทั้งเลือกวิธีการตรวจสอบ เช่น Substantive Judgement, Hierarchical Cluster Analysis หรือ DETECT เป็นต้น โดย DIMTEST ต้องป้อนชุดของข้อสอบอย่างน้อย 2 ส่วน ชุดแรก เรียกว่า Partitioned subtest (PT) เป็นชุดข้อมูลที่ ถือว่าวัดได้ดีที่สุด ซึ่งระบุโดยคอมพิวเตอร์  $\theta_Y$  โดยคะแนน (number-correct scores) ใช้คำนวณ ความแปรปรวนร่วมระหว่างข้อสอบ (inter-item covariance) ชุดที่สอง เรียกว่า Assessment test (AT) ซึ่งเป็นส่วนประกอบข้อสอบที่ถือว่าวัดได้ดีที่สุด (Reckase, 2009; อนันดา สันฐิตวิณิชย์, 2556)

## 2. DETECT

DETECT พัฒนาโดย Zhang และ Stout (1999) เป็นวิธีการทางสถิติสำหรับกำหนด จำนวนมิติ เพื่อแสดงความสัมพันธ์ในเมทริกซ์คะแนนสอบภายใต้สมมติฐานของโครงสร้างอย่างง่าย โดยประมาณ (Approximate simple structure) โปรแกรมที่ใช้จัดการกับวิธีนี้ คือ DETECT ย่อมาจาก Dimensionality evaluation to enumerate contributing traits โปรแกรม DETECT อาศัย ทฤษฎีเดียวกับ DIMTEST ซึ่งวิธีการนี้ไม่ใช่พารามเมตริก โดยวิธี DETECT เป็นการหากลุ่มข้อสอบที่มี ลักษณะเดียวกันซึ่งวัดได้ดีที่สุดในทิศทางของพื้นที่พหุมิติซึ่งแตกต่างมากกว่าทิศทางของการวัดที่ดี ที่สุดสำหรับแบบสอบทั้งหมด

ข้อตกลงเบื้องต้นของวิธี DETECT ได้แก่ ฟังก์ชันการตอบข้อสอบเพิ่มขึ้นในทางเดียว เมื่อพิกัด  $\theta$  เพิ่มขึ้น และความเป็นอิสระของข้อสอบ ซึ่งจะเป็นไปตามข้อตกลงเบื้องต้นนี้เมื่อการ กำหนดความน่าจะเป็นของ String ของคะแนนบน  $\theta$  มาจากความน่าจะเป็นของคะแนนสอบแต่ละ ข้อบน  $\theta$  นอกจากนี้ยังมีข้อตกลงเบื้องต้นที่ไม่เหมือนกับโมเดล MIRT สำหรับข้อสอบที่ตรวจให้ คะแนนสองค่า คือ ความเป็นอิสระของข้อสอบแบบคู่ (pairwise local independence) กำหนดไว้ ว่า  $\text{cov}(U_i, U_j | \theta = 0) = 0$  สำหรับ  $\theta$  ทั้งหมด และ  $i \neq j$  เมื่อ  $i$  และ  $j$  คือ จำนวนข้อสอบใน แบบสอบ นิยามของข้อตกลงเบื้องต้นนี้สอดคล้องกับการประมาณค่าที่ใช้ใน NOHARM (Reckase,



2009) โปรแกรมนี้สามารถวิเคราะห์ข้อสอบที่มีการตรวจให้คะแนนสองค่า (0, 1) มากที่สุด 120 ข้อกับผู้สอบไม่เกิน 6,000 คน (อนันดา สันฐิตินิพนธ์, 2556)

### 3. Pararell Analysis

Pararell Analysis เป็นวิธีการกำหนดคะแนนที่พัฒนาโดย Ledesma และ Valero-Mora ในปี 2007 วิธีการนี้ประกอบด้วย 2 ขั้นตอนหลัก ได้แก่ (1) การวิเคราะห์มิติด้วยโปรแกรม TESTFACT เพื่อคำนวณค่า eigenvalue:  $n$  ค่าแรกของเมทริกซ์สหสัมพันธ์ระหว่างข้อสอบ (2) จำลองชุดข้อมูลซึ่งไม่มีความสัมพันธ์ระหว่างข้อสอบ แต่มีสัดส่วนของการตอบข้อสอบถูกเหมือนกับข้อมูลจริงและกลุ่มตัวอย่างเหมือนกัน หลังจากนั้นวิเคราะห์เพื่อหาค่า eigenvalue โดยใช้การวิเคราะห์เดียวกับข้อมูลจริง และเนื่องจากการจำลองมีการแจกแจงความยากเหมือนกับข้อมูลจริง การวิเคราะห์ข้อมูลเหล่านี้จะมีแนวโน้มขององค์ประกอบความยากเหมือนกับข้อมูลจริง

### 4. Difference-Chi square

วิธี Difference-Chi square พัฒนาโดย Schilling และ Bock ในปี 2009 ผู้วิจัยใช้การทดสอบความแตกต่าง  $\chi^2$  เพื่อทดสอบความสอดคล้องของโมเดลด้วยมิติ  $m$  และ  $m + 1$  เพื่อกำหนดจำนวนแกนพิกัดที่จำเป็นสำหรับการสร้างความสัมพันธ์ในเมทริกซ์การตอบข้อสอบ โดยการทดสอบความแตกต่าง  $\chi^2$  ใช้โปรแกรม TESTFACT สำหรับเมทริกซ์การตอบสองข้อสอบด้วยมิติ  $m$  และ  $m + 1$  ซึ่งโปรแกรมจะให้ค่าสถิติความสอดคล้อง  $\chi^2$  goodness of fit เป็นดังสมการ

$$\chi^2 = 2 \sum_{i=1}^s r_i \log \hat{P}_i - 2 \sum_{i=1}^s r_i \log \hat{P}_i \quad (2.63)$$

เมื่อ  $\chi^2$  คือ สถิติความแตกต่าง  $\chi^2$

$s$  คือ จำนวนของ String ของคำตอบที่สังเกตได้ในเมทริกซ์คะแนนที่สังเกตได้

$r_l$  คือ ความถี่ของ String ของคำตอบที่สังเกตได้  $l$

$\hat{P}_l$  คือ ความน่าจะเป็น (marginal probability) ของ String ของคำตอบ  $l$  ซึ่งถูกประมาณด้วยมิติ  $m$

$\hat{P}_l$  คือ ความน่าจะเป็น (marginal probability) ของ String ของคำตอบ  $l$  ซึ่งถูก

ประมาณด้วยมิติ  $m + 1$

โดยระดับขององศาอิสระสำหรับความแตกต่าง  $\chi^2$  คือ จำนวนข้อสอบ  $n$  ลบจำนวนของมิติ นั่นคือ  $n - m$

### สรุป

การกำหนดจำนวนมิติมีความจำเป็นในการวิเคราะห์เมทริกซ์ของคำตอบ จำนวนของมิติถือว่าเป็นจำนวนแกนพิกัดตั้งฉาก แต่ไม่ใช่จำนวนโครงสร้างที่วัดในแบบสอบ โดยจำนวนของโครงสร้างเป็นจำนวนของคอมโพสิทออ้างอิงที่สอดคล้องกัน (corresponding reference) ซึ่งถูกระบุด้วยกลุ่มของข้อสอบอย่างมีความหมาย กลุ่มของข้อสอบควรมีคุณลักษณะใกล้เคียงมากกับข้อตกลงเบื้องต้นสำหรับการเชื่อมโยง (coherence) ในทางปฏิบัติ จำนวนของชุดความเชื่อมโยงของข้อสอบระบุถึงคะแนนย่อย (subscore) อย่างมีความหมาย โดยตำแหน่งของผู้สอบถูกระบุโดยคอมโพสิทออ้างอิงที่

สอดคล้องกัน ซึ่งต้องระบุด้วยองศาความถูกต้องหรือความเที่ยงที่มีประสิทธิภาพ โดยทั่วไปจำนวนของแกนพิกัดใช้ในการอธิบายโครงสร้างของข้อมูลจากปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบซึ่งน้อยกว่าหรือเท่ากับชุดข้อสอบที่เชื่อมโยงกัน (coherence item sets) และถ้าจำนวนของแกนพิกัดมีค่าน้อยกว่าจำนวนของชุดข้อสอบที่เชื่อมโยงกัน (coherence item sets) ทิศทางของคอมพิวเตอร์อ้างอิงที่อยู่ในชุดข้อสอบจะตกอยู่ในระนาบเดียวกัน

วิธีการกำหนดจำนวนแกนพิกัดซึ่งจำเป็นต่อการนำเสนอสหสัมพันธ์ในเมทริกซ์ข้อมูลมีหลายวิธี ในกรณีที่เหมาะสมคือ ข้อมูลสามารถจำลองได้ด้วยโมเดลการตอบสนองข้อสอบแบบเอกมิติ วิธี DIMTEST เป็นวิธีที่ดีสำหรับการทดสอบสมมติฐานดังกล่าว ถ้าปฏิเสธสมมติฐานจะดำเนินการวิเคราะห์ขั้นตอนสำหรับการกำหนดจำนวนแกนพิกัด เช่น จำนวนของ  $\theta_S$  เพื่อใช้ในการประมาณ ส่วนวิธีการระบุชุดของข้อสอบที่เชื่อมโยงกันใช้การจำแนกกลุ่ม (cluster analysis) ซึ่งมีอยู่หลากหลายวิธี

## 1.5 โปรแกรมที่ใช้จำลองข้อมูลพหุมิติและการกำหนดค่าพารามิเตอร์ของข้อสอบ

### 1.5.1) โปรแกรมที่ใช้จำลองข้อมูลแบบพหุมิติ

#### 1. Package “mvtnorm” ในโปรแกรม R

Package “mvtnorm” สามารถใช้คำนวณ (1) การแจกแจงปกติแบบหลายตัวแปร (multivariate) และความน่าจะเป็น t (2) Quantiles (3) ส่วนเบี่ยงและความหนาแน่นอย่างสุ่ม ปัจจุบันพัฒนาถึงเวอร์ชัน 1.0-5 งานวิจัยเกี่ยวกับการปรับเทียบคะแนน MIRT ที่ใช้งานโปรแกรมดังกล่าว คืองานวิจัยของ Zhang (2012) ฟังก์ชันพื้นฐานของแพ็คเกจนี้ คือ pmvmorm ซึ่งใช้คำนวณความน่าจะเป็นแบบปกติแบบพหุตัวแปร (multivariate normal probabilities) และ pmvt ใช้คำนวณความน่าจะเป็น t แบบพหุตัวแปร (multivariate t probabilities), ค่าเฉลี่ยที่กำหนดเอง, เมทริกซ์สหสัมพันธ์ และขอบเขตการอินทิเกรต (hyperrectangular integration regions) (Genz et al., 2016)

#### 2. Package “mirt” ในโปรแกรม R

Package “mirt” เวอร์ชัน 1.19 ใช้ประมาณค่าพารามิเตอร์ของข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่า โดยใช้โมเดลคุณลักษณะแฝงแบบเอกมิติและพหุมิติโดยการประมาณโมเดลเชิงสำรวจและโมเดลเชิงยืนยันด้วยวิธี Quadrature (EM) หรือ Stochastic (MHRM) แพ็คเกจดังกล่าวสามารถจำลองข้อมูลการตอบข้อสอบแบบพหุมิติสำหรับโมเดล MIRT แบบทดแทนได้และแบบทดแทนไม่ได้ โดยใช้คำสั่ง simdata() ซึ่งจะต้องระบุค่าความชัน ค่าจุดตัดและประเภทของข้อสอบ อย่างไรก็ตามไม่สามารถกำหนดค่ามุมระหว่างมิติได้

### 3. โปรแกรม MATLAB

งานวิจัยของศิวัชทัศน์ สุขสุวรรณ, พัชรี จันทร์เพ็ง และ สัจวรรณ ังดกระโทก (2555) ที่ศึกษาการเชื่อมโยงสเกลแบบพหุมิติ และงานวิจัยของพัชรี จันทร์เพ็ง (2550) ใช้โปรแกรม MATLAB ในการสร้างค่าพารามิเตอร์ที่แท้จริง และแบบแผนการตอบข้อสอบ โดยผู้วิจัยเลือกใช้ค่าอำนาจจำแนก (MDISC) และค่าความยาก (MDIFF) จากงานวิจัยของ Yon (2006) รวมถึงขนาดของมุมในแต่ละระดับซึ่งผู้วิจัยกำหนดให้เท่ากัน นอกจากนี้ยังใช้โปรแกรมดังกล่าวในการจำลองการแจกแจงความสามารถแบบปกติ 2 ตัวแปร

### 4. โปรแกรม WinGen

โปรแกรม WinGen ใช้จำลองค่าความสามารถและการตอบสนองข้อสอบที่สอดคล้องกับ (1) โมเดล IRT แบบตรวจให้คะแนนสองค่า (แบบ 1, 2 และ 3 พารามิเตอร์) และตรวจให้หลายค่า (อย่างเช่น Partial credit model, Generalized partial credit model, Graded response model, Rating scale model และ Nominal response model) (2) โมเดลที่ไม่ใช่พารามิเตอร์ (3) โมเดลการตอบสนองข้อมูลแบบพหุมิติ: โมเดล multidimensional compensatory models (มากที่สุด 5 มิติ) โดยโปรแกรม WinGen สามารถสร้างชุดค่าพารามิเตอร์ข้อสอบและชุดค่าพารามิเตอร์ความสามารถผู้สอบเพื่อสร้างข้อมูลการตอบข้อสอบจากการแจกแจงที่หลากหลาย ไม่ว่าจะเป็น การแจกแจงแบบปกติ การแจกแจงแบบเอกรูป การแจกแจงบีตา และการแจกแจง lognormal โปรแกรมดังกล่าวยังสามารถใช้ไฟล์ค่าพารามิเตอร์จากโปรแกรม BILOG-MG, Parscale และ MULTLOG (Han & Hambleton, 2014)

### 4. SimuMIRT

SimuMIRT เป็นโปรแกรมที่อยู่ในชุดเครื่องมือของ BMIRT (BMIRT toolkit) ซึ่งใช้ได้กับโมเดลการตอบสนองข้อสอบพหุมิติแบบแบบทดแทนได้ SimulateRwo1 ใช้สำหรับโมเดล IRT แบบทดแทนได้เมื่อรู้ค่าความสามารถของผู้สอบ ส่วน SimulatGRRwo ใช้สำหรับโมเดล GRM และ SimulateGRRwo1 ใช้สำหรับโมเดล GRM เมื่อทราบค่าความสามารถของผู้สอบ และใช้สำหรับโมเดล IRT แบบทดแทนไม่ได้ (Noncompensatory MIRT model) โดย SimulateNonCompensatoryRwo ใช้สำหรับโมเดล IRT แบบทดแทนไม่ได้ ส่วน SimulateNonCompensatoryRwo1 ใช้สำหรับโมเดล IRT แบบทดแทนไม่ได้เมื่อรู้ค่าความสามารถของผู้สอบ นอกจากนี้ยังใช้ได้กับโมเดลผลกระทบของผู้ประเมิน (rater effect models) โดย SimulateRaterRwo ใช้สำหรับโมเดล IRT แบบทดแทนได้ที่มีผลกระทบของผู้ประเมิน ส่วน SimulateRaterRwo1 ใช้สำหรับโมเดล IRT แบบทดแทนได้ที่มีผลกระทบของผู้ประเมินและเมื่อรู้ค่าความสามารถของผู้สอบ (Yao, 2016)

### 1.5.2) การกำหนดค่าพารามิเตอร์ของข้อสอบในการจำลองข้อมูลแบบพหุมิติ

การจำลองค่าพารามิเตอร์ที่แท้จริงต้องกำหนดค่าพารามิเตอร์รวม ได้แก่ ค่าอำนาจจำแนกรวม (MDISC) และค่าความยากรวม (MDIFF) ค่าความยากสำหรับโมเดล IRT ควรอยู่ในช่วง -2.5 ถึง 2.5 ส่วนค่าอำนาจจำแนกควรอยู่ในช่วง 0.5-2.5 (ศิริชัย กาญจนวาสี, 2555) เมื่อพิจารณางานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT ที่ใช้การจำลองข้อมูล พบว่า การปรับเทียบคะแนนแบบสอบรูปแบบผสมภายใต้รูปแบบผู้สอบกลุ่มสุ่มเลือกใช้ค่าอำนาจจำแนกรวมและค่าความยากรวมของ Roussos และคณะ (1998) มาปรับใช้ (E. Lee, 2013) การปรับเทียบคะแนนข้อสอบที่ตรวจให้คะแนนสองค่าภายใต้รูปแบบ NEAT เลือกใช้ค่าอำนาจจำแนกรวมและค่าความยากรวมของ Min (2003) มาปรับใช้ (O. Zhang, 2012) รายละเอียดดังตาราง 2.4 และเมื่อพิจารณาการเชื่อมโยงสเกลแบบพหุมิติที่ใช้การจำลองข้อมูล พบว่า งานวิจัยของพัชรี จันทร์เพ็ง (2550) และงานวิจัยของศิวทัศน์ (2555) และงานวิจัยของศิวทัศน์ สุขสุวรรณ และคณะ (2555) ทั้ง 2 งานเลือกใช้ค่าของ Yon (2006) โดยงานวิจัยทั้ง 4 งานศึกษาการเชื่อมโยงสเกลแนวตั้งภายใต้รูปแบบ NEAT จึงกำหนดค่าอำนาจจำแนกรวมและค่าความยากรวมสำหรับนักเรียนในระดับต่ำกว่าและสูงกว่า ส่วนงานวิจัยของ Wei (2008) ที่เชื่อมโยงสเกลระหว่างข้อสอบที่ตรวจให้คะแนนสองค่าเลือกใช้ ค่า MDISC, MDIFF และทิศทางจากงานวิจัยที่ผ่านมาของ Reckase (1985), Reckase (1997) และ Reckase & McKinley (1991) รายละเอียดดังตาราง 2.4 และสรุปค่า MDISC และ MDIFF ได้ดังตาราง 2.3

จากงานวิจัยการเชื่อมโยงสเกลและปรับเทียบคะแนนแบบสอบพหุมิติที่ใช้การจำลองข้อมูลเพื่อจำลองแบบแผนการตอบและค่าพารามิเตอร์ของข้อสอบ ผู้วิจัยส่วนใหญ่มักกำหนดค่าอำนาจจำแนก (MDISC) และค่าความยาก (MDIFF) ไว้ 5 ระดับ และ 10 เพื่อให้ครอบคลุมระดับหรือช่วงของค่าความยากทั้งหมด และเนื่องจากผู้วิจัยสนใจศึกษาการปรับเทียบคะแนนด้วยโมเดล MIRT ภายใต้รูปแบบ NEAT การเลือกใช้ค่าอำนาจจำแนกรวมและค่าความยากของ Min (2003) จึงมีความเหมาะสม

ตาราง 2.2 สรุปค่าพารามิเตอร์ในการจำลองข้อมูลแบบพหุมิติที่ใช้ในงานวิจัย

ค่า	MDISC			MDIFF				
-10.5				ระดับ 1				
-2.0								
-1.5							ระดับ 1	
-1.0				ระดับ 2			ระดับ 2	
-0.5					ระดับ 1			
0.0				ระดับ 3	ระดับ 2	ระดับ 1	ระดับ 3	
0.2	ระดับ 1							
0.4	ระดับ 2	ระดับ 1						
0.5						ระดับ 2		
0.6	ระดับ 3		ระดับ 1					
0.8	ระดับ 4	ระดับ 2	ระดับ 2					
0.9								
1.0	ระดับ 5		ระดับ 3	ระดับ 4	ระดับ 3		ระดับ 4	
1.2	ระดับ 6	ระดับ 3	ระดับ 4					
1.3			ระดับ 5					
1.4	ระดับ 7							
1.5						ระดับ 3	ระดับ 5	
1.6	ระดับ 8	ระดับ 4						
1.8	ระดับ 9							
2.0	ระดับ 10	ระดับ 5			ระดับ 4			
2.2					ระดับ 5			
2.5						ระดับ 4		
3.0						ระดับ 5		
10.5				ระดับ 5				

**ตาราง 2.3** การกำหนดค่าพารามิเตอร์ในการจำลองข้อมูลแบบพหุมิติสำหรับงานวิจัยเกี่ยวกับการเปรียบเทียบคะแนนและการเชื่อมโยงสเกลสำหรับแบบสอบพหุมิติ

ผู้วิจัย	บริบท	การกำหนดค่าพารามิเตอร์ในการจำลองข้อมูล																																				
<b>การเปรียบเทียบคะแนนสำหรับแบบสอบพหุมิติ</b>																																						
Lee (2013) หรือ Lee, Lee & Brennan (2014)	- Random group - DICHO - APSS	ค่าพารามิเตอร์สำหรับแบบสอบ Y (M2PL) ปรับจาก Roussos และคณะ (1998) ค่า MDIFF สำหรับแบบสอบ Y ถูกจำลองขึ้นมาอย่างสุ่มจากการแจกแจงเอกรูป ได้ค่าดังตารางค่าอำนาจจำแนกและความยาก ดังนี้ <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>ระดับ</th> <th>ช่วง MDIFF</th> <th>MDISC</th> <th>จำนวนข้อสอบจาก 30 ข้อ</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>-10.5 ถึง -1.25</td> <td>0.4</td> <td>4</td> </tr> <tr> <td>2</td> <td>-1.25 ถึง -0.50</td> <td>1.6</td> <td>6</td> </tr> <tr> <td>3</td> <td>-0.5 ถึง 0.50</td> <td>1.2</td> <td>10</td> </tr> <tr> <td>4</td> <td>0.50 ถึง 1.25</td> <td>0.8</td> <td>6</td> </tr> <tr> <td>5</td> <td>1.25 ถึง 10.50</td> <td>2.0</td> <td>4</td> </tr> </tbody> </table>	ระดับ	ช่วง MDIFF	MDISC	จำนวนข้อสอบจาก 30 ข้อ	1	-10.5 ถึง -1.25	0.4	4	2	-1.25 ถึง -0.50	1.6	6	3	-0.5 ถึง 0.50	1.2	10	4	0.50 ถึง 1.25	0.8	6	5	1.25 ถึง 10.50	2.0	4												
ระดับ	ช่วง MDIFF	MDISC	จำนวนข้อสอบจาก 30 ข้อ																																			
1	-10.5 ถึง -1.25	0.4	4																																			
2	-1.25 ถึง -0.50	1.6	6																																			
3	-0.5 ถึง 0.50	1.2	10																																			
4	0.50 ถึง 1.25	0.8	6																																			
5	1.25 ถึง 10.50	2.0	4																																			
		ค่าพารามิเตอร์สำหรับแบบสอบ X (M2PL) ปรับจาก Roussos และคณะ (1998) ค่า MDISC ของแบบสอบ Y เหมือนกับแบบสอบ X แต่ค่าเฉลี่ยของ MDIFF สำหรับแบบสอบ Y สูงกว่า ค่าอำนาจจำแนก และความยาก ดังตาราง <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>ระดับ</th> <th>ช่วง MDIFF (แตกต่างกัน้อย)</th> <th>MDIFF (แตกต่างกันมาก)</th> <th>MDSIC</th> <th>จำนวนข้อสอบจาก 30 ข้อ</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>-1.45 ถึง -1.20</td> <td>-1.30 ถึง -1.05</td> <td>0.4</td> <td>4</td> </tr> <tr> <td>2</td> <td>-1.20 ถึง -0.45</td> <td>-1.05 ถึง -0.30</td> <td>1.6</td> <td>6</td> </tr> <tr> <td>3</td> <td>-0.45 ถึง 0.55</td> <td>-0.30 ถึง 0.70</td> <td>1.2</td> <td>10</td> </tr> <tr> <td>4</td> <td>0.55 ถึง 1.30</td> <td>0.70 ถึง 1.45</td> <td>0.8</td> <td>6</td> </tr> <tr> <td>5</td> <td>1.30 ถึง 10.55</td> <td>1.45 ถึง 1.70</td> <td>2.0</td> <td>4</td> </tr> </tbody> </table>	ระดับ	ช่วง MDIFF (แตกต่างกัน้อย)	MDIFF (แตกต่างกันมาก)	MDSIC	จำนวนข้อสอบจาก 30 ข้อ	1	-1.45 ถึง -1.20	-1.30 ถึง -1.05	0.4	4	2	-1.20 ถึง -0.45	-1.05 ถึง -0.30	1.6	6	3	-0.45 ถึง 0.55	-0.30 ถึง 0.70	1.2	10	4	0.55 ถึง 1.30	0.70 ถึง 1.45	0.8	6	5	1.30 ถึง 10.55	1.45 ถึง 1.70	2.0	4						
ระดับ	ช่วง MDIFF (แตกต่างกัน้อย)	MDIFF (แตกต่างกันมาก)	MDSIC	จำนวนข้อสอบจาก 30 ข้อ																																		
1	-1.45 ถึง -1.20	-1.30 ถึง -1.05	0.4	4																																		
2	-1.20 ถึง -0.45	-1.05 ถึง -0.30	1.6	6																																		
3	-0.45 ถึง 0.55	-0.30 ถึง 0.70	1.2	10																																		
4	0.55 ถึง 1.30	0.70 ถึง 1.45	0.8	6																																		
5	1.30 ถึง 10.55	1.45 ถึง 1.70	2.0	4																																		
Zhang (2012)	- DICHO - NEAT	จำลองข้อมูล 2,000 ครั้ง สำหรับแต่ละเงื่อนไข โดยผู้วิจัยกำหนดค่าความอำนาจจำแนกและค่าความยากไว้ 10 ระดับดังนี้ <table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th>ระดับ</th> <th>MDISC</th> <th>MDIFF</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0.2</td> <td>-2.0</td> </tr> <tr> <td>2</td> <td>0.4</td> <td>-1.5</td> </tr> <tr> <td>3</td> <td>0.6</td> <td>1.0</td> </tr> <tr> <td>4</td> <td>0.8</td> <td>-0.5</td> </tr> <tr> <td>5</td> <td>1</td> <td>0.0</td> </tr> <tr> <td>6</td> <td>1.2</td> <td>0.0</td> </tr> <tr> <td>7</td> <td>1.4</td> <td>-1</td> </tr> <tr> <td>8</td> <td>1.6</td> <td>0.5</td> </tr> <tr> <td>9</td> <td>1.8</td> <td>1.5</td> </tr> <tr> <td>10</td> <td>2.0</td> <td>2.0</td> </tr> <tr> <td>ค่าเฉลี่ย</td> <td>1.1</td> <td>0.0</td> </tr> </tbody> </table>	ระดับ	MDISC	MDIFF	1	0.2	-2.0	2	0.4	-1.5	3	0.6	1.0	4	0.8	-0.5	5	1	0.0	6	1.2	0.0	7	1.4	-1	8	1.6	0.5	9	1.8	1.5	10	2.0	2.0	ค่าเฉลี่ย	1.1	0.0
ระดับ	MDISC	MDIFF																																				
1	0.2	-2.0																																				
2	0.4	-1.5																																				
3	0.6	1.0																																				
4	0.8	-0.5																																				
5	1	0.0																																				
6	1.2	0.0																																				
7	1.4	-1																																				
8	1.6	0.5																																				
9	1.8	1.5																																				
10	2.0	2.0																																				
ค่าเฉลี่ย	1.1	0.0																																				

ผู้วิจัย	บริบท	การกำหนดค่าพารามิเตอร์ในการจำลองข้อมูล																																							
		<p>ผู้วิจัยปรับค่าอำนาจจำแนกและค่าความยากจากงานวิจัยของ Min (2003) กำหนดค่าดังกล่าวไว้ 5 ระดับ ได้แก่</p> <table border="1"> <thead> <tr> <th>ระดับ</th> <th>MDISC</th> <th>MDIFF</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0.4</td> <td>-1.5</td> </tr> <tr> <td>2</td> <td>0.8</td> <td>1.0</td> </tr> <tr> <td>3</td> <td>1.2</td> <td>0.0</td> </tr> <tr> <td>4</td> <td>1.6</td> <td>-1.0</td> </tr> <tr> <td>5</td> <td>2.0</td> <td>1.5</td> </tr> <tr> <td>ค่าเฉลี่ย</td> <td>1.2</td> <td>0.0</td> </tr> </tbody> </table>	ระดับ	MDISC	MDIFF	1	0.4	-1.5	2	0.8	1.0	3	1.2	0.0	4	1.6	-1.0	5	2.0	1.5	ค่าเฉลี่ย	1.2	0.0																		
ระดับ	MDISC	MDIFF																																							
1	0.4	-1.5																																							
2	0.8	1.0																																							
3	1.2	0.0																																							
4	1.6	-1.0																																							
5	2.0	1.5																																							
ค่าเฉลี่ย	1.2	0.0																																							
<b>การเชื่อมโยงสเกลสำหรับแบบสอบพหุมิติ</b>																																									
Roussos และคณะ (1998 อ้างอิงใน Lee, 2013)	- DICHO	<p>ค่าอำนาจจำแนกรวม (MDISC) และค่าความยากรวม (MDIFF) เป็นดังตาราง</p> <table border="1"> <thead> <tr> <th>ระดับ</th> <th>MDISC</th> <th>MDIFF</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0.4</td> <td>-10.5</td> </tr> <tr> <td>2</td> <td>0.8</td> <td>1.0</td> </tr> <tr> <td>3</td> <td>1.2</td> <td>0.0</td> </tr> <tr> <td>4</td> <td>1.6</td> <td>-1.0</td> </tr> <tr> <td>5</td> <td>2.0</td> <td>10.5</td> </tr> <tr> <td>ค่าเฉลี่ย</td> <td>1.2</td> <td>0.0</td> </tr> </tbody> </table>	ระดับ	MDISC	MDIFF	1	0.4	-10.5	2	0.8	1.0	3	1.2	0.0	4	1.6	-1.0	5	2.0	10.5	ค่าเฉลี่ย	1.2	0.0																		
ระดับ	MDISC	MDIFF																																							
1	0.4	-10.5																																							
2	0.8	1.0																																							
3	1.2	0.0																																							
4	1.6	-1.0																																							
5	2.0	10.5																																							
ค่าเฉลี่ย	1.2	0.0																																							
พัชรี จันทร์เพ็ญ (2550); ศิวะทัศน์ สุขสุวรรณ (2555); ศิวะทัศน์ สุขสุวรรณ และคณะ (2555) ศิวทัศน์ พัทธจันทร์เพ็ญ และสังวรรณ จักรระโทก	- DICHO, MIXED - NEAT (ผู้สอบ 3 กลุ่ม ความสามารถต่างกัน 3 ระดับ) - สร้างสเกลแนวตั้งแบบพหุมิติ (Multidimensional vertical scaling)	<p>ผู้วิจัยใช้ค่าอำนาจจำแนกรวม (MDISC) และค่าความยากรวม (MDIFF) ของ Yon (2006) โดยค่าอำนาจจำแนกและค่าความยากสำหรับนักเรียนระดับที่สูงกว่าและต่ำกว่า เป็นดังตาราง</p> <table border="1"> <thead> <tr> <th rowspan="2">ระดับ</th> <th colspan="2">นักเรียนระดับที่สูงกว่า</th> <th colspan="2">นักเรียนระดับที่ต่ำกว่า</th> </tr> <tr> <th>MDISC</th> <th>MDIFF</th> <th>MDISC</th> <th>MDIFF</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>0.6</td> <td>-0.5</td> <td>0.6</td> <td>0.0</td> </tr> <tr> <td>2</td> <td>0.8</td> <td>2.0</td> <td>0.8</td> <td>2.5</td> </tr> <tr> <td>3</td> <td>1.0</td> <td>1.0</td> <td>1.0</td> <td>1.5</td> </tr> <tr> <td>4</td> <td>1.2</td> <td>0.0</td> <td>1.2</td> <td>0.5</td> </tr> <tr> <td>5</td> <td>1.3</td> <td>2.2</td> <td>1.3</td> <td>3.0</td> </tr> <tr> <td>ค่าเฉลี่ย</td> <td>1.0</td> <td>0.9</td> <td>1.0</td> <td>1.5</td> </tr> </tbody> </table>	ระดับ	นักเรียนระดับที่สูงกว่า		นักเรียนระดับที่ต่ำกว่า		MDISC	MDIFF	MDISC	MDIFF	1	0.6	-0.5	0.6	0.0	2	0.8	2.0	0.8	2.5	3	1.0	1.0	1.0	1.5	4	1.2	0.0	1.2	0.5	5	1.3	2.2	1.3	3.0	ค่าเฉลี่ย	1.0	0.9	1.0	1.5
ระดับ	นักเรียนระดับที่สูงกว่า			นักเรียนระดับที่ต่ำกว่า																																					
	MDISC	MDIFF	MDISC	MDIFF																																					
1	0.6	-0.5	0.6	0.0																																					
2	0.8	2.0	0.8	2.5																																					
3	1.0	1.0	1.0	1.5																																					
4	1.2	0.0	1.2	0.5																																					
5	1.3	2.2	1.3	3.0																																					
ค่าเฉลี่ย	1.0	0.9	1.0	1.5																																					

## ตอนที่ 2 การเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

การเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ ประกอบด้วย มโนทัศน์เกี่ยวกับการเปรียบเทียบคะแนน การออกแบบการเก็บรวบรวมข้อมูลสำหรับการเปรียบเทียบคะแนน ความสำคัญของการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ กระบวนการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ การประเมินประสิทธิภาพของการเปรียบเทียบคะแนน และงานวิจัยเกี่ยวกับกระบวนการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ มีรายละเอียดดังนี้

### 2.1 มโนทัศน์เกี่ยวกับการเปรียบเทียบคะแนน

มโนทัศน์เกี่ยวกับการเปรียบเทียบคะแนน ประกอบด้วย ความหมายของการเปรียบเทียบคะแนน และเงื่อนไขของการเปรียบเทียบคะแนน มีรายละเอียดดังนี้

#### 2.1.1) ความหมายของการเปรียบเทียบคะแนน

การเปรียบเทียบคะแนนเป็นกระบวนการทางสถิติซึ่งใช้ในการปรับคะแนนของแบบสอบแต่ละฉบับให้สามารถแลกเปลี่ยนคะแนนระหว่างกันได้ (interchangeably) หรืออาจกล่าวได้ว่าการปรับเทียบคะแนนเป็นการปรับความยากที่แตกต่างกันระหว่างแบบสอบ โดยแบบสอบเหล่านั้นมีความยากและเนื้อหาคล้ายคลึงกัน (Kolen & Brennan, 2004) สอดคล้องกับนิยามการปรับเทียบคะแนนที่ระบุว่า การปรับเทียบคะแนนเป็นการแปลงระบบคะแนนของแบบสอบต่างฉบับซึ่งวัดเนื้อหาเดียวกันให้สามารถเปรียบเทียบกันได้โดยตรง โดยแปลงคะแนนของแบบสอบฉบับหนึ่งให้อยู่บนสเกลของแบบสอบอีกฉบับหนึ่ง (ศิริชัย กาญจนวาสี, 2555) ดังนั้นจึงสรุปความหมายของการปรับเทียบคะแนนว่าเป็นกระบวนการทางสถิติที่ใช้การแปลงคะแนนของแบบสอบฉบับหนึ่งไปยังแบบสอบอีกฉบับหนึ่ง เพื่อให้สามารถแลกเปลี่ยนคะแนนและเปรียบเทียบคะแนนระหว่างแบบสอบได้

กรณีที่นักเรียน 2 คนทำแบบสอบคนละฉบับในการทดสอบต่างเวลากัน เมื่อนำคะแนนจากการทดสอบไปใช้ในการเข้าเรียนต่อในระดับมหาวิทยาลัย นักเรียนคนแรกได้คะแนนสูงกว่านักเรียนคนที่ 2 นักเรียนคนแรกอาจเก่งกว่านักเรียนอีกคนหนึ่ง แต่ถ้านักเรียนคนแรกทำแบบสอบที่ง่ายกว่านักเรียนอีกคนหนึ่ง การนำคะแนนของนักเรียนทั้งสองคนมาเปรียบเทียบกันจึงไม่ยุติธรรม เพราะความแตกต่างระหว่างคะแนนของนักเรียนทั้งคู่อาจเกิดจากแบบสอบทั้งสองฉบับมีความยากแตกต่างกัน จึงนิยมใช้การปรับเทียบคะแนนในการทดสอบเข้าเรียนต่อระดับมหาวิทยาลัยเพื่อแก้ปัญหาดังกล่าว กล่าวคือ ถ้าแบบสอบทั้งสองฉบับสามารถปรับเทียบกันได้ ความแตกต่างของคะแนนทั้งสองคนนั้นไม่ได้เกิดจากความยากที่แตกต่างระหว่างแบบสอบ

แบบสอบคู่ขนานกันที่มีโครงสร้างและคุณลักษณะทางสถิติ (statistical specification) เหมือนกัน แต่ก็มีความเป็นไปได้ที่แบบสอบทั้งสองฉบับมีความยากแตกต่างกัน ดังนั้นการปรับเทียบ



คะแนนจึงนำมาใช้ในการเปรียบเทียบความยากที่แตกต่างกันได้ ซึ่งความแตกต่างนั้นไม่ได้มาจากเนื้อหาของแบบสอบ

### 2.1.2) เงื่อนไขของการเปรียบเทียบคะแนน

Doran และ Holland (2000 cited in Holland, 2007) และ Kolen & Brennan (2004) กำหนดข้อตกลงที่ทำให้การเปรียบเทียบคะแนนมีประสิทธิภาพไว้ดังนี้ 1) โครงสร้างเท่าเทียมกัน (equal construct) แบบสอบที่นำมาเปรียบเทียบกันควรมีโครงสร้างเหมือนกัน 2) ความเที่ยงเท่าเทียมกัน (equal reliability) แบบสอบที่นำมาเปรียบเทียบกันควรมีความเที่ยงเหมือนกัน 3) ความสมมาตร (symmetry) ฟังก์ชันการปรับเทียบสำหรับการปรับเทียบคะแนนแบบสอบ Y ไปยังแบบสอบ X เป็นฟังก์ชันผกผันของการปรับเทียบคะแนนแบบสอบ X ไปยังแบบสอบ Y 4) ความเสมอภาค (equity) ผู้สอบที่ทำแบบสอบซึ่งใช้ในการปรับเทียบไม่มีความแตกต่างกัน คะแนนจริงของผู้สอบมีการแจกแจงของคะแนนที่ถูกแปลง (converted score) จากแบบสอบฉบับ X และแบบสอบฉบับ Y เหมือนกัน นั่นคือ ค่าเฉลี่ยส่วนเบี่ยงเบนมาตรฐาน และลักษณะการแจกแจงของคะแนนที่สังเกตได้ซึ่งถูกแปลงจากแบบสอบฉบับ X และแบบสอบฉบับ Y เหมือนกัน 5) ความไม่แปรปรวนของประชากร (population invariance) การเลือกประชากรย่อย (subpopulation) เพื่อใช้ประมาณฟังก์ชันการปรับเทียบคะแนนระหว่างแบบสอบ X และ Y ไม่ควรมีผลต่อการปรับเทียบคะแนน กล่าวคือ ฟังก์ชันการปรับเทียบคะแนนที่ใช้เชื่อมโยงสเกลแบบสอบ X ไปยังแบบสอบ Y ไม่แปรเปลี่ยนตามประชากร

เงื่อนไขสำคัญของการปรับเทียบคะแนน คือ เนื้อหาของแบบสอบที่นำมาเปรียบเทียบกันควรเหมือนกัน (Kolen & Brennan, 2004) เมื่อพิจารณาในเชิงปฏิบัติ ข้อตกลงด้านโครงสร้างเท่าเทียมกันและความเที่ยง เท่าเทียมกัน นั่นคือ แบบสอบมีแผนผังข้อสอบเหมือนกัน (specification) และทำการทดสอบภายใต้เงื่อนไขเดียวกัน ส่วนข้อตกลงด้านความสมมาตรสามารถตรวจสอบได้ด้วยวิธีการถดถอยเพื่อทำนายคะแนนแบบสอบ Y จากแบบสอบ X ส่วนข้อตกลงด้านความเสมอภาคเป็นเหตุผลของการกำหนดข้อตกลงด้านโครงสร้างและความเที่ยงที่เท่าเทียมกัน แต่ข้อตกลงด้านความเสมอภาคตรวจสอบยาก ขณะที่ข้อตกลงด้านความไม่แปรปรวนของประชากรสามารถใช้ในเชิงปฏิบัติได้ง่ายและเป็นเหตุผลของการกำหนดข้อตกลงด้านโครงสร้างและความเที่ยงที่เท่าเทียมกัน (Holland, 2007)

## 2.2 การออกแบบการเก็บรวบรวมข้อมูลสำหรับการปรับเทียบคะแนน

การออกแบบการเก็บรวบรวมข้อมูลสำหรับการเชื่อมโยงสเกลและการปรับเทียบคะแนน ประกอบด้วย รูปแบบของการเก็บรวบรวมข้อมูลสำหรับการปรับเทียบคะแนน และการเลือกใช้รูปแบบของการออกแบบการเก็บรวบรวมข้อมูลสำหรับการปรับเทียบคะแนน มีรายละเอียดดังนี้

### 2.2.1) รูปแบบของการเก็บรวบรวมข้อมูลสำหรับการเปรียบเทียบคะแนน

รูปแบบของการเก็บรวบรวมข้อมูลสำหรับการเปรียบเทียบคะแนนมีหลายรูปแบบ และการเลือกกลุ่มผู้สอบมาใช้ในการเปรียบเทียบคะแนนควรเป็นตัวแทนของกลุ่มผู้สอบที่ทำแบบสอบภายใต้สภาพจัดการทดสอบทั่วไป การออกแบบการเก็บรวบรวมข้อมูลสำหรับการเปรียบเทียบมี 3 รูปแบบ กำหนดให้มีการแปลงคะแนนของแบบสอบ Y ไปยังคะแนนสเกล (scale score) และแบบสอบ X เป็นแบบสอบฉบับใหม่ที่จะเปรียบเทียบกับแบบสอบ Y (Kolen & Brennan, 2004)

#### 1. รูปแบบผู้สอบกลุ่มสุ่ม (random group design)

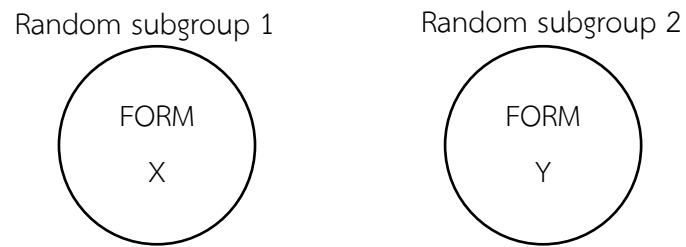
รูปแบบผู้สอบกลุ่มสุ่ม ผู้สอบถูกจัดให้ทำแบบสอบแต่ละรูปแบบอย่างสุ่มโดยใช้กระบวนการเวียน (spiraling process) ลักษณะหนึ่งของกระบวนการเวียน คือ แจกแบบสอบ X และ Y สลับกัน ผู้สอบคนที่ 1 ได้รับแบบสอบ X คนที่ 2 ได้รับแบบสอบ Y คนที่ 3 ได้รับแบบสอบ X สลับไปเรื่อย ๆ กระบวนการเวียนนี้อาจเทียบเคียงได้กับรูปแบบผู้สอบกลุ่มเท่าเทียมกันอย่างสุ่ม (randomly equivalent groups) ที่ทำแบบสอบ X และ Y เมื่อเลือกใช้การออกแบบการเก็บรวบรวมข้อมูลรูปแบบนี้ ความแตกต่างระหว่างระดับความสามารถของผู้สอบที่มีต่อแบบสอบทั้งสองฉบับจึงเป็นตัวชี้วัดของความแตกต่างของความยากระหว่างแบบสอบ

ลักษณะของการนำรูปแบบผู้สอบกลุ่มสุ่มไปใช้ คือ ผู้สอบแต่ละคนทำแบบสอบเพียงฉบับเดียว และสามารถเปรียบเทียบแบบสอบฉบับใหม่มากกว่า 1 ฉบับได้ในเวลาเดียวกันด้วยการรวมแบบสอบฉบับใหม่ที่เพิ่มเข้ามาในกระบวนการเวียน รูปแบบผู้สอบกลุ่มสุ่มต้องดำเนินการทดสอบด้วยแบบสอบทั้งสองฉบับในเวลาเดียวกันซึ่งยากต่อการนำไปใช้ในบางสถานการณ์ และกลุ่มผู้สอบควรมีขนาดใหญ่เพราะผู้สอบที่แตกต่างกันทำแบบสอบแต่ละฉบับซึ่งนำมาใช้ในการเปรียบเทียบคะแนนการนำกระบวนการเวียนมาใช้ มีประเด็นที่ควรพิจารณา คือ ผู้สอบไม่ควรนั่งสอบในลักษณะที่ส่งผลเสียต่อกระบวนการเวียน เช่น นั่งสอบอย่างเป็นระบบ ชาย-หญิง กล่าวคือ ผู้ชายทั้งหมดอาจได้รับแบบสอบ X ส่วนผู้หญิงได้รับแบบสอบ Y เป็นต้น และถ้าจำนวนผู้สอบในห้องสอบเป็นเลขคี่ ผู้สอบย่อมได้รับแบบสอบ X มากกว่าแบบสอบ Y

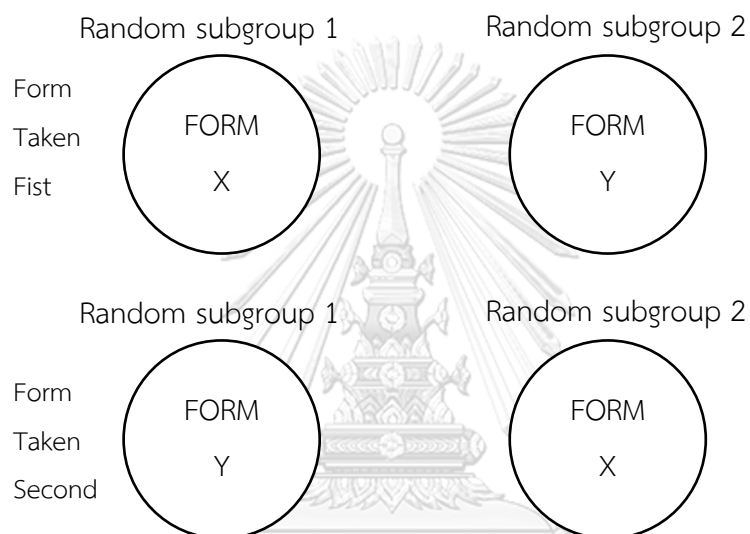
#### 2. รูปแบบผู้สอบกลุ่มเดียว (single group design)

รูปแบบผู้สอบกลุ่มเดียว ผู้สอบแต่ละคนทำแบบสอบ X และ Y กล่าวคือ เมื่อผู้สอบทำแบบสอบ X แล้วก็ทำแบบสอบ Y ต่อ ผู้สอบอาจเกิดความเมื่อยล้าจากการทดสอบแล้วอาจส่งผลให้แบบสอบ Y ยากกว่าแบบสอบ X หรือในทางกลับกันถ้าผู้สอบเกิดความคุ้นเคยกับแบบสอบอาจส่งผลให้แบบสอบ Y ง่ายกว่าแบบสอบ X สะท้อนให้เห็นถึงผลของการเรียงลำดับแบบสอบ (order effect) ดังนั้นรูปแบบผู้สอบกลุ่มเดียวจึงไม่นิยมนำไปใช้มากนัก

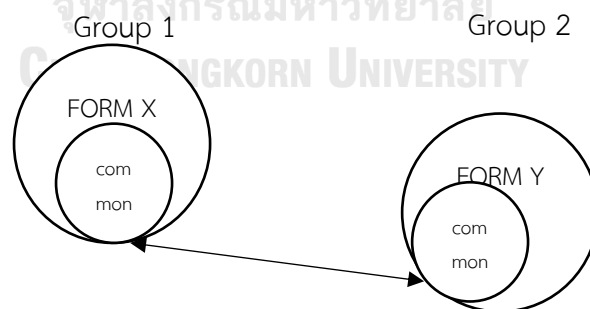
### รูปแบบผู้สอบกลุ่มสุ่ม



### รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล



### รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม



ภาพ 2.12 การออกแบบการเก็บรวบรวมข้อมูลสำหรับการเปรียบเทียบคะแนน

### 3. รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล (single group with counterbalancing design)

การจัดลำดับของการทดสอบให้มีความสมดุลเป็นวิธีหนึ่งที่ใช้ลดผลการเรียงลำดับของแบบสอบ (order effect) วิธีการหนึ่ง คือ สร้างอนุสารแบบสอบ (test booklets) ประกอบด้วยแบบสอบ X และ Y โดยครึ่งหนึ่งของอนุสารแบบสอบเริ่มจากแบบสอบ X แล้วตามด้วยแบบสอบ Y และอีกครึ่งหนึ่งของอนุสารแบบสอบเริ่มจากแบบสอบ Y แล้วตามด้วยแบบสอบ X แบบสอบทั้งสองส่วนควรจับเวลาสอบแยกกัน กระบวนการเวียนแบบนี้ทำให้สามารถเปรียบเทียบผู้สอบที่ทำแบบสอบ Y ก่อนไปยังผู้สอบที่ทำแบบสอบ X ก่อนได้

รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุลเป็นดังภาพ 2.12 ในการปรับเทียบคะแนนแบบสอบ X ไปยังแบบสอบ Y ใช้เพียงข้อมูลจากแบบสอบครั้งแรก (เช่น ข้อมูลแบบสอบ X จากกลุ่มย่อย 1 และข้อมูล แบบสอบ Y จากกลุ่มย่อย แต่ข้อมูลจากแบบสอบครึ่งหลังสามารถนำไปใช้พิจารณาความสัมพันธ์ของการปรับเทียบคะแนน (equating relationship) ระหว่างแบบสอบเมื่อทำการทดสอบด้วยแบบสอบฉบับหนึ่งก่อน ถ้าผลของการทำแบบสอบ X หลังจากทำแบบสอบ Y เหมือนกับผลของการทำแบบสอบ Y หลังจากทำแบบสอบ X แสดงว่าความสัมพันธ์ของการปรับเทียบคะแนนระหว่างการทำทดสอบด้วยแบบสอบฉบับหนึ่งก่อนเหมือนกัน แต่ถ้ามีผลของการจัดลำดับการทดสอบ ความสัมพันธ์ของการปรับเทียบคะแนนทั้งสองกรณีย่อมแตกต่างกันและอาจทำให้เกิดความไม่เสถียรในการปรับเทียบคะแนน

การควบคุมผลของการเรียงลำดับแบบสอบ (order effect) สามารถใช้รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุลได้ ซึ่งใช้ผู้สอบที่มีขนาดเล็กกว่ารูปแบบผู้สอบกลุ่มสุ่มเพราะผู้สอบทำแบบสอบทั้งสองฉบับ ในทางปฏิบัติอาจใช้รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุลแทนรูปแบบผู้สอบกลุ่มสุ่ม เมื่อสามารถจัดการให้ผู้สอบทำแบบสอบทั้งสองฉบับ ไม่ต้องการให้เกิดผลของการเรียงลำดับแบบสอบ และไม่สามารถหากกลุ่มตัวอย่างที่มากพอสำหรับรูปแบบผู้สอบกลุ่มเดียว

การปรับเทียบคะแนนภายใต้รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล โดยใช้การปรับเทียบคะแนนด้วยคะแนนสังเกตได้มี 4 ลักษณะ ได้แก่ (1) ปรับเทียบแบบสอบฉบับ X และ Y โดยใช้รูปแบบผู้สอบกลุ่มสุ่มสำหรับผู้สอบที่ทำแบบสอบฉบับ X ก่อนแล้วตามด้วยฉบับ Y (2) ปรับเทียบแบบสอบฉบับ X และ Y โดยใช้รูปแบบผู้สอบกลุ่มสุ่มสำหรับผู้สอบที่ทำแบบสอบฉบับ Y ก่อนแล้วตามด้วยฉบับ X (3) ปรับเทียบแบบสอบฉบับ X และ Y โดยใช้รูปแบบผู้สอบกลุ่มเดียวสำหรับผู้สอบที่ทำแบบสอบฉบับ X ก่อนแล้วตามด้วยฉบับ Y และ (4) ปรับเทียบแบบสอบฉบับ X และ Y โดยใช้รูปแบบผู้สอบกลุ่มเดียวสำหรับผู้สอบที่ทำแบบสอบฉบับ Y ก่อนแล้วตามด้วยฉบับ X

ถ้าผลการปรับเทียบคะแนนจากการปรับเทียบคะแนนข้างต้นต่างกัน (ไม่เกี่ยวกับความคลาดเคลื่อนของการสุ่ม) แล้วแบบสอบทั้งสองฉบับส่งผลต่อการปรับเทียบคะแนนแบบข้อ 2 ต่างกัน

ควรเลือกใช้รูปแบบผู้สอบกลุ่มสุ่ม ไม่ได้รับผลของลำดับการทำแบบสอบ โดยปัญหาของการเปรียบเทียบคะแนนแบบข้อ 1 มีเพียงกลุ่มตัวอย่างที่ใช้ค่อนข้างเล็ก แต่ถ้าผลการเปรียบเทียบคะแนนจากการเปรียบเทียบคะแนนข้างต้นเหมือนกัน แล้วแบบสอบทั้งสองฉบับส่งผลต่อการเปรียบเทียบคะแนนแบบข้อ 2 กรณีนี้ข้อมูลทั้งหมดนำไปใช้ในการเปรียบเทียบคะแนน ซึ่ง Davier et.al (2003) เสนอแนะให้ใช้วิธีล็อกเชิงเส้น (loglinear) ในการเปรียบเทียบอิกวิเปอร์เซ็นต์ภายใต้รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล

#### 4. รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม (common-item nonequivalent group design: NEAT)

รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วมนำมาใช้เมื่อไม่สามารถจัดการทดสอบวันหนึ่งด้วยแบบสอบมากกว่าหนึ่งฉบับได้ เนื่องจากเหตุผลด้านความปลอดภัยในการทดสอบหรือปัญหาในเชิงปฏิบัติ ในรูปแบบนี้ แบบสอบ X และ Y มีชุดของข้อสอบร่วมและกลุ่มผู้สอบที่แตกต่างกันทำแบบสอบแต่ละฉบับ เช่น ผู้สอบกลุ่มหนึ่งทำการทดสอบในปีแรกด้วยแบบสอบ X ส่วนผู้สอบอีกกลุ่มทำแบบสอบในปีอื่นด้วยแบบสอบ Y ข้อสอบร่วมมี 2 รูปแบบ คือ ข้อสอบร่วมภายใน (internal common item) ซึ่งเป็นตัวแทนของเนื้อหาและคุณลักษณะทางสถิติของแบบสอบฉบับเก่า ซึ่งคะแนนจากข้อสอบร่วมภายในใช้เป็นคะแนนของผู้สอบและข้อสอบร่วมภายนอก (external common item) ซึ่งคะแนนจากข้อสอบร่วมภายนอกไม่ได้ใช้เป็นคะแนนของผู้สอบและมักจับเวลาข้อสอบร่วมภายนอกแยกออกมา

การสะท้อนความแตกต่างระหว่างกลุ่มผู้สอบ ชุดข้อสอบร่วมควรเป็นตัวแทนของแบบสอบทั้งฉบับในด้านเนื้อหาและคุณลักษณะทางสถิติ กล่าวคือ ชุดข้อสอบร่วมควรเป็นแบบสอบทั้งฉบับแบบสั้น (mini version) ซึ่งคล้ายคลึงกับแบบสอบฉบับใหม่และเก่า และข้อสอบร่วมของแบบสอบทั้งฉบับควรมีจำนวนเท่ากัน สำหรับรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วมไม่ควรพิจารณาว่ากลุ่มผู้สอบที่ทำแบบสอบ X มีความเท่าเทียมกับกลุ่มผู้สอบที่ทำแบบสอบ Y ความแตกต่างระหว่างค่าเฉลี่ยของแบบสอบ X และ Y เป็นการรวมความแตกต่างของกลุ่มผู้สอบและความแตกต่างของแบบสอบ

รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วมนิยมใช้อย่างกว้างขวางเนื่องจากรูปแบบนี้ใช้ข้อสอบฉบับเดียวในการทดสอบ ขณะที่รูปแบบผู้สอบกลุ่มสุ่มต้องใช้แบบสอบหลายฉบับ ส่วนรูปแบบผู้สอบกลุ่มเดียวต้องให้ผู้สอบแต่ละคนทำแบบสอบทั้งสองฉบับ นอกจากการออกแบบการเก็บรวบรวมข้อมูลสำหรับการเปรียบเทียบคะแนน 3 รูปแบบนี้ ยังมีรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วมสำหรับการเชื่อมโยงสเกล ซึ่งข้อสอบร่วมที่ใช้ไม่จำเป็นตัวแทนของเนื้อหาของแบบสอบ X และ Y ดังนั้นการเชื่อมโยงสเกลโดยใช้รูปแบบนี้จึงเป็นประเภทคะแนนที่สอดคล้องกัน (concordance) (Kolen, 2007)

### ลักษณะและสัดส่วนของข้อสอบรวม

ข้อสอบรวมเป็นคะแนนรวมที่ใช้ปรับความแตกต่างระหว่างความสามารถของผู้สอบก่อนการปรับความแตกต่างระหว่างความยากของแบบสอบด้วยการปรับเทียบคะแนน โดยข้อสอบรวมส่วนใหญ่มีกรรมอยู่ในแบบสอบแต่ละฉบับ เรียกว่า ข้อสอบสอบรวมภายใน ข้อสอบรวมที่ดีควรวัดโครงสร้างเดียวกับแบบสอบแต่ละฉบับ กล่าวคือ ข้อสอบรวมภายในควรมีความสัมพันธ์สูงกับแบบสอบที่ใช้ในการปรับเทียบคะแนน และแบบสอบรวมไม่ควรสั้นเกินไป ควรสร้างข้อสอบรวมให้ยาวพอที่จะเป็นตัวแทนเนื้อหาในแบบสอบ (Dorans, Moses, & Eignor, 2010)

การสร้างแบบสอบแต่ละฉบับต้องใช้เวลาและค่าใช้จ่าย จึงต้องกำหนดความยาวของข้อสอบสอบรวมที่เหมาะสม กล่าวคือ ได้แนะนำว่า เมื่อแบบสอบนั้นมีจำนวนข้อสอบทั้งหมด 40 ข้อ ควรใช้ข้อสอบรวมอย่างต่ำ 20% (Angoff, Thorndike, & Lindquist, 1971) และกรณีแบบสอบมีความยาวมาก จำนวนข้อสอบรวมที่เพียงพอ คือ 30% Wright (1974 cited in Gao, 2004) ระบุว่า สัดส่วนข้อสอบรวม 10% และ 20% มีความประสิทธิภาพในการปรับเทียบคะแนนสำหรับกลุ่มตัวอย่าง 400 คน และถ้าข้อสอบรวมมีคุณภาพสูง สัดส่วนข้อสอบรวม 10% ก็มีความเพียงพอ McKinley & Reckase (1981) ศึกษาผลของกลุ่มตัวอย่างและความยาวของแบบสอบรวมที่ส่งผลต่อประสิทธิภาพของการเชื่อมโยงสเกล โดยแบบสอบแต่ละฉบับมีข้อสอบที่ตรวจให้คะแนนสองค่า จำนวน 50 ข้อ ที่มีสัดส่วนความยาวข้อสอบรวม 3 แบบ ได้แก่ 5, 15 และ 25 ข้อ คิดเป็น 10%, 30% และ 50% และขนาดกลุ่มตัวอย่างมี 5 รูปแบบ ได้แก่ 100, 300, 500, 1,000 และ 2,000 คน พบว่า ข้อสอบรวมจำนวน 15 ข้อ คิดเป็น 30% มีประสิทธิภาพมากที่สุด และสัดส่วนข้อสอบรวม 30% เมื่อกลุ่มตัวอย่างมีขนาด 2,000 ทำให้การเชื่อมโยงสเกลมีความเสถียร

นอกจากนี้ยังมีงานวิจัยที่ระบุสัดส่วนข้อสอบรวมที่น้อยที่สุดที่ทำให้การปรับเทียบคะแนนข้อสอบที่ตรวจให้คะแนนสองค่ามีประสิทธิภาพ เพื่อประหยัดค่าใช้จ่ายในการทดสอบและเพิ่มความปลอดภัยของคลังข้อสอบ เมื่อแบบสอบมีความยาวแตกต่างกันโดยใช้การจำลองข้อมูล พบว่า (1) ความยาวแบบสอบเป็น 600 ข้อ สัดส่วนข้อสอบรวมที่มีประสิทธิภาพ คือ 1/15 คิดเป็น 6% (2) ความยาวของแบบสอบเป็น 300 ข้อ สัดส่วนข้อสอบรวมที่มีประสิทธิภาพ คือ 1/12 คิดเป็น 8% (3) ความยาวของแบบสอบเป็น 240 และ 200 ข้อ สัดส่วนข้อสอบรวมที่มีประสิทธิภาพ คือ 1/10 คิดเป็น 10% (4) ความยาวของแบบสอบเป็น 160, 120, 100 และ 80 ข้อ สัดส่วนส่วนข้อสอบรวมที่มีประสิทธิภาพ คือ 1/5 คิดเป็น 20% และ (5) ความยาวแบบสอบเป็น 60 และ 40 ข้อ สัดส่วนข้อสอบรวมที่มีประสิทธิภาพ คือ 1/4 และ 1/3 คิดเป็น 25% และ 33% ตามลำดับ (Xiong, Ye, Ding, & Luo, 2013)

## 2.2.2) การเลือกใช้รูปแบบของการออกแบบการเก็บรวบรวมข้อมูลสำหรับการเปรียบเทียบคะแนน

การเลือกใช้รูปแบบของการออกแบบการเก็บรวบรวมข้อมูลสำหรับการเปรียบเทียบคะแนน ควรพิจารณาความซับซ้อนของการจัดการทดสอบและการสร้างแบบสอบ และข้อตกลงเบื้องต้นทางสถิติสรุปได้ดังตาราง 2.4

**ตาราง 2.4** การเปรียบเทียบการออกแบบการเก็บรวบรวมข้อมูลแต่ละรูปแบบในปัจจุบันต่าง ๆ

รูปแบบ	ความซับซ้อนของ การจัดการทดสอบ	ความซับซ้อนของ การสร้างแบบสอบ	ข้อตกลงเบื้องต้นทางสถิติ
Random groups	ปานกลาง – ต้องมี กระบวนการเวียน สำหรับแบบ สอบมากกว่า 1 ฉบับ	ไม่มี	เล็กน้อย – สุ่มกลุ่มผู้สอบในการทำ แบบสอบแต่ละฉบับ
Single group with counterbalancing	มาก – ผู้สอบแต่ละคนทำแบบ สอบสองฉบับ และลำดับของ การทำแบบสอบต้องสมดุล	ไม่มี	ปานกลาง – ต้องไม่มีผลของ การเรียงลำดับแบบสอบ และสุ่ม กลุ่มผู้สอบในการทำแบบสอบ
Common-item nonequivalent groups	ไม่มี – สามารถจัดการทดสอบ ได้แบบทั่วไป	ต้องเป็นตัวแทน – ต้องสร้างข้อสอบ ร่วม	มาก – ข้อสอบร่วมวัดเนื้อหา เดียวกับแบบสอบทั้งสองฉบับ กลุ่ม ผู้สอบมีความคล้ายคลึงกัน และต้อง เป็นไปตามข้อตกลงทางสถิติ
Common item to an IRT calibrated item pool	ไม่มี – สามารถจัดการทดสอบ ได้แบบทั่วไป	ต้องเป็นตัวแทน – ต้องสร้างข้อสอบ ร่วม	มาก - ข้อสอบร่วมวัดเนื้อหา เดียวกับแบบสอบทั้งสองฉบับ กลุ่ม ผู้สอบมีความคล้ายคลึงกัน และต้อง เป็นไปตามข้อตกลงของโมเดลการ ตอบสนองข้อสอบ

## 2.3 กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ

กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (UIRT) มี 2 แบบ ได้แก่ กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (UIRT observed score equating) และกระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (UIRT true score equating)

### 2.3.1) กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ

การปรับเทียบคะแนนสังเกตได้ด้วยโมเดล UIRT มีขั้นตอนดังนี้

1. คำนวณ Conditional observed score distributions;  $f(x|\theta_i)$  ในแต่ละระดับความสามารถ ( $\theta_i$ ) โดยใช้สมการปรับแก้ (Peterson, 2014)

ข้อสอบที่ตรวจให้คะแนนสองค่า: ประยุกต์ใช้สมการ Lord-Wingersky (1984)

$$\begin{aligned}
 f_r(x|\theta_j) &= f_{r-1}(x|\theta_j)(1 - P_r) && \text{เมื่อ } x = 0 \\
 f_r(x|\theta_j) &= f_{r-1}(x|\theta_j)(1 - P_r) + f_{r-1}(x - 1|\theta_j)P_r && \text{เมื่อ } 0 < x < r \\
 f_r(x|\theta_j) &= f_{r-1}(x - 1|\theta_j)P_r && \text{เมื่อ } x = r
 \end{aligned}$$

ข้อสอบที่ตรวจให้คะแนนหลายค่า: ประยุกต์ใช้สมการ Hanson & Thissen (1995)

$$f_r(x|\theta_j) = \sum_{k=1}^{K_j} f_{r-1}(x - W_{rk}) P_{ik}(\theta_j) \quad \text{เมื่อ } \min_r < x < \max_r$$

เมื่อ  $\min_r$  คือ คะแนนต่ำสุดที่เป็นไปได้ภายหลังการเพิ่มของข้อสอบข้อที่  $r$

$\max_r$  คือ คะแนนสูงสุดที่เป็นไปได้ภายหลังการเพิ่มของข้อสอบข้อที่  $r$

2. นำ Conditional observed score distributions ( $f(x|\theta_j)$ ) คูณด้วยความหนาแน่นของความสามารถ (ability density:  $g(\vartheta)$ )

3. รวมผลทั้งหมดในแต่ละระดับความสามารถได้ และการแจกแจงตามขอบของคะแนนสังเกตได้สำหรับแบบสอบแต่ละฉบับ ดังสมการ

$$f(x) = \int_{\vartheta} f(x|\theta)g(\vartheta)d\vartheta$$

เมื่อ  $g(\vartheta)$  คือ การแจกแจงความสามารถ  $\vartheta$

ถ้าการแจกแจงความสามารถ  $\vartheta$  เป็นแบบไม่ต่อเนื่องด้วยจุดการประมาณพื้นที่ (quadrature point) อย่างเท่าๆกัน จะได้ว่า

$$f(x) = \sum_j f(x|\theta_j) g(\theta_j)$$

ถ้าการแจกแจงความสามารถ  $\vartheta$  เป็นแบบไม่ต่อเนื่องด้วยจำนวนความสามารถแบบไม่จำกัด ( $N$  ความสามารถ) จะได้ว่า

$$f(x) = \frac{1}{N} \sum_j f(x|\theta_j)$$

4. เมื่อกำหนด Marginal observed score distributions ( $f(x)$  และ  $f(y)$ ) สำหรับแบบสอบแต่ละฉบับแล้ว จึงสามารถเปรียบเทียบคะแนนของแบบสอบทั้งสองฉบับด้วยวิธีอิกวิเปอร์เซ็นไทล์ได้

### 2.3.2) กระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิตี

กระบวนการการปรับเทียบคะแนนจริงด้วยโมเดล UIRT ทำให้คะแนนจริงของแบบสอบสองฉบับเชื่อมโยงกันได้โดยใช้นิยามทฤษฎีการตอบสนองข้อสอบแบบเอกมิตี มีขั้นตอนดังนี้ (Kolen & Brennan, 2004; Peterson, 2014)

1. ระบุคะแนนจริงจากแบบสอบฉบับใหม่ ( $\tau_A$ ) โดยค่าคะแนนจริงดังกล่าวมีค่าอยู่ระหว่างผลรวมของค่าพารามิเตอร์ (lower asymptote parameters) และจำนวนข้อสอบทั้งหมดในแบบสอบฉบับใหม่ ( $N_A$ ) นั่นคือ

$$\sum_{j:A} c_j < \tau_A < N_A$$

โดยผลรวมของความน่าจะเป็นที่จะตอบข้อสอบข้อนั้นได้ถูกต้อง คือ คะแนนจริง ณ ระดับความสามารถเฉพาะนั้น ๆ (particular ability level:  $\theta_i$ ) ดังสมการ

$$\tau_A(\theta_i) = \sum_{j:A} p_{ij}(\theta_i; a_j, b_j, c_j)$$



2. ระบุ  $\theta_i$  ที่สอดคล้องกับคะแนนจริงของแบบสอบฉบับใหม่ โดยใช้วิธีของ Newton-Raphson โดยนำค่า  $\theta_i$  แทนลงในนิยามของคะแนนจริงของแบบสอบฉบับดั้งเดิม ดังสมการ

$$\tau_B(\theta_i) = \sum_{j:B} p_{ij}(\theta_i; a_j, b_j, c_j)$$

3. อินทิเกรตคะแนนของแบบสอบฉบับใหม่เพื่อหาค่าคะแนนจริงที่สอดคล้องกันในแบบสอบฉบับดั้งเดิม

### สรุปหลักการเกี่ยวกับการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ

เมื่อเปรียบเทียบการปรับเทียบคะแนนจริงและคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ พบว่า การปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ คำนวณง่ายกว่าและไม่จำเป็นต้องใช้การแจกแจงความสามารถของกลุ่มตัวอย่าง แต่ในทางปฏิบัติ ทราบเพียงคะแนนสังเกตได้ ไม่ทราบคะแนนจริง และถ้าค่าพารามิเตอร์โอกาสในการเดาอยู่ในทฤษฎีการตอบสนองข้อสอบ จะไม่มีความเท่าเทียมในคะแนนต่ำและคะแนน number-correct สูง

การปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ สะท้อนความสัมพันธ์ของการปรับเทียบสำหรับคะแนนสังเกตได้ เมื่อเปรียบเทียบการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองแบบเอกมิติ พบว่า การปรับเทียบคะแนนสังเกตได้ดังกล่าวมีการคำนวณซับซ้อนและต้องใช้การแจกแจงความสามารถของกลุ่มตัวอย่าง แต่เมื่อกลุ่มตัวอย่างสองกลุ่มมีความเท่าเทียมกันอย่างสุ่ม พบว่า ผลของการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองแบบเอกมิติเหมือนกับการปรับเทียบด้วยวิธีอีควิเปอร์เซ็นไทล์ และแม้ว่าการปรับเทียบคะแนนทั้งสองแบบมีความแตกต่างกันในเชิงทฤษฎี แต่ผลของการปรับเทียบคะแนนจากการปรับเทียบคะแนนทั้ง 2 แบบในรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบรวม (NEAT) มีความคล้ายคลึงกัน

### 2.4 กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองแบบพหุมิติ

การปรับเทียบคะแนนสำหรับแบบสอบที่วัดในมิติเดียวหรือคุณลักษณะเดียวสามารถใช้กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ แต่ถ้าใช้การวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติกับแบบสอบที่วัดมากกว่าหนึ่งคุณลักษณะย่อมทำให้ผลการประมาณความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบมีความลำเอียงสูง เพื่อแก้ปัญหาดังกล่าวจึงมีการพัฒนาทฤษฎีการตอบสนองข้อสอบแบบพหุมิติขึ้น นำไปสู่การพัฒนาการปรับเทียบคะแนนโดยใช้โมเดล MIRT

กรอบแนวคิดของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองแบบพหุมิติ (MIRT framework) คล้ายคลึงกับการปรับเทียบคะแนนตามทฤษฎีการตอบสนองแบบเอกมิติ (UIRT equating) คือ ความสามารถของผู้สอบสามารถรายงานได้หลายรูปแบบ เช่น คะแนน number-correct และค่า

ความสามารถตามทฤษฎีการตอบสนองข้อสอบ (MIRT ability estimate) แต่กระบวนการปรับเทียบคะแนนทั้งสองรูปแบบก็มีความแตกต่างกัน โดยค่าความสามารถตามทฤษฎีการตอบสนองข้อสอบตามกรอบแนวคิด UIRT ( $\theta$ ) เป็นสเกลาร์ และเมื่อเชื่อมโยงสเกลตามทฤษฎีการตอบสนองแบบเอกมิติ (UIRT scale linking) แล้ว ค่าความสามารถตามทฤษฎีการตอบสนองข้อสอบ ( $\hat{\theta}$ ) จากแบบสอบทั้งสองฉบับอยู่ในเมตริกซ์เดียวและสามารถเปรียบเทียบกันได้ ขณะที่ตามกรอบแนวคิด MIRT ค่าความสามารถ ( $\hat{\theta}$ ) เป็นเวกเตอร์ ( $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m]$ )

นอกจากนี้การนำเสนอความสัมพันธ์ระหว่างตำแหน่งของผู้สอบในสเกลความสามารถและความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องตามกรอบแนวคิด UIRT อยู่ในลักษณะจุดตัด (cutoff point) ในโค้งคุณลักษณะข้อสอบ (item characteristic curve: ICC) ขณะที่ตามกรอบแนวคิด MIRT การนำเสนอความสัมพันธ์ระหว่างตำแหน่งของผู้สอบในสเกลความสามารถและความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องอยู่ในลักษณะคอนทัวร์จุดตัด (cutoff contour) ในพื้นผิวคุณลักษณะข้อสอบ (item characteristic surface: ICS) กล่าวคือ ถ้าผู้สอบที่มีเวกเตอร์ความสามารถตกอยู่ในคอนทัวร์เดียวกันสำหรับข้อสอบข้อหนึ่งแสดงว่าความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบข้อนั้นได้ถูกต้องเท่ากัน และเมื่อพิจารณาตามหลักคณิตศาสตร์ เวกเตอร์ความสามารถที่ตกอยู่ในคอนทัวร์เดียวกันย่อมมีความเท่าเทียมกันเมื่อมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องเท่ากัน แต่ความเท่าเทียมของเวกเตอร์ความสามารถเหล่านี้ไม่สามารถเปรียบเทียบกันได้โดยตรงจนกว่าจะมีค่า p-value เท่ากัน นอกจากนี้เมื่อแบบสอบพหุมิติทั้งสองฉบับอยู่ในเมตริกซ์เดียวกันหลังจากการเชื่อมโยงสเกลตามโมเดล MIRT แล้ว ชุดของค่าพารามิเตอร์ของข้อสอบสำหรับแบบสอบทั้งสองฉบับยังแตกต่างกันอยู่ หลังจากการเชื่อมโยงสเกลตามโมเดล MIRT แบบสอบทั้งสองฉบับจึงยังไม่ได้มีความสัมพันธ์กันโดยตรง ซึ่งเป็นข้อจำกัดที่สำคัญของการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองแบบพหุมิติ (Zhang, 2012) จากข้อจำกัดดังกล่าวจึงมีการศึกษากระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (Full MIRT score equating procedure: FMIRT) ประกอบด้วย 3 กระบวนการ ได้แก่ (1) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (Full MIRT observed score equating procedure: MOSE) (2) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูปโดยใช้การประมาณเอกมิติ (unidimensional approximation to MIRT observed score equating procedure: AOSE) และ (3) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูปโดยใช้การประมาณเอกมิติ (unidimensional approximation to MIRT true score equating procedure: ATSE)

การคำนวณ Unidimensional approximation item parameter estimates ถือว่า การแจกแจงความสามารถแบบหลายมิติในประชากรของผู้สอบเป็นการแจกแจงปกติแบบพหุตัวแปร (multivariate normal distribution:  $\theta \sim MVN(0, \Sigma)$ ) และน้ำหนักของข้อสอบมีความเท่าเทียมกัน ส่วนข้อตกลง

เบื้องต้นของกระบวนการ MOSE คือ การแจกแจงการประมาณพื้นที่แบบพหุตัวแปร (multivariate quadrature distribution) เป็นการแจกแจงแบบปกติมาตรฐานด้วยแกนที่ไม่สัมพันธ์กัน ( $\theta \sim MVN(0, \Sigma)$ ) ซึ่งการปรับเทียบคะแนนภายใต้รูปแบบผู้สอบกลุ่มเท่าเทียมกันอย่างสุ่ม ถือว่าผู้สอบทั้งสองกลุ่มถูกสุ่มมาจากประชากรกลุ่มเดียวกันจึงมีความเท่าเทียมกัน ดังนั้นเมื่อมีการประมาณค่าพารามิเตอร์แบบแยกจากกัน (separate calibration) สำหรับผู้สอบแต่ละกลุ่มภายใต้รูปแบบดังกล่าว จึงไม่ต้องใช้การเชื่อมโยงสเกล เนื่องจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานระหว่างกลุ่มเท่ากันในกระบวนการประมาณค่าพารามิเตอร์ ดังนั้นการปรับเทียบคะแนนสังเกตได้และคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติภายใต้รูปแบบผู้สอบกลุ่มสุ่มไม่จำเป็นต้องใช้การเชื่อมโยงสเกล (MIRT scaling linking) แต่เมื่อประมาณค่าพารามิเตอร์แบบแยกกันสำหรับผู้สอบแต่ละกลุ่มภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม (NEAT) ควรใช้การเชื่อมโยงสเกลเพื่อทำให้ Scale indeterminacy มีความชัดเจนยิ่งขึ้น (Bradley G. Crossman & Lee, 2013)

#### 2.4.1) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (Full MIRT observed score equating procedure: MOSE)

กระบวนการ MOSE เป็นการขยายแนวคิดของกระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (UIRTO equating) กล่าวคือ กระบวนการ MOSE มีแนวคิดใกล้เคียงกับกรอบแนวคิด UIRT โดยการแจกแจงคะแนนสังเกตได้ Number-correct scores สำหรับผู้สอบเป็นผลรวมของความสามารถที่ได้จากการแจกแจงทวินามผสม (compound binomial distribution) จากสมการกระทำซ้ำ (recursion formula) ของ Lord และ Wingersky (1984) แต่ MOSE มีความแตกต่างจากการปรับเทียบ UIRTO คือ การแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขในแต่ละกลุ่มของระดับความสามารถ ( $f_r(x|\theta)$ ) ถูกกำหนดในผลรวมของแต่ละระดับความสามารถในพื้นที่ความสามารถ (O. Zhang, 2012) โดยกระบวนการปรับเทียบคะแนน MOSE สำหรับข้อสอบแบบหลายตัวเลือก ข้อสอบแบบสร้างคำตอบ และข้อสอบรูปแบบผสม มีรายละเอียดดังนี้

- กระบวนการ MOSE สำหรับข้อสอบแบบหลายตัวเลือก (ตรวจให้คะแนนสองค่า) มีขั้นตอนดังนี้ (Bradley G. Crossman & Lee, 2013; Peterson, 2014)

1. คำนวณ Conditional observed score distributions ในแต่ละกลุ่มของระดับความสามารถ ( $f_r(x|\theta)$ ) เมื่อ  $\theta$  คือ เวกเตอร์ของความสามารถ (ability vector)  $\hat{\theta} = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m]$  และ  $r$  คือ ดัชนีข้อสอบ (item index) ขณะที่กรอบแนวคิด UIRT เริ่มจากการกำหนด Conditional observed score distributions ( $f(x|\theta_i)$ ) ในแต่ละระดับความสามารถ ( $\theta_i$ ) ความน่าจะเป็นที่จะได้ 1 คะแนนจากข้อสอบข้อ 1 คือ  $f_1(x = 1|\theta_j) = P_1$   
ความน่าจะเป็นที่จะได้ 0 คะแนนจากข้อสอบข้อ 1 คือ  $f_1(x = 0|\theta_j) = 1 - P_1$

แต่ถ้า  $r > 1$  คำนวณอัลกอริทึม (algorithm) โดยใช้อัลกอริทึมของ Lord-Wingersky และใช้  
 เวกเตอร์ของระดับความสามารถแทนระดับความสามารถเดี่ยว (single ability) ดังสมการ  
 (Brossman and Lee, 2013)

$$\begin{aligned} f_r(x|\theta_j) &= f_{r-1}(x|\theta_j)(1 - P_r) & x = 0 \\ f_r(x|\theta_j) &= f_{r-1}(x|\theta_j)(1 - P_r) + f_{r-1}(x - 1|\theta_j)P_r & 0 < x < r \\ f_r(x|\theta_j) &= f_{r-1}(x - 1|\theta_j)P_r & x = r \end{aligned}$$

เมื่อ  $f_r(x|\theta_j)$  คือ Conditional observed score distributions

2. คำนวณ Marginal observed score distributions ด้วยการคูณการแจกแจง  
 คะแนนอย่างมีเงื่อนไขด้วยความหนาแน่นความสามารถแบบหลายตัวแปร (multivariate) และ  
 รวมผลลัพธ์ทั้งหมดของความสามารถแฝง  $m$  ดังสมการ

$$f(x) = \int_{\theta_1} \int_{\theta_2} \cdots \int_{\theta_m} f(x|\theta) \psi(\theta) d(\theta)$$

หรือ  $f(x) = \sum_{\theta_1} \sum_{\theta_2} \cdots \sum_{\theta_m} f(x|\theta) \psi(\theta)$

$m$  คือ จำนวนของมิติ

การแจกแจงมาตรฐานของตัวแปรพหุด้วยแกนที่ไม่มีความสัมพันธ์กัน (multivariate  
 standard normal distribution with uncorrelated axes:  $\theta \sim MVN(\mathbf{0}, I)$ ) ใช้คำนวณความ  
 หนาแน่นของความสามารถแบบตัวแปรพหุ (multivariate ability density)

3. ปรับเทียบคะแนนแบบสอบทั้งสองฉบับใช้วิธีอีควิเปอร์เซ็นไทล์ (เช่นเดียวกับกรณีเอกมิติ)

- กระบวนการ MOSE สำหรับข้อสอบแบบสร้างคำตอบ (ตรวจให้คะแนนหลายค่า) มีขั้นตอน  
 เหมือนกับข้อสอบแบบหลายตัวเลือก แต่มีความแตกต่าง ดังนี้

1. Conditional observed score distributions ประยุกต์ใช้สูตรของ Hanson (1994)  
 และ Thissen และคณะ (1995 อ้างอิงใน Peterson, 2014) ดังนั้น Conditional observed score  
 distributions สำหรับความสามารถแฝง  $m$  สำหรับข้อสอบข้อที่  $i$  ด้วยลำดับขั้นของคำตอบ (response  
 categories,  $K$ ) คือ ความน่าจะเป็นของผู้สอบ  $j$  ที่จะได้คะแนนในลำดับขั้นที่  $k$  บนเงื่อนไขของ  
 เวกเตอร์ความสามารถ  $m$  ( $\theta_j$ ) ดังสมการ

$$f_r(U_i = W_{rk}|\theta_j) = P_{ik}(\theta_j)$$

เมื่อ  $W_{rk}$  คือ ฟังก์ชันการให้คะแนนของลำดับขั้นที่  $k$  สำหรับข้อสอบข้อที่  $r$  โดยที่ค่า  $K$   
 แปรเปลี่ยนไปตามข้อสอบ

ถ้า  $r > 1$  ความน่าจะเป็นของคะแนน  $x$  ภายหลังจากทดสอบข้อที่  $r$  เป็นดังสมการ

$$f_r(x|\theta_j) = \sum_{k=1}^{K_j} f_{r-1}(x - W_{rk}) P_{ik}(\theta_j) \quad \text{เมื่อ } \min_r < x < \max_r$$

เมื่อ  $\min_r$  คือ คะแนนต่ำสุดที่เป็นไปได้ภายหลังการเพิ่มของข้อสอบข้อที่  $r$

$\max_r$  คือ คะแนนสูงสุดที่เป็นไปได้ภายหลังการเพิ่มของข้อสอบข้อที่  $r$

ถ้าเทอม  $x - W_{rk} < \min_r$  หรือ  $x - W_{rk} > \max_r$  จะได้  $f_{r-1}(x - W_{rk}) = 0$

2. การคำนวณ Marginal observed score distributions คำนวณจากการคูณการแจกแจงอย่างมีเงื่อนไขโดยความหนาแน่นของความสามารถและอินทิเกรต/รวมผลลัพธ์ทั้งหมดของ  $m$  ทักษะ ดังสมการ

$$f(x) = \int_{\theta_1} \int_{\theta_2} \cdots \int_{\theta_m} f(x|\theta) \psi(\theta) d(\theta)$$

หรือ

$$f(x) = \sum_{\theta_1} \sum_{\theta_2} \cdots \sum_{\theta_m} f(x|\theta) \psi(\theta)$$

- กระบวนการปรับเทียบคะแนน FMIRT สำหรับแบบสอบรูปแบบผสม มีขั้นตอนคล้ายคลึงกับข้อสอบแบบหลายตัวเลือก แต่มีความแตกต่าง ดังนี้ (E. Lee, 2013)

1. การคำนวณ Conditional observed score distributions เป็นการรวมสมการเรียกซ้ำ (recursive formula) ของข้อสอบที่ตรวจให้คะแนนหลายสองค่าและหลายค่าบนเวกเตอร์เงื่อนไขความสามารถ ( $\theta_j$ )

2. การคำนวณ Marginal observed score distributions ใช้การแจกแจงของตัวแปร 3 ตัว คือ  $q(\theta_G, \theta_M, \theta_F)$  ดังสมการ

$$\begin{aligned} f(x) &= \iiint_{-\infty}^{\infty} f(x|\theta_G, \theta_M, \theta_F) g(\theta_G, \theta_M, \theta_F) d\theta_G d\theta_M d\theta_F \\ &= \sum_{\theta_G} \sum_{\theta_M} \sum_{\theta_F} f(x|\theta_G, \theta_M, \theta_F) q(\theta_G, \theta_M, \theta_F) \end{aligned}$$

เมื่อ  $q(\theta_G, \theta_M, \theta_F)$  คือ ฟังก์ชันความหนาแน่นของการแจกแจงปกติของตัวแปร 3 ตัวสำหรับคุณลักษณะแฝงทั่วไป คุณลักษณะแฝงเฉพาะข้อสอบแบบหลายตัวเลือก และ คุณลักษณะแฝงเฉพาะข้อสอบแบบสร้างคำตอบ

3. ปรับเทียบคะแนนแบบสอบทั้งสองฉบับใช้วิธีอีควิเปอร์เซ็นไทล์ (เช่นเดียวกับกรณีเอกมิติ)

เมื่อเปรียบเทียบสมการที่ใช้ในกระบวนการ MOSE ระหว่างข้อสอบแบบหลายตัวเลือก สร้างคำตอบ และรูปแบบผสม พบว่า

1. การคำนวณ Marginal observed score distributions สำหรับข้อสอบทั้ง 3 รูปแบบใช้นิยามเดียวกัน คือ ผลรวมของการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขโดยคุณด้วยความหนาแน่นของความสามารถ/ทักษะแฝง  $m$  แต่ความหนาแน่นของความสามารถ/ทักษะแฝงสำหรับข้อสอบรูปแบบผสมประกอบด้วย 3 ตัวแปร ได้แก่ คุณลักษณะแฝงทั่วไป คุณลักษณะแฝงเฉพาะข้อสอบแบบหลายตัวเลือก และ คุณลักษณะแฝงเฉพาะข้อสอบแบบสร้างคำตอบ ( $q(\theta_G, \theta_M, \theta_F)$ )

2. การคำนวณ Conditional observed score distribution ของข้อสอบทั้ง 3 รูปแบบแตกต่างกัน สำหรับข้อสอบแบบหลายตัวเลือกประยุกต์ใช้อัลกอริทึมของ Lord-Wingersky (Bradley G.

Brossman & Lee, 2013) ขณะที่ข้อสอบแบบสร้างคำตอบประยุกต์ใช้สมการเวียนซ้ำ (recursion formula) ของ Hanson (1994) และ Thissen และคณะ (1995) ส่วนข้อสอบรูปแบบผสมเป็นการรวมกันของสมการเวียนซ้ำสำหรับข้อสอบที่มีการตรวจให้คะแนน 2 ค่าและหลายค่าบนเวกเตอร์ความสามารถ ( $\theta_j$ )

ตาราง 2.5 เปรียบเทียบสมการที่ใช้ในกระบวนการ MOSE ของข้อสอบทั้ง 3 รูปแบบ

รูปแบบ	Conditional observed score distribution	Marginal observed score distribution
DICHO	$f_r(x \theta_j) = f_{r-1}(x \theta_j)(1 - P_r)$ <p>เมื่อ <math>x = 0</math></p> $f_r(x \theta_j) = f_{r-1}(x \theta_j)(1 - P_r) + f_{r-1}(x - 1 \theta_j)P_r$ <p>เมื่อ <math>0 &lt; x &lt; r</math></p> $f_r(x \theta_j) = f_{r-1}(x - 1 \theta_j)P_r$ <p>เมื่อ <math>x = r</math></p>	$f(x) = \int_{\theta_1} \int_{\theta_2} \cdots \int_{\theta_m} f(x \theta)\psi(\theta)d(\theta)$ $= \sum_{\theta_1} \sum_{\theta_2} \cdots \sum_{\theta_m} f(x \theta)\psi(\theta)$
POLY	$f_r(x \theta_j)$ $= \sum_{k=1}^{K_j} f_{r-1}(x - W_{rk}) P_{ik}(\theta_j)$	$f(x) = \int_{\theta_1} \int_{\theta_2} \cdots \int_{\theta_m} f(x \theta)\psi(\theta)d(\theta)$ $= \sum_{\theta_1} \sum_{\theta_2} \cdots \sum_{\theta_m} f(x \theta)\psi(\theta)$
MIXED	ผลรวมของสมการเรียกซ้ำ (recursive formula) ของข้อสอบที่ตรวจให้คะแนนหลายสองค่าและหลายค่าบนเวกเตอร์เงื่อนไขความสามารถ ( $\theta_j$ )	$f(x) = \iiint_{-\infty}^{\infty} f(x \theta_G, \theta_M, \theta_F)g(\theta_G, \theta_M, \theta_F)d\theta_G d\theta_M d\theta_F$ $= \sum_{\theta_G} \sum_{\theta_M} \sum_{\theta_F} f(x \theta_G, \theta_M, \theta_F)q(\theta_G, \theta_M, \theta_F)$

จุฬาลงกรณ์มหาวิทยาลัย

#### 2.4.2) กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูโดยใช้ในการประมาณเอกมิติ (Unidimensionalized MIRT Equating)

การประมาณแบบเอกมิติสำหรับการปรับเทียบคะแนนจริงและคะแนนเกดได้แบบพหุมิติอาศัยอัลกอริทึมการประมาณเอกมิติ (Jinming Zhang & Stout, 1999) เพื่อคำนวณค่าพารามิเตอร์ข้อสอบแบบเอกมิติและการแจกแจงความสามารถโดยการกำหนดทิศทางของการวัดที่ดีที่สุด (direction of best measurement) ซึ่งทิศทางการวัดที่ดีที่สุดคือ ทิศทางที่ฟังก์ชันสารสนเทศคะแนนแบบพหุมิติสูงสุดครอบคลุมทุกทิศทางที่เป็นไปได้ทั้งหมด โดยฟังก์ชันสารสนเทศคะแนนแบบพหุมิติ (multidimensional score information function) คือ รากที่สองฟังก์ชันสารสนเทศข้อสอบแบบพหุมิติ (multidimensional item information function) ดังสมการ (E. Lee, 2013)

$$S_Y(\theta) = \sqrt{\frac{[V_Y P(\theta)]^2}{P(\theta)Q(\theta)}}$$

เมื่อ  $\gamma$  คือ เวกเตอร์ของมุมที่ทำกับแกนพิกัดซึ่งแสดงทิศทางจากจุด  $\theta$

$\nabla_\gamma$  คือ อนุพันธ์ (directional derivative) ในทิศทาง  $\gamma$

$P(\theta)$  คือ ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง

$Q(\theta)$  คือ ความน่าจะเป็นในการตอบข้อสอบไม่ถูกต้อง

ทิศทางการวัดที่ดีที่สุด คือ ค่าเฉลี่ยของฟังก์ชันสารสนเทศคะแนนแบบพหุมิติครอบคลุมทุกทิศทาง ซึ่งให้เวกเตอร์มาตรฐานของน้ำหนักในแต่ละมิติ ดังนั้นทิศทางการวัดที่ดีที่สุดจึงคำนวณสเกลแบบเอกมิติโดยคอมโพสิทเชิงเส้นของคุณลักษณะ

- การประมาณแบบเอกมิติสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า

โมเดล Generalized multidimensional compensatory model คือ

$$p_{ij}(\theta_i) = H_j(\mathbf{a}_j^T \theta_i + d_j) \equiv H_j(\sum_{k=1}^{\delta} a_{kj} \theta_k + d_j)$$

เมื่อ  $\mathbf{a}_j^T$  คือ เวกเตอร์ของค่าพารามิเตอร์อำนาจจำแนก โดย  $\mathbf{a}_j^T = (a_{j1}, a_{j2}, \dots, a_{j\delta})$  ซึ่งไม่เท่ากับศูนย์และไม่ใช่ค่าลบ

$d_j$  คือ ดัชนีที่เกี่ยวข้องกับค่าพารามิเตอร์ความยาก

$H_j(\cdot)$  คือ ฟังก์ชันการเชื่อมโยง โดย  $H_j(x)$  เป็นฟังก์ชันที่ไม่ลดลง (monotonically non-decreasing function) ด้วย  $H_j(x) \geq 0$  และ  $H_j(x^*) \geq 0$  ไม่เป็นศูนย์

โมเดล Generalized multidimensional compensatory model ใช้ได้กับโมเดล Multidimensional compensatory logistic model แบบ 3 พารามิเตอร์ และโมเดล Multidimensional compensatory normal ogive model แบบ 3 พารามิเตอร์ (O. Zhang, 2012)

Zhang และ Shout (1999) กล่าวว่า ชุดข้อสอบภายใต้โมเดล MIRT สามารถประมาณได้อย่างใกล้เคียงโดยใช้โมเดล UIRT ด้วยคอมโพสิทความสามารถแบบเอกมิติที่ประมาณได้ ( $\theta_\alpha$ ) และค่าพารามิเตอร์ของข้อสอบแบบเอกมิติที่ประมาณได้ โดยคอมโพสิทความสามารถ:  $\theta_\alpha$  ของเวกเตอร์ความสามารถแบบพหุมิติ  $\theta$  ( $\theta = \theta_1, \theta_2, \dots, \theta_m$ ) เป็นการรวม  $\theta$  เชิงเส้นแบบมาตรฐาน นั่นคือ

$$\theta_\alpha = \hat{\mathbf{a}}^T \theta = \alpha^t \theta = \sum_{k=1}^{\delta} \alpha_j \theta_j$$

เมื่อ  $\hat{\mathbf{a}}^T$  คือ เมทริกซ์ค่าประมาณอำนาจจำแนกของข้อสอบแบบพหุมิติ

$\alpha$  คือ ทิศทางของคอมโพสิท  $\theta_\alpha$  ( $\theta = \theta_1, \theta_2, \dots, \theta_m$ ) หรือการประมาณแบบเอกมิติของเวกเตอร์ความสามารถแบบพหุมิติ โดย  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_d)^t$

โดยกำหนดให้  $Var(\theta_\alpha) = 1$  และผลรวมของทิศทางของคอมโพสิทเท่ากับ 1 ( $\sum_{k=1}^{\delta} \alpha_j = 1$ )

ดังนั้นทิศทางสำหรับคอมโพสิทเชิงเส้น ( $\theta_\alpha$ ) สามารถคำนวณจาก

$$\hat{\alpha}_k = \omega \sum_{j=1}^N w_j E \left\{ H_j(\hat{\mathbf{a}}^T \hat{\theta}) \left[ \sum_{j=1}^N w_j^2 H_j(\hat{\mathbf{a}}^T \hat{\theta}) [1 - H_j(\hat{\mathbf{a}}^T \hat{\theta})] \right]^{-1/2} \right\} \hat{a}_{jk}$$

$$\hat{\alpha}_k = \omega \sum_{j=1}^N w_j E \left[ \frac{H_j(\hat{\mathbf{a}}_j^T \boldsymbol{\Theta})}{\sqrt{\text{var}(Y|\boldsymbol{\Theta})}} \right] \hat{a}_{jk}, k = 1, 2, \dots, \delta$$

เมื่อ  $\boldsymbol{\Theta}$  คือ เวกเตอร์คุณลักษณะแฝงสมบูรณ์ (complete latent trait vector) เช่น เวกเตอร์คุณลักษณะแฝงของผู้สอบทุกคน

$w_j$  คือ น้ำหนักคะแนน

$\omega$  คือ ค่าคงที่เชิงบวก (positive constant) ซึ่งเป็นผลรวมของคอมโพสิต  $\Theta_\alpha$  เท่ากับ 1 ( $\sum_{j=1}^d \alpha_j = 1$ )

$E$  คือ ตัวดำเนินการที่คาดหวัง (expectation operator)

ภายใต้ข้อตกลงเบื้องต้นที่ทุกเทอมมีน้ำหนักเท่ากัน

$$w_j E \left[ \frac{H_j(\hat{\mathbf{a}}_j^T \boldsymbol{\Theta})}{\sqrt{\text{var}(Y|f\boldsymbol{\Theta})}} \right], j = 1, 2, \dots, N$$

ดังนั้นสมการทิศทางของคอมโพสิตเชิงเส้น คือ  $\hat{\alpha}_k = \frac{\sum_{j=1}^N \hat{a}_{jk}}{\sqrt{\sum_{k=1}^{\delta} (\sum_{j=1}^N \hat{a}_{jk})^2}}$

เมื่อ  $N$  คือ จำนวนข้อสอบทั้งหมด

โดยทิศทางของคอมโพสิตเชิงเส้นเป็นสิ่งสำคัญในการประมาณแบบเอกมิตี เทอม  $w_j E \left[ \frac{H_j(\hat{\mathbf{a}}_j^T \boldsymbol{\Theta})}{\sqrt{\text{var}(Y|\boldsymbol{\Theta})}} \right]$

พิจารณาว่าเป็น “น้ำหนักส่วนประกอบ” (compound weight) สำหรับข้อสอบเพื่อใช้กำหนดทิศทางของข้อสอบ โดยแต่ละเทอมถูกกำหนดโดยค่าน้ำหนักคะแนน ( $w_j$ ) และน้ำหนักโมเดล (เชิงทฤษฎี)

ของข้อสอบ  $E \left[ \frac{H_j(\hat{\mathbf{a}}_j^T \boldsymbol{\Theta})}{\sqrt{\text{var}(Y|\boldsymbol{\Theta})}} \right]$  ซึ่งถูกกำหนดโดยอนุพันธ์ของฟังก์ชันเชื่อมโยง (link function) และ

เวกเตอร์ค่าพารามิเตอร์ของอำนาจจำแนก (O. Zhang, 2012)

ขั้นตอนการประมาณค่าพารามิเตอร์ข้อสอบที่ตรวจให้คะแนนสองค่า (โมเดล 3 พารามิเตอร์และ 2 พารามิเตอร์) เป็นดังนี้ (E. Lee, 2013)

1) ประมาณค่าพารามิเตอร์แบบพหุมิติ

2) ประมาณทิศทางของการวัดที่ดีที่สุดในระดับแบบสอบสำหรับแบบสอบฉบับ X และ Y เวกเตอร์ของน้ำหนักที่สอดคล้องกับทิศทางของการวัดที่ดีที่สุดในระดับแบบสอบดังสมการ

$$\hat{\alpha}_k = \frac{\sum_{j=1}^N \hat{a}_{jk}}{\sqrt{\sum_{k=1}^{\delta} (\sum_{j=1}^N \hat{a}_{jk})^2}}$$

เมื่อ  $N$  คือ จำนวนข้อสอบทั้งหมดในแบบสอบ  $\delta$  คือ จำนวนมิติ

3) ประมาณค่าพารามิเตอร์แบบเอกมิตีที่สอดคล้อง



3.1) ประมาณค่าพารามิเตอร์อำนาจจำแนกแบบเอกมิติ

$$\hat{a}_{\alpha j} = (1 + \hat{\sigma}_{\alpha j}^2)^{-1/2} \hat{\mathbf{a}}_j^T \hat{\Sigma} \hat{\boldsymbol{\alpha}}$$

3.2) ประมาณค่าพารามิเตอร์ความยากแบบเอกมิติ

$$\hat{b}_{\alpha j} = \frac{-\hat{a}_{\alpha j}}{\hat{a}_{\alpha j}}$$

3.3) ประมาณค่าพารามิเตอร์โอกาสในการเดา (โมเดล 3 พารามิเตอร์เท่านั้น)

$$\hat{c}_{\alpha j} = \hat{c}_j$$

ความแปรปรวนของทิศทางสำหรับคอมโพสิทเชิงเส้น (standardized

linear composite:  $\boldsymbol{\theta}_\alpha$ ) เป็นดังสมการ

$$\hat{\sigma}_{\alpha j}^2 = \hat{\mathbf{a}}_j^T \hat{\Sigma} \hat{\mathbf{a}}_j - (\hat{\mathbf{a}}_j^T \hat{\Sigma} \hat{\boldsymbol{\alpha}})^2$$

เมื่อ  $\hat{\mathbf{a}}_j^T$  คือ เมทริกซ์ประมาณค่าอำนาจจำแนกของข้อสอบแบบพหุมิติ

สัมประสิทธิ์สำหรับทิศทางของคอมโพสิทเชิงเส้น ( $\boldsymbol{\alpha}$ ) เป็นการประมาณความสามารถแบบเอกมิติ ( $\boldsymbol{\theta}_\alpha$ ) ซึ่งถือว่าการประมาณความสามารถพหุมิติแบบเอกมิติ ซึ่งการประมาณแบบเอกมิติสามารถใช้กับขั้นตอนการเชื่อมโยงสเกลด้วยโมเดล MIRT หรือการปรับเทียบคะแนนด้วยโมเดล MIRT หลังจากการเชื่อมโยงสเกลด้วยโมเดล MIRT

คะแนนจริงของโมเดลการประมาณแบบเอกมิติสำหรับคอมโพสิทเชิงเส้น ( $\boldsymbol{\theta}_\alpha$ ) แทนด้วย  $\tau_\alpha$  โดยคะแนนจริงนี้เกี่ยวข้องกับคอมโพสิทเชิงเส้น ( $\boldsymbol{\theta}_\alpha$ ) ซึ่งก็คือ ผลรวมของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องในแต่ละระดับความสามารถ (composite ability level) ดังสมการ

$$\tau_\alpha = \xi(\boldsymbol{\theta}_\alpha) = \xi(\sum_{k=1}^{\hat{\sigma}} \alpha_k \theta_k)$$

- การประมาณแบบเอกมิติสำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า

งานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT ยังไม่มีการศึกษากับแบบสอบรูปแบบผสมและข้อสอบที่ตรวจให้คะแนนหลายค่า จึงไม่มีการระบุสมการสำหรับทิศทางของการวัดที่ดีที่สุดระดับแบบสอบ ( $\hat{\boldsymbol{\alpha}}_k$ ) และค่าพารามิเตอร์แบบเอกมิติ ( $\hat{a}_{\alpha j}$ ,  $\hat{b}_{\alpha j}$  และ  $\hat{c}_{\alpha j}$ ) สำหรับแบบสอบดังกล่าวไว้อย่างชัดเจน แต่ Zhang (2007) ได้ศึกษาความความแปรปรวนร่วมอย่างมีเงื่อนไขสำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า ผู้วิจัยระบุข้อสอบที่สร้างขึ้นภายใต้โมเดล Generalized multidimensional compensatory model โดยค่าฟังก์ชันการตอบสนองในแต่ละระดับ (item category response function) สำหรับข้อสอบข้อที่  $k$  เป็นดังสมการ (Jimmy Zhang, 2007)

$$\mu_j(\boldsymbol{\theta}) = H_j(\mathbf{a}_j^T \boldsymbol{\theta}_i) \equiv H_j(\sum_{k=1}^{\hat{\sigma}} a_{kj} \theta_k)$$

1. กระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ โดยใช้การประมาณแบบเอกมิติ (unidimensional approximation to MIRT true score equating procedure: ATSE)

หลังจากดำเนินการประมาณค่าแบบเอกมิติ (unidimensional approximation) แล้ว สามารถใช้กระบวนการ ATSE ปรับเทียบคะแนนจริงคอมโพสิท (composite true score;  $T_\alpha$ ) ของแบบสอบทั้งสองฉบับ โดยกระบวนการ ATSE เป็นการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ การปรับเทียบคะแนนจริงรวม (composite true score:  $T_\alpha$ ) บนแบบสอบพหุมิติมีขั้นตอนดังนี้

1. ระบุคะแนนจริงของแบบสอบฉบับใหม่ ( $T_{\alpha A}$ ) ที่เชื่อมโยงกับคะแนนจริงรวมของแบบสอบฉบับอ้างอิงต้องการ

2. ระบุระดับความสามารถ (unidimensional composite ability level,  $\theta_{\alpha}$ ) ที่สอดคล้องกับคะแนนจริงของแบบสอบใหม่คำนวณซ้ำโดยใช้ วิธีของ Newton-Raphson) จากสมการ

$$func(\theta_{\alpha i}) = \tau_{\alpha A} - \sum_{j:B} p_{ij}(\theta_{\alpha i}; a_{\alpha j}, b_{\alpha j}, c_j)$$

3. ระบุคะแนนจริงคอมโพสิท (composite true score) ของแบบสอบฉบับอ้างอิงที่สอดคล้องกับคะแนนจริงของแบบสอบฉบับใหม่ โดยเริ่มจากคำนวณ  $\theta_{\alpha i}$  ก่อนแล้วแทนค่าดังกล่าวลงในสมการคะแนนจริงของแบบสอบฉบับอ้างอิง โดยสมการคะแนนจริงของแบบสอบฉบับอ้างอิง คือ

$$\tau_{\alpha B}(\theta_{\alpha i}) = \sum_{j:B} p_{ij}(\theta_{\alpha i}; a_{\alpha j}, b_{\alpha j}, c_j)$$

2. กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณแบบเอกมิติ (unidimensional approximation to MIRT observed score equating procedure: AOSE)

กระบวนการ AOSE คล้ายคลึงกับการปรับเทียบคะแนนที่สังเกตได้แบบเอกมิติ (unidimensional observed score equating) มีขั้นตอนดังนี้

1. ประมาณค่าพารามิเตอร์ของข้อสอบและค่าความสามารถด้วยวิธีการที่อธิบายไว้ข้างต้น หลังจากนั้นกำหนดการแจกแจงตามเงื่อนไข (conditional distribution:  $f(x|\theta_\alpha)$ ) ณ แต่ละระดับของความสามารถ ( $\theta_\alpha$ ) โดยใช้สูตรเวียนซ้ำของ Lord-Wingersky

2. นำ Conditional distribution คูณด้วยการแจกแจงความสามารถแบบเอกมิติที่ถูกประมาณในประชากรของผู้สอบ

3. รวมพื้นที่ของความสามารถเอกมิติที่ถูกประมาณทั้งหมด ซึ่ง Marginal distribution ถูกคำนวณในแบบสอบแต่ละฉบับอีกครั้ง หลังจากนั้นปรับเทียบแบบสอบเหล่านี้โดยใช้วิธีอีควิเปอร์เซ็นไทล์แบบดั้งเดิม

เมื่อพิจารณาขั้นตอนของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติและพหุมิติทั้ง 4 กระบวนการ พบว่า ขั้นตอนของการปรับเทียบคะแนนสังเกตได้แบบเอกมิติและพหุมิตินี้มีความคล้ายคลึงกัน ได้แก่ (1) คำนวณ Conditional observed score distributions (2) นำการแจกแจง

คะแนนจากข้อ 1 คูณด้วยความหนาแน่นของความสามารถ (ability density) (3) จำนวน Marginal observed score distributions ด้วยการอินทิเกรตผลลัพธ์ทั้งหมดของข้อ 3 สำหรับแบบสอบทั้งสองฉบับ และ (4) ปรับเทียบคะแนนโดยใช้วิธีอีควิเปอร์เซ็นไทล์แบบดั้งเดิม เช่นเดียวกับขั้นตอนการปรับเทียบคะแนนจริงสำหรับข้อมูลเอกมิตีและพหุมิตีที่มีความคล้ายคลึงกัน ได้แก่ (1) ระบุคะแนนจริงของแบบสอบฉบับใหม่ (2) ระบุระดับความสามารถที่สอดคล้องกับคะแนนจริงของแบบสอบฉบับใหม่ (3) ระบุคะแนนจริง (composite true score) ของแบบสอบฉบับอ้างอิงที่สอดคล้องกับคะแนนจริงของแบบสอบฉบับใหม่ โดยผู้วิจัยสรุปขั้นตอนของการปรับเทียบคะแนนทั้ง 4 กระบวนการไว้ดังตาราง แต่อย่างไรก็ตามการปรับเทียบคะแนนสำหรับแบบสอบพหุมิตีและเอกมิตีย่อมมีความแตกต่างกัน ซึ่งผู้วิจัยจะกล่าวในส่วนต่อไป

**ตาราง 2.6** สรุปขั้นตอนของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิตีและพหุมิตี

ขั้นตอนของ UIRT equating procedure		ขั้นตอนของ MIRT equating procedure		
UIRTO		MOSE	Unidimensional approximation	
UIRTT			AOSE	ATSE
1. จำนวน Conditional observed score distributions; $(f(x \theta_j))$ ในแต่ละระดับความสามารถ $(\theta_j)$	1. ระบุคะแนนจริงจากแบบสอบฉบับใหม่ $(\tau_A)$ โดยคะแนนจริงคือผลรวมของความน่าจะเป็นในการตอบข้อสอบได้ ถูกต้องสำหรับข้อสอบแต่ละข้อ	1. จำนวน Conditional observed score distributions ในแต่ละกลุ่มของระดับความสามารถ $(f_r(x \theta))$ เมื่อ $\theta$ คือเวกเตอร์ของความสามารถ โดยใช้ Lord-Wingersky algorithm ตามกรอบแนวคิด MIRT - ใช้ R code ของ Brossman ปี 2010)	หลังจากการประมาณค่าพารามิเตอร์ข้อสอบและการแจกแจงความสามารถแบบเอกมิตี - ใช้ R นำค่าที่ประมาณได้มารวมไว้ในโปรแกรม Equating Recipe 1. จำนวน Conditional observed score distributions ในแต่ละระดับความสามารถ (composite ability level) โดยใช้ Lord-Wingersky algorithm	หลังจากการประมาณค่าพารามิเตอร์ข้อสอบและการแจกแจงความสามารถแบบเอกมิตี - ใช้ R นำค่าที่ประมาณได้มารวมไว้ในโปรแกรม Equating Recipe 1. ระบุคะแนนจริงของแบบสอบฉบับใหม่ที่สอดคล้องกับคะแนนจริงของแบบสอบฉบับอ้างอิง
2. นำการแจกแจงคะแนนจากข้อ 1 คูณด้วยความหนาแน่นของความสามารถ (ability density)	2. ระบุ Particular ability level $(\theta_j)$ ที่สอดคล้องกับคะแนนจริงของแบบสอบฉบับใหม่	2. นำการแจกแจงคะแนนจากข้อ 1 คูณด้วยความหนาแน่นของความสามารถ (multivariate ability density) - ใช้ R code ของ Brossman ปี 2010)	2. นำการแจกแจงจากข้อ 1 คูณด้วย Unidimensional ability distribution ที่ประมาณได้ของประชากรผู้สอบ	2. ระบุ Unidimensional composite ability level $(\theta_j)$ ที่สอดคล้องกับ Composite true score ของแบบสอบฉบับใหม่ โดยใช้กระบวนการทำซ้ำ

ขั้นตอนของ UIRT equating procedure		ขั้นตอนของ MIRT equating procedure		
UIRTO	UIRTT	MOSE	Unidimensional approximation	
			AOSE	ATSE
		* ability density มาจาก การจำลองข้อมูล ซึ่งอาจใช้ แพ็คเกจ mvtnorm ใน R		
3. คำนวณ Marginal observed score distributions ( $f(x)$ ) แบบสอบ ทั้งสองฉบับด้วยการ อินทิเกรตผลลัพธ์ จากข้อ 2 ทั้งหมดใน แต่ละระดับ ความสามารถสำหรับ แบบสอบทั้งสองฉบับ	3. ระบุคะแนน จริงของแบบ สอบฉบับอ้างอิง ที่สอดคล้องกับ $\theta$	3. คำนวณ Marginal observed score distributions ( $f(x)$ ) ด้วย การอินทิเกรตผลลัพธ์จากข้อ 2 ทั้งหมดในแต่ละระดับ ความสามารถ - ใช้ R code ของ Brossman ปี 2010) - ใช้ flexMIRT	3. คำนวณ Marginal observed score distributions สำหรับ แบบสอบทั้งสองฉบับด้วย การอินทิเกรตผลลัพธ์จาก ข้อ 2 ทั้งหมดในแต่ละ ระดับความสามารถ	3. ระบุ Composite true score ของแบบ สอบฉบับเก่าที่ สอดคล้องกับ Composite true score ของแบบสอบ ฉบับใหม่ ตามแนว นิยมของคะแนนจริง ใน UIRT
4. ปรับเทียบคะแนน ของแบบสอบทั้งสอง ฉบับด้วยวิธีอีควิ เปอร์เซ็นต์แบบ ดั้งเดิม		4. ใช้ Marginal observed score distributions มา ปรับเทียบคะแนนของแบบ สอบทั้งสองฉบับด้วยวิธีอีควิ เปอร์เซ็นต์แบบดั้งเดิม - RAGE-RGEQUATE สำหรับ รูปแบบ RG - Equating Recipe ใช้ได้ทุก รูปแบบ แต่เป็น C function	4. ปรับเทียบคะแนนของ แบบสอบทั้งสองฉบับด้วย วิธีอีควิเปอร์เซ็นต์แบบ ดั้งเดิม	

\* UIRTO = UIRT observed score equating, UIRTT = UIRT true score equating, MOSE = Full MIRT observed score equating procedure, AOSE= Unidimensional approximation of MIRT observed score equating, ATSE = Unidimensional approximation of MIRT true score equating, EE= Equipercentile equating method

เมื่อเปรียบเทียบกระบวนการปรับเทียบคะแนนสังเกตได้ระหว่างโมเดล UIRT และโมเดล MIRT พบว่าขั้นตอนการปรับเทียบคะแนนสังเกตได้ระหว่าง 2 กระบวนการเหมือนกัน คือ (1) คำนวณ Conditional observed score distributions (2) คำนวณ Marginal observed score distributions และรวมผลลัพธ์ทั้งหมด โดยการแจกแจงร่วมกันของคะแนนสังเกตได้สำหรับแบบสอบแต่ละฉบับหาได้จากผลคูณการแจกแจงอย่างมีเงื่อนไขคูณด้วยความหนาแน่นของความสามารถแบบพหุตัวแปร (multivariate ability density;  $\psi(\theta)$ ) และ (3) ปรับเทียบคะแนนแบบสอบทั้งสองฉบับด้วยวิธีอีควิเปอร์เซ็นต์ แต่พบว่ามี ความแตกต่างกันดังตาราง 2.7

ตาราง 2.7 เปรียบเทียบกระบวนการปรับเทียบคะแนนสังเกตได้ด้วยโมเดล UIRT และโมเดล MIRT

ประเด็น	UIRT observed equating procedure	FMIRT observed equating procedure
1. ความสามารถ	เป็นข้อมูลแบบสเกลาร์ ( $\theta_j$ )	เป็นข้อมูลแบบเวกเตอร์ ( $\theta_j$ )
2. ความน่าจะเป็นของการตอบข้อสอบได้	ความน่าจะเป็นของการตอบข้อสอบได้ ถูกต้องของข้อสอบข้อใดๆของบุคคลที่ $j$ ใน 1 มิติ	ความน่าจะเป็นของการตอบข้อสอบได้ ถูกต้องของข้อสอบข้อใดๆของบุคคลที่ $j$ มิติทั้งหมด $m$ มิติ
3. Conditional observed score distribution	ข้อสอบที่ตรวจให้คะแนนสองค่า ใช้สมการ Lord & Wingersky (1984) พิจารณาความสามารถของบุคคล 1 มิติ ข้อสอบที่ตรวจให้คะแนนหลายค่า ใช้สมการ Hanson & Thissen (1995) พิจารณาความสามารถของบุคคล 1 มิติ ข้อสอบรูปแบบผสม เป็นผลรวมของสมการเรียกซ้ำ (recursive formula) ของข้อสอบที่ตรวจให้คะแนนหลายสองค่าและหลายค่าบนเงื่อนไขความสามารถ 1 มิติ	ข้อสอบที่ตรวจให้คะแนนสองค่า พิจารณาความสามารถของบุคคลหลายมิติ ข้อสอบที่ตรวจให้คะแนนสองค่า ใช้สมการ Hanson & Thissen (1995) พิจารณาความสามารถของบุคคลหลายมิติ ข้อสอบรูปแบบผสม เป็นผลรวมของสมการเรียกซ้ำ (recursive formula) ของข้อสอบที่ตรวจให้คะแนนหลายสองค่าและหลายค่าบนเวกเตอร์เงื่อนไขความสามารถ ( $\theta_j$ ) หลายมิติ
4. Marginal observed score distribution	เป็นผลรวมของการแจกแจงคะแนนอย่างมีเงื่อนไขคุณกับความหนาแน่นของค่าความสามารถในมิติเดียว (แบบตัวแปร)	เป็นผลรวมของการแจกแจงคะแนนอย่างมีเงื่อนไขคุณกับความหนาแน่นของค่าความสามารถในหลายมิติ (แบบหลายตัวแปรในกรณีของ SS-MIRT เป็นแบบสองตัวแปร)

## 2.5 การประเมินประสิทธิภาพของการปรับเทียบคะแนน

ประสิทธิภาพของการปรับเทียบคะแนนพิจารณาได้จากดัชนีต่าง ๆ ไม่ว่าจะเป็น ความคลาดเคลื่อนมาตรฐานในการปรับเทียบคะแนน (SE) ความลำเอียงของการปรับเทียบคะแนน (bias) รากที่สองของความคลาดเคลื่อนยกกำลังสอง (RMSE) ความคลาดเคลื่อนยกกำลังสองเฉลี่ย (MSE) รากที่สองของความแตกต่างยกกำลังสอง (RMSD) และ Difference That Matters (DTM) มีรายละเอียดดังนี้

### 2.5.1 ความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน (standard error of equating: SE)

การระบุขนาดของความคลาดเคลื่อนอย่างสุ่มในการปรับเทียบคะแนนที่เกิดจากการสุ่มตัวอย่างพิจารณาจากความคลาดเคลื่อนมาตรฐานในการปรับเทียบคะแนน (SE) ค่า SE ในแต่ละจุดคะแนนเป็นดังสมการ (E. Lee, 2013; O. Zhang, 2012)

$$SE_i = \sqrt{\frac{1}{N} \sum_{k=1}^N [\hat{e}_{base_k}(x_i) - \bar{e}_{base}(x_i)]^2}$$

เมื่อ  $\hat{e}_{base}(x_i)$  คือ คะแนนดิบที่ถูกปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา  
 $\bar{e}_{base}(x_i)$  คือ ค่าเฉลี่ยของคะแนนดิบที่ถูกปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา  
 $N$  คือ จำนวนของการทำซ้ำ

ค่าความคลาดเคลื่อนในการปรับเทียบคะแนนสามารถพิจารณาในรูปของสัมประสิทธิ์การแปรผัน (coefficient of variance) ดังสมการ

$$CVSE = \frac{SE_i}{\bar{e}_{base}(x_i)} \times 100$$

### 2.5.2 ความลำเอียงของการปรับเทียบคะแนน (bias)

การระบุขนาดของความคลาดเคลื่อนอย่างเป็นระบบในการปรับเทียบคะแนนพิจารณาจากความลำเอียงของการปรับเทียบคะแนน ความลำเอียงของการปรับเทียบคะแนน คือ ค่าเฉลี่ยของความแตกต่างระหว่างฟังก์ชันของกระบวนการ/วิธีการปรับเทียบคะแนนที่ศึกษากับฟังก์ชันของการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์ ซึ่งความลำเอียงของการปรับเทียบคะแนนในแต่ละจุดคะแนน ดังสมการ

$$Bias_i = \frac{\sum_{k=1}^N [\hat{e}_{base}(x_i) - e_{base}(x_i)]}{N} = \frac{\sum_{k=1}^N [\hat{e}_{base}(x_i)]}{N} - \frac{\sum_{k=1}^N [e_{base}(x_i)]}{N}$$

เมื่อ  $N$  คือ จำนวนของการทำซ้ำ  
 $x_i$  คือ คะแนนที่  $i$   
 $e_{base}(x_i)$  คือ คะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนเกณฑ์  
 $\hat{e}_{base}(x_i)$  คือ คะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา

ดังนั้น ความลำเอียงในแต่ละจุดคะแนน เป็นดังสมการ

$$Bias_i = \overline{\hat{e}_{base}(x_i)} - \overline{e_{base}(x_i)}$$

เมื่อ  $\overline{\hat{e}_{base}(x_i)}$  คือ ค่าเฉลี่ยของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา

$\overline{e_{base}(x_i)}$  คือ ค่าเฉลี่ยของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนเกณฑ์

ค่าความลำเอียงสามารถพิจารณาในรูปของค่าความลำเอียงสัมพัทธ์ (relative bias) ดังสมการ

$$RB = \frac{|\overline{\hat{e}_{base}(x_i)} - \overline{e_{base}(x_i)}|}{\sigma_{base}(x_i)} \times 100$$

### 2.5.3) รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (root mean square error: RMSE)

รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง คือ รากที่สองของผลรวมของความคลาดเคลื่อนและความลำเอียงยกกำลังสอง

$$RMSE_i = \sqrt{SEE_i^2 + Bias_i^2}$$

#### วิธีการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์

การคำนวณความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน (SE) ความลำเอียงของการปรับเทียบคะแนน (bias) และรากที่สองของความคลาดเคลื่อนยกกำลังสอง (RMSE) ต้องระบุการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์เพื่อหาความสัมพันธ์การปรับเทียบคะแนนที่แท้จริง (true equating relationships) ในสถานการณ์จริงเราไม่ทราบความสัมพันธ์ของการปรับเทียบประชากรดังกล่าว จึงต้องประมาณความสัมพันธ์ของการปรับเทียบประชากรจากกลุ่มตัวอย่างที่เท่าเทียม (target equivalent) งานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT ที่ผ่านมามีส่วนใหญ่ใช้วิธีอีควิเปอร์เซ็นไทล์ที่ปรับโค้งให้ราบเรียบ (pre-smoothed equipercentile equating method) และวิธี Identity equating

การปรับเทียบคะแนนด้วยโมเดล MIRT สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT ใช้วิธีอีควิเปอร์เซ็นไทล์ที่ปรับโค้งให้ราบเรียบเป็นเกณฑ์เพื่อหาความสัมพันธ์ของการปรับเทียบคะแนนประชากร โดยวิธีอีควิเปอร์เซ็นไทล์ที่ปรับโค้งให้ราบเรียบนิยมใช้เป็นเกณฑ์เนื่องจากไม่จำเป็นต้องคำนึงถึงมิติของข้อมูล และผลการปรับเทียบคะแนนด้วยวิธีดังกล่าวยังสอดคล้อง (align) อย่างมากกับวิธีการปรับเทียบคะแนนด้วยโมเดล MIRT เนื่องจากลดค่า likelihood ที่จะทำให้เกิดความลำเอียง นอกจากนี้วิธีอีควิเปอร์เซ็นไทล์ดังกล่าวยังมีความคล้ายคลึงกับวิธี IRT เนื่องจากสหสัมพันธ์ระหว่างของการปรับเทียบคะแนนเป็นเส้นโค้ง (Brossman, 2010) ส่วนรูปแบบผู้สอบกลุ่มสุ่ม ผู้วิจัยเลือกใช้วิธี Identity equating เป็นเกณฑ์ (Peterson, 2014) เช่นเดียวกับ

งานวิจัยที่ศึกษาการปรับเทียบคะแนนสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม ซึ่งเลือกใช้วิธี Identity equating เช่นกัน (E. Lee, 2013)

วิธี Identity equating ถือว่าคะแนนแบบสอบฉบับ X เท่าเทียมกับคะแนนแบบสอบฉบับ Y ที่เท่ากัน (identical score) ดังนั้นคะแนน Y ที่ปรับเทียบไปยัง X จึงเท่ากับคะแนน X ดังนั้นถ้านำคะแนน X ลบด้วยคะแนน X ที่ปรับเทียบไปยัง Y ย่อมเท่ากับศูนย์ (กราฟ identity จะเป็นเส้นตรง) (Peterson, 2014) ดังนั้นวิธี Identity equating จึงเป็นการปรับเทียบคะแนนไปยังตัวเอง (equating a test to itself) แต่อย่างไรก็ตามวิธีการนี้มีข้อจำกัด คือ ใช้ได้กับวิธีการปรับเทียบคะแนนที่มีจำนวนพารามิเตอร์ไม่มากนัก และ Kolen & Brennan (2014) แนะนำว่าควรใช้วิธีการนี้เพื่อตรวจสอบวิธีปรับเทียบคะแนนที่ไม่มีประสิทธิภาพจะเหมาะสมกว่าการระบุวิธีการปรับเทียบคะแนนที่ดีที่สุด

นอกจากนี้การปรับเทียบคะแนนด้วยโมเดล MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT (O. Zhang, 2012) ใช้วิธีการประมาณค่าความถี่ (frequency estimation method) เป็นเกณฑ์ โดยผู้วิจัยให้เหตุผลว่าวิธีนี้ใช้เพียงคะแนนรวมและงานวิจัยดังกล่าวไม่ละเมิดข้อตกลงเบื้องต้นของวิธีการประมาณค่าความถี่ โดยวิธีการดังกล่าวถือว่าการแจกแจงของแบบสอบทั้งสองฉบับที่มีเงื่อนไขเป็นคะแนนข้อสอบรวมไม่แปรผันไปตามประชากร (invariant population)

โดยสรุป การปรับเทียบคะแนนด้วยโมเดล MIRT ที่ผ่านมาส่วนใหญ่ใช้วิธีอีควิเปอร์เซ็นต์ที่ปรับโค้งให้ราบเรียบ (pre-smoothed equipercentile equating method) เป็นเกณฑ์เพื่อหาความสัมพันธ์ของการปรับเทียบคะแนนประชากร (true equating relationships) ซึ่งวิธีดังกล่าวนำไปใช้กับแบบสอบรูปแบบผสม ข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม และรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกัน โดยใช้ข้อสอบรวม และเนื่องจากงานวิจัยครั้งนี้ต้องการเปรียบเทียบประสิทธิภาพของกระบวนการ MOSE ในเงื่อนไขต่าง ๆ เพื่อหาวิธีการปรับเทียบคะแนนที่เหมาะสมสำหรับบริบทที่ศึกษา ไม่ได้มีจุดประสงค์เพียงเพื่อตรวจสอบวิธีปรับเทียบที่ไม่มีประสิทธิภาพ ดังนั้นงานวิจัยในครั้งนี้จึงเลือกใช้วิธีอีควิเปอร์เซ็นต์ที่ปรับโค้งให้ราบเรียบ (pre-smoothed equipercentile equating method) เป็นเกณฑ์

วิธีปรับโค้งให้ราบเรียบ (smoothing method) เป็นการประมาณการแจกแจงเชิงประจักษ์และความสัมพันธ์อีควิเปอร์เซ็นต์ที่มีการแจกแจงไม่ราบเรียบซึ่งเป็นลักษณะเฉพาะของประชากร โดยคาดหวังว่าผลจากการประมาณดังกล่าวจะมีความแม่นยำมากกว่าการแจกแจงที่ยังไม่ได้ปรับให้ราบเรียบ (unsmoothed distribution) (Kolen & Brennan, 2004)

วิธี Presmoothing เป็นการปรับการแจกแจงของคะแนนให้เป็นโค้งราบเรียบ มีหลักสำคัญ คือ Moment preservation ที่ระบุว่า การแจกแจงที่ราบเรียบต้องมี Central moment เหมือนกับการแจกแจงที่สังเกตได้ กล่าวคือ ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของการแจกแจงที่ปรับให้เป็นโค้งราบเรียบและการแจกแจงที่ยังไม่ได้ปรับให้เป็นโค้งราบเรียบต้องเท่ากัน โดยวิธี Presmoothing อาจใช้วิธี Polynomial log-linear model หรือวิธี Strong true score distribution model หลังจาก



Presmoothing ข้อมูลด้วยสองวิธีดังกล่าวแล้ว สามารถปรับเทียบคะแนนจากแบบสอบ X ไปยังคะแนนจากแบบสอบ Y ได้ โดยใช้การแจกแจงที่ถูกปรับให้เป็นโค้งราบเรียบแล้ว

วิธี Polynomial log-linear method ใช้โมเดลล็อกเชิงเส้น (log-linear model) ที่มีดีกรีเหมาะสมในการประมาณการแจกแจงคะแนนสอบ โดยพิจารณาความสอดคล้องระหว่างฟังก์ชัน Polynomial กับล็อกของความหนาแน่นของกลุ่มตัวอย่าง (sample density) โมเดลล็อกเชิงเส้นเป็นดังสมการ  $\log(N_x f(x)) = \omega_0 + \omega_1 x_i^1 + \omega_2 x_i^2 + \dots + \omega_3 x_i^C$

จากสมการข้างต้น ล็อกของความหนาแน่นแสดงอยู่ในรูปของดีกรี Polynomial ที่ต่ำที่สุด คือ C โดยค่าพารามิเตอร์  $\beta$  ประมาณได้จากวิธี Maximum likelihood และผลจากการแจกแจงที่สอดคล้อง (fitted distribution) มีคุณสมบัติ Moment preservation เนื่องจากโมเมนต์ C ลำดับแรกของการแจกแจงที่สอดคล้องเหมือนกับโมเมนต์ C ของการแจกแจงของกลุ่มตัวอย่าง (sample distribution) กล่าวคือ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการแจกแจงที่สอดคล้องเหมือนกับค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการแจกแจงของกลุ่มตัวอย่าง

การตรวจสอบความสอดคล้องของการแจกแจงคะแนนสำหรับวิธี Log-linear ควรพิจารณาทั้งดัชนีทางสถิติและกราฟ กล่าวคือ (1) เปรียบเทียบกราฟระหว่างการแจกแจงที่สอดคล้องกับการแจกแจงของกลุ่มตัวอย่าง ซึ่งการแจกแจงที่สอดคล้องควรมีความราบเรียบ (smooth) แต่ไม่แตกต่างจากการแจกแจงของกลุ่มตัวอย่างมากเกินไป (2) เปรียบเทียบค่าไค-สแควร์:  $\chi^2_C - \chi^2_{C+1}$  เช่น เปรียบเทียบค่าไค-สแควร์ของ  $C = 2$  และ  $3$  กับดีกรีอิสระ 1 ถ้ามีความแตกต่างกันอย่างมีนัยสำคัญแสดงว่า โมเดลที่มีดีกรี C มากกว่า (เช่น  $C = 3$ ) มีความสอดคล้องมากกว่าโมเดลที่มีดีกรีน้อยกว่า ( $C = 2$ ) และ (3) เปรียบเทียบดัชนีความสอดคล้อง  $\chi^2_C$  ในภาพรวม โดยเลือกดีกรีที่ไม่มีนัยสำคัญ

แพ็คเกจ “**equate**” โมเดลล็อกเชิงเส้นถูกทำให้สอดคล้องโดยใช้ฟังก์ชัน smoothed = “**lodlinear**” โดยวิธีการระบุโมเดลล็อกเชิงเส้นในฟังก์ชัน presmoothing ที่มีประสิทธิภาพมากที่สุด คือ การระบุเทอมของ polynomial สูงสุดสำหรับตัวแปรแต่ละตัวในแต่ละระดับของปฏิสัมพันธ์ด้วยการระบุ degree=list(c(3,3), c(1,1)) หรือเขียนแบบสั้นว่า degree = list (3,1)

#### 2.5.4) ความคลาดเคลื่อนยกกำลังสองเฉลี่ย (mean square error: MSE)

ความคลาดเคลื่อนยกกำลังสองเฉลี่ย คือ ผลรวมของค่าความลำเอียง ( $Bias_i$ ) และความแปรปรวน ( $VAR_i$ ) ดังสมการ (W.-C. Lee & Brossman, 2012)

$$MSE_i = Bias_i + VAR_i$$

ความแปรปรวน ณ จุดคะแนนดิบ ( $VAR_i$ ) คำนวณจาก

$$VAR_i = \frac{1}{R} \sum_{r=1}^R \left[ \hat{e}_{xr} - \left( \frac{1}{R} \sum_{r=1}^R \hat{e}_{xr} \right) \right]^2$$

เมื่อ  $\hat{e}_{xr}$  คือ คะแนนที่ถูกปรับเทียบโดยประมาณ ณ คะแนนที่  $x$  ในการทำซ้ำครั้งที่  $r$

$e_x$  คือ คะแนนจริงที่ถูกปรับเทียบ ณ คะแนนที่  $x$

จะเห็นว่าค่ารากที่สองของค่าความคลาดเคลื่อนยกกำลังสองเฉลี่ย (MSE) คือ รากที่สองของความคลาดเคลื่อนยกกำลังสอง (RMSE) นั่นเอง

### 2.5.5) รากที่สองของความแตกต่างยกกำลังสอง (root mean square deviation: RMSD)

ความแตกต่างระหว่างการปรับเทียบคะแนนด้วยโมเดล MIRT กับการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์ สามารถพิจารณาได้จากความแตกต่างยกกำลังสองในแต่ละจุดคะแนน ดังสมการ

$$RMSD_i = \sqrt{\frac{1}{N} \sum_{k=1}^N [eq_{Y_{IRT}}(x_i) - eq_{Y_C}(x_i)]^2}$$

เมื่อ  $eq_{Y_{IRT}}(x_i)$  คือ คะแนนดิบ  $i$  ที่ถูกปรับเทียบโดยใช้ MIRT equating procedure

$eq_{Y_C}(x_i)$  คือ คะแนนดิบ  $i$  ที่ถูกปรับเทียบโดยใช้การปรับเทียบคะแนนที่ใช้เป็นเกณฑ์

$w_i$  คือ น้ำหนักที่สัมพันธ์กับคะแนนดิบที่  $i$  บนแบบสอบฉบับใหม่

$z$  คือ คะแนนดิบคอมโพสิทสูงสุด (maximum raw composite score)

$x_j$  คือ คะแนนดิบที่  $i$  บนแบบสอบฉบับใหม่

โดยค่า  $RMSD_i$  เท่ากับผลรวมของความลำเอียงของการปรับเทียบคะแนน (equating bias) และความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน (SEE)

การพิจารณาความแตกต่างระหว่างฟังก์ชันการปรับเทียบคะแนนที่ศึกษากับฟังก์ชันการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์สามารถพิจารณาจากรากที่สองของความแตกต่างยกกำลังสองเฉลี่ยที่ถ่วงน้ำหนักแล้ว (weight root mean square different:  $ARMSD_w$ ) ดังสมการ (Peterson, 2014)

$$ARMSD_w = \sum_{x=1}^N RMSD[eq_{Y_C}(x_j)]P(x_i)$$

เมื่อ  $P(x_i)$  คือ สัดส่วนประชากรของผู้สอบจากประชากรเป้าหมายที่มีคะแนนดิบ  $x_i$  บนแบบสอบที่ได้รับการปรับเทียบ

### 2.5.6) Difference That Matters (DTM)

ความแตกต่างระหว่างคะแนนที่ถูกปรับจากวิธีการปรับเทียบคะแนนที่ศึกษากับวิธีการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์สามารถใช้เปรียบว่าวิธีการปรับเทียบคะแนนใดมีประสิทธิภาพมากกว่าหรือดีที่สุด โดยการเฉลี่ยความแตกต่างดังกล่าวในทุกช่วงคะแนน เพื่อหาค่าสถิติค่าเดียวสำหรับการปรับเทียบคะแนนแต่ละวิธี และความแตกต่างสำหรับการปรับเทียบคะแนนแต่ละวิธีจะถูกนำไปเปรียบเทียบกับ DTM โดย Doran และคณะ (2003 อ้างอิงใน Brossman และ Lee, 2013) ระบุว่า DTM คือ ความแตกต่างของคะแนน (number-correct score) ระหว่างวิธีการปรับเทียบคะแนน ส่วนใหญ่นิยมใช้ค่า  $DTM = 0.5$

## 2.6 โปรแกรมที่ใช้ปรับเทียบคะแนนภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม

### 2.6.1) โปรแกรม Equating Recipes

โปรแกรม Equating Recipes ใช้ภาษา C วิเคราะห์ข้อมูล โดยฟังก์ชัน C ของโปรแกรม สอดคล้องกับแนวคิดของ Kolen และ Brennan (2004) โดยโปรแกรมนี้สามารถใช้กับรูปแบบผู้สอบกลุ่ม สุ่ม ผู้สอบกลุ่มเดียว และผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม วิธีการปรับเทียบคะแนน ได้แก่ วิธีค่าเฉลี่ย วิธีการปรับเทียบเชิงเส้น วิธีอีควิเปอร์เซนไทล์ และวิธีการปรับเทียบคะแนนตามทฤษฎีการ ตอบสนองข้อสอบ ประเภทของการปรับโค้งให้ราบเรียบ ได้แก่ ไม่ปรับโค้งให้ราบเรียบ, Beta binomial presmoothing 2 หรือ 4 พารามิเตอร์, Log-linear presmoothing, Cubic-spline postsmoothing, Kernal และ continuized log-linear (Brennan, Wang, Kim, & Seol, 2009)

### 2.6.2) โปรแกรม IRTEQ

โปรแกรม IRTEQ ใช้ได้กับวิธีค่าเฉลี่ย/ค่าเฉลี่ย วิธีค่าเฉลี่ย/ซิกมา วิธีค่าเฉลี่ย/ซิกมาที่ปรับแก้แล้ว และวิธีโค้งคุณลักษณะแบบสอบ โปรแกรมนี้ใช้ได้กับโมเดลการตอบข้อสอบที่ตรวจให้คะแนนสองค่าและ โมเดลการตอบสนองข้อสอบแบบหลายค่า ได้แก่ Generalized Partial Credit Model, Partial Credit Model, และ Graded Response Model โปรแกรมดังกล่าวยังใช้ได้กับการรวมกัน/การผสมกันของ โมเดล IRT เพื่อประมาณค่าพารามิเตอร์ข้อสอบแบบสอบแต่ละฉบับ และสามารถดึงข้อมูลค่าพารามิเตอร์ ข้อสอบและข้อมูลคะแนนจากโปรแกรม WinGen และ/หรือโปรแกรม PASCAL ได้อีกด้วย (Han, 2009)

### 2.6.3) Package “equate” ในโปรแกรม R

Package “equate” ประกอบด้วยฟังก์ชันสำหรับการเชื่อมโยงสเกลและการปรับเทียบคะแนน สังเกตได้ภายใต้รูปแบบผู้สอบกลุ่มเดียว ผู้สอบกลุ่มเท่าเทียมกัน ผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบ ร่วม โดย Package “equate” ใช้ได้กับฟังก์ชันการเชื่อมโยงสเกลและการปรับเทียบคะแนนต่าง ๆ ได้แก่ ฟังก์ชัน General-linear ฟังก์ชัน synthetic และฟังก์ชัน circle-arc และใช้ได้กับวิธีการปรับเทียบคะแนน แบบดั้งเดิม ไม่ว่าจะเป็นวิธี Levine วิธี Tucker และ วิธี Braun/Holland (Albano, 2014)

การวิจัยครั้งนี้ใช้ Package “equate” ในการปรับเทียบคะแนนด้วยกระบวนการ MOSE ในขั้นตอน ของการปรับเทียบคะแนนด้วยวิธีอีควิเปอร์เซนไทล์แบบดั้งเดิม และการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์ ซึ่ง ก็คือ วิธีอีควิเปอร์เซนไทล์แบบดั้งเดิมที่มีการปรับโค้งให้ราบเรียบก่อน

### 2.6.4) Package “SNSequate” ในโปรแกรม R

Package “equate” ใช้กับ (1) วิธีการปรับเทียบคะแนนสังเกตได้ ได้แก่ วิธีค่าเฉลี่ย วิธีการ ปรับเทียบเชิงเส้น วิธี Local equating วิธี Kernal และวิธีการปรับเทียบอีควิเปอร์เซนไทล์ และ (2) วิธีการ ปรับเทียบคะแนนและเชื่อมโยงสเกลด้วยโมเดล IRT โดยการเชื่อมโยงสเกลใช้วิธีค่าเฉลี่ย-ค่าเฉลี่ย วิธีค่าเฉลี่ย- ซิกมา และวิธีโค้งคุณลักษณะ หลังจากการประมาณค่าพารามิเตอร์และค่าพารามิเตอร์อยู่ในสเกลเดียวกัน

จึงเปรียบเทียบคะแนนด้วยโมเดล IRT ได้ 2 แบบ คือ การเปรียบเทียบคะแนนจริงด้วยโมเดล IRT และการเปรียบเทียบคะแนนสังเกตได้ด้วยโมเดล IRT โดย Package “equate” ใช้ได้กับการเก็บรวบรวมข้อมูลทุกรูปแบบ (Gonzalez, 2014)

### 2.6.5) Package “kequate” ในโปรแกรม R

Package “kequate” ใช้สำหรับวิธี Kernal สำหรับรูปแบบการปรับเทียบคะแนนทุกรูปแบบ ซึ่งแพ็คเกจนี้สามารถค้นหาโมเดลล็อกเชิงเส้น (log-liner) ที่เหมาะสมสำหรับขั้นตอน Pre-smoothing ได้ด้วยฟังก์ชัน glm (Andersson, Branberg, & Wiberg, 2013)

*การวิจัยครั้งนี้ ผู้วิจัยใช้โปรแกรม R ด้วย Package “equate” สำหรับการปรับเทียบคะแนนด้วยกระบวนการ MOSE ในขั้นตอนสุดท้ายซึ่งเป็นการปรับเทียบคะแนนด้วยวิธีอีควิเปอร์เซ็นต์แบบดั้งเดิมด้วย Marginal observed score distributions และใช้สำหรับการปรับเทียบคะแนนที่ใช้เกณฑ์ คือ วิธีอีควิเปอร์เซ็นต์ที่มีการปรับโค้งให้ราบเรียบก่อน*

## 2.7 งานวิจัยเกี่ยวกับกระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

Brossman (2010) และ Brossman และ Lee (2013) พัฒนาการปรับเทียบคะแนนด้วยโมเดล MIRT 3 วิธี สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่มเท่าเทียมกัน ประกอบด้วย กระบวนการปรับเทียบคะแนนด้วยคะแนนสังเกตได้ 2 กระบวนการ ได้แก่ กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (MOSE) และ กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (AOSE) และกระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (ATSE) โดยเปรียบเทียบกับกระบวนการปรับเทียบคะแนนด้วยโมเดล UIRT และวิธีอีควิเปอร์เซ็นต์ (equipercentile equating: EE) งานวิจัยครั้งนี้ใช้ข้อมูลจริงที่มีความเป็นพหุมิติต่ำ

ผลการวิจัยพบว่า เมื่อข้อมูลไม่ได้เป็นเอกมิติแบบเข้มงวด กระบวนการ FMIRT ทั้ง 3 วิธี มีความเหมาะสมมากกว่ากระบวนการปรับเทียบคะแนน UIRT เนื่องจากกระบวนการดังกล่าวให้ผลการปรับเทียบใกล้เคียงกับวิธีอีควิเปอร์เซ็นต์ และมีความคลาดเคลื่อนในการปรับเทียบคะแนนน้อยกว่าเมื่อใช้ข้อมูลแบบพหุมิติ ขณะที่กระบวนการ ATSE และ AOSE ให้ผลการปรับเทียบคะแนนใกล้เคียงกับกระบวนการ MOSE ผู้วิจัยระบุข้อจำกัดและข้อเสนอแนะไว้ว่า (1) ปัจจุบันยังไม่มีกำหนดวิธีการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์อย่างสากล งานวิจัยนี้เลือกใช้วิธีอีควิเปอร์เซ็นต์ทำให้มีข้อจำกัดเกี่ยวกับมาตรฐานของการเปรียบเทียบ (2) ข้อมูลจากการทดสอบจริงมีความเป็นพหุมิติในระดับต่ำถึงปานกลาง ควรศึกษากับระดับความเป็นพหุมิติอื่น ๆ ด้วยการจำลองข้อมูล (3) การวิจัยครั้งนี้มีการจัด

กระทำเพียง 3 ครั้ง กลุ่มตัวอย่างจึงมีขนาดเล็ก หากใช้การจำลองข้อมูลควรมีการทำซ้ำ 100 ครั้งหรือ 1,000 ครั้ง (4) อาจศึกษากับตัวแปรอื่น ๆ และรูปแบบ NEAT

Lee และ Brossman (2012) พัฒนากระบวนการปรับเทียบคะแนนสังเกตได้ โดยใช้กรอบแนวคิด ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติแบบโครงสร้างอย่างง่าย (observed-score equating procedure using a simple-structure multidimensional item response theory: SS-MIRT) สำหรับ รูปแบบผู้สอบกลุ่มสุ่ม โดยใช้แบบสอบรูปแบบผสม ประกอบด้วยกลุ่มของข้อสอบซึ่งวัดความสามารถแตกต่างกัน แต่ข้อสอบภายในกลุ่มวัดความสามารถเด่นเดี่ยวเหมือนกัน และเปรียบเทียบผลการปรับเทียบ ของกระบวนการปรับเทียบคะแนน SS-MIRT กับกระบวนการปรับเทียบคะแนน UIRT และวิธีอีควิเพอร์เซ็นต์ไทล์ งานวิจัยนี้ใช้การจำลองข้อมูล

ผลการวิจัย พบว่า กระบวนการ SS-MIRT ให้ผลการปรับเทียบที่ดีกว่ากระบวนการ UIRT เนื่องจาก ผลของกระบวนการ SS-MIRT ใกล้เคียงกับผลของวิธีอีควิเพอร์เซ็นต์ไทล์มากกว่ากระบวนการ UIRT กล่าวคือ กระบวนการ SS-MIRT มีประสิทธิภาพอยู่ในเกณฑ์ที่ยอมรับได้ เมื่อความสัมพันธ์ระหว่างมิติต่ำ และความลำเอียงมีแนวโน้มเพิ่มขึ้น เมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้น ส่วนกระบวนการ UIRT มีประสิทธิภาพ เมื่อมิติความสามารถมีความสัมพันธ์กันสูง ผู้วิจัยระบุข้อจำกัดและข้อเสนอแนะไว้ว่า (1) ควรศึกษาผลของ ระดับความแตกต่างในโครงสร้างอย่างง่ายที่มีต่อการปรับเทียบคะแนนเมื่อใช้กระบวนการปรับเทียบ SS-MIRT (2) ควรพิจารณาปัจจัยอื่นด้วย เช่น ขนาดของกลุ่มตัวอย่าง จำนวนมิติ และโมเดล IRT อื่น ๆ

Zhang (2012) เปรียบประสิทธิภาพของกระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนอง ข้อสอบแบบพหุมิติสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดย ใช้ข้อสอบรวม (NEAT) โดยใช้วิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT 5 วิธี ได้แก่ วิธี OD วิธี TCF วิธี ICF วิธี M และวิธี NOP และกระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ 3 กระบวนการ ได้แก่ (1) MOSE (2) AOSE และ (3) ATSE โดยการวิจัยครั้งนี้ใช้การจำลองข้อมูลซึ่งมี 4 ขั้นตอนได้แก่ การจำลองคำตอบของผู้สอบ การประมาณค่าพารามิเตอร์แบบแยกกัน การเชื่อมโยงสเกล แบบสอบ และการปรับเทียบคะแนน

ผลการวิจัยพบว่า กระบวนการ AOSE มีประสิทธิภาพในการปรับเทียบคะแนนมากกว่า อีก 2 กระบวนการ และกระบวนการปรับเทียบภายใต้วิธี OD วิธี TCF วิธี ICF มีประสิทธิภาพมากกว่า วิธีการเชื่อมโยงสเกลอีก 2 วิธี โดยการปรับเทียบคะแนนภายใต้วิธี NOP มีประสิทธิภาพต่ำที่สุด นอกจากนี้ ยังพบว่า ความแตกต่างของค่าเฉลี่ยความสามารถของกลุ่มผู้สอบมีผลทางลบต่อการปรับเทียบคะแนน ผู้วิจัยระบุข้อจำกัดและข้อเสนอแนะไว้ว่า (1) งานวิจัยนี้ศึกษาตัวแปรด้านแบบสอบและผู้สอบ จึงควร ศึกษาเกี่ยวกับตัวแปรอื่น ๆ เพิ่มเติม (2) การประมาณค่าพารามิเตอร์ข้อสอบที่ตรวจให้คะแนนสองค่าใช้โปรแกรม TESTFACT ซึ่งมีข้อจำกัดเกี่ยวกับการหมุนแกน จึงควรเลือกการหมุนรูปแบบอื่น ๆ ด้วยการเลือกใช้ โปรแกรมประมาณค่าพารามิเตอร์ที่แตกต่างออกไป โดยเฉพาะอย่างยิ่งเมื่อแบบมีความสัมพันธ์ระหว่างมิติ

สูงภายใต้รูปแบบ NEAT การเลือกการหมุนในการประมาณค่าพารามิเตอร์ด้วยโมเดล MIRT ยิ่งมีความซับซ้อน ซึ่งอาจเลือกใช้โปรแกรมอื่นๆ เช่น Mplus, BMIRT และ IRTPRO (3) การปรับเทียบที่ใช้เป็นเกณฑ์ใช้วิธีการประมาณความถี่ ซึ่งใช้กรอบแนวคิดการปรับเทียบคะแนนสังเกตได้ภายใต้รูปแบบ NEAT อาจเลือกใช้วิธีการปรับเทียบอื่นมาเป็นเกณฑ์ (4) ควรศึกษาประเภทของการหมุนในวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT

Lee (2013) เปรียบเทียบผลของกระบวนการปรับเทียบคะแนนสำหรับแบบสอบพหุมิติ 6 วิธี ได้แก่ (1) UIRTO (2) UIRTT (3) MOSE (4) AOSE (5) ATSE และ (6) EE งานวิจัยนี้ศึกษากับข้อมูลจำลองที่สร้างจากกรอบแนวคิด MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า โดยใช้โมเดล Bifactor ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม พร้อมทั้งศึกษาปัจจัย 4 ด้าน ได้แก่ ความยาวของแบบสอบ ขนาดของกลุ่มตัวอย่าง ความแตกต่างของระดับความยาก และความสัมพันธ์ระหว่างมิติ ซึ่งมีการประมาณค่าพารามิเตอร์สำหรับแบบสอบแต่ละฉบับแยกกัน

ผลการวิจัยพบว่า เมื่อมิติมีความสัมพันธ์กันต่ำ การปรับเทียบคะแนนด้วยกระบวนการ MOSE มีความถูกต้องมากกว่ากระบวนการอื่น ๆ แต่กระบวนการ UIRT 2 กระบวนการให้ผลการปรับเทียบในเกณฑ์ที่ยอมรับได้เมื่อใช้ข้อสอบแบบพหุมิติ วิธีอีควิเปอร์เซ็นไทล์มีขนาดของความคลาดเคลื่อนแบบสุ่มและความคลาดเคลื่อนแบบภาพรวมตลอดเงื่อนไขมากกว่าวิธี IRT โดยผลของกระบวนการปรับเทียบคะแนนคะแนนจริงและคะแนนสังเกตได้ไม่แตกต่างกัน กระบวนการปรับเทียบสำหรับแบบสอบพหุมิติมีความถูกต้องมากขึ้นเมื่อแบบสอบมีความแตกต่างกันน้อย กลุ่มตัวอย่างมีขนาดใหญ่ และแบบสอบยาว ความสัมพันธ์ระหว่างมิติมีผลต่อความคลาดเคลื่อนในการปรับเทียบเพียงเล็กน้อย กระบวนการปรับเทียบทั้ง 4 วิธี ได้แก่ (1) UIRTO (2) UIRTT (3) AOSE (4) ATSE มีความลำเอียงใกล้เคียง ส่วนกระบวนการ MOSE และวิธี EE สำหรับความสัมพันธ์ระหว่างมิติต่ำมีความลำเอียงต่ำกว่าความสัมพันธ์ระหว่างมิติสูง เนื่องจากการแจกแจง Marginal distribution ของแบบสอบ X มีความแตกต่างจากโค้งปกติมากที่สุด เมื่อความสัมพันธ์ระหว่างมิติสูง ผู้วิจัยระบุข้อจำกัดและข้อเสนอแนะไว้ว่า (1) ควรศึกษาในปัจจัยอื่น ๆ เพิ่มเติม (2) ควรศึกษากับโครงสร้างมิติความสามารถอื่น เช่น โครงสร้างที่ซับซ้อน และอาจศึกษากับจำนวนมิติความสามารถที่มากขึ้น (2) ค่าลำเอียงจากวิธีการปรับเทียบคะแนนในเงื่อนไขต่าง ๆ ค่อนข้างต่ำ จึงควรศึกษากับแบบสอบที่มีหลากหลาย เช่น แบบสอบ X มีค่าเฉลี่ยของ MDIFF มากกว่าแบบสอบ Y แต่มีค่า MDISC เท่ากัน (3) งานวิจัยนี้ใช้ชุดค่าพารามิเตอร์ข้อสอบเพียงชุดเดียวในการทำซ้ำทุกครั้งสำหรับกระบวนการปรับเทียบทุกแบบจึงอาจเลือกใช้ค่าพารามิเตอร์ที่แตกต่างออกสำหรับแบบสอบแต่ละฉบับเพื่อศึกษาผลจากรูปแบบของแบบสอบ

Peterson (2014) เปรียบเทียบกระบวนการ MOSE กับวิธีอีควิเปอร์เซ็นไทล์ กระบวนการปรับเทียบด้วยโมเดล UIRT และวิธี Bi-factor equating สำหรับแบบสอบรูปแบบผสมซึ่งวัดคุณลักษณะ 2-4 มิติ ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม และมีการประมาณความสัมพันธ์ระหว่างมิติอย่างอิสระและ

กำหนดให้ความสัมพันธ์เป็นศูนย์ การวิจัยครั้งนี้ใช้ข้อมูลจริงจากการทดสอบขนาดใหญ่และแบบสอบ มีความเป็นพหุมิติต่ำ (ความสัมพันธ์ระหว่างมิติความสามารถสูง)

ผลของการวิจัยพบว่า เมื่อใช้ข้อสอบแบบพหุมิติ ความลำเอียง ความคลาดเคลื่อนมาตรฐาน และ ความคลาดเคลื่อนโดยรวมของกระบวนการ MOSE มีค่าน้อยกว่ากระบวนการอื่น ๆ เมื่อพิจารณาการ จำแนกความสามารถของผู้สอบ พบว่า ผลการจำแนกระหว่างกระบวนการ UIRT, วิธี Bifactor (Bifactor observed-score equating procedure) และกระบวนการ MOSE ไม่มีความแตกต่างกัน ผู้วิจัยระบุข้อจำกัด และข้อเสนอแนะไว้ว่า (1) งานวิจัยนี้ใช้ข้อมูลจากการทดสอบจริง มีความเป็นพหุมิติต่ำ ควรศึกษาเกี่ยวกับ ข้อมูลที่มีความเป็นพหุมิติสูงและมีความหลากหลาย อย่างเช่น ความเป็นพหุมิติต่ำไปจนถึงสูงด้วยการระบุ ความสัมพันธ์ระหว่างมิติเป็น 0.99 ถึง 0.50 เป็นต้น ด้วยการจำลองข้อมูล (2) ควรศึกษาตัวแปรอื่น ๆ เพิ่มเติม (3) ควรใช้วิธีการปรับเทียบอื่น ๆ เป็นเกณฑ์ นอกเหนือจากวิธีอีควิเปอร์เซ็นไทล์แบบดั้งเดิมและ วิธี Identity (3) รูปแบบของค่าอำนาจจำแนกของข้อสอบไม่ได้สอดคล้องกับแบบสอบ

### ประเด็นที่สรุปได้จากงานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT

#### 1) โมเดลการตอบสนองข้อสอบแบบพหุมิติ

โมเดลการตอบสนองข้อสอบแบบพหุมิติที่นิยมศึกษา คือ โมเดล MIRT แบบทดแทนได้ (compensatory MIRT model) โดยการปรับเทียบคะแนนตามโมเดล MIRT สำหรับข้อสอบที่ตรวจให้ คะแนนสองค่านิยมใช้โมเดลโลจิสติกแบบ 2 คือ M2PL ตามลำดับ ส่วนโมเดล MIRT สำหรับแบบสอบ รูปแบบผสมใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (M3PL) และโมเดล Multidimensional grade-response model (MGR) สำหรับข้อสอบตรวจให้คะแนนสองค่าและมากกว่าสองค่าตามลำดับ ส่วนการปรับเทียบ คะแนนภายใต้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติที่มีโครงสร้างอย่างง่าย (SS-MIRT) สำหรับแบบสอบ รูปแบบผสมใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PL) และโมเดล Grade-response model (GRM)

#### 2) รูปแบบการประมาณค่าพารามิเตอร์และโปรแกรมประมาณค่าพารามิเตอร์

รูปแบบการประมาณค่าพารามิเตอร์ของงานวิจัยทั้งหมด (ยกเว้นงานวิจัยของ Brossman (2010) ที่ไม่ได้ระบุไว้) ใช้การประมาณค่าพารามิเตอร์แบบแยกกัน สำหรับรูปแบบผู้สอบกลุ่มสุ่ม ผู้สอบกลุ่มเท่า เทียมกัน และผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม อันเป็นข้อสังเกตสำคัญของงานวิจัยเหล่านี้

โปรแกรมประมาณค่าพารามิเตอร์ที่นิยมใช้สำหรับข้อสอบที่ตรวจให้คะแนนสองค่ามากที่สุด คือ TESTFACT โปรแกรมนี้ไม่สามารถใช้โมเดลโลจิสติก 3 พารามิเตอร์เพราะไม่สามารถประมาณค่าพารามิเตอร์ โอกาสในการเดาได้ ส่วนโปรแกรมที่นิยมใช้สำหรับแบบสอบรูปแบบผสม คือ โปรแกรม flexMIRT

### 3) รูปแบบการออกแบบการเก็บรวบรวมข้อมูลและการเชื่อมโยงสเกล

รูปแบบการออกแบบการเก็บรวบรวมข้อมูลที่นิยมใช้ คือรูปแบบผู้สอบกลุ่มสุ่ม รองลงมาคือรูปแบบผู้สอบกลุ่มเท่าเทียมและรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม โดยรูปแบบ NEAT ใช้กับข้อสอบที่ตรวจให้คะแนนสองค่าเท่านั้น ยังไม่มีการศึกษากับแบบสอบรูปแบบผสม

### 4) ความยาวของแบบสอบ

แบบสอบรูปแบบผสมของงานวิจัยทั้ง 2 งาน มาจากการทดสอบจริง (advance place exam: AP) และส่วนใหญ่ใช้สัดส่วนคะแนนข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าเป็น 70:30 แบบสอบวิชาประวัติศาสตร์งานวิจัยของ Lee & Brossman (2012) ใช้สัดส่วนคะแนน 70:30 เช่นเดียวกับแบบสอบวิชาภาษาอังกฤษในงานวิจัยของ Peterson (2014) ที่ใช้สัดส่วนคะแนน 70:30 ส่วนแบบสอบวิชาเคมี และวิชาภาษาสเปน ใช้สัดส่วนคะแนนเป็น 60:40 และ 80:20 โดยการปรับเทียบคะแนนภายใต้รูปแบบ NEAT ใช้สัดส่วนคะแนนข้อสอบร่วมภายใน คือ 50%

### 5) โครงสร้างมิติความสามารถ และความสัมพันธ์ระหว่างมิติ

โครงสร้างมิติความสามารถที่นิยมใช้ คือ โครงสร้างมิติที่ซับซ้อน (CS) และโครงสร้างมิติที่ไม่ซับซ้อน (APSS) เมื่อพิจารณาแบบสอบจากการทดสอบจริง พบว่า แบบสอบส่วนใหญ่มีความเป็นพหุมิติต่ำจนถึงปานกลาง กล่าวคือ แบบสอบมีความสัมพันธ์ระหว่างมิติในระดับสูงจนถึงปานกลาง อย่างเช่นงานวิจัยของ Peterson ที่ใช้แบบสอบทั้ง 3 วิชามีความสัมพันธ์ระหว่างมิติตั้งแต่ 0.8-0.9 และงานวิจัยของ Brossman (2010) ที่ระบุว่า แบบสอบจากการทดสอบจริงมีความเป็นพหุมิติต่ำถึงปานกลาง ขณะที่งานวิจัยที่ใช้การจำลองข้อมูลมีการกำหนดระดับความสัมพันธ์ระหว่างมิติต่ำและปานกลาง ( $r = 0$  และ  $0.5$ ) และความสัมพันธ์ระหว่างมิติสูงและปานกลาง ( $r = 0.8$  และ  $0.5$ ) โดยงานวิจัยที่ศึกษาตัวแปรความสัมพันธ์ระหว่างมิติ พบว่า การปรับเทียบคะแนนด้วยโมเดล MIRT มีความคลาดเคลื่อนมากขึ้นเมื่อความสัมพันธ์ระหว่างมิติเพิ่มขึ้น (ความเป็นพหุมิติต่ำ)

### 6) ขนาดของกลุ่มตัวอย่าง

ขนาดของกลุ่มตัวอย่างที่ใช้ในการวิจัยส่วนใหญ่มากกว่า 2,000 คน โดยกลุ่มตัวอย่างที่น้อยที่สุดสำหรับแบบสอบแต่ละฉบับ คือ จำนวน 2,000 คน

### 7) วิธีการปรับเทียบคะแนนในขั้นตอนของการปรับเทียบคะแนนด้วยโมเดล MIRT และวิธีการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์

วิธีการปรับเทียบคะแนนในขั้นตอนของการปรับเทียบคะแนนด้วยโมเดล MIRT ของงานวิจัยทั้งหมดใช้วิธีอีควิเปอร์เซ็นไทล์แบบดั้งเดิมซึ่งไม่มีการปรับโค้งให้ราบเรียบ (unsmoothed equipercentile method) ส่วนวิธีการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์ที่นิยมมากที่สุด คือ วิธีอีควิเปอร์เซ็นไทล์แบบดั้งเดิมที่มีการปรับโค้งให้ราบเรียบก่อน (pre-smoothed equipercentile method) รองลงมาคือ วิธี Identity equating และ วิธีการประมาณความถี่ (frequency estimation)



## 8) การประเมินประสิทธิภาพของการปรับเทียบคะแนน

ดัชนีสำหรับการประเมินประสิทธิภาพของการปรับเทียบคะแนนที่นิยมใช้ คือ ค่า Difference That Matters (DTM) มักใช้ร่วมกับค่าดัชนีตัวอื่น ๆ เพื่อเป็นเกณฑ์ในการเปรียบเทียบว่าวิธีการปรับเทียบคะแนนใดมีประสิทธิภาพมากที่สุด รองลงมา คือ ค่าความลำเอียง (bias) ค่าความคลาดเคลื่อนมาตรฐาน (SE) รากที่สองของความแตกต่างยกกำลังสอง (RMSD) รากที่สองของความคลาดเคลื่อนยกกำลังสอง (RMSE) และ ความคลาดเคลื่อนยกกำลังสองเฉลี่ย (MSE) ความแปรปรวน (Variance)

## 9) ตัวแปรที่ศึกษา

งานวิจัยที่ผ่านมามักเป็นการเปรียบเทียบประสิทธิภาพระหว่างกระบวนการปรับเทียบแบบต่าง ๆ ร่วมกับตัวแปรอื่น แบ่งเป็น ด้านแบบสอบ: โครงสร้างของแบบสอบ ความสัมพันธ์ระหว่างมิติ ความยาวของแบบสอบ ความแตกต่างของความยาก และด้านกลุ่มตัวอย่าง: ความแตกต่างระหว่างกลุ่มผู้สอบ

## 10) ผลการวิจัยโดยสรุป

งานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT พบว่า กระบวนการ FMIRT ทั้ง 3 วิธี ได้แก่ MOSE, AOSE และ ATSE มีประสิทธิภาพมากกว่าการปรับเทียบคะแนนด้วยโมเดล UIRT และเมื่อเปรียบเทียบประสิทธิภาพของกระบวนการ FMIRT ทั้ง 3 วิธี พบว่า การปรับเทียบคะแนนด้วยกระบวนการ AOSE มีประสิทธิภาพมากที่สุด สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT แต่ก็อย่างไรก็ตามการปรับเทียบคะแนนสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม พบว่า กระบวนการ MOSE มีประสิทธิภาพมากที่สุด นอกจากนี้การปรับเทียบคะแนนรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม เลือกใช้กระบวนการ MOSE เพื่อเปรียบเทียบประสิทธิภาพกับการปรับเทียบคะแนนด้วยโมเดล UIRT และวิธี Bi-factor ซึ่งพบว่า กระบวนการ MOSE มีประสิทธิภาพมากที่สุด สะท้อนให้เห็นกระบวนการ MOSE มีความเหมาะสมที่จะนำมาปรับเทียบคะแนนแบบสอบพหุมิติ

## 11) ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

งานวิจัยที่ใช้การจำลองข้อมูลส่วนใหญ่ ระบุว่า ควรศึกษาในตัวแปรอื่น ๆ เพิ่มเติม เมื่อพิจารณาตัวแปรที่ศึกษาในงานวิจัยที่ใช้ข้อมูลจริง 2 งาน และการจำลองข้อมูล 3 งาน พบว่า ตัวแปรที่ยังไม่ได้มีการศึกษา คือ ตัวแปรด้านข้อสอบรวม อย่างเช่น สัดส่วนของข้อสอบรวม และกรณีที่ใช้แบบสอบรูปแบบผสมควรศึกษาสัดส่วนของข้อสอบรวมแต่ละประเภทด้วย

งานวิจัยที่ใช้ข้อมูลจริงจากการทดสอบส่วนใหญ่ ระบุว่า ข้อมูลมีความเป็นพหุมิติต่ำถึงปานกลาง นั่นคือ ความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับปานกลางถึงสูง จึงอาจศึกษากับข้อมูลที่มีความเป็นพหุมิตีปานกลางถึงสูง และศึกษาความสัมพันธ์ระหว่างมิติในหลายระดับด้วยการจำลองข้อมูล

ตาราง 2.8 สรุปประเด็นของงานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT

ประเด็น	Broosman (2010)	Lee & Broosman (2012)	Zhang (2012)	Lee (2013)	Peterson (2014)
<b>MIRT model</b>					
M2PL/M3PL	M2PL		M2PL	M2PL	
M3PL+MGR					/
3PL+GRM		/			
<b>โปรแกรมประมาณค่า</b>	งานวิจัยทั้งหมดประมาณค่าพารามิเตอร์แบบแยกกัน (ยกเว้น Broosman ที่ไม่ได้ระบุ)				
TESTFACT	/	N/A	/		
flexMIRT				/	/
<b>รูปแบบ</b>					
NEAT			/		
ผู้สอบกลุ่มเท่าเทียม	/				
ผู้สอบกลุ่มสุ่ม		/		/	/
<b>โครงสร้างมิติ</b>					
SS	N/A	/			
APSS			2 มิติ	2 มิติ	
CS			2 มิติ		2, 4 มิติ
<b>ความสัมพันธ์ระหว่างมิติ</b>	ต่ำ	N/A	0, 0.5	0.8, 0.5	0.8-0.9
<b>ขนาดกลุ่มตัวอย่าง</b>	2,500	3,000	2,000	2,000, 6,000	>100,000
<b>วิธีที่ใช้เป็นเกณฑ์</b>					
Unsmoothed EE					
Presmoothed EE	/			/	/
Post-smoothed EE					
Frequency estimation			/		
Identity				/	
<b>ดัชนี</b>					
SEE			/	/	
RMSE				/	
RMSD			/		/
DTM	/	/		/	/
BIAS		/	/	/	
MSE		/			

ตาราง 2.8 (ต่อ)

ประเด็น	Broosman (2010)	Lee & Broosman (2012)	Zhang (2012)	Lee (2013)	Peterson (2014)
ความยาวแบบสอบ(คะแนน)	40, 50, 48	MC:CR 70:27	40	30,60	MC:CR 75:52 70:20 54:27
ความยาวแบบสอบรวม (คะแนน)	N/A	N/A	20 (50%)	N/A	N/A
<b>ด้านแบบสอบ</b>					
ความสัมพันธ์ระหว่างมิติ		/		/	/
โครงสร้างแบบสอบ			/		/
ความยาวของแบบสอบ				/	
ความแตกต่างของความยาก				/	
<b>ด้านผู้สอบ</b>					
ความแตกต่างของความสามารถ			/		/
ขนาดของกลุ่มตัวอย่าง				/	

### ตอนที่ 3 การเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

#### 3.1 มโนทัศน์เกี่ยวกับการเชื่อมโยงสเกล

##### 3.1.1) ความหมายของการเชื่อมโยง

การปรับเทียบคะแนนเป็นการปรับความแตกต่างของความยากระหว่างแบบสอบที่มีเนื้อหาเหมือนกัน แต่ในบางสถานการณ์ไม่สามารถทำการปรับเทียบคะแนน เช่น กรณีที่แบบสอบมีเนื้อหาและ/หรือคุณลักษณะของค่าสถิติแตกต่างกัน รวมไปถึงการเชื่อมโยงการทดสอบแบบเขียนตอบและการทดสอบโดยใช้คอมพิวเตอร์เป็นฐาน การเชื่อมโยงคลังข้อสอบสำหรับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ การเชื่อมโยงข้อคำถามของที่อยู่ในทดสอบการปฏิบัติ (performance assessment) การสร้างความสัมพันธ์ระหว่างคะแนนของแบบสอบเรียกว่า การเชื่อมโยง (linking) หรืออาจกล่าวได้ว่าการเชื่อมโยงเป็นกระบวนการนำคะแนนจากแบบสอบสองฉบับหรือมากกว่าวางบนสเกลเดียวกัน ตัวอย่างอย่างง่ายของการเชื่อมโยง คือ ความสัมพันธ์ระหว่างองศาเซลเซียส (C) และองศาฟาเรนไฮต์ (F) นั่นคือ  $F = (9/5)C + 32$  หรือ  $C = (5/9)(F - 32)$  สมการทั้งสองนี้เป็นรูปแบบหนึ่งของการเชื่อมโยงการวัดคุณสมบัตินี้ ซึ่งยอมให้วงอุณหภูมิองศาเซลเซียสบนสเกลขององศาฟาเรนไฮต์

##### 3.1.2) ประเภทของการเชื่อมโยงโดย Misley/Linn (Misley/Linn taxonomy)

Misley และ Linn ได้ระบุประเภทของการเชื่อมโยงไว้ 4 ประเภท ได้แก่

###### 1. Equating

การปรับเทียบคะแนนเป็นการสร้างความสัมพันธ์ของคะแนนระหว่างแบบสอบต่างฉบับกัน แต่แบบสอบเหล่านั้นมีเนื้อหาและคุณลักษณะทางสถิติเหมือนกัน กล่าวคือ แบบสอบทั้งสองฉบับมีกรอบแนวคิดและแผนผังแบบสอบเหมือนกัน

###### 2. Calibration

Calibration เป็นการเชื่อมโยงสำหรับแบบสอบที่มีกรอบแนวคิดและแผนผังแบบสอบต่างกัน ซึ่งมีหลายลักษณะ ได้แก่

- Calibration เป็นความสัมพันธ์ระหว่างแบบสอบที่มีแผนผังเนื้อหาเหมือนกัน แต่คุณลักษณะทางสถิติต่างกัน ตัวอย่างที่พบบ่อย คือ แบบสอบที่มีความยาวต่างกัน

- Calibration เป็นความสัมพันธ์ระหว่างแบบสอบที่มีแผนผังเนื้อหา หรือมีคุณลักษณะทางสถิติต่างกัน ตัวอย่างเช่น แบบสอบที่สร้างสำหรับผู้สอบระดับต่างๆ และต้องนำคะแนนของแบบสอบเหล่านั้นมาวางบนสเกลเดียวกัน ซึ่งเรียกว่า Vertical scaling

- Calibration ที่เป็นการใช้ระเบียบวิธี (ที่พบบ่อยคือ โมเดล IRT) เพื่อนำข้อสอบทั้งหมดในโดเมนต่างๆมาวางบนสเกลร่วม และถ้าโมเดลมีความสอดคล้องตามข้อตกลงเบื้องต้น คะแนนของความสามารถของผู้สอบอาจนำมาเปรียบเทียบกันได้

### 3. Projection

ความแตกต่างระหว่าง Projection กับ Equating หรือ Calibration คือ (1) Projection มีทิศทางเดียว (2) ใช้รูปแบบผู้สอบกลุ่มเดียว (3) ไม่จำเป็นต้องโครงสร้างเหมือนกัน โดยความสัมพันธ์ของ Projection เป็นสมการถดถอย

### 4. Moderation

Moderation อาจเรียกว่า “การจับคู่การแจกแจง (distribution matching)” ซึ่งใช้ได้กับรูปแบบผู้สอบกลุ่มเดียว ผู้สอบกลุ่มสุ่ม และผู้สอบกลุ่มไม่เท่าเทียมกัน ตัวอย่างความสัมพันธ์แบบสอดคล้องกัน (concordance relationship) เช่น ผู้สอบกลุ่มเดียวกันทำแบบสอบคนละฉบับ เช่น การเชื่อมโยงแบบสอบ SAT และ ACT บางครั้งเรียกว่า การสเกลเพื่อเปรียบเทียบผลสัมฤทธิ์ ซึ่งเป็นประเภทหนึ่งของการเชื่อมโยงที่มีกรอบแนวคิดต่างกันแต่มีโครงสร้างทั่วไปคล้ายคลึงกัน

### 3.2 แนวทางการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบหรือโมเดล IRT นิยมใช้อย่างกว้างขวางเพราะโมเดลดังกล่าวมีประสิทธิภาพมากกว่าวิธีการเชื่อมโยงสเกลที่ไม่ใช้โมเดล IRT คุณลักษณะความไม่แปรเปลี่ยนของโมเดล IRT ทำให้การเชื่อมโยงสเกลและการประมาณค่าพารามิเตอร์มีความยืดหยุ่น นอกจากนี้ความคลาดเคลื่อนในโมเดล IRT เป็นความคลาดเคลื่อนอย่างสุ่ม อย่างไรก็ตามค่าพารามิเตอร์ในโมเดล IRT มีคุณสมบัติความไม่แปรเปลี่ยน อาจทำให้เกิด Scale indeterminacy ดังที่ได้กล่าวไว้ในแนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

แนวทางการเชื่อมโยงสเกลด้วยโมเดลทฤษฎีการตอบสนองข้อสอบแบบเอกมิติและพหุมิติมี 3 รูปแบบหลัก ได้แก่ การประมาณค่าพารามิเตอร์พร้อมกัน (concurrent calibration) การประมาณค่าพารามิเตอร์แบบกำหนดค่าพารามิเตอร์ของข้อสอบร่วม (fixed common item parameter calibration: FCIP) และการเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แบบแยกกัน (scale linking after separate calibration)

#### 3.2.1) การประมาณค่าพารามิเตอร์พร้อมกัน (concurrent calibration)

การประมาณค่าพารามิเตอร์พร้อมกัน คือ ข้อมูลการตอบข้อสอบของผู้สอบทุกคนจากการทดสอบที่แตกต่างกันถูกนำมารวมกันและดำเนินการประมาณค่าพารามิเตอร์ข้อสอบในครั้งเดียว (Wingersky & Lord, 1984 cited in Zhang, 2012) ดังนั้นข้อมูลจากการประมาณค่าพารามิเตอร์พร้อมกัน คือ ค่าพารามิเตอร์ข้อสอบและค่าความสามารถย่ออยู่ในเมตริกซ์เดียวกัน จึงไม่จำเป็นต้องใช้การแปลงสเกลหรือการเชื่อมโยงสเกล

3.2.2) การประมาณค่าพารามิเตอร์ข้อสอบแบบกำหนดข้อสอบร่วม (fixed common item parameter calibration: FCIP)

การประมาณค่าพารามิเตอร์ข้อสอบแบบกำหนดข้อสอบร่วมมักนำไปใช้กับสถานการณ์การทดสอบที่ทราบค่าพารามิเตอร์ของข้อสอบร่วมและค่าความสามารถของผู้สอบร่วม โดยค่าพารามิเตอร์จากการทดสอบครั้งต่าง ๆ จะถูกนำมาวางบนเมตริกซ์เดียวกันซึ่งก็คือเมตริกซ์สำหรับข้อมูลการทดสอบครั้งที่ผ่านมา ดังนั้นเมื่อใช้วิธี FCIP จึงไม่จำเป็นต้องใช้การแปลงสเกลหรือการเชื่อมโยงสเกล เพราะค่าพารามิเตอร์ของข้อสอบและค่าความสามารถจากการทดสอบที่แตกต่างอยู่บนเมตริกซ์เดียวกันสำหรับข้อมูลการทดสอบก่อนหน้านี้

### 3.2.3) การเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แล้ว (scale linking after separate calibration)

การเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แล้วใช้เมื่อข้อมูลการตอบข้อสอบมาจากแบบสอบหรือการทดสอบที่แตกต่างกัน และค่าพารามิเตอร์ข้อสอบถูกประมาณแบบแยกกัน จึงจำเป็นต้องใช้การแปลงสเกลหรือการเชื่อมโยงสเกล เพื่อให้ค่าพารามิเตอร์จากแบบสอบครั้งหรือฉบับต่าง ๆ ถูกแปลงให้อยู่ในเมตริกซ์เดียวกัน โดยวิธีการเชื่อมโยงสเกลหลังจากการประมาณค่าพารามิเตอร์แล้วแบ่งออกเป็น 2 แบบ คือ แบบเอกมิติ และแบบพหุมิติ

## 3.3 งานวิจัยเกี่ยวกับแนวทางการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบ

งานวิจัยที่เปรียบเทียบการประมาณค่าพารามิเตอร์แบบพร้อมกัน และการเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แบบแยกกันแล้ว (ได้แก่ วิธีค่าเฉลี่ย/ค่าเฉลี่ย วิธีค่าเฉลี่ย/ซิกมา วิธีของ Heabara และวิธีของ Stocking-Lord) สำหรับแบบสอบรูปแบบผสมและข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT พบว่า วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันมีประสิทธิภาพมากกว่า (มีความคลาดเคลื่อนน้อยกว่า) วิธีการเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แบบแยกกัน อีกทั้งยังมีความแกร่งเมื่อใช้กับแบบสอบต้องพิจารณาผลกระทบของรูปแบบ สรุปได้ดังตาราง 2.9 และรายละเอียดมีดังนี้

### 3.3.1) การเชื่อมโยงสเกลด้วยโมเดล UIRT สำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า

อัญชลี ศรีกลชาญ (2552) เปรียบเทียบวิธีการเชื่อมโยงสเกลด้วยโมเดล IRT 2 วิธี ได้แก่ วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันและวิธีโค้งคุณลักษณะ สำหรับแบบสอบรูปแบบผสม การวิจัยครั้งนี้ใช้การจำลองข้อมูล ปัจจัยที่ศึกษามี 3 ปัจจัย ได้แก่ (1) รูปแบบการเก็บรวบรวมข้อมูล ได้แก่ รูปแบบ NEAT และรูปแบบผู้สอบกลุ่มเท่าเทียม (2) สัดส่วนคะแนนของข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าในแบบสอบมี 4 สัดส่วน ได้แก่ 50:50, 60:40, 70:30 และ 80:20 (3) สัดส่วนคะแนนของข้อสอบตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าในแบบสอบร่วม ได้แก่ 3:1, 1:1 และ 1:3

ผลการวิจัย พบว่า เมื่อพิจารณาแบบผู้สอบกลุ่มเท่าเทียมกัน พบว่า วิธีโค้งคุณลักษณะและวิธีประมาณค่าพารามิเตอร์แบบพร้อมกันมีความคลาดเคลื่อนต่ำและให้ผลการเชื่อมโยงสเกลใกล้เคียงกัน ส่วนรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกัน วิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันมีความ

**คลาดเคลื่อนต่ำกว่าวิธีโค้งคุณลักษณะ** เมื่อพิจารณาสัดส่วนของข้อสอบทั้งสองชนิดในแบบสอบและแบบสอบรวม พบว่า เมื่อสัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่าในแบบสอบเพิ่มขึ้นและสัดส่วนข้อสอบรวมที่ตรวจให้คะแนนสองค่าในแบบสอบรวมลดลง ประสิทธิภาพของวิธีโค้งคุณลักษณะและวิธีประมาณค่าพารามิเตอร์แบบพร้อมกันมีความคลาดเคลื่อนต่ำลง เมื่อพิจารณาค่าเฉลี่ย MSE สำหรับสัดส่วนของข้อสอบรวมทั้งสองชนิดในแบบสอบที่แตกต่างกัน 4 สัดส่วน พบว่า การเชื่อมโยงสเกลสำหรับแบบสอบทั้ง 4 สัดส่วนและแบบสอบรวมทั้ง 6 สัดส่วน มีค่าเฉลี่ย MSE ไม่แตกต่างกัน และเมื่อพิจารณาค่าเฉลี่ย MSE สำหรับรูปแบบการเก็บรวบรวมข้อมูลที่แตกต่างกัน พบว่า การเชื่อมโยงสเกลสำหรับรูปแบบการเก็บรวบรวมข้อมูลแบบต่างๆ มีค่าเฉลี่ย MSE แตกต่างกัน

Kim (2004 อ้างอิงใน อัญชลี ศรีกรลชาญ, 2552) เปรียบเทียบประสิทธิภาพและความแกร่งระหว่างวิธีประมาณค่าพารามิเตอร์แบบพร้อมกัน และวิธีการเชื่อมโยงสเกลด้วยโมเดล IRT หลังจากประมาณค่าพารามิเตอร์แบบแยกกัน 4 วิธี ได้แก่ วิธีค่าเฉลี่ย/ค่าเฉลี่ย วิธีค่าเฉลี่ย/ซิกมา วิธีของ Heabara และวิธีของ Stocking-Lord สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT

ผลการวิจัย พบว่า **วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันให้ผลการปรับเทียบคะแนนดีกว่า** วิธีประมาณค่าพารามิเตอร์แบบแยกกัน อีกทั้งยังมีความแกร่งสำหรับแบบสอบที่ต้องพิจารณาผลกระทบของรูปแบบ แต่เมื่อพิจารณาเฉพาะวิธีประมาณค่าพารามิเตอร์แบบแยกกัน พบว่า วิธีของ Heabara และวิธีของ Stocking-Lord ซึ่งจัดว่าเป็นวิธีโค้งคุณลักษณะมีความแกร่งและให้ผลการเชื่อมโยงสเกลที่ดีกว่าวิธีค่าเฉลี่ย/ค่าเฉลี่ย และวิธีค่าเฉลี่ย/ซิกมา ซึ่งจัดว่าเป็นวิธีโมเมนต์

Kim & Kolen (2006) เปรียบเทียบความแกร่งของระหว่างวิธีประมาณค่าพารามิเตอร์แบบพร้อมกันกับวิธีการเชื่อมโยงสเกลด้วยโมเดล IRT หลังการประมาณค่าพารามิเตอร์แบบแยกกัน 4 วิธี ได้แก่ วิธีค่าเฉลี่ย/ค่าเฉลี่ย วิธีค่าเฉลี่ย/ซิกมา วิธีของ Heabara และวิธีของ Stocking-Lord ภายใต้รูปแบบ NEAT แบบสอบฉบับใหม่และฉบับเก่าเป็นแบบสอบรูปแบบผสม ประกอบด้วยข้อสอบแบบหลายตัวเลือก 36 ข้อ และข้อสอบแบบสร้างคำตอบ 15 ข้อ แบบสอบทั้งสองฉบับประกอบด้วยข้อสอบรวมแบบหลายตัวเลือก 12 ข้อ และข้อสอบรวมแบบสร้างคำตอบ 5 ข้อ ปัจจัยที่ศึกษามี 3 ด้าน ได้แก่ (1) ระดับของผลกระทบรูปแบบ (format effects) ซึ่งก็คือ ระดับความสัมพันธ์ระหว่างความสามารถมิติที่ 1, 2 และ 3 ระดับ ได้แก่ ความสัมพันธ์ระดับสูง ปานกลาง และไม่มีความสัมพันธ์ (2) ประเภทของแบบสอบรูปแบบผสม 2 แบบ ได้แก่ ช่วงกว้างและช่วงแคบ และ (3) ระดับความไม่เท่าเทียมระหว่างผู้สอบทั้งสองกลุ่ม 3 ระดับ ได้แก่ สูง ปานกลาง และต่ำ

ผลการวิจัย พบว่า **วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันมีประสิทธิภาพมากกว่า** วิธีการเชื่อมโยงสเกลทั้ง 4 วิธี และมีความแกร่งต่อผลกระทบรูปแบบแต่ละระดับ (ไม่มี ปานกลาง และมาก) และเมื่อพิจารณารณที่ที่ไม่มีผลกระทบรูปแบบ พบว่า วิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันมีความแม่นยำมากกว่าวิธีประมาณค่าพารามิเตอร์แบบแยกกัน แต่อย่างไรก็ตามวิธีประมาณค่าพารามิเตอร์พร้อม

กันมีประสิทธิภาพดีกว่าวิธีโค้งคุณลักษณะเพียงเล็กน้อย โดยค่าความแกร่งของวิธีเชื่อมโยงสเกลทั้ง 2 วิธี ไม่ได้ความคงที่สำหรับข้อสอบทุกประเภท นอกจากนี้ยังพบว่า การประมาณค่าพารามิเตอร์แบบแยกกัน ทำให้ได้ค่าพารามิเตอร์ของผู้สอบแต่ละกลุ่มสำหรับข้อสอบร่วมแต่ละฉบับ จึงสามารถนำผลที่ได้ไปใช้ในการวินิจฉัยได้ ขณะที่วิธีประมาณค่าพารามิเตอร์แบบพร้อมกัน ค่าพารามิเตอร์ของข้อสอบร่วมจากแบบสอบทั้งสองฉบับจะรวมเป็นชุดเดียวกัน

Kim, Choi, Lee และ Um (2008) เปรียบเทียบประสิทธิภาพของการปรับเทียบคะแนนภายใต้รูปแบบผู้สอบกลุ่มสุ่มและรูปแบบ NEAT โดยใช้ข้อมูลจริงจากการทดสอบ สำหรับแบบสอบรูปแบบผสม วิธีการปรับเทียบคะแนนด้วยโมเดล IRT ทั้ง 4 วิธี ประกอบด้วย วิธีการปรับเทียบคะแนนสำหรับรูปแบบ NEAT 2 วิธี (ข้อสอบร่วมมีเฉพาะข้อสอบแบบหลายตัวเลือก) ได้แก่ วิธีประมาณค่าพารามิเตอร์แบบพร้อมกัน และวิธีประมาณค่าพารามิเตอร์พร้อมกันด้วยกลุ่มผู้สอบหลายกลุ่ม และวิธีการปรับเทียบคะแนนสำหรับรูปแบบผู้สอบกลุ่มสุ่ม 2 วิธี ได้แก่ วิธีประมาณค่าพารามิเตอร์แบบแยกกันภายใต้รูปแบบผู้สอบกลุ่มสุ่ม และวิธีประมาณค่าพารามิเตอร์แบบพร้อมกันภายใต้รูปแบบผู้สอบกลุ่มสุ่ม

ผลการวิจัย พบว่า วิธีการปรับเทียบคะแนนทั้ง 4 วิธีให้ผลการปรับเทียบคะแนนที่ดีและใกล้เคียงกันภายใต้รูปแบบผู้สอบกลุ่มสุ่ม ขนาดของกลุ่มตัวอย่างมีผลต่อวิธีการปรับเทียบคะแนนที่แตกต่างกัน และมีผลต่อวิธีการปรับเทียบคะแนนสำหรับรูปแบบผู้สอบกลุ่มสุ่มและรูปแบบ NEAT แตกต่างกันในทางปฏิบัติรูปแบบ NEAT ให้ผลการปรับเทียบคะแนนไม่ถูกต้องมากขึ้นเมื่อข้อสอบร่วมไม่มีคุณภาพหรือข้อสอบร่วมมีความยาวสั้น

Tian (2011) เปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลด้วยโมเดล IRT 2 วิธี ได้แก่ วิธีประมาณค่าพารามิเตอร์แบบแยกกัน (วิธีของ Stocking-Lord) และวิธีประมาณค่าพารามิเตอร์แบบพร้อมกัน สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT งานวิจัยครั้งนี้ใช้การจำลองข้อมูล ประสิทธิภาพของวิธีการเชื่อมโยงสเกลพิจารณาจากประสิทธิภาพในการกู้คืนค่าพารามิเตอร์ของข้อสอบ (ค่า MSE และส่วนเบี่ยงเบนมาตรฐาน) และความแม่นยำในการประมาณคะแนน (ค่า RMSD) ปัจจัยที่ศึกษามี 2 ปัจจัย คือ จำนวนข้อสอบร่วม ได้แก่ 5, 10 และ 15 ข้อ และประเภทของข้อสอบร่วม 3 แบบ ได้แก่ ข้อสอบแบบหลายตัวเลือกอย่างเดียว ข้อสอบแบบสร้างคำตอบอย่างเดียว และรูปแบบผสม

ผลการวิจัย พบว่า วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันมีประสิทธิภาพในการกู้คืนค่าพารามิเตอร์มากกว่าวิธีของ Stocking-Lord เนื่องจากมีค่า MSE น้อยกว่า โดยความยาวของข้อสอบร่วมมีผลต่อประสิทธิภาพในการกู้คืนค่าพารามิเตอร์ของวิธีการเชื่อมโยงสเกลทั้ง 2 วิธี ข้อสอบร่วมที่มีความยาวเพิ่มขึ้นจะมีค่า MSE ลดลง และประเภทของข้อสอบร่วมมีผลต่อประสิทธิภาพในการกู้คืนค่าพารามิเตอร์ของวิธีการเชื่อมโยงทั้ง 2 วิธี โดยข้อสอบร่วมแบบสร้างคำตอบอย่างเดียวมีค่า MSE น้อยที่สุด และค่า MSE มีค่าลดลงเมื่อจำนวนข้อสอบร่วมแบบสร้างคำตอบมากขึ้น เมื่อพิจารณาความแม่นยำในการประมาณคะแนน พบว่า วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันมีความแม่นยำในการประมาณ



คะแนนมากกว่าวิธีประมาณค่าพารามิเตอร์แบบแยกกัน กล่าวคือ เมื่อจำนวนและประเภทของข้อสอบร่วมเท่ากัน วิธีประมาณค่าพารามิเตอร์พร้อมกันมีค่า RMSD น้อยกว่า แต่ความแตกต่างของค่า RMSD ระหว่างวิธีการเชื่อมโยงสเกลทั้ง 2 วิธีน้อยมาก

### 3.3.2) การเชื่อมโยงสเกลด้วยโมเดล UIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า

Keller และ Keller (2015) เปรียบเทียบประสิทธิภาพของวิธีการสเกลคะแนนด้วยโมเดล UIRT 3 วิธี ได้แก่ วิธี Fixed common item parameter (FCIP) วิธีของ Stocking-Lord และวิธีการปรับค่าพารามิเตอร์พร้อมกัน (CC) เมื่อเนื้อหาของแบบสอบมีการเปลี่ยนแปลง การวิจัยศึกษานี้ศึกษา 4 ตัวแปร ได้แก่ (1) องค์ประกอบของรูปแบบของแบบสอบ 4 แบบที่มีโดเมนเนื้อหาบางส่วนต่างกัน แต่มีความยากคู่ขนานกัน (2) องค์ประกอบของข้อสอบร่วมภายใน (3) วิธีการเชื่อมโยงสเกล (4) การแจกแจงความสามารถของผู้สอบ 3 แบบ ได้แก่ เจื่อนไขสมมติฐานซึ่งความสามารถไม่เปลี่ยนแปลง เจื่อนไขการเลื่อนของค่าเฉลี่ย และเจื่อนไขการเลื่อนของส่วนเบี่ยงเบนมาตรฐาน

ผลการวิจัย พบว่า เมื่อเปรียบเทียบประสิทธิภาพระหว่างวิธีการเชื่อมโยงสเกล **วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันให้ผลการเชื่อมโยงสเกลสเถียรมากกว่า**วิธีของ Stocking-Lord กรณีกลุ่มเท่าเทียมกันหรือมีการเลื่อนของความเบ้ พบว่า วิธี FCIP มีความแกร่งมากที่สุดเมื่อมีการเปลี่ยนแปลงเนื้อหา กรณีการเลื่อนของค่าเฉลี่ย วิธีประมาณค่าพารามิเตอร์แบบพร้อมกัน (CC) มีประสิทธิภาพดีกว่า

Hanson, McGraw-Hill และ Béguin (2002) เปรียบเทียบประสิทธิภาพระหว่างวิธีประมาณค่าพารามิเตอร์แบบพร้อมกันและวิธีเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แบบแยกกัน (4 วิธี ได้แก่ วิธีค่าเฉลี่ย/ค่าเฉลี่ย วิธีค่าเฉลี่ย/ซิกมา วิธีของ Heabara และวิธีของ Stocking-Lord) สำหรับแบบสอบจำนวน 60 ข้อภายใต้รูปแบบ NEAT ปัจจัยที่ศึกษา ได้แก่ (1) โปรแกรมประมาณค่า 2 โปรแกรม คือ MULTLOG และ BILOG-MG (2) กลุ่มตัวอย่าง 2 ระดับ คือ 3,000 และ 1,000 คน (3) ความยาวของข้อสอบร่วม คือ 20 และ 10 ข้อ และ (4) รูปแบบการเก็บรวบรวมข้อมูล ได้แก่ รูปแบบผู้สอบเท่าเทียม และรูปแบบ NEAT

ผลการวิจัย พบว่า **การประมาณค่าพารามิเตอร์แบบพร้อมกันมีความคลาดเคลื่อนต่ำกว่าการประมาณค่าพารามิเตอร์แบบแยกกัน** โดยการประมาณค่าพารามิเตอร์แบบพร้อมกันมีคลาดเคลื่อนลดลงเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ และเมื่อพิจารณาเฉพาะวิธีการเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แบบแยกกัน พบว่า วิธีโมเมนต์ทั้ง 2 วิธี ได้แก่ วิธีค่าเฉลี่ย/ค่าเฉลี่ย และวิธีค่าเฉลี่ย/ซิกมา มีความคลาดเคลื่อนมากกว่าวิธีคั้งคุณลักษณะทั้ง 2 วิธี ได้แก่ วิธีของ Heabara และวิธีของ Stocking-Lord

ตาราง 2.9 สรุปประเด็นสำคัญของงานวิจัยเกี่ยวกับประสิทธิภาพของวิธีการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิตี ภายใต้รูปแบบ NEAT

ประเด็น		อัญชิตี ศรีภักดี ชาม (2552)	Kim (2004)	Kim & Kolen (2006)	Kim & คณະ (2008)	Tian (2011)	Hanson et al., (2002)	Keller & Keller (2015)
โมเดล	3PL	/	/	/	/	/	/	/
	GPC	/	/	/	/	/		
รูปแบบ	NEAT	/	/	/	/	/	/	/
	EG				/			
วิธีการ เชื่อมโยง	วิธีปรับค่าพารามิเตอร์พร้อมกัน	/	/	/	/	/	/	/
	วิธีค่าเฉลี่ย/ค่าเฉลี่ย		/	/	/		/	
	วิธีค่าเฉลี่ย/ซิกมา		/	/	/		/	
	วิธีโค้งคุณลักษณะ	/						
	วิธีของ Heabara		/	/	/		/	
	วิธีของ Stocking-Lord		/	/	/	/	/	
	วิธี FCIP							/
ค่าสถิติ	RMSE							/
	MSE	/				/	/	
	Bias						/	/
	RMSD				/	/		
	อื่นๆ			/	/			/
ปัจจัย	สัดส่วนข้อสอบรวม					/	/	
	ประเภทของข้อสอบรวม					/		
	สัดส่วนข้อสอบรวมแต่ละประเภท	/						
	องค์ประกอบของข้อสอบรวม							
	ผลกระทบของรูปแบบ		/	/				
	สัดส่วนข้อสอบแต่ละประเภท	/						
	องค์ประกอบของแบบสอบ							/
	รูปแบบการเก็บรวบรวมข้อมูล	/			/			
	ขนาดของกลุ่มตัวอย่าง						/	
ความแตกต่างของความสามารถ			/			/	/	

### 3.4 การเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

การนำโมเดล MIRT model ไปใช้ต้องพิจารณาเกี่ยวกับระบบพิกัด (coordinate systems) ซึ่งไม่มีกฎเกณฑ์ตายตัว Interdiminancy ทั้ง 3 ส่วน สำหรับการสร้างระบบพิกัด ได้แก่ จุดกำเนิดของพื้นที่หน่วยของการวัดในแต่ละแกน และการกำหนดทิศทางแกนของระบบพิกัด (Reckase, 2009) มีข้อตกลงเบื้องต้น คือ ตำแหน่งของผู้สอบและคุณลักษณะของข้อสอบไม่แปรเปลี่ยน

การกำหนดระบบพิกัด โปรแกรมการประมาณ MIRT แต่ละโปรแกรมใช้ระบบของตัวเอง การประมาณค่าพารามิเตอร์แบบแยกกันสำหรับแบบสอบฉบับ X และ Y บนแกนพิกัดใช้โปรแกรมคอมพิวเตอร์กรณีสองรูปแบบผู้สอบกลุ่มสุ่ม จุดกำเนิดและหน่วยของการวัดค่าพารามิเตอร์ที่ประมาณได้สำหรับแบบสอบแต่ละฉบับเหมือนกัน แต่ทิศทางของแกนของระบบพิกัดอาจไม่เหมือนกัน ดังนั้นการหมุนต้องกำหนดให้อยู่แนวเดียวกับแกนพิกัด (coordinate axes) ของแบบสอบทั้งสองฉบับ กรณีสองรูปแบบ NEAT ใช้ข้อสอบร่วมในการกำหนดการหมุน กรณีสองรูปแบบผู้สอบกลุ่มเดียวใช้ข้อสอบร่วมในการกำหนดการหมุน กรณีสองรูปแบบผู้สอบกลุ่มสุ่มไม่มีทั้งข้อสอบร่วมและผู้สอบร่วม

โมเดลการตอบสนองข้อสอบแบบพหุมิติปรับแก้เดลต้ามิติ ดังสมการ (Zhang, 1996; Zhang & Stout, 1999a; Zhang & Wang, 1998 cited in Brossman & Lee, 2013)

$$p_{ij}(\theta_i) = c_j + (1 - c_j)H_j(\mathbf{a}_j^T \theta_i + d_j) = c_j + (1 - c_j)H_j(\sum_{k=1}^{\delta} a_{jk} \theta_k + d_j)$$

เมื่อ  $\mathbf{a}_j^T = (a_{j1}, a_{j2}, \dots, a_{j\delta})$ ,  $a_{j1}, a_{j2}, \dots, a_{j\delta}$  ไม่ได้มีค่าและไม่เป็นศูนย์ทั้งหมด

$H_j(x)$  คือ ฟังก์ชันที่ไม่ลดลง (non-decreasing [link] function)

$H_j(x) \geq 0$  ทุกค่าของ  $x$  และ  $H_j(x) H_{j^*}(x^*)$  ไม่เป็นศูนย์ เพราะว่า  $(x, x^*)$  มีค่าแปรผันสำหรับ  $j^* = 1, 2, \dots, n$  (เครื่องหมาย \* แทน ฟังก์ชันความแตกต่าง)

$H_j(\cdot)$  คือ ฟังก์ชันเชื่อมโยงของตระกูล (family) ของโมเดล

โมเดลการตอบสนองข้อสอบปรับแก้เดลต้ามิติที่ใช้ทั่วไปมี 2 โมเดล ได้แก่ โมเดลโลจิสติก 3 พารามิเตอร์ปรับแก้แบบพหุมิติ (M3PL) และโมเดล ogive ปกติปรับแก้แบบพหุมิติ

โมเดล M3PL คือ

$$p_{ij}(\theta_i) = c_j + (1 - c_j) \left( \frac{\exp[1.7(\mathbf{a}_j^T \theta_i + d_j)]}{1 + \exp[1.7(\mathbf{a}_j^T \theta_i + d_j)]} \right)$$

$$H_j(\theta_i) = c_j + (1 - c_j) \Phi(\mathbf{a}_j^T \theta_i + d_j)$$

ฟังก์ชันเชื่อมโยงที่สอดคล้องกัน คือ

$$H_j(\theta_i) = c_j + (1 - c_j) \Phi(\cdot)$$

เมื่อ  $\Phi(\cdot)$  คือ ฟังก์ชันการแจกแจงปกติมาตรฐาน

กรอบแนวคิด UIRT ประมาณค่าสัมประสิทธิ์ 2 ค่า ได้แก่ ค่า A และ B เพื่อปรับให้เข้ากับความแตกต่างใน (1) จุดกำเนิด (origin) (2) หน่วยของการวัด (unit of measurement) ส่วนการเชื่อมโยงสเกลด้วย MIRT ประมาณค่าสัมประสิทธิ์สเกลาร์และเมทริกซ์ เพื่อปรับให้เข้ากับความแตกต่าง (1) การแทนที่

(translation) ซึ่งคล้ายกับ “จุดกำเนิด” ของ UIRT (2) การเปลี่ยนขนาด (dilation) ซึ่งคล้ายกับหน่วยของการวัด (3) การหมุน และ (4) สหสัมพันธ์ ขณะที่การประมาณค่าพารามิเตอร์ภายใต้กรอบแนวคิด MIRT อาจแทนที่ค่าประมาณความสามารถและค่าประมาณพารามิเตอร์ของข้อสอบบนสเกลที่แตกต่างกัน เนื่องจากความไม่แน่นอนของการหมุน (rotational indeterminacy) ความไม่แน่นอนของสหสัมพันธ์ (correlation indeterminacy) และความไม่แน่นอนในจุดกำเนิดและหน่วยของการวัด

นิพจน์ (expression) ทางคณิตศาสตร์สำหรับการแปลงค่าประมาณพารามิเตอร์และค่าประมาณความสามารถจากสเกลใหม่ (A) ไปยังสเกลตั้งต้น (B) ใช้สมการทั่วไป คือ

$$\begin{aligned}\theta_{Bi} &= T^{-1}\theta_{Ai} + \beta, \\ \mathbf{a}_{Bj}^T &= \mathbf{a}_{Aj}^T T, \\ d_{Bj} &= d_{Aj} - \mathbf{a}_{Aj}^T T \beta, \\ c_{Bj} &= c_{Aj}\end{aligned}$$

เมื่อ  $\theta_i$  คือ เวกเตอร์ของค่าประมาณความสามารถพหุมิติ  
 $\mathbf{a}_j^T$  คือ เวกเตอร์ของค่าพารามิเตอร์อำนาจจำแนก  
 $d_j$  คือ ดัชนีที่สัมพันธ์กับพารามิเตอร์ความยาก  
 $T$  คือ เมทริกซ์มุมแหลม หรือเมทริกซ์มุมฉากเพื่อคำนวณ Indeterminacy ของการหมุน (rotational indeterminacy) และ Indeterminacy ของการเปลี่ยนขนาด (dilation indeterminacy)

$\beta$  คือ เมทริกซ์เพื่อคำนวณ Indeterminacy ของการแทนที่ (translation indeterminacy) สมการดังกล่าวขึ้นอยู่กับรูปแบบของการเชื่อมโยงสเกล ดังนั้น

$\beta$  อาจเท่ากับ 0

$T$  อาจเป็นผลลัพธ์ของเมทริกซ์การหมุนและพารามิเตอร์การเปลี่ยนขนาดเดียว หรืออาจผลลัพธ์ของเมทริกซ์การหมุนและพารามิเตอร์การเปลี่ยนขนาดแบบตั้งฉาก ซึ่งคำนวณสำหรับพารามิเตอร์การเปลี่ยนขนาดที่แยกกันในแต่ละมิติ หรืออาจเป็นเมทริกซ์การหมุนอย่างง่าย

Zhang (2012) อธิบายวิธีการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองแบบพหุมิติ (MIRT linking method) 6 วิธี ได้แก่ วิธีของ Hirsch วิธีของ Thompson, Nearing และ Davey วิธีของ Li และ Lisztz วิธีของ Min วิธีของ Reckase และ Martineau วิธีของ Oshima, Davey และ Lee

### 3.4.1) วิธีของ Hirsch (Hirsch's Method)

วิธีของ Hirsch พัฒนาขึ้นในปี 1989 เป็นวิธีแรกที่ใช้ในการเชื่อมโยงสเกลแบบพหุมิติ สำหรับรูปแบบผู้สอบร่วม และรูปแบบข้อสอบร่วม กระบวนการของวิธีการนี้มี 4 ขั้นตอน ได้แก่

1. การประมาณค่าพารามิเตอร์และการประมาณความสามารถ
2. การระบุเวกเตอร์: Common orthogonal basis vector สำหรับแบบสอบทั้งสองฉบับ เพื่อให้ค่าพารามิเตอร์ของแบบสอบทั้งสองฉบับอยู่ในพื้นที่มิติความสามารถเดียวกัน โดยใช้เมทริกซ์

การแปลง (transformation matrix) เชื่อมโยงระหว่างเวกเตอร์ความสามารถและเวกเตอร์ Common orthogonal basis vector ดังนั้นเวกเตอร์ common orthogonal basis vector จึงเป็นเมทริกซ์ร่วมระหว่างแบบสอบทั้งสองฉบับ ผลลัพธ์ที่ได้จากเมทริกซ์การแปลง คือ เวกเตอร์ค่าพารามิเตอร์อำนาจจำแนกของแบบสอบที่ถูกปรับเทียบ ซึ่งมี basis vector ที่มีมุมเท่ากับ basis vector ของแบบสอบฉบับตั้งต้น ถือว่าเป็นการหมุนครั้งแรกในวิธีการนี้ แต่อย่างไรก็ตาม Min (2003) แนะนำว่า มีโปรแกรมที่สำหรับประมาณค่าพารามิเตอร์แบบพหุมิติ เช่น NOHARM, TESTFECT และ ConQuest ซึ่งได้มิตินี้ความสามารถแบบตั้งฉาก จึงไม่จำเป็นต้องมีขั้นตอนการระบุ Common basis vector

3. จัดตำแหน่งเวกเตอร์ของแบบสอบทั้งสองฉบับผ่านการหมุนแบบตั้งฉาก โดยเป็นการหมุนครั้งที่สองผ่านเมทริกซ์การแปลง ซึ่งเป็นการหมุนค่าประมาณความสามารถแบบ Orthogonal Procrustes rotation โดยเป็นการหมุนค่าประมาณความสามารถของแบบสอบที่ถูกปรับเทียบในขั้น 2 ไปยังค่าประมาณความสามารถของแบบสอบที่ฉบับตั้งต้นในขั้น 2 เพื่อให้ได้ค่าความแตกต่างยกกำลังสองน้อยที่สุด (เช่น  $\sum \sum (\hat{\theta}_{kEj} - \hat{\theta}_{kBj})^2$ ) ซึ่งเป็นการแก้ปัญหา Rotation indeterminacy

4. ใช้การแปลงเชิงเส้น ปรับเทียบค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานในแต่ละมิติสำหรับแบบสอบทั้งสองฉบับ และในขั้นตอนนี้จัดการกับปัญหา Dilation indeterminacy และ Translation indeterminacy

### 3.4.2) วิธีของ Thompson, Nearing และ Davey (Thompson, Nearing and Davey's Method)

วิธีของ Thompson, Nearing และ Davey พัฒนาขึ้นในปี 1996 สำหรับรูปแบบผู้สอบกลุ่มเท่าเทียมกันอย่างสม่ำเสมอมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากัน จึงไม่ต้องแก้ปัญหา Dilation indeterminacy และ Translation indeterminacy แต่แบบสอบทั้ง 2 ฉบับที่นำมาปรับเทียบกันต้องมีโครงสร้างเหมือนกันมีคุณลักษณะ (trait) ที่ต้องใช้ในแบบสอบใกล้เคียงกันมากในขั้นตอนระบุเมทริกซ์การแปลงใช้การหมุน Orthogonal Procrustes rotation เพื่อให้ได้ matrix trace perator ยกกำลังสองน้อยที่สุด

### 3.4.3) วิธีของ Li และ Lisztz (Li and Lisztz's (LL) Method)

วิธีของ Li และ Lisztz พัฒนาขึ้นในปี 2000 สำหรับรูปแบบ NEAT วิธีการแก้ปัญหา MIRT scale indeterminacy คือ กระบวนการ Best composite transformation procedure โดยจัดการกับปัญหาการหมุน การเลื่อนขนาน และการเปลี่ยนขนาดแยกกัน กระบวนการ Best composite transformation procedure ใช้ในการหาส่วนประกอบในการเชื่อมโยง (linking component) 3 ส่วน ได้แก่

1. เมทริกซ์การหมุน (rotation matrix) โดยใช้ Orthogonal Procrustes rotation
2. เวกเตอร์การเลื่อนขนาน (translation vector) ที่มีความแตกต่างระหว่างความยากของแบบสอบทั้งสองฉบับยกกำลังสองน้อยที่สุด

3. ค่าคงที่การเปลี่ยนขนาดศูนย์กลาง (central dilation constant) โดยใช้ Trace method เพื่อให้ผลรวมของความแตกต่างระหว่าง Centralize discrimination parameter matrices จากเมทริกซ์การหมุนของแบบสอบทั้งสองฉบับยกกำลังสองน้อยที่สุด

#### 3.4.4) วิธีของ Min (Min's (M) Method)

วิธีของ Min พัฒนาขึ้นในปี 2003 สำหรับรูปแบบ NEAT เพื่อแก้ไขจุดอ่อนของวิธี LL โดย Min เสนอให้ใช้เมทริกซ์การเลื่อนขนาดแทนค่าคงที่การเลื่อนขนาด (dilation constant) ส่วนการหาเมทริกซ์การหมุน และเวกเตอร์การเลื่อนขนาดของวิธีนี้เหมือนกับวิธี LL แต่ใช้เทคนิค eigen value และ eigen vector

#### 3.4.5) วิธีของ Reckase และ Martineau (Reckase and Martineau's (NOP) Method)

วิธีของ Reckase และ Martineau พัฒนาขึ้นในปี 2004 สำหรับรูปแบบ NEAT และอาศัยการหมุนแบบ Non-orthogonal Procrustes transformation มีจุดเด่น 2 ประการได้แก่ 1) เมทริกซ์อำนาจจำแนกของข้อสอบในแต่ละมิติจากแบบสอบฉบับที่ถูกปรับเทียบมีมิติสอดคล้องสูงกับเมทริกซ์ของแบบสอบฉบับดั้งเดิม 2) ไม่จำเป็นต้องใช้ Dilation vector

#### 3.4.6) วิธีของ Oshima, Davey และ Lee (Oshima, Davey and Lee's (ODL) Method)

วิธีของ Oshima, Davey และ Lee พัฒนาขึ้นในปี 2000 สำหรับรูปแบบ NEAT ซึ่งวิธี ODL มีความแตกต่างจากวิธีการเชื่อมโยงสเกลอื่นๆ โดยพัฒนาจากแนวคิด IRT จึงมองว่าเป็นการขยายวิธีการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองแบบเอกมิติ ซึ่งมีการหมุนสองครั้ง ถ้าความสัมพันธ์ระหว่างมิติของแบบสอบทั้งสองฉบับเหมือนกัน เมทริกซ์ของการหมุนมีลักษณะตั้งฉาก วิธีการนี้ยังดำเนินการเปลี่ยนขนาดสเกล (scale dilation) ร่วมกับการหมุนทันที และไม่จำเป็นต้องใช้ค่าพารามิเตอร์การเปลี่ยนขนาด วิธีการ ODL มี 4 กระบวนการที่จะทำให้ฟังก์ชันเป้าหมายมีค่าน้อยที่สุด ได้แก่

##### 1) Direct (OD) procedure

กระบวนการนี้เป็นส่วนขยายแบบพหุตัวแปรของวิธีการเชื่อมโยงสเกลที่มีค่าไค-สแควร์น้อยที่สุดแบบเอกมิติ (unidimensional minimum chi-square linking method) โดยฟังก์ชันการเชื่อมโยงสเกลของกระบวนการ OD พิจารณาผลรวมที่น้อยที่สุดของความแตกต่างยกกำลังระหว่างองค์ประกอบที่สอดคล้องกันของเมทริกซ์พารามิเตอร์อำนาจจำแนกของข้อสอบสำหรับแบบสอบดั้งเดิมและเมทริกซ์พารามิเตอร์ของข้อสอบที่ถูกแปลงแล้วสำหรับแบบสอบที่ถูกปรับเทียบ และเวกเตอร์ความยากสำหรับแบบสอบฉบับดั้งเดิมและเวกเตอร์ความยากที่ถูกแปลงแล้วสำหรับแบบสอบที่ถูกปรับเทียบ

##### 2) Equated function procedure

กระบวนการนี้เป็นส่วนขยายแบบพหุมิติของวิธีค่าเฉลี่ย/ค่าเฉลี่ย โดยกระบวนการดังกล่าวต้องหาผลรวมที่น้อยที่สุดของความแตกต่างยกกำลังระหว่างฟังก์ชันที่ระบุในชุดของส่วนประกอบในเมทริกซ์อำนาจจำแนกและเวกเตอร์ความยาก

### 3) Test characteristic function (TCF) procedure

กระบวนการนี้เป็นส่วนขยายแบบพหุมิติของวิธี ของ Stocking และ Lord โดยกระบวนการนี้ต้องการหาผลรวมที่น้อยที่สุดของความแตกต่างระหว่างพื้นผิวคุณลักษณะของแบบสอบแบบพหุมิติ (TCS) ดังสมการ

$$f(\mathbf{A}, \boldsymbol{\beta}) = \frac{1}{q} \sum_{\boldsymbol{\theta}} \mathbf{W}_{\boldsymbol{\theta}} (\sum_{i=1}^n P_{1i}(\boldsymbol{\theta}) - \sum_{i=1}^n P_{2i}^*(\boldsymbol{\theta}))^2$$

เมื่อ  $P(\cdot)$  คือ ฟังก์ชันการตอบสนองข้อสอบสำหรับโมเดลพหุมิติด้วยชุดค่าพารามิเตอร์ที่สอดคล้องกัน  $(\mathbf{a}_{1i}, d_{1i})$  และ  $(\mathbf{a}_{2i}^*, d_{2i}^*)$

$\mathbf{a}_{1i}$  คือ เมทริกซ์ค่าพารามิเตอร์อำนาจจำแนกข้อสอบสำหรับแบบสอบฉบับดั้งเดิม

$\mathbf{a}_{2i}^*$  คือ เมทริกซ์ค่าพารามิเตอร์ข้อสอบที่ได้รับการแปลงสำหรับแบบสอบฉบับปรับเทียบ

$d_{1i}$  คือ เวกเตอร์ค่าความยากสำหรับแบบสอบฉบับฐาน

$d_{2i}^*$  คือ เวกเตอร์ค่าความยากสำหรับแบบสอบฉบับฐานสำหรับแบบสอบฉบับปรับเทียบ

$n$  คือ จำนวนข้อสอบรวม

กระบวนการ TCF ใช้วิธีการประมาณแบบต่อเนื่องในการอินทิเกรต และกำหนด  $q$  เป็นจำนวนกริดในการประมาณพื้นที่ (quadrature grids) สำหรับการจับคู่เวกเตอร์  $\boldsymbol{\theta}$  และการประมาณพื้นที่ความสามารถคำนวณจากการรวมกันภายนอกวงเล็บ โดยกระบวนการ TCS มี 2 รูปแบบ ได้แก่

1. รูปแบบถ่วงน้ำหนัก (weight version) การถ่วงน้ำหนักใช้ค่าความแตกต่างยกกำลังสองของผลรวมของเวกเตอร์น้ำหนักที่สอดคล้องกับค่าความสามารถที่แตกต่างกัน เพื่อให้ทราบว่ามีบางบริเวณของสเกลความสามารถมีความสำคัญมากกว่าส่วนอื่น

2. รูปแบบไม่ถ่วงน้ำหนัก (un-weight version) ทำให้ค่าความสามารถต่าง ๆ มีน้ำหนักเท่ากัน ( $\mathbf{W}_{\boldsymbol{\theta}} = 1$ )

งานวิจัยที่ศึกษาการเชื่อมโยงสเกลแบบพหุมิติสำหรับแบบสอบรูปแบบผสมใช้รูปแบบไม่มีการถ่วงน้ำหนัก (Yao & Boughton, 2009) โดยนักวิจัยเรียกกระบวนการนี้ว่า TRF ย่อมาจาก Test response function โดยความแตกต่างระหว่าง TRF เป็นดังสมการ

$$\min \left\{ \frac{1}{Q^D} \sum_{i=1}^{Q^D} [TRF(\boldsymbol{\theta}_i, \boldsymbol{\beta}_i) - TRF(\boldsymbol{\theta}_i, \boldsymbol{\beta}^*)]^2 \right\}$$

เมื่อ  $TRF(\boldsymbol{\theta}_i, \boldsymbol{\beta}_i) = \frac{1}{J_1 + \sum_{j=1}^{J_2} (K_j - 1)} \sum_{i=1}^N \left[ \sum_{j=1}^{J_1} P_{ij1} + \sum_{j=1}^{J_2} \sum_{k=1}^{K_j} (k - 1) P_{ijk} \right]$

เมื่อ  $Q^D$  คือ จุดในการประมาณพื้นที่ทั้งหมด (quadrature points)

$D$  คือ จำนวนของมิติ

$P_{ijk}$  คือ ความน่าจะเป็นของการตอบในขั้น  $k - 1$  สำหรับแบบสอบที่มีการตรวจให้คะแนนหลายค่า  $j$  ข้อ ของผู้สอบที่มีเวกเตอร์ความสามารถ  $\boldsymbol{\theta}_i$  ซึ่งความน่าจะเป็น  $P_{ijk}$  หาได้จาก Partial credit model (M-2PCM)

$J_1$  คือ จำนวนของข้อสอบที่ตรวจให้คะแนนสองค่า

$J_2$  คือ จำนวนของข้อสอบที่ตรวจให้คะแนนหลายค่า

กระบวนการนี้ใช้โปรแกรม LinkMIRT ที่พัฒนาโดย Yao ปี 2004 เพื่อหาเมทริกซ์การแปลง และเวกเตอร์ตำแหน่ง (location vector) โดยการกำหนดจุดกำเนิด จำนวนในการประมาณพื้นที่และจำนวนของการวนซ้ำมีผลต่อความถูกต้องในการเชื่อมโยงสเกล

#### 4) Item characteristic function (ICF) procedure

กระบวนการนี้มีความคล้ายคลึงกับกระบวนการ TCF ซึ่งต้องการหาความแตกต่างยกกำลังสองที่น้อยที่สุดของผลรวมของฟังก์ชันคุณลักษณะข้อสอบ (ICF) ระหว่างแบบสอบทั้งสองฉบับ กระบวน ICF เป็นส่วนขยายของวิธีการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองแบบเอกมิติของ Haebara เป้าหมายของกระบวนการนี้ คือ ทำให้ค่าพารามิเตอร์ของข้อสอบจากแบบสอบฉบับที่ส่วนถูกปรับเทียบคล้ายกับฉบับตั้งต้นมากที่สุดด้วยการเลือกเมทริกซ์การหมุนและเวกเตอร์การเลื่อนขนาน ที่เหมาะสมที่สุด

#### สรุปหลักการเกี่ยวกับการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองแบบพหุมิติ

วิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT ทั้ง 6 วิธีมีความแตกต่างกันในด้าน (1) รูปแบบการเก็บรวบรวมข้อมูล (2) การแก้ปัญหา Rotation indeterminacy (ใช้แนวคิดการวิเคราะห์องค์ประกอบหรือแนวคิด IRT) (3) วิธีการหมุน (4) การมี/ไม่มีค่าพารามิเตอร์การเปลี่ยนขนาด และ (5) ประเภทของค่าพารามิเตอร์การเปลี่ยนขนาด ซึ่งสามารถสรุปหลักการสำคัญของแต่ละวิธีได้ดังนี้

วิธีของ Hirsch เป็นวิธีการแรกที่ได้รับการพัฒนาซึ่งให้แนวทางที่สำคัญในการพัฒนาวิธีการต่อๆมา แต่มีความซับซ้อนและยากต่อการนำไปใช้จริง กระบวนการนี้ต้องกำหนดเมทริกซ์ร่วมไปจนถึงระบบสเกล การเชื่อมโยงสเกลจากแบบสอบทั้งสองฉบับ โดย Hirsch ระบุการแก้ปัญหการหมุน การเปลี่ยนขนาดและการเลื่อนขนานของการเชื่อมโยงสเกลด้วยโมเดล MIRT หลังจากนั้น Thompson ได้พัฒนาการเชื่อมโยงสเกลด้วยโมเดล MIRT ในปี 1997 สำหรับรูปแบบผู้สอบกลุ่มเท่าเทียมกัน วิธีการนี้ไม่จำเป็นต้องการเปลี่ยนขนาดและการเลื่อนขนาน แต่มีข้อตกลงเบื้องต้นที่สำคัญ คือ แบบสอบทั้งสองต้องมีความเป็นคู่ขนานอย่างแท้จริง และในการวิเคราะห์เชิงมิติใช้การวิเคราะห์กลุ่ม (cluster analysis) โดยคุณลักษณะในกลุ่มเดียวกันมีคอสโพลิทเหมือนกัน

การเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม (NEAT) มีหลายวิธี ได้แก่ วิธี LL วิธี M วิธี ODL และวิธี NOP โดยวิธี LL แก้ปัญหา Indeterminacy ทั้ง 3 ด้านแบบแยกกันโดยใช้เวกเตอร์การเลื่อนขนาน ค่าพารามิเตอร์การเปลี่ยนขนาด และ Orthogonal Procrustes rotation matrix ขณะที่วิธี M พัฒนาวิธี LL โดยแทนค่าคงที่การเปลี่ยนขนาดในวิธี LL ด้วย Diagonal dilation matrix ส่วนวิธี NOP ใช้การหมุนแบบไม่ตั้งฉาก (non-orthogonal rotation: NOP) เพื่อลดข้อจำกัดของวิธี M และเมื่อพิจารณาวิธีการเชื่อมโยงสเกลตามโมเดล MIRT วิธีต่างๆสำหรับรูปแบบ NEAT พบว่า วิธี ODL และวิธี NOP ใช้การหมุนแบบไม่ตั้งฉากได้ เพื่อแก้ปัญหา Rotation indeterminacy



โดยแก้ปัญหา Rotation indeterminacy และ Dilation indeterminacy ทั้งนี้ วิธี ODL และวิธี NOP จึงไม่มีค่าพารามิเตอร์การเปลี่ยนขนาด ขณะที่วิธี LL และวิธี M ต้องใช้การหมุนแบบตั้งฉากเท่านั้น

**ตาราง 2.10** สรุปเงื่อนไขของวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT หลังจากการประมาณค่าพารามิเตอร์

MIRT linking Method	วิธีการหมุนแกน		linking component			รูปแบบ
	OP	NOP	Rotation matrix	Translation vector	Central dilation	
Hirsch's Method	/		/	/		ผู้สอบร่วม/ ข้อสอบร่วม
Thompson, Nearing & Davey's Method	/					ผู้สอบกลุ่ม เท่าเทียมกัน
LL Method	/		/	/	/	NEAT
NOP Method		/	/	/		NEAT
ODL Method		/	/	/		NEAT

โปรแกรมเชื่อมโยงสเกลภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม

### 1) LinkMIRT

LinkMIRT เป็นโปรแกรมในชุดเครื่องมือ BMIRT (BMIRT toolkit) สำหรับเชื่อมโยงชุดค่าพารามิเตอร์ 2 ชุดตามกรอบแนวคิด MIRT โดยใช้ได้กับวิธีของ Stocking and Lord วิธีค่าเฉลี่ย/ค่าเฉลี่ย และวิธีค่าเฉลี่ย/ซิกมา สำหรับรูปแบบข้อสอบร่วม ผู้สอบร่วม ผู้สอบกลุ่มสุ่ม และผู้สอบกลุ่มเท่าเทียมกัน งานวิจัยของศิวทัศน์ สุขสุวรรณ, พัชรี จันทร์เพ็ง และสังวร ังดกระโทก (2555) ใช้โปรแกรมนี้กับวิธีการเชื่อมโยง TCF (Yao, 2016)

### 2) Package “plink” ในโปรแกรม R

Package “plink” ใช้สำหรับการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองข้อสอบสำหรับแบบสอบรูปแบบผสมภายใต้รูปแบบ NEAT โดยแพคเกจนี้ใช้ได้กับโมเดลการตอบสนองข้อสอบแบบเอกมิติ และวิธีการเชื่อมโยงสเกล 4 วิธี ได้แก่ วิธีค่าเฉลี่ย/ซิกมา วิธีค่าเฉลี่ย/ค่าเฉลี่ย วิธีของ Haebra และวิธีของ Stocking-Lord โดยแพคเกจนี้สามารถคำนวณเมทริกซ์การหมุน (rotation matrix) และเมทริกซ์การเลื่อนตำแหน่ง (translation matrix) เพื่อนำเมทริกซ์ดังกล่าวมาใช้ในการแปลงค่าพารามิเตอร์ระหว่างแบบสอบต่อไป (Weeks, 2010)

### 3.5 งานวิจัยที่เกี่ยวข้องกับการเชื่อมโยงสเกลตามทฤษฎีการตอบสนองแบบพหุมิติ

งานวิจัยที่เกี่ยวข้องกับการเชื่อมโยงด้วยโมเดล MIRT แบ่งเป็น 2 ประเด็นหลักตามประเด็นรูปแบบของแบบสอบ คือ แบบสอบรูปแบบผสม และข้อสอบที่ตรวจให้คะแนนสองค่า โดยงานวิจัยทั้งหมดใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติแบบทดแทนได้

#### 3.5.1) การเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับแบบสอบรูปแบบผสม

Patz และ Yao (2007) ได้พัฒนาวิธีการเชื่อมโยงสเกลแนวตั้ง (*vertical scaling*) ด้วยโมเดล MIRT สำหรับผู้สอบ 5 กลุ่ม (เกรด 3 4 5 6 และ 7) ผู้วิจัยพัฒนาอัลกอริทึม MCMC สำหรับโมเดลเพื่อใช้ในการสเกลแบบพร้อมกันสำหรับข้อสอบแบบหลายตัวเลือกและแบบสร้างคำตอบ (โมเดล M3PL และ M2PPC ตามลำดับ) โดยแบบสอบแต่ละฉบับมีข้อสอบร่วมกัน แบบสอบร่วมมีทั้งข้อสอบหลายตัวเลือกอย่างเดียวและรูปแบบผสม การประมาณค่าพารามิเตอร์ภายใต้อัลกอริทึม MCMC ใช้โปรแกรม BMIRT ของ Yao (2003) ผู้วิจัยใช้การประมาณค่าพารามิเตอร์แบบพร้อมกันสำหรับแบบสอบรูปแบบผสมภายใต้ผู้สอบหลายกลุ่ม

ผลการวิจัย พบว่า การประมาณค่าพารามิเตอร์ของข้อมูลจริงโดยใช้โปรแกรม BMIRT มีความสอดคล้องกับโมเดล และปรากฏความเป็นพหุมิติและระดับความเป็นพหุมิติสูงขึ้นเมื่อ (1) ข้อสอบแบบหลายตัวเลือกและแบบสร้างคำตอบถูกสเกลพร้อมกัน และ (2) ข้อมูลจากแต่ละเกรดถูกประมาณพร้อมกัน ผู้วิจัยระบุข้อจำกัดและข้อเสนอแนะว่า งานวิจัยนี้ใช้ข้อมูลจริง ช่วงค่าความยากและอำนาจจำแนกจึงไม่ได้ครอบคลุมมากนัก ดังนั้นผลการวิจัยจึงนำไปใช้อ้างอิงได้เฉพาะบริบทการทดสอบที่มีค่าความยากและอำนาจจำแนกคล้ายคลึงกับงานวิจัยนี้เท่านั้น

Yao & Boughton (2009) ศึกษาความถูกต้องของการเชื่อมโยงสเกลแบบพหุมิติสำหรับแบบสอบรูปแบบผสม สำหรับรูปแบบ NEAT โดยใช้ฟังก์ชันคุณลักษณะแบบสอบ (TCF) ซึ่งใช้ข้อมูลจำลองแบบ 2 มิติที่มาจากข้อมูลค่าพารามิเตอร์ของการทดสอบขนาดใหญ่ การประมาณค่าพารามิเตอร์ใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติเชิงยืนยันผ่านโปรแกรม BMIRT งานวิจัยนี้ศึกษากับชุดข้อสอบร่วม 12 เงื่อนไข ซึ่งมีทั้งข้อสอบร่วมรูปแบบผสม และข้อสอบร่วมหลายตัวเลือกเพียงอย่างเดียว และแต่ละเงื่อนไขมีการรวมกันของโครงสร้างอย่างง่ายและโครงสร้างซับซ้อนที่แตกต่างกัน โดยจำนวนข้อสอบร่วมมากที่สุด คือ 60 ข้อ (ใช้เป็นเกณฑ์) รองลงมา คือ 32, 24, 16, 8 และ 6 ขึ้นตามลำดับ การวิจัยครั้งนี้มีขนาดกลุ่มตัวอย่าง 3,000 คน และความถูกต้องของการปรับเทียบคะแนนพิจารณาจากค่า RMSE และค่าความลำเอียง

ผลการวิจัย พบว่า ผลการเชื่อมโยงสเกลในแต่ละเงื่อนไขมีประสิทธิภาพดี ยกเว้นเงื่อนไขข้อสอบรวม 6 ข้อ ในภาพรวมชุดข้อสอบรวมที่ประกอบด้วยโครงสร้างอย่างง่ายทั้ง 2 มิติมีประสิทธิภาพมากกว่าชุดข้อสอบรวมที่มีเฉพาะโครงสร้างแบบซับซ้อนหรือโครงสร้างอย่างง่ายเพียงอย่างเดียว และเมื่อจำนวนข้อสอบแบบสร้างคำตอบเพิ่มขึ้น ทำให้ประสิทธิภาพของการเชื่อมโยงสเกลในพื้นที่สองมิติเพิ่มขึ้น ซึ่งเมื่อพิจารณาเฉพาะเงื่อนไขข้อสอบรวมรูปแบบผสม พบว่า เงื่อนไขข้อสอบรวมที่มีค่า RMSE และความลำเอียงต่ำเมื่อเทียบกับเงื่อนไขเกณฑ์ (ข้อสอบรวม 60 ข้อ) คือ (1) กรณีข้อสอบรวม 16 ข้อ ประกอบด้วยข้อสอบหลายตัวเลือกและสร้างคำตอบ คือ 12 ข้อ (12 คะแนน) และ 4 ข้อ (16 คะแนน) คิดเป็นอัตราส่วน 43:57 และ (2) ข้อสอบรวม 32 ข้อ ประกอบด้วย ข้อสอบหลายตัวเลือกและสร้างคำตอบ คือ 24 ข้อ (24 คะแนน) และ 8 ข้อ (32 คะแนน) คิดเป็นอัตราส่วน 43:57 อย่างไรก็ตามกรณีข้อสอบรวม 8 ข้อ ประกอบด้วย ข้อสอบหลายตัวเลือกและสร้างคำตอบ คือ 6 ข้อ (6 คะแนน) และ 2 ข้อ (4 คะแนน) ก็ให้ผลการเชื่อมโยงสเกลดี นอกจากนี้เมื่อพิจารณาด้านค่าเฉลี่ยของประชากร พบว่า ประชากรที่มีค่าเฉลี่ยแตกต่างกันและประชากรมีความสัมพันธ์กันสูงมีความคลาดเคลื่อนในการประมาณมากกว่าประชากรมีค่าเฉลี่ยใกล้เคียงกับความยากของแบบสอบและประชากรมีความสัมพันธ์ระหว่างมิติสูง ผู้วิจัยระบุข้อเสนอแนะว่า งานวิจัยต่อไปควรศึกษากับข้อมูลพหุมิติที่มีมากกว่า 2 มิติ

ศิวะทัศน์ สุขสุวรรณ (2555) เปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลแนวตั้งแบบพหุมิติ 2 วิธี ได้แก่ วิธีโค้งคุณลักษณะแบบสอบ (TCF) และการเชื่อมโยงสเกลด้วยวิธีการหมุนแบบไม่ตั้งฉาก (NOP) สำหรับแบบสอบรูปแบบผสม และศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงสเกล (2 วิธี) โครงสร้างมิติความสามารถ (2 รูปแบบ ได้แก่ โครงสร้างที่ไม่ซับซ้อนและโครงสร้างซับซ้อน) สัดส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า (3 รูปแบบ ได้แก่ 60:40, 70:30 และ 80:20) และสัดส่วนข้อสอบรวม (3 รูปแบบ ได้แก่ 20%:, 25% และ 3) ที่ส่งผลต่อประสิทธิภาพของการเชื่อมโยงสเกล เงื่อนไขที่ศึกษาทั้งหมด คือ 36 เงื่อนไข

ผลการวิจัย พบว่า วิธีการเชื่อมโยงสเกล โครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบรวมแต่ละประเภท และสัดส่วนคะแนนข้อสอบรวมร่วมกันส่งผลต่อประสิทธิภาพของการเชื่อมโยงสเกลอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.5 การเชื่อมโยงสเกลด้วยวิธี NOP เมื่อโครงสร้างมิติความสามารถไม่ซับซ้อน เงื่อนไขสัดส่วนคะแนนข้อสอบรวมแต่ละประเภท 80%:20% และสัดส่วนคะแนนข้อสอบรวม 20% มีประสิทธิภาพมากกว่าเงื่อนไขอื่นๆ และเมื่อโครงสร้างมิติความสามารถซับซ้อน เงื่อนไขสัดส่วนคะแนนข้อสอบรวมแต่ละประเภท 80%:20% และสัดส่วนคะแนนข้อสอบรวม 20% มีประสิทธิภาพมากกว่าเงื่อนไขอื่นๆ เมื่อพิจารณาประสิทธิภาพของวิธีการเชื่อมโยงสเกลโดยรวม พบว่า การเชื่อมโยงสเกลด้วยวิธี NOP เมื่อโครงสร้างมิติความสามารถไม่ซับซ้อน สัดส่วนคะแนนข้อสอบรวมแต่ละประเภท 80%:20% และสัดส่วนคะแนนข้อสอบรวม 20% มีประสิทธิภาพมากกว่าวิธี TCF ผู้วิจัยระบุข้อจำกัดและข้อเสนอแนะว่า (1) ควรศึกษากับวิธีการเชื่อมโยงอื่น ๆ (2) การประมาณค่าพารามิเตอร์ด้วยโปรแกรม

BMIRT มักมีความคลาดเคลื่อนสูงเสมอ โดยเฉพาะอย่างยิ่งเมื่อค่าอำนาจจำแนกของข้อสอบมีแนวโน้มสูงกว่าความเป็นจริง จึงควรตรวจสอบว่ารูปแบบการตอบที่คำนวณจากค่าพารามิเตอร์ที่แท้จริงเป็นไปตามเงื่อนไขหรือไม่ (3) ควรศึกษากับข้อมูลที่มีหลายมิติ

ศิวะทัศน์ สุขสุวรรณ, พัทรี จันทร์เพ็ง สัจจวรรณ ังตกระโทก และภณัฐ ก้วยเจริญพานิชก์ (2555) ศึกษาผลของสัดส่วนของน้ำหนักคะแนนของข้อสอบร่วมที่มีต่อประสิทธิภาพในการเชื่อมโยงสเกลด้วยวิธี NOP สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT ด้วยการจำลองข้อมูลด้วยกลุ่มตัวอย่าง 3,000 คน สัดส่วนของข้อสอบร่วมที่ศึกษามี 3 แบบ ได้แก่ 20%, 25% และ 30% โดยมีโครงสร้างมิติความสามารถแบบไม่ซับซ้อน (APSS) งานวิจัยนี้ใช้การประมาณค่าพารามิเตอร์ผ่านโปรแกรม BMIRTI

ผลการวิจัย พบว่า สัดส่วนของข้อสอบร่วมส่งผลต่อประสิทธิภาพในการเชื่อมโยงสเกลอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.5 โดยการเชื่อมโยงสเกลด้วยสัดส่วนของข้อสอบร่วม 25% มีประสิทธิภาพมากที่สุด แต่งานวิจัยนี้ศึกษาเฉพาะวิธี NOP สำหรับโครงสร้างมิติความสามารถแบบไม่ซับซ้อน (APSS) ผู้วิจัยระบุข้อเสนอแนะว่า ควรศึกษาการวิธีการเชื่อมโยงสเกลอื่น ๆ อย่างเช่น วิธี TCF สำหรับโครงสร้างมิติความสามารถแบบซับซ้อน (CS)

ศิวะทัศน์ สุขสุวรรณ, พัทรี จันทร์เพ็ง และสัจจวรรณ ังตกระโทก (2555) ศึกษาประสิทธิภาพของการเชื่อมโยงสเกลแนวตั้งแบบพหุมิติสำหรับแบบสอบรูปแบบผสมด้วยวิธี TCF ภายใต้รูปแบบ NEAT ใน 2 ตัวแปร ได้แก่ (1) สัดส่วนน้ำหนักคะแนนของข้อสอบร่วม 3 แบบ ได้แก่ 20%, 25% และ 30% และ (2) มิติโครงสร้างความสามารถ 2 แบบ ได้แก่ โครงสร้างแบบไม่ซับซ้อน (APSS) และโครงสร้างแบบซับซ้อน (CS) สัดส่วนของคะแนนระหว่างข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่า คือ 70%:30% ซึ่งใช้การจำลองข้อมูลกับกลุ่มตัวอย่าง 3,000 คน งานวิจัยนี้ประมาณค่าพารามิเตอร์ผ่านโปรแกรม BMIRTI

ผลการวิจัย พบว่า สัดส่วนน้ำหนักคะแนนของข้อสอบร่วมและโครงสร้างมิติความสามารถไม่ร่วมกันส่งผลต่อประสิทธิภาพของการเชื่อมโยงสเกล แต่สัดส่วนน้ำหนักคะแนนของข้อสอบร่วมและโครงสร้างมิติความสามารถต่างส่งผลต่อประสิทธิภาพของการเชื่อมโยงสเกล โดยการเชื่อมโยงสเกลในโครงสร้างมิติความสามารถแบบไม่ซับซ้อนมีประสิทธิภาพมากกว่าโครงสร้างมิติความสามารถแบบซับซ้อน และการเชื่อมโยงสเกลด้วยสัดส่วนน้ำหนักของข้อสอบร่วม 25% มีประสิทธิภาพมากกว่าสัดส่วนอื่น ๆ แต่วิธี TCF ยังไม่มีความชัดเจนเกี่ยวกับการแปลงค่าพารามิเตอร์ที่สัมพันธ์กับความยาก ผู้วิจัยระบุข้อเสนอแนะว่า ควรเปรียบเทียบวิธี TCF กับวิธีการเชื่อมโยงสเกลอื่น ๆ อย่างเช่น วิธี NOP ภายใต้โมเดล IRT อื่น ๆ เช่น Grade Response Model และ Generalize Partial Credit Model รายละเอียดงานวิจัยสรุปได้ดังตาราง 2.11

## ประเด็นที่สรุปได้จากงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับแบบสอบรูปแบบผสม

### 1) โมเดลการตอบสนองข้อสอบแบบพหุมิติ

โมเดลการตอบสนองข้อสอบแบบพหุมิติที่นิยมศึกษามากที่สุด คือ โมเดล MIRT แบบทดแทนได้ ข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่าใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (M3PL) และโมเดล Partial credit model แบบ 2 พารามิเตอร์ (M2PPC) ตามลำดับ

### 2) วิธีการเชื่อมโยงสเกล และการประมาณค่าพารามิเตอร์

งานวิจัยมีการเปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยง อย่างเช่น วิธี TCF และ NOP และมิงงานวิจัยที่ศึกษาประสิทธิภาพการประมาณค่าพารามิเตอร์แบบพร้อมกัน ซึ่งให้ผลการประมาณค่าพารามิเตอร์ที่ดีและสอดคล้องกับโมเดล แต่ยังไม่มีการศึกษาในเชิงเปรียบเทียบระหว่างการประมาณค่าพารามิเตอร์แบบพร้อมกันกับการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธีการเชื่อมโยงสเกล โปรแกรมเชื่อมโยงสเกลระหว่างแบบสอบ คือ LinkMIRT และส่วนการประมาณค่าพารามิเตอร์สำหรับแบบสอบรูปแบบผสมดำเนินการด้วยโปรแกรม BIMIRT และโปรแกรม BMIRTII ด้วยวิธี Markov chain Monte Carlo (MCMC)

### 3) รูปแบบการออกแบบการเก็บรวบรวมข้อมูล

รูปแบบการออกแบบการเก็บรวบรวมข้อมูลที่ศึกษา คือ รูปแบบ NEAT

### 4) ลักษณะของแบบสอบและแบบสอบร่วม

แบบสอบรูปแบบผสมประกอบด้วยข้อสอบที่ตรวจให้คะแนนสองค่ามากกว่าข้อสอบที่ตรวจให้คะแนนหลายค่า สัดส่วนคะแนนระหว่างข้อสอบที่ตรวจให้คะแนนสองค่ามากกว่าข้อสอบที่ตรวจให้คะแนนหลายส่วนใหญ่ คือ 70%:30% รองลงมา คือ 85%:25% และสัดส่วนจำนวนข้อสอบแต่ละประเภทที่ใช้ คือ 75%:25% (คิดเป็นสัดส่วนคะแนนสำหรับข้อสอบแต่ละประเภทได้เป็น 50:50)

ข้อสอบร่วมที่ใช้เป็นข้อสอบร่วมภายในมีสัดส่วนอยู่ในช่วง 20%-50% โดยการเชื่อมโยงสเกลด้วยวิธี TCF สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT พบว่า สัดส่วนแบบสอบร่วม 25% มีความเหมาะสมมากที่สุด (ศิวะทัศน์, พัชรี จันทรพิง และสังวรณ์ ังคระโทก, 2555) งานวิจัยส่วนใหญ่ใช้สัดส่วนคะแนนของข้อสอบร่วมที่ตรวจให้คะแนนสองค่ามากกว่าข้อสอบที่ตรวจให้คะแนนหลายค่า โดยสัดส่วนของข้อสอบร่วมแต่ละประเภทดังกล่าวมีทั้งน้อยกว่าและมากกว่าสัดส่วนของข้อสอบแต่ละประเภทในแบบสอบ นอกจากนี้ยังมีงานวิจัยที่ศึกษาโครงสร้างมิติความสามารถ พบว่า ข้อสอบร่วมที่มีโครงสร้างอย่างง่าย (APSS) ทั้ง 2 มิติมีประสิทธิภาพมากกว่าข้อสอบร่วมที่มีโครงสร้างซับซ้อน (CS) เพียงอย่างเดียว หรือมีข้อสอบร่วมที่มีโครงสร้างอย่างง่าย (APSS) จำนวนน้อย และจำนวนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าคิดเป็น 25% ของจำนวนข้อสอบร่วมทั้งหมดมีประสิทธิภาพมากกว่าการใช้ข้อสอบร่วมที่ตรวจให้คะแนนสองค่าเพียงอย่างเดียว

### 5) โครงสร้างมิติความสามารถ

โครงสร้างมิติความสามารถ (dimensional structure) ที่นิยมใช้ คือ โครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อน (CS) เช่นเดียวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับแบบสอบที่ประกอบด้วยข้อสอบที่มีการตรวจให้คะแนนสองค่า

### 6) ขนาดของกลุ่มตัวอย่าง

ขนาดของกลุ่มตัวอย่างที่ใช้ในการวิจัย คือ 2,500 และ 3,000 คน

### 7) การประเมินประสิทธิภาพของการเชื่อมโยงสเกล

ประสิทธิภาพของการเชื่อมโยงสเกลส่วนใหญ่พิจารณาจากค่าความลำเอียง (bias) และรากที่สองของความคลาดเคลื่อนยกกำลังสอง (RMSE)

### 8) ตัวแปรที่ศึกษา

งานวิจัยส่วนใหญ่ศึกษาและเปรียบเทียบประสิทธิภาพของการวิธีการเชื่อมโยงสเกลร่วมกับตัวแปรอื่นๆ (1) ด้านกลุ่มตัวอย่าง: ค่าเฉลี่ยของประชากรและเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วม (2) ด้านข้อสอบรวม: ความยาวของข้อสอบรวม ประเภทและสัดส่วนคะแนนของข้อสอบรวม โครงสร้างมิติความสามารถ และโครงสร้างของชุดข้อสอบ (3) ด้านแบบสอบ: โครงสร้างมิติความสามารถ ความเท่าเทียมของความยาก

### 9) ผลการวิจัยโดยสรุป

งานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT มีการศึกษาประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบพร้อม ซึ่งพบว่า วิธีการเชื่อมโยงสเกลแนวตั้ง (vertical scaling) ที่มีการประมาณค่าพารามิเตอร์พร้อมกันให้ข้อมูลที่สอดคล้องกับโมเดล ขณะที่การเปรียบเทียบประสิทธิภาพระหว่างวิธีการเชื่อมโยงสเกล (วิธี TCF และวิธี NOP) พบว่า วิธี NOP มีประสิทธิภาพมากกว่าในการเชื่อมโยงสเกลแนวตั้ง แต่เมื่อพิจารณาการเชื่อมโยงสเกลแบบสอบพหุมิติรูปแบบผสมด้วยวิธี TCF ภายใต้รูปแบบ NEAT พบว่า วิธี TCF มีประสิทธิภาพในการเชื่อมโยงสเกลสำหรับแบบสอบรูปแบบผสมเช่นกัน

### 10) ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

งานวิจัยที่ใช้ข้อมูลจริงเสนอให้มีการศึกษากับแบบสอบที่มีความยากและอำนาจจำแนกครอบคลุมและหลากหลายด้วยการจำลองข้อมูล งานวิจัยส่วนใหญ่เสนอให้ศึกษากับการเชื่อมโยงอื่นๆและโมเดลอื่นๆที่ยังไม่ได้ทำการศึกษา และเมื่อพิจารณางานวิจัยที่ผ่านมาพบว่า ยังไม่มีการศึกษาภายใต้โมเดล Grade Response Model และ Generalize Partial Credit Model นอกจากนี้ยังไม่มีเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบพร้อมกันและแบบแยกกันสำหรับแบบสอบพหุมิติรูปแบบผสม

ตาราง 2.11 สรุปประเด็นสำคัญของงานวิจัยการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT

ประเด็น	Patz และ Yao (2007)	Yao & Boughton (2009)	ศิวัชรัตน์ สุสุวรรณ (2555)	ศิวัชรัตน์ สุสุวรรณ และคณะ (2555)	ศิวัชรัตน์ สุสุวรรณ และคณะ (2555)
MIRT model	M3PL+M2PPC				
วิธีเชื่อมโยงสเกล					
Concurrent calibration	/				
TCF		/	/	/	/
NOP			/		
ความยาวแบบสอบ MC:CR (คะแนน)	35:18, 40:13, 41:13, 42:13, 38:13	60 45:15ข้อ	80%:20% 70%:30% 60%:40%	70%:30%	70%:30%
ความยาวแบบสอบ ร่วม (คะแนน)	- MC= 18 - MC:CR = 18:4, 19:4	DICHO (ข้อ) = 6, 8, 16, 32, 60 DICHO:POLY = 3:1 (ข้อ)		20%, 25%, 30% DICHO:POLLY =70:30	20%, 25%, 30% DICHO:POLY=70 :30
โครงสร้างมิติ ความสามารถ	-	-SS (2 มิติ) -CS (2 มิติ) r=0, 0.5	-APSS (2 มิติ) -CS (2 มิติ)	-APSS (2 มิติ)	-APSS (2 มิติ) -CS (2 มิติ)
ขนาดกลุ่มตัวอย่าง	2,500 แต่ละ เกรด	3,000	3,000 แต่ละ ระดับ	3,000	3,000
โปรแกรมเชื่อมโยง	N/A	LinkMIRT	LinkMIRT	LinkMIRT	LinkMIRT
ดัชนี	- AIC - BIC	- BIAS - RMSE	- RMSE - F-test	- BIAS - RMSE	- BIAS - RMSE
ตัวแปรที่ศึกษา					
- วิธีการเชื่อมโยง	/				
- โครงสร้างมิติ ความสามารถ			/		/
- สัดส่วนคะแนน ข้อสอบร่วม			/	/	/
- โครงสร้างข้อสอบร่วม		/			
- ประเภทข้อสอบร่วม		/			
- สัดส่วนข้อสอบร่วม		/	/		
- ความแตกต่างของ กลุ่มตัวอย่าง		/			

### 3.5.2) งานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า

#### 1. การเปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบพร้อมกัน และการประมาณค่าพารามิเตอร์แบบแยกกันด้วยการเชื่อมโยงสเกล

งานวิจัยในประเด็นดังกล่าว พบว่า **วิธีการประมาณค่าพารามิเตอร์แบบพร้อมกัน มีประสิทธิภาพมากกว่าวิธีการเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์แบบแยกกัน** รายละเอียดดังนี้

Lin (2008) เปรียบเทียบประสิทธิภาพระหว่าง (1) **วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันด้วยโมเดล MIRT (โปรแกรม BMIRT)** (2) **วิธีการเชื่อมโยงสเกลหลังจากประมาณค่าพารามิเตอร์ด้วยโมเดล MIRT (โปรแกรม NOHARM)** วิธีการเชื่อมโยงสเกลที่ใช้คือ วิธีของ Min (3) **วิธีการวิเคราะห์องค์ประกอบแบบพร้อมกัน (โปรแกรม Mplus)** และ (4) **วิธีประมาณค่าพารามิเตอร์แบบพร้อมกันด้วยโมเดล UIRT (โปรแกรม BILOG)** สำหรับ**ข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT** งานวิจัยครั้งนี้ใช้การจำลองข้อมูลเพื่อศึกษาปัจจัย 4 ปัจจัย ได้แก่ ความตั้งใจเชิงโครงสร้าง ความเท่าเทียมของโครงสร้าง ความเท่าเทียมของความยาก และความเท่าเทียมของผู้สอบ โดยความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับปานกลางไปจนถึงสูง คือ 0.5, 0.7 และ 0.9 ประสิทธิภาพของวิธีการเหล่านี้พิจารณาจากการกู้คืนค่าพารามิเตอร์ และการประมาณคะแนนจริงของผู้สอบพิจารณาจากค่าความลำเอียงและส่วนเบี่ยงเบนมาตรฐาน

ผลการวิจัย พบว่า การวิเคราะห์องค์ประกอบแบบพร้อมกันมีประสิทธิภาพในการกู้คืนค่าพารามิเตอร์มากกว่าวิธีที่ใช้โมเดล MIRT ทั้งสองวิธี แต่วิธีที่ใช้โมเดล MIRT ทั้งสองวิธีมีประสิทธิภาพในการประมาณคะแนนจริงของผู้สอบมากกว่าวิธีวิเคราะห์องค์ประกอบแบบพร้อมกัน และวิธีที่ใช้โมเดล UIRT มีความแกร่งต่อการละเมิดข้อตกลงเบื้องต้นของความเป็นเอกมิติ **เมื่อพิจารณาผู้สอบสองกลุ่มที่ไม่เท่าเทียมกัน พบว่า วิธีประมาณค่าพารามิเตอร์แบบแยกกันด้วยโมเดล MIRT มีความลำเอียงมากกว่า** และมีความเสถียรต่ำกว่าในการกู้คืนค่าพารามิเตอร์ จึงจำเป็นต้องอาศัยการเชื่อมโยงสเกล แต่เมื่อใช้การเชื่อมโยงสเกลแล้วพบว่ามีความลำเอียงมากและเสถียรต่ำ สะท้อนให้เห็นว่าวิธี Min อาจไม่เหมาะสมกับบริบทของงานวิจัย แต่อย่างไรก็ตามผลการวิจัยไม่สามารถอ้างอิงไปยังสถานการณ์ที่มีมิติมีความสัมพันธ์กันต่ำ

Simon (2008) เปรียบเทียบประสิทธิภาพของการเชื่อมโยงสเกลด้วยโมเดล MIRT แบบแยกกัน (separate MIRT linking) 4 วิธี ได้แก่ วิธี TCF วิธี ICF วิธี DO และวิธี M และการประมาณค่าพารามิเตอร์แบบพร้อมกัน (concurrent calibration) สำหรับ**รูปแบบผู้สอบกลุ่มเท่าเทียมกัน** ตัวแปรอิสระที่ศึกษา 5 ตัวแปร ได้แก่ (1) **ขนาดของกลุ่มตัวอย่าง** คือ 500, 1,000 และ 3,000 คน (2) **ความยาวของแบบสอบ** คือ 60 ข้อที่มีข้อสอบร่วม 20 ข้อ (ประมาณ 30%) และ 40 ข้อ ที่มีข้อสอบร่วม 20 ข้อ (คิดเป็น 50%) (3) **ความเท่าเทียมของกลุ่มตัวอย่างสองกลุ่ม** คือ กลุ่มเท่าเทียมกันและกลุ่มไม่เท่าเทียมกัน (4) **ความสัมพันธ์ระหว่างมิติความสามารถ** คือ 0.5, 0.7 และ 0.9 และ**วิธีการเชื่อมโยงสเกล MIRT** คือวิธี TCF วิธี ICF วิธี DO และวิธี M และการประมาณค่าพารามิเตอร์แบบพร้อมกัน



ผลการวิจัยพบว่า **วิธีการประมาณค่าพารามิเตอร์พร้อมกันมีประสิทธิภาพมากกว่าวิธีการเชื่อมโยงสเกลแบบแยกกัน** (แม้กระทั่งในรูปแบบ NEAT ที่มีส่วนเบี่ยงเบนมาตรฐาน 0.5 และมีความสามารถมีความสัมพันธ์กันสูง) กรณีกลุ่มตัวอย่างใหญ่ขึ้น การประมาณค่าพารามิเตอร์แบบพร้อมกันมีประสิทธิภาพมากกว่าวิธีการเชื่อมโยงสเกลแบบแยกกัน (โดยเฉพาะอย่างยิ่งในแบบสอบฉบับสั้น) เมื่อพิจารณาการเชื่อมโยงสเกลแบบแยกกันในรูปแบบ NEAT พบว่า **วิธี ICF มีประสิทธิภาพมากกว่าวิธีการเชื่อมโยงสเกลแบบแยกกันอีก 3 วิธี** ส่วนวิธี M มีประสิทธิภาพต่ำที่สุด แต่ในรูปแบบผู้สอบกลุ่มเท่าเทียมกัน **วิธีการเชื่อมโยงสเกลแบบแยกกันทั้ง 4 วิธีมีประสิทธิภาพใกล้เคียงกัน** ผู้วิจัยให้ข้อเสนอแนะว่า (1) ควรศึกษากับข้อสอบที่วัดหลายมิติ เนื่องจากในงานวิจัยนี้ ข้อสอบแต่ละข้อวัดเพียงมิติเดียว (2) ควรเลือกใช้วิธีการเชื่อมโยง เพื่อเปรียบเทียบประสิทธิภาพของวิธีเหล่านั้น (3) ควรศึกษากับโมเดล Non-compensatory MIRT (4) ควรศึกษาประสิทธิภาพของโปรแกรมต่างๆที่ใช้ในการประมาณค่าพารามิเตอร์ เช่น TESTFACT MIRTE และ MULTIFACT (5) ควรศึกษาการประมาณความสามารถของผู้สอบ เนื่องจากการประเมินประสิทธิภาพของการเชื่อมโยงสเกลในงานวิจัยนี้ใช้เพียงค่าพารามิเตอร์ของข้อสอบเท่านั้น (6) ควรศึกษากลุ่มตัวอย่าง 2 กลุ่มที่มีความแตกต่างกันทั้งในขนาดกลุ่มตัวอย่างและความสัมพันธ์ระหว่างมิติความสามารถ เนื่องจากวิจัยนี้ศึกษาในเงื่อนไขของกลุ่มตัวอย่าง 3 ระดับ และความสัมพันธ์ระหว่างมิติความสามารถ 3 ระดับ และการเชื่อมโยงสเกลกลุ่มตัวอย่างทั้ง 2 กลุ่มอยู่ภายใต้เงื่อนไขเดียวกัน

## 2. การเปรียบเทียบประสิทธิภาพระหว่างวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT

พัชรี จันทร์เพ็ง (2550) ได้เปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT 2 วิธี ได้แก่ วิธีของ Min (การหมุนแกนแบบตั้งฉาก) และวิธี NOP (การหมุนแบบไม่ตั้งฉาก) สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT โดยใช้จำนวนข้อสอบรวมภายใน 30% งานวิจัยนี้ใช้การจำลองข้อมูลเพื่อศึกษากับตัวแปรต่างๆ ได้แก่ (1) ระดับความสัมพันธ์ระหว่างมิติ 3 ระดับ ได้แก่ ไม่มีความสัมพันธ์ มีความสัมพันธ์ระดับปานกลาง และมีความสัมพันธ์ระดับสูง และ (2) โครงสร้างมิติความสามารถ 2 แบบ ได้แก่ โครงสร้างที่ไม่ซับซ้อน (APSS) และ โครงสร้างที่ซับซ้อน (CS) นอกจากนี้ยังศึกษาปฏิสัมพันธ์ระหว่างวิธีการเชื่อมโยงสเกล โครงสร้างมิติความสามารถ และระดับความสัมพันธ์ระหว่างมิติความสามารถ ดังนั้นเงื่อนไขที่ทำการศึกษามีทั้งหมด 12 เงื่อนไข

ผลการวิจัย พบว่า วิธีของ Min และวิธี NOP มีประสิทธิภาพใกล้เคียงกัน โดยเฉพาะอย่างยิ่งเมื่อมิติความสามารถไม่มีความสัมพันธ์กันและมีความสัมพันธ์ระหว่างปานกลางกับโครงสร้างที่ไม่ซับซ้อน แต่ถ้ามิติความสามารถมีความสัมพันธ์กันมากขึ้นกับโครงสร้างที่ซับซ้อน พบว่า วิธี NOP มีประสิทธิภาพมากกว่าวิธีของ Min อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.5 นอกจากนี้ยังพบว่าวิธีการเชื่อมโยงสเกล โครงสร้างมิติความสามารถ และความสัมพันธ์ระหว่างมิติความสามารถร่วมกันส่งผลต่อความลำเอียงของการเชื่อมโยงสเกลของการแปลงค่าพารามิเตอร์จากระดับกลางไประดับสูงอย่างมีนัยสำคัญทางสถิติที่

ระดับ 0.5 แต่ไม่ส่งผลต่อค่า RMSE ในทุกระดับการแปลงคะแนน และประสิทธิภาพของวิธีการเชื่อมโยงสเกลขึ้นอยู่กับโครงสร้างมิติความสามารถและความสัมพันธ์ระหว่างมิติความสามารถ

Hirsch (1989: บทคัดย่อ) ศึกษาวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT โดยใช้โมเดลโลจิสติกแบบ 2 โมเดล (M2PL) สำหรับรูปแบบ NEAT ซึ่งการวิจัยนี้ใช้ทั้งการข้อมูลจริงและการจำลองข้อมูล เพื่อเปรียบเทียบความคลาดเคลื่อนของการปรับเทียบคะแนนระหว่างคะแนนจริงและความสามารถที่ประมาณได้ ผลการวิจัย พบว่า เมื่อเปรียบเทียบกับความสามารถที่แท้จริง ถือว่าการปรับเทียบคะแนนมีประสิทธิภาพ แต่ความคงที่ของการประมาณความสามารถยังไม่ดีนัก (ไม่เป็นที่น่าพอใจ)

Davey, Oshima และ Lee (1994) ศึกษาประสิทธิภาพของการเชื่อมโยงสเกลด้วยโมเดล MIRT โดยใช้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ สำหรับรูปแบบ NEAT งานวิจัยนี้ใช้การจำลองข้อมูลความสามารถ 2 ชุดที่แตกต่างกัน เพื่อศึกษาคุณลักษณะ 2 มิติ ผู้วิจัยเชื่อมโยงสเกลของกลุ่มปรับเทียบและกลุ่มฐานให้อยู่สเกลเดียวกัน ผลการวิจัย พบว่า การใช้วิธีความแตกต่างของเวกเตอร์ค่าเฉลี่ยมีประสิทธิภาพมากกว่าการใช้วิธีความแตกต่างของเมทริกซ์ความแปรปรวนและความแปรปรวนร่วม นอกจากนี้ยังพบว่า การแปลงคะแนนแบบมูมฉากทำให้การประมาณค่าพารามิเตอร์มีความใกล้เคียงกับค่าพารามิเตอร์ที่แท้จริงน้อยกว่าการแปลงคะแนนแบบมูมแหลม ผู้วิจัยให้ข้อเสนอแนะว่า ควรทำการตรวจสอบความถูกต้องของผลที่ได้จากโปรแกรม NOHRM และควรศึกษาสมการการเชื่อมโยงสเกลแบบต่างๆว่าส่งผลอย่างไรต่อประสิทธิภาพของการเชื่อมโยงสเกล

Thompson, Nering และ Davey (1997) พัฒนารูปแบบวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับรูปแบบข้อสอบร่วมหรือผู้สอบร่วม (ไม่มีข้อสอบร่วม) โดยกำหนดคอมโพสิทอ้างอิง (reference composite) ด้วยการวิเคราะห์เนื้อหาอย่างละเอียดตามแนวคิดของ Reckase และคณะ (1997) และประยุกต์ใช้การหมุน Orthogonal Procrustes ผลการวิจัย พบว่า วิธีการที่ใช้ระบุการจัดกลุ่มเนื้อหาที่คล้ายคลึงจากแบบสอบหลายฉบับที่แตกต่างส่งผลต่อประสิทธิภาพของการปรับเทียบคะแนน ผู้วิจัยให้ข้อเสนอแนะว่า ควรพัฒนารูปแบบวิธีการเชื่อมโยงสเกลให้มีความชัดเจนและเป็นรูปธรรมเพื่อนำไปใช้กับข้อมูลจริง

Li และ Lissitz (2000) ประเมินวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT 3 วิธี โดยใช้โมเดลโลจิสติก 2 พารามิเตอร์ สำหรับรูปแบบ NEAT ซึ่งการวิจัยนี้ทำการจำลองคุณลักษณะแบบพหุมิติ ซึ่งเป็นข้อมูลพหุมิติที่มีการแจกแจงปกติ (MVN, (0,1)) การเชื่อมโยงสเกลประกอบด้วย 1) Orthogonal Procrustes rotation 2) Translation transformation และ 3) Single dilation ผลการวิจัย พบว่า วิธีการเชื่อมโยงสเกล MIRT ที่ไม่มีความลำเอียงและเป็นตัวประมาณที่คงที่ คือ การรวมกันของ Procrustes rotation, Rotation of trace และการประมาณแบบกำลังสองน้อยที่สุด (least squares estimation) ซึ่งใช้ในการประมาณเมทริกซ์การแปลงแบบตั้งฉาก (rotational transformation matrix) ค่าพารามิเตอร์การเปลี่ยนแปลงขนาด (dilation parameter) และ ชุดของค่าพารามิเตอร์การเลื่อนขนาน (translation parameter) โดย Rotation of trace เป็นวิธีที่ดีที่สุดในการประมาณค่าพารามิเตอร์ การเปลี่ยน

ขนาด รองลงมา คือ Ratio of eigenvalue และ MTRS ขณะที่วิธียกกำลังสองน้อยที่สุดมีความคลาดเคลื่อนน้อยกว่า MTRS ในการประมาณชุดของค่าพารามิเตอร์การเลื่อนขนาน ผู้วิจัยให้ข้อเสนอแนะว่าควรพัฒนาการเชื่อมโยงสเกลสำหรับแบบสอบที่มีจำนวนมิติมากขึ้น และศึกษาในหลายๆเงื่อนไข เช่น จำนวนข้อสอบรวม ขนาดกลุ่มตัวอย่าง ตัวประมาณค่าพารามิเตอร์ โมเดลการเชื่อมโยงสเกล เป็นต้น

Oshima, Davey และ Lee (2000) ประเมินวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT 4 วิธี โดยใช้โมเดลโลจิสติก 2 พารามิเตอร์ ได้แก่ (1) วิธี Direct (2) วิธี Equated Function (3) วิธี TCF และ (4) วิธี ICF สำหรับรูปแบบ NEAT ซึ่งการวิจัยนี้ใช้การจำลองข้อมูลเพื่อเปรียบเทียบการประมาณค่าพารามิเตอร์ของแต่ละวิธี ผลการวิจัยพบว่า วิธี TCF และวิธี ICF มีความคงที่และสามารถประมาณค่าพารามิเตอร์การเชื่อมโยงสเกลที่แท้จริงได้มากกว่าอีกสองวิธี นอกจากนี้ยังพบว่าวิธีการเชื่อมโยงสเกลทุกวิธีมีประสิทธิภาพอยู่ในเกณฑ์ที่รับได้ ดังนั้นในการเลือกวิธีการเชื่อมโยงสเกลที่ดีที่สุดจึงขึ้นอยู่กับจุดประสงค์ของการเชื่อมโยงสเกล ผู้วิจัยให้ข้อเสนอแนะว่าควรศึกษาในเงื่อนไขอื่นๆ อย่างเช่น จำนวนมิติ และควรพัฒนาเทคนิค/วิธีการเชื่อมโยงสเกล รวมถึงกระบวนการตรวจสอบเทคนิค/วิธีการเชื่อมโยงสเกลเหล่านั้น

Min (2003 อ้างอิงใน พัชรี จันทร์, 2550) ได้เปรียบเทียบวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT 3 วิธีโดยใช้โมเดลโลจิสติก 2 พารามิเตอร์ ได้แก่ (1) วิธี OLD (2) วิธี LL และ (3) วิธี M สำหรับรูปแบบ NEAT วิธีการเชื่อมโยงสเกลที่ใช้มีการแก้ปัญหา Scalar dilation parameter ซึ่งวิธี LL และวิธี M มีการหมุนแบบตั้งฉาก ส่วนวิธี OLD มีการหมุนแบบไม่ตั้งฉาก ผลการวิจัย พบว่ากรณีโครงสร้างแบบสอบที่ไม่ซับซ้อน (APSS) วิธี OLD ค่าอำนาจจำแนกที่ได้มีความคลาดเคลื่อนน้อยกว่าวิธี M แต่กรณีโครงสร้างแบบสอบซับซ้อน (MS) วิธี LL ค่าอำนาจจำแนกที่ได้มีความคลาดเคลื่อนมากกว่าอีกสองวิธี นอกจากนี้ยังพบว่า วิธี M ให้ค่าเมทริกซ์ (diagonal matrix) ที่มีความคงที่ และเมื่อพิจารณาผลในภาพรวม พบว่า วิธี M มีความคลาดเคลื่อนของค่าพารามิเตอร์อำนาจจำแนกน้อยกว่าอีกสองวิธี แต่วิธี M ไม่สามารถนำไปใช้กับข้อมูลที่มีจำนวนมิติมากเกินไป ผู้วิจัยให้ข้อเสนอแนะว่าควรพัฒนาวิธีการเชื่อมโยงสเกลที่สามารถใช้ได้กับข้อมูลที่มีหลายมิติมากขึ้น

Reckase และ Martineau (2004 อ้างอิงใน พัชรี จันทร์, 2550 และ Zhang, 2012) ศึกษาวิธีการเชื่อมโยงสเกลในแนวตั้งด้วยโมเดล MIRT โดยใช้โมเดลโลจิสติก 3 พารามิเตอร์ สำหรับรูปแบบ NEAT เพื่อนำผลการเชื่อมโยงสเกลไปศึกษาพัฒนาการภาคตัดขวางของผลสัมฤทธิ์ทางการเรียนของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 วิชาวิทยาศาสตร์ ซึ่งการเชื่อมโยงสเกลครั้งนี้ประกอบด้วย (1) Orthogonal Procrustes rotation (2) Translation transformation และ (3) Single dilation งานวิจัยนี้ได้พัฒนาวิธี M ที่มีข้อจำกัดในการนำไปใช้กับข้อมูลที่มีซับซ้อนเกินไปด้วยการแปลงคะแนนแบบไม่ตั้งฉาก (Non-orthogonal Procrustes rotation) ซึ่งใช้การแปลงคะแนนแบบหมุนแหลมเพื่อลดความยุ่งยากในการคำนวณ ผลการวิจัย พบว่า โมเดลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติสามารถสะท้อนให้เห็นถึงความซับซ้อนของการศึกษาภาคตัดขวางของผลสัมฤทธิ์ทางการเรียน ซึ่งสะท้อนให้เห็นถึงความ

แตกต่างของความรู้และทักษะของระดับความสามารถในแต่ละระดับชั้นของนักเรียนอย่างชัดเจน เนื่องจาก พัฒนาการของผลสัมฤทธิ์ทางการเรียนมีลักษณะไม่เป็นแบบแผน หลากหลาย โดยเฉพาะในระดับ 6 ไปยังระดับ 7 ผู้วิจัยให้ข้อเสนอแนะว่า ควรศึกษาข้อมูลที่มีจำนวนมากขึ้น และพัฒนาวิธีการเชื่อมโยง สเกลมีเหมาะสมกับข้อมูลเหล่านั้น

Yon (2006: บทคัดย่อ) เปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลแนวดิ่งด้วย โมเดล MIRT 2 วิธี ได้แก่ วิธี TCF และวิธี NOP สำหรับรูปแบบ NEAT ซึ่งใช้ข้อมูลจริงและข้อมูลจำลอง โดยศึกษากับโครงสร้างมิติความสามารถ 2 ลักษณะ คือ APSS และ CS ผลการวิจัยพบว่า ในการจำลอง ข้อมูล การเชื่อมโยงสเกลสำหรับโครงสร้าง APSS มีความถูกต้องมากกว่าโครงสร้าง CS และวิธี NOP ให้ผลการประมาณค่าพารามิเตอร์ความยากถูกต้องมากกว่าวิธี TCF เนื่องจากค่า RMSE ต่ำ แต่วิธี NOP ก็มีความลำเอียงสูงขึ้น เมื่อพิจารณาข้อมูลจริง วิธี NOP มีความคลาดเคลื่อนน้อยกว่าวิธี TCF และในภาพรวม การเชื่อมโยงสเกลแบบพหุมิติในแนวดิ่งสำหรับโครงสร้าง CS มีความถูกต้องน้อยกว่าโครงสร้าง APSS แต่ถ้าต้องการเชื่อมโยงสเกลสำหรับโครงสร้าง APSS ควรเลือกใช้วิธี NOP ผู้วิจัยให้ข้อเสนอแนะว่า ควรเปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลที่ลักษณะการหมุนแกนแตกต่างกัน

Wei (2008) เปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT 4 วิธี ได้แก่ (1) วิธี Direct (2) วิธี Equated Function (3) วิธี TCF และ (4) วิธี ICF สำหรับรูปแบบ NEAT โดย ใช้การจำลองข้อมูลประเมินประสิทธิภาพของวิธีการเชื่อมโยงสเกลเหล่านั้นภายใต้เงื่อนไขจำนวน 192 เงื่อนไข ซึ่งประกอบด้วย 4 ปัจจัย ได้แก่ โครงสร้างแบบสอบ ขนาดของกลุ่มตัวอย่าง ความยาวของแบบ สอบ การแจกแจงความสามารถของผู้สอบ โดยพิจารณาประสิทธิภาพดังกล่าวจากความแตกต่างระหว่าง ค่าประมาณพารามิเตอร์ของข้อสอบในกลุ่มฐานและค่าประมาณพารามิเตอร์ของข้อสอบที่ถูกแปลงแล้วใน กลุ่มปรับเทียบ (Diff)

ผลการวิจัย พบว่า การเชื่อมโยงสเกลทั้ง 4 วิธีสำหรับโครงสร้าง APSS มีประสิทธิภาพ ดีกว่าโครงสร้าง CS เนื่องจากการประมาณโครงสร้างแบบสอบและค่าพารามิเตอร์ของข้อสอบสำหรับ โครงสร้าง APSS มีความถูกต้องมากกว่าโครงสร้าง CS ในกรณีโครงสร้างที่ไม่ซับซ้อน วิธีการเชื่อมโยง สเกลทั้ง 4 วิธีมีประสิทธิภาพใกล้เคียงกัน แต่ในกรณีโครงสร้างแบบสอบซับซ้อน พบว่า (1) วิธี Equated Function ไม่มีประสิทธิภาพในการเชื่อมโยงสเกล (2) ประสิทธิภาพของวิธีการเชื่อมโยงทั้ง 3 วิธีที่เหลือ ขึ้นอยู่กับเงื่อนไขของการทดสอบ (3) วิธี Direct method มีประสิทธิภาพในการเชื่อมโยงสเกลมากที่สุด ผู้วิจัยให้ข้อเสนอแนะว่า (1) ควรเปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลทั้ง 4 วิธี ในกรณี เฉพาะ อย่างเช่น วิธี TCF และวิธี ICF ควรใช้ระดับความสามารถที่จุดตั้งต้นหรือระดับใดควรใช้ Equally spaced grid theta method หรือ Empirical theta method และถ้าเลือกใช้ Empirical theta method ควรใช้ผู้สอบกลุ่มฐาน กลุ่มปรับเทียบ หรือทั้งสองกลุ่ม และวิธีการใดดีที่สุด ส่วนวิธี Equated Function อาจศึกษาว่าควรใช้ค่าประมาณพารามิเตอร์ข้อสอบเท่าใด ควรใช้เกณฑ์ในการพิจารณาเลือกชุดข้อสอบที่

เหมาะสม และควรใช้ฟังก์ชันใดในการเชื่อมโยงสเกล (2) ควรศึกษาเกณฑ์การประเมินประสิทธิภาพของวิธีการเชื่อมโยงสเกลแต่ละวิธี (3) ควรประเมินประสิทธิภาพของโปรแกรมการประมาณค่าในโมเดล MIRT (4) ควรพัฒนาวิธีการเชื่อมโยงสเกล โดยอาจใช้ทั้งหลักการของโมเดลการตอบสนองข้อสอบแบบพหุมิติ และการวิเคราะห์องค์ประกอบผสมกัน

**ประเด็นที่สรุปได้จากงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า**

### 1) โมเดลการตอบสนองข้อสอบแบบพหุมิติ

โมเดลการตอบสนองข้อสอบแบบพหุมิติที่นิยมศึกษามากที่สุด คือ โมเดล MIRT แบบทดแทนได้ ข้อสอบที่ตรวจให้คะแนนสองค่านิยมใช้โมเดลแบบ 2 พารามิเตอร์มากกว่า 3 พารามิเตอร์

### 2) วิธีการเชื่อมโยงสเกล และการประมาณค่าพารามิเตอร์

งานวิจัยส่วนใหญ่ศึกษาประสิทธิภาพระหว่างวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT ขณะที่งานวิจัยบางส่วนเปรียบเทียบระหว่างวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันและการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธีการเชื่อมโยงสเกล ซึ่งการประมาณค่าพารามิเตอร์ใช้โปรแกรม NOHARM และ TESTFACT

### 3) รูปแบบการออกแบบการเก็บรวบรวมข้อมูล

งานวิจัยส่วนใหญ่ศึกษาภายใต้รูปแบบ NEAT และมีบางส่วนศึกษารูปแบบผู้สอบกลุ่มเท่าเทียม

### 4) ความยาวของแบบสอบและแบบสอบร่วม

ความยาวของแบบสอบที่ใช้ในงานวิจัยมีความหลากหลาย แต่ความยาวของแบบสอบที่น้อยที่สุด คือ 30 ข้อ ซึ่งสอดคล้องกับคำแนะนำของ Kolen และ Brennan (2004) ที่ระบุว่าควรใช้แบบสอบอย่างน้อย 30-40 ข้อ ข้อสอบร่วมที่ใช้เป็นข้อสอบร่วมภายใน สัดส่วนคะแนนข้อสอบร่วมที่นิยมใช้ คือ 20%, 25% และ 30% นอกจากนี้ยังใช้สัดส่วนคะแนนข้อสอบร่วม 15%

### 5) ตัวแปรที่ศึกษา

งานวิจัยเกี่ยวกับวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT ส่วนใหญ่มักเปรียบเทียบประสิทธิภาพของวิธีการเชื่อมโยงสเกลต่างๆ ร่วมกับตัวแปรอื่นๆ ไม่ว่าจะเป็น (1) ด้านกลุ่มตัวอย่าง: ขนาดของกลุ่มตัวอย่าง และความแตกต่างระหว่างกลุ่มตัวอย่าง/การแจกแจงความสามารถระหว่างผู้สอบ (2) ด้านข้อสอบร่วม: ความยาวของข้อสอบร่วม และโครงสร้างของชุดข้อสอบร่วม (3) ด้านแบบสอบ: ความยาวของแบบสอบ ความเท่าเทียมของโครงสร้างแบบสอบ ความเท่าเทียมของความยาก และความสัมพันธ์ระหว่างมิติ งานวิจัยส่วนใหญ่พิจารณาประสิทธิภาพของการวิธีการเชื่อมโยงสเกลจากค่าความลำเอียง (bias) และรากที่สองของความแตกต่างยกกำลังสอง (RMSE)

## 6) ผลการวิจัยโดยสรุป

งานวิจัยที่เปรียบเทียบประสิทธิภาพของการประมาณค่าพารามิเตอร์แบบพร้อมกัน และการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธีการเชื่อมโยงสเกล สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า พบว่า วิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันมีประสิทธิภาพมากกว่าการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธีการเชื่อมโยงสเกล

## 7) ข้อเสนอแนะสำหรับงานวิจัยครั้งต่อไป

งานวิจัยที่เปรียบเทียบประสิทธิภาพระหว่างการประมาณค่าพารามิเตอร์และแบบแยกกันศึกษาเฉพาะความสัมพันธ์ระหว่างมิติความสามารถระดับปานกลางไปจนถึงสูง จึงควรศึกษาในบริบทความสัมพันธ์ระหว่างมิติความสามารถต่ำ (ความเป็นพหุมิติสูง) งานวิจัยที่เปรียบเทียบประสิทธิภาพระหว่างวิธีการเชื่อมโยงสเกลระบุข้อเสนอแนะสำคัญ คือ ควรตรวจสอบความถูกต้องผลที่ได้จากการประมาณค่าพารามิเตอร์ด้วยโปรแกรมต่างๆ ควรพัฒนาวิธีการเชื่อมโยงสเกลหรือศึกษาวิธีการเชื่อมโยงสเกลอื่นๆนอกเหนือจากบริบทที่ศึกษา นอกจากนี้ยังแนะนำให้ศึกษาเกณฑ์ที่ใช้ในการประเมินประสิทธิภาพของวิธีการเชื่อมโยงสเกลแต่ละวิธี

**ตาราง 2.12** สรุปประเด็นของงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า

งานวิจัย	MIRT model	โปรแกรม Calibration	Design	ความยาวของแบบสอบ	ความยาวแบบสอบรวม
<b>เปรียบเทียบระหว่างประสิทธิภาพวิธีประมาณค่าพารามิเตอร์แบบพร้อมกันและแยกกัน</b>					
Lin (2008)	M2PL	BMIRT-พร้อมกัน NOHARM-แยกกัน	NEAT	60	20
Simon (2008)	M3PL	TESTFACT	EG, NEAT	60	20
<b>เปรียบเทียบประสิทธิภาพระหว่างวิธีการเชื่อมโยงสเกล</b>					
พัชรี จันทรเพ็ง (2550)	M3PL	TESTFACT	NEAT	100	30 (30%)
Hirsch (1989)	M2PL	MIRTE	ผู้สอบร่วม	40	N/A
Davey, et al., (1996)	M2PL	NOHARM	NEAT	N/A	40
Thompson et al., (1997)	M3PL	NOHARM	EG	200	N/A
Li (1997) Li & Lissitz (2000)	M2PL	TESTFACT	NEAT	N/A	15,25
Oshima et al., (2000)	M2PL	NOHARM	NEAT	40	N/A
Min (2007)	M2PL	NOHARM	NEAT	N/A	20
Reckase & Martineau (2004)	M2PL	TESTFACT	NEAT	50	25
Yon (2006)	M2PL	TESTFACT	NEAT	30	30
Wei (2008)	M2PL	NOHARM	NEAT	40/20	N/A

ตาราง 2.13 สรุปตัวแปรที่ศึกษาของงานวิจัยเกี่ยวกับการเชื่อมโยงสเกดด้วยโมเดล MIRT สำหรับ  
ข้อสอบที่ตรวจให้คะแนนสองค่า

ประเด็น	Lin	Simon	Hirsch	Davey et al.	Thompson et al.	Li & Lissitz	Oshima et al	Min	Reckase & Martineau	Yon	Wei
<b>ด้านแบบสอบ</b>											
- ลักษณะตั้งฉากเชิงโครงสร้าง	/										
- รูปแบบโครงสร้าง											/
- ความเท่าเทียมของโครงสร้าง	/										
- ความเท่าเทียมของความยาก	/										
- ความสัมพันธ์ระหว่างมิติ		/									
<b>ด้านแบบสอบรวม</b>											
- ประเภท											
- ความยาว						/					/
<b>ด้านรูปแบบการเชื่อมโยงสเกด</b>						/					
<b>ด้านกลุ่มตัวอย่าง</b>											
- ความแตกต่างของความสามารถ	/	/									/
- ขนาด		/				/					/
<b>ด้านการตรวจสอบมิติ</b>									/		
<b>ด้านกรวดมนมหาวิทยาลัย</b>											
- ขนาด		/				/					/
<b>ด้านการตรวจสอบมิติ</b>									/		

## ตอนที่ 4 สังเคราะห์งานวิจัยเกี่ยวกับการปรับเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล MIRT และกรอบแนวคิดการวิจัย

### 4.1 สังเคราะห์งานวิจัยเกี่ยวกับการปรับเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล MIRT

#### 4.1.1) ผลการวิจัยโดยสรุปของการปรับเทียบคะแนนด้วยโมเดล MIRT

กระบวนการปรับเทียบคะแนนด้วยโมเดล MIRT (FMIRT equating procedure) ประกอบด้วย กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูปแบบ (MOSE) กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณแบบเอกมิติ (AOSE) และกระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณแบบเอกมิติ (ATSE) ผลการวิจัยที่ผ่านมา พบว่า การปรับเทียบคะแนนข้อสอบที่ตรวจให้คะแนนสองค่า สำหรับผู้สอบกลุ่มเท่าเทียมกัน ด้วยกระบวนการ FMIRT ทั้ง 3 วิธี มีประสิทธิภาพมากกว่าการปรับเทียบด้วยโมเดล UIRT และกระบวนการทั้ง 3 วิธีให้ผลการปรับเทียบใกล้เคียงกัน (Bradley Grant Brossman, 2010)

การปรับเทียบคะแนนข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT พบว่า กระบวนการ AOSE มีประสิทธิภาพมากที่สุด โดยงานวิจัยนี้มีการเชื่อมโยงสเกลระหว่างแบบสอบและเปรียบเทียบประสิทธิภาพการปรับเทียบคะแนนภายใต้วิธีการเชื่อมโยงสเกลต่าง ๆ พบว่า กระบวนการปรับเทียบคะแนนด้วยโมเดล MIRT ภายใต้วิธี TCF และวิธี ICF มีความคลาดเคลื่อนต่ำกว่าวิธี NOP และวิธี OD (O. Zhang, 2012) อย่างไรก็ตามเมื่อพิจารณาการปรับเทียบคะแนนข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบกลุ่มสุ่ม พบว่า กระบวนการ MOSE มีประสิทธิภาพมากที่สุด (E. Lee, 2013) สอดคล้องกับการปรับเทียบคะแนนแบบสอบรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่ม ที่เลือกใช้ กระบวนการ MOSE เพื่อเปรียบเทียบประสิทธิภาพกับกระบวนการ UIRT และวิธี Bi-factor พบว่า กระบวนการ MOSE มีประสิทธิภาพมากที่สุด (Peterson, 2014) ดังนั้นกระบวนการ MOSE จึงมีความเหมาะสมที่จะนำมาปรับเทียบคะแนนแบบสอบพหุมิติ

#### 4.1.2) ข้อจำกัดและข้อเสนอแนะของงานวิจัยการปรับเทียบคะแนนด้วยโมเดล MIRT

1. การปรับเทียบคะแนนด้วยโมเดล MIRT ทั้งหมดยกเว้นงานวิจัยของ Brossman (2010) มีการประมาณค่าพารามิเตอร์แบบแยกกัน แต่เมื่อพิจารณางานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT และ UIRT พบว่า การประมาณค่าพารามิเตอร์แบบพร้อมกันมีความคลาดเคลื่อนต่ำกว่าการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธีการเชื่อมโยงต่าง ๆ ดังนั้นผู้วิจัยจึงสนใจศึกษาการปรับเทียบคะแนนด้วยโมเดล MIRT เมื่อมีการประมาณค่าพารามิเตอร์แบบพร้อมกัน

2. การปรับเทียบคะแนนด้วยโมเดล MIRT มีการศึกษาข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่ม รูปแบบ NEAT และรูปแบบผู้สอบกลุ่มเท่าเทียม และศึกษาแบบสอบรูปแบบ



ผสม ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม แต่ยังไม่มีการปรับเทียบคะแนนแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT ผู้วิจัยจึงสนใจศึกษาในประเด็นดังกล่าว

3. ตัวแปรอิสระที่ศึกษามีเฉพาะตัวแปรด้านแบบสอบและผู้สอบ แต่ยังไม่มีการศึกษาด้านข้อสอบร่วม/แบบสอบร่วม สำหรับรูปแบบ NEAT และเมื่อพิจารณางานวิจัยเกี่ยวกับการเชื่อมโยงสเกลภายใต้รูปแบบ NEAT พบว่า มีการศึกษาตัวแปรด้านข้อสอบร่วม ประเด็นสำคัญที่ควรพิจารณา คือ สัดส่วนข้อสอบร่วม และสัดส่วนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่า

4. โมเดล MIRT สำหรับข้อสอบที่ตรวจให้คะแนนหลายค่าที่ใช้ในงานวิจัยที่ผ่านมา คือ โมเดล Multidimensional grade-response model (MGR) แต่ยังไม่มีการศึกษาในโมเดลอื่น อย่างเช่น โมเดล Multidimensional generalized partial credit model (MGPC) ซึ่งโมเดลนี้ยังไม่มีการศึกษาการเชื่อมโยงสเกลแบบสอบพหุมิติรูปแบบผสมเช่นกัน

#### 4.1.3) ผลการวิจัยโดยสรุปของการเชื่อมโยงสเกลด้วยโมเดล MIRT

##### 1. การเปรียบเทียบประสิทธิภาพระหว่างการประมาณค่าพารามิเตอร์แบบพร้อมกันและการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธีการเชื่อมโยงสเกล

งานวิจัยในประเด็นดังกล่าว พบว่า การเชื่อมโยงสเกลข้อสอบที่ตรวจให้คะแนนสองค่าด้วยการประมาณค่าพารามิเตอร์แบบพร้อมกันมีประสิทธิภาพมากกว่าการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธีการเชื่อมโยงสเกล (Lin, 2008; Simon, 2008) สอดคล้องกับผลการวิจัยสำหรับแบบสอบรูปแบบผสมที่พบว่า ข้อมูลจากการประมาณค่าพารามิเตอร์มีสอดคล้องกับโมเดล (Patz & Yao, 2007)

##### 2. การเปรียบเทียบประสิทธิภาพระหว่างวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT

งานวิจัยที่เปรียบเทียบประสิทธิภาพระหว่างวิธีการเชื่อมโยงสเกลต่าง ๆ มีความหลากหลาย และผลการวิจัยมีความแตกต่างกันไปขึ้นอยู่กับบริบทที่ศึกษา เนื่องจากผลการวิจัยของการปรับเทียบคะแนนด้วยโมเดล MIRT สำหรับรูปแบบ NEAT พบว่า กระบวนการปรับเทียบคะแนนด้วยโมเดล MIRT ภายใต้วิธี TCF และวิธี ICF มีความคลาดเคลื่อนต่ำกว่าวิธี NOP และวิธี OD (O. Zhang, 2012) ดังนั้นผู้วิจัยจึงให้ความสนใจไปที่วิธี TCF และ ICF เป็นหลัก ซึ่งงานวิจัยที่ผ่านมา พบว่า การเชื่อมโยงสเกลสำหรับแบบสอบรูปแบบผสมด้วยวิธี TCF ภายใต้รูปแบบ NEAT ให้ผลการเชื่อมโยงที่ดี (Yao & Boughton, 2009; ศิวทัศน์ สุขสุวรรณ, พัชรี จันทร์เพ็ง และสังวรณ์ รัตกระโทก, 2555) ด้วยเหตุนี้วิธี TCF จึงมีความเหมาะสมที่จะนำมาใช้เชื่อมโยงสเกลแบบสอบพหุมิติรูปแบบผสม ภายใต้รูปแบบ NEAT

#### 4.1.4) ข้อจำกัดและข้อเสนอแนะของงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT

การเชื่อมโยงสเกลมีการเปรียบเทียบการประมาณค่าพารามิเตอร์แบบพร้อมกันและการประมาณค่าพารามิเตอร์แบบแยกกันด้วยการเชื่อมโยงสเกล สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า แต่ยังไม่มีการศึกษาในประเด็นดังกล่าว สำหรับแบบสอบรูปแบบผสม

#### 4.1.5) ประเด็นที่ควรพิจารณาในการปรับเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล MIRT สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT

##### 1. โครงสร้างมิตិความสามารถ

งานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT ส่วนใหญ่ศึกษาโครงสร้างมิติความสามารถ 2 แบบ คือ โครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างซับซ้อน (CS) แบบ 2 มิติ (E. Lee, 2013; Peterson, 2014; O. Zhang, 2012) ส่วนโครงสร้างอย่างง่าย (SS) ไม่นิยมศึกษามากนัก เนื่องจากเป็นไปได้ยากที่ข้อสอบข้อหนึ่งจะอยู่ในมิติใดมิติหนึ่งเพียงมิติเดียว (Jinming Zhang & Stout, 1999) โดยงานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT ศึกษาในโครงสร้าง SS เพียงงานเดียว ซึ่งผู้วิจัยระบุว่าเป็นการวิจัยนำร่องเพื่อพัฒนาการปรับเทียบคะแนนสำหรับแบบสอบพหุมิติ (W.-C. Lee & Brossman, 2012) ดังนั้นผู้วิจัยจึงสนใจศึกษาโครงสร้างมิติความสามารถแบบ APSS และ CS แบบ 2 มิติ

##### 2. ความสัมพันธ์ระหว่างมิติความสามารถ

งานวิจัยเกี่ยวกับการปรับเทียบคะแนนด้วยโมเดล MIRT มีการศึกษาตัวแปรความสัมพันธ์ระหว่างมิติความสามารถ พบว่า เมื่อความสัมพันธ์ระหว่างมิติความสามารถสูงขึ้น การปรับเทียบคะแนนมีค่าความลำเอียงและความคลาดเคลื่อนมาตรฐานเพิ่มขึ้น (E. Lee, 2013; W.-C. Lee & Brossman, 2012) สอดคล้องกับผลการวิจัยการเชื่อมโยงสเกลข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT พบว่า เมื่อความสัมพันธ์ระหว่างมิติสูงขึ้น ค่า RMSE ของค่าพารามิเตอร์อำนาจจำแนกเพิ่มขึ้นในทุกเงื่อนไขของขนาดกลุ่มตัวอย่าง และค่าความลำเอียงที่ได้จากการประมาณค่าพารามิเตอร์ยิ่งเพิ่มขึ้น (Lin, 2008; Simon, 2008) งานวิจัยที่ผ่านมาให้ผลการวิจัยไปในทิศทางเดียวกัน กล่าวคือ เมื่อความสัมพันธ์ระหว่างมิติสูงขึ้น ประสิทธิภาพของการเชื่อมโยงสเกลและการปรับเทียบคะแนนลดลง จึงไม่จำเป็นต้องศึกษาในประเด็นดังกล่าว โดยงานวิจัยที่ศึกษาตัวแปรดังกล่าวมีการกำหนดเป็นความสัมพันธ์ระดับสูง กลาง ต่ำ เช่น .8, .5 และ 0 แต่เมื่อพิจารณางานวิจัยที่ใช้ข้อมูลจากการทดสอบจริง พบว่า แบบสอบส่วนใหญ่มักมีความสัมพันธ์ระหว่างมิติความสามารถตั้งแต่ปานกลางไปจนถึงสูง (Bradley Grant Brossman, 2010) และเนื่องจากผู้วิจัยไม่ได้สนใจศึกษาตัวแปรนี้ จึงกำหนดให้ความสัมพันธ์ระหว่างมิติความสามารถอยู่ในระดับปานกลาง คือ .5

##### 3. ความยาวของข้อสอบแต่ละประเภทในแบบสอบรูปแบบผสม

เมื่อพิจารณาการทดสอบขนาดใหญ่ในบริบทของประเทศไทย การทดสอบ CU-TAD ประกอบด้วยข้อสอบที่ตรวจให้คะแนนสองค่าจำนวน 30 ข้อ และข้อสอบที่ตรวจให้คะแนนหลายค่า 6 ข้อ คะแนนเต็ม 70 คะแนน คิดเป็นสัดส่วนข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าเป็น 30:70 คะแนน ส่วนการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) ระดับมัธยมศึกษาชั้นปีที่ 6 ประจำปีการศึกษา 2559 มีทั้งข้อสอบปรนัย 5 ตัวเลือก 1 คำตอบ (0, 1 คะแนน) และข้อสอบปรนัย 5 ตัวเลือก 2 คำตอบและหลายคำตอบ (0, 2 คะแนน) จึงไม่ถือว่ามิติข้อสอบที่ตรวจให้คะแนนหลายค่าผสมอยู่ เมื่อพิจารณาการทดสอบจริงในบริบทของต่างประเทศ อย่างเช่น การทดสอบภาษาอังกฤษ TOEFL ประกอบด้วย 4 ส่วน คือ

การอ่าน (ข้อละ 1 คะแนน 30 ข้อ) การฟัง (ข้อละ 1 คะแนน 30 ข้อ) การพูด 6 ข้อ (คะแนนตั้งแต่ 0-4 คะแนนเต็ม 24 คะแนน แต่ปรับสเกลเป็น 30 คะแนน) การเขียน (คะแนนตั้งแต่ 0-5 คะแนนเต็ม 30 คะแนน) นั่นคือ สัดส่วนคะแนนที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าเป็น 50:50 คะแนน

เมื่อพิจารณาการเปรียบเทียบคะแนนแบบสอบพหุมิติรูปแบบผสมวิชาประวัติศาสตร์ และ วิชาภาษาอังกฤษจากการทดสอบจริง (Advance place exam) ใช้สัดส่วนคะแนนข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าเป็น 70%:30% (W.-C. Lee & Brossman, 2012; Peterson, 2014) ขณะที่การเชื่อมโยงสเกลด้วยโมเดล MIRT โดยใช่การจำลองข้อมูลส่วนใหญ่กำหนดสัดส่วนคะแนนข้อสอบดังกล่าวเป็น 70%:30% (Patz & Yao, 2007; ศิวะทัศน์ สุขสุวรรณ, 2555) และเนื่องจากการทดสอบภายในประเทศ เช่น CU-TAD และ O-NET กำหนดคะแนนเต็มเป็น 100 คะแนน ด้วยเหตุนี้ผู้วิจัยจึงกำหนดสัดส่วนคะแนนข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าของแบบสอบเป็น 70%:30% คะแนนเต็มของแบบสอบ คือ 100 คะแนน โดยคะแนนเต็มของข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่า คือ 70 และ 30 คะแนนตามลำดับ ผู้วิจัยกำหนดคะแนนเต็มข้อสอบที่ตรวจให้คะแนนหลายค่าเป็น 2 คะแนน เพื่อให้สามารถศึกษาสัดส่วนข้อสอบรวมแต่ละประเภทได้ตามเงื่อนไขที่กำหนดไว้มากที่สุด

#### 4. สัดส่วนข้อสอบรวม

การสร้างแบบสอบแต่ละฉบับต้องใช้เวลาและค่าใช้จ่าย จึงต้องกำหนดความยาวข้อสอบรวมที่เหมาะสม โดยสัดส่วนข้อสอบรวมที่มีประสิทธิภาพในการเปรียบเทียบคะแนนสำหรับกลุ่มตัวอย่าง 400 คน คือ 10% และ 20% และถ้าข้อสอบรวมมีคุณภาพสูง สัดส่วนข้อสอบรวม 10% ก็มีความเพียงพอ (Gao, 2004; McKinley & Reckase, 1981) และเมื่อพิจารณาการเชื่อมโยงสเกลด้วย MIRT พบว่า สัดส่วนคะแนนข้อสอบรวมที่เหมาะสมสำหรับการเชื่อมโยงสเกลแบบสอบพหุมิติรูปแบบผสม ที่มีโครงสร้างมิติตามความสามารถแบบซับซ้อน ด้วยวิธี TCF คือ 25% (ศิวะทัศน์ สุขสุวรรณ, พิชรี จันทรพิง และสังวรรณ ังคกระโทก, 2555) นอกจากนี้ยังมีงานวิจัยที่ระบุสัดส่วนข้อสอบรวมที่น้อยที่สุดที่ทำให้การเปรียบเทียบคะแนนข้อสอบที่ตรวจให้คะแนนสองค่ามีประสิทธิภาพ เมื่อความยาวของแบบสอบเป็น 160, 120, 100 และ 80 ข้อ คือ 20% และเมื่อความยาวแบบสอบเป็น 60 และ 40 ข้อ สัดส่วนข้อสอบรวมที่มีประสิทธิภาพ คือ 25% และ 33% ตามลำดับ (Xiong et al., 2013) อย่างไรก็ตามการเปรียบเทียบคะแนนด้วยโมเดล MIRT ภายใต้รูปแบบ NEAT ใช้สัดส่วนคะแนนข้อสอบรวม 50% ซึ่งถือว่าเป็นสัดส่วนจำนวนมาก อีกทั้งยังไม่มีศึกษาเกี่ยวกับสัดส่วนข้อสอบรวม ด้วยเหตุนี้ผู้วิจัยจึงสนใจศึกษาความยาวข้อสอบรวมที่สั้นลงเพื่อเพิ่มทางเลือกให้แก่ผู้สร้างแบบสอบพหุมิติตัวด้วยสัดส่วนข้อสอบรวม 3 แบบ คือ 10%, 20% และ 30% ตามลำดับ

#### 5. อัตราส่วนข้อสอบรวมที่ตรวจให้คะแนนสองค่า

การเปรียบเทียบคะแนนและการเชื่อมโยงสเกลแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT ควรใช้ข้อสอบรวมรูปแบบผสม เพื่อลดข้อจำกัดของข้อสอบแต่ละประเภท การเชื่อมโยงสเกลแบบสอบพหุมิติรูปแบบผสมมีการศึกษาอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40, 70:30

และ 80:20 (Patz & Yao, 2007; Yao & Boughton, 2009; ศิวทัศน์ สุขสุวรรณ, 2555; ศิวทัศน์ สุขสุวรรณ, พัทธี จันทร์เพ็ง และสังวรณัจฉกรระโทก, 2555) และการเชื่อมโยงสเกลด้วยโมเดล MIRT ส่วนใหญ่นิยมใช้สัดส่วนคะแนนข้อสอบร่วมดังกล่าวเป็น 70:30 แต่ไม่มีการศึกษาในประเด็นดังกล่าวสำหรับการเปรียบเทียบคะแนนด้วยโมเดล MIRT

งานวิจัยเกี่ยวกับอัตราส่วนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า พบว่า การเชื่อมโยงสเกลด้วยวิธีเชิงเส้นมีความคลาดเคลื่อนลดลง เมื่อใช้แบบสอบร่วมรูปแบบผสมและข้อสอบที่ตรวจให้คะแนนสองค่าลดลง (อัญชลิ ศรีกลชาญ, 2552) สอดคล้องกับการเชื่อมโยงสเกลด้วยโมเดล IRT ที่พบว่า ค่า MSE ลดลงเมื่อข้อสอบร่วมแบบสร้างคำตอบเพิ่มขึ้น (Kim, 2004 อ้างอิงในอัญชลิ ศรีกลชาญ, 2552) และเนื่องจากงานวิจัยการเชื่อมโยงสเกล MIRT ศึกษาเฉพาะอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่ามากกว่าตรวจให้คะแนนหลายค่า ผู้วิจัยจึงสนใจศึกษาอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนหลายค่ามากกว่าตรวจให้คะแนนสองค่าร่วมด้วย ดังนั้นอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าที่สนใจศึกษา 4 แบบ คือ 40:60, 50:50, 60:40 และ 70:30

#### 6. วิธีการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์

วิธีการปรับเทียบคะแนนที่ใช้เป็นเกณฑ์ที่นิยมมากที่สุด คือ วิธีอีควิเปอร์เซ็นไทล์ที่มีการปรับโค้งให้ราบเรียบก่อน (pre-smoothed equipercentile method) รองลงมาคือ วิธี Identity equating และ วิธีการประมาณความถี่ (frequency estimation) โดยวิธีอีควิเปอร์เซ็นไทล์ที่ปรับโค้งให้ราบเรียบไม่จำเป็นต้องคำนึงถึงมิติของข้อมูล และผลการปรับเทียบคะแนนด้วยวิธีดังกล่าวยังสอดคล้อง (align) อย่างมากกับวิธีการปรับเทียบคะแนนด้วยโมเดล MIRT เนื่องจากลดค่า likelihood ที่จะทำให้เกิดความลำเอียง นอกจากนี้วิธีอีควิเปอร์เซ็นไทล์ดังกล่าวยังมีความคล้ายคลึงกับวิธี IRT เนื่องจากสหสัมพันธ์ระหว่างการปรับเทียบคะแนนเป็นเส้นโค้ง ส่วนรูปแบบผู้สอบกลุ่มสุ่ม ผู้วิจัยเลือกใช้วิธี Identity equating เป็นเกณฑ์ (Peterson, 2014) ด้วยเหตุนี้ผู้วิจัยจึงเลือกใช้วิธีอีควิเปอร์เซ็นไทล์ที่ปรับโค้งให้ราบเรียบเป็นเกณฑ์ในการเปรียบเทียบประสิทธิภาพ

#### 7. ดัชนีสำหรับการประเมินประสิทธิภาพของการปรับเทียบคะแนน

ดัชนีที่ใช้เปรียบเทียบประสิทธิภาพของการปรับเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล MIRT มีหลากหลาย โดยดัชนีที่นิยมใช้ คือ ค่าความลำเอียง (bias) และความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน (SE) และพิจารณาประสิทธิภาพในภาพรวมของค่าความลำเอียงและความคลาดเคลื่อนมาตรฐานจากรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ด้วยเหตุนี้ผู้วิจัยจึงพิจารณาประสิทธิภาพของการปรับเทียบคะแนนจากค่าความลำเอียงสัมพัทธ์ สัมประสิทธิ์การแปรผันของความคลาดเคลื่อนมาตรฐาน (CSVE) และค่า RMSE

## 8. จำนวนการทำซ้ำ

เมื่อพิจารณาจำนวนการทำซ้ำในการจำลองข้อมูลที่ผ่านมา พบว่า จำนวนการทำซ้ำมากที่สุด คือ 500 ครั้ง รองลงมา คือ 200, 100, 50, 25, 20 และ 10 ครั้งตามลำดับ โดยจำนวนการทำซ้ำเฉลี่ย คือ 132 ครั้ง โดยจำนวนการทำซ้ำที่นิยมใช้ คือ 20 ครั้ง (3 งานวิจัย) รองลงมา คือ 50 และ 100 ครั้ง (อย่างละ 2 งานวิจัย) แต่อย่างไรก็ตาม Harwell, Hsu และ Kirisci (1996) อ้างอิงใน พัชรี จันทรเพ็ญ, (2550) พบว่า การศึกษาโดยใช้โมเดล IRT เป็นฐาน ควรมีการทำซ้ำอย่างน้อย 20 ครั้ง รายละเอียดดังตาราง 2.15 แต่อย่างไรก็ตามยังไม่มีการระบุจำนวนการทำซ้ำที่เหมาะสมสำหรับการเปรียบเทียบคะแนนไว้อย่างชัดเจน เนื่องจากการวิจัยครั้งนี้ต้องใช้เวลาจำลองข้อมูล 22 เดือนไซ ซึ่งใช้เวลาในการดำเนินการแต่ละรอบประมาณ 1 ชั่วโมงเป็นอย่างต่ำ ผู้วิจัยจึงสามารถกระทำซ้ำข้อมูลได้ 40 เดือนไซในแต่ละรอบ

ตาราง 2.14 จำนวนการทำซ้ำในงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลและปรับเทียบคะแนนด้วยโมเดล MIRT

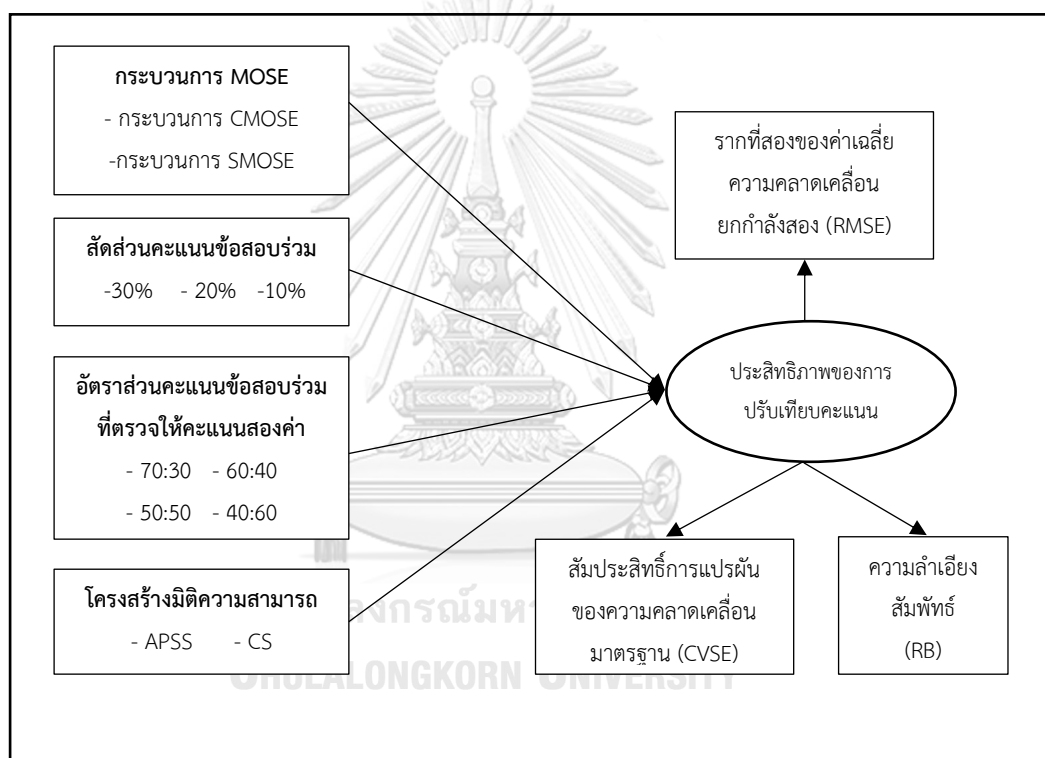
ผู้วิจัย (ปี)	จำนวนการทำซ้ำ (ครั้ง)
การเชื่อมโยงสเกลด้วยโมเดล MIRT	
Yon (2006)	10
Oshima, Davey และ Lee (2000); Yao และ Boughton (2009); พัชรี จันทรเพ็ญ (2550), ศิวะทัศน์ สุขสุวรรณ (2555)	20
Min (2003); Min (2007)	25
Simon (2008)	50
Li และ Lisztz (2000)	100
Li (1997)	200
Yao (2008); Wei (2008)	500
การปรับเทียบคะแนนด้วยโมเดล MIRT	
Lee, Lee & Brennan (2014)	50
Zhang (2012)	200
ค่าเฉลี่ย	132

## 9. ขนาดตัวอย่าง

การปรับเทียบคะแนนด้วยโมเดล MIRT ส่วนใหญ่ใช้กลุ่มตัวอย่างขนาดใหญ่มากกว่า 2,000 คน ขนาดกลุ่มตัวอย่างแต่ละกลุ่มที่น้อยที่สุด คือ จำนวน 2,000 คน (O. Zhang, 2012) ซึ่งสอดคล้องกับแนวคิดของการปรับเทียบคะแนนตามโมเดล IRT ที่ระบุไว้ว่าการปรับเทียบคะแนนด้วยโมเดล IRT ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ (Bradley Grant Brossman, 2010) ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยใช้ขนาดตัวอย่างกลุ่มละ 3,000 คน สำหรับแบบสอบแต่ละฉบับ

## 4.2 กรอบแนวคิดการวิจัย

จากผลการวิจัยของการเปรียบเทียบคะแนนและการเชื่อมโยงสเกลด้วยโมเดล MIRT รวมถึงข้อเสนอแนะและข้อจำกัดของงานวิจัยเหล่านั้น ผู้วิจัยจึงสนใจเปรียบเทียบประสิทธิภาพกระบวนการ MOSE 2 แบบ คือ กระบวนการ MOSE ที่มีการประมาณค่าพารามิเตอร์แบบพร้อมกัน (CMOSE) และกระบวนการ MOSE ที่มีการประมาณค่าพารามิเตอร์แบบแยกกัน และใช้การเชื่อมโยงสเกลด้วยวิธี TCF (SMOSE) สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วมเมื่อโครงสร้างมิตีความสามารถ สัดส่วนคะแนนข้อสอบร่วม และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแตกต่างกัน ด้วยการจำลองข้อมูลแบบมอนติคาร์โล กรอบแนวคิดการวิจัยเป็นดังนี้



ภาพ 2.13 กรอบแนวคิดในการวิจัย

### บทที่ 3

#### วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของกระบวนการ MOSE 2 แบบ ได้แก่ กระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบพร้อมกัน (CMOSE) และกระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบแยกกัน และใช้การเชื่อมโยงสเกลด้วยวิธี TCF (SMOSE) สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม (NEAT) ซึ่งศึกษาในเงื่อนไขต่าง ๆ ได้แก่ โครงสร้างมิติความสามารถ 2 แบบ สัดส่วนคะแนนข้อสอบร่วม 3 แบบ และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า 4 แบบ โดยใช้การจำลองข้อมูลแบบมอนติคาร์โลด้วยโมเดลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ โดยโมเดลสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า คือ โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (M3PL) และโมเดลสำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า คือ Multidimensional generalize partial credit (MGPC) ประสิทธิภาพของการปรับเทียบคะแนนพิจารณาจาก ค่าความลำเอียงสัมพัทธ์ (RB) สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (CVSE) และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) และการนำเสนอวิธีดำเนินการวิจัยโดยละเอียด แบ่งออกเป็น 5 ขั้นตอน ดังนี้

ตอนที่ 1 การจำลองข้อมูล

ตอนที่ 2 การตรวจสอบความถูกต้องของการจำลองข้อมูล

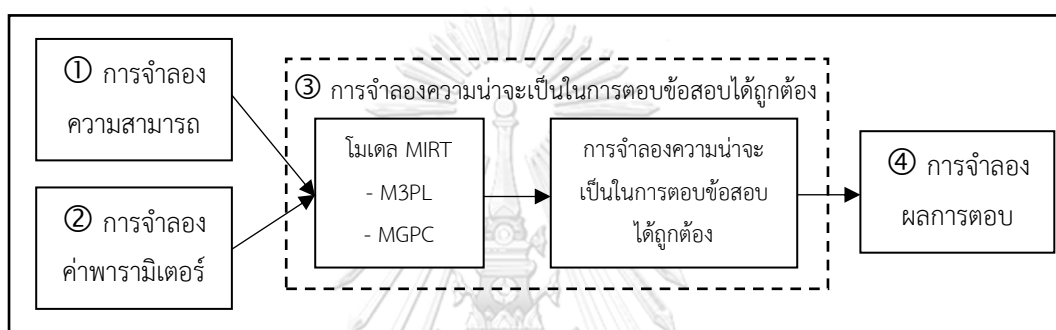
ตอนที่ 3 การวัดประสิทธิภาพของการปรับเทียบคะแนน

ตอนที่ 4 การวิเคราะห์ประสิทธิภาพของการปรับเทียบคะแนน

ตอนที่ 5 ขั้นตอนการดำเนินงานวิจัย

## ตอนที่ 1 การจำลองข้อมูล

การจำลองข้อมูลการตอบข้อสอบของผู้สอบ เริ่มจาก การจำลองความสามารถของผู้สอบ และการจำลองค่าพารามิเตอร์ของข้อสอบ หลังจากนั้นจึงค่าความสามารถและค่าพารามิเตอร์ที่จำลองได้แทนในโมเดลการตอบสนองข้อสอบแบบพหุมิติสำหรับข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่า ซึ่งก็คือ โมเดล M3PL และ MGPC ตามลำดับ ซึ่งจะได้ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง สำหรับข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่าตามโมเดล M3PL และ MGPC ตามลำดับ แล้วใช้ความน่าจะเป็นดังกล่าวเป็นตัวกำหนดผลการตอบข้อสอบ โดยผลการตอบข้อสอบที่ตรวจให้คะแนนสองค่า คือ 0, 1 คะแนน และผลการตอบข้อสอบที่ตรวจให้คะแนนหลายค่า คือ 0, 1 และ 2 ดังภาพ 3.1



ภาพ 3.1 ขั้นตอนการจำลองข้อมูล

ในการนำเสนอเงื่อนไขการจำลองข้อมูลจึงแบ่งเป็น 4 ส่วนตามภาพ 3.1 ได้แก่ (1) การจำลองค่าความสามารถของผู้สอบ (2) การจำลองค่าพารามิเตอร์ของข้อสอบ (3) การจำลองความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง และ (4) การจำลองแบบแผนการตอบ

### 1.1 การจำลองค่าความสามารถของผู้สอบ

การจำลองค่าความสามารถกลุ่มละ 3,000 คน โดยลักษณะการแจกแจงความสามารถของผู้สอบทั้งสองกลุ่มเป็นการแจกแจงแบบปกติ 2 ตัวแปร (bivariate normal distribution) ประกอบด้วยค่าเฉลี่ยและค่าความแปรปรวน-ความแปรปรวนร่วม งานวิจัยนี้ถือว่าแบบสอบแต่ละฉบับใช้สำหรับผู้สอบแต่ละกลุ่มที่มีความแตกต่างกัน โดยผู้วิจัยกำหนดเวกเตอร์ค่าเฉลี่ยและเมทริกซ์ค่าความแปรปรวน-ความแปรปรวนร่วมของกลุ่มหนึ่ง คือ  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$  และ  $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$  ขณะที่เวกเตอร์ค่าเฉลี่ยและเมทริกซ์ค่าความแปรปรวน-ความแปรปรวนร่วมของกลุ่มสอง คือ  $\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$  และ  $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$  ดังนั้นความสามารถของผู้สอบแต่ละกลุ่มจึงสร้างจากข้อมูลที่มีการแจกแจงปกติแบบสองตัวแปร คือ  $N(\mu_g, \sigma_g)$  เมื่อ  $g$  คือ กลุ่มผู้สอบ ในที่นี้ คือ กลุ่ม 1 และกลุ่ม 2 โดยแต่ละกลุ่มมีระดับความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ .5 ดังตาราง 3.1 ซึ่งผู้วิจัยดำเนินการผ่าน Package “mvtnorm” ในโปรแกรม R ด้วยฟังก์ชัน pmvnorm



ตาราง 3.1 การกระจายค่าความสามารถและความสัมพันธ์ระหว่างมิติสำหรับผู้สอบแต่ละกลุ่ม

ผู้สอบ	เวกเตอร์ ค่าเฉลี่ย ( $\mu$ )	ความแปรปรวน-ความ แปรปรวนร่วม ( $\Sigma$ )	ความสัมพันธ์ ระหว่างมิติความสามารถ
กลุ่ม 1	$\begin{bmatrix} 0 \\ 0 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	$r = 0.50$
กลุ่ม 2	$\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$	$\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$	

## 1.2 การจำลองค่าพารามิเตอร์ของข้อสอบ

ในนำเสนอการจำลองค่าพารามิเตอร์ของข้อสอบแบ่งเป็น 2 ส่วนได้แก่ (1) สัดส่วนคะแนนข้อสอบรวมและอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า และ (2) การจำลองค่าพารามิเตอร์ของข้อสอบ

### 1.2.1) สัดส่วนคะแนนข้อสอบรวมและอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า

งานวิจัยครั้งเปรียบเทียบประสิทธิภาพของกระบวนการ MOSE 2 รูปแบบ ภายใต้เงื่อนไข ได้แก่ ได้แก่ โครงสร้างมิติความสามารถ 2 แบบ ได้แก่ โครงสร้างอย่างง่าย และโครงสร้างที่ซับซ้อน สัดส่วนคะแนนข้อสอบรวม 3 แบบ ได้แก่ 10%, 20% และ 30% และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 4 แบบ ได้แก่ 40:60, 50:50, 60:40 และ 70:30 ซึ่งแบบสอบรูปแบบผสม ประกอบด้วย ข้อสอบที่ตรวจให้คะแนนสองค่า (0 และ 1 คะแนน) และข้อสอบที่ตรวจให้คะแนนหลายค่า (0, 1 และ 2 คะแนน) จำนวน 15 ข้อ คะแนนเต็ม 30 คะแนน (ข้อละ 2 คะแนน) โดยแบบสอบทั้งฉบับมีสัดส่วนคะแนนข้อสอบที่ตรวจให้คะแนนสองค่า คือ 70%:30% อย่างไรก็ตามมีเงื่อนไขอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าที่ไม่สามารถกำหนดจำนวนข้อสอบและคะแนนได้ตรงตามเงื่อนไขพอดี ดังนี้

1. เงื่อนไขอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 50:50 สำหรับสัดส่วนคะแนนข้อสอบรวม 30% ได้คะแนนข้อสอบรวมแต่ละประเภทเป็น 15 คะแนน : 15 คะแนน ซึ่งจำนวนข้อสอบที่ตรวจให้คะแนนหลายค่าไม่ลงตัว จึงกำหนดเป็นคะแนนที่ใกล้เคียง คือ 14 คะแนน : 16 คะแนน

2. เงื่อนไขอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30 สำหรับสัดส่วนคะแนนข้อสอบรวม 30% จะได้คะแนนข้อสอบรวมแต่ละประเภทเป็น 21 คะแนน : 9 คะแนน ซึ่งจำนวนข้อสอบที่ตรวจให้คะแนนหลายค่าไม่ลงตัว จึงกำหนดเป็นคะแนนที่ใกล้เคียง คือ 22 คะแนน : 8 คะแนน

3. เงื่อนไขอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 50:50 สำหรับสัดส่วนคะแนนข้อสอบรวม 10% จะได้คะแนนข้อสอบรวมแต่ละประเภทเป็น 5 คะแนน : 5 คะแนน ซึ่งจำนวนข้อสอบที่ตรวจให้คะแนนหลายค่าไม่ลงตัว เมื่อกำหนดเป็นคะแนนที่ใกล้เคียง คือ 4 คะแนน : 6 คะแนน จะไปซ้ำซ้อนกับสัดส่วน 60:40 จึงไม่ได้ศึกษาในเงื่อนไขดังกล่าว รายละเอียดดังตาราง 3.2

แบบสอบรูปแบบผสมที่ใช้สำหรับการเปรียบเทียบคะแนน ภายใต้รูปแบบ NEAT มี 2 ฉบับ แบบสอบแต่ละฉบับมีคะแนนเต็ม 100 คะแนน ประกอบด้วย คะแนนเต็มของข้อสอบที่ตรวจให้ คะแนนสองค่าและตรวจให้คะแนนหลายค่า คือ 70 และ 30 คะแนนตามลำดับ โดยแบบสอบทั้งสองฉบับ มีข้อสอบร่วมกัน ดังนั้นแบบสอบที่ใช้ในการวิจัยครั้งนี้จึงประกอบด้วย 3 ส่วนหลัก คือ (1) ข้อสอบเฉพาะสำหรับแบบสอบ X (2) ข้อสอบเฉพาะสำหรับแบบสอบ Y และ (3) ข้อสอบร่วม เนื่องจากสัดส่วนคะแนน ข้อสอบร่วมแตกต่างกันตามเงื่อนไขที่ศึกษา ผู้วิจัยจึงกำหนดรายละเอียดของคะแนนและจำนวนข้อสอบ ร่วมในแต่ละกลุ่มก่อน จากนั้นจึงกำหนดรายละเอียดดังกล่าวสำหรับข้อสอบเฉพาะให้สอดคล้องกับ สัดส่วนข้อสอบร่วมในแต่ละเงื่อนไข

ตาราง 3.2 อัตราส่วนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแยกตามสัดส่วนคะแนนข้อสอบร่วม

สัดส่วน ข้อสอบร่วม	อัตราส่วนคะแนนข้อสอบร่วม ที่ตรวจให้คะแนนสองค่า	คะแนน (ข้อ)		
		DICHOTOMOUS	POLYTOMOUS	รวม
10% (10 คะแนน)	40:60 (4 คะแนน:6 คะแนน)	4 คะแนน (4 ข้อ)	6 คะแนน (3 ข้อ)	10 คะแนน (7 ข้อ)
	60:40 (6 คะแนน:4 คะแนน)	6 คะแนน (6 ข้อ)	4 คะแนน (2 ข้อ)	10 คะแนน (8 ข้อ)
	70:30 (8 คะแนน:2 คะแนน)	8 คะแนน (8 ข้อ)	2 คะแนน (1 ข้อ)	10 คะแนน (9 ข้อ)
20% (20 คะแนน)	40:60 (8 คะแนน:12 คะแนน)	8 คะแนน (8 ข้อ)	12 คะแนน (6 ข้อ)	20 คะแนน (14 ข้อ)
	50:50 (10 คะแนน:10 คะแนน)	10 คะแนน (10 ข้อ)	10 คะแนน (5 ข้อ)	20 คะแนน (15 ข้อ)
	60:40 (12 คะแนน:8 คะแนน)	12 คะแนน (12 ข้อ)	8 คะแนน (4 ข้อ)	20 คะแนน (16 ข้อ)
	70:30 (14 คะแนน:6 คะแนน)	14 คะแนน (14 ข้อ)	6 คะแนน (3 ข้อ)	20 คะแนน (17 ข้อ)
30% (30 คะแนน)	40:60 (10 คะแนน:20 คะแนน)	10 คะแนน (10 ข้อ)	20 คะแนน (10 ข้อ)	30 คะแนน (20 ข้อ)
	50:50 (14 คะแนน:16 คะแนน)	14 คะแนน (14 ข้อ)	16 คะแนน (8 ข้อ)	30 คะแนน (22 ข้อ)
	60:40 (18 คะแนน:12 คะแนน)	18 คะแนน (18 ข้อ)	12 คะแนน (6 ข้อ)	30 คะแนน (24 ข้อ)
	70:30 (22 คะแนน:8 คะแนน)	22 คะแนน (22 ข้อ)	8 คะแนน (4 ข้อ)	30 คะแนน (26 ข้อ)

### (1) สัดส่วนคะแนนข้อสอบร่วมสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS)

ลักษณะของโครงสร้างที่ไม่ซับซ้อน (APSS) คือ ข้อสอบแบ่งออกเป็น 2 กลุ่มย่อย ลักษณะเวกเตอร์ของข้อสอบในมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ( $\theta_1$ ) ตั้งแต่ 0 - 15 องศา ส่วนเวกเตอร์ของ ข้อสอบในมิติที่ 2 ( $\theta_2$ ) ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 76-90 องศา โดยคะแนนรวมของข้อสอบร่วมในแต่ละ กลุ่มเท่ากัน แต่ในบางเงื่อนไขมีจำนวนข้อสอบร่วมในแต่ละกลุ่มต่างกัน เนื่องจากจำนวนข้อสอบร่วมที่มี การตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าในแต่ละกลุ่มต่างกัน

- อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า 40:60 ในสัดส่วนคะแนนข้อสอบ ร่วม 10%, 20% และ 30% คือ 4:6 (4 ข้อ : 3 ข้อ), 8:12 (8 ข้อ : 6 ข้อ) และ 10:20 (10 ข้อ : 10 ข้อ) ตามลำดับ โดยจำนวนข้อสอบร่วมทั้งหมดถูกแบ่งออกเป็นสองกลุ่มเท่ากัน ยกเว้นเงื่อนไขสัดส่วนข้อสอบ

รวม 10% ที่มีจำนวนข้อสอบรวมในแต่ละกลุ่มต่างกัน โดยกลุ่ม 1 และ 2 มีข้อสอบ 4 ข้อ (5 คะแนน) และ 3 ข้อ (5 คะแนน) ตามลำดับ

- อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 50:50 ในสัดส่วนคะแนนข้อสอบรวม 20% และ 30% คือ 10:10 (10 ข้อ : 5 ข้อ) และ 14:16 (14 ข้อ : 8 ข้อ) ตามลำดับ เงื่อนไขสัดส่วนข้อสอบรวม 20% มีจำนวนข้อสอบรวมในแต่ละกลุ่มต่างกัน โดยกลุ่ม 1 และกลุ่ม 2 มีข้อสอบ 7 ข้อ (10 คะแนน) และ 8 ข้อ (10 คะแนน) ตามลำดับ ส่วนสัดส่วนข้อสอบรวม 30% จำนวนข้อสอบรวมแต่ละกลุ่มเท่ากัน แบ่งเป็นกลุ่มละ 11 ข้อ (15 คะแนน) อย่างไรก็ตามไม่มีการศึกษาสัดส่วนดังกล่าวในเงื่อนไขสัดส่วนคะแนนข้อสอบรวม 10% เนื่องจากคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าและหลายค่าที่ได้ คือ 5 คะแนน : 5 คะแนน ไม่สอดคล้องกับคะแนนเต็มของข้อสอบรวมที่ตรวจให้คะแนนหลายค่า

- อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 60:40 ในสัดส่วนคะแนนข้อสอบรวม 10%, 20% และ 30% คือ 6:4 (6 ข้อ : 2 ข้อ), 12:8 (12 ข้อ : 4 ข้อ) และ 18:12 (18 ข้อ : 6 ข้อ) ตามลำดับ โดยคะแนนและจำนวนข้อสอบรวมทั้งหมดถูกแบ่งออกเป็นสองกลุ่มเท่ากัน

- อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30 ในสัดส่วนคะแนนข้อสอบรวม 10%, 20% และ 30% คือ 8:2 (8 ข้อ : 1 ข้อ), 14:6 (14 ข้อ : 3 ข้อ) และ 22:8 (22 ข้อ : 4 ข้อ) เงื่อนไขสัดส่วนข้อสอบรวม 10% มีจำนวนข้อสอบในแต่ละกลุ่มต่างกัน โดยมีติ 1 และมติ 2 มีข้อสอบ 4 ข้อ (5 คะแนน) และ 5 ข้อ (5 คะแนน) ตามลำดับ และเงื่อนไขสัดส่วนข้อสอบรวม 20% ที่มีจำนวนข้อสอบในแต่ละกลุ่มต่างกัน โดยกลุ่ม 1 และกลุ่ม 2 มีข้อสอบ 9 ข้อ (10 คะแนน) และ 8 ข้อ (10 คะแนน) ตามลำดับ รายละเอียดดังตาราง 3.3

ตาราง 3.3 สัดส่วนคะแนนข้อสอบรวมและข้อสอบเฉพาะในแต่ละเงื่อนไข สำหรับโครงสร้างที่ไม่ซับซ้อน

ข้อสอบรวม						ข้อสอบเฉพาะ				
สัดส่วนคะแนน	ขนาดของมุมจาก		ลำดับข้อ (คะแนน)			ขนาดของมุมจาก		ลำดับข้อ (คะแนน)		
	มติ 1 ( $\theta_1$ )	มติ 2 ( $\theta_2$ )	DICHO (คะแนน)	POLY (คะแนน)	รวม (คะแนน)	มติ 1 ( $\theta_1$ )	มติ 2 ( $\theta_2$ )	DICHO (คะแนน)	POLY (คะแนน)	รวม (คะแนน)
สัดส่วนคะแนนข้อสอบรวม 10%						สัดส่วนคะแนนข้อสอบเฉพาะ 90%				
40:60 (4:6)	0-15 76-90	90- $\theta_1$	1-3 (3) 4 (1)	5 (2) 6-7 (4)	4 ข้อ (5) 3 ข้อ (5)	0-15 76-90	90- $\theta_1$	1-31 (31) 32-62 (31)	63-69 (14) 70-76 (14)	38 ข้อ (45) 38 ข้อ (45)
60:40 (6:4)	0-15 76-90	90- $\theta_1$	1-3 (3) 4-6 (3)	7 (2) 8 (2)	4 ข้อ (5) 4 ข้อ (5)					
70:30 (8:2)	0-15 76-90	90- $\theta_1$	1-3 (3) 4-8 (5)	9 (2) -	4 ข้อ (5) 5 ข้อ (5)					

ตาราง 3.3 (ต่อ)

ข้อสอบรวม						ข้อสอบเฉพาะ				
สัดส่วน คะแนน DICHO	ขนาดของมุมจาก		ลำดับข้อ (คะแนน)			ขนาดของมุมจาก		ลำดับข้อ (คะแนน)		
	มิติ 1 ( $\theta_1$ )	มิติ 2 ( $\theta_2$ )	DICHO (คะแนน)	POLY (คะแนน)	รวม (คะแนน)	มิติ 1 ( $\theta_1$ )	มิติ 2 ( $\theta_2$ )	DICHO (คะแนน)	POLY (คะแนน)	รวม (คะแนน)
สัดส่วนคะแนนข้อสอบรวม 20%						สัดส่วนคะแนนข้อสอบเฉพาะ 80%				
40:60 (8:12)	0-15	76-90	1-4 (4)	9-11 (6)	7 ข้อ(10)	0-15	90- $\theta_1$	1-28 (28)	57-62 (12)	34 ข้อ (40)
	76-90	0-15	5-8 (4)	12-14 (6)	7 ข้อ(10)					
50:50 (10:10)	0-15	76-90	1-4 (4)	11-13 (6)	7 ข้อ(10)	76-90		29-56 (56)	63-68 (12)	34 ข้อ (40)
	76-90	0-15	5-10 (6)	14-15 (4)	8 ข้อ(10)					
60:40 (12:8)	0-15	76-90	1-6 (6)	13-14 (4)	8 ข้อ(10)	76-90		29-56 (56)	63-68 (12)	34 ข้อ (40)
	76-90	0-15	7-12 (6)	15-16 (4)	8 ข้อ(10)					
70:30 (14:6)	0-15	76-90	1-8 (8)	15 (2)	9 ข้อ(10)	76-90		29-56 (56)	63-68 (12)	34 ข้อ (40)
	76-90	0-15	9-14 (6)	16-17 (4)	8 ข้อ(10)					
สัดส่วนคะแนนข้อสอบรวม 30%						สัดส่วนคะแนนข้อสอบเฉพาะ 70%				
40:60 (10:20)	0-15	76-90	1-5 (5)	11-15(10)	10 ข้อ(15)	0-15	90- $\theta_1$	1-25 (25)	51-55 (10)	30 ข้อ (35)
	76-90	0-15	6-10 (5)	16-20(10)	10 ข้อ (15)					
50:50 (14:16)	0-15	76-90	1-7 (7)	15-22 (8)	11 ข้อ (15)	76-90		26-50 (25)	56-60 (10)	30 ข้อ (35)
	76-90	0-15	8-14 (7)	23-30 (8)	11 ข้อ (15)					
60:40 (18:12)	0-15	76-90	1-9 (9)	19-24 (6)	12 ข้อ (15)	76-90		26-50 (25)	56-60 (10)	30 ข้อ (35)
	76-90	0-15	10-18 (9)	25-29 (6)	12 ข้อ (15)					
70:30 (22:8)	0-15	76-90	1-11 (11)	23-26 (4)	13 ข้อ (15)	76-90		26-50 (25)	56-60 (10)	30 ข้อ (35)
	76-90	0-15	12-22(11)	27-30 (4)	13 ข้อ (15)					

## (2) สัดส่วนคะแนนข้อสอบรวมสำหรับโครงสร้างที่ซับซ้อน

อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าในแต่ละมิติสำหรับโครงสร้างที่ซับซ้อน (CS) เหมือนกับสัดส่วนในโครงสร้างที่ไม่ซับซ้อน (APSS) แต่โครงสร้างที่ซับซ้อน แบ่งแบบสอบออกเป็น 4 ชุดย่อยซึ่งมีขนาดของมุมที่วัดจากแต่ละกลุ่มแตกต่างกัน โดยกำหนดมุมของโครงสร้างที่ซับซ้อนตามงานวิจัยดังกล่าว โดยแบ่งแบบสอบออกเป็น 4 ชุดย่อย โดยแบบสอบย่อย 2 ชุดแรกแยกเป็นมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 0-15 องศา และมิติที่ 2 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 76-90 องศา ส่วนแบบสอบย่อยอีก 2 ชุดวัดในลักษณะมิติรวม โดยชุดแรกมีความใกล้เคียงกับมิติที่ 1 มากกว่ามิติที่ 2 ทำมุมกับแกนที่ 1 ประมาณ 31-45 องศา ส่วนชุดสอง มีความใกล้เคียงกับมิติที่ 2 มากกว่ามิติที่ 1 ทำมุมกับแกนมิติที่ 1 ตั้งแต่ 46-60 องศา

เนื่องจากจำนวนข้อสอบรวมและคะแนนเต็มข้อสอบรวมไม่สามารถแบ่งเป็น 4 ชุดย่อยได้พอดี ดังนั้นจำนวนข้อสอบและคะแนนเต็มของข้อสอบรวมในแต่ละกลุ่มจึงไม่เท่ากัน และข้อสอบรวมที่ตรวจให้คะแนนหลายค่าไม่สามารถแบ่งเป็นแบบสอบ 4 ชุดย่อยได้ จึงทำให้ข้อสอบรวมประเภทดังกล่าวอยู่ในบางกลุ่มเท่านั้น รายละเอียดดังตาราง 3.4

ตาราง 3.4 สัดส่วนคะแนนข้อสอบรวมและข้อสอบเฉพาะในแต่ละเงื่อนไข สำหรับโครงสร้างที่ซับซ้อน

ข้อสอบรวม						ข้อสอบเฉพาะ				
สัดส่วน คะแนน	ขนาดของมุม จาก		ลำดับข้อ (คะแนน)			ขนาดของมุมจาก		ลำดับข้อ (คะแนน)		
	มิติ 1 ( $\theta_1$ )	มิติ 2 ( $\theta_2$ )	DICHO	POLY	รวม (คะแนน)	มิติ 1 ( $\theta_1$ )	มิติ 2 ( $\theta_2$ )	DICHO	POLY	รวม (คะแนน)
สัดส่วนข้อสอบรวม 10%						สัดส่วนข้อสอบเฉพาะ 90%				
40:60 (4 : 6)	0-15	$\theta_1$	1 (1)	5 (2)	2 ข้อ (3)	0-15	$\theta_1$	1-16 (16)	64-66 (6)	19 ข้อ (22)
	31-45		2-3 (2)		2 ข้อ (2)					
	46-60			6 (2)	2 ข้อ (2)					
	76-90		4 (1)	7 (2)	2 ข้อ (3)					
60:40 (6 : 4)	0-15	$\theta_1$	1 (1)	7 (2)	2 ข้อ (3)	31-45	$\theta_1$	17-31 (15)	67-70 (8)	19 ข้อ (23)
	31-45		2-3 (2)		2 ข้อ (2)					
	46-60		4-5 (2)		2 ข้อ (2)					
	76-90		6 (1)	8 (2)	2 ข้อ (3)					
70:30 (8 : 2)	0-15	$\theta_1$	1 (1)	9 (2)	2 ข้อ (3)	46-60	$\theta_1$	32-47 (15)	71-74 (8)	19 ข้อ (23)
	31-45		2-3 (2)		3 ข้อ (3)					
	46-60		4-5 (2)		2 ข้อ (2)					
	76-90		6-8 (3)		2 ข้อ (3)					
40:60 (8 : 12)	0-15	$\theta_1$	1 (1)	9-10 (4)	3 ข้อ (5)	0-15	$\theta_1$	1-14 (14)	57-59 (6)	17 ข้อ (20)
	31-45		2-4 (3)	11 (2)	4 ข้อ (5)					
	46-60		5-7 (3)	12 (2)	4 ข้อ (5)					
	76-90		8 (1)	13-14 (4)	3 ข้อ (5)					
50:50 (10 : 10)	0-15	$\theta_1$	1 (1)	11-12 (4)	3 ข้อ (5)	31-45	$\theta_1$	15-28 (14)	60-62 (6)	17 ข้อ (20)
	31-45		2-4 (3)	13 (2)	4 ข้อ (5)					
	46-60		5-7 (3)	14 (2)	4 ข้อ (5)					
	76-90		8-10 (3)	15 (2)	4 ข้อ (5)					
60:40 (12 : 8)	0-15	$\theta_1$	1-3 (3)	13 (2)	4 ข้อ (5)	46-60	$\theta_1$	29-42 (14)	63-65 (6)	17 ข้อ (20)
	31-45		4-6 (3)	14 (2)	4 ข้อ (5)					
	46-60		7-9 (3)	15 (2)	4 ข้อ (5)					
	76-90		10-12 (3)	16 (2)	4 ข้อ (5)					
70:30 (14 : 6)	0-15	$\theta_1$	1-5 (5)		5 ข้อ (5)	76-90	$\theta_1$	43-56 (14)	66-68 (6)	17 ข้อ (20)
	31-45		6-8 (3)	15 (2)	5 ข้อ (5)					
	46-60		9-11 (3)	16 (2)	4 ข้อ (5)					
	76-90		12-14 (3)	17 (2)	4 ข้อ (5)					
สัดส่วนข้อสอบรวม 20%						สัดส่วนข้อสอบเฉพาะ 80%				

ตาราง 3.4 (ต่อ)

ข้อสอบรวม						ข้อสอบเฉพาะ				
สัดส่วน คะแนน	ขนาดของมุม จาก		ลำดับข้อ (คะแนน)			ขนาดของมุมจาก		ลำดับข้อ (คะแนน)		
	มิติ 1 ( $\theta_1$ )	มิติ 2 ( $\theta_2$ )	DICHO	POLY	รวม (คะแนน)	มิติ 1 ( $\theta_1$ )	มิติ 2 ( $\theta_2$ )	DICHO	POLY	รวม (คะแนน)
สัดส่วนข้อสอบรวม 30%						สัดส่วนข้อสอบเฉพาะ 70%				
40:60 (8 : 12)	0-15	$\theta_1$	1-3 (3)	11-13 (6)	6 ข้อ (9)	0-15	$\theta_1$	1-12 (12)	51-52 (4)	14 ข้อ (16)
	31-45		4-5 (2)	13-15 (4)	6 ข้อ (6)					
	46-60		6-7 (2)	16-17 (4)	6 ข้อ (6)					
	76-90		8-10 (3)	18-20 (6)	6 ข้อ (9)					
50:50 (10 : 10)	0-15	$\theta_1$	1-3 (3)	15-16 (4)	5 ข้อ (7)	31-45		13-25 (13)	53-55 (6)	16 ข้อ (19)
	31-45		4-7 (4)	17-18 (4)	6 ข้อ (8)					
	46-60		8-11 (4)	19-20 (4)	6 ข้อ (8)					
	76-90		12-14 (3)	21-22 (4)	5 ข้อ (7)					
60:40 (18 : 12)	0-15	$\theta_1$	1-5 (5)	19-20 (4)	7 ข้อ (9)	46-60		26-38 (13)	56-58 (6)	16 ข้อ (19)
	31-45		6-9 (4)	21 (2)	5 ข้อ (6)					
	46-60		10-13 (4)	22 (2)	5 ข้อ (6)					
	76-90		14-18 (5)	23-24 (4)	6 ข้อ (9)					
70:30 (22 : 8)	0-15	$\theta_1$	1-7 (7)	23 (2)	8 ข้อ (9)	76-90		39-50 (12)	59-60 (4)	14 ข้อ (16)
	31-45		8-11 (4)	24 (2)	6 ข้อ (6)					
	46-60		12-15 (4)	25 (2)	6 ข้อ (6)					
	76-90		16-22 (7)	26 (2)	8 ข้อ (9)					

### 1.2.2) การจำลองค่าพารามิเตอร์ของข้อสอบ

การจำลองค่าพารามิเตอร์ของข้อสอบใช้ข้อมูล 2 ส่วน คือ (1) ค่าอำนาจจำแนกรวม และค่าความยากรวม และ (2) ทิศทางการวัดที่ดีที่สุด (มุมระหว่างเวกเตอร์) หลังจากนั้นจึงนำค่าทั้ง 2 ส่วนมาคำนวณค่าพารามิเตอร์ข้อสอบ โดยดำเนินการผ่านโปรแกรม R มีรายละเอียดดังนี้

(1) ระบุค่าอำนาจจำแนกรวม (MDISC) และค่าความยากรวม (MDIFF) ซึ่งผู้วิจัยเลือกใช้ค่าอำนาจจำแนกรวมและค่าความยากของ Min (2003 อ้างอิงใน Zhang, 2012) ดังตาราง 3.5 โดยกระทำซ้ำตามรูปแบบดังกล่าวจนครบตามจำนวนข้อสอบที่กำหนดไว้ในแต่ละเงื่อนไข โดยค่าอำนาจจำแนกรวมและค่าความยากรวมในแต่ละเงื่อนไขระบุไว้ใน ดังภาคผนวก ก (ตาราง 1-11 สำหรับโครงสร้าง APSS และตาราง 12-22 สำหรับโครงสร้าง CS)

ตาราง 3.5 การกระจายค่าความยากรวมและค่าอำนาจจำแนกรวมใน 5 ระดับ

ระดับ	ค่าอำนาจจำแนกรวม (MDISC)	ค่าความยากรวม (MDIFF)
1	.4	-1.5
2	.8	1.0
3	1.2	0
4	1.6	-1.0
5	2.0	1.5

(2) **ระบุทิศทางการวัดที่ดีที่สุด (direction of best measurement)** ซึ่งก็คือมุมระหว่างเวกเตอร์สำหรับข้อสอบ ( $\alpha_{ik}$ ) โดยค่ามุมของข้อสอบแต่ละข้อในแต่ละกลุ่มย่อยถูกสุ่มด้วยการแจกแจงสม่ำเสมอ (uniform distribution) ตามช่วงที่กำหนด ซึ่งค่ามุมสำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อน (CS) แบ่งข้อสอบออกเป็น 2 กลุ่มย่อยและ 4 กลุ่มย่อย ดังตาราง 3.6

ตาราง 3.6 ขนาดของมุมเมื่อวัดจากมิติที่ 1 และมิติที่ 2 สำหรับโครงสร้าง APSS และ CS

แบบสอบย่อย	โครงสร้าง APSS		โครงสร้าง CS	
	ขนาดของมุมเมื่อวัดจาก		ขนาดของมุมเมื่อวัดจาก	
	มิติที่ 1 ( $\theta_1$ )	มิติที่ 2 ( $\theta_2$ )	มิติที่ 1 ( $\theta_1$ )	มิติที่ 2 ( $\theta_2$ )
1	0-15 องศา	76-90 องศา	0-15 องศา	76-90 องศา
2	76-90 องศา	0-15 องศา	31-45 องศา	46-60 องศา
3	-	-	46-60 องศา	31-45 องศา
4	-	-	76-90 องศา	0-15 องศา

(3) **คำนวณค่าพารามิเตอร์ของข้อสอบ** จากค่าความยากรวม ค่าอำนาจจำแนกรวม (ข้อ 1.1) และค่ามุม (ข้อ 1.2) ดังสมการ

$$\begin{aligned} a_{1i} &= MDISC_i \times \cos\alpha_{1i} \\ a_{2i} &= MDISC_i \times \sin\alpha_{1i} = MDISC_i \times \cos\alpha_{2i} \\ d_i &= -MDISC_i \times MDIFF_i \end{aligned}$$

การจำลองค่าพารามิเตอร์สำหรับโครงสร้างมิติความสามารถทั้งสองแบบมีการทำซ้ำจำนวน 40 รอบในแต่ละเงื่อนไขสำหรับแบบสอบแต่ละฉบับ

### 1.3 การจำลองความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง

การจำลองความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง เป็นการนำค่าความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบที่จำลองได้ในข้อ 1.1 และ 1.2 มาคำนวณความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องสำหรับผู้สอบคนที่  $j$  ของข้อสอบข้อที่  $i$  โดยความน่าจะเป็นในตอบข้อสอบที่ตรวจให้คะแนนสองค่าได้ถูกต้อง คำนวณภายใต้โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (M3PL) ดังสมการ

$$P(U_{ij} = 1 | \theta_j, \mathbf{a}_i, c_i, d_i) = c_i + (1 - c_i) \frac{e^{\mathbf{a}_i \theta_j + d_i}}{1 + e^{\mathbf{a}_i \theta_j + d_i}}$$

เมื่อ  $P(U_{ij} = 1 | \theta_j, \mathbf{a}_i, d_i)$  คือ ความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้อง ( $U_{ij} = 1$ ) สำหรับผู้สอบ  $j$  ของข้อสอบข้อที่  $i$

$\mathbf{a}_i$  คือ  $m \times 1$  เวกเตอร์ของพารามิเตอร์อำนาจจำแนกของข้อสอบ

$d_i$  คือ ค่าพารามิเตอร์ที่เป็นสเกลาร์ซึ่งสัมพันธ์กับความยากของข้อสอบ

$\theta_j$  คือ  $m \times 1$  เวกเตอร์ของพารามิเตอร์ความสามารถสำหรับผู้สอบ

$m$  คือ จำนวนของมิติความสามารถ

ส่วนความน่าจะเป็นในตอบข้อสอบข้อสอบที่ตรวจให้คะแนนหลายค่าได้ถูกต้องคำนวณภายใต้จากโมเดล Multidimensional generalized partial credit แบบ 2 พารามิเตอร์ (MGPC) ดังสมการ

$$P(u_{ij} = k | \theta_j) = \frac{e^{k \mathbf{a}_i \theta_j - \sum_{u=0}^k \beta_{iu}}}{\sum_{v=0}^{K_i} e^{v \mathbf{a}_i \theta_j - \sum_{u=0}^v \beta_{iu}}}$$

เมื่อ  $K_i$  คือ คะแนนสูงสุดสำหรับข้อสอบข้อที่  $i$  โดยคะแนนต่ำสุดกำหนดให้เป็น 0

$K_i + 1$  คือ ระดับคะแนนทั้งหมด

$k$  คือ คะแนนของผู้สอบสำหรับข้อสอบข้อที่  $i$

$\beta_{iu}$  คือ ค่าพารามิเตอร์เทรชโฮลด์ (threshold) สำหรับระดับคะแนน  $u$

### 1.4 การจำลองผลการตอบข้อสอบ

การจำลองผลการตอบข้อสอบประกอบด้วย (1) การจำลองผลการตอบข้อสอบที่ตรวจให้คะแนนสองค่า และ (2) การจำลองผลการตอบข้อสอบที่ตรวจให้คะแนนหลายค่า มีรายละเอียดดังนี้

#### 1.3.1) การจำลองผลการตอบข้อสอบที่ตรวจให้คะแนนสองค่า

การจำลองผลการตอบข้อสอบที่ตรวจให้คะแนนสองค่าข้อที่  $i$  สำหรับผู้สอบคนที่  $j$  ที่มีความสามารถ  $\theta$  ซึ่งก็คือ  $U_{ij}$  เป็นการเทียบการสุ่มแบบเอกรูป (uniform) เขียนเป็น  $y$  ที่อยู่ในช่วง  $(0,1)$  ไปจนถึงความน่าจะเป็นในการตอบข้อสอบถูก ดังนี้



$$U_{ij} = \begin{cases} 0, & y > P_{ij} \\ 1, & y \leq P_{ij} \end{cases}$$

ในทำนองเดียวกัน การจำลองผลการตอบข้อสอบที่ตรวจให้คะแนนหลายค่าข้อที่  $i$  ในรายการคำตอบที่  $k$  สำหรับผู้สอบคนที่  $j$  ที่มีความสามารถ  $\theta$  ซึ่งก็คือ  $U_{ijk}$  เป็นการเทียบการสุ่มแบบเอกรูป (uniform)

$$U_{ijk} = \begin{cases} 0, & 0 < y \leq P_{ij1} \\ 1, & P_{ij1} < y \leq P_{ij2} \\ 2, & P_{ij2} < y < 1 \end{cases}$$

## ตอนที่ 2 การตรวจสอบความถูกต้องของการจำลองข้อมูล

การตรวจสอบความถูกต้องของการจำลองข้อมูล แบ่งเป็น 2 ประเด็น ได้แก่

2.1 การตรวจสอบความถูกต้องของค่ามูที่จำลองขึ้น

2.2 การตรวจสอบความถูกต้องของค่าความสามารถของผู้สอบที่จำลองขึ้น

### 2.1 การตรวจสอบความถูกต้องของค่ามูที่จำลองขึ้น

การตรวจสอบความถูกต้องของค่ามูที่จำลองขึ้นโดยใช้โปรแกรม R ผู้วิจัยได้เลือกทดลองจำลองค่ามูสำหรับแบบสอบที่มีโครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อน (CS) ภายใต้เงื่อนไขสัดส่วนคะแนนข้อสอบรวม 30% และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30 จำนวน 20 รอบ เนื่องจากเงื่อนไขดังกล่าวมีจำนวนข้อสอบมากที่สุด คือ 86 ข้อ หลังจากนั้นนำค่ามูที่ได้มาวิเคราะห์ด้วยสถิติบรรยาย ได้แก่ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน เพื่อพิจารณาว่าค่ามูที่จำลองขึ้นสำหรับแบบสอบที่มีโครงสร้างดังกล่าวอยู่ในช่วงที่กำหนดไว้หรือไม่ โดยการผลตรวจสอบดังกล่าว มีรายละเอียดดังนี้

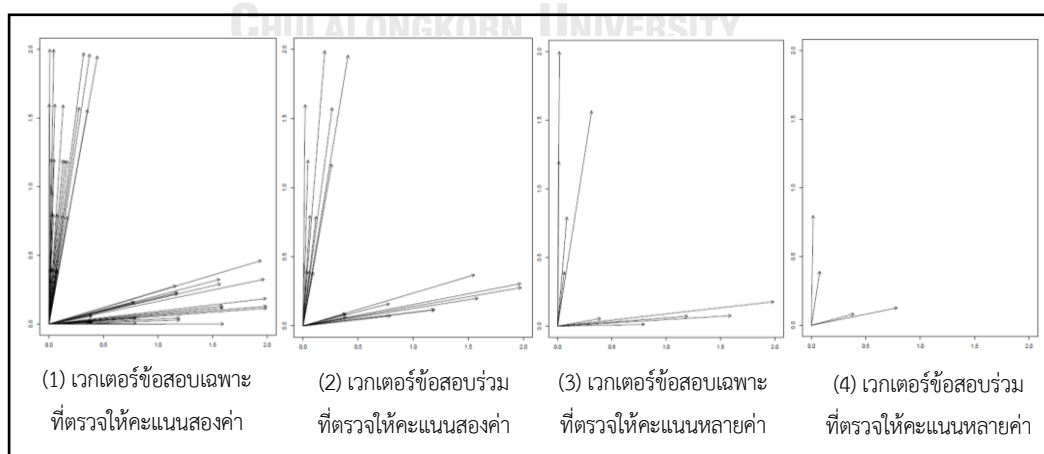
#### 2.1.1) ผลการตรวจสอบความถูกต้องของค่ามูที่จำลองขึ้น สำหรับโครงสร้างที่ไม่ซับซ้อน

แบบสอบที่มีโครงสร้างที่ไม่ซับซ้อนอย่าง ภายใต้เงื่อนไขสัดส่วนคะแนนข้อสอบรวม 30% และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30 ประกอบด้วย ข้อสอบเฉพาะจำนวน 60 ข้อ (คิดเป็น 70 คะแนน) ซึ่งประกอบด้วยข้อสอบแบบหลายตัวเลือกจำนวน 50 ข้อ (50 คะแนน) และข้อสอบแบบสร้างคำตอบจำนวน 10 ข้อ (20 คะแนน) และข้อสอบรวมจำนวน 26 ข้อ (30 คะแนน) ซึ่งประกอบด้วย ข้อสอบแบบหลายตัวเลือกจำนวน 22 ข้อ (22 คะแนน) และข้อสอบแบบสร้างคำตอบจำนวน 4 ข้อ (8 คะแนน) โดยการตรวจสอบความถูกต้องของค่ามูที่จำลองขึ้น สำหรับแบบสอบที่มีโครงสร้างที่ไม่ซับซ้อน เป็นการพิจารณาว่าข้อสอบกลุ่มแรกจำนวน 43 ข้อ มีค่ามูอยู่ในช่วง 0-15 องศา และข้อสอบกลุ่มสองจำนวน 43 ข้อ มีค่ามูอยู่ในช่วง 76-90 องศา หรือไม่

จากการจำลองค่ามุมสำหรับแบบสอบที่มีโครงสร้างอย่างง่ายภายใต้เงื่อนไขดังกล่าว จำนวน 20 รอบ พบว่า ในข้อสอบกลุ่มแรก ค่ามุมน้อยที่สุดของข้อสอบ คือ .02 (ข้อที่ 1) และค่ามุมมากที่สุด คือ 14.99 องศา (ข้อที่ 25) เป็นไปตามเงื่อนไขที่กำหนด และเมื่อพิจารณาค่ามุมจากแกนมิติที่ 1 และแกนมิติที่ 2 พบว่า ข้อสอบทุกข้อมีค่ามุมจากแกนมิติที่ 1 และ 2 รวมกันได้ 90 องศา อย่างเช่น ในข้อที่ 1 ค่ามุมน้อยที่สุดจากแกนมิติที่ 1 คือ .02 และค่ามุมมากที่สุดจากแกนมิติที่ 2 คือ 89.98 รวมกันได้ 90 องศา ส่วนค่ามุมมากที่สุดจากแกนมิติที่ 1 คือ 2.06 และค่ามุมน้อยที่สุดจากแกนมิติที่ 2 คือ 88.62 รวมกันได้ 90 องศาเช่นกัน

เมื่อพิจารณาข้อสอบกลุ่มสอง ค่ามุมที่น้อยที่สุด คือ 76.05 (ข้อ 26) และค่ามุมที่มากที่สุด คือ 89.96 (ข้อ 86) เป็นไปตามเงื่อนไขที่กำหนด และเมื่อพิจารณาค่ามุมจากแกนมิติที่ 1 และแกนมิติที่ 2 พบว่า ข้อสอบทุกข้อมีค่ามุมจากแกนมิติที่ 1 และ 2 รวมกันได้ 90 องศา อย่างเช่น ในข้อที่ 86 ค่ามุมน้อยที่สุดจากแกนมิติที่ 1 คือ 76.05 และค่ามุมมากที่สุดจากแกนมิติที่ 2 คือ 13.95 รวมกันได้ 90 องศา ส่วนค่ามุมมากที่สุดจากแกนมิติที่ 1 คือ 77.13 และค่ามุมน้อยที่สุดจากแกนมิติที่ 2 คือ 12.87 รวมกันได้ 90 องศาเช่นกัน จากข้อค้นพบดังกล่าว จึงสรุปได้ว่าการจำลองค่ามุมสำหรับแบบสอบที่มีโครงสร้างอย่างง่ายจึงมีความถูกต้องและเป็นไปตามเงื่อนไข รายละเอียดของค่ามุมที่จำลองขึ้นในแต่ละข้อ ดังภาคผนวก ข (ตาราง 23-24)

จากการตรวจสอบความถูกต้องของมุมที่จำลองขึ้น 20 รอบ ผู้วิจัยได้สร้างตัวอย่างกราฟแสดงเวกเตอร์ของแบบสอบที่มีโครงสร้างอย่างง่ายจากการจำลองรอบหนึ่ง ซึ่งแบ่งแบบสอบออกเป็น 2 ชุดย่อย ประกอบด้วย (1) เวกเตอร์ของข้อสอบเฉพาะที่ตรวจให้คะแนนสองค่า 50 ข้อ (2) เวกเตอร์ของข้อสอบรวมที่ตรวจให้คะแนนสองค่า 22 ข้อ (3) เวกเตอร์ของข้อสอบเฉพาะที่ตรวจให้คะแนนหลายค่า 10 ข้อ และ (4) เวกเตอร์ของข้อสอบรวมที่ตรวจให้คะแนนหลายค่า 4 ข้อ ดังแผนภาพ 3.2



ภาพ 3.2 ตัวอย่างกราฟแสดงเวกเตอร์ของแบบสอบที่มีโครงสร้างที่ไม่ซับซ้อน

### 2.1.2) ผลการตรวจสอบความถูกต้องของค่ามุมที่จำลองขึ้น สำหรับโครงสร้างที่ซับซ้อน

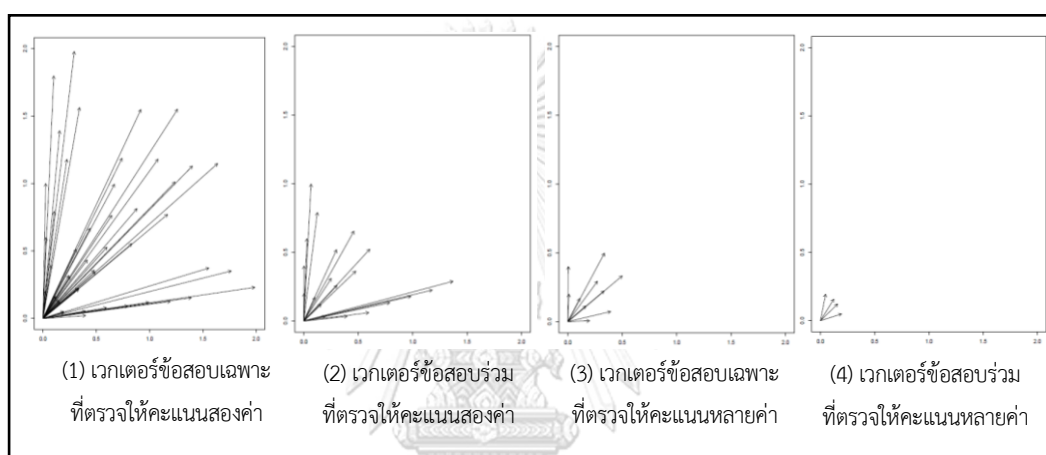
แบบสอบที่มีโครงสร้างที่ซับซ้อน (CS) ภายใต้เงื่อนไขสัดส่วนคะแนนข้อสอบรวม 30% และสัดส่วนคะแนนข้อสอบรวมแต่ละประเภท 70:30 ประกอบด้วย ข้อสอบเฉพาะจำนวน 60 ข้อ (70 คะแนน) ซึ่งประกอบด้วยข้อสอบแบบหลายตัวเลือกจำนวน 50 ข้อ (50 คะแนน) และข้อสอบแบบสร้างคำตอบจำนวน 10 ข้อ (20 คะแนน) และข้อสอบรวมจำนวน 26 ข้อ (30 คะแนน) ซึ่งประกอบด้วย ข้อสอบแบบหลายตัวเลือกจำนวน 22 ข้อ (22 คะแนน) และข้อสอบแบบสร้างคำตอบจำนวน 4 ข้อ (8 คะแนน) โดยการตรวจสอบความถูกต้องของค่ามุมที่จำลองขึ้นสำหรับแบบสอบที่มีโครงสร้างอย่างง่าย เป็นการพิจารณาว่าข้อสอบกลุ่มแรกมีค่ามุมอยู่ในช่วง 0-15 องศา ข้อสอบกลุ่มสองมีค่ามุมอยู่ในช่วง 31-45 องศา ข้อสอบกลุ่มสามมีค่ามุมอยู่ในช่วง 45-60 องศา และข้อสอบกลุ่มสี่มีค่ามุมอยู่ในช่วง 76-90 องศา หรือไม่

จากการจำลองค่ามุมสำหรับแบบสอบที่มีโครงสร้างที่ซับซ้อนภายใต้เงื่อนไขดังกล่าว จำนวน 20 รอบ พบว่า ในข้อสอบกลุ่มแรก ค่ามุมน้อยที่สุดของข้อสอบกลุ่มแรกเริ่มจาก .11 (ข้อที่ 1) และค่ามุมมากที่สุด คือ 14.88 องศา (ข้อที่ 12) เป็นไปตามเงื่อนไขที่กำหนด และเมื่อพิจารณาค่ามุมจากแกนมิติที่ 1 และแกนมิติที่ 2 พบว่า ข้อสอบทุกข้อมีค่ามุมจากแกนมิติที่ 1 และ 2 รวมกันได้ 90 องศา อย่างเช่น ข้อที่ 1 ค่ามุมน้อยที่สุดจากแกนมิติที่ 1 คือ .11 และค่ามุมมากที่สุดจากแกนมิติที่ 2 คือ 89.89 รวมกันได้ 90 องศา ส่วนค่ามุมมากที่สุดจากแกนมิติที่ 1 คือ 6.45 และค่ามุมน้อยที่สุดจากแกนมิติที่ 2 คือ 83.55 รวมกันได้ 90 องศาเช่นกัน เมื่อพิจารณาข้อสอบกลุ่มสอง ค่ามุมน้อยที่สุด คือ 31.01 (ข้อ 13) และค่ามุมมากที่สุด คือ 44.95 (ข้อ 25) เป็นไปตามเงื่อนไขที่กำหนด และเมื่อพิจารณาค่ามุมจากแกนมิติที่ 1 และแกนมิติที่ 2 พบว่า ข้อสอบทุกข้อมีค่ามุมจากแกนมิติที่ 1 และ 2 รวมกันได้ 90 องศา อย่างเช่น ข้อที่ 13 ค่ามุมน้อยที่สุดจากแกนมิติที่ 1 คือ 31.01 และค่ามุมมากที่สุดจากแกนมิติที่ 2 คือ 58.99 รวมกันได้ 90 องศา ส่วนค่ามุมมากที่สุดจากแกนมิติที่ 1 คือ 34.76 และค่ามุมน้อยที่สุดจากแกนมิติที่ 2 คือ 55.24 รวมกันได้ 90 องศาเช่นกัน

เมื่อพิจารณาข้อสอบกลุ่มสาม ค่ามุมน้อยที่สุด คือ 46.00 (ข้อ 26) และค่ามุมมากที่สุด คือ 59.90 (ข้อ 38) เป็นไปตามเงื่อนไขที่กำหนด และเมื่อพิจารณาค่ามุมจากแกนมิติที่ 1 และแกนมิติที่ 2 พบว่า ข้อสอบทุกข้อมีค่ามุมจากแกนมิติที่ 1 และ 2 รวมกันได้ 90 องศา อย่างเช่น ข้อที่ 26 ค่ามุมน้อยที่สุดจากแกนมิติที่ 1 คือ 46.00 และค่ามุมมากที่สุดจากแกนมิติที่ 2 คือ 44.00 รวมกันได้ 90 องศา ส่วนค่ามุมมากที่สุดจากแกนมิติที่ 1 คือ 50.88 และค่ามุมน้อยที่สุดจากแกนมิติที่ 2 คือ 39.12 รวมกันได้ 90 องศาเช่นกัน และเมื่อพิจารณาข้อสอบกลุ่มสี่ ค่ามุมน้อยที่สุด คือ 76.00 (ข้อ 39) และค่ามุมมากที่สุด คือ 86.69 (ข้อ 50) เป็นไปตามเงื่อนไขที่กำหนด และเมื่อพิจารณาค่ามุมจากแกนมิติที่ 1 และแกนมิติที่ 2 พบว่า ข้อสอบทุกข้อมีค่ามุมจากแกนมิติที่ 1 และ 2 รวมกันได้ 90 องศา อย่างเช่น ข้อที่ 39 ค่ามุมน้อยที่สุดจากแกนมิติที่ 1 คือ 76.00 และค่ามุมมากที่สุดจากแกนมิติที่ 2 คือ 14.00 รวมกันได้ 90 องศา ส่วนค่ามุมมากที่สุดจากแกนมิติที่ 1 คือ 78.51 และค่ามุมน้อยที่สุดจากแกนมิติที่ 2 คือ 11.49 รวมกันได้ 90 องศา

องศาเช่นกัน จากข้อค้นพบดังกล่าวจึงสรุปได้ว่า การจำลองค่ามุมสำหรับแบบสอบที่มีโครงสร้างแบบ ซับซ้อนจึงมีความถูกต้องและเป็นไปตามเงื่อนไข รายละเอียดของค่ามุมที่จำลองขึ้นในแต่ละข้อ ดังภาคผนวก ข (ตาราง 25-28)

จากการตรวจสอบความถูกต้องของมุมที่จำลองขึ้น 20 รอบ ผู้วิจัยได้สร้างตัวอย่างกราฟ แสดงเวกเตอร์ของแบบสอบที่มีโครงสร้างที่ซับซ้อนจากการจำลองรอบหนึ่ง ซึ่งแบ่งแบบสอบออกเป็น 4 ชุดย่อย ประกอบด้วย (1) เวกเตอร์ของข้อสอบเฉพาะที่ตรวจให้คะแนนสองค่า 50 ข้อ (2) เวกเตอร์ของ ข้อสอบรวมที่ตรวจให้คะแนนสองค่า 22 ข้อ (3) เวกเตอร์ของข้อสอบเฉพาะที่ตรวจให้คะแนนหลายค่า 10 ข้อ และ (4) เวกเตอร์ของข้อสอบรวมที่ตรวจให้คะแนนหลายค่า 4 ข้อ ดังแผนภาพ 3.3



ภาพ 3.3 ตัวอย่างกราฟแสดงเวกเตอร์ของแบบสอบที่มีโครงสร้างที่ซับซ้อน

## 2.2 การตรวจสอบความถูกต้องของค่าความสามารถของผู้สอบที่จำลองขึ้น

การตรวจสอบความถูกต้องของค่าความสามารถของผู้สอบที่จำลองขึ้นโดยใช้โปรแกรม R ผู้วิจัย ได้เลือกทดลองจำลองค่าความสามารถสำหรับผู้สอบทั้งสองกลุ่ม เพื่อพิจารณาว่าค่าความสามารถดังกล่าวเป็นไปตามเงื่อนไขที่กำหนดหรือไม่ กล่าวคือ การแจกแจงความสามารถของผู้สอบทั้งสองกลุ่ม เป็นการแจกแจงแบบปกติ 2 ตัวแปร (bivariate normal distribution) โดยเวกเตอร์ค่าเฉลี่ยและ เมทริกซ์ค่าความแปรปรวน-ความแปรปรวนร่วมของกลุ่มหนึ่ง คือ  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$  และ  $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$  ส่วนเวกเตอร์ค่าเฉลี่ยและเมทริกซ์ค่าความแปรปรวน-ความแปรปรวนร่วมของกลุ่มสอง คือ  $\begin{bmatrix} .5 \\ .5 \end{bmatrix}$  และ  $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$  และความสัมพันธ์ระหว่างมิติความสามารถเท่ากับ .5 มีรายละเอียดดังนี้

### 2.2.1) ผลการตรวจสอบความถูกต้องของค่าความสามารถของผู้สอบกลุ่มแรกที่จำลองขึ้น

การตรวจสอบความถูกต้องของค่าความสามารถของผู้สอบกลุ่มแรกที่จำลองขึ้น ประกอบด้วย 2 ส่วน ได้แก่ ค่าเฉลี่ยของความสามารถมิติที่ 1 และมิติที่ 2 และความสัมพันธ์ระหว่างมิติความสามารถ

#### (1) ค่าเฉลี่ยของความสามารถมิติที่ 1 และมิติที่ 2

เมื่อพิจารณาค่าสถิติเชิงบรรยายของค่าความสามารถของผู้สอบกลุ่มแรกที่จำลองขึ้นในแต่ละรอบ (20 รอบ) ดังภาคผนวก ค (ตาราง 29) พบว่า ค่าเฉลี่ยของความสามารถมิติที่ 1 มีค่าตั้งแต่ -.03 ถึง .02 ส่วนค่าเฉลี่ยของความสามารถมิติที่ 2 มีค่าตั้งแต่ -.03 ถึง .05 และเมื่อเปรียบเทียบเวกเตอร์ค่าเฉลี่ยของความสามารถมิติที่ 1 และมิติที่ 2 สำหรับผู้สอบกลุ่มแรกด้วยสถิติทดสอบ Hotelling T-square โดยมีสมมติฐานการทดสอบ คือ

$$H_0: \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ และ } H_1: \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \neq \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

ผลการวิเคราะห์พบว่า ไม่ปฏิเสธ  $H_0$  กล่าวคือ เวกเตอร์ค่าเฉลี่ยของค่าความสามารถมิติที่ 1 และมิติที่ 2 สำหรับผู้สอบกลุ่มแรกเท่ากับ  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$  ( $p = .167$ )

#### (2) ความสัมพันธ์ระหว่างมิติความสามารถ

เมื่อพิจารณาค่าสถิติเชิงบรรยายของความสัมพันธระหว่างมิติความสามารถของผู้สอบกลุ่มแรกที่จำลองขึ้นในแต่ละรอบ (20 รอบ) ดังภาคผนวก ค (ตาราง 29) พบว่า ความแปรปรวนของค่าความสามารถมิติที่ 1 มีค่าตั้งแต่ .96 ถึง 1.07 ส่วนความแปรปรวนของค่าความสามารถมิติที่ 2 มีค่าตั้งแต่ .97 ถึง 1.05 โดยความสามารถของผู้สอบทั้งสองมิติมีความสัมพันธ์ตั้งแต่ .45 ถึง .53 และเมื่อเปรียบเทียบความสัมพันธ์ระหว่างมิติจากทดสอบแต่ละรอบด้วยสถิติทดสอบ T-test โดยมีสมมติฐานการทดสอบ คือ

$$H_0: \rho = 0.5 \text{ และ } H_1: \rho \neq 0.5$$

ผลการวิเคราะห์พบว่า ไม่ปฏิเสธ  $H_0$  กล่าวคือ ความสามารถมิติที่ 1 และ 2 ของผู้สอบกลุ่มแรกเท่ากับ .50 ( $p = .592$ )

จากข้อค้นพบดังกล่าวจึงสรุปได้ว่า การจำลองค่าความสามารถของผู้สอบกลุ่มแรกเป็นไปตามเงื่อนไขที่กำหนด กล่าวคือเวกเตอร์ค่าเฉลี่ยและเมทริกซ์ค่าความแปรปรวน-ความแปรปรวนร่วมของกลุ่ม

แรกเท่ากับ  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$  และ  $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

## 2.2.2) ผลการตรวจสอบความถูกต้องของค่าความสามารถของผู้สอบกลุ่มสองที่จำลองขึ้น

การตรวจสอบความถูกต้องของค่าความสามารถของผู้สอบกลุ่มสองที่จำลองขึ้น ประกอบด้วย 2 ส่วน ได้แก่ ค่าเฉลี่ยของความสามารถมิติที่ 1 และมิติที่ 2 และความสัมพันธ์ระหว่างมิติความสามารถ

### (1) ค่าเฉลี่ยของความสามารถมิติที่ 1 และมิติที่ 2

เมื่อพิจารณาค่าสถิติเชิงบรรยายขอและค่าความสามารถของผู้สอบกลุ่มสองที่จำลองขึ้น ในแต่ละรอบ (20 รอบ) ดังภาคผนวก ค (ตาราง 30) พบว่า ค่าเฉลี่ยของความสามารถมิติที่ 1 มีค่าตั้งแต่ .45 ถึง .54 ส่วนค่าเฉลี่ยของความสามารถมิติที่ 2 มีค่าตั้งแต่ .47 ถึง .54 และเมื่อเปรียบเทียบเวกเตอร์ค่าเฉลี่ยของความสามารถมิติที่ 1 และมิติที่ 2 สำหรับผู้สอบกลุ่มแรกด้วยสถิติทดสอบ Hotelling T-square โดยมีสมมติฐานการทดสอบ คือ

$$H_0: \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix} \text{ และ } H_1: \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \neq \begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$$

ผลการวิเคราะห์พบว่า ยอมรับ  $H_0$  กล่าวคือ เวกเตอร์ค่าเฉลี่ยของค่าความสามารถมิติที่ 1 และมิติที่ 2 สำหรับผู้สอบกลุ่มสองเท่ากับ  $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$  ( $p = .589$ )

### (2) ความสัมพันธ์ระหว่างมิติความสามารถ

เมื่อพิจารณาค่าสถิติเชิงบรรยายของความสัมพันธ์ระหว่างมิติความสามารถของผู้สอบกลุ่มสองที่จำลองขึ้นในแต่ละรอบ (20 รอบ) ดังภาคผนวก ค (ตาราง 30) พบว่า ความแปรปรวนของค่าความสามารถมิติที่ 1 มีค่าตั้งแต่ .96 ถึง 1.08 ส่วนความแปรปรวนของค่าความสามารถมิติที่ 2 มีค่าตั้งแต่ .97 ถึง 1.03 โดยความสามารถของผู้สอบทั้งสองมิติมีความสัมพันธ์ตั้งแต่ .47 ถึง .52 และเมื่อเปรียบเทียบความสัมพันธ์ระหว่างมิติจากทดสอบแต่ละรอบด้วยสถิติทดสอบ T-test โดยมีสมมติฐานการทดสอบ คือ

$$\text{จู่พา } H_0: \rho = 0.5 \text{ และ } H_1: \rho \neq 0.5$$

ผลการวิเคราะห์พบว่า ไม่ปฏิเสธ  $H_0$  กล่าวคือ ความสามารถมิติที่ 1 และ 2 ของผู้สอบกลุ่มแรกเท่ากับ .50 ( $p = .541$ )

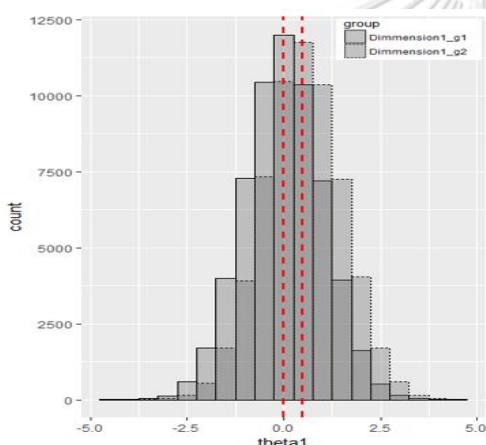
จากข้อค้นพบดังกล่าวจึงสรุปได้ว่า การจำลองค่าความสามารถของผู้สอบกลุ่มสองเป็นไปตามเงื่อนไขที่กำหนด กล่าวคือเวกเตอร์ค่าเฉลี่ยและเมทริกซ์ค่าความแปรปรวน-ความแปรปรวนร่วมของกลุ่ม

แรกเท่ากับ  $\begin{bmatrix} 0.5 \\ 0.5 \end{bmatrix}$  และ  $\begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$

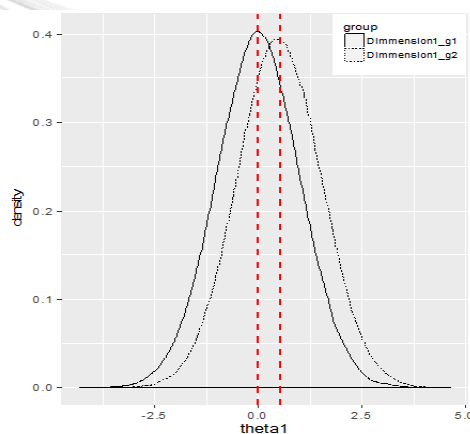
ตาราง 3.7 ผลการทดสอบความถูกต้องของค่าความสามารถสำหรับผู้สอบทั้งสองกลุ่ม

มิติ	ค่าสถิติบรรยายของค่าความสามารถของ ผู้สอบกลุ่มแรก			ค่าสถิติบรรยายของค่าความสามารถ ของผู้สอบกลุ่มสอง		
	N	Mean	Variance	N	Mean	Variance
1	6000	-.01	1.00	6000	.50	1.00
2	6000	.00	1.00	6000	.50	1.00
Hotelling T-square	$T^2 = 1.787$ , $df1 = 2$ , $df2 = 59998$ , $p = .167$			$T^2 = .530$ , $df1 = 2$ , $df2 = 59998$ , $p = .589$		
Bartlett's test	$X^2 = 17107.395$ , $df = 2$ , $p = .000$			$X^2 = 17109.587$ , $df = 2$ , $p = .000$		

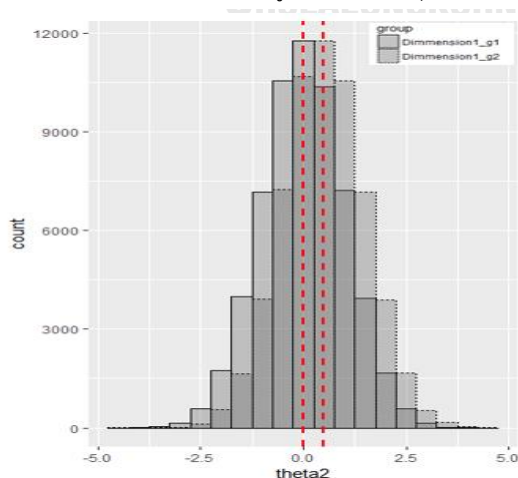
นอกจากนี้ผู้วิจัยได้สร้างฮิสโทแกรมและโค้งปกติแสดงความแตกต่างของค่าความสามารถในแต่ละมิติระหว่างผู้สอบทั้งสองกลุ่ม ซึ่งแสดงให้เห็นว่าผู้สอบกลุ่มแรกมีค่าความสามารถเฉลี่ยในมิติที่ 1 และ 2 สูงกว่าผู้สอบกลุ่มสองเท่ากับ .5 ดังภาพ 3.4 – 3.7



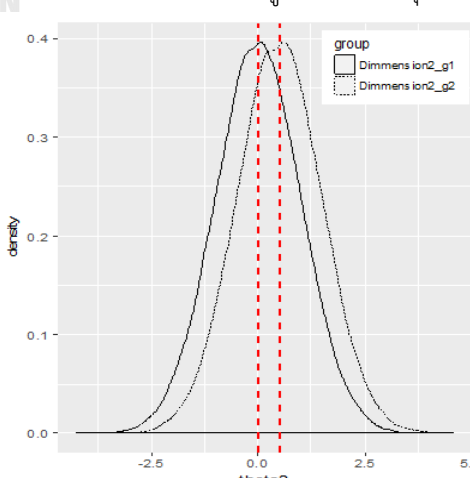
ภาพ 3.4 ฮิสโทแกรมแสดงความสามารถมิติที่ 1 ระหว่างผู้สอบทั้งสองกลุ่ม



ภาพ 3.5 เส้นโค้งปกติแสดงความสามารถมิติที่ 1 ระหว่างผู้สอบทั้งสองกลุ่ม



ภาพ 3.7 ฮิสโทแกรมแสดงความสามารถมิติที่ 2 ระหว่างผู้สอบทั้งสองกลุ่ม



ภาพ 3.6 เส้นโค้งปกติแสดงความสามารถมิติที่ 2 ระหว่างผู้สอบทั้งสองกลุ่ม

### ตอนที่ 3 การวัดประสิทธิภาพของการปรับเทียบคะแนน

ประสิทธิภาพของการปรับเทียบคะแนนพิจารณาจาก (1) ค่าความลำเอียงสัมพัทธ์ (RB) (2) สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนในการปรับเทียบคะแนน (CVSE) และ (3) รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ซึ่งจะอธิบายรายละเอียดไว้ในหัวข้อ 3.1-3.3 ตามลำดับ นอกจากนี้ยังสามารถพิจารณาความสามารถในการประมาณค่า SEE ของคะแนนปรับเทียบแต่ละคะแนนได้จาก ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ ( $RMSE_k$ )

การวัดประสิทธิภาพกำหนดให้วิธีการปรับเทียบที่ใช้เป็นเกณฑ์ คือวิธีอีควิเปอร์เซ็นต์ไทล์ที่ปรับโค้งให้ราบเรียบ (*presmoothed equipercentile equating method*) ซึ่งใช้ Polynomial log-linear model ปรับโค้งการแจกแจงให้ราบเรียบด้วยดีกรี (degree) ต่าง ๆ เพื่อหาโมเดลที่เหมาะสมกับข้อมูลมากที่สุด โดยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ได้จากการปรับโค้งแล้วต้องมีค่าเท่ากับข้อมูลเดิมที่ยังไม่ได้ทำการปรับ หลังจากนั้นจึงนำข้อมูล ที่ได้รับการปรับโค้งแล้วไปใช้ในการปรับเทียบคะแนนด้วยวิธีอีควิเปอร์เซ็นต์ไทล์ ซึ่งดำเนินการผ่านแพคเกจ “equate” ในโปรแกรม โดยโมเดลล็อกเชิงเส้นถูกทำให้สอดคล้องโดยใช้ฟังก์ชัน smoothed=“loglinear” โดยวิธีการระบุโมเดลล็อกเชิงเส้นในฟังก์ชัน presmoothing ที่มีประสิทธิภาพมากที่สุด คือ การระบุเทอมของ polynomial สูงสุดสำหรับตัวแปรแต่ละตัวในแต่ละระดับของปฏิสัมพันธ์ด้วยการระบุ degree=list(c(3,3), c(1,1)) หรือเขียนแบบสั้นว่า degree = list (3,1) (Albano, 2014)

#### 3.1 ค่าความลำเอียงสัมพัทธ์

ค่าความลำเอียงสัมพัทธ์ (relative bias: RB) คือ ค่าสัมบูรณ์ของผลต่างระหว่างค่าเฉลี่ยของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา ( $\hat{e}_{base}(x_i)$ ) และค่าเฉลี่ยของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนเกณฑ์ ( $\overline{e_{base}(x_i)}$ ) เมื่อเทียบกับส่วนเบี่ยงเบนมาตรฐานของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนเกณฑ์ ซึ่งสะท้อนความแม่นยำในการปรับเทียบคะแนน ดังสมการ

$$RB = \frac{|\hat{e}_{base}(x_i) - \overline{e_{base}(x_i)}|}{\sigma_{base}(x_i)} \times 100$$

เมื่อ	$x_i$	คือ คะแนนที่ $i$
	$e_{base}(x_i)$	คือ คะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนเกณฑ์
	$\hat{e}_{base}(x_i)$	คือ คะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา
	$\sigma_{base}(x_i)$	คือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา



### 3.2 สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน

สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (coefficient of variance of standard error of equating: CVSE) หมายถึง การกระจายของค่าความคลาดเคลื่อนมาตรฐานในการปรับเทียบคะแนน (SE) ซึ่งสะท้อนความถูกต้องในการปรับเทียบคะแนน ดังสมการ

$$CVSE = \frac{SE_i}{\hat{e}_{base}(x_i)} \times 100$$

โดยค่าความคลาดเคลื่อนมาตรฐาน (SE) คือ ขนาดของความคลาดเคลื่อนอย่างสุ่มในการปรับเทียบคะแนนซึ่งเกิดจากการสุ่มตัวอย่าง ค่าความคลาดเคลื่อนในแต่ละจุดคะแนนเป็นดังสมการ

$$SE_i = \sqrt{\frac{1}{N} \sum_{k=1}^N [\hat{e}_{base_k}(x_i) - \overline{\hat{e}_{base}(x_i)}]^2}$$

เมื่อ  $\hat{e}_{base}(x_i)$  คือ คะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา  
 $\overline{\hat{e}_{base}(x_i)}$  คือ ค่าเฉลี่ยของคะแนนปรับเทียบจากกระบวนการปรับเทียบคะแนนที่ศึกษา

### 3.3 รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง

รากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (root mean square error: RMSE) คือ รากที่สองของผลรวมของความแปรปรวนของความคลาดเคลื่อนมาตรฐานและความลำเอียงยกกำลังสอง ซึ่งสะท้อนประสิทธิภาพโดยรวม ดังสมการ

$$RMSE_i = \sqrt{SE_i^2 + Bias_i^2}$$

### 3.4 ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ

ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ ( $RMSE_k$ ) วัดได้จาก รากที่สองของความคลาดเคลื่อนกำลังสอง ซึ่งสะท้อนความสามารถของตัวประมาณในการประมาณค่าความคลาดเคลื่อนมาตรฐาน (SE) ของคะแนนปรับเทียบแต่ละคะแนนได้อย่างแม่นยำและถูกต้อง ดังสมการ

$$RMSE_k = \sqrt{\frac{\sum_{i=1}^{rep} (SE_{ik} - MCSE_k)^2}{rep-1}}$$

เมื่อ  $\widehat{SE}_{ik}$  คือ ค่า SE ที่ประมาณได้ในแต่ละจุดคะแนน  $k$  ในแต่ละรอบ  $i$   
 $\overline{MCSE}_k$  คือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนปรับเทียบ (40 รอบ) ในแต่ละระดับคะแนน  $k$

#### ตอนที่ 4 การวิเคราะห์ประสิทธิภาพของการปรับเทียบคะแนน

การเปรียบเทียบประสิทธิภาพระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT เมื่อโครงสร้างมิตีความสามารถ สัดส่วนคะแนนข้อสอบรวมและอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า แตกต่างกัน พิจารณาจากค่าความลำเอียงสัมพัทธ์ (RB) สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (SE) และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ด้วยสถิติวิเคราะห์ 3 แบบ ได้แก่

4.1 สถิติบรรยาย (descriptive statistics) เพื่ออธิบายลักษณะของค่า RB ค่า CVSE และค่า RMSE ประกอบด้วย ค่าสูงสุด ค่าต่ำสุด ค่าเฉลี่ยส่วนเบี่ยงเบนมาตรฐาน และความแปรปรวน

4.2 สถิติวิเคราะห์ความแปรปรวน (ANOVA) เพื่อตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยของตัวแปรตาม คือ ค่า RMSE

4.3 สถิติวิเคราะห์ตัวแปรพหุนาม (MANOVA) เพื่อตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยของตัวแปรตาม ได้แก่ ค่า RB ค่า CVSE

#### ตอนที่ 5 ขั้นตอนการดำเนินงานวิจัย

การวิจัยครั้งนี้ใช้การจำลองข้อมูลแบบมอนติคาร์โล ด้วยขนาดตัวอย่างกลุ่มละ 3,000 คนและจำนวนการทำซ้ำ 40 รอบ โดยขั้นตอนการดำเนินการสำหรับการปรับเทียบคะแนนมี 4 ขั้นตอนหลัก ได้แก่ (1) การจำลองข้อมูลค่าพารามิเตอร์ข้อสอบและผลการตอบข้อสอบ (2) การประมาณค่าพารามิเตอร์ข้อสอบ (3) การเชื่อมโยงสเกลด้วยวิธี TCF และ (4) การปรับเทียบคะแนน

##### 5.1 การจำลองค่าพารามิเตอร์ของข้อสอบและแบบแผนการตอบข้อสอบ

ขั้นตอนนี้เป็นการสร้างรูปแบบการตอบข้อสอบสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า คือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน ด้วยโมเดล M3PL และข้อสอบที่ตรวจให้คะแนนหลายค่า คือ 0, 1 และ 2 คะแนน ด้วยโมเดล MGPC ซึ่งมีการจำลองข้อมูล 4 ส่วน คือ (1) การจำลองค่าความสามารถ (2) การจำลองค่าพารามิเตอร์ของข้อสอบ (3) การจำลองความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง และ (4) การจำลองผลการตอบข้อสอบ โดยผู้วิจัยได้อธิบายรายละเอียดของการจำลองข้อมูลค่าพารามิเตอร์และผลการตอบข้อสอบไว้ในหัวข้อ 1.2.2, 1.3 และ 1.4

## 5.2 การประมาณค่าพารามิเตอร์ข้อสอบ

การประมาณค่าพารามิเตอร์สำหรับแบบสอบพหุมิติรูปแบบผสม (MIRT calibration) ด้วยโมเดล M3PL และโมเดล MGPC ใช้ Package “mirt” ในโปรแกรม R ด้วยอัลกอริทึม EM

1. การประมาณค่าพารามิเตอร์แบบแยกกัน (separate calibration) ใช้คำสั่ง `mirt ()` โดยกำหนดแบบสอบที่ต้องการประมาณค่าพารามิเตอร์ และจำนวนข้อสอบแต่ละประเภท

2. การประมาณค่าพารามิเตอร์แบบพร้อมกัน (concurrent calibration) ใช้คำสั่ง `multipleGroup ()` โดยกำหนดจำนวนข้อสอบของแบบสอบแต่ละฉบับ จำนวนข้อสอบแต่ละประเภท และจำนวนผู้สอบแต่ละกลุ่มตามเงื่อนไข

## 5.3 การเชื่อมโยงสเกลระหว่างแบบสอบ

กระบวนการ MOSE ที่มีการประมาณค่าพารามิเตอร์แบบแยกกัน และใช้การเชื่อมโยงสเกล ด้วยวิธี Test characteristic function (TCF) แบบไม่ถ่วงน้ำหนัก ซึ่งค่าความสามารถต่าง ๆ มีน้ำหนักเท่ากัน ( $W_0 = 1$ ) เช่นเดียวกับงานวิจัยของ Yao และ Boughton (2009) โดยวิธี TCF ต้องการหาผลรวมน้อยที่สุดของความแตกต่างระหว่างพื้นผิวคุณลักษณะของแบบสอบแบบพหุมิติ ดังสมการ

$$\min \left\{ \frac{1}{Q^D} \sum_{i=1}^{Q^D} [TRF(\theta_i, \beta_i) - TRF(\theta_i, \beta^*)]^2 \right\}$$

$$\text{เมื่อ } TRF(\theta_i, \beta_i) = \frac{1}{J_1 + \sum_{j=1}^{J_2} (K_j - 1)} \sum_{i=1}^N \left[ \sum_{j=1}^{J_1} P_{ij1} + \sum_{j=1}^{J_2} \sum_{k=1}^{K_j} (k - 1) P_{ijk} \right]$$

เมื่อ  $Q^D$  คือ จุดในการประมาณพื้นที่ทั้งหมด (quadrature points)

$D$  คือ จำนวนของมิติ

$P_{ijk}$  คือ ความน่าจะเป็นของการตอบในขั้น  $k - 1$  สำหรับแบบสอบที่ตรวจให้คะแนนหลายค่า  $j$  ข้อ ของผู้สอบที่มีเวกเตอร์ความสามารถ  $\theta_i$  ซึ่งความน่าจะเป็น  $P_{ijk}$  หาได้จากโมเดล MGPC

$J_1$  คือ จำนวนของข้อสอบที่ตรวจให้คะแนนสองค่า

$J_2$  คือ จำนวนของข้อสอบที่ตรวจให้คะแนนหลายค่า

การเชื่อมโยงสเกลด้วยวิธี TCF ต้องกำหนดเมทริกซ์การหมุน (rotation matrix) ขนาด  $2 \times 2$  (จำนวนมิติ  $\times$  จำนวนมิติ) และเมทริกซ์การเลื่อนตำแหน่ง (translation matrix) ขนาด  $2 \times 1$  (จำนวนมิติ  $\times 1$ ) ซึ่งผู้วิจัยคำนวณเมทริกซ์ดังกล่าวโดยใช้แพ็คเกจ `plink` ในโปรแกรม R ตามแนวคิดของ Oshima, Davey and Lee (2000, cited in Zhang, 2012) ที่มีการหมุนแบบไม่ตั้งฉาก (non-orthogonal) หลังจากนั้นจึงนำเมทริกซ์ที่ได้มาใช้ในการแปลงค่าพารามิเตอร์ ผ่านโปรแกรม R ด้วยการประยุกต์ใช้ R codes

จากงานวิจัยของ Simon (2008) ในขั้นตอนนี้ผู้วิจัยได้เพิ่มการประเมินประสิทธิภาพโดยรวม (ค่า RMSE) ของการเชื่อมโยงสเกลสำหรับค่าพารามิเตอร์ต่าง ซึ่งผลการประเมินในแต่ละเงื่อนไขอธิบายไว้ในภาคผนวก ง

#### 5.4 การเปรียบเทียบคะแนนระหว่างแบบสอบ

กระบวนการเปรียบเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (MOSE) มี 3 ขั้นตอน ดำเนินการผ่านโปรแกรม R โดยขั้นตอนสุดท้ายต้องมีการจำลองความหนาแน่นของค่าความสามารถ มีรายละเอียดดังนี้

**5.4.1) คำนวน Conditional observed score distributions ในแต่ละกลุ่มของระดับความสามารถ** สำหรับข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่า หลังจากนั้นนำค่าที่ได้มารวมกันสำหรับแบบสอบรูปแบบผสม โดยใช้โปรแกรม R ด้วยการประยุกต์ใช้คำสั่งของ Brossman (2010) มีรายละเอียดดังนี้

1. คำนวน Conditional observed score distributions สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า โดยใช้ Lord-Wingersky algorithm ตามกรอบแนวคิด MIRT ดังสมการ (Brossman, 2010)

$$\begin{aligned} f_r(x|\theta_j) &= f_{r-1}(x|\theta_j)(1 - P_r) & x = 0 \\ f_r(x|\theta_j) &= f_{r-1}(x|\theta_j)(1 - P_r) + f_{r-1}(x - 1|\theta_j)P_r & 0 < x < r \\ f_r(x|\theta_j) &= f_{r-1}(x - 1|\theta_j)P_r & x = r \end{aligned}$$

เมื่อ  $f_r(x|\theta_j)$  คือ Conditional observed score distributions

$\theta$  คือ เวกเตอร์ของความสามารถเวกเตอร์ของความสามารถ  $m$  (ability vector)

$r$  คือ ดัชนีข้อสอบ (item index)

2. คำนวน Conditional observed score distributions สำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า โดยประยุกต์ใช้สูตรของ Hanson (1994) และ Thissen และคณะ (1995 อ้างอิงใน Peterson, 2014) ดังสมการ

$$f_r(x|\theta_j) = \sum_{k=1}^{K_j} f_{r-1}(x - W_{rk}) P_{ik}(\theta_j) \quad \text{เมื่อ } \min_r < x < \max_r$$

เมื่อ  $f_r(x|\theta_j)$  คือ Conditional observed score distributions สำหรับความสามารถแฝง  $m$  สำหรับข้อสอบข้อที่  $i$  ด้วยลำดับชั้นของคำตอบ (response categories,  $k$ )

$W_{rk}$  คือ ฟังก์ชันการให้คะแนนของลำดับชั้นที่  $k$  สำหรับข้อสอบข้อที่  $r$  โดยที่ค่า  $K$  แปรเปลี่ยนไปตามข้อสอบ

$\min_r$  คือ คะแนนต่ำสุดที่เป็นไปได้ภายหลังการเพิ่มของข้อสอบข้อที่  $r$

$\max_r$  คือ คะแนนสูงสุดที่เป็นไปได้ภายหลังการเพิ่มของข้อสอบข้อที่  $r$

3. คำนวณ Conditional observed score distributions สำหรับแบบสอบรูปแบบผสม ด้วยการรวมผลลัพธ์การแจกแจงของข้อสอบที่ตรวจให้คะแนนสองค่าและตรวจให้คะแนนหลายค่าบนเวกเตอร์เงื่อนไขความสามารถ ( $\theta_j$ )

5.4.2) คำนวณ Marginal observed score distributions สำหรับแบบสอบรูปแบบผสม โดยนำ Conditional observed score distributions (ข้อ 1.3) คูณด้วยความหนาแน่นความสามารถแบบพหุตัวแปร (multivariate ability density:  $\psi(\theta)$ ) และรวมผลลัพธ์ทั้งหมดของความสามารถแฝง  $m$  ดังสมการ

$$f(x) = \sum_{\theta_1} \sum_{\theta_2} \cdots \sum_{\theta_m} f(x|\theta)\psi(\theta)$$

การจำลองความหนาแน่นความสามารถแบบพหุตัวแปร (multivariate ability density:  $\psi(\theta)$ ) ทุกความสามารถแฝง  $m$  กำหนดให้มีการแจกแจงมาตรฐานที่ไม่มีความสัมพันธ์กัน (multivariate standard normal distribution with uncorrelated axes:  $\theta \sim MVN(\mathbf{0}, I)$ ) ซึ่งคำนวณจาก Package “mvtnorm” ในโปรแกรม R ด้วยฟังก์ชัน dmvtnorm

5.4.3) ปรับเทียบคะแนน โดยใช้ Marginal observed score distributions ในการปรับเทียบคะแนนระหว่างแบบสอบด้วยวิธีอิกวิเปอร์เซ็นไทล์แบบดั้งเดิมผ่าน Package “equate” โปรแกรม R

## บทที่ 4

### ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ใช้การจำลองข้อมูลแบบมอนติคาร์โล เพื่อเปรียบเทียบประสิทธิภาพในการปรับเทียบคะแนนแบบสอบพหุมิติรูปแบบผสม ระหว่างกระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบพร้อมกัน (CMOSE) และกระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบแยกกัน และใช้การเชื่อมโยงสเกลด้วยวิธี TCF (SMOSE) สำหรับแบบสอบพหุมิติ ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกัน โดยใช้ข้อสอบร่วม (NEAT) เมื่อโครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบร่วม และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า แตกต่างกัน

ประสิทธิภาพในการปรับเทียบคะแนนสามารถพิจารณาได้จาก (1) ความลำเอียงสัมพัทธ์ (RB) (2) สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (CVSE) และ (3) รากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (root mean square: RMSE) สะท้อนถึงความแม่นยำ ความถูกต้อง ประสิทธิภาพโดยรวมของการปรับเทียบคะแนน ตามลำดับ การวิเคราะห์ประสิทธิภาพในการปรับเทียบคะแนนใช้สถิติวิเคราะห์ความแปรปรวน (ANOVA) เพื่อตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยของค่า RMSE และสถิติวิเคราะห์ตัวแปรพหุนาม (MANOVA) เพื่อตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยของ ค่า RB และค่า CVSE

การนำเสนอผลการวิเคราะห์ เริ่มจาก ผลปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพของการปรับเทียบคะแนน แล้วจึงอธิบายผลการวิเคราะห์ตามวัตถุประสงค์การวิจัย และสรุปประสิทธิภาพของการปรับเทียบแยกตามวัตถุประสงค์ นอกจากนี้ผู้วิจัยได้วิเคราะห์ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบเพิ่มเติมไว้ท้ายบท การนำเสนอแบ่งเป็น 7 ตอน ดังนี้

ตอนที่ 1 ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพของการปรับเทียบคะแนน

ตอนที่ 2 ประสิทธิภาพระหว่างกระบวนการ CMOSE กับกระบวนการ SMOSE

ตอนที่ 3 ประสิทธิภาพของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบร่วมแตกต่างกัน

ตอนที่ 4 ประสิทธิภาพของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแตกต่างกัน

ตอนที่ 5 ประสิทธิภาพของกระบวนการ MOSE เมื่อโครงสร้างมิติความสามารถแตกต่างกัน

ตอนที่ 6 สรุปประสิทธิภาพของการปรับเทียบคะแนนแยกตามวัตถุประสงค์

ตอนที่ 7 ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ

## ตอนที่ 1 ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพของการปรับเทียบคะแนน

การนำเสนอผลปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพของการปรับเทียบคะแนน แบ่งเป็น 2 ส่วน คือ (1) ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพโดยรวม (RMSE) และ (2) ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อความแม่นยำในการปรับเทียบคะแนน (RB) และความถูกต้องในการปรับเทียบคะแนน (CVSE) รายละเอียดดังนี้

### 1.1 ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพโดยรวม (RMSE)

การวิเคราะห์ประสิทธิภาพโดยรวมของการปรับเทียบคะแนนจำแนกตามเงื่อนไขใช้การวิเคราะห์ความแปรปรวน (ANOVA) ซึ่งพบว่า ความแปรปรวนของค่า RMSE ระหว่างกลุ่มแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ( $p = .000$ ) ดังตาราง 4.1 และเมื่อทดสอบอิทธิพลระหว่างตัวแปรอิสระที่มีต่อค่า RMSE พบว่า ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระ 4 ตัว ได้แก่ รูปแบบของกระบวน MOSE (MOSE) โครงสร้างมิติความสามารถ (DS) สัดส่วนคะแนนข้อสอบรวม (ANCHOR) และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า (DICH0) ต่อค่า RMSE อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) อย่างไรก็ตามปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระทั้ง 4 ตัว มีผลต่อค่า RMSE เพียงร้อยละ 0.7 ( $\eta^2 = .007$ ) ซึ่งไม่มีผลในเชิงปฏิบัติมากนัก

เมื่อพิจารณาขนาดอิทธิพลของปฏิสัมพันธ์ระหว่างตัวแปรอิสระ 3 ตัวแปร และ 2 ตัวแปร พบว่าขนาดของปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรมีค่าต่ำกว่าร้อยละ 1.00 ( $\eta^2 < 0.010$ ) ยกเว้นปฏิสัมพันธ์อิทธิพลของสัดส่วนคะแนนข้อสอบรวมกับอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า (ANCHOR x DICH0) ต่อค่า RMSE ที่มีขนาดอิทธิพลร้อยละ 1.80 ( $\eta^2 = .018$ ) และมีปฏิสัมพันธ์ดังภาพ 4.7 ซึ่งค่าเฉลี่ยของขนาดอิทธิพลของปฏิสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 11 แบบ คือ .005 คิดเป็นร้อยละ .50 ด้วยเหตุนี้จึงต้องพิจารณาปฏิสัมพันธ์อิทธิพลของสัดส่วนคะแนนข้อสอบรวมกับอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า และวิเคราะห์อิทธิพลอย่างง่าย (simple-effect) โดยใช้ปฏิสัมพันธ์ดังกล่าวต่อไป

ตาราง 4.1 ผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ

Source	Type III Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
MOSE	4623.072	1	4623.072	3330.010	.000	.431
DS	45.609	1	45.609	32.852	.000	.007
ANCHOR	1966.235	2	983.117	708.142	.000	.244
DICHO	27.726	3	9.242	6.657	.000	.005
MOSE * DS	.331	1	.331	.238	.626	.000
MOSE * ANHCOR	21.819	2	10.910	7.858	.000	.004
MOSE * DICHO	28.671	3	9.557	6.884	.000	.005
DS * ANCHOR	.909	2	.455	.327	.721	.000
DS * DICHO	11.336	3	3.779	2.722	.043	.002
<b>ANCHOR * DICHO</b>	<b>111.722</b>	<b>5</b>	<b>22.344</b>	<b>16.095</b>	<b>.000</b>	<b>.018</b>
MOSE * DS * ANCOR	51.101	2	25.550	18.404	.000	.008
MOSE * DS * DICHO	31.087	3	10.362	7.464	.000	.005
MOSE * ANCHOR * DICHO	21.670	5	4.334	3.122	.008	.004
DS * ANCHOR * DICHO	24.136	5	4.827	3.477	.004	.004
MOSE * DS * ANCOR * DICHO	45.809	5	9.162	6.599	.000	.007
<b>Levene's Test</b>	<b>F-test = 55.385</b>	<b>df1 =43</b>	<b>df2 =4400</b>	<b>p = .000</b>		

หมายเหตุ : \*  $p < 0.05$

## 1.2 ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อความแม่นยำในการปรับเทียบคะแนน (RB) และความถูกต้องในการปรับเทียบคะแนน (CVSE)

การวิเคราะห์ความแม่นยำและความถูกต้องในการปรับเทียบคะแนนจำแนกตามเงื่อนไข ใช้การวิเคราะห์ความแปรปรวนพหุนาม (MANOVA) จากการตรวจสอบความเป็นเอกพันธ์ของความแปรปรวน พบว่า เมทริกซ์ความแปรปรวน-ความแปรปรวนของตัวแปรอิสระทั้ง 4 ตัวระหว่างกลุ่มแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .00$ ) อย่างไรก็ตามเมทริกซ์ค่าความแปรปรวนไม่เท่ากันส่งผลกระทบต่อคุณภาพของสถิติทดสอบน้อย เมื่อกลุ่มตัวอย่างย่อยมีขนาดกัน (Hair et al., 2010) ซึ่งงานวิจัยครั้งนี้กำหนดขนาดตัวอย่างแต่ละกลุ่มเท่ากัน คือ 3,000 คน นอกจากนี้เมื่อตรวจสอบความสัมพันธ์ระหว่างตัวแปรตาม พบว่า ตัวแปรตามทั้ง 2 ตัวมีความสัมพันธ์กันเพียงพอที่จะใช้สถิติวิเคราะห์ตัวแปรพหุนามได้ กล่าวคือ เมทริกซ์สหสัมพันธ์ของเทอมความคลาดเคลื่อนของตัวแปรไม่ใช่เมทริกซ์เอกลักษณะ



### 1.2.1) ผลการทดสอบความแปรปรวนพหุนาม

จากการละเมิดข้อตกลงเบื้องต้นความเป็นเอกพันธ์ของความแปรปรวนระหว่างกลุ่มตัวอย่างที่มีขนาดเท่ากัน จึงควรใช้สถิติ Hottelling T square เพื่อทดสอบความแปรปรวนพหุนาม (Hakstian, Roed, & Lind, 1979 อ้างอิงใน พัชรี จันทร์เพ็ญ, 2550) โดยการทดสอบความแตกต่างระหว่างกลุ่มของค่าความเอียงสัมพัทธ์ (RB) และค่าสัมประสิทธิ์การแปรผันของความคลาดเคลื่อนในการปรับเทียบคะแนน (CVSE) เมื่อรูปแบบกระบวนการ MOSE โครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบร่วม และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า แตกต่างกัน พบว่า มีปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระ 4 ตัวแปรอิสระต่อค่า RB และค่า CVSE โดยเฉลี่ยต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 รายละเอียดดังตาราง 4.2

เมื่อพิจารณาขนาดอิทธิพลของปฏิสัมพันธ์ระหว่างตัวแปรอิสระที่ส่งผลต่อค่า RB และค่า CVSE จากค่า Partial Eta Squared ของสถิติทดสอบ Hottelling T square พบว่า ขนาดอิทธิพล (effect size) ต่ำที่สุดและมากที่สุด คือ ร้อยละ 0.3 (DS\*ANCHOR) และร้อยละ 2.2 (ANCHOR\*DICHO) ตามลำดับ ซึ่งมีขนาดของอิทธิพลของปฏิสัมพันธ์ระหว่างตัวแปรอิสระโดยเฉลี่ย คือ .005 กล่าวคือ ปฏิสัมพันธ์ระหว่างตัวแปรอิสระมีอิทธิพลต่อค่า RMSE โดยเฉลี่ยร้อยละ .10 ( $\eta^2_{interaction} = .010$ )

เมื่อพิจารณาขนาดอิทธิพลของตัวแปรอิสระ (main effect) พบว่า ตัวแปรอิสระที่มีผลต่อค่า RB และค่า CVSE มากที่สุด คือ รูปแบบของกระบวนการ MOSE ด้วยขนาดอิทธิพลร้อยละ 17.00 ( $\eta^2_{MOSE} = .170$ ) รองลงมา คือ สัดส่วนคะแนนข้อสอบร่วม ด้วยขนาดอิทธิพลร้อยละ 9.1 ( $\eta^2_{ANCHOR} = .091$ ) โครงสร้างมิติความสามารถ ด้วยขนาดอิทธิพลร้อยละ 8.5 ( $\eta^2_{DS} = .085$ ) และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า ด้วยขนาดอิทธิพลร้อยละ 0.9 ( $\eta^2_{DICHO} = .009$ )

ตาราง 4.2 ผลการทดสอบความแปรปรวนพหุนาม

Effect	สถิติทดสอบ	Value	F	Hypothesis df	p	Partial Eta Squared
MOSE	Pillai's Trace	.170	451.969 <sup>b</sup>	2.000	.000	.170
	Wilks' Lambda	.830	451.969 <sup>b</sup>	2.000	.000	.170
	Hotelling's Trace	.205	451.969 <sup>b</sup>	2.000	.000	.170
	Roy's Largest Root	.205	451.969 <sup>b</sup>	2.000	.000	.170
DS	Pillai's Trace	.085	205.263 <sup>b</sup>	2.000	.000	.085
	Wilks' Lambda	.915	205.263 <sup>b</sup>	2.000	.000	.085
	Hotelling's Trace	.093	205.263 <sup>b</sup>	2.000	.000	.085
	Roy's Largest Root	.093	205.263 <sup>b</sup>	2.000	.000	.085
ANCHOR	Pillai's Trace	.167	200.475	4.000	.000	.084
	Wilks' Lambda	.833	210.447 <sup>b</sup>	4.000	.000	.087
	Hotelling's Trace	.201	220.456	4.000	.000	.091
	Roy's Largest Root	.200	441.029 <sup>c</sup>	2.000	.000	.167
DIICHO	Pillai's Trace	.017	12.799	6.000	.000	.009
	Wilks' Lambda	.983	12.842 <sup>b</sup>	6.000	.000	.009
	Hotelling's Trace	.018	12.884	6.000	.000	.009
	Roy's Largest Root	.017	24.512 <sup>c</sup>	3.000	.000	.016
MOSE x DS	Pillai's Trace	.014	31.302 <sup>b</sup>	2.000	.000	.014
	Wilks' Lambda	.986	31.302 <sup>b</sup>	2.000	.000	.014
	Hotelling's Trace	.014	31.302 <sup>b</sup>	2.000	.000	.014
	Roy's Largest Root	.014	31.302 <sup>b</sup>	2.000	.000	.014
MOSE x ANCHOR	Pillai's Trace	.023	25.953	4.000	.000	.012
	Wilks' Lambda	.977	26.093 <sup>b</sup>	4.000	.000	.012
	Hotelling's Trace	.024	26.234	4.000	.000	.012
	Roy's Largest Root	.024	51.792 <sup>c</sup>	2.000	.000	.023
MOSE x DICHO	Pillai's Trace	.010	7.549	6.000	.000	.005
	Wilks' Lambda	.990	7.564 <sup>b</sup>	6.000	.000	.005
	Hotelling's Trace	.010	7.578	6.000	.000	.005
	Roy's Largest Root	.010	14.484 <sup>c</sup>	3.000	.000	.010
DS x ANCHOR	Pillai's Trace	.006	6.094	4.000	.000	.003
	Wilks' Lambda	.994	6.100 <sup>b</sup>	4.000	.000	.003
	Hotelling's Trace	.006	6.107	4.000	.000	.003
	Roy's Largest Root	.006	12.109 <sup>c</sup>	2.000	.000	.005
DS x DICHO	Pillai's Trace	.007	5.451	6.000	.000	.004
	Wilks' Lambda	.993	5.460 <sup>b</sup>	6.000	.000	.004
	Hotelling's Trace	.007	5.468	6.000	.000	.004
	Roy's Largest Root	.007	10.811 <sup>c</sup>	3.000	.000	.007

ตาราง 4.2 (ต่อ)

Effect	สถิติทดสอบ	Value	F	Hypothesis df	p	Partial Eta Squared
ANCHOR x DICH0	Pillai's Trace	.044	19.822	10.000	.000	.022
	Wilks' Lambda	.956	20.031 <sup>b</sup>	10.000	.000	.022
	Hotelling's Trace	.046	20.240	10.000	.000	.022
	Roy's Largest Root	.045	39.847 <sup>c</sup>	5.000	.000	.043
MOSE x DS x ANCHOR	Pillai's Trace	.023	25.152	4.000	.000	.011
	Wilks' Lambda	.977	25.285 <sup>b</sup>	4.000	.000	.011
	Hotelling's Trace	.023	25.417	4.000	.000	.011
	Roy's Largest Root	.023	50.258 <sup>c</sup>	2.000	.000	.022
MOSE x DS x DICH0	Pillai's Trace	.011	8.051	6.000	.000	.005
	Wilks' Lambda	.989	8.069 <sup>b</sup>	6.000	.000	.005
	Hotelling's Trace	.011	8.087	6.000	.000	.005
	Roy's Largest Root	.011	15.768 <sup>c</sup>	3.000	.000	.011
MOSE x ANCHOR x DICH0	Pillai's Trace	.027	11.905	10.000	.000	.013
	Wilks' Lambda	.973	11.977 <sup>b</sup>	10.000	.000	.013
	Hotelling's Trace	.027	12.050	10.000	.000	.014
	Roy's Largest Root	.027	23.664 <sup>c</sup>	5.000	.000	.026
DS x ANCHOR x DICH0	Pillai's Trace	.026	11.750	10.000	.000	.013
	Wilks' Lambda	.974	11.809 <sup>b</sup>	10.000	.000	.013
	Hotelling's Trace	.027	11.868	10.000	.000	.013
	Roy's Largest Root	.025	22.391 <sup>c</sup>	5.000	.000	.025
MOSE x DS x ANCHOR x DICH0	Pillai's Trace	.025	11.343	10.000	.000	.013
	Wilks' Lambda	.975	11.400 <sup>b</sup>	10.000	.000	.013
	Hotelling's Trace	.026	11.456	10.000	.000	.013
	Roy's Largest Root	.025	21.735 <sup>c</sup>	5.000	.000	.024
<b>Box's M = 4987.824</b>		F-test =38.379	df1 = 129	df2=9591528.522	p = .000	
<b>Likelihood ratio = .000</b>		Approx. Chi-square = 11988.279		df = 2	p = .000	

หมายเหตุ : \*  $p < 0.05$

### 1.2.2) ผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ

จากการทดสอบความเท่ากันของความแปรปรวนของตัวแปรตามทั้ง 2 ตัว ด้วย Levene's test พบว่า ความแปรปรวนของความลำเอียงสัมพัทธ์ (RB) และสัมประสิทธิ์การแปรผันของความคลาดเคลื่อน (CVSE) มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) เมื่อรูปแบบของกระบวนการ MOSE โครงสร้างมิตិความสามารถ สัดส่วนคะแนนข้อสอบรวม และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าต่างกัน รายละเอียดดังตาราง 4.3

#### 1. การทดสอบอิทธิพลระหว่างตัวแปรอิสระที่มีต่อความแม่นยำ

จากการทดสอบอิทธิพลระหว่างตัวแปรอิสระที่มีต่อความแม่นยำในการเปรียบเทียบคะแนน (ค่า RB) พบว่า มีปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระทั้ง 4 ตัวแปร ได้แก่ รูปแบบของกระบวนการ MOSE โครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบรวม และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า ต่อค่า RB อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ 2.20 ( $\eta^2 = .022$ ) ซึ่งมีผลในเชิงปฏิบัติ ด้วยเหตุนี้จึงไม่จำเป็นต้องพิจารณาปฏิสัมพันธ์ระหว่างตัวแปรในระดับ 3 และ 2 ตัวแปร รวมถึงอิทธิพลหลักของตัวแปรอิสระแต่ละตัว ดังนั้นการวิเคราะห์ความแม่นยำในการเปรียบเทียบคะแนนของตัวแปรอิสระแต่ละตัว จึงต้องพิจารณาตัวแปรอื่น ๆ ร่วมด้วย

#### 2. การทดสอบอิทธิพลระหว่างตัวแปรอิสระที่มีต่อความถูกต้อง

จากการทดสอบอิทธิพลระหว่างตัวแปรอิสระที่มีต่อความถูกต้องในการเปรียบเทียบคะแนน (ค่า CVSE) เมื่อพิจารณาปฏิสัมพันธ์ในระดับ 4 ตัวแปร พบว่า ไม่มีปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระทั้ง 4 ตัวที่ส่งผลต่อค่า CVSE ( $p = .311$ ) เมื่อพิจารณาปฏิสัมพันธ์ในระดับ 3 ตัวแปร พบว่า มีปฏิสัมพันธ์อิทธิพลระหว่างโครงสร้างมิติความสามารถ (DS) สัดส่วนคะแนนข้อสอบรวม (ANCHOR) และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า (TYPE) ต่อค่า CVSE อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .030$ ) อย่างไรก็ตามขนาดอิทธิพลของปฏิสัมพันธ์ดังกล่าวมีเพียงร้อยละ 0.30 ( $\eta^2 = .003$ ) ซึ่งไม่มีผลในเชิงปฏิบัติ และเมื่อพิจารณาปฏิสัมพันธ์ในระดับ 2 ตัวแปร พบว่า มีปฏิสัมพันธ์อิทธิพลระหว่างรูปแบบของกระบวนการ MOSE กับสัดส่วนคะแนนข้อสอบรวมต่อค่า CVSE ( $p = .001$ ) อย่างไรก็ตามขนาดอิทธิพลของปฏิสัมพันธ์ดังกล่าวมีเพียงร้อยละ 0.30 ( $\eta^2 = .003$ ) ซึ่งไม่มีผลในเชิงปฏิบัติ ดังนั้นจึงสามารถพิจารณาอิทธิพลหลักของตัวแปรอิสระ

เมื่อพิจารณาอิทธิพลหลักของตัวแปรอิสระ (main effect) พบว่า ตัวแปรอิสระที่มีผลต่อเวกเตอร์ค่าเฉลี่ยของค่า CVSE ระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ได้แก่ รูปแบบของกระบวนการ MOSE โครงสร้างมิติความสามารถ และสัดส่วนคะแนนข้อสอบรวม ( $p = .000$ ) โดยมีขนาดอิทธิพลร้อยละ 6.90, .30 และ .90 ตามลำดับ ( $\eta^2 = .069, .003$  และ .009 ตามลำดับ) ขณะที่อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าไม่มีผลต่อค่า CVSE อย่างมีนัยสำคัญ ( $p = .150$ )

ตาราง 4.3 ผลการทดสอบอิทธิพลระหว่างตัวแปรอิสระ

Source		Type III Sum of Squares	df	Mean Square	F	p	Partial Eta Squared
MOSE	RB	2338774.085	1	2338774.085	325.881	.000	.069
	CVSE	39695.406	1	39695.406	292.727	.000	.062
DS	RB	2896043.987	1	2896043.987	403.531	.000	.084
	CVSE	1981.519	1	1981.519	14.612	.000	.003
ANCOR	RB	4690842.239	2	2345421.120	326.808	.000	.129
	CVSE	5337.905	2	2668.952	19.682	.000	.009
DICHO	RB	512782.524	3	170927.508	23.817	.000	.016
	CVSE	720.714	3	240.238	1.772	.150	.001
METHOD x DS	RB	436774.834	1	436774.834	60.860	.000	.014
	CVSE	197.595	1	197.595	1.457	.227	.000
METHOD x ANCOR	RB	742695.666	2	371347.833	51.743	.000	.023
	CVSE	1840.561	2	920.281	6.786	.001	.003
METHOD x DICHO	RB	266993.216	3	88997.739	12.401	.000	.008
	CVSE	309.767	3	103.256	.761	.516	.001
DS x ANCOR	RB	168940.267	2	84470.133	11.770	.000	.005
	CVSE	751.784	2	375.892	2.772	.063	.001
DS x DICHO	RB	207552.969	3	69184.323	9.640	.000	.007
	CVSE	53.994	3	17.998	.133	.941	.000
ANCOR x DICHO	RB	1391595.148	5	278319.030	38.781	.000	.042
	CVSE	1081.718	5	216.344	1.595	.158	.002
METHOD x DS x ANCOR	RB	666152.456	2	333076.228	46.410	.000	.021
	CVSE	183.715	2	91.858	.677	.508	.000
METHOD x DS x DICHO	RB	319564.813	3	106521.604	14.843	.000	.010
	CVSE	202.374	3	67.458	.497	.684	.000
METHOD x ANCOR x DICHO	RB	805817.167	5	161163.433	22.456	.000	.025
	CVSE	434.756	5	86.951	.641	.668	.001
DS x ANCOR x DICHO	RB	797467.403	5	159493.481	22.224	.000	.025
	CVSE	1674.810	5	334.962	2.470	.030	.003
METHOD x DS x ANCOR x DICHO	RB	700227.720	5	140045.544	19.514	.000	.022
	CVSE	807.709	5	161.542	1.191	.311	.001
Levene's test	RB	F-test = 126.679	df1 = 43	df2 = 4400	p = .000		
	CVSE	F-test = 5.968	df1 = 43	df2 = 4400	p = .000		

หมายเหตุ : \*  $p < 0.05$

## ตอนที่ 2 ประสิทธิภาพระหว่างกระบวนการ CMOSE กับกระบวนการ SMOSE

การนำเสนอผลการเปรียบเทียบประสิทธิภาพระหว่างกระบวนการ CMOSE กับกระบวนการ SMOSE (วัตถุประสงค์ข้อ 1) แบ่งเป็น 3 ส่วนหลัก คือ (1) ผลด้านประสิทธิภาพโดยรวม (RMSE) (2) ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB) และ (3) ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE) รายละเอียดดังนี้

### 2.1 ผลด้านประสิทธิภาพโดยรวม (RMSE)

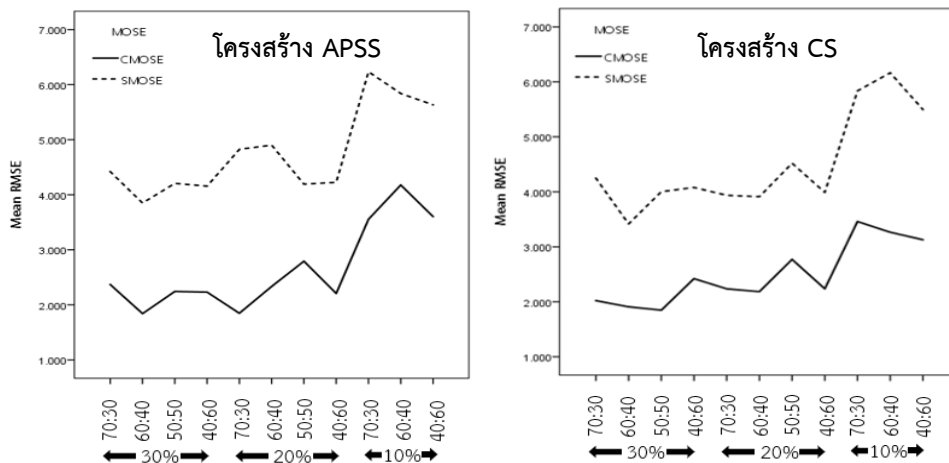
เมื่อพิจารณาประสิทธิภาพโดยรวมของการปรับเทียบคะแนนสำหรับแบบสอบทั้ง 2 โครงสร้าง (ภาพ 4.1) พบว่า ค่า RMSE เฉลี่ยของกระบวนการ CMOSE ต่ำกว่ากระบวนการ SMOSE ทุกเงื่อนไข เช่นเดียวกับผลการกระจายของค่า RMSE ที่พบว่า ค่า RMSE ของกระบวนการ CMOSE มีการกระจายน้อยกว่ากระบวนการ SMOSE ในทุกเงื่อนไข ดังภาพ 4.2-4.3

เมื่อพิจารณากระบวนการ CMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) จากตาราง 4.4 พบว่า มีค่า RMSE เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ด้วยอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 ( $\overline{RMSE} = 1.840$ ) ส่วนกระบวนการ CMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า มีค่า RMSE เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ด้วยอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 50:50 ( $\overline{RMSE} = 1.848$ )

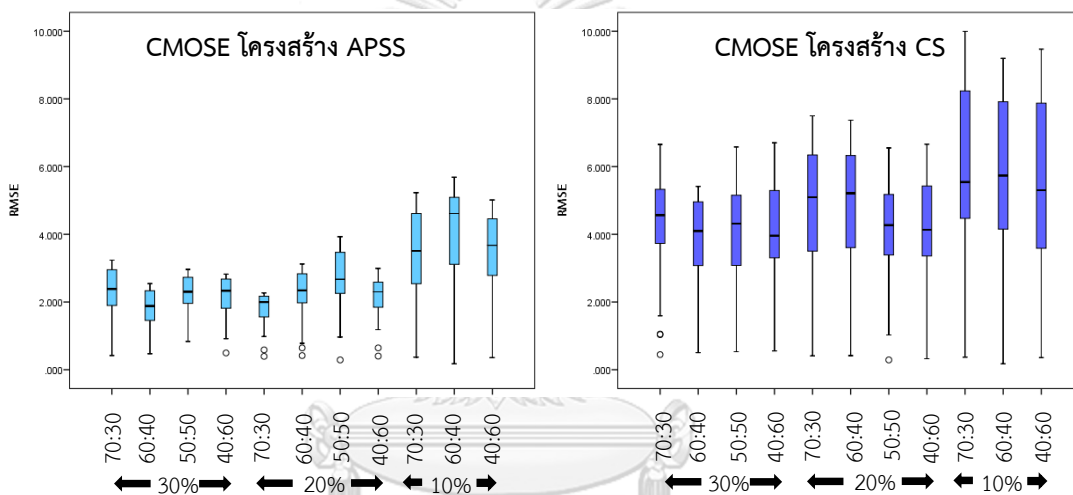
เมื่อพิจารณากระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า ค่า RMSE เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ด้วยอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 ( $\overline{RMSE} = 3.850$ ) ส่วนกระบวนการ SMOSE สำหรับโครงสร้างแบบซับซ้อน (CS) พบว่า มีค่า RMSE เฉลี่ยต่ำที่สุดเมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ด้วยอัตราส่วนคะแนนข้อสอบที่ตรวจให้คะแนนสองค่า คือ 60:40 ( $\overline{RMSE} = 3.417$ )

เมื่อพิจารณาความแตกต่างของค่า RMSE (ค่า d) ระหว่างกระบวนการ MOSE ทั้ง 2 แบบ พบว่า เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10% ค่า RMSE เฉลี่ยของกระบวนการ SMOSE ส่วนใหญ่แตกต่างจากกระบวนการ CMOSE น้อยกว่าสัดส่วนคะแนนข้อสอบรวมอื่น ๆ

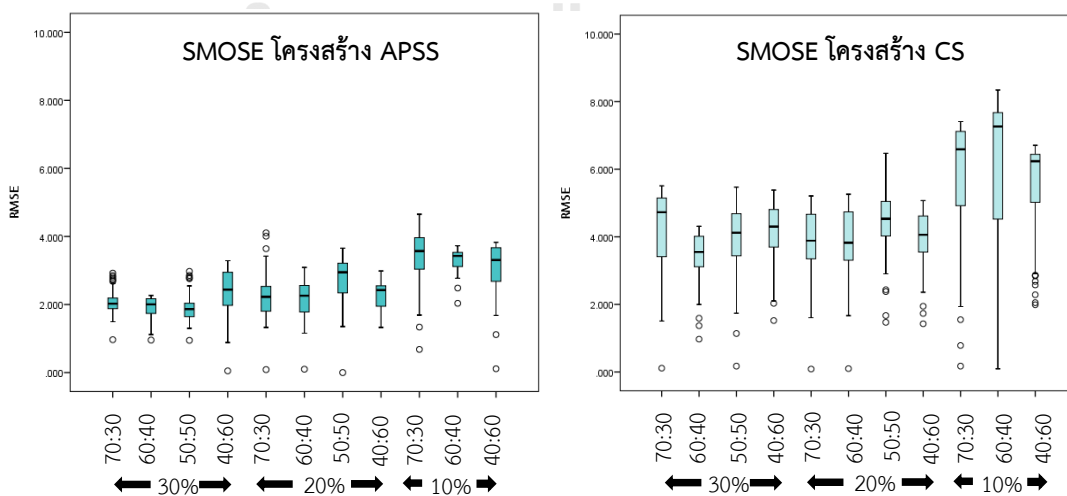
ผลของสถิติบรรยายดังกล่าวสอดคล้องกับผลการวิเคราะห์ความแปรปรวน (ANOVA) จากตาราง 4.1 ที่พบว่า เวกเตอร์ค่าเฉลี่ยของประสิทธิภาพโดยรวมระหว่างรูปแบบของกระบวนการ MOSE มีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ 43.10 สะท้อนให้เห็นว่า กระบวนการ CMOSE มีประสิทธิภาพโดยรวมมากกว่ากระบวนการ SMOSE



ภาพ 4.1 ค่า RMSE ของกระบวนการ MOSE สำหรับโครงสร้าง APSS และ CS



ภาพ 4.2 แผนภูมิ Box plot ค่า RMSE ของกระบวนการ CMOSE โครงสร้าง APSS และ CS



ภาพ 4.3 แผนภูมิ Box plot ค่า RMSE ของกระบวนการ SMOSE โครงสร้าง APSS และ CS

ตาราง 4.4 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของประสิทธิภาพโดยรวมของการปรับเทียบคะแนน

โครงสร้าง	ข้อสอบ รวม	ข้อสอบรวม แต่ละประเภท	ค่าเฉลี่ย		ส่วนเบี่ยงเบนมาตรฐาน		$d = \frac{ \mu_s - \mu_c }{SD_c}$
			CMOSE	SMOSE	CMOSE	SMOSE	
APSS	30%	70:30	2.369	4.420	.643	1.463	3.190
		60:40	1.840	3.850	.525	1.288	3.829
		50:50	2.243	4.206	.570	1.454	3.444
		40:60	2.231	4.156	.511	1.461	3.767
	20%	70:30	1.847	4.825	.393	1.850	7.578
		60:40	2.334	4.902	.566	1.751	4.537
		50:50	2.793	4.192	.760	1.447	1.841
		40:60	2.207	4.225	.471	1.473	4.285
	10%	70:30	3.554	6.234	1.150	2.191	2.330
		60:40	4.178	5.839	1.194	2.320	1.391
		40:60	3.604	5.633	1.019	2.433	1.991
CS	30%	70:30	2.021	4.249	.356	1.117	6.258
		60:40	1.908	3.417	.338	0.756	4.464
		50:50	1.848	4.003	.398	0.983	5.415
		40:60	2.421	4.079	.570	0.995	2.909
	20%	70:30	2.236	3.937	.551	0.887	3.087
		60:40	2.185	3.912	.486	0.937	3.553
		50:50	2.772	4.520	.573	1.179	3.051
		40:60	2.236	3.990	.469	0.812	3.740
	10%	70:30	3.461	5.836	.803	1.718	2.958
		60:40	3.266	6.165	.506	2.059	5.729
		40:60	3.129	5.496	.653	1.424	3.625



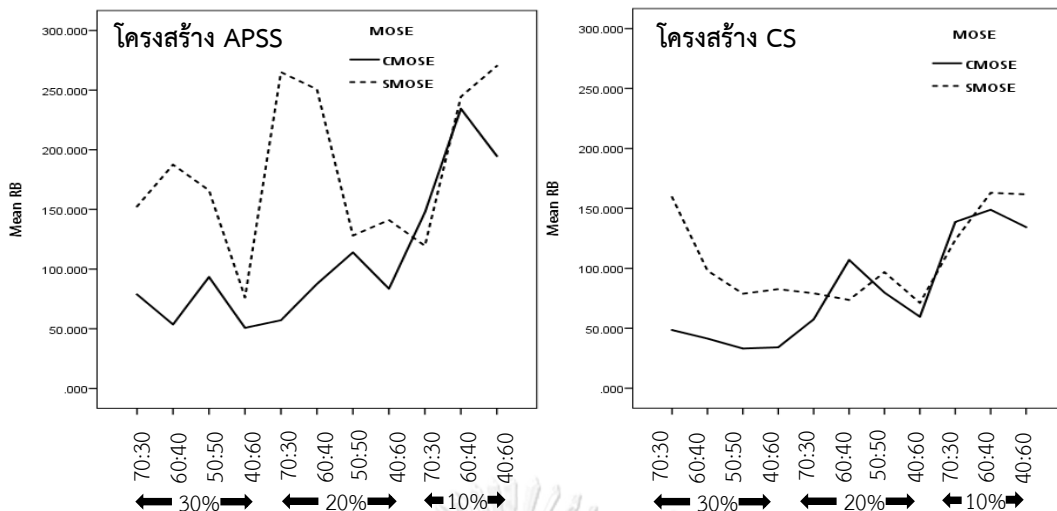
## 2.2 ผลด้านความแม่นยำในการเปรียบเทียบคะแนน (RB)

เมื่อพิจารณาความแม่นยำในการเปรียบเทียบคะแนนสำหรับแบบสอบทั้ง 2 โครงสร้าง (ภาพ 4.4) พบว่า ค่าความลำเอียงสัมพัทธ์ (RB) เฉลี่ยของกระบวนการ CMOSE ส่วนใหญ่ต่ำกว่ากระบวนการ SMOSE โดยมีเพียง 3 เงื่อนไขที่กระบวนการ SMOSE มีค่า RB ต่ำกว่ากระบวนการ CMOSE ได้แก่ (1) โครงสร้างที่ไม่ซับซ้อน (APSS) ที่มีสัดส่วนคะแนนข้อสอบรวม คือ 10% และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 (2) โครงสร้างที่ซับซ้อน (CS) สัดส่วนคะแนนข้อสอบรวม คือ 20% และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 และ (3) โครงสร้างที่ซับซ้อน (CS) สัดส่วนคะแนนข้อสอบรวม คือ 10% และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 ขณะที่เมื่อพิจารณาการกระจายของค่า RB พบว่า ค่า RB ของกระบวนการ CMOSE มีการกระจายน้อยกว่ากระบวนการ SMOSE ทุกเงื่อนไข ดังภาพ ภาพ 4.5-4.6

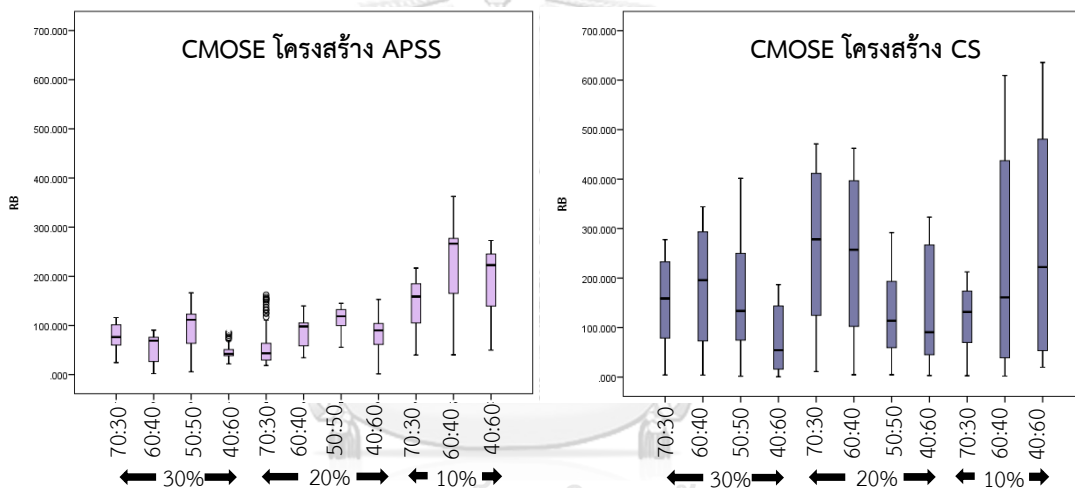
เมื่อพิจารณากระบวนการ CMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) จากตาราง 4.5 พบว่า ค่า RB เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ที่มีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 40:60 ( $\overline{RB} = 50.752$ ) ส่วนกระบวนการ CMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า ค่า RB เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ที่มีอัตราส่วนคะแนนข้อสอบที่ตรวจให้คะแนนสองค่า คือ 50:50 ( $\overline{RB} = 33.067$ )

เมื่อพิจารณากระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า ค่า RB เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ที่มีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 40:60 ( $\overline{RB} = 76.276$ ) ส่วนกระบวนการ SMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า มีค่า RB เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% ด้วยอัตราส่วนคะแนนข้อสอบที่ตรวจให้คะแนนสองค่า คือ 40:60 ( $\overline{RB} = 73.616$ )

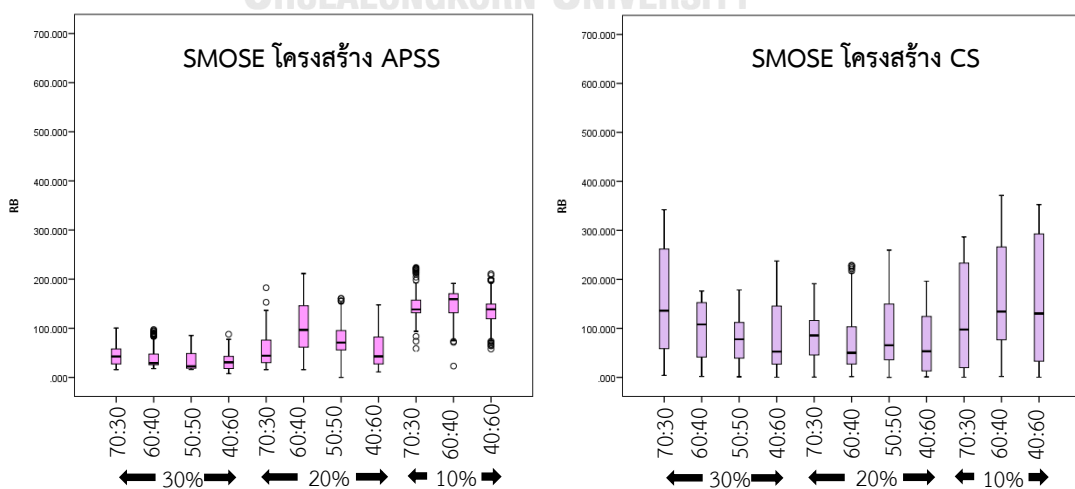
เมื่อพิจารณาผลการวิเคราะห์ความแปรปรวนพหุนาม (MONOVA) จากตาราง 4.3 พบว่า มีปฏิสัมพันธ์ระหว่างตัวแปรอิสระทั้ง 4 ตัวต่อค่า RB อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยมีขนาดอิทธิพลร้อยละ 2.2 ซึ่งค่า RB ของกระบวนการ CMOSE ส่วนใหญ่ต่ำกว่ากระบวนการ SMOSE มีเพียงบางเงื่อนไขที่มีผลการวิเคราะห์แตกต่างไปดังที่ได้กล่าวไว้ในข้างต้น อย่างไรก็ตามกระบวนการ CMOSE มีการกระจายน้อยกว่าทุกเงื่อนไข สะท้อนให้เห็นว่า กระบวนการ CMOSE มีความแม่นยำมากกว่ากระบวนการ SMOSE



ภาพ 4.4 ค่า RB ของกระบวนการ MOSE สำหรับโครงสร้าง APSS และ CS



ภาพ 4.5 แผนภูมิ Box plot ค่า RB ของกระบวนการ CMOSE โครงสร้าง APSS และ CS



ภาพ 4.6 แผนภูมิ Box plot ค่า RB ของกระบวนการ SMOSE โครงสร้าง APSS และ CS

ตาราง 4.5 ค่าสถิติบรรยายของความแม่นยำและความถูกต้องในการเปรียบเทียบคะแนน

ค่าสถิติ	สัดส่วนข้อสอบรวม	สัดส่วน DICH0	โครงสร้างไม่ซับซ้อน (APSS)				โครงสร้างซับซ้อน (CS)			
			CMOSE		SMOSE		CMOSE		SMOSE	
			RB	CVSE	RB	CVSE	RB	CVSE	RB	CVSE
ค่าเฉลี่ย	30%	70:30	78.753	7.627	152.491	12.419	48.488	7.182	159.368	13.174
		60:40	53.577	6.792	187.519	12.092	41.356	6.532	98.072	10.695
		50:50	93.385	5.967	166.001	12.490	33.067	7.108	78.800	13.112
		40:60	50.752	6.996	76.276	11.978	34.112	8.037	82.620	12.317
	20%	70:30	57.125	5.905	265.010	10.508	57.438	9.139	79.126	14.981
		60:40	87.542	7.828	250.634	12.013	107.002	7.528	73.616	13.676
		50:50	113.988	7.948	128.087	12.227	79.738	9.458	96.789	14.158
		40:60	83.541	7.799	141.129	12.158	59.530	8.418	71.167	14.284
	10%	70:30	147.429	8.554	119.471	19.874	138.699	9.532	123.747	17.772
		60:40	234.403	7.243	244.484	12.746	148.856	8.548	162.993	18.709
		40:60	194.700	7.842	270.244	12.532	134.315	8.716	161.689	17.104
	ส่วนเบี่ยงเบนมาตรฐาน	30%	70:30	23.818	10.969	84.893	10.792	26.833	8.738	109.853
60:40			27.849	9.220	111.272	12.620	24.508	7.331	56.562	9.296
50:50			36.766	7.787	120.765	11.588	20.431	9.441	47.877	11.598
40:60			21.948	7.379	66.600	9.784	26.389	8.800	68.199	10.907
20%		70:30	39.789	6.205	148.847	8.268	36.720	14.912	42.764	20.076
		60:40	29.502	9.849	150.781	9.634	55.237	11.352	66.096	12.765
		50:50	21.573	9.642	86.151	10.657	2.382	12.917	80.722	12.045
		40:60	33.080	11.053	114.677	11.495	41.206	11.472	65.682	14.357
10%		70:30	46.824	9.407	61.883	20.428	51.952	12.899	104.185	13.865
		60:40	77.664	5.510	210.583	7.668	30.717	12.264	115.313	16.527
		40:60	65.266	8.926	222.297	10.923	35.678	10.745	127.955	16.290
ค่าต่ำสุด		30%	70:30	24.578	.000	4.317	.000	15.811	0.000	3.982
	60:40		1.929	.000	3.953	.000	17.701	0.000	1.734	.000
	50:50		5.775	.690	1.722	.000	16.042	0.000	1.181	.000
	40:60		21.842	.673	0.603	.000	7.648	0.000	0.135	.000
	20%	70:30	18.446	.000	11.053	.000	15.811	0.000	0.345	.000
		60:40	34.250	.000	4.565	.000	15.811	0.000	1.440	.000
		50:50	55.571	.000	4.621	.000	0.000	0.000	0.000	.000
		40:60	1.596	.000	3.063	.000	11.167	0.000	0.939	.000
	10%	70:30	39.936	.000	2.636	.000	0.197	0.000	0.251	.000
		60:40	40.360	.000	2.066	.000	23.083	0.000	1.759	.000
		40:60	49.968	.000	19.814	.000	29.315	0.000	0.202	.000

ตาราง 4.5 (ต่อ)

ค่าสถิติ	สัดส่วนข้อสอบรวม	สัดส่วน DICO	โครงสร้างไม่ซับซ้อน (APSS)				โครงสร้างซับซ้อน (CS)			
			CMOSE		SMOSE		CMOSE		SMOSE	
			RB	CVSE	RB	CVSE	RB	CVSE	RB	CVSE
ค่าสถิติ	30%	70:30	116.057	57.718	277.736	53.090	176.458	34.614	342.105	105.338
		60:40	90.456	43.631	344.110	62.942	97.463	29.263	176.140	40.394
		50:50	166.354	45.903	401.596	72.701	85.220	38.786	178.503	54.817
		40:60	116.197	35.299	186.920	56.945	186.771	40.151	237.343	56.563
ค่าสูงสุด	20%	70:30	162.335	26.547	471.079	41.606	182.537	74.599	191.134	143.856
		60:40	139.672	45.144	462.213	44.440	211.339	47.331	228.801	63.746
		50:50	144.986	50.681	292.041	71.243	10.875	58.656	259.759	54.245
		40:60	153.095	56.733	323.256	57.327	147.577	47.971	196.208	67.416
ค่าสถิติ	10%	70:30	216.772	53.642	212.533	139.256	238.939	53.247	286.770	58.587
		60:40	362.418	22.982	609.439	25.571	191.284	53.072	371.334	118.405
		40:60	272.806	50.820	635.694	69.930	211.038	49.585	352.591	114.482

### 2.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE)

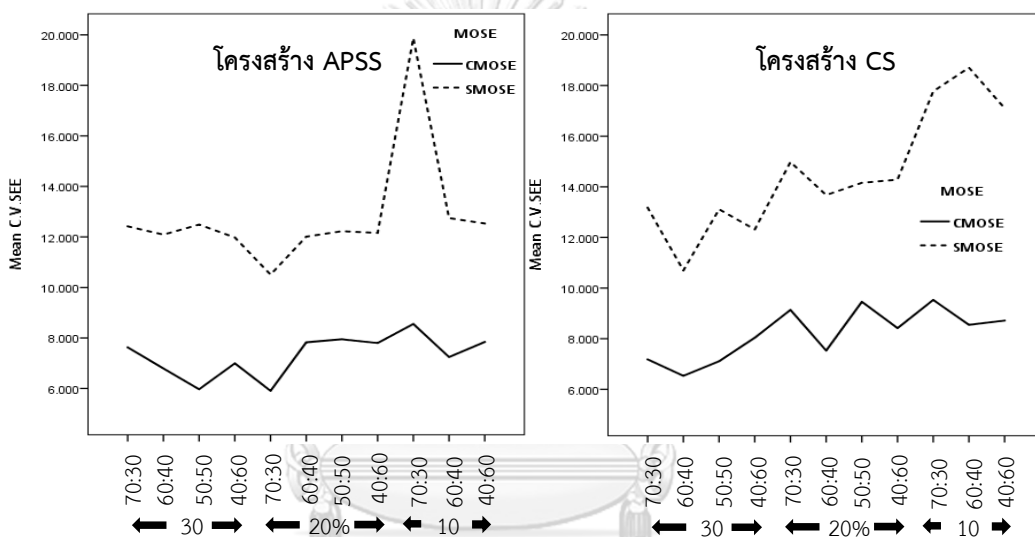
เมื่อพิจารณาความถูกต้องในการปรับเทียบคะแนนสำหรับแบบสอบทั้ง 2 โครงสร้าง (ภาพ 4.7) พบว่า ค่า CVSE เฉลี่ยของกระบวนการ CMOSE ต่ำกว่ากระบวนการ SMOSE ทุกเงื่อนไข เช่นเดียวกับผลการกระจายของค่า CVSE ที่พบว่า ค่า CVSE ของกระบวนการ CMOSE มีการกระจายน้อยกว่ากระบวนการ SMOSE ทุกเงื่อนไข ดังภาพ 4.8-4.9

เมื่อพิจารณากระบวนการ CMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) จากตาราง 4.5 พบว่า ค่า CVSE เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% ที่มีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 ( $\overline{CVSE} = 5.905$ ) ส่วนกระบวนการ CMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า ค่า CVSE เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ที่มีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 ( $\overline{CVSE} = 6.532$ )

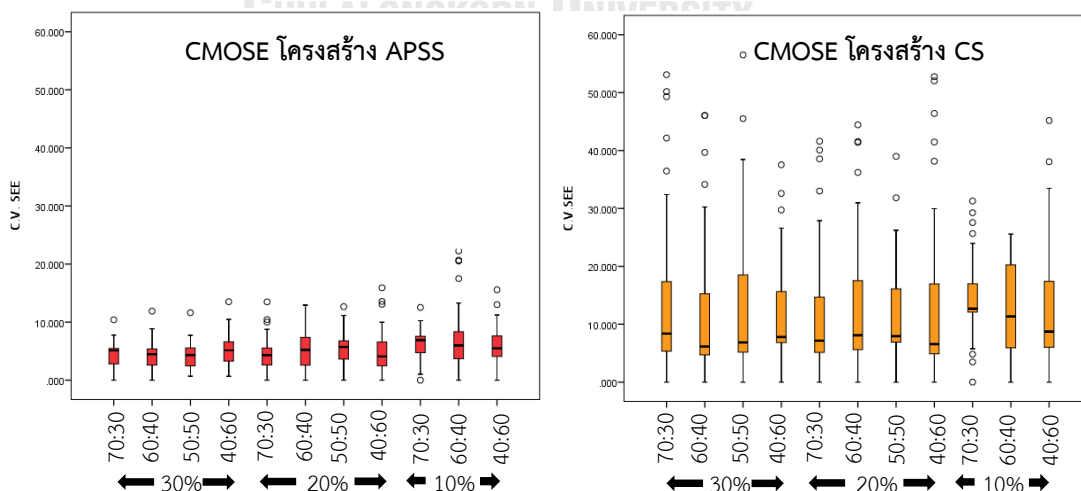
กระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า มีค่า CVSE เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% ที่มีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 ( $\overline{CVSE} = 10.508$ ) ส่วนกระบวนการ SMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า มีค่า CVSE เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ที่มีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 ( $\overline{CVSE} = 10.695$ )

ผลของสถิติบรรยายดังกล่าวสอดคล้องกับผลการวิเคราะห์ความแปรปรวนพหุนาม (MANOVA) จากตาราง 4.3 ที่พบว่า กระบวนการ CMOSE มีเวกเตอร์ค่าเฉลี่ยของค่า CVSE ต่ำกว่า กระบวนการ SMOSE อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ 6.2 สะท้อนให้เห็นว่า กระบวนการ CMOSE มีความถูกต้องในการปรับเทียบคะแนนมากกว่า กระบวนการ SMOSE

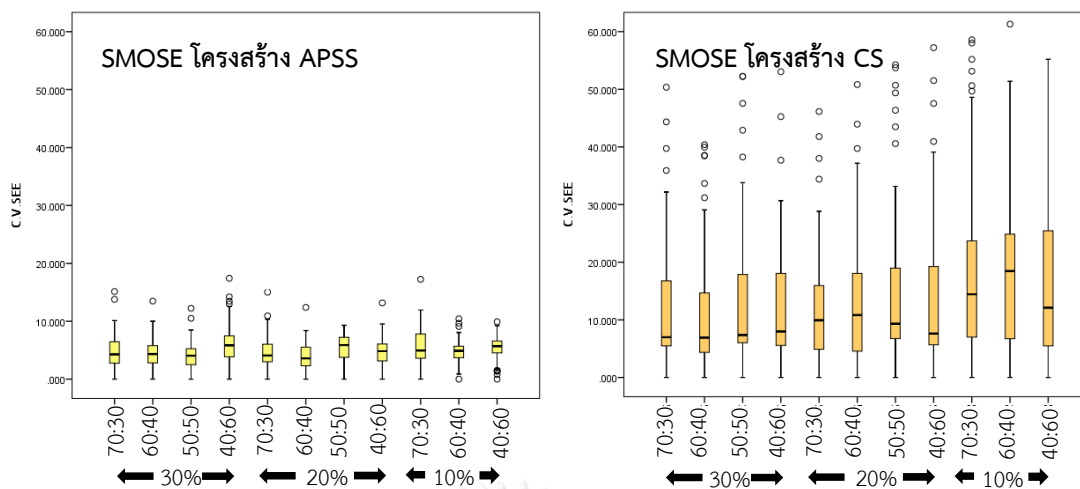
จากผลการเปรียบเทียบประสิทธิภาพในการปรับเทียบคะแนนระหว่างกระบวนการ CMOSE และ SMOSE สามารถสรุปได้ว่า กระบวนการ CMOSE มีประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องมากกว่ากระบวนการ SMOSE



ภาพ 4.7 ค่า CVSE ของกระบวนการ MOSE สำหรับโครงสร้าง APSS และ CS



ภาพ 4.8 แผนภูมิ Box plot ค่า CVSE ของกระบวนการ CMOSE โครงสร้าง APSS และ CS



ภาพ 4.9 แผนภูมิ Box plot ค่า CVSE ของกระบวนการ SMOSE โครงสร้าง APSS และ CS

### ตอนที่ 3 ประสิทธิภาพของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบรวมแตกต่างกัน

การนำเสนอผลการศึกษาประสิทธิภาพของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบรวมแตกต่างกัน คือ สัดส่วน 30%, 20% และ 10% (วัตถุประสงค์ข้อ 2) แบ่งเป็น 3 ส่วนหลัก คือ (1) ผลด้านประสิทธิภาพโดยรวม (RMSE) (2) ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB) และ (3) ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE) รายละเอียดดังนี้

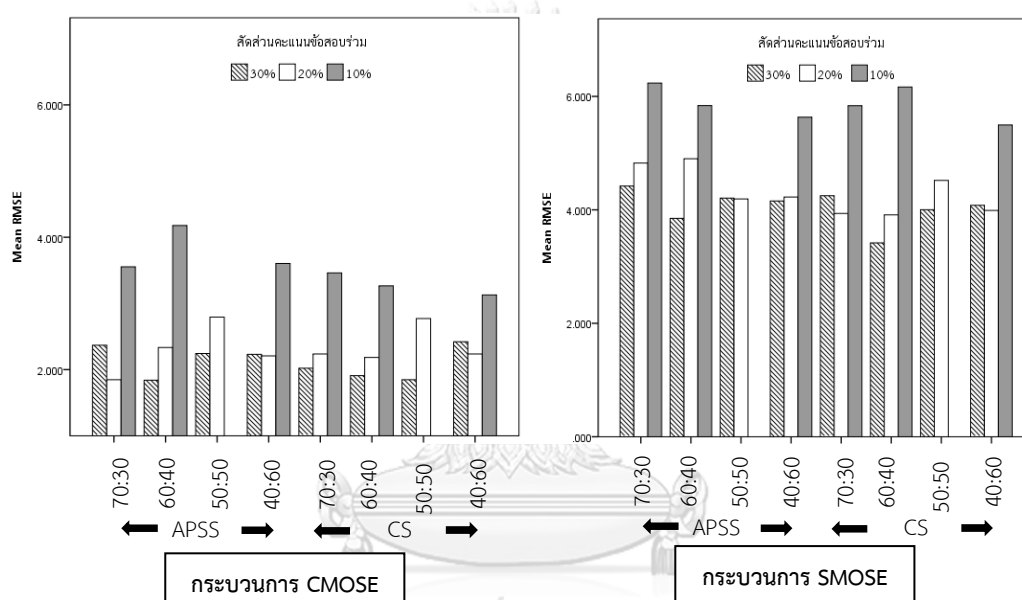
#### 3.1 ผลด้านประสิทธิภาพโดยรวม (RMSE)

เมื่อพิจารณาประสิทธิภาพโดยรวมสำหรับกระบวนการ *CMOSE* สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) (ภาพ 4.10) พบว่า ค่า RMSE เฉลี่ยลดลง เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้น (ถ้าอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 และ 50:50) โดยค่า RMSE เฉลี่ยของสัดส่วนคะแนนข้อสอบรวม 10% มากที่สุดทุกเงื่อนไข ส่วนกระบวนการ *CMOSE* สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า ค่า RMSE เฉลี่ยลดลง เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้น (ยกเว้นอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 40:60 ซึ่งมีค่า RMSE เฉลี่ยต่ำที่สุดเมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20%, 30% และ 10% ตามลำดับ) โดยค่า RMSE เฉลี่ยของสัดส่วนคะแนนข้อสอบรวม 10% มากที่สุดในทุกเงื่อนไข (สถิติเชิงบรรยายโดยละเอียดของค่า RMSE ดังตาราง 4.7)

เมื่อพิจารณาประสิทธิภาพโดยรวมสำหรับกระบวนการ *SMOSE* สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า ค่า RMSE เฉลี่ยลดลง เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้นทุกเงื่อนไข ส่วนกระบวนการ *SMOSE* สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า ค่า RMSE เฉลี่ยลดลง เมื่อสัดส่วน

คะแนนข้อสอบรวมเพิ่มขึ้น (ถ้าอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 และ 50:50)

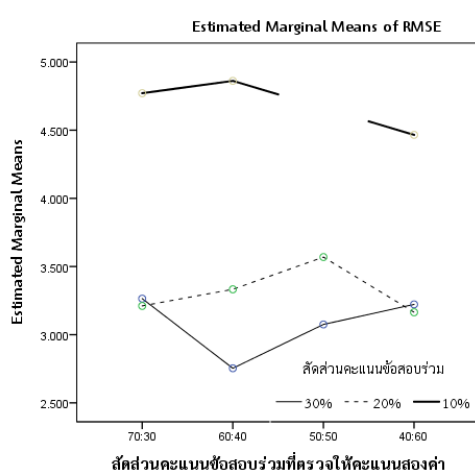
จากสถิติเชิงบรรยายข้างต้น ผลการวิจัยส่วนใหญ่ พบว่า กระบวนการ MOSE มีค่า RMSE เฉลี่ยลดลง เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้น และผลการวิจัยบางส่วน พบว่า กระบวนการ MOSE มีค่า RMSE ต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20%, 30% และ 10% ตามลำดับ โดยที่สัดส่วนคะแนนข้อสอบรวม 10% มีค่า RMSE มากที่สุดทุกเงื่อนไข สอดคล้องกับผลการกระจายของค่า RMSE ที่พบว่า สัดส่วนคะแนนข้อสอบรวม 10% มีการกระจายมากที่สุด ขณะที่ สัดส่วนคะแนนข้อสอบรวม 30% และ 20% มีการกระจายใกล้เคียงกัน ดังภาพ 4.2-4.3



ภาพ 4.10 แผนภูมิแท่งเปรียบเทียบค่า RMSE สำหรับสัดส่วนคะแนนข้อสอบรวม

เมื่อพิจารณาผลการวิเคราะห์ความแปรปรวน (ANOVA) ที่พบว่า มีปฏิสัมพันธ์อิทธิพลของ สัดส่วนคะแนนข้อสอบรวมและ อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า (ANCHOR x DICHO) ต่อค่า RMSE ด้วยขนาดอิทธิพล ร้อยละ 1.80 ( $\eta^2 = .018$ ) จึงวิเคราะห์อิทธิพลอย่างง่าย (simple-effect) ของสัดส่วนคะแนนข้อสอบรวม และจากแผนภาพปฏิสัมพันธ์ 4.11 พบว่า ผลการวิจัย แบ่งเป็น 2 แบบ คือ (1) กรณีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนค่า คือ 60:40 และ 50:50 พบว่า ค่า RMSE มากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10%, 30% และ 20% ตามลำดับ สอดคล้องกับผลการทดสอบอิทธิพลอย่างง่าย ที่พบว่า กระบวนการ MOSE มีค่า RMSE เฉลี่ยลดลง เมื่อ สัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้น อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ถ้าอัตราส่วนคะแนน ข้อสอบรวมที่ตรวจให้คะแนนค่า คือ 60:40 และ 50:50 ดังตาราง 4.6

(2) กรณีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนค่า คือ 70:30 และ 40:60 พบว่าค่า RMSE มากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10%, 30% และ 20% ตามลำดับ อย่างไรก็ตามก็ตามความแตกต่างของค่า RMSE สำหรับสัดส่วนคะแนนข้อสอบรวม 20% และ 30% น้อยมากสอดคล้องกับผลการทดสอบอิทธิพลอย่างง่าย ที่พบว่า ค่า RMSE ของสัดส่วนคะแนนข้อสอบรวม 10% มากกว่าสัดส่วนคะแนนข้อสอบรวม 30% และ 20% ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ขณะที่ค่า RMSE ของสัดส่วนคะแนนข้อสอบรวม 30% มากกว่าสัดส่วน 20% อย่างไม่มีนัยสำคัญทางสถิติ



ภาพ 4.11 ปฏิสัมพันธ์อิทธิพลระหว่าง ANCHOR X DICHO ต่อค่า RMSE

จากข้อค้นพบข้างต้น ผลการวิจัยส่วนใหญ่

พบว่า กระบวนการ MOSE มีประสิทธิภาพโดยรวมมากขึ้น เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้น และผลการวิจัยบางส่วน พบว่า กระบวนการ MOSE มีประสิทธิภาพมากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20%, 30% และ 10% ตามลำดับ สะท้อนให้เห็นว่า การปรับเทียบคะแนนด้วยสัดส่วนคะแนนข้อสอบรวม 30% และ 20% ทำให้มีประสิทธิภาพโดยรวมสูง ขณะที่สัดส่วนคะแนนข้อสอบรวม 10% ทำให้การปรับเทียบคะแนนมีประสิทธิภาพโดยรวมต่ำที่สุด

ตาราง 4.6 ผลวิเคราะห์อิทธิพลอย่างง่ายของตัวแปรสัดส่วนข้อสอบรวม

DICHOTOMOUS	ANCHOR	Mean Difference (I-J)	Std. Error	p-value	95% Confidence Interval for Difference	
					Lower Bound	Upper Bound
70:30	30%	.053	.083	1.000	-.145	.252
	10%	-1.507*	.083	.000	-1.705	-1.308
	20%	-1.560*	.083	.000	-1.759	-1.362
60:40	30%	-.579*	.083	.000	-.778	-.381
	10%	-2.108*	.083	.000	-2.306	-1.909
	20%	-1.529*	.083	.000	-1.727	-1.330
50:50	30%	-.494*	.083	.000	-.657	-.332
	10%	-1.243*	.083	.000	-1.442	-1.045
40:60	30%	.058	.083	1.000	-.141	.256
	20%	-1.301*	.083	.000	-1.500	-1.103

หมายเหตุ : \*  $p < 0.05$



ตาราง 4.7 ค่าสถิติบรรยายของประสิทธิภาพโดยรวมของการปรับเทียบคะแนน

โครงสร้าง	ข้อสอบรวม	ข้อสอบรวมแต่ละประเภท	CMOSE					SMOSE				
			M	SD	MIN	MAX	VAR	M	SD	MIN	MAX	VAR
แบบไม่ซับซ้อน (APSS)	30%	70:30	2.369	.643	.418	3.238	.413	4.420	1.463	.447	6.657	2.141
		60:40	1.840	.525	.470	2.544	.275	3.850	1.288	.504	5.416	1.660
		50:50	2.243	.570	.837	2.964	.325	4.206	1.454	.532	6.580	2.113
		40:60	2.231	.511	.495	2.824	.261	4.156	1.461	.554	6.703	2.134
	20%	70:30	1.847	.393	.394	2.270	.154	4.825	1.850	.410	7.501	3.421
		60:40	2.334	.566	.417	3.123	.320	4.902	1.751	.416	7.371	3.066
		50:50	2.793	.760	.289	3.928	.578	4.192	1.447	.289	6.553	2.093
		40:60	2.207	.471	.402	2.994	.222	4.225	1.473	.323	6.662	2.170
	10%	70:30	3.554	1.150	.370	5.228	1.324	6.234	2.191	.370	9.996	4.801
		60:40	4.178	1.194	.176	5.686	1.426	5.839	2.320	.176	9.196	5.384
		40:60	3.604	1.019	.358	5.016	1.039	5.633	2.433	.358	9.470	5.919
	แบบซับซ้อน (CS)	30%	70:30	2.021	.356	.113	2.924	.127	4.249	1.117	.113	5.509
60:40			1.908	.338	.175	2.264	.114	3.417	0.756	.175	4.313	.571
50:50			1.848	.398	.175	2.974	.158	4.003	0.983	.175	5.469	.967
40:60			2.421	.570	.050	3.290	.325	4.079	0.995	.050	5.384	.989
20%		70:30	2.236	.551	.088	4.107	.304	3.937	0.887	.088	5.209	.787
		60:40	2.185	.486	.100	3.091	.236	3.912	0.937	.100	5.261	.879
		50:50	2.772	.573	.000	3.654	.329	4.520	1.179	.000	6.474	1.390
		40:60	2.236	.469	.094	2.987	.220	3.990	0.812	.094	5.074	.660
10%		70:30	3.461	.803	.175	4.653	.645	5.836	1.718	.175	7.412	2.950
		60:40	3.266	.506	.094	3.729	.256	6.165	2.059	.094	8.342	4.240
		40:60	3.129	.653	.113	3.829	.426	5.496	1.424	0.113	6.708	2.028

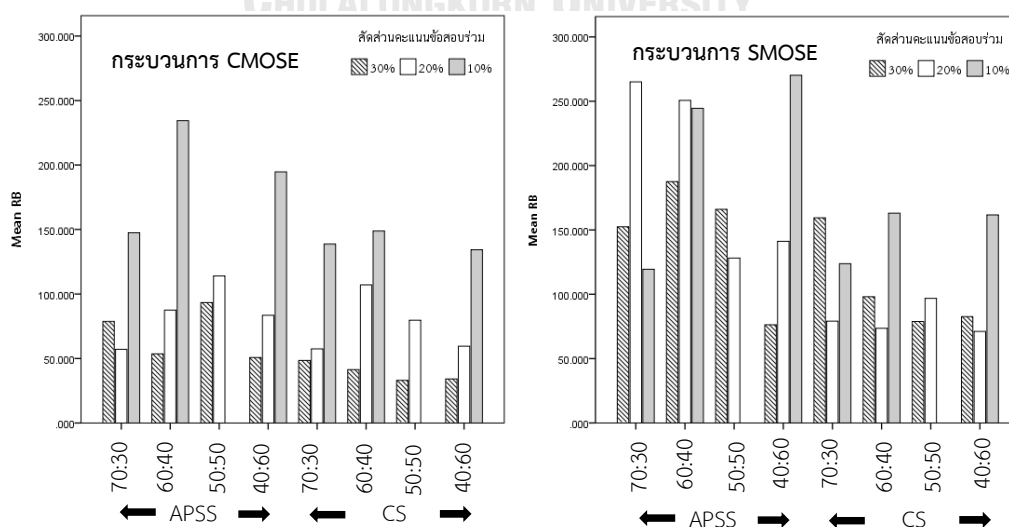
### 3.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB)

เมื่อพิจารณาความแม่นยำในการปรับเทียบคะแนนสำหรับกระบวนการ CMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) (ภาพ 4.12) พบว่า ค่า RB เฉลี่ยลดลง เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้นทุกเงื่อนไข (ยกเว้นอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30) โดยค่า RB เฉลี่ยของสัดส่วนคะแนนข้อสอบรวม 10% มากที่สุดทุกเงื่อนไข ส่วนกระบวนการ CMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า ค่า RB เฉลี่ยลดลง เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้นทุกเงื่อนไข โดยค่า RB เฉลี่ยของสัดส่วนคะแนนข้อสอบรวม 10% มากที่สุดทุกเงื่อนไข

เมื่อพิจารณาความแม่นยำในการปรับเทียบคะแนนสำหรับกระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) (ภาพ 4.12) พบว่า ส่วนมากมีค่า RB เฉลี่ยมากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% (ถ้าอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 และ 60:40)

ขณะที่ส่วนมากมีค่า RB เฉลี่ยต่ำที่สุด เมื่ออัตราส่วนคะแนนข้อสอบรวม คือ 30% (ถ้าอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 และ 40:60) โดยมีข้อค้นพบที่แตกต่าง คือ กรณีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 พบว่า มีค่า RB มีค่าต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10%, 30% และ 20% ตามลำดับ ส่วนกระบวนการ SMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) (ภาพ 4.12) พบว่า ค่า RB เฉลี่ยมากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10%, 30% และ 20% ตามลำดับ (ยกเว้นเมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสอง คือ 50:50 ซึ่งพบว่า มีค่า RB เฉลี่ยต่ำลง เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้น) และมีค่า RB เฉลี่ยต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20%

จากผลการวิจัยข้างต้นจะเห็นว่า มีข้อค้นพบสำคัญ 2 แบบได้แก่ (1) ค่า RB เฉลี่ยลดลง เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้นทุกเงื่อนไข และ (2) ค่า RB มีค่ามากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10%, 30% และ 20% และมีข้อค้นพบที่แตกต่าง คือ มีค่า RB มีค่าต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10%, 30% และ 20% ตามลำดับ (ถ้าอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30) สะท้อนให้เห็นว่าการพิจารณาค่า RB ในสัดส่วนคะแนนข้อสอบรวมแต่ละแบบควรพิจารณาร่วมกับตัวแปรอิสระอื่น ๆ สอดคล้องกับผลการวิเคราะห์ความแปรปรวนพหุคูณ (MANOVA) ที่พบว่า มีปฏิสัมพันธ์อิทธิพลระหว่างรูปแบบของกระบวนการ MOSE กับตัวแปรอิสระอื่น ๆ ต่อค่า RB อย่างมีนัยสำคัญที่ระดับ .05 ด้วยขนาดอิทธิพลร้อยละ 2.2 ดังตาราง 4.3 อย่างไรก็ตามเมื่อพิจารณาการกระจายของค่า RB พบว่า ค่า RB เฉลี่ยของสัดส่วนคะแนนข้อสอบรวม 10% ส่วนใหญ่มีการกระจายมากที่สุด ดังภาพ 4.5-4.6 ด้วยเหตุนี้จึงมีแนวโน้มว่า สัดส่วนคะแนนข้อสอบรวม 30% และ 20% มีความแม่นยำในการเปรียบเทียบคะแนน ขณะที่ควรหลีกเลี่ยงสัดส่วนคะแนนข้อสอบรวม 10%



ภาพ 4.12 แผนภูมิแท่งเปรียบเทียบค่า RB สำหรับสัดส่วนคะแนนข้อสอบรวม

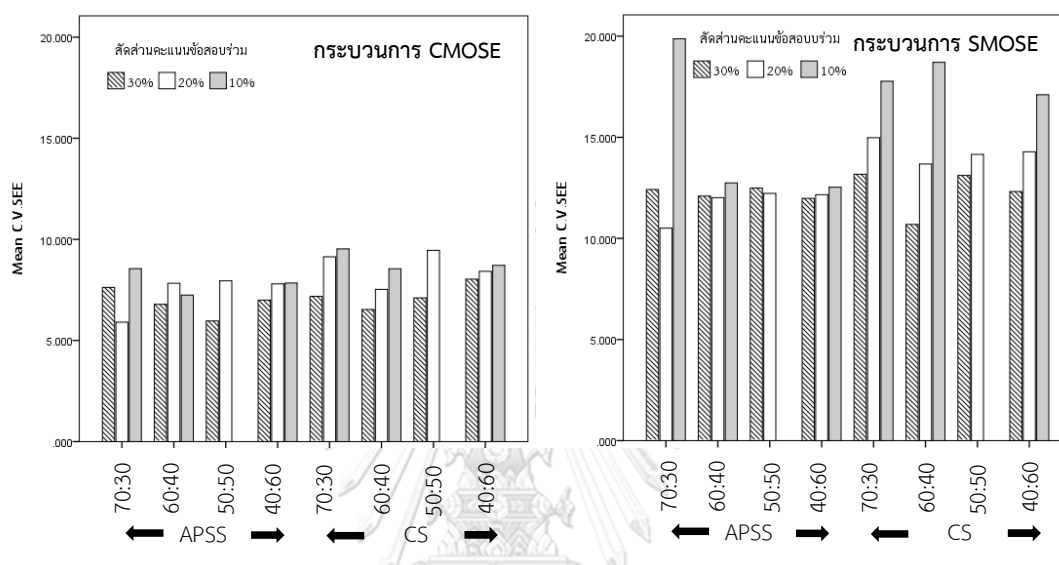
### 3.3 ผลด้านความถูกต้องในการเปรียบเทียบคะแนน (CVSE)

เมื่อพิจารณาความถูกต้องในการเปรียบเทียบคะแนนสำหรับกระบวนการ CMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) (ภาพ 4.13) พบว่า มีค่า CVSE เฉลี่ยต่ำลง เมื่อสัดส่วนคะแนนข้อสอบร่วมเพิ่มขึ้น (ถ้าอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 50:50 และ 40:60) โดยค่า CVSE เฉลี่ยของสัดส่วนคะแนนข้อสอบร่วม 30% ต่ำที่สุด (ยกเว้นอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 70:30 ซึ่งพบว่า สัดส่วนคะแนนข้อสอบร่วม 20% มีค่า CVSE เฉลี่ยต่ำที่สุด) ค่า CVSE เฉลี่ยของสัดส่วนคะแนนข้อสอบร่วม 10% มากที่สุด (ยกเว้นอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 60:40 ซึ่งพบว่า สัดส่วนคะแนนข้อสอบร่วม 20% มีค่า CVSE เฉลี่ยมากที่สุด) ส่วนกระบวนการ CMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า มีค่า CVSE เฉลี่ยต่ำลง เมื่อสัดส่วนคะแนนข้อสอบร่วมเพิ่มขึ้นทุกเงื่อนไข

เมื่อพิจารณาความถูกต้องในการเปรียบเทียบคะแนนสำหรับกระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า มีค่า CVSE มากที่สุด เมื่อสัดส่วนคะแนนข้อสอบร่วม คือ 20%, 30% และ 10% ตามลำดับ (ยกเว้นเมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 40:60 ซึ่งพบว่า มีค่า CVSE ต่ำลง เมื่อสัดส่วนคะแนนข้อสอบร่วมเพิ่มขึ้น) ส่วนกระบวนการ SMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า CVSE เฉลี่ยต่ำลง เมื่อสัดส่วนคะแนนข้อสอบร่วมเพิ่มขึ้นทุกเงื่อนไข

เมื่อพิจารณาความถูกต้องในภาพรวม พบว่า ค่า CVSE เฉลี่ยส่วนใหญ่มากที่สุด เมื่อสัดส่วนคะแนนข้อสอบร่วม คือ 10% อย่างไรก็ตาม ค่า CVSE เฉลี่ยในแต่ละเงื่อนไขค่อนข้างใกล้เคียงกัน (โดยเฉพาะกระบวนการ CMOSE) เช่นเดียวกับการกระจายของค่า CVSE ดังภาพ 4.7-4.8 ที่พบว่า สัดส่วนคะแนนข้อสอบร่วมรูปแบบต่าง ๆ มีการกระจายใกล้เคียงกัน มีเพียงกรณีกระบวนการ SMOSE สำหรับโครงสร้าง CS ที่พบว่า ค่า CVSE ของสัดส่วนคะแนนข้อสอบร่วม 10% มีการกระจายมากที่สุดอย่างชัดเจน สอดคล้องกับผลวิเคราะห์อิทธิพลหลักของตัวแปรอิสระ (main effect) ที่พบว่า สัดส่วนคะแนนข้อสอบร่วมมีผลต่อเวกเตอร์ค่าเฉลี่ยของค่า CVSE ระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ .90 ( $\eta^2 = .009$ ) ซึ่งอาจไม่มีผลในเชิงปฏิบัติมากนัก ดังตาราง 4.3 ทำให้แนวโน้มของผลด้านความถูกต้องสำหรับเงื่อนไขสัดส่วนคะแนนข้อสอบร่วมชัดเจนน้อยกว่าผลด้านประสิทธิภาพโดยรวม (ค่า RMSE มีขนาดอิทธิพลร้อยละ 24.4) ด้วยเหตุนี้ จึงมีแนวโน้มว่า สัดส่วนคะแนนข้อสอบร่วม 30% และ 20% มีความแม่นยำในการเปรียบเทียบคะแนน ขณะที่ควรหลีกเลี่ยงสัดส่วนคะแนนข้อสอบร่วม 10%

จากผลการศึกษาประสิทธิภาพในการปรับเทียบของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบร่วมแตกต่างกัน มีแนวโน้มสัดส่วนคะแนนข้อสอบร่วม 30% และ 20% มีประสิทธิภาพในการปรับเทียบคะแนน และควรหลีกเลี่ยงการใช้สัดส่วนคะแนนข้อสอบร่วม 10% โดยสัดส่วนคะแนนข้อสอบร่วมมีผลต่อประสิทธิภาพโดยรวมมากที่สุด (ค่า RMSE) รองลงมา คือ ความถูกต้องและความแม่นยำในการปรับเทียบคะแนน ตามลำดับ (ค่า RB และค่า CVSE)



ภาพ 4.13 แผนภูมิแท่งเปรียบเทียบค่า CVSE สำหรับสัดส่วนคะแนนข้อสอบร่วม

#### ตอนที่ 4 ประสิทธิภาพของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแตกต่างกัน

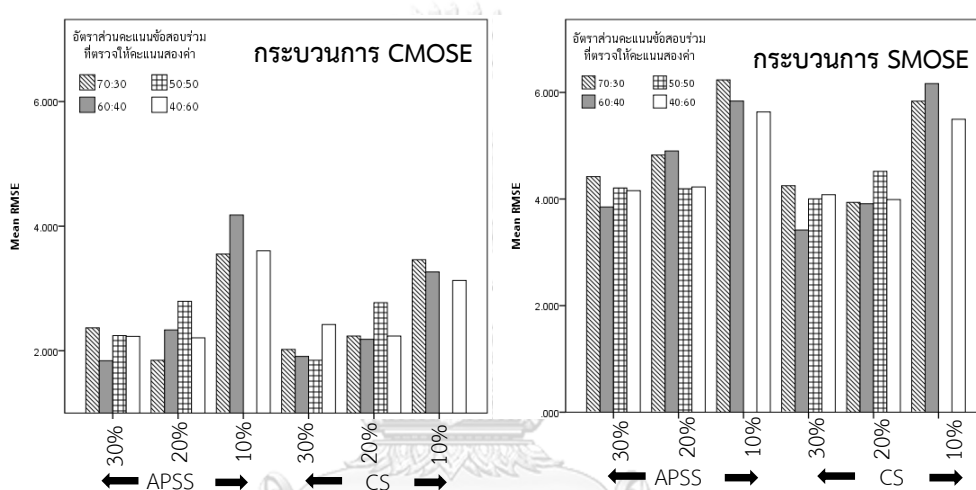
การนำเสนอผลการศึกษาประสิทธิภาพโดยรวมของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแตกต่างกัน คือ อัตราส่วนฯ 40:60, 50:50, 60:40 และ 70:30 (วัตถุประสงค์ข้อ 3) แบ่งเป็น 3 ส่วนหลัก คือ (1) ผลด้านประสิทธิภาพโดยรวม (RMSE) (2) ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB) และ (3) ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE) รายละเอียดดังนี้

##### 4.1 ผลด้านประสิทธิภาพโดยรวม (RMSE)

เมื่อพิจารณาประสิทธิภาพโดยรวมสำหรับกระบวนการ CMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) (ภาพ 4.14) พบว่า มีค่า RMSE เฉลี่ยส่วนใหญ่ต่ำที่สุด เมื่ออัตราส่วนคะแนนข้อสอบร่วมคะแนนตรวจให้คะแนนสองค่า คือ 70:30 ส่วนกระบวนการ CMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า ไม่มีแนวโน้มของค่า RMSE เฉลี่ยมากที่สุดหรือต่ำที่สุดอย่างชัดเจน

เมื่อพิจารณาความถูกต้องในการปรับเทียบคะแนนสำหรับกระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า มีค่า RMSE เฉลี่ยส่วนใหญ่มากที่สุด เมื่อสัดส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 70:30 ส่วนกระบวนการ SMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า ค่า RMSE ส่วนใหญ่ต่ำที่สุด เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า 60:40

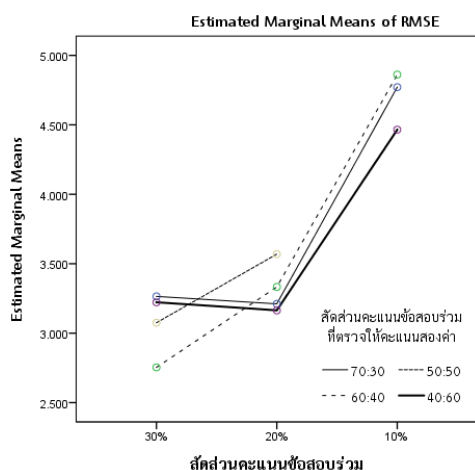
ผลการวิเคราะห์สถิติเชิงบรรยายข้างต้น จะเห็นว่า ค่าเฉลี่ย RMSE ไม่มีแนวโน้มชัดเจนว่าเพิ่มขึ้นหรือลดลง เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าเป็นอย่างไร สอดคล้องกับการกระจายของค่า RMSE ที่พบว่า ค่า RMSE ของอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่ามีการกระจายในแต่ละเงื่อนไขใกล้เคียงกันมาก ดังภาพ 4.2-4.3



ภาพ 4.14 แผนภูมิแท่งเปรียบเทียบค่า RMSE สำหรับสัดส่วนคะแนนข้อสอบร่วม

เมื่อพิจารณาผลการวิเคราะห์ความแปรปรวน (ANOVA) พบว่า มีปฏิสัมพันธ์อิทธิพลของสัดส่วนคะแนนข้อสอบร่วมและ อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า (ANCHOR x DICHO) ต่อค่า RMSE ด้วยขนาดอิทธิพลร้อยละ 1.80 ( $\eta^2 = .018$ ) และเมื่อพิจารณาแผนภาพปฏิสัมพันธ์ 4.15 กรณีสัดส่วนคะแนนข้อสอบร่วม 30% พบว่า มีค่า RMSE มากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 70:30, 40:60, 50:50 และ 60:40 ตามลำดับ โดยค่า RMSE ระหว่างสัดส่วนฯ 70:30 50:50 และ 40:60 แตกต่างกันน้อย สอดคล้องกับผลการวิเคราะห์อิทธิพลอย่างง่าย ที่พบว่า (1) ค่า RMSE ของอัตราส่วนฯ 70:30 มากกว่าอัตราส่วนฯ 60:40 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ (3) ค่า RMSE ของอัตราส่วนฯ 70:30 มากกว่าอัตราส่วนฯ 50:50 และ 40:60 ตามลำดับ อย่างไม่มีนัยสำคัญ ดังตาราง 4.8

เมื่อพิจารณาสัดส่วนคะแนนข้อสอบรวม 20% พบว่า มีค่า RMSE มากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 50:50, 60:40, 70:30 และ 40:60 ตามลำดับ โดยค่า RMSE ระหว่างอัตราส่วนฯ 60:40, 70:30 และ 40:60 มีความแตกต่างกันน้อย ดังภาพ 4.14 สอดคล้องกับผลการวิเคราะห์หือทธิพลอย่างง่าย ที่พบว่า (1) ค่า RMSE ของอัตราส่วนฯ 50:50 มากกว่าอัตราส่วนฯ 70:30, 40:60 และ 60:40 ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ (2) ค่า RMSE ของอัตราส่วนฯ 40:60 ต่ำกว่าอัตราส่วนฯ 60:40 และ 70:30 ตามลำดับอย่างไม่มีนัยสำคัญทางสถิติ



ภาพ 4.15 ปฏิสัมพันธ์อิทธิพลของ ANCHOR X DICHO ต่อค่า RMSE

เมื่อพิจารณาสัดส่วนคะแนนข้อสอบรวม 10% พบว่า ค่า RMSE มากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40, 70:30 และ 40:60 ตามลำดับ โดยค่า RMSE ระหว่างอัตราส่วนฯ 60:40 กับ 70:30 แตกต่างกันน้อย สอดคล้องกับผลการวิเคราะห์หือทธิพลอย่างง่าย ที่พบว่า (1) ค่า RMSE ของอัตราส่วนฯ 40:60 ต่ำกว่าอัตราส่วนฯ 60:40 และ 70:30 ตามลำดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ (2) ค่า RMSE ของอัตราส่วนฯ 60:40 มากกว่าอัตราส่วนฯ 70:30 อย่างไม่มีนัยสำคัญ ดังตาราง 4.8

ผลการวิเคราะห์ข้างต้นสะท้อนให้เห็นว่า อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าไม่ได้มีแนวโน้มของประสิทธิภาพโดยรวมชัดเจนว่าเพิ่มขึ้นหรือลดลงอย่างไร เนื่องจากค่า RMSE แตกต่างกันในแต่ละสัดส่วนข้อสอบรวม ซึ่งอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่ามีอิทธิพลต่อค่าประสิทธิภาพโดยรวมเพียงร้อยละ .50 อย่างไรก็ตามเมื่อพิจารณาในภาพรวมมีแนวโน้มว่า อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 40:60 มีประสิทธิภาพโดยรวมมากกว่าอัตราส่วนอื่น ๆ

ตาราง 4.8 ผลวิเคราะห์หือทธิพลอย่างง่ายของอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า

ANCHOR	DICHOTOMOUS	Mean Difference (I-J)	Std. Error	p-value	95% Confidence Interval for Difference		
					Lower Bound	Upper Bound	
30%	70:30	60:40	.511*	.083	.000	.292	.729
		50:50	.189	.083	.135	-.029	.408
		40:60	.042	.083	1.000	-.176	.261
	60:40	50:50	-.321*	.083	.001	-.540	-.103
		40:60	-.468*	.083	.000	-.687	-.249
	50:50	40:60	-.147	.083	.459	-.366	.072
20%	70:30	60:40	-.122	.083	.850	-.341	.097
		50:50	-.358*	.083	.000	-.577	-.139
		40:60	.047	.083	1.000	-.172	.265
	60:40	50:50	-.236*	.083	.026	-.455	-.018
		40:60	.168	.083	.253	-.050	.387
	50:50	40:60	.405*	.083	.000	.186	.624
10%	70:30	60:40	-.090	.083	.827	-.289	.108
		40:60	.306*	.083	.001	.107	.504
	60:40	40:60	.396*	.083	.000	.198	.595

หมายเหตุ : \*  $p < 0.05$

#### 4.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB)

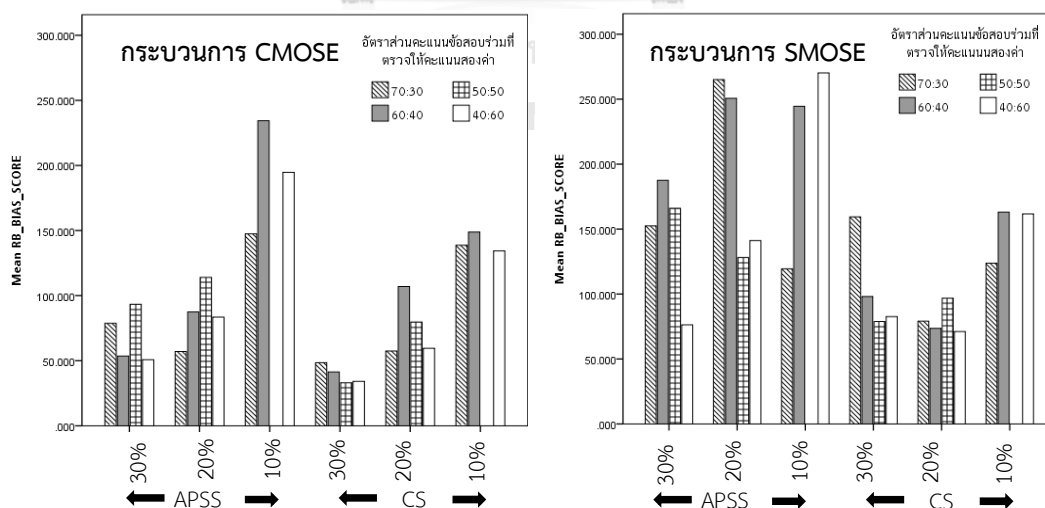
เมื่อพิจารณาความแม่นยำในการปรับเทียบคะแนนสำหรับกระบวนการ CMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) (ภาพ 4.16) พบว่า มีค่า RB เฉลี่ยมากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าเท่ากับหลายค่า คือ 50:50 และมีค่า RB เฉลี่ยต่ำที่สุด เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 (ยกเว้นสัดส่วนคะแนนข้อสอบรวม 30% ซึ่งพบว่า อัตราส่วนฯ คือ 40:60 มีค่า RB เฉลี่ยต่ำที่สุด) ส่วนกระบวนการ CMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า มีค่า RB เฉลี่ยมากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 (ยกเว้นสัดส่วนคะแนนข้อสอบรวม 30% ซึ่งพบว่า มีค่า RB เฉลี่ยมากที่สุด เมื่ออัตราส่วนฯ คือ 70:30) ขณะค่า RB เฉลี่ยต่ำที่สุด แตกต่างไปตามเงื่อนไข

เมื่อพิจารณาความแม่นยำในการปรับเทียบคะแนนสำหรับกระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า ไม่มีแนวโน้มของค่า RMSE เฉลี่ยมากที่สุดหรือต่ำที่สุดอย่างชัดเจน โดยอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่ามากกว่า (70:30 และ 60:40) มีค่า RB

เฉลี่ยมากกว่าอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนหลายค่ามากกว่า (40:60) เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% และ 20% ส่วนกระบวนการ SMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า พบว่า ไม่มีแนวโน้มของค่า RB เฉลี่ยมากที่สุดหรือต่ำที่สุดอย่างชัดเจน

ผลการวิเคราะห์สถิติเชิงบรรยายข้างต้น จะเห็นว่า ค่าเฉลี่ย RB ไม่มีแนวโน้มชัดเจนว่า เพิ่มขึ้นหรือลดลง เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าเป็นอย่างไร สอดคล้องกับการกระจายของค่า RB ที่พบว่า ค่า RB ของอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่ามีการกระจายแตกต่างกันในแต่ละเงื่อนไข แต่ไม่สามารถระบุแนวโน้มว่าอัตราส่วน แบบใดที่มีการกระจายมากที่สุดหรือน้อยที่สุด ดังภาพ 4.5-4.6 สอดคล้องกับผลการวิเคราะห์ความแปรปรวนพหุนาม (MANOVA) ที่พบว่า มีปฏิสัมพันธ์อิทธิพลระหว่างรูปแบบของกระบวนการ MOSE กับตัวแปรอิสระอื่น ๆ ต่อค่า RB อย่างมีนัยสำคัญที่ระดับ .05 ด้วยขนาดอิทธิพลร้อยละ 2.2 ดังตาราง 4.3 จึงต้องเปรียบเทียบค่า RB ของอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าสำหรับกระบวนการ MOSE ทั้ง 2 แบบ ร่วมกับตัวแปรอื่น ๆ อย่างไรก็ตามสามารถระบุแนวโน้มค่า RB เฉลี่ยมากที่สุดและต่ำที่สุดได้ชัดเจนกว่าค่า RMSE เนื่องจากอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนน มีขนาดอิทธิพลต่อค่า RB (ร้อยละ 1.6) มากกว่าค่า RMSE (ร้อยละ .05)

ผลการวิเคราะห์ข้างต้นสะท้อนให้เห็นว่า อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าไม่ได้มีแนวโน้มของความแม่นยำในการเปรียบเทียบคะแนนชัดเจนว่าเพิ่มขึ้นหรือลดลงอย่างไรก็ตามเมื่อพิจารณาในภาพรวม มีแนวโน้มว่าอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 40:60 มีความแม่นยำในการเปรียบเทียบคะแนนมากกว่าอัตราส่วนอื่น ๆ



ภาพ 4.16 แผนภูมิแท่งเปรียบเทียบค่า RB สำหรับอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า



#### 4.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE)

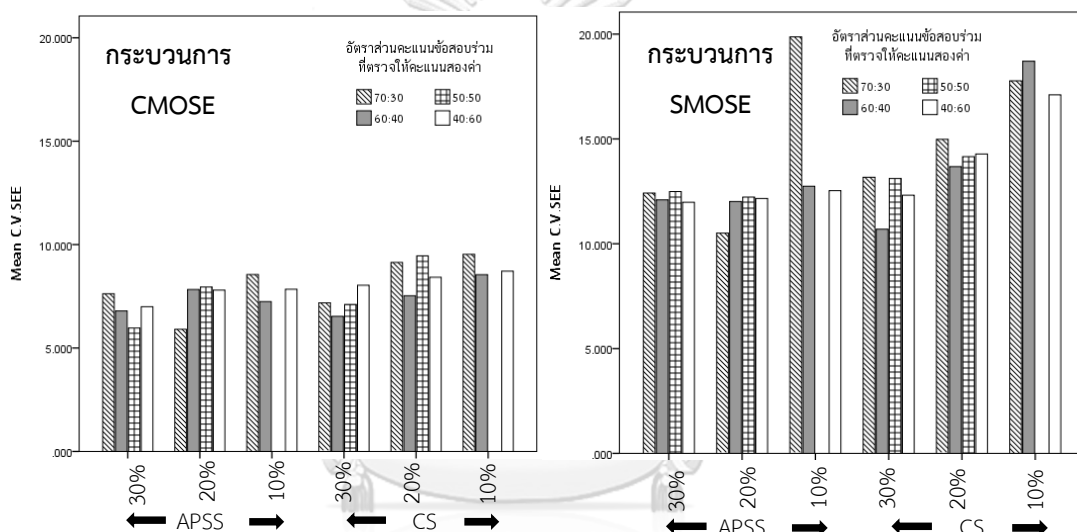
เมื่อพิจารณาความถูกต้องในการปรับเทียบคะแนนสำหรับกระบวนการ CMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) (ภาพ 4.17) พบว่า มีค่า CVSE เฉลี่ยมากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 70:30 (ยกเว้นสัดส่วนคะแนนข้อสอบร่วม 20% ซึ่งพบว่าอัตราส่วนฯ 50:50 มีค่า CVSE เฉลี่ยมากที่สุด) ขณะที่ค่า CVSE เฉลี่ยต่ำที่สุดแตกต่างกันตามเงื่อนไข ส่วนกระบวนการ CMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า ส่วนใหญ่มีค่า CVSE เฉลี่ยต่ำที่สุด เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 60:40 ขณะที่ค่า CVSE เฉลี่ยมากที่สุดแตกต่างกันตามเงื่อนไข

เมื่อพิจารณาความแม่นยำในการปรับเทียบคะแนนสำหรับกระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า มีค่า CVSE เฉลี่ยมากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าคือ 50:50 ขณะที่ค่า CVSE เฉลี่ยต่ำที่สุดแตกต่างกันตามเงื่อนไข ส่วนกระบวนการ SMOSE สำหรับโครงสร้างที่ซับซ้อน (CS) พบว่า ค่า CVSE เฉลี่ยมากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 70:30 (ยกเว้นสัดส่วนคะแนนข้อสอบร่วม 10% ซึ่งพบว่า อัตราส่วนฯ 60:40 มีค่า CVSE มากที่สุด) และ ค่า CVSE เฉลี่ยต่ำที่สุด เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าคือ 60:40 (ยกเว้นสัดส่วนคะแนนข้อสอบร่วม 10% ซึ่งพบว่าอัตราส่วนฯ 40:60 มีค่า CVSE ต่ำที่สุด)

นอกจากนี้ยังพบว่า ค่าเฉลี่ย CVSE ของอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าส่วนใหญ่ใกล้เคียงกัน ไม่มีแนวโน้มชัดเจนว่าเพิ่มขึ้นหรือลดลง เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าเป็นอย่างไร เช่นเดียวกับการกระจายของค่า CVSE ที่พบว่า ค่า CVSE ของอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่ามีการกระจายในแต่ละเงื่อนไขใกล้เคียงกันมาก ดังภาพ 4.5-4.6 สอดคล้องกับผลการวิเคราะห์อิทธิพลหลักของตัวแปรอิสระ (main effect) ที่พบว่า เวกเตอร์ ค่าเฉลี่ยของ CVSE สำหรับอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแต่ละแบบไม่แตกต่างกัน ( $p = .150$ ) ดังตาราง 4.3

ผลการวิเคราะห์ข้างต้นสะท้อนให้เห็นว่า ความถูกต้องของอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแต่ละแบบไม่ได้แตกต่างกันอย่างมีนัยสำคัญ และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าไม่ได้มีแนวโน้มของความแม่นยำในการปรับเทียบคะแนนชัดเจนว่าเพิ่มขึ้นหรือลดลงอย่างไร อย่างไรก็ตามเมื่อพิจารณาในภาพรวม มีแนวโน้มว่าอัตราส่วนฯ 40 มีความถูกต้องมากที่สุด รองลงมา คือ อัตราส่วนฯ 40:60 โดยอัตราส่วนฯ ทั้งสองแบบไม่แตกต่างกันมาก

จากผลการศึกษาประสิทธิภาพในการเปรียบเทียบของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าแตกต่างกัน พบว่า ไม่สามารถระบุแนวโน้มของประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องในการเปรียบเทียบคะแนนว่ามีขนาดเพิ่มขึ้น เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนเพิ่มขึ้นหรือลดลงอย่างไร โดยอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่ามีแนวโน้มว่ามีผลต่อความแม่นยำในการเปรียบเทียบคะแนน (ค่า RB) มากกว่าประสิทธิภาพโดยรวม (ค่า RMSE) แต่อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าไม่มีผลต่อความเที่ยงในการเปรียบเทียบคะแนน (ค่า CVSE) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 อย่างไรก็ตามมีแนวโน้มว่าอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 40:60 มีประสิทธิภาพในการเปรียบเทียบคะแนนมากกว่าอัตราส่วนอื่น ๆ



ภาพ 4.17 แผนภูมิแท่งเปรียบเทียบค่า CVSE สำหรับอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า

## ตอนที่ 5 ประสิทธิภาพของกระบวนการ MOSE เมื่อโครงสร้างมิติต่างกัน

การนำเสนอผลการศึกษาประสิทธิภาพโดยรวมของกระบวนการ MOSE เมื่อโครงสร้างมิติต่างกัน คือ โครงสร้างที่ไม่ซับซ้อน (APSS) และโครงสร้างที่ซับซ้อน (CS) (วัตถุประสงค์ข้อ 4) แบ่งเป็น 3 ส่วนหลัก คือ (1) ผลด้านประสิทธิภาพโดยรวม (RMSE) (2) ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB) และ (3) ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE) รายละเอียดดังนี้

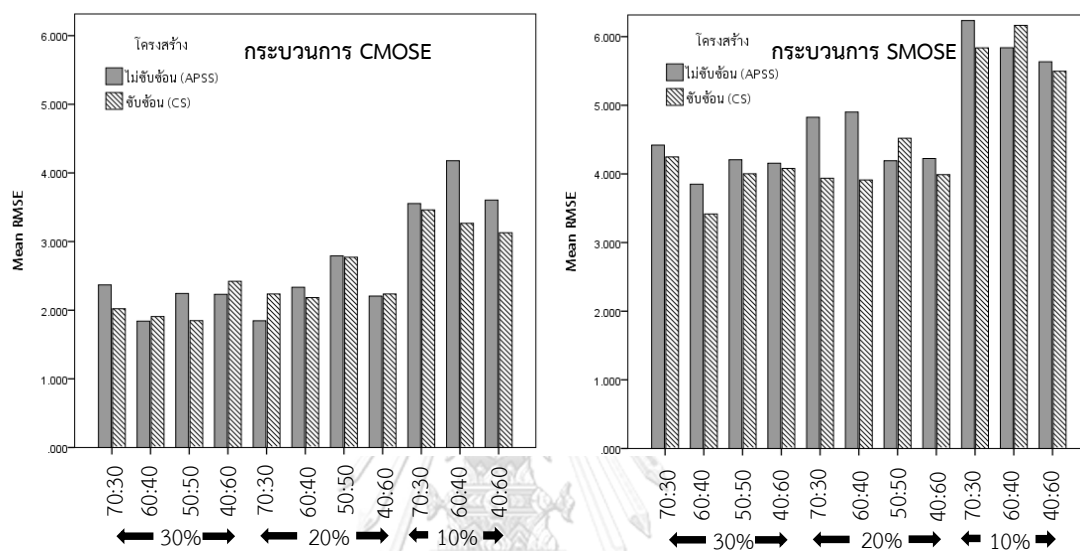
### 5.1 ผลด้านประสิทธิภาพโดยรวม (RMSE)

เมื่อพิจารณาประสิทธิภาพโดยรวมสำหรับกระบวนการ CMOSE เมื่อสัดส่วนคะแนนข้อสอบรวม 10% (ภาพ 4.18) พบว่า โครงสร้าง CS ส่วนใหญ่มีค่า RMSE เฉลี่ยต่ำกว่าโครงสร้าง APSS โดยโครงสร้าง CS มีค่า RMSE เฉลี่ยต่ำกว่าทุกเงื่อนไข ขณะที่สัดส่วนคะแนนข้อสอบรวม คือ 30% พบว่า โครงสร้าง CS มีค่า RMSE เฉลี่ยต่ำกว่า เมื่ออัตราส่วนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 และ 50:50 และเมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% พบว่า แบบสอบที่มีโครงสร้าง CS มีค่า RMSE เฉลี่ยต่ำกว่า เมื่ออัตราส่วนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 และ 50:50

เมื่อพิจารณาประสิทธิภาพโดยรวมสำหรับกระบวนการ SMOSE พบว่า โครงสร้าง CS ส่วนใหญ่มีค่า RMSE เฉลี่ยต่ำกว่าโครงสร้าง APSS มีเพียง 2 เงื่อนไขที่โครงสร้าง APSS มีค่า RMSE เฉลี่ยต่ำกว่า คือ (1) สัดส่วนคะแนนข้อสอบรวม คือ 20% ที่มีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 50:50 และ (2) สัดส่วนคะแนนข้อสอบรวม คือ 10% ที่มีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40

ผลการวิเคราะห์สถิติเชิงบรรยายข้างต้น จะเห็นว่า โครงสร้างที่ซับซ้อน (CS) ส่วนใหญ่มีค่า RMSE ต่ำกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) ขณะที่ผลการกระจายของค่า RMSE พบว่า ค่า RMSE ของโครงสร้าง APSS ส่วนใหญ่มีการกระจายมากกว่าโครงสร้าง CS และเมื่อพิจารณาสัดส่วนคะแนนข้อสอบรวม คือ 10% พบว่า การกระจายของค่า RMSE ระหว่างโครงสร้างมิติต่างกันทั้ง 2 แบบแตกต่างกันมาก ขณะที่เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% และ 20% พบว่า การกระจายของค่า RMSE ระหว่างโครงสร้างมิติต่างกันทั้ง 2 แบบแตกต่างกันน้อยดังภาพ 4.2-4.3 สอดคล้องกับการวิเคราะห์อิทธิพลระหว่างตัวแปรอิสระ ที่พบว่า โครงสร้างที่ซับซ้อน (CS) มีค่า RMSE ต่ำกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลเพียงร้อยละ .7 ( $\eta^2 = .007$ ) จึงมีบางเงื่อนไขที่ให้ข้อค้นพบแตกต่างออกไป

ผลการวิเคราะห์ข้างต้นสะท้อนให้เห็นว่า **ประสิทธิภาพโดยรวมระหว่างโครงสร้างมิต** **ความสามารถทั้ง 2 แบบอาจไม่ได้แตกต่างกันในเชิงปฏิบัติ** เนื่องจากประสิทธิภาพโดยรวมระหว่าง **โครงสร้างมิตความสามารถทั้ง 2 แบบส่วนใหญ่ใกล้เคียงกัน** แต่เมื่อพิจารณาในภาพรวม **โครงสร้างที่** **ซับซ้อน (CS) ส่วนใหญ่มีประสิทธิภาพโดยรวมมากกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) แต่มีบางกรณีที่ได้** **ข้อค้นพบแตกต่างออกไป**



ภาพ 4.18 กราฟแท่งเปรียบเทียบค่า RMSE สำหรับโครงสร้างมิตความสามารถ

## 5.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB)

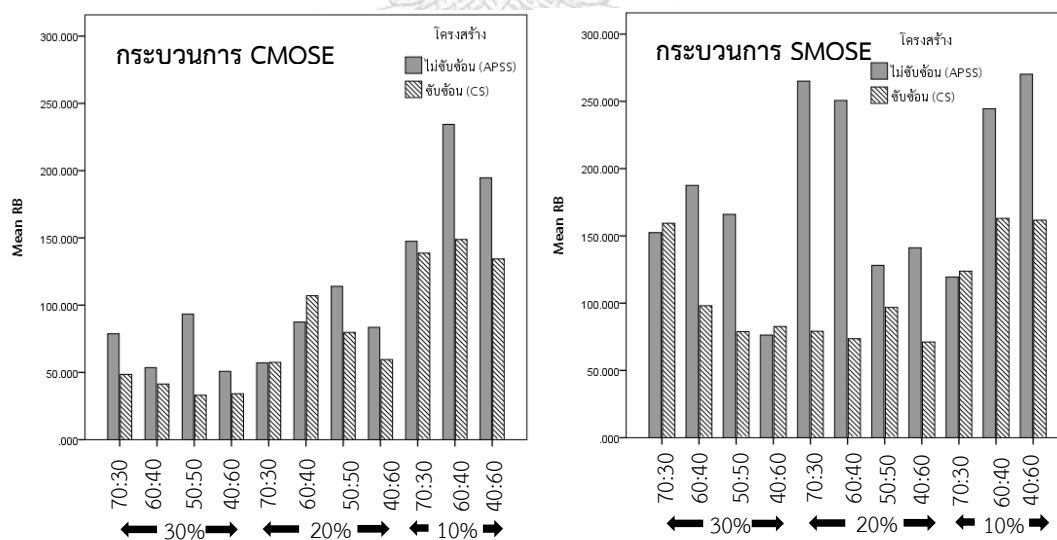
เมื่อพิจารณาความแม่นยำในการปรับเทียบคะแนนสำหรับกระบวนการ CMOSE เมื่อสัดส่วน **คะแนนข้อสอบรวม คือ 30% และ 10%** (ภาพ 4.19) พบว่า โครงสร้าง CS มีค่า RB เฉลี่ยต่ำกว่า โครงสร้าง APSS ทุกเงื่อนไข ขณะที่สัดส่วนคะแนนข้อสอบรวม คือ 20% พบว่า โครงสร้าง CS มีค่า RB เฉลี่ยต่ำกว่าโครงสร้าง APSS เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 50:50 และ 40:60 สอดคล้องกับผลการกระจายของค่า RB เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% และ 10% ที่ พบว่า ค่า RB ของโครงสร้าง CS มีการกระจายน้อยกว่าโครงสร้าง APSS ขณะที่เมื่อสัดส่วนคะแนน ข้อสอบรวม 20% พบว่า ค่า RB ของโครงสร้าง APSS มีการกระจายน้อยกว่าโครงสร้าง CS ดังภาพ 4.5-4.6

เมื่อพิจารณาประสิทธิภาพโดยรวมสำหรับกระบวนการ SMOSE เมื่อสัดส่วนคะแนนข้อสอบ **รวม คือ 20%** พบว่า โครงสร้าง CS มีค่า RB ต่ำกว่าโครงสร้าง APSS ทุกเงื่อนไข ขณะที่สัดส่วนคะแนน ข้อสอบรวม คือ 30% พบว่า โครงสร้าง CS มีค่า RB เฉลี่ยต่ำกว่าโครงสร้าง APSS เมื่ออัตราส่วน **คะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 และ 50:50** และเมื่อสัดส่วนคะแนนข้อสอบ **รวม คือ 10%** พบว่า โครงสร้าง CS มีค่า RB เฉลี่ยต่ำกว่าโครงสร้าง APSS (ยกเว้นอัตราส่วนคะแนน

ข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 70:30) สอดคล้องกับผลการกระจายของค่า RB ที่พบว่า ค่า RB ของโครงสร้าง CS มีการกระจายน้อยกว่าโครงสร้าง APSS (ยกเว้นเมื่อสัดส่วนคะแนนข้อสอบร่วม 30% และ 10% ที่มีอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า 70:30)

ผลการวิเคราะห์สถิติเชิงบรรยายข้างต้น จะเห็นว่า **โครงสร้าง CS ส่วนใหญ่มีค่า RB เฉลี่ยต่ำกว่า** และมีการกระจายน้อยกว่าโครงสร้าง APSS แต่มีบางกรณีที่ได้ข้อค้นพบแตกต่างออกไป สอดคล้องกับสอดคล้องกับผลการวิเคราะห์ความแปรปรวนพหุนาม (MANOVA) ที่พบว่า มีปฏิสัมพันธ์อิทธิพลระหว่างรูปแบบของกระบวนการ MOSE กับตัวแปรอิสระอื่น ๆ ต่อค่า RB อย่างมีนัยสำคัญที่ระดับ .05 ด้วยขนาดอิทธิพลร้อยละ 2.2 ดังตาราง 4.3 โดยโครงสร้างมิตีความสามารถมีขนาดอิทธิพลต่อค่า RB ร้อยละ .8 จึงต้องเปรียบเทียบค่า RB ระหว่างโครงสร้างที่ไม่ซับซ้อน (APSS) โครงสร้างที่ซับซ้อน (CS) ของกระบวนการ MOSE ทั้ง 2 แบบร่วมกับตัวแปรอื่น ๆ

ผลการวิเคราะห์ข้างต้นสะท้อนให้เห็นว่า **โครงสร้างที่ซับซ้อน (CS) ส่วนใหญ่มีความแม่นยำในการปรับเทียบคะแนนมากกว่าโครงสร้างที่ไม่ซับซ้อน (APSS)** แต่มีบางกรณีที่พบว่า โครงสร้าง APSS มีค่า RB เฉลี่ยต่ำกว่าโครงสร้าง CS แต่ความแตกต่างระหว่างค่า RB เฉลี่ยของโครงสร้างมิตีความสามารถทั้งสองแบบไม่มากนัก นอกจากนี้ยังพบว่า ความแม่นยำในการปรับเทียบคะแนนระหว่างโครงสร้างมิตีความสามารถมีความแตกต่างกันมากกว่าประสิทธิภาพโดยรวม (โดยเฉพาะกระบวนการ SMOSE)



ภาพ 4.19 แผนภูมิแท่งเปรียบเทียบค่า RB สำหรับโครงสร้างมิตีความสามารถ

### 5.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE)

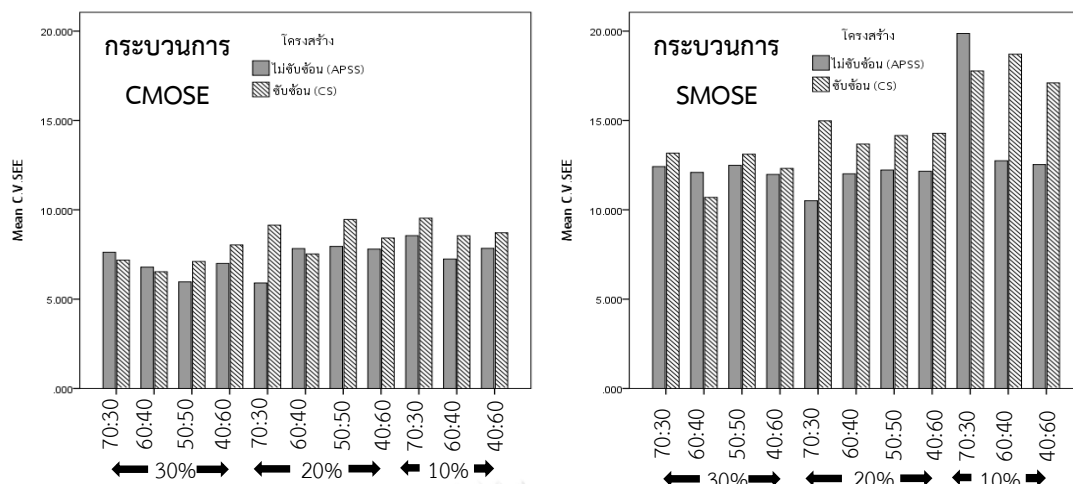
เมื่อพิจารณาความแม่นยำในการปรับเทียบคะแนนสำหรับกระบวนการ CMOSE เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10% (ภาพ 4.20) พบว่า โครงสร้าง APSS มีค่า CVSE เฉลี่ยต่ำกว่าโครงสร้าง CS ทุกเงื่อนไข ขณะที่สัดส่วนคะแนนข้อสอบรวม คือ 30% พบว่า โครงสร้าง APSS มีค่า CVSE เฉลี่ยต่ำกว่าโครงสร้าง CS เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 50:50 และ 40:60 และเมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% พบว่า โครงสร้าง APSS มีค่า CVSE เฉลี่ยต่ำกว่าโครงสร้าง CS (ยกเว้นอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 60:40)

เมื่อพิจารณาประสิทธิภาพโดยรวมสำหรับกระบวนการ SMOSE เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% พบว่า โครงสร้าง APSS มีค่า CVSE เฉลี่ยต่ำกว่าโครงสร้าง CS ทุกเงื่อนไข ขณะที่สัดส่วนคะแนนข้อสอบรวม คือ 30% พบว่า โครงสร้าง APSS มีค่า CVSE เฉลี่ยต่ำกว่าโครงสร้าง CS (ยกเว้นอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 60:40) และเมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10% พบว่า โครงสร้าง APSS มีค่า CVSE เฉลี่ยต่ำกว่าโครงสร้าง CS (ยกเว้นอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30)

ผลการวิเคราะห์สถิติเชิงบรรยายข้างต้น จะเห็นว่า **โครงสร้าง APSS ส่วนใหญ่มีค่า CVSE เฉลี่ยต่ำกว่าและมีการกระจายน้อยกว่าโครงสร้าง CS** (แตกต่างจากแนวโน้มของค่า RMSE และ RB) อย่างไรก็ตามค่า CVSE เฉลี่ยระหว่างโครงสร้าง CS และ APSS ใกล้เคียงกัน เช่นเดียวกับผลการกระจายของค่า CVSE ที่พบว่า ค่า CVSE ระหว่างโครงสร้างมิติตามความสามารถทั้ง 2 แบบใกล้เคียงกัน ดังภาพ 4.8-4.9 สอดคล้องกับผลการวิเคราะห์อิทธิพลหลักของตัวแปรอิสระ (main effect) พบว่า **เวกเตอร์ค่าเฉลี่ยของความถูกต้องระหว่างรูปแบบของโครงสร้างมิติตามความสามารถแตกต่างกันกลุ่มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยมีขนาดอิทธิพลเพียงร้อยละ .30 ซึ่งไม่มีผลในเชิงปฏิบัติ**

ผลการวิเคราะห์ข้างต้นสะท้อนให้เห็นว่า **ความถูกต้องระหว่างโครงสร้างมิติตามความสามารถทั้ง 2 แบบอาจไม่ได้แตกต่างกันในเชิงปฏิบัติ** เนื่องจากประสิทธิภาพโดยรวมระหว่างโครงสร้างมิติตามความสามารถทั้ง 2 แบบส่วนใหญ่ใกล้เคียงกัน แต่เมื่อพิจารณาในภาพรวม **โครงสร้างที่ไม่ซับซ้อน (APSS) ส่วนใหญ่มีความถูกต้องในการปรับเทียบคะแนนมากกว่าโครงสร้างที่ซับซ้อน (CS)**

จากผลการศึกษาประสิทธิภาพในการปรับเทียบของกระบวนการ MOSE เมื่อโครงสร้างมิติตามความสามารถแตกต่างกัน มีแนวโน้มว่าโครงสร้างที่ซับซ้อน (CS) มีความแม่นยำและประสิทธิภาพโดยรวมมากกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) โดยโครงสร้างมิติตามความสามารถมีผลต่อแม่นยำในการปรับเทียบคะแนน (ค่า RB) มากที่สุด เมื่อเทียบกับความถูกต้องและประสิทธิภาพโดยรวมในการปรับเทียบคะแนน (ค่า CVSE และ RMSE)



ภาพ 4.20 แผนภูมิแท่งเปรียบเทียบค่า CVSE สำหรับโครงสร้างมิตีความสามารถ

## ตอนที่ 6 สรุปประสิทธิภาพของการปรับเทียบคะแนนแยกตามวัตถุประสงค์

การนำเสนอประสิทธิภาพของการปรับเทียบคะแนนในส่วนนี้ เป็นการพิจารณาประสิทธิภาพของการปรับเทียบคะแนนร่วมกันทั้งด้านประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องแยกตามวัตถุประสงค์ทั้ง 4 ข้อ ดังนี้ (สรุปผลการวิจัยแยกตามวัตถุประสงค์ และตัวแปรตามจะอธิบายไว้ในบทที่ 5)

1. กระบวนการ CMOSE มีประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องมากกว่ากระบวนการ SMOSE
2. มีแนวโน้มว่าสัดส่วนคะแนนข้อสอบรวม 30% และ 20% มีประสิทธิภาพในการปรับเทียบคะแนน และควรหลีกเลี่ยงการใช้สัดส่วนคะแนนข้อสอบรวม 10%
3. ไม่สามารถระบุแนวโน้มของประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องในการปรับเทียบคะแนนว่ามีขนาดเพิ่มขึ้น เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนเพิ่มขึ้นหรือลดลงอย่างไร เมื่อพิจารณาในภาพรวม มีแนวโน้มว่าอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 40:60 มีประสิทธิภาพในการปรับเทียบคะแนนมากกว่าอัตราส่วนอื่น ๆ
4. มีแนวโน้มว่าโครงสร้างที่ซับซ้อน (CS) มีความแม่นยำและประสิทธิภาพโดยรวมมากกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) แต่ความเที่ยงของการปรับเทียบคะแนนระหว่างโครงมิตีความสามารถทั้ง 2 แบบไม่แตกต่างกันอย่างมีนัยสำคัญ

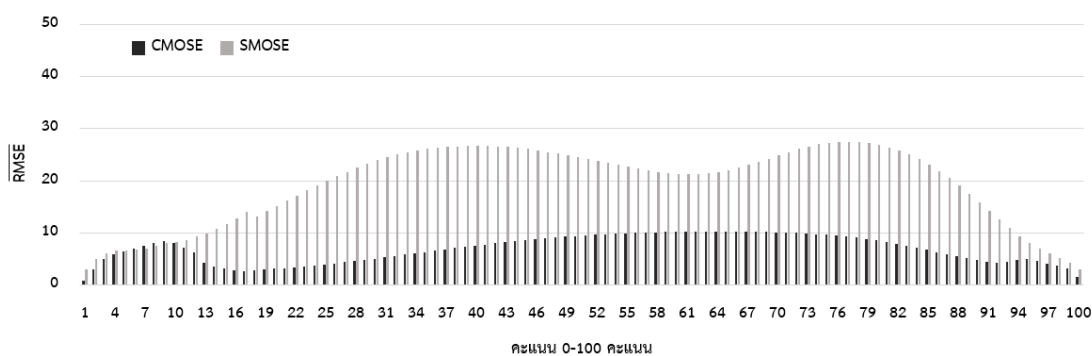
## ตอนที่ 7 ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ

ในส่วนนี้เป็นการวิเคราะห์เพิ่มเติมจากวัตถุประสงค์หลักของการวิจัยทั้ง 4 ข้อ โดยวิเคราะห์ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ (mean square error of estimated standard error) เพื่ออธิบายประสิทธิภาพในการประมาณค่า SE ของคะแนนปรับเทียบแต่ละระดับคะแนน ภายใต้เงื่อนไขการจำลองที่แตกต่างกันทั้ง 22 สถานการณ์ ประสิทธิภาพดังกล่าววัดจากค่าเฉลี่ยรากที่สองของความคลาดเคลื่อนกำลังสอง ( $\overline{RMSE}$ ) โดยที่  $\overline{RMSE} = \frac{\sum_{m=1}^{22} RMSE_{km}}{m}$  เมื่อ  $m$  คือ เงื่อนไขการจำลองข้อมูลที่มีค่าเท่ากับ 1, 2, ..., 22 และส่วนเบี่ยงเบนมาตรฐานของรากที่สองของความคลาดเคลื่อนกำลังสอง ( $SD_{RMSE}$ ) โดยที่  $SD_{RMSE} = \sqrt{\frac{\sum_{m=1}^{22} (RMSE_{km} - \overline{RMSE})^2}{k-1}}$  และเพื่อให้การแปลผลมีความหมายจึงนำเสนอค่าสัมประสิทธิ์การแปรผันของ RMSE ( $CV_{RMSE}$ ) ประกอบการแปลผล รายละเอียดเป็นดังนี้

เมื่อพิจารณาค่าเฉลี่ยของประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบแต่ละคะแนนในภาพรวมของ 22 เงื่อนไข ( $\overline{RMSE}$ ) จากภาพ 4.21 พบว่า ค่า  $\overline{RMSE}$  สำหรับกระบวนการ CMOSE ต่ำกว่ากระบวนการ SMOSE กล่าวคือ ความเบี่ยงเบนในการประมาณค่า SE สำหรับกระบวนการ CMOSE ต่ำกว่ากระบวนการ SMOSE

เมื่อพิจารณากระบวนการ CMOSE พบว่า ค่า  $\overline{RMSE}$  ต่ำที่สุดและมากที่สุด คือ .908 และ 10.125 ตามลำดับ โดยช่วงคะแนน 61-70 มีค่า  $\overline{RMSE}$  ต่ำที่สุด คือ 10.125-10.279 และช่วงคะแนน 71-80 มีค่า  $\overline{RMSE}$  มากที่สุด คือ 15.359-16.996 ดังตาราง 4.9 (พิจารณาประสิทธิภาพของการประมาณค่า SE แต่ละคะแนนได้จากตาราง 33 ในภาคผนวก จ)

เมื่อพิจารณากระบวนการ SMOSE พบว่า ค่า  $\overline{RMSE}$  ต่ำที่สุดและมากที่สุด คือ 3.045 และ 27.481 ตามลำดับ โดยช่วงคะแนน 71-80 มีค่า  $\overline{RMSE}$  ต่ำที่สุด คือ 25.523-27.481 และช่วงคะแนน 51-60 มีค่า  $\overline{RMSE}$  มากที่สุด คือ 151.951-191.321 นอกจากนี้มีข้อสังเกตว่า ช่วงกลางคะแนนของกระบวนการ SMOSE มีความเบี่ยงเบนในการประมาณค่า SE มากกว่าคะแนนช่วงต้นและช่วงปลาย



ภาพ 4.21 ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ



ตาราง 4.9 ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบในภาพรวม

ช่วง คะแนน	CMOSE			SMOSE		
	$\overline{RMSE}$	$SD_{RMSE}$	$CV_{RMSE}$	$\overline{RMSE}$	$SD_{RMSE}$	$CV_{RMSE}$
1-10	.908-8.171	0.578-3.686	.334-13.583	3.045-8.344	4.062-8.344	14.607-34.615
11-20	2.743-7.211	.712-4.838	.421-23.407	8.684-14.184	5.346-10.415	28.775-108.466
21-30	3.114-5.084	.747-1.584	.559-2.508	16.199-23.948	6.301-7.538	39.702-56.817
31-40	5.333-7.575	1.688-2.417	2.849-5.844	24.616-26.703	7.727-8.434	59.713-71.140
41-50	7.800-9.399	2.475-2.978	6.128-8.870	24.495-12.133	8.644-12.133	74.726-147.207
51-60	9.929-10.211	3.038-3.582	9.229-12.830	21.293-24.138	12.568-13.699	151.951-191.321
61-70	10.125-10.279	3.638-4.043	13.236-16.349	21.024-24.908	9.851-13.243	97.037-175.372
71-80	8.551-10.060	3.919-4.125	15.359-16.996	25.523-27.481	9.934-10.751	98.688-115.574
81-90	4.823-7.575	1.688-2.417	2.849-5.844	15.791-26.432	6.930-10.267	48.021-105.411
91-100	1.618-4.939	.653-2.290	.426-5.245	3.102-14.178	1.605-5.971	2.576-35.656

จากผลการวิเคราะห์ตามวัตถุประสงค์ ที่พบว่า (1) ไม่สามารถระบุแนวโน้มของประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องในการปรับเทียบคะแนนว่ามีขนาดเพิ่มขึ้น เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนเพิ่มขึ้นหรือลดลงอย่างไร และ (2) ความเที่ยงของการปรับเทียบคะแนนระหว่างเครื่องมือความสามารถทั้ง 2 แบบไม่แตกต่างกันอย่างมีนัยสำคัญ ผู้วิจัยจึงเปรียบเทียบประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบในแต่ละเงื่อนไข ภายใต้ 2 ตัวแปร ได้แก่ กระบวนการ MOSE และสัดส่วนคะแนนข้อสอบร่วม เพิ่มเติมไว้ในภาคผนวก ฉ

## บทที่ 5

### สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีจุดประสงค์ เพื่อเปรียบเทียบประสิทธิภาพระหว่างกระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์แบบพร้อมกัน (CMOSE) และกระบวนการ MOSE ประมาณค่าพารามิเตอร์แบบแยกกัน และใช้การเชื่อมโยงสเกลด้วยวิธี TCF (SMOSE) และเพื่อศึกษาประสิทธิภาพของกระบวนการ MOSE ทั้ง 2 แบบ เมื่อสัดส่วนคะแนนข้อสอบรวม อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า และโครงสร้างมิติความสามารถแตกต่างกัน โดยสัดส่วนคะแนนข้อสอบรวม มี 3 แบบ ได้แก่ 10%, 20% และ 30% อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า มี 3 แบบ ได้แก่ 40:60, 50:50, 60:40 และ 70:30 และโครงสร้างมิติความสามารถมี 2 แบบ ได้แก่ โครงสร้างที่ไม่ซับซ้อน และโครงสร้างที่ซับซ้อน ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบรวม

ขั้นตอนการจำลองข้อมูลแบบมอนติคาร์โล ประกอบด้วย 4 ขั้นตอนหลัก ได้แก่ (1) จำลองค่าพารามิเตอร์ของข้อสอบและแบบแผนการตอบข้อสอบ (2) การประมาณค่าพารามิเตอร์ตามโมเดลทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (3) การเชื่อมโยงสเกล (สำหรับกระบวนการ MOSE ที่ใช้การเชื่อมโยงสเกล) และ (4) การปรับเทียบคะแนน ภายใต้กลุ่มตัวอย่าง 2 กลุ่ม กลุ่มละ 3,000 คน ที่มีการกระทำซ้ำ 40 รอบ โดยดำเนินการผ่านโปรแกรม R

ประสิทธิภาพในการปรับเทียบคะแนนสามารถพิจารณาได้จาก (1) ค่าความลำเอียงสัมพัทธ์ (RB) (2) สัมประสิทธิ์การแปรผันของค่าความคลาดเคลื่อนมาตรฐาน (CVSE) และ (3) ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) สะท้อนถึงความแม่นยำ ความถูกต้อง และประสิทธิภาพโดยรวมของการปรับเทียบคะแนน ตามลำดับ โดยการวิเคราะห์ข้อมูลประกอบด้วย 3 ส่วน คือ (1) สถิติบรรยาย เพื่ออธิบายลักษณะของค่า RMSE ค่า RB และค่า CVSE (2) สถิติวิเคราะห์ความแปรปรวน (ANOVA) เพื่อตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยของประสิทธิภาพโดยรวมของการปรับเทียบคะแนน (ค่า RMSE) และ (3) สถิติวิเคราะห์ตัวแปรพหุนาม (MANOVA) เพื่อตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยของความแม่นยำและความถูกต้องในการปรับเทียบคะแนน (ค่า RB และค่า CVSE)

การนำเสนอสรุปผลการวิจัย และอภิปราย แบ่งเป็น 6 ส่วน ประกอบด้วย ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพของการปรับเทียบคะแนน ผลการวิจัยแบ่งตามวัตถุประสงค์ (4 ข้อ) นอกจากนี้ยังอภิปรายเพิ่มเติมในประเด็นที่น่าสนใจ ได้แก่ วิธีอีควิเปอร์เซ็นไทล์ที่มีการปรับโค้งให้ราบเรียบ (pre-smoothed equipercentile equating method) และการกำหนดวิธีปรับเทียบคะแนนเกณฑ์ ส่วนข้อเสนอแนะ ประกอบด้วย ข้อเสนอแนะในการนำไปปฏิบัติ ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป และข้อเสนอแนะเชิงทฤษฎี มีรายละเอียดดังนี้

## สรุปผลการวิจัย

### 1. ปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อประสิทธิภาพของการปรับเทียบคะแนน

#### 1.1 ผลด้านประสิทธิภาพโดยรวม (RMSE)

มีปฏิสัมพันธ์อิทธิพลของสัดส่วนคะแนนข้อสอบร่วมกับอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าต่อประสิทธิภาพโดยรวมในการปรับเทียบคะแนน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ 1.8

#### 1.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB)

มีปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระทั้ง 4 ตัวแปร ได้แก่ รูปแบบของกระบวนการ MOSE โครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบร่วม และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าต่อความแม่นยำอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ 2.2

#### 1.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE)

ไม่มีปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระต่อความถูกต้องในการปรับเทียบคะแนนในเชิงปฏิบัติ

### 2. ประสิทธิภาพระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE

#### 2.1 ผลด้านประสิทธิภาพโดยรวมของการปรับเทียบคะแนน (RMSE)

การปรับเทียบคะแนนด้วยกระบวนการ CMOSE มีประสิทธิภาพโดยรวมมากกว่ากระบวนการ SMOSE อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ 43.1 โดยประสิทธิภาพโดยรวมของกระบวนการ CMOSE มีการกระจายน้อยกว่ากระบวนการ SMOSE ทุกเงื่อนไข

#### 2.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB)

จากปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระทั้ง 4 ตัวแปรต่อความแม่นยำในการปรับเทียบคะแนน เมื่อพิจารณาในภาพรวม พบว่า กระบวนการ CMOSE ส่วนใหญ่มีความแม่นยำในการปรับเทียบคะแนนมากกว่ากระบวนการ SMOSE ซึ่งบางกรณีมีผลการวิจัยแตกต่างออกไป อย่างไรก็ตามความแม่นยำในการปรับเทียบคะแนนของกระบวนการ CMOSE มีการกระจายน้อยกว่ากระบวนการ SMOSE ทุกเงื่อนไข

#### 2.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE)

กระบวนการ CMOSE มีความถูกต้องในการปรับเทียบคะแนนมากกว่ากระบวนการ SMOSE อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ 6.2 โดยความถูกต้องในการปรับเทียบคะแนนของกระบวนการ CMOSE มีการกระจายน้อยกว่ากระบวนการ SMOSE ทุกเงื่อนไข

จากผลการวิจัยข้างต้นสะท้อนให้เห็นว่า กระบวนการ CMOSE มีแนวโน้มว่าประสิทธิภาพมากกว่ากระบวนการ SMOSE ทั้งในด้านความแม่นยำ ความถูกต้อง และประสิทธิภาพโดยรวม

### 3. ประสิทธิภาพของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบรวมแตกต่างกัน

#### 3.1 ผลด้านประสิทธิภาพโดยรวมของการปรับเทียบคะแนน (RMSE)

จากปฏิสัมพันธ์อิทธิพลของสัดส่วนคะแนนข้อสอบรวมกับอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าต่อประสิทธิภาพโดยรวมในการปรับเทียบคะแนน เมื่อวิเคราะห์อิทธิพลอย่างง่าย พบว่า (1) กรณีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 และ 50:50 กระบวนการ MOSE มีประสิทธิภาพโดยรวมมากขึ้น เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้น อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) และ (2) กรณีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 และ 40:60 พบว่า ประสิทธิภาพโดยรวมของการปรับเทียบคะแนนมากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20%, 30% และ 10% ตามลำดับ แต่ประสิทธิภาพโดยรวมของสัดส่วนคะแนนข้อสอบรวม 20% และ 30% แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ โดยสัดส่วนคะแนนข้อสอบรวม 10% มีประสิทธิภาพโดยรวมต่ำที่สุดทุกเงื่อนไข

#### 3.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน (RB)

จากปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระทั้ง 4 ตัวแปรต่อความแม่นยำในการปรับเทียบคะแนน เมื่อพิจารณาในภาพรวม มีข้อค้นพบสำคัญ 2 แบบได้แก่ (1) ค่า RB เฉลี่ยลดลง เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้นทุกเงื่อนไข และ (2) ค่า RB มีค่ามากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10%, 30% และ 20% อย่างไรก็ตาม ค่า RB เฉลี่ยของสัดส่วนคะแนนข้อสอบรวม 10% ส่วนใหญ่มีการกระจายมากที่สุด ดังนั้นมีแนวโน้มว่า สัดส่วนคะแนนข้อสอบรวม 30% และ 20% มีความแม่นยำในการปรับเทียบคะแนน ขณะที่ควรหลีกเลี่ยงสัดส่วนคะแนนข้อสอบรวม 10%

#### 3.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน (CVSE)

ความถูกต้องในการปรับเทียบคะแนนระหว่างรูปแบบของสัดส่วนคะแนนข้อสอบรวมแตกต่างกันอย่างน้อยหนึ่งกลุ่มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ .90 เมื่อพิจารณาในภาพรวม พบว่า ค่า CVSE เฉลี่ยส่วนใหญ่มากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10% อย่างไรก็ตาม ค่า CVSE เฉลี่ยในแต่ละเงื่อนไขค่อนข้างใกล้เคียงกัน โดยสัดส่วนคะแนนข้อสอบรวมรูปแบบต่าง ๆ มีการกระจายใกล้เคียงกัน มีเพียงกระบวนการ SMOSE สำหรับโครงสร้าง CS ที่พบว่า ค่า CVSE ของสัดส่วนคะแนนข้อสอบรวม 10% มีการกระจายมากที่สุดอย่างชัดเจน ดังนั้นจึงมีแนวโน้มว่า สัดส่วนคะแนนข้อสอบรวม 30% และ 20% มีความแม่นยำในการปรับเทียบคะแนน

จากผลการวิจัยข้างต้นสะท้อนให้เห็นว่า มีแนวโน้มสัดส่วนคะแนนข้อสอบรวม 30% และ 20% มีประสิทธิภาพในการปรับเทียบคะแนน และควรหลีกเลี่ยงการใช้สัดส่วนคะแนนข้อสอบรวม 10% โดยสัดส่วนคะแนนข้อสอบรวมมีผลต่อประสิทธิภาพโดยรวมมากที่สุด (ค่า RMSE) รองลงมา คือ ความถูกต้องและความแม่นยำในการปรับเทียบคะแนน ตามลำดับ (ค่า RB และค่า CVSE)

#### 4. ประสิทธิภาพของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าแตกต่างกัน

##### 4.1 ผลด้านประสิทธิภาพโดยรวมของการปรับเทียบคะแนน

จากปฏิสัมพันธ์อิทธิพลของสัดส่วนคะแนนข้อสอบรวมกับอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า ที่ส่งผลต่อประสิทธิภาพโดยรวมในการปรับเทียบคะแนน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อวิเคราะห์อิทธิพลอย่างง่าย พบว่า (1) เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% ประสิทธิภาพโดยรวมมากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40, 50:50, 40:60 และ 70:30 ตามลำดับ ซึ่งประสิทธิภาพโดยรวมระหว่างสัดส่วนฯ 70:30 และ 40:60 แตกต่างกันอย่างไม่มีนัยสำคัญ (2) เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% พบว่า ประสิทธิภาพโดยรวมมากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 40:60, 70:30, 50:50 และ 60:40 ตามลำดับ ซึ่งประสิทธิภาพโดยรวมระหว่างอัตราส่วนฯ 70:30, 50:50 และ 40:60 แตกต่างกันอย่างไม่มีนัยสำคัญ และสัดส่วนคะแนนข้อสอบรวม คือ 10% พบว่า ประสิทธิภาพโดยรวมมากที่สุด เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 40:60, 70:30 และ 60:40 ตามลำดับ ซึ่งประสิทธิภาพโดยรวมระหว่างอัตราส่วนฯ 60:40 และ 70:30 % มีความแตกต่างกันอย่างไม่มีนัยสำคัญ

อย่างไรก็ตามเมื่อพิจารณาในภาพรวม พบว่า ไม่มีแนวโน้มชัดเจนว่า อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่ามีประสิทธิภาพโดยรวมเพิ่มขึ้นหรือลดลงอย่างไร โดยอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60 มีแนวโน้มว่ามีประสิทธิภาพโดยรวมมากที่สุดเมื่อเทียบกับสัดส่วนอื่น

##### 4.2 ผลด้านความแม่นยำในการปรับเทียบคะแนน

จากปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระทั้ง 4 ตัวแปรต่อความแม่นยำในการปรับเทียบคะแนน เมื่อพิจารณาในภาพรวม พบว่า ไม่มีแนวโน้มชัดเจนว่า อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่ามีประสิทธิภาพโดยรวมเพิ่มขึ้นหรือลดลงอย่างไร โดยอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60 มีแนวโน้มว่ามีความแม่นยำมากที่สุดเมื่อเทียบกับสัดส่วนอื่น ๆ

##### 4.3 ผลด้านความถูกต้องในการปรับเทียบคะแนน

ความถูกต้องสำหรับอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าแต่ละแบบไม่แตกต่างกันอย่างมีนัยสำคัญ ( $p = .150$ ) โดยความถูกต้องสำหรับในแต่ละอัตราส่วนค่อนข้างใกล้เคียงกันและมีการกระจายไม่แตกต่างกันมากนัก อย่างไรก็ตาม เมื่อพิจารณาในภาพรวม พบว่า อัตราส่วนฯ 60:40 มีแนวโน้มของความถูกต้องมากที่สุด รองลงมา คือ อัตราส่วนฯ 40:60 ซึ่งความถูกต้องระหว่างอัตราส่วนฯ ทั้ง 2 รูปแบบไม่แตกต่างกันมากนัก

จากผลการวิจัยข้างต้นสะท้อนให้เห็นว่า อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า ไม่ได้มีผลต่อประสิทธิภาพของการเปรียบเทียบคะแนนในเชิงปฏิบัติมากนัก และไม่สามารถระบุแนวโน้มของประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องว่า มีขนาดเพิ่มขึ้น เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าเพิ่มขึ้นหรือลดลง อย่างไรก็ตาม หากต้องระบุอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าที่มีประสิทธิภาพในการเปรียบเทียบคะแนน ควรเลือกใช้อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 40:60

## 5. ประสิทธิภาพของกระบวนการ MOSE เมื่อโครงสร้างมิติความสามารถแตกต่างกัน

### 5.1 ผลด้านประสิทธิภาพโดยรวมของการเปรียบเทียบคะแนน

จากผลการวิเคราะห์ห้อธิพหลหลัก พบว่า โครงสร้างที่ซับซ้อน (CS) มีประสิทธิภาพโดยรวมมากกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ .7 จึงมีบางเงื่อนไขที่ให้ออกค้นพบแตกต่างออกไป อย่างไรก็ตามเมื่อพิจารณาในภาพรวมพบว่า โครงสร้าง CS ส่วนใหญ่มีประสิทธิภาพโดยรวมมากกว่าโครงสร้าง APSS โดยประสิทธิภาพโดยรวมของโครงสร้าง CS ส่วนใหญ่มีการกระจายน้อยกว่าโครงสร้าง APSS ซึ่งกรณีโครงสร้าง APSS มีประสิทธิภาพโดยรวมมากกว่าโครงสร้าง CS พบว่า ความแตกต่างระหว่างโครงสร้างทั้ง 2 แบบไม่มากนัก

### 5.2 ผลด้านความแม่นยำในการเปรียบเทียบคะแนน

จากปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระทั้ง 4 ตัวแปรต่อความแม่นยำในการเปรียบเทียบคะแนน เมื่อพิจารณาในภาพรวม พบว่าโครงสร้าง CS ส่วนใหญ่มีความแม่นยำมากกว่าโครงสร้าง APSS และกรณีโครงสร้าง APSS มีความแม่นยำมากกว่าโครงสร้าง CS พบว่า ความแม่นยำของโครงสร้างมิติความสามารถทั้งสองแบบแตกต่างกันไม่มากนัก

### 5.3 ผลด้านความถูกต้องในการเปรียบเทียบคะแนน

จากผลการวิเคราะห์ห้อธิพหลหลัก พบว่า โครงสร้างที่ไม่ซับซ้อน (APSS) โครงสร้างที่ซับซ้อน (CS) มีค่า RMSE ต่ำกว่า อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ .3 ซึ่งไม่มีผลในเชิงปฏิบัติ โดยความถูกต้องระหว่างโครงสร้างมิติความสามารถทั้ง 2 แบบในแต่ละเงื่อนไขใกล้เคียงกัน อย่างไรก็ตามเมื่อพิจารณาในภาพรวม พบว่า โครงสร้าง APSS ส่วนใหญ่มีความถูกต้องมากกว่าโครงสร้าง CS (แตกต่างจากผลด้านความแม่นยำและประสิทธิภาพโดยรวม)

จากผลการวิจัยข้างต้นสะท้อนให้เห็นว่า โครงสร้างมิติความสามารถไม่ได้ส่งผลต่อประสิทธิภาพในการเปรียบเทียบคะแนนมากนัก อย่างไรก็ตามโครงสร้างที่ซับซ้อน (CS) มีแนวโน้มว่า ประสิทธิภาพโดยรวมและความแม่นยำมากกว่า

## อภิปรายผล

การวิจัยในครั้งนี้พบว่า มีปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระที่ส่งต่อประสิทธิภาพโดยรวม และความแม่นยำ (ค่า RMSE และค่า RB) ดังนั้นการอภิปรายผลจึงเริ่มจากการอภิปรายผลเกี่ยวกับปฏิสัมพันธ์ระหว่างตัวแปรอิสระ แล้วจึงอภิปรายของผลการวิจัยแยกตามวัตถุประสงค์ รวมถึงอภิปรายเพิ่มเติมเกี่ยวกับวิธีอีควิเปอร์เซนไทล์ที่มีการปรับโค้งให้ราบเรียบ (pre-smoothed equipercentile equating method) ต่อไป

### 1. ปฏิสัมพันธ์ระหว่างตัวแปรอิสระที่ส่งผลกระทบต่อประสิทธิภาพในการปรับเทียบคะแนน

ผลการวิจัย พบว่า (1) มีปฏิสัมพันธ์อิทธิพลของสัดส่วนคะแนนข้อสอบร่วมกับอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าต่อประสิทธิภาพโดยรวมในการปรับเทียบคะแนน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ 1.80 และ (2) มีปฏิสัมพันธ์อิทธิพลระหว่างรูปแบบของกระบวนการ MOSE โครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบร่วม และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าต่อความแม่นยำอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ( $p = .000$ ) ด้วยขนาดอิทธิพลร้อยละ 2.20 ซึ่งงานวิจัยที่ผ่านมา พบว่า มีปฏิสัมพันธ์อิทธิพลระหว่างตัวแปรอิสระที่มีต่อประสิทธิภาพในการปรับเทียบคะแนน เช่นกัน ไม่ว่าจะเป็น งานวิจัยของพัชร จันทรเพ็ง (2550) ที่พบว่า (1) วิธีการเชื่อมโยงสเกล โครงสร้างมิติความสามารถ และความสัมพันธ์ระหว่างมิติความสามารถร่วมกันส่งผลกระทบต่อค่าความลำเอียง (bias) (2) วิธีการเชื่อมโยงสเกล และโครงสร้างมิติความสามารถร่วมกันส่งผลกระทบต่อประสิทธิภาพโดยรวมในการปรับเทียบคะแนน (ค่า RMSE) และ (3) วิธีการเชื่อมโยงสเกล และความสัมพัทธ์ระหว่างมิติร่วมกันส่งผลกระทบต่อประสิทธิภาพโดยรวมในการปรับเทียบคะแนน เช่นเดียวกับ ศิวะทัศน์ สุขสุวรรณ (2555) ที่พบว่า วิธีการเชื่อมโยงสเกล โครงสร้างมิติความสามารถ สัดส่วนคะแนนข้อสอบร่วม และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า ร่วมกันส่งผลกระทบต่อค่าความลำเอียงและประสิทธิภาพโดยรวมในการปรับเทียบคะแนน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และงานวิจัยของ Zhang (2012) ที่พบว่า วิธีการเชื่อมโยงสเกล และความแตกต่างของความสามารถระหว่างกลุ่มผู้สอบร่วมกันส่งผลกระทบต่อค่าความลำเอียงและค่า RMSD และวิธีการปรับเทียบคะแนนและความแตกต่างของความสามารถระหว่างกลุ่มผู้สอบ ร่วมกันส่งผลกระทบต่อค่าความลำเอียงและค่า RMSD อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

## 2. ประสิทธิภาพระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE

จากผลการเปรียบเทียบภาพระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE พบว่า กระบวนการ CMOSE มีความถูกต้องและประสิทธิภาพโดยรวมมากกว่ากระบวนการ SMOSE ทุกเงื่อนไข และเมื่อพิจารณาถึงความแม่นยำพบว่า กระบวนการ CMOSE ส่วนใหญ่มีความแม่นยำมากกว่า กระบวนการ SMOSE โดยกระบวนการ SMOSE มีความแม่นยำมากกว่ามีเพียง 3 กรณี ได้แก่ (1) โครงสร้างที่ไม่ซับซ้อน (APSS) สัดส่วนคะแนนข้อสอบรวม 10% และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 (2) โครงสร้างที่ซับซ้อน (CS) สัดส่วนคะแนนข้อสอบรวม 20% และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 60:40 และ (3) โครงสร้างที่ซับซ้อน (CS) สัดส่วนคะแนนข้อสอบรวม 10% และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 โดยความแตกต่างด้านความแม่นยำระหว่าง CMOSE และกระบวนการ SMOSE (เทียบกับส่วนเบี่ยงเบนมาตรฐานของกระบวนการ CMOSE) ทั้ง 3 กรณี คือ .597, .604 และ .288 ตามลำดับ ซึ่งต่างกันน้อยกว่า 1SD (ตาราง 4.10) ขณะที่กรณีกระบวนการ CMOSE มีความแม่นยำมากกว่ากระบวนการ SMOSE พบว่า ความแตกต่างของความแม่นยำระหว่างกระบวนการทั้ง 2 แบบค่อนข้างมาก โดยกรณีโครงสร้างที่ไม่ซับซ้อนและซับซ้อน มีค่าเฉลี่ยของความแตกต่างดังกล่าว คือ 2.370 และ 1.289 ด้วยเหตุนี้จึงสามารถสรุปได้ว่ากระบวนการ CMOSE มีประสิทธิภาพในการปรับเทียบคะแนนมากกว่า กระบวนการ SMOSE ซึ่งสอดคล้องกับสมมติฐานที่ระบุว่ากระบวนการ CMOSE มีความลำเอียง (bias) และค่า RMSE ต่ำกว่ากระบวนการ SMOSE

กระบวนการ CMOSE มีประสิทธิภาพมากกว่ากระบวนการ SMOSE เนื่องจากการประมาณค่าพารามิเตอร์จากแบบสอบ 2 ฉบับในครั้งเดียว ทำให้ตัวอย่างในการประมาณค่าพารามิเตอร์มีขนาดใหญ่ ทำให้การคลาดเคลื่อนในการประมาณค่าต่ำกว่า อีกทั้งยังช่วยลดความคลาดเคลื่อนในการปรับเทียบคะแนนอันเนื่องจากการเชื่อมโยงสเกล (S.-H. Kim & Cohen, 1997; Seonghoon Kim & Kolen, 2006) ขณะที่การประมาณค่าพารามิเตอร์แบบแยกกันต้องใช้การเชื่อมโยงสเกล เพื่อนำค่าพารามิเตอร์ที่ประมาณได้ไว้บนสเกลของข้อสอบรวม ทำให้เกิดความคลาดเคลื่อนอย่างสุ่ม (random error) เนื่องจากการสุ่มตัวอย่างและการประมาณค่าสัมประสิทธิ์ของการเชื่อมโยงสเกล ส่งผลให้ค่าพารามิเตอร์หลังจากการเชื่อมโยงมีความแปรปรวน (Meng, 2007)

ผลการวิจัยดังกล่าวสอดคล้องกับงานวิจัยที่ผ่านมาหลายงาน ไม่ว่าจะเป็น งานวิจัยที่เปรียบเทียบประสิทธิภาพของการเชื่อมโยงสเกลด้วยโมเดล MIRT ระหว่างวิธีประมาณค่าพารามิเตอร์พร้อมกันและแยกกัน ซึ่งพบว่า การประมาณค่าพารามิเตอร์พร้อมกัน มีค่า RMSE และค่าความลำเอียงสำหรับค่าอำนาจจำแนก ( $\alpha_1$  และ  $\alpha_2$ ) ต่ำกว่าการประมาณค่าพารามิเตอร์แยกกันและใช้การเชื่อมโยงสเกล โดยสัดส่วนคะแนนข้อสอบรวม คือ 33.33% (Simon, 2008) สอดคล้องกับผลการวิจัยที่พบว่าการปรับเทียบคะแนนภายใต้รูปแบบ NEAT ด้วยวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกัน



มีประสิทธิภาพมากกว่าวิธีการประมาณค่าพารามิเตอร์แบบแยกกัน (ด้วยวิธีการเชื่อมโยงสเกลทั้ง 2 วิธี) เนื่องจากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันมีความลำเอียงในประมาณค่า  $a_1$ ,  $a_2$  และ  $d$  ต่ำกว่าและมีความเสถียรมากกว่าวิธีการประมาณค่าพารามิเตอร์แบบแยกกัน เช่นเดียวกับงานวิจัยที่เปรียบเทียบประสิทธิภาพของการเชื่อมโยงสเกลด้วยโมเดล IRT สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT ซึ่งพบว่า การประมาณค่าพารามิเตอร์พร้อมกันมีความคลาดเคลื่อนยกกำลังสองเฉลี่ย (mean square error: MSE) ต่ำกว่าการประมาณค่าพารามิเตอร์แยกกันและใช้การเชื่อมโยงสเกลโดยสัดส่วนคะแนนข้อสอบรวม คือ 33.33% (Seonghoon Kim & Kolen, 2006) และการประมาณค่าพารามิเตอร์แบบพร้อมกันมีค่า RMSD ต่ำกว่าการประมาณค่าพารามิเตอร์แบบแยกกันด้วยวิธีการเชื่อมโยงสเกลของ Stocking-Lord เนื่องจากค่าความสามารถของผู้สอบจากการประมาณค่าพารามิเตอร์แบบพร้อมกันมีความใกล้เคียงกับความสามารถที่แท้จริงของผู้สอบมากกว่าการประมาณค่าพารามิเตอร์แบบแยกกัน (Tian, 2011) นอกจากนี้ยังพบว่าการประมาณค่าพารามิเตอร์พร้อมกันมีความคลาดเคลื่อนต่ำกว่าวิธีโค้งคุณลักษณะ (อัญชลี ศรีภักชญา, 2552) และสอดคล้องกับผลการวิจัยสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT ที่พบว่า การประมาณค่าพารามิเตอร์พร้อมกันมีประสิทธิภาพมากกว่าการประมาณค่าพารามิเตอร์แยกกันและเชื่อมโยงสเกลด้วยวิธี Stocking-Lord เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 21.43% (Keller & Keller, 2015) และการประมาณค่าพารามิเตอร์พร้อมกันมีความคลาดเคลื่อนยกกำลังสองเฉลี่ย (MSE) ต่ำกว่าการประมาณค่าพารามิเตอร์แยกกันและใช้การเชื่อมโยงสเกลด้วยวิธีต่าง ๆ เมื่อสัดส่วนข้อสอบรวม คือ 20% (Bradley A. Hanson & Anton A. Béguin, 2002)

เมื่อพิจารณากรณีกระบวนการ SMOSE มีความแม่นยำมากกว่ากระบวนการ CMOSE ทั้ง 3 กรณี จะสังเกตว่า ส่วนใหญ่อยู่ในสัดส่วนคะแนนข้อสอบรวม 10% ซึ่งเมื่อพิจารณาความแตกต่างระหว่างกระบวนการ MOSE ทั้ง 2 แบบในสัดส่วนคะแนนข้อสอบรวม 10% พบว่า ความแม่นยำระหว่างกระบวนการ CMOSE และ SMOSE มีความแตกต่างกันน้อยกว่าสัดส่วนอื่น ๆ โดยกรณีสัดส่วนข้อสอบรวม 10% ที่มีอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30 มีจำนวนข้อสอบรวมเพียง 9 ข้อ ประกอบด้วย ข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่าจำนวน 8 ข้อ และ 1 ข้อ ตามลำดับ ซึ่งข้อสอบรวมที่ตรวจให้คะแนนหลายค่าที่มีข้อเดียว อาจทำให้ข้อสอบรวมดังกล่าวไม่เป็นตัวแทนเนื้อหาของแบบสอบ (Kolen & Brennan, 2004) อันจะส่งผลต่อความแม่นยำของข้อสอบรวม นำไปสู่การละเมิดข้อตกลงเบื้องต้นของการปรับเทียบคะแนน ซึ่งการประมาณค่าพารามิเตอร์แบบแยกกันสามารถช่วยลดผลกระทบของการกำหนดข้อมูลจำเพาะของโมเดลไม่ถูกต้อง (model misspecification) (Kim & Cohen, 1998 cited in Tian, 2011) โดยงานวิจัยเกี่ยวกับการปรับเทียบคะแนนแบบสอบด้วยโมเดล UIRT ซึ่งพบว่าเมื่อข้อมูลมีลักษณะของการกำหนดข้อมูลจำเพาะของโมเดลไม่ถูกต้อง การประมาณค่าพารามิเตอร์แบบแยกกันและใช้การเชื่อมโยงสเกลมีความคลาดเคลื่อน (ค่า MSE) ต่ำกว่าการประมาณ

ค่าพารามิเตอร์แบบพร้อมกัน (Béguin, Hanson, & Glas, 2000) และสอดคล้องกับอีกงานวิจัยหนึ่งซึ่งพบว่า การประมาณค่าพารามิเตอร์พร้อมกันมีความลำเอียงยกกำลังสอง (squared bias) มากกว่าวิธีการประมาณค่าพารามิเตอร์แยกกัน และเชื่อมโยงสเกลด้วยวิธี Mean/Sigma และวิธี Stocking-Lord เมื่อสัดส่วนข้อสอบร่วม คือ 10% (Bradley A. Hanson & Anton A. Béguin, 2002)

ด้วยเหตุนี้การปรับเทียบคะแนนแบบสอบพหุมิติรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม ควรเลือกใช้กระบวนการ MOSE ที่มีการประมาณค่าพารามิเตอร์แบบพร้อมกัน (CMOSE) อย่างไรก็ตามการประมาณค่าพารามิเตอร์แบบแยกกัน อาจนำมาใช้ในกรณีที่ต้องใช้ค่าพารามิเตอร์ข้อสอบสำหรับผู้สอบทั้งสองกลุ่ม เพื่อใช้ในการวินิจฉัยประเด็นปัญหาต่าง ๆ เช่น การระบุ DIF ของความแม่นยำและความถูกต้องของข้อสอบร่วม และการระบุ outlier ของค่าพารามิเตอร์ของข้อสอบที่ได้จากการประมาณแต่ละครั้ง สะท้อนถึงความแตกต่างระหว่างกลุ่ม และตัดข้อสอบร่วมที่ไม่มีคุณภาพ ซึ่งการประมาณค่าพารามิเตอร์พร้อมกัน ทำให้มีค่าพารามิเตอร์เพียงชุดเดียวกัน ไม่สามารถตรวจสอบปัญหาดังกล่าวได้ (Hanson, CTB/McGraw-Hill, & Béguin, 2002; Kolen & Brennan, 2004)

### 3. ประสิทธิภาพของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบร่วมแตกต่างกัน

จากผลการศึกษาประสิทธิภาพของกระบวนการ MOSE เมื่อสัดส่วนคะแนนข้อสอบร่วมแตกต่างกัน เมื่อพิจารณาในภาพรวมของกระบวนการ MOSE สำหรับเงื่อนไขสัดส่วนคะแนนข้อสอบร่วม พบว่า (1) ส่วนใหญ่มีประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องมากขึ้นมากที่สุด เมื่อสัดส่วนคะแนนข้อสอบร่วมเพิ่มขึ้น ซึ่งสอดคล้องกับสมมติฐานที่ตั้งไว้ (กระบวนการ MOSE ทั้ง 2 แบบ น่าจะมีค่า CVSE และค่า RMSE ลดลง เมื่อสัดส่วนคะแนนข้อสอบร่วมเพิ่มขึ้น) โดยสัดส่วนคะแนนข้อสอบร่วม 10% ส่วนใหญ่มีประสิทธิภาพโดยรวม ความแม่นยำและความถูกต้องต่ำกว่าสัดส่วนอื่น ๆ อย่างชัดเจน (2) บางส่วนประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องมากที่สุด เมื่อสัดส่วนคะแนนข้อสอบร่วม คือ 20%, 30% และ 10% ตามลำดับ และ (3) มีข้อค้นพบที่แตกต่าง คือ กระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน และอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 70:30 พบว่า มีความแม่นยำมากที่สุดเมื่อสัดส่วนคะแนนข้อสอบร่วม คือ 10%, 30% และ 20% ตามลำดับ

จากผลการวิจัยข้างต้นสะท้อนให้เห็นว่าการปรับเทียบคะแนนด้วยสัดส่วนคะแนนข้อสอบร่วม 30% และ 20% มีประสิทธิภาพมากกว่าสัดส่วนคะแนนข้อสอบร่วม 10% เนื่องจากเมื่อสัดส่วนคะแนนข้อสอบร่วมมีเพียง 10% ทำให้ข้อสอบร่วมมีจำนวนน้อยมาก เมื่อพิจารณาจำนวนข้อสอบสำหรับโครงสร้างที่ซับซ้อน ซึ่งแบ่งแบบสอบออกเป็น 4 กลุ่ม พบว่า สัดส่วนคะแนนข้อสอบร่วมสองค่า คือ 70:30, 60:40 และ 40:60 มีข้อสอบร่วมที่ตรวจให้คะแนนหลายค่า 1, 2 และ 3 ข้อ ตามลำดับ ทำให้แบบสอบกลุ่มย่อยแต่ละกลุ่มมีข้อสอบมีเพียงข้อเดียวหรือไม่มีข้อสอบ และเมื่อพิจารณาโครงสร้างที่ไม่ซับซ้อน ซึ่งแบ่งแบบสอบออกเป็น 2 กลุ่มย่อย พบว่า สัดส่วนคะแนนข้อสอบร่วมสองค่า คือ 70:30,

60:40 และ 40:60 ทำให้แบบสอบกลุ่มย่อยบางกลุ่มมีจำนวนข้อสอบมีเพียงข้อเดียว ส่งผลให้ข้อสอบรวมดังกล่าวไม่เป็นตัวแทนของเนื้อหา และคุณลักษณะทางสถิติของทั้งฉบับ ซึ่งในกรณีของการจำลองข้อมูลนั้น ถ้าการสุ่มเพียงค่าเดียวแล้วได้ค่าสุดโต่ง (outlier) ย่อมทำให้ค่าดังกล่าวเบี่ยงเบนไปจากค่าปกติ ซึ่งการสร้างข้อสอบรวมมีหลักสำคัญ คือ ข้อสอบรวมต้องมีความยาวเหมาะสม และเป็นตัวแทนทั้งด้านเนื้อหา และความยากของแบบสอบทั้งฉบับ (Kolen & Brennan, 2004; ศิริชัย กาญจนวาสิ, 2555) โดย Angoff และคณะ (1971) แนะนำว่า แบบสอบมีข้อสอบ 40 ข้อ ควรใช้จำนวนข้อสอบรวมอย่างต่ำร้อยละ 20 ของข้อสอบทั้งหมด และแบบสอบมีข้อสอบมากกว่า 40 ข้อ ควรใช้ข้อสอบรวมอย่างต่ำร้อยละ 30 ของข้อสอบทั้งหมด และยังมีข้อเสนอแนะที่ระบุว่า สัดส่วนข้อสอบรวมที่มีประสิทธิภาพในการเปรียบเทียบคะแนนสำหรับกลุ่มตัวอย่าง 400 คน คือ 10% และ 20% (Gao, 2004; McKinley & Reckase, 1981) โดยการเปรียบเทียบคะแนนส่วนใหญ่ใช้จำนวนข้อสอบรวมประมาณ 20-25% ของแบบสอบทั้งฉบับ (Kolen & Brennan, 2004)

งานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล UIRT สำหรับแบบสอบรูปแบบผสมที่มีจำนวนข้อสอบรวมแตกต่างกัน 3 รูปแบบ ได้แก่ 5, 10 และ 15 ข้อ คิดเป็น 10%, 20% และ 30% ตามลำดับ พบว่า (1) ค่า MSE ของการกู้คืนค่าพารามิเตอร์  $a$  และ  $b$  ลดลง เมื่อสัดส่วนข้อสอบรวมเพิ่มขึ้น (2) ค่า RMSD ลดลงเมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้น (เมื่อสัดส่วนข้อสอบรวมแต่ละประเภทคงที่) (Tian, 2011) และการเชื่อมโยงสเกลด้วยวิธี TCF สำหรับแบบสอบรูปแบบผสมที่มีโครงสร้างที่ไม่ซับซ้อนและซับซ้อน เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% 25% และ 30% พบว่า สัดส่วนคะแนนข้อสอบรวม 25% มีค่า Bias และค่า RMSE ต่ำที่สุด นั่นคือ การเชื่อมโยงสเกลในโครงสร้าง APSS และสัดส่วนน้ำหนักคะแนนข้อสอบรวม 25% มีประสิทธิภาพมากกว่าสัดส่วนอื่น ๆ (ศิระทัศน์ สุขสุวรรณ และคณะ, 2555)

เมื่อพิจารณาจำนวนข้อสอบรวมทั้งหมดของงานวิจัยครั้งนี้ สำหรับสัดส่วนคะแนนข้อสอบรวม 10%, 20% และ 30% คือ 7-9, 14-17 และ 20-26 ข้อ ตามลำดับ (จำนวนข้อสอบทั้งหมดในแต่ละสัดส่วน คือ 83-86 ข้อ, 82-85 ข้อ และ 80-86 ข้อ ตามลำดับ) เทียบเคียงกับงานวิจัยของ Yao และ Boughton (2009) ซึ่งพิจารณาประสิทธิภาพจากค่า RMSE และความลำเอียง พบว่า ผลการเชื่อมโยงสเกลในแต่ละเงื่อนไขมีประสิทธิภาพดี ยกเว้นเงื่อนไขข้อสอบรวม 6 ข้อ และเมื่อพิจารณาเฉพาะข้อสอบรวมรูปแบบผสม พบว่า จำนวนข้อสอบรวมที่มีประสิทธิภาพเมื่อเทียบกับเกณฑ์ คือ 16 ข้อ และ 32 ข้อ (คิดเป็น 28 คะแนน และ 56 คะแนน) จะเห็นว่าสอดคล้องกับจำนวนข้อสอบรวมในงานวิจัยครั้งนี้ สำหรับสัดส่วนข้อสอบรวม 20% และ 30% คือ 14-17 ข้อ และ 20-26 ข้อ ตามลำดับ อย่างไรก็ตาม Yao และ Boughton ยังพบว่า ข้อสอบรวม 8 ข้อ ก็ให้ผลการเชื่อมโยงสเกลดี ซึ่งสอดคล้องกับจำนวนข้อสอบรวม คือ 10% ของงานวิจัยครั้งนี้ คือ 7-9 ข้อ

งานวิจัยเกี่ยวกับการเชื่อมโยงสเกลแบบสอบพหุมิติอีกงานหนึ่ง (มีเฉพาะข้อสอบที่ตรวจให้คะแนนสองค่า) พบว่า สัดส่วนข้อสอบรวมน้อยที่สุดที่มีประสิทธิภาพ เมื่อความยาวของแบบสอบเป็น 160, 120, 100 และ 80 ข้อ คือ สัดส่วนข้อสอบรวม 20% เมื่อคิดเป็นจำนวนข้อสอบรวม คือ 20 ข้อ และ 16 ข้อ (Xiong et al., 2013) สอดคล้องกับงานวิจัยครั้งนี้ที่มีจำนวนข้อสอบทั้งหมด 86-80 ข้อ ซึ่งพบว่า จำนวนข้อสอบรวมสำหรับสัดส่วนคะแนนข้อสอบรวม 20% คือ 14-17 ข้อ ด้วยเหตุนี้จึงสามารถประมาณจำนวนข้อสอบรวมน้อยที่สุดที่มีประสิทธิภาพสำหรับสัดส่วนคะแนนข้อสอบรวม 10% คือ อย่างต่ำ 8-9 ข้อ ขณะที่จำนวนข้อสอบรวมน้อยที่สุดสำหรับสัดส่วนคะแนนข้อสอบรวม 20% และ 30% คือ อย่างต่ำ 14-16 ข้อ

ส่วนข้อค้นพบที่แตกต่างชัดเจน คือ กระบวนการ SMOSE สำหรับโครงสร้างที่ไม่ซับซ้อน และ อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า คือ 70:30 พบว่า มีความแม่นยำมากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10%, 30% และ 20% ตามลำดับ กล่าวคือ ค่า RB มีค่ามากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20%, 30% และ 10% ตามลำดับ โดยค่า RB ของสัดส่วนคะแนนข้อสอบรวม คือ 20% มีค่ามากที่สุดอย่างชัดเจน ขณะที่ความแม่นยำของสัดส่วน 30% และ 10% มีค่าใกล้เคียงกัน ซึ่งอาจเป็นเพราะสัดส่วนคะแนนข้อสอบรวม 20% มีจำนวนข้อสอบรวมในแบบสอบแต่ละกลุ่ม คือ 1 และ 2 ข้อ ซึ่งค่าที่สุ่มได้อาจเป็นค่าความยากที่ยากเกิน หรือง่ายเกินไป หรือค่าที่สุ่มได้อาจไม่ตัวแทนของแบบสอบ ในขณะที่เดียวกันสัดส่วนคะแนนข้อสอบรวม 10% จำนวนข้อสอบรวมในแบบสอบแต่ละกลุ่ม คือ 1 ข้อ อาจสุ่มได้ค่าที่ดีและเป็นตัวแทนของแบบสอบ และเมื่อพิจารณาความแตกต่างระหว่างค่ามุมของแบบสอบย่อยแต่ละกลุ่ม ดังตาราง 5.1 พบว่า ค่าเฉลี่ยของความแตกต่างดังกล่าวระหว่างข้อสอบรวมและข้อสอบเฉพาะมากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม 20%, 30 และ 10% ตามลำดับ (.74, .61 และ .57 ตามลำดับ) ด้วยเหตุนี้สัดส่วนคะแนนข้อสอบรวม 20% มีความแม่นยำต่ำที่สุดเป็นเพราะความคลาดเคลื่อนเนื่องจากการสุ่มตัวอย่างในการจำลองข้อมูล

จากผลการวิจัยสะท้อนให้เห็นว่า การเปรียบเทียบคะแนนแบบสอบพหุมิติรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบรวม ควรใช้สัดส่วนคะแนนข้อสอบรวม 20%-30% และ หลีกเลี่ยงสัดส่วนคะแนนข้อสอบรวม 10%

ตาราง 5.1 ความแตกต่างระหว่างค่ามุมในแต่ละกลุ่ม สำหรับสัดส่วนคะแนนข้อสอบรวม

สัดส่วน	แบบสอบเฉพาะ X				แบบสอบเฉพาะ Y				แบบสอบรวม (A)			
	กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
30%	8.04	81.72	81.96	7.85	6.18	83.29	83.82	6.29	7.11	82.72	82.89	7.28
20%	7.50	82.66	82.50	7.34	7.18	82.12	82.82	7.88	6.30	82.62	83.61	6.93
10%	8.06	82.86	81.94	7.14	6.25	83.40	83.75	6.60	7.94	82.06		
<b>ความแตกต่างของสัดส่วนคะแนนข้อสอบรวม 30%</b>												
X-A	0.93	0.57	0.93	0.57	ค่าเฉลี่ยความแตกต่าง = 0.61							
Y-A	0.66	0.22	0.56	0.42								
<b>ความแตกต่างของสัดส่วนคะแนนข้อสอบรวม 20%</b>												
X-A	1.20	0.03	1.11	0.41	ค่าเฉลี่ยความแตกต่าง = 0.74							
Y-A	0.89	0.50	0.79	0.95								
<b>ความแตกต่างของสัดส่วนคะแนนข้อสอบรวม 10%</b>												
X-A	0.11	0.81	-	-	ค่าเฉลี่ยความแตกต่าง = 0.57							
Y-A	1.69	1.34	-	-								

#### 4. ประสิทธิภาพของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าแตกต่างกัน

จากผลการศึกษาประสิทธิภาพของกระบวนการ MOSE เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าแตกต่างกัน ซึ่งพบว่า ในภาพรวมไม่สามารถระบุแนวโน้มของประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องว่า มีขนาดเพิ่มขึ้น เมื่อสัดส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนเพิ่มขึ้น หรือลดลงอย่างไร จากผลการวิจัยแบ่งการอภิปราย ดังนี้

- แนวโน้มของประสิทธิภาพโดยรวม ความแม่นยำ และความถูกต้องว่ามีขนาดเพิ่มขึ้น  
เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าเพิ่มขึ้นหรือลดลงยังไม่ชัดเจน นอกจากนี้ อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าที่มีประสิทธิภาพมากที่สุดมีทั้งรูปแบบที่สัดส่วนคะแนนอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่ามากกว่าหลายค่า และอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าน้อยกว่าหลายค่า ซึ่งไม่เป็นไปตามสมมติฐานที่ตั้งไว้ คือ “กระบวนการ MOSE ทั้ง 2 แบบ น่าจะมีค่า RB และ ค่า CVSE เพิ่มขึ้น เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าเพิ่มขึ้น” อย่างไรก็ตามเมื่อพิจารณาผลการวิจัยในภาพรวม พบว่า สัดส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60 สำหรับกระบวนการ MOSE มีแนวโน้มว่าความแม่นยำมากที่สุด เนื่องจากอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60 ส่วนใหญ่มีค่าเฉลี่ยของค่า RB ที่สุดเมื่อเทียบกับสัดส่วนอื่น ๆ

จากผลในภาพรวมที่พบว่า อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า 40:60 ส่วนใหญ่มีความแม่นยำมากที่สุด อาจเกิดจากสัดส่วนดังกล่าวมีจำนวนข้อสอบที่ตรวจให้คะแนนหลายค่ามากที่สุด ส่งผลให้จำนวนขั้นการตอบเพิ่มขึ้น ซึ่ง Kim (2004 cited in Tian, 2011) อธิบายความสัมพันธ์ระหว่างผลของการปรับเทียบคะแนนและประเภทของข้อสอบสอบร่วมไว้ว่า จำนวนของขั้นการตอบ (response categories) ส่งผลต่อกระบวนการเชื่อมโยงสเกล กล่าวคือ ถ้าจำนวนขั้นของการตอบเพิ่มขึ้น ผลของการเชื่อมโยงสเกลจะมีความถูกต้องแม่นยำ (accurate) มากขึ้น และเมื่อพิจารณางานวิจัยของ Yao และ Boughton (2009) ซึ่งพบว่า เจื่อนไขข้อสอบร่วมที่มีค่า RMSE และความลำเอียงต่ำเมื่อเทียบกับเกณฑ์ คือ (1) กรณีข้อสอบร่วม 16 ข้อ ประกอบด้วย ข้อสอบหลายตัวเลือกและสร้างคำตอบ คือ 12 ข้อ (12 คะแนน) และ 4 ข้อ (16 คะแนน) คิดเป็นอัตราส่วน 43:57 และ (2) ข้อสอบร่วม 32 ข้อ ประกอบด้วย ข้อสอบหลายตัวเลือกและสร้างคำตอบ คือ 24 ข้อ (24 คะแนน) และ 8 ข้อ (32 คะแนน) คิดเป็นอัตราส่วน 43:57 ซึ่งใกล้เคียงกับอัตราส่วนของงานวิจัยครั้งนี้ คือ 40:60

นอกจากนี้งานวิจัยที่ผ่านมา พบว่า ประสิทธิภาพของการปรับเทียบคะแนน/เชื่อมโยงสเกลเพิ่มขึ้น เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าลดลง อย่างเช่น งานวิจัยที่ศึกษาการปรับเทียบคะแนนด้วยโมเดล UIRT ระหว่างวิธีประมาณค่าพารามิเตอร์แบบแยกกัน (วิธีของ Stocking-Lord) และวิธีประมาณค่าพารามิเตอร์แบบพร้อมกัน สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT ซึ่งพบว่า ประเภทของข้อสอบร่วมมีผลต่อประสิทธิภาพในการกู้คืนค่าพารามิเตอร์ของวิธีการเชื่อมโยงทั้ง 2 วิธี โดยข้อสอบร่วมแบบสร้างคำตอบอย่างเดียวมีค่า MSE น้อยที่สุด และค่า MSE มีค่าลดลงเมื่อจำนวนข้อสอบร่วมแบบสร้างคำตอบมากขึ้น (Tian, 2011) งานวิจัยที่ศึกษาการปรับเทียบคะแนนด้วยโมเดล UIRT เมื่ออัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 50:50, 60:40, 70:30 และ 80:20 ซึ่งพบว่า ประสิทธิภาพของวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความคลาดเคลื่อนต่ำลงเมื่อสัดส่วนของข้อสอบที่ตรวจให้คะแนนสองค่าในแบบสอบเพิ่มขึ้นและสัดส่วนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าในแบบสอบร่วมลดลง (อัญชลี ศรีกลชาญ, 2552) เช่นเดียวกับงานวิจัยที่ศึกษาการเชื่อมโยงสเกลแบบแนวตั้ง ซึ่งพบว่า ความถูกต้องของการเชื่อมโยงสเกลเพิ่มขึ้น เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ขึ้น ใช้ข้อสอบร่วมรูปแบบผสม จำนวนข้อสอบร่วมเพิ่มขึ้น และข้อสอบร่วมที่ตรวจให้คะแนนหลายค่าเพิ่มขึ้น (Meng, 2007)

- ความถูกต้องระหว่างรูปแบบของอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าไม่แตกต่างกัน อาจเป็นเพราะจำนวนข้อสอบร่วมในแต่ละสัดส่วนไม่ได้แตกต่างกันนัก ทำให้ผลด้านความถูกต้อง ซึ่งพิจารณาความแตกต่างระหว่างคะแนนปรับเทียบจากวิธีศึกษา กับค่าเฉลี่ยของคะแนนปรับเทียบจากวิธีศึกษาในแต่ละสัดส่วนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ โดยเฉพาะอย่างยิ่งจำนวนข้อสอบร่วมที่ตรวจให้คะแนนหลายค่าในสัดส่วนคะแนนข้อสอบร่วม 20% และ 10% ที่ต่างกันเพียงข้อเดียว กล่าวคือ เมื่อสัดส่วนคะแนนข้อสอบร่วม 10% จำนวนข้อสอบร่วมที่ตรวจให้คะแนนหลาย

ค่าสำหรับสัดส่วน 70:30, 60:40 และ 40:60 คือ 1, 2 และ 3 ข้อ ตามลำดับ เมื่อสัดส่วนคะแนนข้อสอบรวม 20% จำนวนข้อสอบรวมที่ตรวจให้คะแนนหลายค่าสำหรับสัดส่วน 70:30, 60:40, 50:50 และ 40:60 คือ 3, 4, 5 และ 6 ข้อ ตามลำดับ และเมื่อสัดส่วนคะแนนข้อสอบรวม 20% จำนวนข้อสอบรวมที่ตรวจให้คะแนนหลายค่าสำหรับสัดส่วน 70:30, 60:40, 50:50 และ 40:60 คือ 3, 4, 5 และ 6 ข้อ ตามลำดับ ขณะที่จำนวนข้อสอบรวมที่ตรวจให้คะแนนหลายค่าในสัดส่วน 20% และ 10% ต่างกันเพียง 2 ข้อ โดยจำนวนข้อสอบรวมที่ตรวจให้คะแนนหลายค่าสำหรับสัดส่วน 70:30, 60:40, 50:50 และ 40:60 คือ 4, 6, 8 และ 10 ข้อ ตามลำดับ

จากผลการวิจัยข้างต้น สะท้อนให้เห็นว่า อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า ไม่ได้มีผลต่อประสิทธิภาพในการเปรียบเทียบคะแนนมากนัก อย่างไรก็ตาม หากต้องระบุอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าที่มีประสิทธิภาพในการเปรียบเทียบคะแนน พบว่า การเปรียบเทียบคะแนนด้วยกระบวนการ MOSE สำหรับสัดส่วนคะแนนข้อสอบรวม 30% และ 20% ควรเลือกใช้อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60 เนื่องจากในภาพรวม พบว่า สัดส่วนดังกล่าวมีประสิทธิภาพในภาพรวม ความแม่นยำ และความถูกต้องมากกว่าสัดส่วนอื่น ๆ แต่เมื่อพิจารณาเฉพาะกระบวนการ CMOSE สำหรับสัดส่วนคะแนนข้อสอบรวม 30% และ 20% สามารถเลือกใช้อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าได้ทั้ง 40:60 และ 60:40

##### 5. ประสิทธิภาพของกระบวนการ MOSE เมื่อโครงสร้างมิติตามความสามารถแตกต่างกัน

จากผลการศึกษาประสิทธิภาพของกระบวนการ MOSE เมื่อโครงสร้างมิติตามความสามารถแตกต่างกัน พบว่า ในภาพรวม โครงสร้างที่ซับซ้อน (CS) ส่วนใหญ่มีประสิทธิภาพโดยรวมและความแม่นยำมากกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) แต่ความถูกต้องระหว่างโครงสร้างมิติตามความสามารถทั้ง 2 แบบ ไม่มีความแตกต่างกันในเชิงปฏิบัติ (เนื่องจากมีขนาดอิทธิพลเพียงร้อยละ .30)

ผลการวิจัยแบ่งการอภิปราย ดังนี้

- โครงสร้างที่ซับซ้อน (CS) ส่วนใหญ่มีประสิทธิภาพโดยรวมและความแม่นยำมากกว่าโครงสร้างที่ไม่ซับซ้อน (ASS) ซึ่งไม่สอดคล้องกับสมมติฐานที่ตั้งไว้ คือ กระบวนการ MOSE สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) น่าจะมีค่า RB และค่า RMSE มากกว่าโครงสร้างที่ซับซ้อน (CS) อย่างไรก็ตามมีงานวิจัยบางส่วนเป็นไปตามข้อค้นพบของงานวิจัยครั้งนี้ กล่าวคือ การเชื่อมโยงสเกลด้วยโมเดล MIRT จากระดับที่ 1 ไป ระดับที่ 2 และระดับที่ 2 ไประดับที่ 3 สำหรับโครงสร้าง APSS ส่วนใหญ่มีค่า RMSE มากกว่าโครงสร้าง CS แต่ผลของการวิจัยดังกล่าวมีบางส่วนที่ไม่สอดคล้องกับงานวิจัยในครั้งนี้ กล่าวคือ โครงสร้าง CS ส่วนใหญ่มีความลำเอียงมากกว่าโครงสร้าง APSS สำหรับการเชื่อมโยงสเกลในระดับที่ 1 ไประดับที่ 2 (ศิวัชต์ศน์ สุขสุวรรณ, 2555) เช่นเดียวกับงานวิจัยของพัชรี จันทร์เพ็ง (2550) ที่พบว่าผลการวิจัยบางส่วนสอดคล้องกับงานวิจัยครั้งนี้ โดยการเชื่อมโยงสเกลด้วยวิธี M และวิธี NOP จากระดับที่ 2 ไประดับที่ 3

สำหรับโครงสร้าง APSS มีค่า RMSE มากกว่าโครงสร้าง CS เมื่อความสัมพันธ์ระหว่างมิติ คือ 0.5 และผลการวิจัยบางส่วนแตกต่างกัน คือ การเชื่อมโยงสเกลด้วยวิธี M และวิธี NOP จากระดับที่ 2 ไประดับที่ 3 สำหรับโครงสร้าง CS มีความลำเอียงมากกว่าโครงสร้าง APSS

เมื่อพิจารณาการแบ่งกลุ่มของแบบสอบที่มีโครงสร้าง CS จะเห็นว่า แบบสอบที่มีโครงสร้าง CS แบ่งแบบสอบออกเป็น 4 ชุดย่อย โดยแบบสอบ 2 ชุดย่อยจะเน้นการวัดความสามารถมิติที่ 1 และมิติที่ 2 ขณะที่แบบสอบอีก 2 ชุดย่อยวัดความสามารถในภาพรวมระหว่างมิติ ซึ่งอาจทำให้วัดความสามารถได้ครอบคลุมมากขึ้นทั้งความสามารถเฉพาะในแต่ละมิติและความสามารถร่วมกันทั้งสองมิติ ซึ่งงานวิจัยของพัชรี จันทร์เพ็ง (2550) ได้วิเคราะห์ความเที่ยงของข้อมูลที่จำลอง พบว่า เมื่อเพิ่มจำนวนข้อสอบรวมจาก 20 ข้อเป็น 30 ข้อ ทำให้ความเที่ยงของแบบสอบรวมเพิ่มขึ้น โดยแบบสอบที่มีโครงสร้าง CS มีความเที่ยงมากกว่าโครงสร้าง APSS ด้วยเหตุนี้โครงสร้าง CS ส่วนใหญ่จึงมีประสิทธิภาพโดยรวมและความแม่นยำมากกว่าโครงสร้าง APSS

- ความถูกต้องระหว่างโครงสร้างมิติความสามารถทั้ง 2 แบบไม่มีความแตกต่างกันในเชิงปฏิบัติ (เนื่องจากมีขนาดอิทธิพลเพียงร้อยละ .30) สอดคล้องกับผลการวิจัยของ Zhang (2012) ที่ศึกษาการปรับเทียบคะแนนแบบสอบพหุมิติ สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT ซึ่งพบว่า โครงสร้างมิติความสามารถมีผลน้อยมากต่อประสิทธิภาพของการปรับเทียบคะแนน อย่างไรก็ตาม โครงสร้าง APSS ส่วนใหญ่มีความถูกต้องมากกว่าโครงสร้าง CS เนื่องจากการประมาณค่าพารามิเตอร์สำหรับโครงสร้าง APSS มีความถูกต้องมากกว่าและง่ายกว่าโครงสร้าง CS สอดคล้องกับงานวิจัยที่ศึกษาวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT 4 วิธี วิธี ได้แก่ (1) Direct (2) Equated Function (3) TCF และ (4) ICF ภายใต้รูปแบบ NEAT ที่พบว่า โครงสร้างที่ไม่ซับซ้อนมีค่า MSE ต่ำกว่าโครงสร้างที่ซับซ้อน แต่อย่างไรก็ตามผลการเชื่อมโยงระหว่างโครงสร้างทั้งสองแบบใกล้เคียงกัน เมื่อกลุ่มตัวอย่างมีขนาดใหญ่ คือ 2,000 คน (Wei, 2008)

จากข้อค้นพบข้างต้น สะท้อนให้เห็นว่า ผลการวิจัยเกี่ยวกับโครงสร้างมิติความสามารถมีความหลากหลาย และงานวิจัยเดียวกันมีทั้งผลที่เหมือนกันและต่างกัน ด้วยเหตุนี้โครงสร้างมิติความสามารถอาจไม่มีอิทธิพลมากนักต่อประสิทธิภาพในการปรับเทียบคะแนน อย่างไรก็ตามในภาพรวมผลการวิจัยครั้งนี้พบว่า โครงสร้างที่ซับซ้อน (CS) มีความแม่นยำและมีประสิทธิภาพโดยรวมมากกว่าโครงสร้างที่ไม่ซับซ้อน (APSS) และแม้ว่าโครงสร้างที่ไม่ซับซ้อน (APSS) จะมีความถูกต้องมากกว่า แต่ค่าความถูกต้องระหว่างโครงสร้างมิติความสามารถค่อนข้างใกล้เคียงกัน ซึ่งการปรับเทียบคะแนนแบบสอบพหุมิตินี้ยังไม่มีการศึกษาเงื่อนไขโครงสร้างมิติความสามารถ ดังนั้นหากต้องการเลือกรูปแบบของโครงสร้างมิติความสามารถ ในเบื้องต้นควรเลือกแบบสอบที่มีโครงสร้างที่ซับซ้อน แต่ถ้าต้องการใช้โครงสร้างที่ไม่ซับซ้อน ควรเลือกสัดส่วนคะแนนข้อสอบรวม 30% เนื่องจากโครงสร้างที่ไม่ซับซ้อนสำหรับสัดส่วนคะแนนข้อสอบรวม 30% มีความแม่นยำ ความถูกต้อง และประสิทธิภาพโดยรวมมากกว่าสัดส่วนอื่น ๆ



## 6. อภิปรายเพิ่มเติมในประเด็นที่น่าสนใจ

การอภิปรายเพิ่มเติมในประเด็นที่น่าสนใจ ได้แก่ วิธีอีควิเปอร์เซ็นไทล์ที่มีการปรับโค้งให้ราบเรียบ (presmoothed equipercentile equating method) และการกำหนดวิธีปรับเทียบคะแนนเกณฑ์ รายละเอียดดังนี้

### 1. วิธีอีควิเปอร์เซ็นไทล์ที่มีการปรับโค้งให้ราบเรียบ

การปรับเทียบคะแนนด้วยวิธีอีควิเปอร์เซ็นไทล์จะมีประสิทธิภาพเพิ่มขึ้น ด้วยวิธีการปรับโค้ง (smoothing method) ซึ่งใช้เพื่อปรับโค้งคะแนนของแบบสอบที่มีลักษณะผิดปกติอย่างเป็นระบบ (systematic irregularities) ที่อาจเกิดจากความแปรปรวนในการสุ่มตัวอย่าง (sampling variability) และ/หรือความคลาดเคลื่อนอย่างเป็นระบบที่เกิดจากการให้คะแนน โดยคาดหวังว่าการปรับโค้งจะทำให้การประมาณค่ามีความแม่นยำมากกว่าขึ้น (Kolen & Brennan, 2004)

งานวิจัยที่ผ่านมาส่วนใหญ่ พบว่า วิธีอีควิเปอร์เซ็นไทล์ที่มีการปรับโค้งให้ราบเรียบมีความคลาดเคลื่อนมาตรฐานต่ำกว่าวิธีอีควิเปอร์เซ็นไทล์แบบดั้งเดิม และโดยทั่วไปคะแนนช่วงต้นและช่วงท้ายจะมีจำนวนผู้สอบน้อยกว่าเมื่อเทียบกับคะแนนช่วงกลาง จึงอาจทำให้มีความคลาดเคลื่อนของคะแนนช่วงต้นและท้ายมีค่าสูง ซึ่งงานวิจัยครั้งนี้ พบว่า คะแนนช่วงต้น มีค่า CVSE มากกว่าคะแนนช่วงกลาง เนื่องจากคะแนนช่วงต้นมีการแจกแจงน้อย กล่าวคือ มีผู้สอบจำนวนน้อยที่ได้คะแนนช่วงต้น (โดยเฉพาะช่วง 1-6 คะแนน) แต่ถ้ามีการปรับโค้งโดยมีข้อตกลงเบื้องต้น คือ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานก่อน-หลังปรับโค้งต้องเหมือนเดิม ย่อมทำให้การแจกแจงของข้อมูลจากแบบสอบแต่ละฉบับมีลักษณะปกติ และเมื่อทำการปรับเทียบคะแนนจากข้อมูลที่ได้รับปรับโค้งแล้ว ผลการปรับเทียบที่ได้จึงมีความคลาดเคลื่อนลดลง (Brossman, 2010; Kolen & Brennan, 2004)

กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MOSE) สำหรับงานวิจัยครั้งนี้ใช้วิธีอีควิเปอร์เซ็นไทล์ในการปรับเทียบคะแนนระหว่างแบบสอบ โดยไม่มีการปรับโค้งข้อมูลทั้งสองชุดก่อนการปรับเทียบคะแนน เช่นเดียวกับงานวิจัยที่ผ่านมา ดังนั้นหากต้องการพัฒนากระบวนการ MOSE ให้มีประสิทธิภาพมากยิ่งขึ้น อาจเพิ่มการปรับโค้งข้อมูลจากแบบสอบแต่ละฉบับที่จะใช้ในการปรับเทียบคะแนน ซึ่งก็คือ *Marginal observed scored distribution* หลังจากนั้นจึงนำข้อมูลที่ได้รับการปรับโค้งแล้วมาปรับเทียบคะแนนด้วยวิธีอีควิเปอร์เซ็นไทล์ต่อไป

## 2) การกำหนดวิธีเปรียบเทียบคะแนนเกณฑ์

การประเมินหรือเปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบคะแนนต่าง ๆ ยังไม่มีการกำหนดวิธีเกณฑ์อย่างชัดเจน ซึ่งงานวิจัยที่ศึกษาการปรับเทียบคะแนนด้วยโมเดล MIRT ก็ใช้วิธีเกณฑ์แตกต่างกัน อย่างไรก็ตามวิธีหนึ่งที่นำมาใช้เป็นเกณฑ์ในเปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบคะแนนแบบสอบพหุมิติ คือ วิธีอีควิเปอร์เซ็นไทล์ที่ปรับโค้งให้ราบเรียบ โดย Brossman (2010) ระบุว่า วิธีอีควิเปอร์เซ็นไทล์มีความเหมาะสมที่จะใช้เป็นเกณฑ์ เนื่องจากไม่จำเป็นต้องคำนึงถึงมิติของข้อมูล และผลการปรับเทียบคะแนนด้วยวิธีดังกล่าวยังสอดคล้อง (align) อย่างมากกับวิธีการปรับเทียบคะแนนด้วยโมเดล MIRT เนื่องจากลดค่า likelihood ที่ทำให้เกิดความลำเอียง นอกจากนี้วิธีอีควิเปอร์เซ็นไทล์ยังมีความคล้ายคลึงกับวิธี IRT เนื่องจากสหสัมพันธ์ระหว่างของการปรับเทียบคะแนนเป็นเส้นโค้ง และเมื่อนำวิธีการปรับโค้ง (smoothing method) มาใช้ร่วมด้วยยิ่งทำให้วิธีเกณฑ์ดังกล่าวมีประสิทธิภาพเพิ่มขึ้น (Kolen & Brennan, 2004)

ดังนั้นการวิจัยครั้งนี้จึงกำหนดให้วิธีอีควิเปอร์เซ็นไทล์ที่มีการปรับโค้งให้ราบเรียบ (presmoothed equipercentile equating method) เป็นเกณฑ์ในการเปรียบเทียบและศึกษาประสิทธิภาพของกระบวนการ CMOSE และกระบวนการ SMOSE เพื่อคำนวณความลำเอียงสัมพัทธ์ (ค่า RB) อย่างไรก็ตาม หากต้องการกำหนดวิธีการปรับเทียบคะแนนเกณฑ์ให้มีความน่าเชื่อถือยิ่งขึ้น อาจเปรียบเทียบวิธีเกณฑ์หลาย ๆ วิธี ว่าวิธีการใดมีแม่นยำและถูกต้องมากที่สุด



## ข้อเสนอแนะ

### 1. ข้อเสนอแนะในการนำไปปฏิบัติ

การปรับเทียบคะแนนแบบสอบพหุมิติรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกัน โดยใช้ข้อสอบร่วม โดยแบบสอบประกอบด้วยข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่าที่มีอัตราส่วนคะแนน 70:30 หรือใกล้เคียง ควรเลือกใช้

1.1 กระบวนการ CMOSE กล่าวคือ ประมาณค่าพารามิเตอร์ระหว่างแบบสอบแต่ละฉบับพร้อมกัน (concurrent calibration) ภายใต้โมเดลการตอบสนองข้อสอบแบบพหุมิติ และคำนวณเวกเตอร์ความสามารถแบบพหุตัวแปร (multivariate distribution) หลังจากนั้นจึงนำค่าพารามิเตอร์และความสามารถเหล่านั้นมาคำนวณ Marginal observed score distributions เพื่อใช้ในการปรับเทียบคะแนนต่อไป

1.2 สัดส่วนคะแนนข้อสอบร่วม คือ 20% หรือ 30% โดยกรณีแบบสอบมีจำนวน 80-86 ข้อหรือใกล้เคียง สามารถใช้ผลการวิจัยครั้งนี้เป็นแนวทางในการระบุจำนวนข้อสอบร่วม กล่าวคือ จำนวนข้อสอบร่วมน้อยที่สุด สำหรับสัดส่วนคะแนนข้อสอบร่วม 20% และ 30% คือ อย่างต่ำ 14-16 ข้อ แต่กรณีจำเป็นต้องใช้สัดส่วนคะแนนข้อสอบร่วม 10% อาจกำหนดจำนวนข้อสอบร่วมอย่างต่ำ คือ 8-9 ข้อ โดยต้องสร้างข้อสอบร่วมให้มีคุณภาพสูง กล่าวคือ เป็นตัวแทนของแบบสอบฉบับทั้งด้าน โครงสร้าง ความเที่ยง และความยาก อย่างไรก็ตาม หากคะแนนเต็มของข้อสอบที่ตรวจให้คะแนนหลายค่าเพิ่มขึ้น จำนวนข้อสอบร่วมอาจลดลงได้ แต่ถ้าคะแนนเต็มของข้อสอบที่ตรวจให้คะแนนหลายค่าลดลง จำนวนข้อสอบร่วมอาจเพิ่มขึ้นได้เช่นกัน

1.3 อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 40:60 หรือ 60:40 สำหรับสัดส่วนคะแนนข้อสอบร่วม คือ 20% หรือ 30% เนื่องจากประสิทธิภาพระหว่างอัตราส่วนทั้ง 2 แบบใกล้เคียงกัน แต่ถ้าสัดส่วนคะแนนข้อสอบร่วม คือ 10% ควรเลือกใช้อัตราส่วน 40:60 อย่างไรก็ตามอัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่าไม่ได้ส่งผลต่อประสิทธิภาพของการปรับเทียบคะแนนมากนัก ดังนั้นการใช้อัตราส่วนดังกล่าวกับสถานการณ์จริงสามารถปรับเปลี่ยนได้ตามความเหมาะสม

1.4 โครงสร้างมิติความสามารถ คือ โครงสร้างที่ซับซ้อน (CS) แต่ถ้าต้องการใช้แบบสอบที่มีโครงสร้างที่ไม่ซับซ้อน (APSS) ควรเลือกใช้อัตราส่วนคะแนนข้อสอบร่วมที่ตรวจให้คะแนนสองค่า คือ 40:60 อย่างไรก็ตาม โครงสร้างมิติความสามารถไม่ได้ส่งผลต่อประสิทธิภาพของการปรับเทียบคะแนนมากนัก หากสถานการณ์การทดสอบจริงไม่สามารถกำหนดลักษณะโครงสร้างมิติความสามารถได้ยังสามารถใช้แนวทางการของงานวิจัยครั้งนี้ในการกำหนดตัวแปรต่าง ๆ ได้เช่นกัน

กรณีต้องปรับเทียบคะแนนด้วยกระบวนการ SMOSE ซึ่งมีการประมาณค่าพารามิเตอร์ของแบบสอบสองฉบับหรือแบบสอบแต่ละฉบับแยกกัน (separate calibration) สามารถใช้แนวทางข้างต้นใน

การกำหนดสัดส่วนคะแนนข้อสอบรวม อัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า และ โครงสร้างมิตีความสามารถได้เช่นกัน

## 2. ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

2.1 ควรนำกระบวนการ CMOSE ไปทดลองใช้กับแบบสอบพหุมิติ ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบรวม (ข้อมูลจริง) ซึ่งเมื่อวิเคราะห์คุณภาพของแบบสอบแล้ว แบบสอบที่ใช้จริงอาจเป็นโครงสร้างที่ไม่ซับซ้อนหรือซับซ้อนก็ได้ โดยสัดส่วนคะแนนข้อสอบรวมและอัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่าอาจแตกต่างกันไปจากสถานการณ์จำลอง ซึ่งถ้าผลของวิจัยพบว่า กระบวนการ CMOSE มีประสิทธิภาพดี ย่อมเป็นการยืนยันประสิทธิภาพและขยายขอบเขตการใช้งานของกระบวนการ CMOSE ให้มีความหลากหลาย แต่ถ้าผลการวิจัยพบว่า กระบวนการ CMOSE ยังมีข้อจำกัด ก็จะได้ข้อค้นพบที่เป็นประโยชน์ในเลือกใช้กระบวนการดังกล่าวให้เหมาะสมกับบริบทของการเปรียบเทียบคะแนนต่อไป

2.2 การเปรียบเทียบคะแนนแบบสอบพหุมิติกับข้อมูลจริง ในขั้นตอนของการประมาณค่าพารามิเตอร์ อาจเลือกใช้โปรแกรมการวิเคราะห์ข้อสอบสำเร็จรูป อย่างเช่น โปรแกรม flexMIRT ซึ่งนิยมใช้ในการวิจัยเรื่องการเปรียบเทียบคะแนนแบบสอบพหุมิติรูปแบบผสม เนื่องจาก ง่ายต่อการใช้งาน ไม่ต้องเขียนคำสั่งเพื่อวิเคราะห์ข้อสอบ ซึ่งผู้วิจัยได้ทดลองใช้โปรแกรม flexMIRT และ แพคเกจ “mirt” กับข้อมูลแบบสอบจำนวนหนึ่ง พบว่า โปรแกรมและแพคเกจดังกล่าวได้ผลการวิเคราะห์ข้อสอบเหมือนกัน แต่ โปรแกรม flexMIRT ใช้เวลาดำเนินการน้อยกว่า แต่ถ้าต้องการจำลองข้อมูลที่มีเงื่อนไขหลากหลาย หรือต้องการกำหนดค่ามุมระหว่างมิติของข้อสอบ โปรแกรม flexMIRT ไม่สามารถทำได้ จำเป็นต้องเขียนคำสั่งในโปรแกรม R หรือโปรแกรม MATLAB แต่ถ้าการจำลองข้อมูลเป็นเพียงการกำหนดช่วงค่าพารามิเตอร์ สามารถเลือกใช้โปรแกรม flexMIRT ได้

2.3 ควรศึกษากับโมเดลตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติแบบทดแทนไม่ได้ (non-compensatory model) โดยผู้สอบที่มีความสามารถในมิติหนึ่งต่ำไม่สามารถทดแทนได้ จุดอ่อนนี้ได้ด้วยความสามารถ/จุดแข็งในอีกมิติหนึ่งได้ เนื่องจากงานวิจัยในครั้งนี้และงานวิจัยที่ผ่านมาศึกษากับโมเดล MIRT แบบทดแทนได้เท่านั้น

2.4 การเปรียบเทียบคะแนนแบบสอบพหุมิติรูปแบบผสมที่ประมาณค่าพารามิเตอร์แบบแยกกัน อาจเลือกใช้วิธีการเชื่อมโยงสเกลอื่น ๆ อย่างเช่น วิธี Item characteristic function (ICF) ซึ่งเป็นการเชื่อมโยงสเกลที่ให้ประสิทธิภาพที่ดีเช่นกัน โดย Zhang (2012) ที่ศึกษาการเปรียบเทียบคะแนนแบบสอบพหุมิติ สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT พบว่า การเชื่อมโยงสเกลด้วยวิธี TCF และ ICF มีประสิทธิภาพดีกว่าวิธี OD วิธี M และวิธี NOP

2.5 ควรศึกษากับแบบสอบสั้น หรือแบบสอบที่มีจำนวนข้อสอบน้อยลง เพื่อให้สอดคล้องกับสถานการณ์การทดสอบวัดผลสัมฤทธิ์ทางการเรียนในระดับชั้นเรียนมากขึ้น เนื่องจากงานวิจัยครั้งนี้ศึกษาแบบสอบยาวที่มีจำนวน 86-80 ข้อ เพื่อให้สามารถศึกษาเงื่อนไขต่าง ๆ ได้ครบถ้วน โดยการใช้แบบสอบสั้นไม่จำเป็นต้องมีคะแนนรวม 100 คะแนน หรือกรณีที่ต้องการคะแนนเต็ม 100 คะแนน อาจปรับคะแนนเต็มของข้อสอบที่ตรวจให้คะแนนหลายค่าเพิ่มขึ้น เพื่อลดจำนวนข้อสอบลง

2.6 ควรศึกษากับข้อสอบที่ตรวจให้คะแนนหลายค่าที่มีหลายระดับคะแนนมากขึ้น อย่างเช่น 5 คะแนน (0, 1, 2, 3, 4 และ 5 คะแนน) เป็นต้น เนื่องจากงานวิจัยครั้งนี้ศึกษากับข้อสอบที่ตรวจให้คะแนนสองค่าเพียง 3 ระดับ คือ 0, 1 และ 2 คะแนน ซึ่งศึกษาได้ทั้งการจำลองข้อมูลหรือข้อมูลจริง

2.7 การเปรียบเทียบคะแนนในขั้นตอนการคำนวณ Conditional observed score distributions และ Marginal observed score distributions ใช้เวลานาน ดังนั้นในการจำลองข้อมูลที่มีการกระทำซ้ำหลายรอบ หรือมีจำนวนผู้สอบและข้อสอบจำนวนมาก ควรเผื่อเวลาสำหรับการดำเนินการที่มากพอ

### 3. ข้อเสนอแนะเชิงทฤษฎี

กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MOSE) ในปัจจุบันใช้วิธีอีควิเปอร์เซ็นไทล์แบบดั้งเดิมในการปรับเทียบคะแนนระหว่างแบบสอบ โดยข้อมูลที่ใช้ในการปรับเทียบอยู่ในลักษณะของการแจกแจง คือ Marginal observed score distribution ซึ่งจากศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง อาจพัฒนาประสิทธิภาพของกระบวนการ MOSE ให้ดียิ่งขึ้นด้วยการใช้วิธีการปรับโค้ง (smoothing method) มาปรับโค้งการแจกแจงของข้อมูลจะใช้ในการปรับเทียบคะแนนให้มีลักษณะปกติ โดยเฉพาะอย่างยิ่งเมื่อการแจกแจงของข้อมูลที่มีความผิดปกติอย่างเป็นระบบที่อาจเกิดมาจากการสุ่มตัวอย่างหรือการให้คะแนน วิธีการปรับโค้งดังกล่าวจะทำให้การแจกแจงมีลักษณะปกติมากขึ้น โดยที่ค่าเฉลี่ยและส่วนเบี่ยงเบนไม่แตกต่างจากเดิม หลังจากนั้นจึงนำ Marginal observed score distribution ที่ได้รับการปรับโค้งแล้ว มาปรับเทียบคะแนนด้วยวิธีอีควิเปอร์เซ็นไทล์

## รายการอ้างอิง

### ภาษาไทย

- พัชรี จันทรเพ็ญ. (2550). การเปรียบเทียบคุณภาพของวิธีการเชื่อมโยงคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติภายใต้การหมุนแกน โครงสร้างเชิงมิติและระดับความสัมพันธ์ที่แตกต่างกัน. (ปริญาญาดุษฎีบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่ (4 ed.). กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ศิวะทัศน์ สุขสุวรรณ. (2555). การศึกษาคุณภาพของการเชื่อมโยงคะแนนแนวตั้งแบบพหุมิติสำหรับแบบทดสอบรูปแบบผสม. (ดุษฎีบัณฑิต), มหาวิทยาลัยขอนแก่น, บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น.
- ศิวะทัศน์ สุขสุวรรณ, พัชรี จันทรเพ็ญ, & สัจจวรรณ ังดกระโทก. (2555). คุณภาพของการเชื่อมโยงคะแนนแนวตั้งแบบพหุมิติสำหรับแบบทดสอบ รูปแบบผสมด้วยวิธีโค้งลักษณะแบบทดสอบ. วารสารศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น, 35(2), 109-118.
- อนันดา สันฐิตวิณชัย. (2556). การพัฒนาลักษณะเฉพาะของแบบสอบที่บูรณาการระหว่างตัวชี้วัดความสามารถด้านการอ่าน การคิดวิเคราะห์ และการเขียน กับตัวชี้วัดการประเมินสาระการเรียนรู้: การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ. (ปริญาญาดุษฎีบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย.
- อัญชลี ศรีกลชาญ. (2552). คุณภาพของการปรับเทียบคะแนนสำหรับแบบสอบรูปแบบผสม : การประยุกต์ใช้การปรับเทียบตามทฤษฎีการตอบสนองข้อสอบ ด้วยวิธีโค้งคุณลักษณะและการปรับค่าพารามิเตอร์พร้อมกัน. (ปริญาญาดุษฎีบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย.

### ภาษาอังกฤษ

- Albano, A. D. (2014). equate: An R Package for Observed-Score Linking and Equating. <http://CRAN.R-project.org/package=equate>.
- Andersson, B., Branberg, K., & Wiberg, M. (2013). Test equating using the Kernel method with the R package kequate. *Journal of Statistical Software*, 55(6).
- Angoff, W. H., Thorndike, R. L., & Lindquist, E. F. (1971). *Educational measurement* (2nd ed.). Washington: American Council on Education.

- Béguin, A., Hanson, B. A., & Glas, C. A. W. (2000). *Effect of multidimensionality on separate and concurrent estimation in IRT equating*. Paper presented at the the 2000 annual meeting of the National Council of Measurement in Education, New Orleans, IL.  
<https://www.researchgate.net/publication/270584999>
- Bradley A. Hanson, C. M.-H., & Anton A. Béguin. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26(1), 3–24.
- Brennan, R. L., Wang, T., Kim, S., & Seol, J. (2009). *Equating Recipes*. Retrieved from Center for Advanced Studies in Measurement and Assessment (CASMA):
- Brossman, B. G. (2010). *Observed score and true score equating procedures for multidimensional item response theory*. (Doctor of Philosophy), University of Iowa. Retrieved from <http://ir.uiowa.edu/etd/469>
- Brossman, B. G., & Lee, W.-C. (2013). Observed Score and True Score Equating Procedures for Multidimensional Item Response Theory. *Applied Psychological Measurement*, 37(6), 460–481. doi:10.1177/0146621613484083
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6).
- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). *Principles and Practices of Test Score Equating*. Retrieved from Princeton, New Jersey:
- Gao, H. (2004). *The effect of different anchor tests on the accuracy of test equating for test adaptation*. (Doctor of Philosophy), Ohio university.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., . . . Hothorn, T. (2016). Package ‘mvtnorm’. <http://mvtnorm.R-forge.R-project.org>
- Gonzalez, J. (2014). SNSequate: standard and nonstandard statistical models and methods for test equating. *Journal of Statistical Software*, 59(7).
- Han, K. T. (2009). IRTEQ: windows application that Implements item response theory scaling and equating. *Applied Psychological Measurement*, 33(6).

- Han, K. T., & Hambleton, R. K. (2014). *User's Manual for WinGen3: Windows Software that Generates IRT Model Parameters and Item Responses*. Retrieved from Amherst, MA:
- Hanson, B. A., CTB/McGraw-Hill, & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-Item equating design. *Applied Psychological Measurement, 26*(1), 3–24.
- Holland, P. W. (2007). A Framework and History for Score Linking. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and Aligning Scores and Scales* New York: Springer-Verlag New York. Retrieved from [https://link.springer.com/content/pdf/10.1007%2F978-0-387-49771-6\\_2.pdf](https://link.springer.com/content/pdf/10.1007%2F978-0-387-49771-6_2.pdf).
- Houts, C. R., & Cai, L. (2015). flexMIRT: flexible multilevel multidimensional item analysis and test scoring user's manual version 3.0RC. from Vector Psychometric Group, LLC <https://www.vpgcentral.com/wp-content/uploads/2015/11/flexMIRTUserManual3.0.pdf>
- Keller, L. A., & Keller, R. R. (2015). The effect of changing content on IRT scaling methods. *Applied Measurement in Education, 28*(2), 99-114.  
doi:10.1080/08957347.2014.1002922
- Kim, S.-H., & Cohen, A. S. (1997). *A comparison of linking and concurrent calibration under the graded response model*. Paper presented at the the Annual Meeting of the American Educational Research Association Chicago, IL.
- Kim, S., & Kolen, M. J. (2006). Robustness to format effects of IRT linking methods for mixed-format tests. *Applied measurement in education, 19*(4), 357-381.
- Kim, S., Walker, M. E., & McHale, F. (2010). Comparisons among designs for equating mixed-format tests in large-scale assessments. *Journal of Educational Measurement, 47*(1), 36-53.
- Kolen, M. J. (2007). Data Collection Designs and Linking Procedures. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales*. New York: Springer Science+Business Media.



- Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, Scaling, and Linking : Methods and Practices*. In S. E. Fienberg & W. J. v. d. Linden (Series Eds.), doi:10.1007/978-1-4939-0317-7
- Lee, E. (2013). *Equating multidimensional tests under a random groups design: a comparison of various equating procedures*. (Doctor of Philosophy), University of Iowa. Retrieved from <http://ir.uiowa.edu/etd/5011>
- Lee, W.-C., & Brossman, B. G. (2012). Observed Score Equating for Mixed-Format Tests Using a Simple-Structure Multidimensional IRT Framework. *Center for Advanced Studies in Measurement and Assessment (CASMA)*, 2(2.2), 115-142.
- Lin, P. (2008). *IRT vs. Factor analysis approaches in analyzing multigroup multidimensional binary data: the effect of structural orthogonality, and the equivalence in test structure, item difficulty, & examinee groups*. (Doctor of Philosophy), University of Maryland
- McKinley, R. L., & Reckase, M. D. (1981). *A comparison of procedures for constructing large item pools* (ED207998). Retrieved from Columbia: <https://eric.ed.gov/?id=ED207998>
- Meng, H. (2007). *A comparison study of IRT calibration methods for mixed-format tests in vertical scaling*. (The Doctor of Philosophy D), The University of Iowa. Retrieved from <https://ir.uiowa.edu/cgi/viewcontent.cgi?article=1523&context=etd>
- Patz, R. J., & Yao, L. (2007). Methods and models for vertical scaling. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and Aligning Scores and Scales* (pp. 253-272): Springer New York.
- Peterson, J. L. (2014). *Multidimensional item response theory observed score equating methods for mixed-format tests*. (Ph.D), University of Iowa.
- Reckase, M. D. (1985). The difficulty of test items that measure more than one ability. *Applied Psychological Measurement Inc*, 9(4), 401-412.
- Reckase, M. D. (2009). *Multidimensional item response theory* *Statistics for Social and Behavioral Sciences*, Retrieved from <http://dx.doi.org/10.1007/978-0-387-89976-3>

- Simon, M. K. (2008). *Comparison of concurrent and separate multidimensional IRT linking of item parameters*. (Doctor of philosophy ), The University of Minnesota Digital Conservancy.
- Suksuwana, S., Junpengb, P., Ngudgratokec, S., & Guayjarernpanishk, P. (2012). *The effect of proportion common item's with mixed format test on multidimensional item response theory linking*. Paper presented at the International Conference on Education and Educational Psychology (ICEEPSY 2012).
- Tian, F. (2011). *A comparison of equating/linking using the Stocking-Lord method and concurrent calibration mixed-format tests in the non-equivalent group common-item design under IRT*. (Doctor of Philosophy), Boston College University.
- Weeks, J. P. (2010). plink: an R package for linking mixed-format tests using IRT-based methods. *Journal of Statistical Software*, 35(12).
- Wei, Y. (2008). *A simulation study on the performance of four multidimensional IRT scale linking methods*. (Doctor of philosophy ), University of Florida.
- Xiong, J., Ye, X., Ding, S., & Luo, F. (2013). The research on anchor ratio of test equating design. In W. Zhang (Ed.), *Advanced Technology in Teaching* (Vol. 163, pp. 413-418): Springer Berlin Heidelberg.
- Xiong, J., Ye, X., Ding, S., & Luo, F. (2012). The Research on Anchor Ratio of Test Equating Design. . In I. Z. W. (Ed.), *Advanced Technology in Teaching* (Vol. 163, pp. 413-418). Berlin, Heidelberg: Springer.
- Yao, L. (2006). The BMIRT toolkit. from Defense Manpower Data Center [www.bmirt.com/media/f5abb5352d553d5fffff807cffffd524.pdf](http://www.bmirt.com/media/f5abb5352d553d5fffff807cffffd524.pdf)
- Yao, L. (2016). The BMIRT Toolkit: Defense Manpower Data Center Retrieved from [www.bmirt.com/media/f5abb5352d553d5fffff807cffffd524.pdf](http://www.bmirt.com/media/f5abb5352d553d5fffff807cffffd524.pdf)
- Yao, L., & Boughton, K. (2009). Multidimensional linking for tests with mixed item types. *Journal of Educational Measurement*, 46(2), 177-197.
- Zhang, J. (2007). Conditional covariance theory and detect for polytomous items. *Psychometrika*, 72(1), 69-91.

- Zhang, J., & Stout, W. (1999). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, 64(2), 129-152.
- Zhang, O. (2012). *Observed score and true score equating for multidimensional item response theory under nonequivalent group anchor test design*. (The Degree of Doctor of Philosophy), University of Florida.





ภาคผนวก

จุฬาลงกรณ์มหาวิทยาลัย  
**CHULALONGKORN UNIVERSITY**



ภาคผนวก ก

ค่าความยากรวม (MDIFF) ค่าอำนาจจำแนกรวม (MDISC)  
และค่ามุมระหว่างเวกเตอร์ในแต่ละเงื่อนไข

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

ตาราง 1 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 10% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	6.75	82.50	83.25	7.50	7.37	82.41	82.63	7.59	6.89	83.69	83.58	6.31
2	-1.5	0.4	7.73	81.84	82.27	8.16	8.62	82.52	81.38	7.48	6.94	83.11		
3	1.0	0.6	7.51	82.54	82.49	7.46	7.86	83.39	82.14	6.61	6.42	83.06		
4	-0.5	0.8	7.29	82.45	82.71	7.55	7.73	81.06	82.27	8.94				
5	0	1.0	9.14	83.23	80.86	6.77	8.06	83.87	81.94	6.13				
6	0	1.2	8.74	82.29	81.26	7.71	7.39	84.49	82.61	5.51				
7	-1.0	1.4	6.62	83.34	83.38	6.66	8.72	83.17	81.28	6.83				
8	0.5	1.6	7.85	82.87	82.15	7.13	8.07	83.47	81.93	6.53				
9	1.5	1.8	8.28	83.42	81.72	6.58	7.21	83.16	82.79	6.84				
10	2.0	2.0	8.34	82.52	81.66	7.48	8.42	83.35	81.58	6.65				
11	-2.0	0.2	7.07	83.56	82.93	6.44	6.55	84.21	83.45	5.79				
12	-1.5	0.4	8.59	83.20	81.41	6.80	6.95	83.10	83.05	6.90				
13	1.0	0.6	7.88	83.02	82.12	6.98	8.06	85.24	81.94	4.76				
14	-0.5	0.8	7.62	82.96	82.38	7.04	9.72	84.42	80.28	5.58				
15	0	1.0	6.40	81.48	83.60	8.52	6.10	80.98	83.90	9.02				
16	0	1.2	6.69	83.08	83.31	6.92	6.93	82.66	83.07	7.34				
17	-1.0	1.4	7.14	83.61	82.86	6.39	6.88	82.90	83.12	7.10				
18	.5	1.6	5.37	83.89	84.63	6.11	6.97	84.19	83.03	5.81				
19	1.5	1.8	5.82	84.77	84.18	5.23	8.25	83.28	81.75	6.72				
20	2.0	2.0	7.28	81.60	82.72	8.40	7.10	83.77	82.90	6.23				
21	-2.0	0.2	7.13	84.22	82.87	5.78	6.44	83.46	83.56	6.54				
22	-1.5	0.4	6.96	82.82	83.04	7.18	7.98	84.33	82.02	5.67				
23	1.0	0.6	6.40	83.09	83.60	6.91	7.74	81.60	82.26	8.40				
24	-5	0.8	6.85	83.39	83.15	6.61	6.15	82.51	83.85	7.49				
25	0	1.0	7.79	82.81	82.21	7.19	9.09	83.06	80.91	6.94				
26	0	1.2	6.52	83.05	83.48	6.95	9.15	82.69	80.85	7.31				
27	-1.0	1.4	8.01	82.42	81.99	7.58	7.14	81.96	82.86	8.04				
28	.5	1.6	6.45	83.12	83.55	6.88	8.17	82.45	81.83	7.55				
29	1.5	1.8	8.22	81.56	81.78	8.44	6.93	83.10	83.07	6.90				
30	2.0	2.0	6.17	83.24	83.83	6.76	7.78	82.07	82.22	7.93				
31	-2.0	0.2	8.28	81.51	81.72	8.49	7.52	83.19	82.48	6.81				
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	7.96	82.01	82.04	7.99	8.66	83.40	81.34	6.60	6.91	83.70	81.12	8.88
2	-1.5	0.4	6.59	84.10	83.41	5.90	6.23	82.38	83.77	7.62			83.09	6.30
3	1.0	0.6	7.24	82.15	82.76	7.85	6.50	82.25	83.50	7.75				
4	-0.5	0.8	8.77	84.23	81.23	5.77	7.58	82.27	82.42	7.73				
5	0	1.0	8.17	84.04	81.83	5.96	9.12	83.98	80.88	6.02				
6	0	1.2	6.81	82.75	83.19	7.25	6.95	83.77	83.05	6.23				
7	-1.0	1.4	7.28	82.70	82.72	7.30	9.55	83.84	80.45	6.16				

ตาราง 2 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 10% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 60:40														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	8.53	83.59	81.47	6.41	6.69	83.51	83.31	6.49	7.47	83.48	82.53	6.52
2	-1.5	0.4	6.90	83.77	83.10	6.23	7.86	83.16	82.14	6.84	7.67	81.87	82.33	8.13
3	1.0	0.6	6.97	82.56	83.03	7.44	5.97	81.40	84.03	8.60	6.43	81.78	83.57	8.22
4	-0.5	0.8	6.78	83.10	83.22	6.90	7.20	81.24	82.80	8.76				
5	0	1.0	7.27	83.44	82.73	6.56	7.96	82.16	82.04	7.84				
6	0	1.2	7.60	80.65	82.40	9.35	7.18	82.73	82.82	7.27				
7	-1.0	1.4	8.29	82.92	81.71	7.08	6.62	83.23	83.38	6.77				
8	0.5	1.6	7.34	83.13	82.66	6.87	7.71	82.60	82.29	7.40				
9	1.5	1.8	7.66	83.34	82.34	6.66	8.45	82.82	81.55	7.18				
10	2.0	2.0	7.98	82.05	82.02	7.95	7.08	81.37	82.92	8.63				
11	-2.0	0.2	6.42	82.75	83.58	7.25	8.62	83.97	81.38	6.03				
12	-1.5	0.4	5.75	84.28	84.25	5.72	8.32	84.18	81.68	5.82				
13	1.0	0.6	5.82	82.05	84.18	7.95	7.34	81.92	82.66	8.08				
14	-0.5	0.8	8.13	83.70	81.87	6.30	7.06	82.87	82.94	7.13				
15	0	1.0	8.70	83.42	81.30	6.58	7.04	82.82	82.96	7.18				
16	0	1.2	8.21	81.97	81.79	8.03	7.18	83.71	82.82	6.29				
17	-1.0	1.4	8.05	82.99	81.95	7.01	8.52	81.64	81.48	8.36				
18	0.5	1.6	8.83	82.86	81.17	7.14	7.18	84.21	82.82	5.79				
19	1.5	1.8	9.16	82.81	80.84	7.19	8.65	84.27	81.35	5.73				
20	2.0	2.0	7.33	83.99	82.67	6.01	5.86	82.35	84.14	7.65				
21	-2.0	0.2	8.26	82.63	81.74	7.37	6.79	82.58	83.21	7.42				
22	-1.5	0.4	7.83	81.72	82.17	8.28	8.28	83.42	81.72	6.58				
23	1.0	0.6	6.51	82.89	83.49	7.11	9.63	82.21	80.37	7.79				
24	-0.5	0.8	7.39	82.02	82.61	7.98	7.85	83.19	82.15	6.81				
25	0	1.0	4.86	83.45	85.14	6.55	8.16	82.62	81.84	7.38				
26	0	1.2	9.34	82.16	80.66	7.84	8.06	81.92	81.94	8.08				
27	-1.0	1.4	7.53	82.54	82.47	7.46	6.47	82.36	83.53	7.64				
28	0.5	1.6	8.32	82.58	81.68	7.42	7.35	83.83	82.65	6.17				
29	1.5	1.8	7.51	84.05	82.49	5.95	8.88	81.86	81.12	8.14				
30	2.0	2.0	8.35	82.88	81.65	7.12	7.31	83.16	82.69	6.84				
31	-2.0	0.2	7.19	83.93	82.81	6.07	7.40	82.21	82.60	7.79				
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	6.48	84.40	83.52	5.60	9.13	83.85	80.87	6.15	8.27	82.78	81.73	7.22
2	-1.5	0.4	7.37	82.64	82.63	7.36	7.57	83.51	82.43	6.49				
3	1.0	0.6	7.21	82.63	82.79	7.37	7.39	82.47	82.61	7.53				
4	-0.5	0.8	7.13	81.44	82.87	8.56	7.11	82.87	82.89	7.13				
5	0	1.0	7.31	82.72	82.69	7.28	7.51	83.83	82.49	6.17				
6	0	1.2	7.66	81.76	82.34	8.24	7.61	83.42	82.39	6.58				
7	-1.0	1.4	6.20	82.36	83.80	7.64	6.39	82.43	83.61	7.57				

ตาราง 3 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 10% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	7.60	82.52	82.40	7.48	8.57	83.08	81.43	6.92	8.94	85.54	83.07	6.31
2	-1.5	0.4	7.67	81.05	82.33	8.95	9.12	82.68	80.88	7.32	9.82	84.39	84.18	
3	1.0	0.6	7.56	84.73	82.44	5.27	6.25	83.72	83.75	6.28	9.15	83.49	81.06	
4	-0.5	0.8	7.39	84.60	82.61	5.40	6.59	83.82	83.41	6.18			80.18	
5	0	1.0	10.68	83.08	79.32	6.92	7.49	83.20	82.51	6.80			80.85	
6	0	1.2	7.21	83.65	82.79	6.35	7.20	81.03	82.80	8.97				
7	-1.0	1.4	7.13	84.46	82.87	5.54	8.39	83.29	81.61	6.71				
8	0.5	1.6	8.67	81.91	81.33	8.09	7.19	82.63	82.81	7.37				
9	1.5	1.8	6.33	82.07	83.67	7.93	6.44	83.86	83.56	6.14				
10	2.0	2.0	9.06	81.80	80.94	8.20	6.92	82.84	83.08	7.16				
11	-2.0	0.2	6.76	82.10	83.24	7.90	8.82	83.76	81.18	6.24				
12	-1.5	0.4	7.75	82.11	82.25	7.89	6.39	82.10	83.61	7.90				
13	1.0	0.6	9.93	82.27	80.07	7.73	7.49	83.27	82.51	6.73				
14	-0.5	0.8	7.11	83.49	82.89	6.51	7.69	82.05	82.31	7.95				
15	0	1.0	8.15	83.59	81.85	6.41	9.11	83.15	80.89	6.85				
16	0	1.2	6.86	83.71	83.14	6.29	6.10	82.24	83.90	7.76				
17	-1.0	1.4	5.50	83.02	84.50	6.98	6.33	84.62	83.67	5.38				
18	0.5	1.6	6.75	82.28	83.25	7.72	7.25	82.28	82.75	7.72				
19	1.5	1.8	8.35	83.98	81.65	6.02	7.09	83.04	82.91	6.96				
20	2.0	2.0	7.87	82.35	82.13	7.65	9.11	83.98	80.89	6.02				
21	-2.0	0.2	7.19	82.14	82.81	7.86	6.79	81.36	83.21	8.64				
22	-1.5	0.4	7.80	83.68	82.20	6.32	7.23	83.31	82.77	6.69				
23	1.0	0.6	6.48	84.43	83.52	5.57	8.48	83.05	81.52	6.95				
24	-0.5	0.8	8.94	82.50	81.06	7.50	7.80	83.61	82.20	6.39				
25	0	1.0	5.75	83.91	84.25	6.09	9.83	83.12	80.17	6.88				
26	0	1.2	7.68	82.74	82.32	7.26	8.13	84.00	81.87	6.00				
27	-1.0	1.4	8.28	82.92	81.72	7.08	5.89	83.27	84.11	6.73				
28	0.5	1.6	6.18	83.33	83.82	6.67	8.52	82.87	81.48	7.13				
29	1.5	1.8	6.73	82.61	83.27	7.39	7.68	82.30	82.32	7.70				
30	2.0	2.0	8.73	82.77	81.27	7.23	10.12	83.35	79.88	6.65				
31	-2.0	0.2	7.24	83.16	82.76	6.84	5.81	83.16	84.19	6.84				
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	6.96	82.82	83.04	7.18	7.05	83.40	82.95	6.60	7.94	82.06	-	-
2	-1.5	0.4	10.46	83.38	79.54	6.62	6.53	82.39	83.47	7.61				
3	1.0	0.6	6.66	82.32	83.34	7.68	7.49	82.94	82.51	7.06				
4	-0.5	0.8	9.71	83.76	80.29	6.24	5.93	83.70	84.07	6.30				
5	0	1.0	6.74	82.19	83.26	7.81	6.41	84.57	83.59	5.43				
6	0	1.2	7.54	82.55	82.46	7.45	5.60	84.02	84.40	5.98				
7	-1.0	1.4	8.31	83.00	81.69	7.00	4.76	82.79	85.24	7.21				



ตาราง 4 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 20% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	7.23	81.19	82.77	8.81	8.03	84.38	81.97	5.62	5.22	83.48	84.78	6.52
2	-1.5	0.4	6.30	82.41	83.70	7.59	7.77	84.80	82.23	5.20	5.55	83.49	84.45	6.51
3	1.0	0.6	8.11	84.16	81.89	5.84	7.99	84.56	82.01	5.44	7.11	81.91	82.89	8.09
4	-0.5	0.8	8.00	82.49	82.00	7.51	7.34	82.57	82.66	7.43	7.71	83.44	82.29	6.56
5	0	1.0	7.65	82.87	82.35	7.13	8.21	82.38	81.79	7.62				
6	0	1.2	7.35	83.56	82.65	6.44	7.78	84.39	82.22	5.61				
7	-1.0	1.4	8.24	82.13	81.76	7.87	9.30	84.70	80.70	5.30				
8	0.5	1.6	7.74	83.37	82.26	6.63	7.01	82.02	82.99	7.98				
9	1.5	1.8	9.28	82.50	80.72	7.50	7.84	83.10	82.16	6.90				
10	2.0	2.0	7.27	84.00	82.73	6.00	8.32	80.73	81.68	9.27				
11	-2.0	0.2	8.74	81.90	81.26	8.10	7.70	82.22	82.30	7.78				
12	-1.5	0.4	6.00	84.31	84.00	5.69	6.24	82.91	83.76	7.09				
13	1.0	0.6	7.10	81.29	82.90	8.71	7.87	84.70	82.13	5.30				
14	-0.5	0.8	9.81	83.48	80.19	6.52	7.00	82.14	83.00	7.86				
15	0	1.0	7.69	83.69	82.31	6.31	8.06	83.66	81.94	6.34				
16	0	1.2	7.41	81.38	82.59	8.62	6.73	82.69	83.27	7.31				
17	-1.0	1.4	6.57	82.79	83.43	7.21	9.35	82.72	80.65	7.28				
18	0.5	1.6	8.58	83.41	81.42	6.59	8.06	82.83	81.94	7.17				
19	1.5	1.8	7.07	82.08	82.93	7.92	7.75	81.63	82.25	8.37				
20	2.0	2.0	7.51	82.83	82.49	7.17	6.99	83.22	83.01	6.78				
21	-2.0	0.2	6.62	82.42	83.38	7.58	7.14	82.58	82.86	7.42				
22	-1.5	0.4	6.37	82.79	83.63	7.21	7.75	81.84	82.25	8.16				
23	1.0	0.6	6.39	82.98	83.61	7.02	7.33	81.55	82.67	8.45				
24	-0.5	0.8	7.24	82.26	82.76	7.74	7.87	82.75	82.13	7.25				
25	0	1.0	8.25	84.23	81.75	5.77	6.36	83.23	83.64	6.77				
26	0	1.2	7.96	82.34	82.04	7.66	7.70	81.64	82.30	8.36				
27	-1.0	1.4	8.46	82.60	81.54	7.40	9.77	84.33	80.23	5.67				
28	0.5	1.6	8.18	81.78	81.82	8.22	7.45	82.71	82.55	7.29				
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	7.11	83.82	82.89	6.18	7.51	83.49	82.49	6.51	8.27	82.80	81.73	7.20
2	-1.5	0.4	7.83	83.43	82.17	6.57	9.03	83.43	80.97	6.57	6.01	84.14	83.99	5.86
3	1.0	0.6	7.07	84.25	82.93	5.75	7.71	82.81	82.29	7.19	8.39	83.95	81.61	6.05
4	-0.5	0.8	7.10	83.55	82.90	6.45	7.38	83.25	82.62	6.75				
5	0	1.0	6.82	82.85	83.18	7.15	7.51	84.09	82.49	5.91				
6	0	1.2	8.01	83.19	81.99	6.81	7.68	83.46	82.32	6.54				

ตาราง 5 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 20% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 50:50														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	6.51	83.28	83.49	6.72	7.86	82.58	82.14	7.42	6.51	83.28	83.49	6.72
2	-1.5	0.4	9.98	82.22	80.02	7.78	8.87	82.31	81.13	7.69	9.98	82.22	80.02	7.78
3	1.0	0.6	6.69	82.24	83.31	7.76	8.16	83.66	81.84	6.34	6.69	82.24	83.31	7.76
4	-0.5	0.8	6.64	83.60	83.36	6.40	8.31	83.13	81.69	6.87	6.64	83.60	83.36	6.40
5	0	1.0	8.55	82.53	81.45	7.47	9.37	83.48	80.63	6.52	8.55	82.53	81.45	7.47
6	0	1.2	6.07	84.69	83.93	5.31	8.01	81.98	81.99	8.02	6.07	84.69	83.93	5.31
7	-1.0	1.4	7.71	82.57	82.29	7.43	6.38	83.33	83.62	6.67	7.71	82.57	82.29	7.43
8	0.5	1.6	8.43	83.84	81.57	6.16	6.84	82.12	83.16	7.88	8.43	83.84	81.57	6.16
9	1.5	1.8	7.06	81.95	82.94	8.05	9.75	81.49	80.25	8.51	7.06	81.95	82.94	8.05
10	2.0	2.0	4.71	83.30	85.29	6.70	7.30	82.57	82.70	7.43	4.71	83.30	85.29	6.70
11	-2.0	0.2	8.31	83.28	81.69	6.72	8.30	82.43	81.70	7.57	8.31	83.28	81.69	6.72
12	-1.5	0.4	7.15	82.89	82.85	7.11	8.96	82.58	81.04	7.42	7.15	82.89	82.85	7.11
13	1.0	0.6	7.17	81.93	82.83	8.07	8.63	83.57	81.37	6.43	7.17	81.93	82.83	8.07
14	-0.5	0.8	8.09	84.93	81.91	5.07	9.04	83.47	80.96	6.53	8.09	84.93	81.91	5.07
15	0	1.0	8.75	81.76	81.25	8.24	8.30	82.26	81.70	7.74	8.75	81.76	81.25	8.24
16	0	1.2	7.98	83.75	82.02	6.25	8.36	81.81	81.64	8.19	7.98	83.75	82.02	6.25
17	-1.0	1.4	6.82	84.05	83.18	5.95	6.82	83.80	83.18	6.20	6.82	84.05	83.18	5.95
18	0.5	1.6	6.96	83.47	83.04	6.53	5.83	84.78	84.17	5.22	6.96	83.47	83.04	6.53
19	1.5	1.8	7.47	81.74	82.53	8.26	6.48	84.39	83.52	5.61	7.47	81.74	82.53	8.26
20	2.0	2.0	6.89	83.08	83.11	6.92	8.01	83.63	81.99	6.37	6.89	83.08	83.11	6.92
21	-2.0	0.2	7.04	83.08	82.96	6.92	9.47	83.90	80.53	6.10	7.04	83.08	82.96	6.92
22	-1.5	0.4	7.79	83.28	82.21	6.72	6.74	80.69	83.26	9.31	7.79	83.28	82.21	6.72
23	1.0	0.6	7.70	83.52	82.30	6.48	5.99	83.56	84.01	6.44	7.70	83.52	82.30	6.48
24	-0.5	0.8	6.31	83.75	83.69	6.25	7.52	83.23	82.48	6.77	6.31	83.75	83.69	6.25
25	0	1.0	8.91	82.27	81.09	7.73	5.87	83.12	84.13	6.88	8.91	82.27	81.09	7.73
26	0	1.2	5.80	82.74	84.20	7.26	10.35	83.00	79.65	7.00	5.80	82.74	84.20	7.26
27	-1.0	1.4	6.04	84.07	83.96	5.93	7.58	81.88	82.42	8.12	6.04	84.07	83.96	5.93
28	0.5	1.6	7.55	82.36	82.45	7.64	6.81	81.96	83.19	8.04	7.55	82.36	82.45	7.64
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	6.15	84.84	83.85	5.16	7.65	82.53	82.35	7.47	7.14	82.75	84.48	7.25
2	-1.5	0.4	8.51	83.55	81.49	6.45	7.68	82.37	82.32	7.63	5.52	83.23	80.12	6.77
3	1.0	0.6	7.73	82.40	82.27	7.60	7.98	83.42	82.02	6.58	9.88	82.86		
4	-0.5	0.8	5.89	81.98	84.11	8.02	7.79	82.85	82.21	7.15				
5	0	1.0	6.89	82.54	83.11	7.46	8.51	83.80	81.49	6.20				
6	0	1.2	6.41	83.42	83.59	6.58	8.56	81.13	81.44	8.87				

ตาราง 6 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 20% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 60:40														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	5.11	82.79	84.89	7.21	7.18	83.86	82.82	6.14	5.11	82.79	84.89	7.21
2	-1.5	0.4	6.84	83.00	83.16	7.00	7.77	84.11	82.23	5.89	6.84	83.00	83.16	7.00
3	1.0	0.6	7.91	84.51	82.09	5.49	8.32	82.35	81.68	7.65	7.91	84.51	82.09	5.49
4	-0.5	0.8	7.33	82.45	82.67	7.55	7.05	83.34	82.95	6.66	7.33	82.45	82.67	7.55
5	0	1.0	6.95	83.09	83.05	6.91	6.28	82.49	83.72	7.51	6.95	83.09	83.05	6.91
6	0	1.2	7.73	83.50	82.27	6.50	7.99	83.50	82.01	6.50	7.73	83.50	82.27	6.50
7	-1.0	1.4	6.97	82.42	83.03	7.58	7.27	83.82	82.73	6.18	6.97	82.42	83.03	7.58
8	0.5	1.6	7.48	83.20	82.52	6.80	7.98	82.71	82.02	7.29	7.48	83.20	82.52	6.80
9	1.5	1.8	7.93	82.80	82.07	7.20	8.14	82.07	81.86	7.93	7.93	82.80	82.07	7.20
10	2.0	2.0	8.36	83.31	81.64	6.69	6.54	83.82	83.46	6.18	8.36	83.31	81.64	6.69
11	-2.0	0.2	6.71	82.12	83.29	7.88	7.32	81.18	82.68	8.82	6.71	82.12	83.29	7.88
12	-1.5	0.4	7.78	82.74	82.22	7.26	9.13	83.61	80.87	6.39	7.78	82.74	82.22	7.26
13	1.0	0.6	7.29	84.54	82.71	5.46	8.12	82.27	81.88	7.73	7.29	84.54	82.71	5.46
14	-0.5	0.8	7.93	82.47	82.07	7.53	7.63	82.81	82.37	7.19	7.93	82.47	82.07	7.53
15	0	1.0	6.96	83.75	83.04	6.25	7.65	83.77	82.35	6.23	6.96	83.75	83.04	6.25
16	0	1.2	9.36	83.93	80.64	6.07	6.72	82.31	83.28	7.69	9.36	83.93	80.64	6.07
17	-1.0	1.4	8.22	84.05	81.78	5.95	6.04	82.26	83.96	7.74	8.22	84.05	81.78	5.95
18	0.5	1.6	7.74	82.75	82.26	7.25	6.77	83.38	83.23	6.62	7.74	82.75	82.26	7.25
19	1.5	1.8	7.89	83.49	82.11	6.51	7.99	82.83	82.01	7.17	7.89	83.49	82.11	6.51
20	2.0	2.0	8.57	84.23	81.43	5.77	8.97	84.70	81.03	5.30	8.57	84.23	81.43	5.77
21	-2.0	0.2	7.62	84.22	82.38	5.78	9.22	82.08	80.78	7.92	7.62	84.22	82.38	5.78
22	-1.5	0.4	7.88	83.13	82.12	6.87	6.97	84.31	83.03	5.69	7.88	83.13	82.12	6.87
23	1.0	0.6	8.85	81.77	81.15	8.23	6.86	83.43	83.14	6.57	8.85	81.77	81.15	8.23
24	-0.5	0.8	7.63	82.36	82.37	7.64	8.50	82.25	81.50	7.75	7.63	82.36	82.37	7.64
25	0	1.0	6.92	83.73	83.08	6.27	7.96	84.37	82.04	5.63	6.92	83.73	83.08	6.27
26	0	1.2	6.08	82.01	83.92	7.99	6.35	85.39	83.65	4.61	6.08	82.01	83.92	7.99
27	-1.0	1.4	8.23	82.44	81.77	7.56	6.87	80.63	83.13	9.37	8.23	82.44	81.77	7.56
28	0.5	1.6	7.94	80.98	82.06	9.02	6.76	84.72	83.24	5.28	7.94	80.98	82.06	9.02
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	5.75	82.77	84.25	7.23	7.37	81.96	82.63	8.04	5.75	82.77	84.25	7.23
2	-1.5	0.4	5.46	81.65	84.54	8.35	7.40	84.21	82.60	5.79	5.46	81.65	84.54	8.35
3	1.0	0.6	6.56	84.18	83.44	5.82	7.22	85.02	82.78	4.98	6.56	84.18	83.44	5.82
4	-0.5	0.8	7.92	83.61	82.08	6.39	6.27	84.75	83.73	5.25	7.92	83.61	82.08	6.39
5	0	1.0	8.50	81.13	81.50	8.87	7.02	82.98	82.98	7.02	8.50	81.13	81.50	8.87
6	0	1.2	7.20	82.66	82.80	7.34	7.62	82.65	82.38	7.20	82.66	82.80	7.34	7.62

ตาราง 7 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 20% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	8.20	83.72	81.80	6.28	9.13	83.16	80.87	6.84	6.29	83.16	83.06	6.84
2	-1.5	0.4	6.64	83.82	83.36	6.18	7.39	82.40	82.61	7.60	6.99	82.93	84.50	7.07
3	1.0	0.6	8.36	83.70	81.64	6.30	7.19	82.96	82.81	7.04	6.94	82.91	83.38	7.09
4	-0.5	0.8	8.03	83.53	81.97	6.47	6.63	84.35	83.37	5.65	5.50	83.39	84.15	6.61
5	0	1.0	7.42	82.47	82.58	7.53	7.44	83.00	82.56	7.00	6.62	81.92	82.76	8.08
6	0	1.2	6.38	83.32	83.62	6.68	5.75	83.46	84.25	6.54	5.85	82.96	83.57	7.04
7	-1.0	1.4	7.07	81.91	82.93	8.09	7.40	82.56	82.60	7.44	7.24	83.71		
8	0.5	1.6	8.19	82.89	81.81	7.11	6.70	82.14	83.30	7.86	6.43	83.01		
9	1.5	1.8	6.19	82.52	83.81	7.48	8.64	83.66	81.36	6.34				
10	2.0	2.0	7.38	83.98	82.62	6.02	6.38	82.96	83.62	7.04				
11	-2.0	0.2	7.76	83.68	82.24	6.32	7.26	82.31	82.74	7.69				
12	-1.5	0.4	7.82	82.99	82.18	7.01	5.96	82.47	84.04	7.53				
13	1.0	0.6	8.53	82.34	81.47	7.66	7.54	84.60	82.46	5.40				
14	-0.5	0.8	6.44	82.38	83.56	7.62	8.31	82.98	81.69	7.02				
15	0	1.0	5.43	82.56	84.57	7.44	6.92	82.50	83.08	7.50				
16	0	1.2	7.21	83.16	82.79	6.84	5.31	83.15	84.69	6.85				
17	-1.0	1.4	8.56	83.72	81.44	6.28	7.76	83.64	82.24	6.36				
18	0.5	1.6	7.37	83.72	82.63	6.28	6.70	83.91	83.30	6.09				
19	1.5	1.8	6.73	84.02	83.27	5.98	8.50	82.48	81.50	7.52				
20	2.0	2.0	6.54	82.71	83.46	7.29	7.35	82.80	82.65	7.20				
21	-2.0	0.2	7.04	81.78	82.96	8.22	8.72	83.14	81.28	6.86				
22	-1.5	0.4	7.46	82.87	82.54	7.13	7.77	83.61	82.23	6.39				
23	1.0	0.6	8.00	83.03	82.00	6.97	7.09	82.77	82.91	7.23				
24	-0.5	0.8	9.19	83.36	80.81	6.64	7.53	82.59	82.47	7.41				
25	0	1.0	6.10	83.63	83.90	6.37	8.41	83.23	81.59	6.77				
26	0	1.2	7.65	82.68	82.35	7.32	7.03	84.04	82.97	5.96				
27	-1.0	1.4	9.42	81.63	80.58	8.37	7.27	82.51	82.73	7.49				
28	0.5	1.6	7.10	83.30	82.90	6.70	7.96	83.10	82.04	6.90				
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	7.50	82.12	82.50	7.88	7.55	82.93	82.45	7.07	6.30	82.62	83.51	7.38
2	-1.5	0.4	6.51	84.15	83.49	5.85	7.30	82.31	82.70	7.69			83.70	6.49
3	1.0	0.6	7.62	81.75	82.38	8.25	7.42	83.31	82.58	6.69				
4	-0.5	0.8	8.31	82.09	81.69	7.91	7.14	80.30	82.86	9.70				
5	0	1.0	7.10	81.74	82.90	8.26	7.05	82.35	82.95	7.65				
6	0	1.2	7.96	84.09	82.04	5.91	6.64	81.53	83.36	8.47				

ตาราง 8 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 30% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	5.91	83.70	84.09	6.30	7.63	83.97	82.37	6.03	5.91	83.70	84.09	6.30
2	-1.5	0.4	7.80	82.01	82.20	7.99	7.46	84.64	82.54	5.36	7.80	82.01	82.20	7.99
3	1.0	0.6	6.41	82.80	83.59	7.20	6.87	81.41	83.13	8.59	6.41	82.80	83.59	7.20
4	-0.5	0.8	8.58	82.53	81.42	7.47	7.78	85.11	82.22	4.89	8.58	82.53	81.42	7.47
5	0	1.0	8.65	82.21	81.35	7.79	8.14	83.50	81.86	6.50	8.65	82.21	81.35	7.79
6	0	1.2	8.01	81.87	81.99	8.13	7.79	83.88	82.21	6.12	8.01	81.87	81.99	8.13
7	-1.0	1.4	6.44	82.92	83.56	7.08	7.53	82.13	82.47	7.87	6.44	82.92	83.56	7.08
8	0.5	1.6	5.61	82.91	84.39	7.09	7.55	82.92	82.45	7.08	5.61	82.91	84.39	7.09
9	1.5	1.8	7.77	82.05	82.23	7.95	8.48	84.33	81.52	5.67	7.77	82.05	82.23	7.95
10	2.0	2.0	7.54	83.27	82.46	6.73	8.75	82.33	81.25	7.67	7.54	83.27	82.46	6.73
11	-2.0	0.2	6.49	83.47	83.51	6.53	6.03	81.85	83.97	8.15	6.49	83.47	83.51	6.53
12	-1.5	0.4	7.87	82.61	82.13	7.39	8.68	82.36	81.32	7.64	7.87	82.61	82.13	7.39
13	1.0	0.6	8.06	84.64	81.94	5.36	8.46	83.29	81.54	6.71	8.06	84.64	81.94	5.36
14	-0.5	0.8	6.95	83.59	83.05	6.41	6.71	82.84	83.29	7.16	6.95	83.59	83.05	6.41
15	0	1.0	7.90	83.38	82.10	6.62	8.33	83.13	81.67	6.87	7.90	83.38	82.10	6.62
16	0	1.2	8.05	82.62	81.95	7.38	7.19	83.78	82.81	6.22	8.05	82.62	81.95	7.38
17	-1.0	1.4	7.67	83.55	82.33	6.45	7.81	83.64	82.19	6.36	7.67	83.55	82.33	6.45
18	0.5	1.6	7.18	82.92	82.82	7.08	7.97	83.91	82.03	6.09	7.18	82.92	82.82	7.08
19	1.5	1.8	6.46	84.58	83.54	5.42	6.89	82.91	83.11	7.09	6.46	84.58	83.54	5.42
20	2.0	2.0	7.30	83.56	82.70	6.44	6.33	83.23	83.67	6.77	7.30	83.56	82.70	6.44
21	-2.0	0.2	6.82	82.95	83.18	7.05	6.80	83.09	83.20	6.91	6.82	82.95	83.18	7.05
22	-1.5	0.4	7.93	84.00	82.07	6.00	8.68	82.45	81.32	7.55	7.93	84.00	82.07	6.00
23	1.0	0.6	6.54	84.03	83.46	5.97	7.15	82.65	82.85	7.35	6.54	84.03	83.46	5.97
24	-0.5	0.8	7.06	83.55	82.94	6.45	6.68	84.24	83.32	5.76	7.06	83.55	82.94	6.45
25	0	1.0	8.19	81.55	81.81	8.45	7.49	81.69	82.51	8.31	8.19	81.55	81.81	8.45
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	9.17	82.21	80.83	7.79	7.65	82.84	82.35	7.16	9.17	82.21	80.83	7.79
2	-1.5	0.4	6.31	83.12	83.69	6.88	7.94	83.54	82.06	6.46	6.31	83.12	83.69	6.88
3	1.0	0.6	8.80	81.97	81.20	8.03	7.41	84.16	82.59	5.84	8.80	81.97	81.20	8.03
4	-0.5	0.8	7.14	84.29	82.86	5.71	7.57	82.49	82.43	7.51	7.14	84.29	82.86	5.71
5	0	1.0	6.97	83.52	83.03	6.48	7.64	83.52	82.36	6.48	6.97	83.52	83.03	6.48

ตาราง 9 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 30% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 50:50														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	8.24	82.74	81.76	7.26	7.51	84.29	82.49	5.71	5.92	82.42	84.08	7.58
2	-1.5	0.4	7.51	83.83	82.49	6.17	7.05	81.74	82.95	8.26	7.00	83.54	83.00	6.46
3	1.0	0.6	7.94	82.47	82.06	7.53	6.83	84.11	83.17	5.89	6.88	82.52	83.12	7.48
4	-0.5	0.8	8.23	82.60	81.77	7.40	8.37	83.60	81.63	6.40	8.18	82.61	81.82	7.39
5	0	1.0	7.32	81.48	82.68	8.52	8.04	82.63	81.96	7.37	6.06	83.00	83.94	7.00
6	0	1.2	7.71	84.33	82.29	5.67	6.49	84.22	83.51	5.78	8.58	81.74	81.42	8.26
7	-1.0	1.4	8.96	83.98	81.04	6.02	8.24	83.37	81.76	6.63	7.28	82.83	82.72	7.17
8	0.5	1.6	7.22	83.64	82.78	6.36	6.60	82.10	83.40	7.90				
9	1.5	1.8	8.67	82.43	81.33	7.57	8.14	82.17	81.86	7.83				
10	2.0	2.0	6.25	83.40	83.75	6.60	6.61	83.22	83.39	6.78				
11	-2.0	0.2	6.55	83.13	83.45	6.87	8.66	85.44	81.34	4.56				
12	-1.5	0.4	6.65	82.11	83.35	7.89	8.27	82.37	81.73	7.63				
13	1.0	0.6	6.55	81.55	83.45	8.45	7.73	81.72	82.27	8.28				
14	-0.5	0.8	9.06	82.78	80.94	7.22	7.41	84.16	82.59	5.84				
15	0	1.0	7.50	83.62	82.50	6.38	6.82	82.84	83.18	7.16				
16	0	1.2	7.73	82.69	82.27	7.31	5.34	83.32	84.66	6.68				
17	-1.0	1.4	7.81	83.93	82.19	6.07	7.76	83.63	82.24	6.37				
18	0.5	1.6	7.43	83.48	82.57	6.52	6.21	83.01	83.79	6.99				
19	1.5	1.8	7.62	82.46	82.38	7.54	8.74	82.97	81.26	7.03				
20	2.0	2.0	7.05	83.76	82.95	6.24	8.03	83.62	81.97	6.38				
21	-2.0	0.2	7.28	83.53	82.72	6.47	4.56	83.58	85.44	6.42				
22	-1.5	0.4	8.64	83.70	81.36	6.30	8.16	83.14	81.84	6.86				
23	1.0	0.6	7.96	83.59	82.04	6.41	7.48	83.77	82.52	6.23				
24	-0.5	0.8	5.92	82.21	84.08	7.79	7.24	83.32	82.76	6.68				
25	0	1.0	7.13	82.41	82.87	7.59	8.67	84.22	81.33	5.78				
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	5.87	82.45	84.13	7.55	7.39	82.15	82.61	7.85	9.04	82.59	80.96	7.41
2	-1.5	0.4	6.95	84.21	83.05	5.79	8.23	83.25	81.77	6.75	6.63	85.10	83.37	4.90
3	1.0	0.6	7.05	82.00	82.95	8.00	9.99	84.55	80.01	5.45	9.12	83.38	80.88	6.62
4	-0.5	0.8	7.21	82.99	82.79	7.01	7.18	84.23	82.82	5.77	6.85	83.47	83.15	6.53
5	0	1.0	8.74	82.83	81.26	7.17	5.83	82.18	84.17	7.82				

ตาราง 10 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 30% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 60:40														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	8.19	83.40	81.81	6.60	7.92	83.44	82.08	6.56	7.71	83.03	82.29	6.97
2	-1.5	0.4	6.89	83.73	83.11	6.27	6.30	83.45	83.70	6.55	3.71	84.32	86.29	5.68
3	1.0	0.6	7.21	80.88	82.79	9.12	7.07	82.75	82.93	7.25	7.60	82.70	82.40	7.30
4	-0.5	0.8	8.10	81.49	81.90	8.51	8.33	81.82	81.67	8.18	7.85	83.47	82.15	6.53
5	0	1.0	8.43	84.78	81.57	5.22	6.79	82.91	83.21	7.09	8.45	82.56	81.55	7.44
6	0	1.2	7.73	82.57	82.27	7.43	8.81	82.88	81.19	7.12	6.88	82.34	83.12	7.66
7	-1.0	1.4	6.65	83.05	83.35	6.95	7.94	83.83	82.06	6.17	6.65	80.79	83.35	9.21
8	0.5	1.6	7.65	83.05	82.35	6.95	8.05	82.93	81.95	7.07	8.39	82.61	81.61	7.39
9	1.5	1.8	7.10	82.59	82.90	7.41	7.40	82.79	82.60	7.21	7.22	84.76	82.78	5.24
10	2.0	2.0	7.33	83.92	82.67	6.08	8.57	81.85	81.43	8.15				
11	-2.0	0.2	8.34	80.91	81.66	9.09	8.45	82.54	81.55	7.46				
12	-1.5	0.4	6.93	81.70	83.07	8.30	7.70	83.13	82.30	6.87				
13	1.0	0.6	7.44	84.04	82.56	5.96	7.42	84.09	82.58	5.91				
14	-0.5	0.8	8.17	84.17	81.83	5.83	6.38	84.84	83.62	5.16				
15	0	1.0	6.73	82.25	83.27	7.75	8.24	83.14	81.76	6.86				
16	0	1.2	6.79	81.73	83.21	8.27	6.68	83.90	83.32	6.10				
17	-1.0	1.4	6.63	84.02	83.37	5.98	7.98	84.57	82.02	5.43				
18	0.5	1.6	8.04	84.81	81.96	5.19	7.83	81.04	82.17	8.96				
19	1.5	1.8	8.82	84.91	81.18	5.09	6.53	82.04	83.47	7.96				
20	2.0	2.0	7.49	80.82	82.51	9.18	8.47	83.37	81.53	6.63				
21	-2.0	0.2	8.02	82.97	81.98	7.03	8.37	82.53	81.63	7.47				
22	-1.5	0.4	7.78	83.91	82.22	6.09	7.09	83.01	82.91	6.99				
23	1.0	0.6	9.30	83.25	80.70	6.75	8.16	83.77	81.84	6.23				
24	-0.5	0.8	7.53	82.95	82.47	7.05	7.04	82.69	82.96	7.31				
25	0	1.0	8.87	82.01	81.13	7.99	6.01	83.79	83.99	6.21				
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	7.68	83.87	82.32	6.13	7.06	83.07	82.94	6.93	7.31	84.16	82.69	5.84
2	-1.5	0.4	7.79	82.02	82.21	7.98	8.61	81.63	81.39	8.37	6.21	82.26	83.79	7.74
3	1.0	0.6	6.72	82.59	83.28	7.41	7.09	81.88	82.91	8.12	5.96	83.81	84.04	6.19
4	-0.5	0.8	6.90	82.13	83.10	7.87	5.87	84.16	84.13	5.84				
5	0	1.0	5.63	81.97	84.37	8.03	7.87	83.56	82.13	6.44				

ตาราง 11 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง APSS

สัดส่วนคะแนนข้อสอบรวม 30% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30														
ข้อสอบที่ตรวจให้คะแนนสองค่า														
ลำดับ	MDIFF	MDISC	ค่ามุมข้อสอบเฉพาะฉบับ X				ค่ามุมข้อสอบเฉพาะฉบับ Y				ข้อสอบรวม			
			กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2		กลุ่ม 1		กลุ่ม 2	
			$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$	$\theta_1$	$\theta_2$
1	-2.0	0.2	8.88	83.61	81.12	6.39	9.76	83.47	80.24	6.53	7.40	83.70	82.60	6.30
2	-1.5	0.4	6.84	83.48	83.16	6.52	7.03	82.08	82.97	7.92	8.09	83.03	81.91	6.97
3	1.0	0.6	6.98	82.53	83.02	7.47	5.45	83.51	84.55	6.49	7.40	83.84	82.60	6.16
4	-0.5	0.8	7.55	81.68	82.45	8.32	7.81	82.47	82.19	7.53	7.52	83.60	82.48	6.40
5	0	1.0	6.93	83.90	83.07	6.10	6.40	83.39	83.60	6.61	8.51	82.99	81.49	7.01
6	0	1.2	7.01	82.69	82.99	7.31	7.60	82.50	82.40	7.50	7.08	83.98	82.92	6.02
7	-1.0	1.4	6.95	83.41	83.05	6.59	7.89	83.29	82.11	6.71	7.35	82.57	82.65	7.43
8	0.5	1.6	7.05	82.34	82.95	7.66	6.76	83.77	83.24	6.23	7.21	82.32	82.79	7.68
9	1.5	1.8	8.86	83.02	81.14	6.98	7.59	82.50	82.41	7.50	8.39	82.89	81.61	7.11
10	2.0	2.0	7.18	82.94	82.82	7.06	5.29	83.71	84.71	6.29	8.05	84.36	81.95	5.64
11	-2.0	0.2	8.60	82.55	81.40	7.45	7.40	82.74	82.60	7.26	7.95	83.71	82.05	6.29
12	-1.5	0.4	8.24	83.20	81.76	6.80	7.18	82.78	82.82	7.22				
13	1.0	0.6	8.33	84.35	81.67	5.65	6.24	82.09	83.76	7.91				
14	-0.5	0.8	5.74	84.85	84.26	5.15	7.48	82.61	82.52	7.39				
15	0	1.0	6.62	83.68	83.38	6.32	7.81	83.69	82.19	6.31				
16	0	1.2	7.04	82.53	82.96	7.47	7.28	83.44	82.72	6.56				
17	-1.0	1.4	7.59	82.58	82.41	7.42	6.46	82.28	83.54	7.72				
18	0.5	1.6	6.90	82.87	83.10	7.13	6.90	82.94	83.10	7.06				
19	1.5	1.8	7.95	83.27	82.05	6.73	8.04	82.91	81.96	7.09				
20	2.0	2.0	7.30	83.98	82.70	6.02	6.97	81.95	83.03	8.05				
21	-2.0	0.2	7.32	82.49	82.68	7.51	8.42	82.98	81.58	7.02				
22	-1.5	0.4	8.36	82.73	81.64	7.27	6.27	82.52	83.73	7.48				
23	1.0	0.6	7.20	83.96	82.80	6.04	7.29	82.33	82.71	7.67				
24	-0.5	0.8	8.56	84.38	81.44	5.62	8.40	84.40	81.60	5.60				
25	0	1.0	7.31	83.60	82.69	6.40	6.71	82.86	83.29	7.14				
ข้อสอบที่ตรวจให้คะแนนหลายค่า														
1	-2.0	0.2	7.70	83.79	82.30	6.21	7.32	82.89	82.68	7.11	7.75	82.63	82.25	7.37
2	-1.5	0.4	7.16	82.39	82.84	7.61	6.67	82.90	83.33	7.10	6.46	82.81	83.54	7.19
3	1.0	0.6	8.12	83.81	81.88	6.19	8.21	82.77	81.79	7.23				
4	-0.5	0.8	7.93	82.99	82.07	7.01	7.59	82.66	82.41	7.34				
5	0	1.0	9.25	82.92	80.75	7.08	8.04	82.98	81.96	7.02				



ตาราง 12 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 10% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60																
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม			
					1	2	3	4	1	2	3	4	1	2	3	4
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	8.58	36.61	52.35	83.76	6.17	36.62	54.37	82.82	6.75	38.75	83.64	
		2	-1.5	0.4	7.12	37.17	53.36	82.92	7.25	36.96	52.44	83.95	37.03			
		3	1.0	0.6	7.21	37.51	53.77	83.50	6.50	37.86	54.53	79.78				
		4	-0.5	0.8	7.87	37.64	52.96	82.40	7.71	38.18	51.31	83.32				
		5	0	1.0	8.74	38.82	53.08	81.68	9.90	38.23	53.31	82.34				
		6	0	1.2	6.70	38.08	51.80	82.67	9.20	38.27	54.05	84.20				
		7	-1.0	1.4	7.20	38.85	53.33	82.03	7.03	38.41	54.62	82.89				
		8	0.5	1.6	6.55	37.39	53.91	83.39	6.47	38.99	53.23	82.10				
		9	1.5	1.8	8.09	36.93	53.35	82.90	6.35	37.42	52.76	82.06				
		10	2.0	2.0	6.00	38.43	52.23	83.77	8.66	37.97	52.07	82.87				
		11	-2.0	0.2	7.37	37.67	52.95	82.83	7.39	37.72	51.15	82.77				
		12	-1.5	0.4	7.27	36.76	53.29	83.09	7.69	38.23	53.48	83.48				
		13	1.0	0.6	7.28	38.37	53.70	82.55	8.13	36.95	52.84	84.11				
		14	-0.5	0.8	8.97	39.29	53.86	82.04	7.09	38.11	52.74	83.06				
		15	0	1.0	6.45	38.25	53.67	82.19	8.59	36.06	53.46	82.80				
		16	0	1.2	7.89			83.02	7.17			83.08				
	วัดจาก $\theta_2$	1	-2.0	0.2	81.42	53.39	37.65	6.24	83.83	53.38	35.63	7.18				
		2	-1.5	0.4	82.88	52.83	36.64	7.08	82.75	53.04	37.56	6.05				
		3	1.0	0.6	82.79	52.49	36.23	6.50	83.50	52.14	35.47	10.22				
		4	-0.5	0.8	82.13	52.36	37.04	7.60	82.29	51.82	38.69	6.68				
		5	0	1.0	81.26	51.18	36.92	8.32	80.10	51.77	36.69	7.66				
		6	0	1.2	83.30	51.92	38.20	7.33	80.80	51.73	35.95	5.80				
		7	-1.0	1.4	82.80	51.15	36.67	7.97	82.97	51.59	35.38	7.11				
		8	0.5	1.6	83.45	52.61	36.09	6.61	83.53	51.01	36.77	7.90				
		9	1.5	1.8	81.91	53.07	36.65	7.10	83.65	52.58	37.24	7.94				
		10	2.0	2.0	84.00	51.57	37.77	6.23	81.34	52.03	37.93	7.13				
		11	-2.0	0.2	82.63	52.33	37.05	7.17	82.61	52.28	38.85	7.23				
		12	-1.5	0.4	82.73	53.24	36.71	6.91	82.31	51.77	36.52	6.52				
		13	1.0	0.6	82.72	51.63	36.30	7.45	81.87	53.05	37.16	5.89				
		14	-0.5	0.8	81.03	50.71	36.14	7.96	82.91	51.89	37.26	6.94				
		15	0	1.0	83.55	51.75	36.33	7.81	81.41	53.94	36.54	7.20				
		16	0	1.2	82.11			6.98	82.83			6.92				
POLYTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	6.30	37.68	52.51	82.56	7.12	36.48	53.17	83.86	7.27	53.95	83.74	
		2	-1.5	0.4	7.62	37.57	54.53	83.70	6.94	37.81	53.22	83.42				
		3	1.0	0.6	7.35	37.67	53.18	83.80	6.97	36.68	53.33	82.60				
		4	-0.5	0.8		37.77	52.96			37.26	53.63					
	วัดจาก $\theta_2$	1	-2.0	0.2	83.70	52.32	37.49	7.44	82.88	53.52	36.83	6.14	82.73	36.05	6.26	
		2	-1.5	0.4	82.38	52.43	35.47	6.30	83.06	52.19	36.78	6.58				
		3	1.0	0.6	82.65	52.33	36.82	6.20	83.03	53.32	36.67	7.40				
		4	-0.5	0.8		52.23	37.04			52.74	36.37					

ตาราง 13 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 10% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 60:40																
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม			
					1	2	3	4	1	2	3	4	1	2	3	4
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	7.88	39.63	52.09	82.34	6.67	38.03	53.31	83.13	8.78	37.50	52.41	83.98
		2	-1.5	0.4	7.14	39.50	52.11	81.11	7.03	37.50	52.36	83.55	38.85	52.84		
		3	1.0	0.6	7.86	36.99	52.04	81.92	8.69	39.49	54.43	84.14				
		4	-0.5	0.8	8.02	36.33	52.61	82.99	8.03	37.85	53.78	82.02				
		5	0	1.0	6.68	38.41	55.11	83.71	8.20	37.11	53.00	84.35				
		6	0	1.2	8.88	38.07	52.34	84.21	5.63	38.73	53.89	82.90				
		7	-1.0	1.4	7.09	37.82	52.52	84.03	7.82	37.15	53.11	82.71				
		8	0.5	1.6	7.85	38.32	52.40	83.24	8.28	38.16	51.54	82.67				
		9	1.5	1.8	7.22	37.74	51.54	82.58	8.12	39.25	52.02	81.64				
		10	2.0	2.0	6.47	36.87	53.60	83.64	7.39	38.54	51.33	83.30				
		11	-2.0	0.2	8.72	38.31	54.12	83.00	6.50	38.46	53.12	83.06				
		12	-1.5	0.4	8.62	37.73	52.62	82.30	9.24	37.18	53.45	81.77				
		13	1.0	0.6	7.41	38.46	52.41	83.18	6.34	37.72	51.73	82.82				
		14	-0.5	0.8	6.65	37.19	51.96	82.53	7.77	37.38	53.43	82.21				
		15	0	1.0	7.60	36.47	54.05	82.94	6.70	37.41	53.94	82.00				
		16	0	1.2	7.41			83.35	7.30			81.93				
	วัดจาก $\theta_2$	1	-2.0	0.2	82.12	50.37	37.91	7.66	83.33	51.97	36.69	6.87	81.22	52.50	37.59	6.02
		2	-1.5	0.4	82.86	50.50	37.89	8.89	82.97	52.50	37.64	6.45	51.15	37.16		
		3	1.0	0.6	82.14	53.01	37.96	8.08	81.31	50.51	35.57	5.86				
		4	-0.5	0.8	81.98	53.67	37.39	7.01	81.97	52.15	36.22	7.98				
		5	0	1.0	83.32	51.59	34.89	6.29	81.80	52.89	37.00	5.65				
		6	0	1.2	81.12	51.93	37.66	5.79	84.37	51.27	36.11	7.10				
		7	-1.0	1.4	82.91	52.18	37.48	5.97	82.18	52.85	36.89	7.29				
		8	0.5	1.6	82.15	51.68	37.60	6.76	81.72	51.84	38.46	7.33				
		9	1.5	1.8	82.78	52.26	38.46	7.42	81.88	50.75	37.98	8.36				
		10	2.0	2.0	83.53	53.13	36.40	6.36	82.61	51.46	38.67	6.70				
		11	-2.0	0.2	81.28	51.69	35.88	7.00	83.50	51.54	36.88	6.94				
		12	-1.5	0.4	81.38	52.27	37.38	7.70	80.76	52.82	36.55	8.23				
		13	1.0	0.6	82.59	51.54	37.59	6.82	83.66	52.28	38.27	7.18				
		14	-0.5	0.8	83.35	52.81	38.04	7.47	82.23	52.62	36.57	7.79				
		15	0	1.0	82.40	53.53	35.95	7.06	83.30	52.59	36.06	8.00				
		16	0	1.2	82.59			6.65	82.70			8.07				
POLYTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	7.30	37.92	53.00	82.38	7.32	37.99	54.50	83.29	8.12			83.86
		2	-1.5	0.4	5.92	38.07	52.37	82.89	7.84	38.38	52.61	82.80				
		3	1.0	0.6	6.42	38.00	53.11	83.24	7.79	37.57	54.05	83.42				
		4	-0.5	0.8		37.44	51.82			39.24	53.20					
	วัดจาก $\theta_2$	1	-2.0	0.2	82.70	52.08	37.00	7.62	82.68	52.01	35.50	6.71	81.88			6.14
		2	-1.5	0.4	84.08	51.93	37.63	7.11	82.16	51.62	37.39	7.20				
		3	1.0	0.6	83.58	52.00	36.89	6.76	82.21	52.43	35.95	6.58				
		4	-0.5	0.8		52.56	38.18			50.76	36.80					

ตาราง 14 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 10% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30																
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม			
					1	2	3	4	1	2	3	4	1	2	3	4
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	8.29	37.17	52.99	82.75	81.71	52.83	37.01	7.25	7.50	36.62	53.82	83.23
		2	-1.5	0.4	9.14	38.89	53.78	81.95	80.86	51.11	36.22	8.05		37.44	51.43	83.66
		3	1.0	0.6	7.14	37.92	52.86	84.41	82.86	52.08	37.14	5.59				82.56
		4	-0.5	0.8	5.93	38.30	52.73	83.63	84.07	51.70	37.27	6.37				
		5	0	1.0	5.97	39.02	53.17	83.67	84.03	50.98	36.83	6.33				
		6	0	1.2	7.94	38.71	52.67	83.46	82.06	51.29	37.33	6.54				
		7	-1.0	1.4	8.20	39.01	54.35	81.24	81.80	50.99	35.65	8.76				
		8	0.5	1.6	8.65	37.95	54.60	82.85	81.35	52.05	35.40	7.15				
		9	1.5	1.8	7.41	37.33	53.63	81.78	82.59	52.67	36.37	8.22				
		10	2.0	2.0	6.52	39.28	51.81	83.68	83.48	50.72	38.19	6.32				
		11	-2.0	0.2	7.67	40.35	52.32	83.19	82.33	49.65	37.68	6.81				
		12	-1.5	0.4	10.53	38.63	53.87	83.12	79.47	51.37	36.13	6.88				
		13	1.0	0.6	5.53	38.64	52.55	85.46	84.47	51.36	37.45	4.54				
		14	-0.5	0.8	8.42	37.70	53.77	84.14	81.58	52.30	36.23	5.86				
		15	0	1.0	6.60	40.31	51.97	81.67	83.40	49.69	38.03	8.33				
		16	0	1.2	7.40			83.33	82.60			6.67				
	วัดจาก $\theta_2$	1	-2.0	0.2	8.01	39.08	51.57	83.06	81.99	50.92	38.43	6.94	82.50	53.38	36.18	6.77
		2	-1.5	0.4	9.58	37.11	55.07	82.63	80.42	52.89	34.93	7.37		52.56	38.57	6.34
		3	1.0	0.6	9.33	35.89	52.39	84.11	80.67	54.11	37.61	5.89				7.44
		4	-0.5	0.8	9.05	39.34	52.04	82.63	80.95	50.66	37.96	7.37				
		5	0	1.0	7.60	38.45	51.04	82.74	82.40	51.55	38.96	7.26				
		6	0	1.2	8.66	36.10	52.38	82.41	81.34	53.90	37.62	7.59				
		7	-1.0	1.4	7.05	39.18	52.71	83.33	82.95	50.82	37.29	6.67				
		8	0.5	1.6	8.22	36.65	51.24	83.82	81.78	53.35	38.76	6.18				
		9	1.5	1.8	5.70	38.51	52.93	84.65	84.30	51.49	37.07	5.35				
		10	2.0	2.0	7.32	37.40	53.80	84.38	82.68	52.60	36.20	5.62				
		11	-2.0	0.2	7.41	39.12	54.55	83.38	82.59	50.88	35.45	6.62				
		12	-1.5	0.4	6.88	36.88	53.84	80.75	83.12	53.12	36.16	9.25				
		13	1.0	0.6	6.89	37.80	52.98	85.23	83.11	52.20	37.02	4.77				
		14	-0.5	0.8	5.93	38.30	53.35	84.71	84.07	51.70	36.65	5.29				
		15	0	1.0	7.06	39.99	53.51	84.11	82.94	50.01	36.49	5.89				
		16	0	1.2	6.12			80.85	83.88			9.15				
POLYTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	4.80	38.02	54.61	82.47	6.37	34.01	52.06	82.74	6.57			
		2	-1.5	0.4	5.41	38.18	52.84	85.11	5.49	38.31	52.65	82.27				
		3	1.0	0.6	8.62	36.81	52.19	85.38	7.38	36.35	52.08	83.37				
		4	-0.5	0.8		37.56	54.27			36.34	51.82					
	วัดจาก $\theta_2$	1	-2.0	0.2	85.20	51.98	35.39	7.53	83.63	55.99	37.94	7.26	83.43			
		2	-1.5	0.4	84.59	51.82	37.16	4.89	84.51	51.69	37.35	7.73				
		3	1.0	0.6	81.38	53.19	37.81	4.62	82.62	53.65	37.92	6.63				
		4	-0.5	0.8		52.44	35.73			53.66	38.18					

ตาราง 15 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 20% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60																
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม			
					1	2	3	4	1	2	3	4	1	2	3	4
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	9.76	37.21	52.55	82.02	8.01	39.90	51.34	81.40	8.58	37.18	55.00	83.94
		2	-1.5	0.4	6.52	39.52	54.04	85.48	9.91	37.21	52.92	82.42	38.56	54.78		
		3	1.0	0.6	7.93	37.65	52.57	84.16	7.12	38.80	52.97	82.93	38.31	52.28		
		4	-0.5	0.8	6.34	38.42	52.22	82.31	5.80	38.19	52.03	81.68				
		5	0	1.0	8.91	38.43	54.26	85.49	3.91	37.26	53.11	83.72				
		6	0	1.2	6.66	37.91	54.13	84.11	6.79	37.64	51.92	85.61				
		7	-1.0	1.4	8.26	38.06	52.44	83.71	8.12	38.78	49.97	81.92				
		8	0.5	1.6	6.83	36.99	52.70	81.81	9.06	36.92	53.10	82.54				
		9	1.5	1.8	6.30	39.30	51.99	83.63	8.21	38.53	52.42	81.84				
		10	2.0	2.0	6.67	39.04	54.89	82.42	7.67	36.54	52.57	84.59				
		11	-2.0	0.2	7.42	38.47	53.08	82.73	8.26	38.76	51.65	82.02				
		12	-1.5	0.4	9.76	35.57	51.95	82.61	4.99	37.59	51.86	82.44				
		13	1.0	0.6	8.83	38.40	53.28	82.59	5.98	39.00	52.32	81.77				
		14	-0.5	0.8	8.73	37.55	53.46	83.96	6.82	39.05	52.21	84.76				
	วัดจาก $\theta_2$	1	-2.0	0.2	80.24	52.79	37.45	7.98	81.99	50.10	38.66	8.60	81.42	52.82	35.00	6.06
		2	-1.5	0.4	83.48	50.48	35.96	4.52	80.09	52.79	37.08	7.58	51.44	35.22		
		3	1.0	0.6	82.07	52.35	37.43	5.84	82.88	51.20	37.03	7.07	51.69	37.72		
		4	-0.5	0.8	83.66	51.58	37.78	7.69	84.20	51.81	37.97	8.32				
		5	0	1.0	81.09	51.57	35.74	4.51	86.09	52.74	36.89	6.28				
		6	0	1.2	83.34	52.09	35.87	5.89	83.21	52.36	38.08	4.39				
		7	-1.0	1.4	81.74	51.94	37.56	6.29	81.88	51.22	40.03	8.08				
		8	0.5	1.6	83.17	53.01	37.30	8.19	80.94	53.08	36.90	7.46				
		9	1.5	1.8	83.70	50.70	38.01	6.37	81.79	51.47	37.58	8.16				
		10	2.0	2.0	83.33	50.96	35.11	7.58	82.33	53.46	37.43	5.41				
		11	-2.0	0.2	82.58	51.53	36.92	7.27	81.74	51.24	38.35	7.98				
		12	-1.5	0.4	80.24	54.43	38.05	7.39	85.01	52.41	38.14	7.56				
		13	1.0	0.6	81.17	51.60	36.72	7.41	84.02	51.00	37.68	8.23				
		14	-0.5	0.8	81.27	52.45	36.54	6.04	83.18	50.95	37.79	5.24				
POLYTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	6.75	37.75	51.96	82.39	83.25	52.25	38.04	7.61	8.95	38.06	54.84	82.88
		2	-1.5	0.4	7.19	38.79	50.80	83.25	82.81	51.21	39.20	6.75	6.18			82.64
		3	1.0	0.6	5.19	38.88	51.88	81.92	84.81	51.12	38.12	8.08				
	วัดจาก $\theta_2$	1	-2.0	0.2	8.75	37.08	52.19	83.70	81.25	52.92	37.81	6.30	81.05	51.94	35.16	7.12
		2	-1.5	0.4	7.55	36.88	54.23	83.07	82.45	53.12	35.77	6.93	83.82			7.36
		3	1.0	0.6	9.90	39.70	52.11	80.76	80.10	50.30	37.89	9.24				

ตาราง 16 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 20% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 50:50																
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม			
					1	2	3	4	1	2	3	4	1	2	3	4
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	7.16	38.66	52.38	80.80	8.26	37.37	50.89	82.88	8.30	37.86	53.96	82.37
		2	-1.5	0.4	6.92	38.35	52.38	81.90	7.27	38.14	51.95	84.41		37.08	52.67	82.73
		3	1.0	0.6	6.73	37.64	51.65	82.29	8.50	38.49	53.18	83.53		37.94	53.52	84.08
		4	-0.5	0.8	7.91	37.42	52.80	84.82	8.00	38.13	51.48	83.04				
		5	0	1.0	7.02	38.56	52.33	83.54	5.92	39.33	52.40	84.49				
		6	0	1.2	8.36	36.74	54.29	82.88	7.24	38.11	52.71	82.82				
		7	-1.0	1.4	6.99	39.00	54.56	82.25	8.03	39.80	52.69	82.53				
		8	0.5	1.6	8.82	37.47	54.40	82.23	6.38	36.73	52.25	82.41				
		9	1.5	1.8	7.66	38.33	55.05	83.32	8.10	36.81	52.78	83.63				
		10	2.0	2.0	6.73	35.98	51.77	82.75	6.81	36.87	52.85	80.68				
		11	-2.0	0.2	6.46	38.12	53.32	84.23	8.14	37.55	53.26	82.29				
		12	-1.5	0.4	7.28	38.64	53.00	81.06	6.02	36.71	54.37	81.50				
		13	1.0	0.6	8.20	38.25	54.08	81.87	7.32	37.07	54.61	81.92				
		14	-0.5	0.8	6.95	36.37	53.02	83.87	8.56	36.37	52.60	82.08				
	วัดจาก $\theta_2$	1	-2.0	0.2	82.84	51.34	37.62	9.20	81.74	52.63	39.11	7.12	81.70	52.14	36.04	7.63
		2	-1.5	0.4	83.08	51.65	37.62	8.10	82.73	51.86	38.05	5.59		52.92	37.33	7.27
		3	1.0	0.6	83.27	52.36	38.35	7.71	81.50	51.51	36.82	6.47		52.06	36.48	5.92
		4	-0.5	0.8	82.09	52.58	37.20	5.18	82.00	51.87	38.52	6.96				
		5	0	1.0	82.98	51.44	37.67	6.46	84.08	50.67	37.60	5.51				
		6	0	1.2	81.64	53.26	35.71	7.12	82.76	51.89	37.29	7.18				
		7	-1.0	1.4	83.01	51.00	35.44	7.75	81.97	50.20	37.31	7.47				
		8	0.5	1.6	81.18	52.53	35.60	7.77	83.62	53.27	37.75	7.59				
		9	1.5	1.8	82.34	51.67	34.95	6.68	81.90	53.19	37.22	6.37				
		10	2.0	2.0	83.27	54.02	38.23	7.25	83.19	53.13	37.15	9.32				
		11	-2.0	0.2	83.54	51.88	36.68	5.77	81.86	52.45	36.74	7.71				
		12	-1.5	0.4	82.72	51.36	37.00	8.94	83.98	53.29	35.63	8.50				
		13	1.0	0.6	81.80	51.75	35.92	8.13	82.68	52.93	35.39	8.08				
		14	-0.5	0.8	83.05	53.63	36.98	6.13	81.44	53.63	37.40	7.92				
POLYTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	7.44	37.18	53.46	83.88	6.28	37.90	52.01	82.56	8.06	39.31	52.65	82.92
		2	-1.5	0.4	8.92	37.90	53.41	82.88	7.94	38.81	51.18	82.32	5.74			
		3	1.0	0.6	6.37	37.57	53.43	83.14	6.93	37.39	53.12	83.50				
	วัดจาก $\theta_2$	1	-2.0	0.2	82.56	52.82	36.54	6.12	83.72	52.10	37.99	7.44	81.94	50.69	37.35	7.08
		2	-1.5	0.4	81.08	52.10	36.59	7.12	82.06	51.19	38.82	7.68	84.26			
		3	1.0	0.6	83.63	52.43	36.57	6.86	83.07	52.61	36.88	6.50				

ตาราง 17 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 20% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 60:40																
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม			
					1	2	3	4	1	2	3	4	1	2	3	4
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	8.99	38.00	52.88	84.15	7.80	38.01	53.08	83.69	7.50	38.39	52.45	82.95
		2	-1.5	0.4	8.35	38.79	53.40	84.08	8.74	38.94	52.25	82.66	6.93	38.60	51.82	83.43
		3	1.0	0.6	6.32	36.66	53.73	82.25	7.88	38.75	51.05	82.65	6.64	39.38	53.30	83.54
		4	-0.5	0.8	6.53	38.55	51.95	80.86	8.66	38.93	52.93	82.98				
		5	0	1.0	9.62	37.40	53.37	82.61	6.18	37.85	52.14	83.58				
		6	0	1.2	4.50	38.05	55.07	82.42	6.00	38.16	52.74	83.21				
		7	-1.0	1.4	6.81	37.22	53.07	82.10	7.96	36.77	51.31	83.55				
		8	0.5	1.6	7.43	36.24	52.30	82.68	7.94	37.35	53.18	83.36				
		9	1.5	1.8	6.56	39.63	52.89	83.34	7.02	36.45	52.40	82.55				
		10	2.0	2.0	6.99	37.74	53.64	83.38	7.76	37.06	52.91	82.98				
		11	-2.0	0.2	8.73	39.22	52.97	84.15	7.02	38.33	52.38	83.55				
		12	-1.5	0.4	7.44	38.32	53.57	84.57	8.83	38.67	52.32	82.46				
		13	1.0	0.6	7.55	37.09	52.58	83.29	8.27	37.58	52.18	83.21				
		14	-0.5	0.8	8.42	37.84	53.46	83.75	8.90	37.91	53.66	82.20				
	วัดจาก $\theta_2$	1	-2.0	0.2	81.01	52.00	37.12	5.85	82.20	51.99	36.92	6.31	82.50	51.61	37.55	7.05
		2	-1.5	0.4	81.65	51.21	36.60	5.92	81.26	51.06	37.75	7.34	83.07	51.40	38.18	6.57
		3	1.0	0.6	83.68	53.34	36.27	7.75	82.12	51.25	38.95	7.35	83.36	50.62	36.70	6.46
		4	-0.5	0.8	83.47	51.45	38.05	9.14	81.34	51.07	37.07	7.02				
		5	0	1.0	80.38	52.60	36.63	7.39	83.82	52.15	37.86	6.42				
		6	0	1.2	85.50	51.95	34.93	7.58	84.00	51.84	37.26	6.79				
		7	-1.0	1.4	83.19	52.78	36.93	7.90	82.04	53.23	38.69	6.45				
		8	0.5	1.6	82.57	53.76	37.70	7.32	82.06	52.65	36.82	6.64				
		9	1.5	1.8	83.44	50.37	37.11	6.66	82.98	53.55	37.60	7.45				
		10	2.0	2.0	83.01	52.26	36.36	6.62	82.24	52.94	37.09	7.02				
		11	-2.0	0.2	81.27	50.78	37.03	5.85	82.98	51.67	37.62	6.45				
		12	-1.5	0.4	82.56	51.68	36.43	5.43	81.17	51.33	37.68	7.54				
		13	1.0	0.6	82.45	52.91	37.42	6.71	81.73	52.42	37.82	6.79				
		14	-0.5	0.8	81.58	52.16	36.54	6.25	81.10	52.09	36.34	7.80				
POLYTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	7.62	37.48	52.98	84.61	8.18	37.44	51.92	82.34	8.09	38.52	53.56	83.07
		2	-1.5	0.4	8.18	37.36	53.07	82.98	7.82	38.66	52.19	83.23				
		3	1.0	0.6	7.95	38.28	51.67	84.10	6.12	37.77	53.10	83.11				
	วัดจาก $\theta_2$	1	-2.0	0.2												
		2	-1.5	0.4	82.38	52.52	37.02	5.39	81.82	52.56	38.08	7.66	81.91	51.48	36.44	6.93
		3	1.0	0.6	81.82	52.64	36.93	7.02	82.18	51.34	37.81	6.77				

ตาราง 18 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 20% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30																
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม			
					1	2	3	4	1	2	3	4	1	2	3	4
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	8.60	37.73	52.50	85.03	7.40	39.51	51.92	82.71	5.30	37.38	52.18	82.71
		2	-1.5	0.4	7.85	38.43	53.59	81.79	7.75	38.07	53.45	83.70	7.30	38.67	52.72	84.85
		3	1.0	0.6	8.44	38.25	53.89	83.95	8.49	37.76	52.36	82.02	7.83	39.39	53.37	84.04
		4	-0.5	0.8	7.36	37.53	53.18	81.69	6.59	38.06	52.96	81.88	6.61			
		5	0	1.0	8.60	37.94	52.60	84.00	7.11	39.19	53.42	82.19	8.55			
		6	0	1.2	8.44	38.91	53.58	83.43	7.55	37.56	53.34	84.03				
		7	-1.0	1.4	8.05	38.69	53.22	82.19	8.65	38.32	54.10	82.32				
		8	0.5	1.6	6.47	37.65	53.63	82.16	8.92	38.80	53.41	82.56				
		9	1.5	1.8	6.90	38.77	53.86	82.50	5.12	37.57	51.67	82.26				
		10	2.0	2.0	6.21	37.18	52.72	83.44	8.35	39.32	53.62	84.21				
		11	-2.0	0.2	7.14	36.45	51.90	83.54	6.61	38.95	54.84	84.37				
		12	-1.5	0.4	7.48	38.15	52.62	82.68	7.41	37.47	53.81	83.53				
		13	1.0	0.6	6.68	36.83	53.17	83.76	6.96	39.90	53.78	82.84				
		14	-0.5	0.8	6.37	37.80	52.23	82.70	7.86	39.12	53.23	83.31				
	วัดจาก $\theta_2$	1	-2.0	0.2	81.40	52.27	37.50	4.97	82.60	50.49	38.08	7.29	84.70	52.62	37.82	7.29
		2	-1.5	0.4	82.15	51.57	36.41	8.21	82.25	51.93	36.55	6.30	82.70	51.33	37.28	5.15
		3	1.0	0.6	81.56	51.75	36.11	6.05	81.51	52.24	37.64	7.98	82.17	50.61	36.63	5.96
		4	-0.5	0.8	82.64	52.47	36.82	8.31	83.41	51.94	37.04	8.12	83.39			
		5	0	1.0	81.40	52.06	37.40	6.00	82.89	50.81	36.58	7.81	81.45			
		6	0	1.2	81.56	51.09	36.42	6.57	82.45	52.44	36.66	5.97				
		7	-1.0	1.4	81.95	51.31	36.78	7.81	81.35	51.68	35.90	7.68				
		8	0.5	1.6	83.53	52.35	36.37	7.84	81.08	51.20	36.59	7.44				
		9	1.5	1.8	83.10	51.23	36.14	7.50	84.88	52.43	38.33	7.74				
		10	2.0	2.0	83.79	52.82	37.28	6.56	81.65	50.68	36.38	5.79				
		11	-2.0	0.2	82.86	53.55	38.10	6.46	83.39	51.05	35.16	5.63				
		12	-1.5	0.4	82.52	51.85	37.38	7.32	82.59	52.53	36.19	6.47				
		13	1.0	0.6	83.32	53.17	36.83	6.24	83.04	50.10	36.22	7.16				
		14	-0.5	0.8	83.63	52.20	37.77	7.30	82.14	50.88	36.77	6.69				
POLYTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	7.64	38.32	52.71	81.37	8.12	37.75	54.12	83.26	37.15	52.48	83.71	
		2	-1.5	0.4	8.36	38.48	52.50	82.28	7.51	37.02	52.57	82.85				
		3	1.0	0.6	9.37	38.37	52.91	82.47	7.86	38.01	53.27	82.13				
	วัดจาก $\theta_2$	1	-2.0	0.2	82.36	51.68	37.29	8.63	81.88	52.25	35.88	6.74	52.85	37.52	6.29	
		2	-1.5	0.4	81.64	51.52	37.50	7.72	82.49	52.98	37.43	7.15				
		3	1.0	0.6	80.63	51.63	37.09	7.53	82.14	51.99	36.73	7.87				

ตาราง 19 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 30% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 40:60																
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม			
					1	2	3	4	1	2	3	4	1	2	3	4
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	7.87	37.47	52.79	82.05	8.15	39.00	52.52	82.58	7.68	37.96	53.46	82.75
		2	-1.5	0.4	5.57	37.18	53.69	82.89	7.36	37.89	53.12	82.04	6.59	36.41	53.29	83.42
		3	1.0	0.6	7.41	38.62	52.89	83.22	8.17	38.97	53.57	80.88	8.42			82.14
		4	-0.5	0.8	7.57	39.89	53.44	82.49	8.26	38.72	52.92	84.23				
		5	0	1.0	7.23	38.22	53.81	82.76	9.48	38.62	52.67	85.10				
		6	0	1.2	5.95	36.06	50.90	83.02	8.96	38.98	54.19	82.81				
		7	-1.0	1.4	6.58	38.47	52.01	82.10	7.59	38.28	52.31	82.76				
		8	0.5	1.6	9.59	37.97	51.21	82.63	7.97	38.86	53.44	82.97				
		9	1.5	1.8	6.41	37.10	54.76	82.73	6.19	39.76	52.83	83.44				
		10	2.0	2.0	8.21	37.07	54.21	81.83	8.16	39.16	52.37	83.34				
		11	-2.0	0.2	7.72	37.90	53.72	82.64	5.66	38.16	54.51	82.57				
		12	-1.5	0.4	8.34	37.55	52.66	83.08	8.12	37.73	53.86	84.43				
		13	1.0	0.6		38.44	52.87			36.91	53.15					
	วัดจาก $\theta_2$	1	-2.0	0.2	82.13	52.53	37.21	7.95	81.85	51.00	37.48	7.42	82.32	52.04	36.54	7.25
		2	-1.5	0.4	84.43	52.82	36.31	7.11	82.64	52.11	36.88	7.96	83.41	53.59	36.71	6.58
		3	1.0	0.6	82.59	51.38	37.11	6.78	81.83	51.03	36.43	9.12	81.58			7.86
		4	-0.5	0.8	82.43	50.11	36.56	7.51	81.74	51.28	37.08	5.77				
		5	0	1.0	82.77	51.78	36.19	7.24	80.52	51.38	37.33	4.90				
		6	0	1.2	84.05	53.94	39.10	6.98	81.04	51.02	35.81	7.19				
		7	-1.0	1.4	83.42	51.53	37.99	7.90	82.41	51.72	37.69	7.24				
		8	0.5	1.6	80.41	52.03	38.79	7.37	82.03	51.14	36.56	7.03				
		9	1.5	1.8	83.59	52.90	35.24	7.27	83.81	50.24	37.17	6.56				
		10	2.0	2.0	81.79	52.93	35.79	8.17	81.84	50.84	37.63	6.66				
		11	-2.0	0.2	82.28	52.10	36.28	7.36	84.34	51.84	35.49	7.43				
		12	-1.5	0.4	81.66	52.45	37.34	6.92	81.88	52.27	36.14	5.57				
		13	1.0	0.6		51.56	37.13			53.09	36.85					
POLYTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	6.55	36.51	53.34	82.91	8.48	39.08	53.71	82.81	7.29	38.79	53.94	82.68
		2	-1.5	0.4	8.27	39.82	53.44	81.85	7.32	37.99	52.27	81.37	8.98	35.75	53.40	82.46
		3	1.0	0.6		39.26	53.99			38.43	52.37		6.62			82.53
	วัดจาก $\theta_2$	1	-2.0	0.2	83.45	53.49	36.66	7.09	81.52	50.92	36.29	7.19	82.71	51.21	36.06	7.32
		2	-1.5	0.4	81.73	50.18	36.56	8.15	82.68	52.01	37.73	8.63	81.02	54.25	36.60	7.54
		3	1.0	0.6		50.74	36.01			51.57	37.63		83.38			7.47



ตาราง 20 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 30% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 50:50																
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม			
					1	2	3	4	1	2	3	4	1	2	3	4
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	8.83	37.69	53.22	84.89	7.62	37.42	52.25	80.77	7.27	36.33	52.92	83.10
		2	-1.5	0.4	6.81	37.53	54.12	83.46	9.28	37.07	53.09	83.35	8.48	37.66	53.89	82.71
		3	1.0	0.6	6.93	38.75	52.86	83.28	7.32	37.93	52.97	82.74	7.90			83.35
		4	-0.5	0.8	7.74	36.21	53.20	81.16	7.94	37.21	53.99	83.85	9.11			83.73
		5	0	1.0	7.16	36.73	52.16	84.04	6.95	37.53	53.40	83.90	7.88			83.64
		6	0	1.2	9.08	37.38	53.17	82.93	7.05	38.90	53.13	83.88				
		7	-1.0	1.4	6.40	37.37	51.34	84.08	8.83	38.63	53.03	84.02				
		8	0.5	1.6	7.22	37.74	53.56	83.15	7.89	38.19	53.00	83.12				
		9	1.5	1.8	8.24	37.43	53.83	80.24	9.54	36.97	51.58	83.08				
		10	2.0	2.0	6.37	38.29	52.45	82.57	7.94	38.85	53.04	82.36				
		11	-2.0	0.2	8.85	37.19	52.84	83.48	7.04	38.79	53.11	82.57				
		12	-1.5	0.4	7.74	39.78	54.64	83.02	7.12	37.66	52.52	82.41				
		13	1.0	0.6			36.99	53.21			39.07	54.51				
	วัดจาก $\theta_2$	1	-2.0	0.2	81.17	52.31	36.78	5.11	82.38	52.58	37.75	9.23	82.73	53.67	37.08	6.90
		2	-1.5	0.4	83.19	52.47	35.88	6.54	80.72	52.93	36.91	6.65	81.52	52.34	36.11	7.29
		3	1.0	0.6	83.07	51.25	37.14	6.72	82.68	52.07	37.03	7.26	82.10			6.65
		4	-0.5	0.8	82.26	53.79	36.80	8.84	82.06	52.79	36.01	6.15	80.89			6.27
		5	0	1.0	82.84	53.27	37.84	5.96	83.05	52.47	36.60	6.10	82.12			6.36
		6	0	1.2	80.92	52.62	36.83	7.07	82.95	51.10	36.87	6.12				
		7	-1.0	1.4	83.60	52.63	38.66	5.92	81.17	51.37	36.97	5.98				
		8	0.5	1.6	82.78	52.26	36.44	6.85	82.11	51.81	37.00	6.88				
		9	1.5	1.8	81.76	52.57	36.17	9.76	80.46	53.03	38.42	6.92				
		10	2.0	2.0	83.63	51.71	37.55	7.43	82.06	51.15	36.96	7.64				
		11	-2.0	0.2	81.15	52.81	37.16	6.52	82.96	51.21	36.89	7.43				
		12	-1.5	0.4	82.26	50.22	35.36	6.98	82.88	52.34	37.48	7.59				
		13	1.0	0.6			53.01	36.79			50.93	35.49				
POLYTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	8.60	38.17	54.59	82.83	8.48	38.91	53.45	83.77	7.81	38.72	52.86	82.71
		2	-1.5	0.4	6.97	35.87	53.48	82.21	5.98	38.14	52.96	84.80	8.67	38.60	52.85	82.19
		3	1.0	0.6		37.30	52.63			38.57	55.06					
	วัดจาก $\theta_2$	1	-2.0	0.2	81.40	51.83	35.41	7.17	81.52	51.09	36.55	6.23	82.19	51.28	37.14	7.29
		2	-1.5	0.4	83.03	54.13	36.52	7.79	84.02	51.86	37.04	5.20	81.33	51.40	37.15	7.81
		3	1.0	0.6		52.70	37.37			51.43	34.94					

ตาราง 21 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 30% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 60:40																	
ประเภท	มุม	ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม				
					1	2	3	4	1	2	3	4	1	2	3	4	
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	8.46	36.29	52.94	83.76	7.08	37.46	53.59	83.39	8.39	38.75	53.73	82.32	
		2	-1.5	0.4	6.70	39.29	52.40	80.90	8.97	36.27	52.72	82.45	6.93	38.07	53.94	81.80	
		3	1.0	0.6	7.27	37.64	51.91	82.83	6.28	37.54	54.67	83.44	7.38	38.39	54.03	84.19	
		4	-0.5	0.8	9.06	39.40	51.95	82.13	8.14	36.91	52.51	82.72	7.41	38.74	51.95	82.60	
		5	0	1.0	8.14	37.70	51.84	82.94	8.28	37.34	52.57	84.54	5.35			84.52	
		6	0	1.2	5.41	38.31	52.85	83.11	8.42	38.92	53.76	82.81					
		7	-1.0	1.4	7.17	38.21	51.70	83.13	7.49	38.14	52.46	82.07					
		8	0.5	1.6	8.70	37.73	53.23	84.20	5.82	38.17	53.22	84.01					
		9	1.5	1.8	7.34	38.43	53.60	83.46	7.07	37.54	53.63	80.92					
		10	2.0	2.0	7.06	37.44	53.32	82.79	6.61	37.34	52.67	82.16					
		11	-2.0	0.2	9.25	36.70	53.42	82.67	9.69	37.82	54.74	84.58					
		12	-1.5	0.4	6.60	39.30	51.93	83.52	7.62	39.13	53.40	82.65					
		13	1.0	0.6		37.80	52.73			38.26	52.72						
	วัดจาก $\theta_2$	1	-2.0	0.2	81.54	53.71	37.06	6.24	82.92	52.54	36.41	6.61	81.61	51.25	36.27	7.68	
		2	-1.5	0.4	83.30	50.71	37.60	9.10	81.03	53.73	37.28	7.55	83.07	51.93	36.06	8.20	
		3	1.0	0.6	82.73	52.36	38.09	7.17	83.72	52.46	35.33	6.56	82.62	51.61	35.97	5.81	
		4	-0.5	0.8	80.94	50.60	38.05	7.87	81.86	53.09	37.49	7.28	82.59	51.26	38.05	7.40	
		5	0	1.0	81.86	52.30	38.16	7.06	81.72	52.66	37.43	5.46	84.65			5.48	
		6	0	1.2	84.59	51.69	37.15	6.89	81.58	51.08	36.24	7.19					
		7	-1.0	1.4	82.83	51.79	38.30	6.87	82.51	51.86	37.54	7.93					
		8	0.5	1.6	81.30	52.27	36.77	5.80	84.18	51.83	36.78	5.99					
		9	1.5	1.8	82.66	51.57	36.40	6.54	82.93	52.46	36.37	9.08					
		10	2.0	2.0	82.94	52.56	36.68	7.21	83.39	52.66	37.33	7.84					
		11	-2.0	0.2	80.75	53.30	36.58	7.33	80.31	52.18	35.26	5.42					
		12	-1.5	0.4	83.40	50.70	38.07	6.48	82.38	50.87	36.60	7.35					
		13	1.0	0.6		52.20	37.27			51.74	37.28						
POLYTONOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	6.97	36.22	54.03	81.79	6.92	38.13	53.20	84.02	8.19	37.63	51.06	81.95	
		2	-1.5	0.4	7.73	37.97	53.47	82.71	6.08	37.02	52.97	82.77	6.92			82.50	
		3	1.0	0.6		37.53	53.83			38.61	54.38						
	วัดจาก $\theta_2$	1	-2.0	0.2	83.03	53.78	35.97	8.21	83.08	51.87	36.80	5.98	81.81	52.37	38.94	8.05	
		2	-1.5	0.4	82.27	52.03	36.53	7.29	83.92	52.98	37.03	7.23	83.08			7.50	
		3	1.0	0.6		52.47	36.17			51.39	35.62						

ตาราง 22 ค่าความยากรวม ค่าอำนาจจำแนกรวม และมุมเฉลี่ยสำหรับแบบสอบที่มีโครงสร้าง CS

สัดส่วนคะแนนข้อสอบรวม 30% เมื่ออัตราส่วนคะแนนข้อสอบรวมที่ตรวจให้คะแนนสองค่า 70:30																	
ประเภท		ลำดับ	MDIFF	MDISC	ข้อสอบเฉพาะ X				ข้อสอบเฉพาะ Y				ข้อสอบรวม				
					1	2	3	4	1	2	3	4	1	2	3	4	
DICHOTOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	6.26	37.93	53.22	81.97	7.89	36.62	51.84	83.50	6.57	38.04	52.20	82.87	
		2	-1.5	0.4	8.11	35.43	52.87	83.69	6.41	40.19	52.82	83.31	7.04	37.81	53.70	82.20	
		3	1.0	0.6	7.82	39.45	53.16	82.66	7.58	37.70	53.98	82.52	7.91	38.62	53.87	82.38	
		4	-0.5	0.8	7.45	37.13	53.97	83.37	9.47	39.16	51.96	81.67	7.22	37.74	54.06	85.09	
		5	0	1.0	8.33	38.44	50.90	82.56	6.10	38.22	54.17	82.87	7.82			84.64	
		6	0	1.2	9.18	37.90	52.02	81.73	9.00	36.32	52.76	83.36	6.56				84.23
		7	-1.0	1.4	6.76	38.06	52.09	84.19	6.10	38.34	53.90	81.99	7.18				81.85
		8	0.5	1.6	7.59	38.49	52.12	82.41	7.04	38.19	52.04	82.00					
		9	1.5	1.8	8.48	37.72	52.92	83.45	8.54	38.51	52.49	82.12					
		10	2.0	2.0	7.79	38.43	53.21	83.83	7.82	38.38	53.05	83.38					
		11	-2.0	0.2	8.61	36.66	53.64	83.73	9.03	38.70	52.05	83.27					
		12	-1.5	0.4	7.58	37.92	53.68	82.01	8.23	37.51	52.52	83.09					
		13	1.0	0.6		38.91	51.92			38.37	50.97						
	วัดจาก $\theta_2$	1	-2.0	0.2	83.74	52.07	36.78	8.03	82.11	53.38	38.16	6.50	83.43	51.96	37.80	7.13	
		2	-1.5	0.4	81.89	54.57	37.13	6.31	83.59	49.81	37.18	6.69	82.96	52.19	36.30	7.80	
		3	1.0	0.6	82.18	50.55	36.84	7.34	82.42	52.30	36.02	7.48	82.09	51.38	36.13	7.62	
		4	-0.5	0.8	82.55	52.87	36.03	6.63	80.53	50.84	38.04	8.33	82.78	52.26	35.94	4.91	
		5	0	1.0	81.67	51.56	39.10	7.44	83.90	51.78	35.83	7.13	82.18			5.36	
		6	0	1.2	80.82	52.10	37.98	8.27	81.00	53.68	37.24	6.64	83.44			5.77	
		7	-1.0	1.4	83.24	51.94	37.91	5.81	83.90	51.66	36.10	8.01	82.82			8.15	
		8	0.5	1.6	82.41	51.51	37.88	7.59	82.96	51.81	37.96	8.00					
		9	1.5	1.8	81.52	52.28	37.08	6.55	81.46	51.49	37.51	7.88					
		10	2.0	2.0	82.21	51.57	36.79	6.17	82.18	51.62	36.95	6.62					
		11	-2.0	0.2	81.39	53.34	36.36	6.27	80.97	51.30	37.95	6.73					
		12	-1.5	0.4	82.42	52.08	36.32	7.99	81.77	52.49	37.48	6.91					
		13	1.0	0.6		51.09	38.08			51.63	39.03						
POLYTONOMOUS	วัดจาก $\theta_1$	1	-2.0	0.2	7.97	38.15	53.09	81.65	7.72	37.74	53.28	83.05	7.16	38.48	52.10	82.81	
		2	-1.5	0.4	8.34	38.55	54.03	82.47	7.41	37.38	53.38	83.54					
		3	1.0	0.6		38.21	52.91			37.19	52.42						
	วัดจาก $\theta_2$	1	-2.0	0.2	82.03	51.85	36.91	8.35	82.28	52.26	36.62	6.95	82.84	51.52	37.90	7.19	
		2	-1.5	0.4	81.66	51.45	35.97	7.53	82.59	52.62	37.58	6.46					
		3	1.0	0.6		51.79	37.09			52.81							



ภาคผนวก ข

ผลการตรวจสอบความถูกต้องของคำมูมที่จำลองขึ้น

เงื่อนไขสัดส่วนคะแนนข้อสอบรวม 30% และสัดส่วนคะแนนข้อสอบรวมแต่ละประเภท 70:30

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY

ตาราง 23 ค่าสถิติของมุมที่จำลองขึ้นสำหรับแบบสอบกลุ่มแรกที่มีโครงสร้างที่ไม่ซับซ้อน (APSS)

แบบสอบย่อยกลุ่มแรก (เวกเตอร์ของข้อสอบในมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ( $\theta_1$ ) ตั้งแต่ 0 - 15 องศา)								
ข้อที่	มุม 0-15 องศา				มุม 90- $\theta_1$ องศา			
	Min	Max	Mean	SD	Min	Max	Mean	SD
ข้อสอบเฉพาะที่ตรวจให้คะแนนสองค่า กลุ่มแรก								
1	.02	2.06	.51	.46	88.62	89.98	89.51	0.42
2	.03	2.27	.97	.69	87.29	89.96	88.95	0.65
3	.40	3.15	1.49	.81	85.19	89.76	88.33	0.97
4	.75	3.83	1.94	.95	84.95	89.25	87.87	0.93
5	1.06	5.07	2.42	1.16	84.89	88.89	87.41	1.04
6	1.34	5.81	2.80	1.32	84.79	88.24	86.70	1.06
7	1.66	6.04	3.41	1.31	84.48	88.14	86.25	1.13
8	1.69	6.70	4.29	1.48	83.84	87.85	85.86	1.23
9	1.72	7.19	4.78	1.49	82.29	87.75	85.11	1.53
10	2.41	8.02	5.27	1.51	82.16	87.42	84.65	1.52
11	3.49	8.47	5.87	1.48	81.26	86.37	83.81	1.52
12	4.35	9.32	6.45	1.48	80.77	85.51	83.31	1.46
13	4.72	9.38	7.05	1.52	80.20	84.96	82.60	1.36
14	4.99	9.88	7.58	1.45	79.63	84.52	82.01	1.48
15	5.27	10.33	8.05	1.53	78.64	84.48	81.31	1.65
16	6.57	11.01	8.69	1.43	78.54	84.28	80.84	1.53
17	6.95	12.32	9.30	1.47	77.61	83.87	80.45	1.58
18	7.12	12.37	10.00	1.37	77.49	83.80	79.83	1.65
19	8.26	12.51	10.52	1.20	76.51	81.55	78.97	1.35
20	9.39	12.66	11.04	.89	76.28	81.45	78.44	1.23
21	9.95	13.25	11.75	.87	76.21	80.95	78.00	1.14
22	10.01	14.04	12.36	.99	76.15	78.62	77.30	0.74
23	11.87	14.19	13.06	.76	75.35	78.16	76.74	0.73
24	12.30	14.75	13.73	.66	75.34	78.14	76.22	0.70
25	13.30	14.99	14.53	.46	75.00	78.04	75.63	0.76
ข้อสอบร่วมที่ตรวจให้คะแนนสองค่า กลุ่มแรก								
51	.07	2.77	1.12	.77	87.23	89.93	88.88	0.77
52	.42	4.84	2.51	1.21	85.16	89.58	87.49	1.21
53	.72	7.49	3.63	1.68	82.51	89.28	86.37	1.68
54	1.70	7.94	4.78	1.84	82.06	88.30	85.22	1.84
55	2.29	9.36	6.25	1.81	80.64	87.71	83.75	1.81

ตาราง 23 (ต่อ)

แบบสอบย่อยกลุ่มแรก (เวกเตอร์ของข้อสอบในมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ( $\theta_1$ ) ตั้งแต่ 0 - 15 องศา)								
ข้อที่	มุม 0-15 องศา				มุม 90- $\theta_1$ องศา			
	Min	Max	Mean	SD	Min	Max	Mean	SD
ข้อสอบรวมที่ตรวจให้คะแนนสองค่า กลุ่มแรก								
56	2.78	10.63	7.50	1.70	79.37	87.22	82.50	1.70
57	4.30	11.44	8.60	1.63	78.56	85.70	81.40	1.63
58	7.04	13.76	10.01	1.91	76.24	82.96	79.99	1.91
59	7.93	14.12	10.92	1.83	75.88	82.07	79.08	1.83
60	9.16	14.80	12.49	1.54	75.20	80.84	77.51	1.54
61	11.45	14.91	13.86	.89	75.09	78.55	76.14	0.89
ข้อสอบเฉพาะที่ตรวจให้คะแนนหลายค่า กลุ่มแรก								
73	.38	6.65	2.67	1.90	83.35	89.62	87.33	1.90
74	.98	8.98	4.61	2.16	81.02	89.02	85.39	2.16
75	1.21	11.05	7.04	2.90	78.95	88.79	82.96	2.90
76	5.72	13.46	9.21	2.50	76.54	84.28	80.79	2.50
77	6.41	14.94	11.97	2.41	75.06	83.59	78.03	2.41
ข้อสอบรวมที่ตรวจให้คะแนนหลายค่า กลุ่มแรก								
83	.25	11.83	4.98	3.41	78.17	89.75	85.02	3.41
84	.91	14.85	9.28	4.10	75.15	89.09	80.72	4.10

ตาราง 24 ค่าสถิติของมุมที่จำลองขึ้นสำหรับแบบสอบกลุ่มสองที่มีโครงสร้างที่ไม่ซับซ้อน (APSS)

แบบสอบย่อยกลุ่มสอง (เวกเตอร์ของข้อสอบในมิติที่ 2 ทำมุมกับแกนมิติที่ 1 ( $\theta_2$ ) ตั้งแต่ 76 - 90 องศา)								
ข้อที่	มุม 76-90 องศา				มุม 90- $\theta_2$ องศา			
	Min	Max	Mean	SD	Min	Max	Mean	SD
ข้อสอบเฉพาะที่ตรวจให้คะแนนสองค่า กลุ่มสอง								
26	76.05	77.13	76.51	0.33	12.87	13.95	13.49	0.33
27	76.15	79.37	77.24	0.78	10.63	13.85	12.76	0.78
28	76.26	79.86	77.95	0.98	10.14	13.74	12.05	0.98
29	76.65	80.75	78.55	1.12	9.25	13.35	11.45	1.12
30	77.19	80.89	79.10	1.15	9.11	12.81	10.90	1.15
31	77.45	82.61	79.88	1.43	7.39	12.55	10.12	1.43
32	77.45	82.63	80.25	1.52	7.37	12.55	9.75	1.52
33	77.49	83.50	81.06	1.62	6.50	12.51	8.94	1.62
34	77.49	83.95	81.42	1.59	6.05	12.51	8.58	1.59
35	78.22	84.06	81.97	1.57	5.94	11.78	8.03	1.57
36	79.70	84.64	82.46	1.48	5.36	10.30	7.54	1.48
37	81.01	85.58	83.14	1.38	4.42	8.99	6.86	1.38
38	81.30	86.79	83.76	1.51	3.21	8.70	6.24	1.51
39	81.65	87.55	84.20	1.62	2.45	8.35	5.80	1.62
40	82.70	87.63	84.97	1.27	2.37	7.30	5.03	1.27
41	83.61	87.70	85.48	1.14	2.30	6.39	4.52	1.14
42	83.63	87.74	85.99	1.06	2.26	6.37	4.01	1.06
43	83.76	88.01	86.39	1.06	1.99	6.24	3.61	1.06
44	83.93	88.23	86.89	1.08	1.77	6.07	3.11	1.08
45	84.13	88.59	87.26	1.05	1.41	5.87	2.74	1.05
46	85.08	89.05	87.66	1.01	.95	4.92	2.34	1.01
47	86.01	89.48	88.18	0.99	.52	3.99	1.82	0.99
48	87.13	89.78	88.76	0.67	.22	2.87	1.24	0.67
49	87.36	89.85	89.13	0.59	.15	2.64	0.87	0.59
50	87.56	89.96	89.54	0.55	.04	2.44	0.46	0.55
ข้อสอบรวมที่ตรวจให้คะแนนสองค่า กลุ่มสอง								
62	76.10	78.91	76.97	0.80	11.09	13.90	13.03	0.80
63	76.61	81.44	78.18	1.39	8.56	13.39	11.82	1.39
64	77.31	82.33	79.45	1.53	7.67	12.69	10.55	1.53
65	77.42	83.68	80.06	1.69	6.32	12.58	9.94	1.69
66	78.46	84.05	81.24	1.63	5.95	11.54	8.76	1.63

ตาราง 24 (ต่อ)

แบบสอบย่อยกลุ่มสอง (เวกเตอร์ของข้อสอบในมิติที่ 2 ทำมุมกับแกนมิติที่ 1 ( $\theta_2$ ) ตั้งแต่ 76 - 90 องศา)								
ข้อที่	มุม 76-90 องศา				มุม 90- $\theta_2$ องศา			
	Min	Max	Mean	SD	Min	Max	Mean	SD
ข้อสอบรวมที่ตรวจให้คะแนนสองค่า กลุ่มสอง								
67	79.82	85.32	82.44	1.55	4.68	10.18	7.56	1.55
68	81.56	87.39	83.95	1.64	2.61	8.44	6.05	1.64
69	82.88	88.38	85.27	1.71	1.62	7.12	4.73	1.71
70	83.91	89.23	86.73	1.77	.77	6.09	3.27	1.77
71	85.09	89.36	87.81	1.51	.64	4.91	2.19	1.51
72	86.18	89.87	88.88	0.97	.13	3.82	1.12	0.97
ข้อสอบเฉพาะที่ตรวจให้คะแนนหลายค่า กลุ่มสอง								
78	76.55	84.83	79.47	2.17	5.17	13.45	10.53	2.17
79	78.26	85.41	81.47	2.10	4.59	11.74	8.53	2.10
80	78.65	86.69	83.14	1.95	3.31	11.35	6.86	1.95
81	80.52	88.70	85.61	2.31	1.30	9.48	4.39	2.31
82	80.95	89.83	87.99	1.98	.17	9.05	2.01	1.98
ข้อสอบรวมที่ตรวจให้คะแนนหลายค่า กลุ่มสอง								
85	76.61	86.00	80.42	2.75	4.00	13.39	9.58	2.75
86	76.83	89.58	85.22	3.63	.42	13.17	4.78	3.63



ตาราง 25 ค่าสถิติของมุมที่จำลองขึ้นสำหรับแบบสอบกลุ่มแรกที่มีโครงสร้างที่ซับซ้อน (CS)

แบบสอบย่อยกลุ่มแรก (เวกเตอร์ของข้อสอบในมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ( $\theta_1$ ) ตั้งแต่ 0 - 15 องศา)								
ข้อที่	มุม 0-15 องศา				มุม 90- $\theta_1$ องศา			
	Min	Max	Mean	SD	Min	Max	Mean	SD
ข้อสอบเฉพาะที่ตรวจให้คะแนนสองค่า กลุ่มแรก								
1	.11	6.45	1.56	1.36	83.55	89.89	88.44	1.36
2	.81	6.49	2.85	1.34	83.51	89.19	87.15	1.34
3	1.16	9.08	4.44	2.12	80.92	88.84	85.56	2.12
4	2.96	9.42	5.84	1.93	80.58	87.04	84.16	1.93
5	3.94	11.03	6.96	1.89	78.97	86.06	83.04	1.89
6	4.58	11.47	8.35	1.84	78.53	85.42	81.65	1.84
7	5.29	11.61	9.37	1.53	78.39	84.71	80.63	1.53
8	8.31	12.99	10.47	1.20	77.01	81.69	79.53	1.20
9	9.17	13.82	11.49	1.30	76.18	80.83	78.51	1.30
10	9.54	14.24	12.34	1.14	75.76	80.46	77.66	1.14
11	11.09	14.72	13.31	0.96	75.28	78.91	76.69	0.96
12	12.66	14.88	14.20	0.69	75.12	77.34	75.80	0.69
ข้อสอบร่วมที่ตรวจให้คะแนนสองค่า กลุ่มแรก								
51	.63	8.29	2.78	2.24	81.71	89.37	87.22	2.24
52	1.74	11.35	5.09	2.84	78.65	88.26	84.91	2.84
53	2.18	11.89	6.71	2.78	78.11	87.82	83.29	2.78
54	4.14	13.99	9.02	2.82	76.01	85.86	80.98	2.82
55	5.21	14.09	9.95	2.56	75.91	84.79	80.05	2.56
56	6.51	14.35	10.37	1.78	75.65	83.49	79.63	1.78
57	8.12	14.59	12.29	1.87	75.41	81.88	77.71	1.87
ข้อสอบเฉพาะที่ตรวจให้คะแนนหลายค่า กลุ่มแรก								
73	.20	8.35	2.83	1.89	81.65	89.80	87.17	1.89
74	1.52	14.30	6.37	3.81	75.70	88.48	83.63	3.81
ข้อสอบร่วมที่ตรวจให้คะแนนหลายค่า กลุ่มแรก								
83	.92	14.99	9.06	4.67	75.01	89.08	80.94	4.67

ตาราง 26 ค่าสถิติของมุมที่จำลองขึ้นสำหรับแบบสอบกลุ่มสองที่มีโครงสร้างที่ซับซ้อน (CS)

แบบสอบย่อยกลุ่มสอง (เวกเตอร์ของข้อสอบในมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ( $\theta_1$ ) ตั้งแต่ 31 - 45 องศา)								
ข้อที่	มุม 31-45 องศา				มุม 90- $\theta_1$ องศา			
	Min	Max	Mean	SD	Min	Max	Mean	SD
ข้อสอบเฉพาะที่ตรวจให้คะแนนสองค่า กลุ่มสอง								
13	31.01	34.76	32.06	1.02	55.24	58.99	57.94	1.02
14	31.39	36.68	33.11	1.56	53.32	58.61	56.89	1.56
15	32.03	37.79	34.29	1.59	52.21	57.97	55.71	1.59
16	32.29	39.07	35.28	1.82	50.93	57.71	54.72	1.82
17	32.96	40.41	36.12	2.16	49.59	57.04	53.88	2.16
18	32.98	41.18	37.12	2.17	48.82	57.02	52.88	2.17
19	33.51	42.17	38.27	2.42	47.83	56.49	51.73	2.42
20	34.19	42.37	39.23	2.20	47.63	55.81	50.77	2.20
21	36.45	43.70	40.45	2.06	46.30	53.55	49.55	2.06
22	37.66	44.21	41.48	1.90	45.79	52.34	48.52	1.90
23	41.00	44.58	42.79	1.22	45.42	49.00	47.21	1.22
24	41.48	44.73	43.54	1.00	45.27	48.52	46.46	1.00
25	42.04	44.95	44.30	0.91	45.05	47.96	45.70	0.91
ข้อสอบร่วมที่ตรวจให้คะแนนสองค่า กลุ่มสอง								
58	31.81	41.34	34.58	2.44	48.66	58.19	55.42	2.44
59	33.41	42.11	37.11	2.15	47.89	56.59	52.89	2.15
60	33.97	43.20	38.80	2.31	46.80	56.03	51.20	2.31
61	36.03	44.11	40.67	2.26	45.89	53.97	49.33	2.26
ข้อสอบเฉพาะที่ตรวจให้คะแนนหลายค่า กลุ่มสอง								
75	31.33	41.25	34.39	2.75	48.75	58.67	53.43	2.75
76	32.05	42.31	36.03	2.55	47.69	57.95	55.97	2.55
77	34.09	44.93	39.38	3.20	45.07	55.91	50.02	3.20
ข้อสอบร่วมที่ตรวจให้คะแนนหลายค่า กลุ่มสอง								
84	31.24	44.23	37.33	4.17	45.77	58.76	52.67	4.17

ตาราง 27 ค่าสถิติของมุมที่จำลองขึ้นสำหรับแบบสอบกลุ่มสามที่มีโครงสร้างที่ซับซ้อน (CS)

แบบสอบย่อยกลุ่มสาม (เวกเตอร์ของข้อสอบในมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ( $\theta_1$ ) ตั้งแต่ 46 - 60 องศา)								
ข้อที่	มุม 46-60 องศา				มุม 90- $\theta_1$ องศา			
	Min	Max	Mean	SD	Min	Max	Mean	SD
ข้อสอบเฉพาะที่ตรวจให้คะแนนสองค่า กลุ่มสาม								
26	46.00	50.88	47.08	1.32	39.12	44.00	42.92	1.32
27	46.22	51.31	48.05	1.38	38.69	43.78	41.95	1.38
28	47.28	51.78	49.22	1.46	38.22	42.72	40.78	1.46
29	47.61	53.12	49.85	1.62	36.88	42.39	40.15	1.62
30	48.40	53.91	50.86	1.52	36.09	41.60	39.14	1.52
31	48.93	55.25	51.99	1.63	34.75	41.07	38.01	1.63
32	49.87	56.13	52.84	1.90	33.87	40.13	37.16	1.90
33	50.84	57.32	53.98	1.84	32.68	39.16	36.02	1.84
34	52.68	58.34	54.87	1.75	31.66	37.32	35.13	1.75
35	52.98	58.90	55.40	1.80	31.10	37.02	34.60	1.80
36	53.26	58.95	56.53	1.56	31.05	36.74	33.47	1.56
37	56.10	59.29	57.80	1.02	30.71	33.90	32.20	1.02
38	56.37	59.90	58.60	1.02	30.10	33.63	31.40	1.02
ข้อสอบร่วมที่ตรวจให้คะแนนสองค่า กลุ่มสาม								
62	46.46	53.50	50.20	2.16	36.50	43.54	39.80	2.16
63	48.01	56.58	51.89	2.08	33.42	41.99	38.11	2.08
64	48.85	56.83	53.87	2.11	33.17	41.15	36.13	2.11
65	51.85	58.50	56.09	2.07	31.50	38.15	33.91	2.07
ข้อสอบเฉพาะที่ตรวจให้คะแนนหลายค่า กลุ่มสาม								
78	47.68	57.95	51.38	2.83	32.05	42.32	38.62	2.83
79	49.82	59.88	55.47	3.19	30.12	40.18	34.53	3.19
80	52.22	59.95	56.48	1.85	30.05	37.78	33.52	1.85
ข้อสอบร่วมที่ตรวจให้คะแนนหลายค่า กลุ่มสาม								
85	46.59	59.46	53.46	5.02	30.54	43.41	36.54	5.02

ตาราง 28 ค่าสถิติของมุมที่จำลองขึ้นสำหรับแบบสอบกลุ่มสี่ที่มีโครงสร้างที่ซับซ้อน (CS)

แบบสอบย่อยกลุ่มสี่ (เวกเตอร์ของข้อสอบในมิติที่ 1 ทำมุมกับแกนมิติที่ 1 ( $\theta_1$ ) ตั้งแต่ 76 - 90 องศา)								
ข้อที่	มุม 76-90 องศา				มุม 90- $\theta_1$ องศา			
	Min	Max	Mean	SD	Min	Max	Mean	SD
ข้อสอบเฉพาะที่ตรวจให้คะแนนสองค่า กลุ่มสี่								
39	76.00	78.51	77.20	0.89	11.49	14.00	12.80	0.89
40	76.64	80.75	78.16	1.12	9.25	13.36	11.84	1.12
41	77.08	81.20	79.22	1.36	8.80	12.92	10.78	1.36
42	77.53	83.19	80.44	1.52	6.81	12.47	9.56	1.52
43	78.15	85.08	81.85	1.96	4.92	11.85	8.15	1.96
44	78.41	85.71	82.93	1.65	4.29	11.59	7.07	1.65
45	81.14	87.41	84.04	1.63	2.59	8.86	5.96	1.63
46	82.42	88.20	84.93	1.62	1.80	7.58	5.07	1.62
47	83.13	88.40	85.70	1.55	1.60	6.87	4.30	1.55
48	84.03	89.36	86.70	1.62	.64	5.97	3.30	1.62
49	84.83	89.75	87.75	1.61	.25	5.17	2.25	1.61
50	86.69	89.97	88.78	1.22	.03	3.31	1.22	1.22
ข้อสอบร่วมที่ตรวจให้คะแนนสองค่า กลุ่มสี่								
66	76.05	82.86	78.64	2.10	7.14	13.95	11.36	2.10
67	76.56	86.49	81.03	2.74	3.51	13.44	8.97	2.74
68	77.65	87.11	83.85	3.42	2.89	12.35	6.15	3.42
69	79.71	88.63	85.53	3.30	1.40	10.29	4.47	3.30
70	79.87	89.14	86.87	2.51	.86	10.13	3.13	2.51
71	80.52	89.96	87.45	2.15	.04	9.48	2.55	2.15
72	82.22	89.98	87.64	2.05	.02	7.78	2.36	2.05
ข้อสอบเฉพาะที่ตรวจให้คะแนนหลายค่า กลุ่มสี่								
81	76.58	87.41	82.62	3.15	2.59	13.42	7.38	3.15
82	80.60	89.78	86.11	2.99	.22	9.40	3.89	2.99
ข้อสอบร่วมที่ตรวจให้คะแนนหลายค่า กลุ่มสี่								
86	76.58	89.98	81.81	4.34	.02	13.42	8.19	4.34



ภาคผนวก ค

ผลการตรวจสอบความถูกต้องของค่าความสามารถที่จำลองขึ้น

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

ตาราง 29 ค่าความสามารถของผู้สอบแต่ละกลุ่มที่จำลองขึ้นในแต่ละรอบ (20 รอบ)

รอบ	ความสามารถของผู้สอบกลุ่ม 1			ความสามารถของผู้สอบกลุ่ม 2		
	$\mu$	$\Sigma$	r	$\mu$	$\Sigma$	r
1	$\begin{bmatrix} -0.01 \\ 0.00 \end{bmatrix}$	$\begin{bmatrix} 1.01 & 0 \\ 0 & 0.98 \end{bmatrix}$	0.50	$\begin{bmatrix} 0.54 \\ 0.50 \end{bmatrix}$	$\begin{bmatrix} 1.04 & 0 \\ 0 & 1.01 \end{bmatrix}$	0.47
2	$\begin{bmatrix} -0.01 \\ -0.02 \end{bmatrix}$	$\begin{bmatrix} 0.98 & 0 \\ 0 & 0.97 \end{bmatrix}$	0.50	$\begin{bmatrix} 0.51 \\ 0.49 \end{bmatrix}$	$\begin{bmatrix} 1.02 & 0 \\ 0 & 1.00 \end{bmatrix}$	0.52
3	$\begin{bmatrix} -0.02 \\ -0.03 \end{bmatrix}$	$\begin{bmatrix} 1.02 & 0 \\ 0 & 1.02 \end{bmatrix}$	0.49	$\begin{bmatrix} 0.51 \\ 0.50 \end{bmatrix}$	$\begin{bmatrix} 0.98 & 0 \\ 0 & 1.00 \end{bmatrix}$	0.50
4	$\begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.98 & 0 \\ 0 & 1.02 \end{bmatrix}$	0.49	$\begin{bmatrix} 0.53 \\ 0.51 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0 \\ 0 & 0.99 \end{bmatrix}$	0.50
5	$\begin{bmatrix} -0.01 \\ -0.04 \end{bmatrix}$	$\begin{bmatrix} 0.99 & 0 \\ 0 & 1.01 \end{bmatrix}$	0.49	$\begin{bmatrix} 0.49 \\ 0.47 \end{bmatrix}$	$\begin{bmatrix} 0.99 & 0 \\ 0 & 0.98 \end{bmatrix}$	0.50
6	$\begin{bmatrix} -0.03 \\ 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.98 & 0 \\ 0 & 1.03 \end{bmatrix}$	0.50	$\begin{bmatrix} 0.53 \\ 0.51 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0 \\ 0 & 1.02 \end{bmatrix}$	0.52
7	$\begin{bmatrix} -0.02 \\ 0.04 \end{bmatrix}$	$\begin{bmatrix} 0.99 & 0 \\ 0 & 1.02 \end{bmatrix}$	0.50	$\begin{bmatrix} 0.50 \\ 0.48 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0 \\ 0 & 0.97 \end{bmatrix}$	0.50
8	$\begin{bmatrix} 0.01 \\ -0.01 \end{bmatrix}$	$\begin{bmatrix} 1.01 & 0 \\ 0 & 1.01 \end{bmatrix}$	0.50	$\begin{bmatrix} 0.51 \\ 0.53 \end{bmatrix}$	$\begin{bmatrix} 1.07 & 0 \\ 0 & 0.99 \end{bmatrix}$	0.50
9	$\begin{bmatrix} -0.01 \\ -0.01 \end{bmatrix}$	$\begin{bmatrix} 1.03 & 0 \\ 0 & 0.99 \end{bmatrix}$	0.48	$\begin{bmatrix} 0.48 \\ 0.50 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0 \\ 0 & 0.98 \end{bmatrix}$	0.50
10	$\begin{bmatrix} 0.01 \\ 0.05 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0 \\ 0 & 0.98 \end{bmatrix}$	0.50	$\begin{bmatrix} 0.52 \\ 0.51 \end{bmatrix}$	$\begin{bmatrix} 0.99 & 0 \\ 0 & 1.00 \end{bmatrix}$	0.49
11	$\begin{bmatrix} 0.02 \\ 0.00 \end{bmatrix}$	$\begin{bmatrix} 1.02 & 0 \\ 0 & 0.99 \end{bmatrix}$	0.52	$\begin{bmatrix} 0.45 \\ 0.47 \end{bmatrix}$	$\begin{bmatrix} 0.98 & 0 \\ 0 & 1.00 \end{bmatrix}$	0.50
12	$\begin{bmatrix} 0.00 \\ 0.01 \end{bmatrix}$	$\begin{bmatrix} 0.97 & 0 \\ 0 & 1.03 \end{bmatrix}$	0.50	$\begin{bmatrix} 0.49 \\ 0.48 \end{bmatrix}$	$\begin{bmatrix} 1.01 & 0 \\ 0 & 0.97 \end{bmatrix}$	0.48
13	$\begin{bmatrix} -0.03 \\ -0.02 \end{bmatrix}$	$\begin{bmatrix} 0.99 & 0 \\ 0 & 1.00 \end{bmatrix}$	0.50	$\begin{bmatrix} 0.50 \\ 0.51 \end{bmatrix}$	$\begin{bmatrix} 0.98 & 0 \\ 0 & 0.97 \end{bmatrix}$	0.50
14	$\begin{bmatrix} 0.00 \\ 0.00 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0 \\ 0 & 0.99 \end{bmatrix}$	0.45	$\begin{bmatrix} 0.49 \\ 0.50 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0 \\ 0 & 0.98 \end{bmatrix}$	0.49
15	$\begin{bmatrix} 0.02 \\ 0.03 \end{bmatrix}$	$\begin{bmatrix} 1.02 & 0 \\ 0 & 0.98 \end{bmatrix}$	0.51	$\begin{bmatrix} 0.48 \\ 0.48 \end{bmatrix}$	$\begin{bmatrix} 1.05 & 0 \\ 0 & 0.98 \end{bmatrix}$	0.50
16	$\begin{bmatrix} -0.03 \\ -0.02 \end{bmatrix}$	$\begin{bmatrix} 1.01 & 0 \\ 0 & 0.98 \end{bmatrix}$	0.50	$\begin{bmatrix} 0.51 \\ 0.49 \end{bmatrix}$	$\begin{bmatrix} 1.03 & 0 \\ 0 & 1.03 \end{bmatrix}$	0.48
17	$\begin{bmatrix} -0.01 \\ 0.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0 \\ 0 & 1.02 \end{bmatrix}$	0.51	$\begin{bmatrix} 0.51 \\ 0.49 \end{bmatrix}$	$\begin{bmatrix} 1.08 & 0 \\ 0 & 1.01 \end{bmatrix}$	0.53
18	$\begin{bmatrix} -0.03 \\ -0.03 \end{bmatrix}$	$\begin{bmatrix} 0.98 & 0 \\ 0 & 1.03 \end{bmatrix}$	0.51	$\begin{bmatrix} 0.52 \\ 0.54 \end{bmatrix}$	$\begin{bmatrix} 0.96 & 0 \\ 0 & 1.01 \end{bmatrix}$	0.48
19	$\begin{bmatrix} -0.01 \\ -0.03 \end{bmatrix}$	$\begin{bmatrix} 1.02 & 0 \\ 0 & 1.05 \end{bmatrix}$	0.53	$\begin{bmatrix} 0.49 \\ 0.51 \end{bmatrix}$	$\begin{bmatrix} 0.98 & 0 \\ 0 & 1.01 \end{bmatrix}$	0.50
20	$\begin{bmatrix} 0.01 \\ -0.02 \end{bmatrix}$	$\begin{bmatrix} 1.05 & 0 \\ 0 & 1.03 \end{bmatrix}$	0.48	$\begin{bmatrix} 0.48 \\ 0.50 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0 \\ 0 & 1.00 \end{bmatrix}$	0.50

ตาราง 30 ค่าสถิติบรรยายของความสามารถของผู้สอบแต่ละกลุ่มที่จำลองขึ้นจำนวน 20 รอบ

รอบ	$\theta_1$	ค่าสถิติบรรยายของความสามารถของผู้สอบกลุ่มแรก						ค่าสถิติบรรยายของความสามารถของผู้สอบกลุ่มสอง					
		Min	Max	Mean	SD	S <sup>2</sup>	r	Min	Max	Mean	SD	S <sup>2</sup>	r
1	$\theta_1$	-3.32	3.55	-0.03	1.02	1.04	0.51	-4.80	4.06	0.53	0.99	0.99	0.50
	$\theta_2$	-4.04	3.28	0.02	1.01	1.01		-2.43	4.01	0.53	1.02	1.03	
2	$\theta_1$	-4.63	3.23	0.00	1.02	1.04	0.50	-2.57	4.11	0.52	1.01	1.02	0.50
	$\theta_2$	-4.09	3.54	-0.04	1.00	1.00		-3.50	4.03	0.51	1.00	0.99	
3	$\theta_1$	-3.28	3.59	0.01	0.98	0.97	0.49	-2.92	4.60	0.49	1.00	1.01	0.50
	$\theta_2$	-4.01	3.48	-0.01	1.02	1.04		-3.22	4.19	0.48	1.01	1.02	
4	$\theta_1$	-3.47	2.90	-0.01	0.98	0.96	0.48	-2.65	3.80	0.49	0.99	0.98	0.49
	$\theta_2$	-3.87	2.99	0.01	0.97	0.95		-3.22	4.2	0.48	0.99	0.97	
5	$\theta_1$	-3.80	3.67	0.00	0.99	0.99	0.47	-3.02	4.49	0.47	1.00	1.00	0.50
	$\theta_2$	-3.55	3.99	0.01	1.02	1.03		-3.95	4.51	0.48	1.01	1.02	
6	$\theta_1$	-3.55	3.64	-0.01	0.99	0.98		-3.24	4.08	0.49	1.00	1.00	0.49
	$\theta_2$	-3.70	3.32	0.01	0.98	0.97		-2.98	4.13	0.50	1.01	1.02	
7	$\theta_1$	-2.98	3.63	-0.02	0.98	0.97		-3.09	4.42	0.49	1.01	1.01	0.47
	$\theta_2$	-3.27	3.85	-0.02	1.00	1.00		-3.65	3.93	0.50	0.99	0.98	
8	$\theta_1$	-3.53	3.36	0.00	1.01	1.02		-3.64	4.21	0.51	1.01	1.02	0.49
	$\theta_2$	-3.08	3.44	0.00	0.98	0.97		-2.81	4.38	0.52	1.03	1.06	
9	$\theta_1$	-4.20	3.17	0.01	1.00	1.00		-2.99	3.65	0.51	1.01	1.02	0.50
	$\theta_2$	-3.50	3.66	-0.01	0.98	0.96		-2.84	4.08	0.52	1.00	1.00	
10	$\theta_1$	-3.01	3.78	-0.03	0.99	0.99		-2.68	3.99	0.55	0.99		0.47
	$\theta_2$	-3.24	3.32	-0.01	1.01	1.02		-3.48	4.14	0.58	1.01		
11	$\theta_1$	-3.25	3.58	0.01	1.01	1.01		-2.49	4.15	0.51	0.99	0.99	0.54
	$\theta_2$	-3.55	3.89	0.01	1.01	1.02		-3.25	3.97	0.50	1.00	1.00	
12	$\theta_1$	-3.85	3.73	-0.03	1.02			-2.98	3.97	0.49	1.01	1.03	0.50
	$\theta_2$	-3.49	3.46	-0.02	1.01	1.02		-2.89	3.91	0.51	0.99	0.99	
13	$\theta_1$	-3.57	3.58	0.00	0.99	0.98		-3.03	4.04	0.50	1.02		0.50
	$\theta_2$	-3.18	3.54	0.01	0.97			-2.94	3.59	0.53	1.00	1.00	
14	$\theta_1$	-4.31	3.40	0.01	1.00	1.00		-3.02	3.85	0.49	1.01	1.02	0.51
	$\theta_2$	-3.55	3.49	0.02	1.01	1.02		-3.05	3.73	0.49	1.01	1.02	
15	$\theta_1$	-3.27	3.29	0.01	1.01	1.01		-2.82	4.45	0.51	1.00	1.01	0.50
	$\theta_2$	-3.28	3.17	-0.01	0.99	0.99		-3.52	3.82	0.52	1.02		
16	$\theta_1$	-3.38	4.00	0.00	1.01	1.02		-3.20	3.64	0.48	1.01	1.01	0.50
	$\theta_2$	-3.11	3.46	-0.01	1.02	1.03		-3.03	4.04	0.48	0.99		
17	$\theta_1$	-3.85	3.54	-0.01	1.00	1.00		-2.64	4.10	0.49	0.98	0.96	0.51
	$\theta_2$	-3.13	3.30	-0.01	0.98	0.97		-3.87	4.74	0.51	1.00	1.00	
18	$\theta_1$	-4.03	3.45	0.00	0.99	0.99		-2.85	4.23	0.49	0.99	0.97	0.49
	$\theta_2$	-3.12	3.29	0.00	0.98	0.97		-3.05	3.49	0.47	0.98	0.96	
19	$\theta_1$	-3.58	3.42	-0.03	1.01	1.02		-3.25	4.41	0.49	1.00	1.00	0.51
	$\theta_2$	-3.20	3.47	0.01	1.01	1.02		-2.81	4.00	0.50	0.98	0.99	
20	$\theta_1$	-3.55	3.61	-0.01	0.99	0.98		-2.94	4.83	0.48	1.01	1.03	0.49
	$\theta_2$	-4.46	2.88	-0.03	0.98	0.97		-3.48	3.97	0.52	1.03	1.05	

ภาคผนวก ง  
ผลการประเมินประสิทธิภาพโดยรวมของการเชื่อมโยงสากล

จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY



### ผลการประเมินประสิทธิภาพของการเชื่อมโยงสเกลด้วยวิธี TCF

การเชื่อมโยงสเกลด้วยวิธี TCF ต้องมีการกำหนดเมทริกซ์การหมุน (rotation matrix) ขนาด  $2 \times 2$  และเมทริกซ์การเลื่อนตำแหน่ง (translation matrix) ขนาด  $2 \times 1$  โดยใช้แพ็คเกจ plink ในโปรแกรม R ตามแนวคิดของ Oshima, Davey and Lee (2000, cited in Zhang, 2012) ที่มีการหมุนแบบไม่ตั้งฉาก (non-orthogonal) ซึ่งค่าเฉลี่ยของเมทริกซ์ดังกล่าวในแต่ละเงื่อนไข โดยเฉลี่ยจากการกระทำซ้ำ 40 รอบ ดังตาราง 31 หลังจากนั้นจึงนำเมทริกซ์ที่ได้มาใช้ในการแปลงค่าพารามิเตอร์ ผ่านโปรแกรม R ด้วยการประยุกต์ใช้ R codes จากงานวิจัยของ Simon (2008) ในขั้นตอนนี้ผู้วิจัยประเมินประสิทธิภาพโดยรวม (ค่า RMSE) จากการเชื่อมโยงสเกลสำหรับค่าพารามิเตอร์ต่าง ๆ ทั้ง 22 เงื่อนไข ดังตาราง 31

เมื่อพิจารณาประสิทธิภาพโดยรวมในเบื้องต้น พบว่า การเชื่อมโยงสเกลข้อสอบที่ตรวจให้คะแนนหลายค่ามีประสิทธิภาพโดยรวมมากกว่าข้อสอบที่ตรวจให้คะแนนสองค่า **เมื่อพิจารณาประสิทธิภาพโดยรวมของข้อสอบที่ตรวจให้คะแนนสองค่า สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า ค่า RMSE มีค่าตั้งแต่ 2.444 - .312 ส่วนโครงสร้างที่ซับซ้อน (CS) พบว่า ค่า RMSE มีค่าตั้งแต่ 3.433 - .364** **เมื่อพิจารณาประสิทธิภาพโดยรวมของข้อสอบที่ตรวจให้คะแนนหลายค่า สำหรับโครงสร้างที่ไม่ซับซ้อน (APSS) พบว่า ค่า RMSE มีค่าตั้งแต่ 1.918 - .056 ส่วนโครงสร้างที่ซับซ้อน (CS) พบว่า ค่า RMSE มีค่าตั้งแต่ 1.866 - .050** โดยค่า RMSE ของพารามิเตอร์  $d$  มากกว่าพารามิเตอร์  $a_1$  และ  $a_2$  ทุกเงื่อนไข โดยค่า RMSE ที่ได้จากการวิจัยครั้งนี้อยู่ในช่วงของค่า RMSE ที่ได้จากงานวิจัยเกี่ยวกับการเชื่อมโยงสเกลด้วยโมเดล MIRT ที่ผ่านมา อย่างเช่น งานวิจัยของ Yao (2008) ที่มีค่า RMSE อยู่ในช่วง 1.458 - .005 ส่วนงานวิจัยของศิวัชต์ศน์ สุขสุวรรณ (2555) มีค่า RMSE อยู่ในช่วง 1.712 - .163 และงานวิจัยของพัชรี จันทร์เพ็ง (2555) มีค่า RMSE อยู่ในช่วง 2.022 - .075

นอกจากนี้จะพบว่า การเชื่อมโยงสเกลข้อสอบที่ตรวจให้คะแนนหลายค่ามีประสิทธิภาพโดยรวมมากกว่าข้อสอบที่ตรวจให้คะแนนสองค่า โดยประสิทธิภาพโดยรวมสำหรับข้อสอบทั้ง 2 แบบมากขึ้น เมื่อสัดส่วนคะแนนข้อสอบรวมเพิ่มขึ้น นอกจากนี้ยังพบว่า ข้อสอบที่ตรวจให้คะแนนสองค่าสำหรับโครงสร้าง APSS มีประสิทธิภาพโดยรวมมากกว่าโครงสร้าง CS แต่ข้อสอบที่ตรวจให้คะแนนหลายค่าสำหรับโครงสร้าง CS มีประสิทธิภาพโดยรวมมากกว่าโครงสร้าง APSS ดังนั้นวิธี TCF จึงมีประสิทธิภาพในการเชื่อมโยงสเกลแบบสอบพหุมี

ตาราง 31 ค่าสัมประสิทธิ์การแปลงสเกลเฉลี่ย (transformation coefficient) ในเงื่อนไขต่าง ๆ

ANCHOR	DICHO	โครงสร้าง APSS		โครงสร้าง CS	
		ค่า A	ค่า B	ค่า A	ค่า B
10%	40:60	[1.030    -.491] [-.084    1.558]	[-.656] [.022]	[1.302    -.203] [-.246    .925]	[-.864] [.916]
	60:40	[1.604    .475] [-.283    .953]	[-.492] [-.037]	[1.316    -.674] [-.332    1.375]	[-.111] [.420]
	70:30	[1.112    -.119] [-.2011    1.111]	[-.833] [.572]	[1.356    .148] [-.711    .707]	[-.154] [.901]
20%	40:60	[1.352    .035] [-.267    .967]	[-.621] [.385]	[1.433    -.247] [-.183    1.477]	[-.243] [.188]
	50:50	[1.236    .146] [-.200    .801]	[-.528] [.802]	[1.54    -.390] [-.304    1.350]	[-.097] [.469]
	60:40	[1.177    .084] [-.334    .747]	[-.292] [.575]	[1.347    -.010] [-.301    0.961]	[-.081] [.539]
	70:30	[1.187    .797] [-.198    .026]	[-.152] [.258]	[1.110    -.196] [-.235    1.038]	[-.582] [.579]
30%	40:60	[1.101    .030] [-.107    .945]	[-.408] [-.032]	[1.223    .262] [-.343    .671]	[-.748] [.456]
	50:50	[1.095    .015] [-.065    .924]	[-.116] [-.263]	[1.223    -.222] [-.099    1.141]	[-.549] [.170]
	60:40	[1.123    .073] [-.102    .920]	[-.023] [.332]	[1.017    -.040] [-.096    .931]	[-.299] [.634]
	70:30	[1.334    .196] [-.222    .852]	[-.192] [.639]	[1.321    .069] [-.298    .841]	[-.158] [.483]

ตาราง 32 ค่า RMSE เฉลี่ยของค่าพารามิเตอร์ที่ได้จากการเชื่อมโยงสเกลในเงื่อนไขต่าง ๆ

ANCHOR	DICHOTOMOUS ANCHOR	โครงสร้าง APSS			โครงสร้าง CS				
		$a_1$	$a_2$	$d$	$a_1$	$a_2$	$d$		
<b>DICHOTOMOUS</b>									
10%	40:60	.945	.905	<b>2.444</b>	1.319	1.488	2.813		
	60:40	.885	.810	2.082	1.228	1.273	<b>3.433</b>		
	70:30	.635	.608	1.712	.817	.891	2.288		
20%	40:60	.619	.605	1.199	.818	.889	1.417		
	50:50	.547	.506	1.174	.569	.615	1.184		
	60:40	.456	.487	1.264	.591	.569	1.458		
	70:30	.391	.403	1.261	.392	.406	1.050		
30%	40:60	.432	.412	.699	.505	.534	.803		
	50:50	.422	.414	.958	.403	.383	.913		
	60:40	.313	<b>.312</b>	1.106	.401	.407	.973		
	70:30	.337	.323	1.000	<b>.364</b>	.381	.982		
<b>POLYTOMOUS</b>									
ANCHOR	DICHOTOMOUS ANCHOR	โครงสร้าง APSS				โครงสร้าง CS			
		$a_1$	$a_2$	$d_1$	$d_2$	$a_1$	$a_2$	$d_1$	$d_2$
10%	40:60	.281	.280	1.461	<b>1.918</b>	.235	.286	1.310	<b>1.866</b>
	60:40	.268	.232	1.132	1.611	.214	.230	1.207	1.529
	70:30	.191	.177	.992	1.415	.141	.148	.890	1.321
20%	40:60	.148	.150	1.042	1.422	.135	.149	.878	1.296
	50:50	.135	.120	.777	1.182	.093	.106	.648	1.010
	60:40	.116	.118	.752	1.106	.088	.087	.639	.876
	70:30	.090	.094	.651	.909	.053	.056	.531	.770
30%	40:60	.100	.102	.620	1.030	.099	.103	.730	1.019
	50:50	.095	.093	.642	.908	.058	.064	.586	.799
	60:40	<b>.056</b>	.060	.572	.800	.053	.053	.493	.716
	70:30	.063	.059	.487	.693	.048	.050	.447	.625

ภาคผนวก จ

ผลของประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบในภาพรวม



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY

ตาราง 33 ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบแต่ละระดับคะแนน

คะแนน	CMOSE			SMOSE		
	M	SD	VAR	M	SD	VAR
1	0.908	0.578	0.334	3.045	4.812	23.156
2	3.031	2.400	5.762	4.985	4.684	21.941
3	4.940	3.910	15.289	6.106	4.417	19.509
4	5.965	3.619	13.095	6.560	4.230	17.894
5	6.514	3.396	11.531	6.637	4.062	16.497
6	7.034	3.207	10.284	6.821	3.822	14.607
7	7.514	3.128	9.787	7.055	4.012	16.095
8	8.018	3.000	9.000	7.519	4.638	21.511
9	8.417	3.505	12.283	8.026	5.295	28.040
10	8.171	3.686	13.583	8.344	5.883	34.615
11	7.211	4.577	20.950	8.684	6.793	46.142
12	6.227	4.838	23.407	9.292	6.757	45.663
13	4.292	3.615	13.068	9.853	6.874	47.247
14	3.649	3.242	10.513	10.807	7.664	58.735
15	3.114	1.781	3.171	11.754	8.389	70.380
16	2.842	0.779	0.607	12.867	9.335	87.140
17	2.743	0.649	0.421	13.980	10.415	108.466
18	2.879	0.712	0.507	13.195	5.364	28.775
19	2.975	0.714	0.509	14.184	5.682	32.281
20	3.114	0.747	0.559	15.179	5.985	35.822
21	3.259	0.791	0.626	16.199	6.301	39.702
22	3.415	0.848	0.719	17.177	6.415	41.148
23	3.586	0.915	0.836	18.187	6.751	45.574
24	3.769	0.991	0.983	19.117	6.998	48.973
25	3.959	1.074	1.153	19.976	7.040	49.564
26	4.164	1.167	1.362	20.839	7.140	50.978
27	4.382	1.266	1.604	21.671	7.194	51.753
28	4.608	1.368	1.873	22.473	7.281	53.008
29	4.843	1.477	2.182	23.236	7.395	54.684
30	5.084	1.584	2.508	23.948	7.538	56.817
31	5.333	1.688	2.849	24.616	7.727	59.713
32	5.586	1.789	3.200	25.171	7.898	62.377
33	5.840	1.886	3.556	25.536	7.918	62.701
34	6.097	1.979	3.915	25.824	7.896	62.346
35	6.353	2.066	4.269	26.086	7.933	62.928
36	6.608	2.147	4.611	26.296	7.979	63.667

ตาราง 33 (ต่อ)

คะแนน	CMOSE			SMOSE		
	M	SD	VAR	M	SD	VAR
37	6.857	2.222	4.939	26.463	8.043	64.694
38	7.101	2.292	5.251	26.583	8.137	66.211
39	7.341	2.356	5.552	26.665	8.263	68.279
40	7.575	2.417	5.844	26.703	8.434	71.140
41	7.800	2.475	6.128	26.682	8.644	74.726
42	8.016	2.531	6.407	26.611	8.894	79.111
43	8.223	2.586	6.686	26.488	9.188	84.417
44	8.421	2.640	6.970	26.314	9.529	90.810
45	8.610	2.694	7.259	26.086	9.901	98.027
46	8.788	2.749	7.556	25.821	10.309	106.275
47	8.957	2.804	7.865	25.519	10.743	115.422
48	9.115	2.861	8.186	25.193	11.205	125.543
49	9.262	2.919	8.521	24.848	11.673	136.253
50	9.399	2.978	8.870	24.495	12.133	147.207
51	9.526	3.038	9.229	24.138	12.568	157.951
52	9.642	3.099	9.601	23.783	12.955	167.822
53	9.748	3.160	9.985	23.427	13.273	176.167
54	9.843	3.222	10.378	23.075	13.528	183.009
55	9.929	3.283	10.779	22.723	13.699	187.653
56	10.004	3.345	11.187	22.376	13.802	190.506
57	10.070	3.405	11.597	22.040	13.832	191.321
58	10.126	3.465	12.009	21.732	13.791	190.184
59	10.173	3.524	12.420	21.474	13.683	187.224
60	10.211	3.582	12.830	21.293	13.504	182.361
61	10.241	3.638	13.236	21.204	13.243	175.372
62	10.262	3.693	13.638	21.236	12.922	166.977
63	10.275	3.745	14.026	21.386	12.518	156.709
64	10.279	3.795	14.406	21.650	12.042	145.022
65	10.276	3.844	14.773	22.032	11.541	133.193
66	10.265	3.890	15.128	22.504	11.015	121.320
67	10.244	3.933	15.469	23.053	10.539	111.064
68	10.215	3.974	15.789	23.651	10.153	103.086
69	10.176	4.011	16.085	24.270	9.909	98.197
70	10.125	4.043	16.349	24.908	9.851	97.037
71	10.060	4.072	16.579	25.523	9.934	98.688

ตาราง 33 (ต่อ)

คะแนน	CMOSE			SMOSE		
	M	SD	VAR	M	SD	VAR
72	9.982	4.096	16.773	26.092	10.112	102.251
73	9.888	4.113	16.915	26.592	10.323	106.570
74	9.775	4.123	16.996	26.998	10.522	110.719
75	9.642	4.125	17.012	27.282	10.669	113.830
76	9.482	4.114	16.926	27.443	10.746	115.470
77	9.295	4.090	16.728	27.487	10.751	115.574
78	9.079	4.050	16.403	27.414	10.682	114.111
79	8.831	3.993	15.946	27.208	10.561	111.535
80	8.551	3.919	15.359	26.882	10.413	108.429
81	8.240	3.828	14.655	26.432	10.267	105.411
82	7.897	3.723	13.860	25.831	10.133	102.675
83	7.531	3.610	13.034	25.072	10.022	100.434
84	7.139	3.487	12.156	24.142	9.916	98.336
85	6.734	3.357	11.268	23.053	9.786	95.764
86	6.325	3.205	10.271	21.845	9.546	91.118
87	5.940	3.023	9.138	20.533	9.194	84.524
88	5.577	2.859	8.175	19.044	8.626	74.410
89	5.203	2.704	7.314	17.474	7.873	61.983
90	4.823	2.507	6.285	15.791	6.930	48.021
91	4.509	2.290	5.245	14.178	5.971	35.656
92	4.359	1.971	3.883	12.521	4.945	24.458
93	4.468	1.697	2.881	10.898	4.088	16.709
94	4.875	1.634	2.669	9.366	3.455	11.937
95	4.939	1.459	2.128	8.055	2.925	8.553
96	4.669	0.997	0.994	6.964	2.573	6.620
97	4.183	0.882	0.778	6.035	2.347	5.508
98	3.776	0.790	0.625	5.190	2.093	4.380
99	3.183	0.653	0.426	4.356	1.788	3.196
100	1.618	0.714	0.510	3.102	1.605	2.576

ภาคผนวก ฉ

ผลของประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบในแต่ละเงื่อนไข



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY



## ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบในแต่ละเงื่อนไข

ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบแต่ละระดับคะแนน วัดจากรากที่สองของความคลาดเคลื่อนกำลังสอง (RMSE<sub>k</sub>) โดยที่  $RMSE_k = \sqrt{\frac{\sum_{i=1}^{rep} (SE_{ik} - MCSE_k)^2}{rep-1}}$  เมื่อ  $k$  คือระดับคะแนนมีค่าเท่ากับ 1, 2, ..., 100 ซึ่งการนำเสนอประสิทธิภาพดังกล่าวแบ่งเป็น 2 ส่วน ได้แก่ (1) ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE และ (2) ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบ เมื่อสัดส่วนคะแนนข้อสอบรวมแตกต่างกัน

### 1. ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบระหว่างกระบวนการ CMOSE และกระบวนการ SMOSE

เมื่อพิจารณาประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบในแต่ละเงื่อนไข จากภาพ 1-4 ในภาคผนวก ฉ สำหรับโครงสร้าง APSS ในภาพรวม พบว่า ส่วนใหญ่การเบี่ยงเบนในการประมาณค่า SE สำหรับกระบวนการ CMOSE ต่ำกว่ากระบวนการ SMOSE ยกเว้น 3 กรณี ที่พบว่าการประมาณค่า SE สำหรับกระบวนการ CMOSE เบี่ยงเบนมากกว่ากระบวนการ SMOSE ได้แก่ (1) อัตราส่วนฯ 70:30 สัดส่วนคะแนนข้อสอบรวม 10% ประมาณช่วงคะแนน 25-85 (2) อัตราส่วนฯ 60:40 สัดส่วนคะแนนข้อสอบรวม ประมาณช่วงคะแนน 50-60 แต่ความแตกต่างดังกล่าวมีเพียงเล็กน้อย และ (3) อัตราส่วนฯ 50:50 สัดส่วนคะแนนข้อสอบรวม 10% ประมาณช่วงคะแนน 45-65

เมื่อพิจารณาโครงสร้าง CS ในภาพรวม พบว่า ส่วนใหญ่การเบี่ยงเบนในการประมาณค่า SE สำหรับกระบวนการ CMOSE ต่ำกว่ากระบวนการ SMOSE ยกเว้น 2 กรณี ที่พบว่าการประมาณค่า SE สำหรับกระบวนการ CMOSE เบี่ยงเบนมากกว่ากระบวนการ SMOSE ได้แก่ (1) อัตราส่วนฯ 70:30 สัดส่วนคะแนนข้อสอบรวม 10% ประมาณช่วงคะแนน 1-10 และ 75-100 และ (2) อัตราส่วนฯ 60:40 สัดส่วนคะแนนข้อสอบรวม 20% และ 10% ประมาณช่วงคะแนน 1-10

### 2. ประสิทธิภาพการประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ เมื่อสัดส่วนคะแนนข้อสอบรวมแตกต่างกัน

เมื่อพิจารณากระบวนการ CMOSE ในภาพรวม (ภาพ 9-10) พบว่า ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบ สำหรับสัดส่วนคะแนนข้อสอบรวมต่าง ๆ ใกล้เคียงกัน โดยส่วนใหญ่ ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบสำหรับสัดส่วนคะแนนข้อสอบรวม 10% ต่ำที่สุด ส่วนกระบวนการ SMOSE ในภาพรวม พบว่า การประมาณค่าความคลาดเคลื่อนมาตรฐานของคะแนนปรับเทียบ สำหรับสัดส่วนคะแนนข้อสอบรวมแต่ละแบบแตกต่างกัน โดยแนวโน้มของการเบี่ยงเบนในการประมาณค่า SE แตกต่างกันในแต่ละช่วงคะแนน

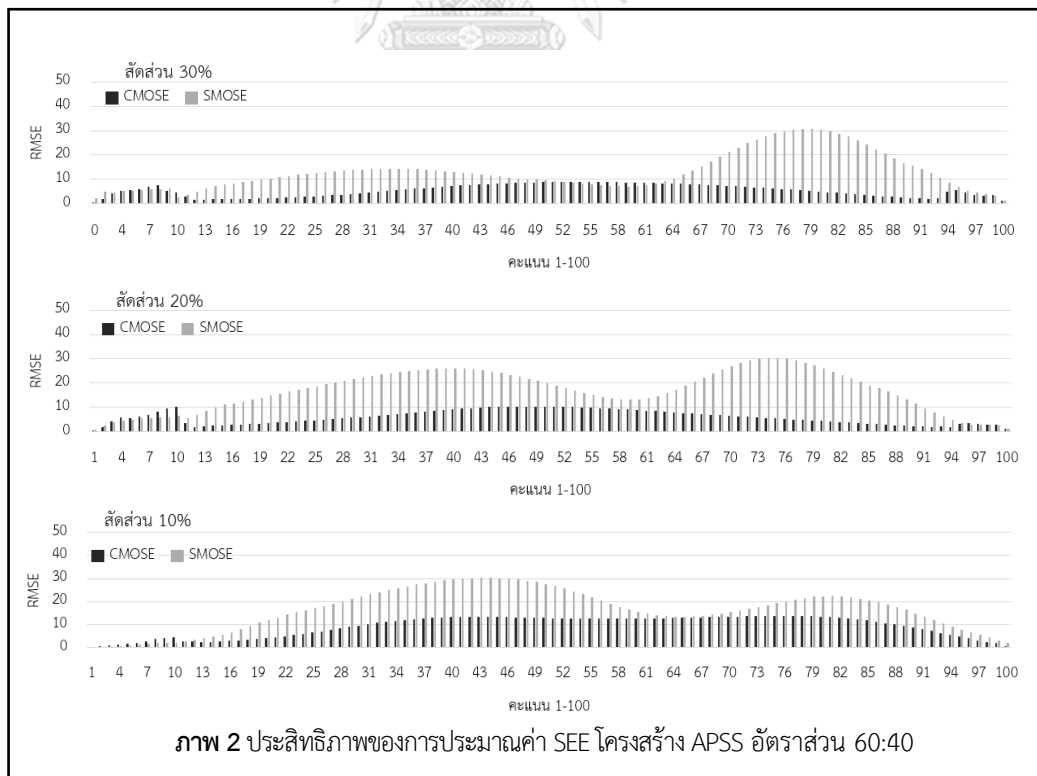
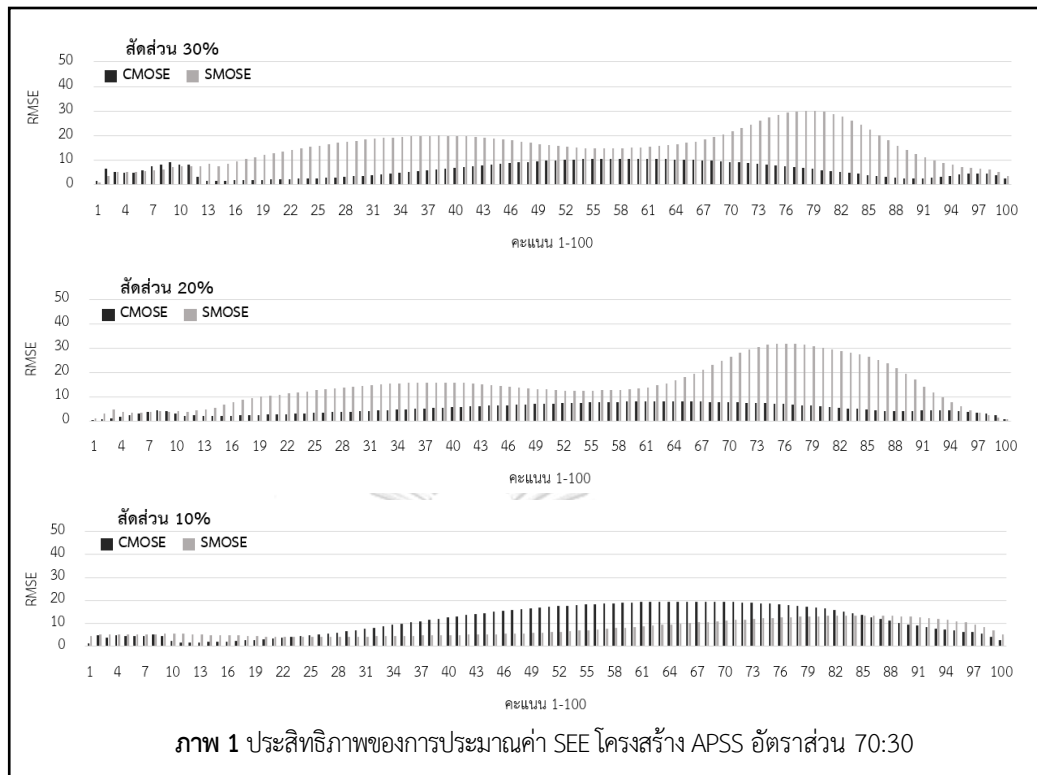


ส่วนช่วงคะแนน 61-80 พบว่า การเบี่ยงเบนในการประมาณค่า SE ต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20%, 10% และ 30% ตามลำดับ

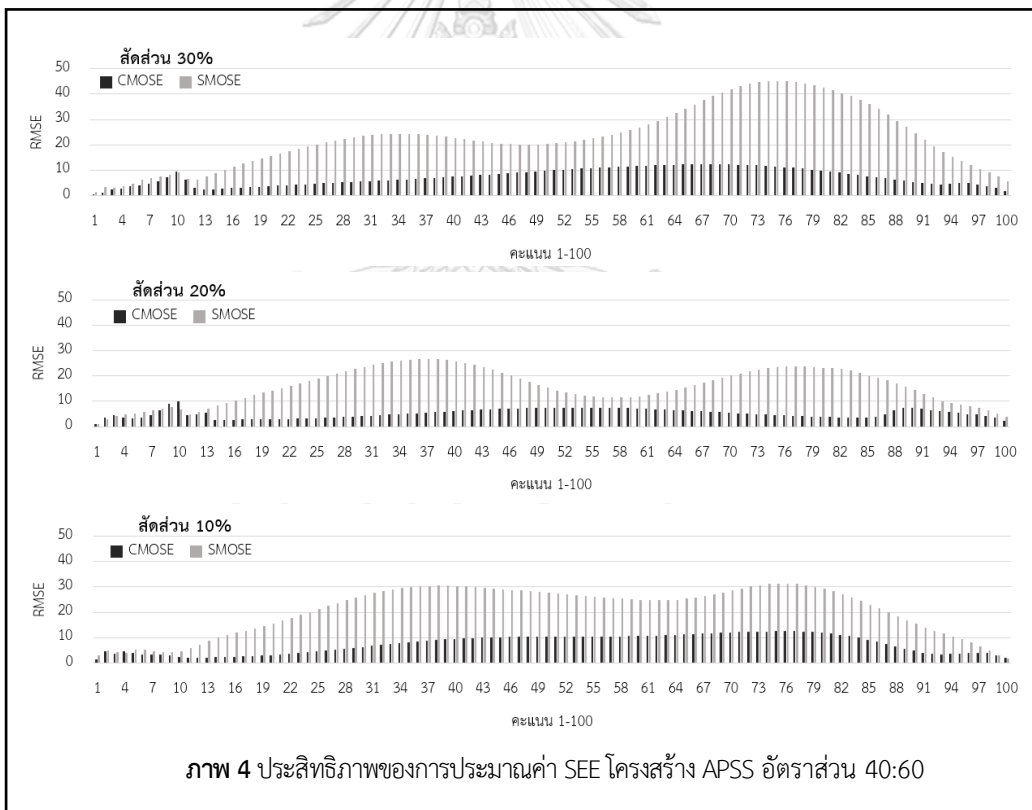
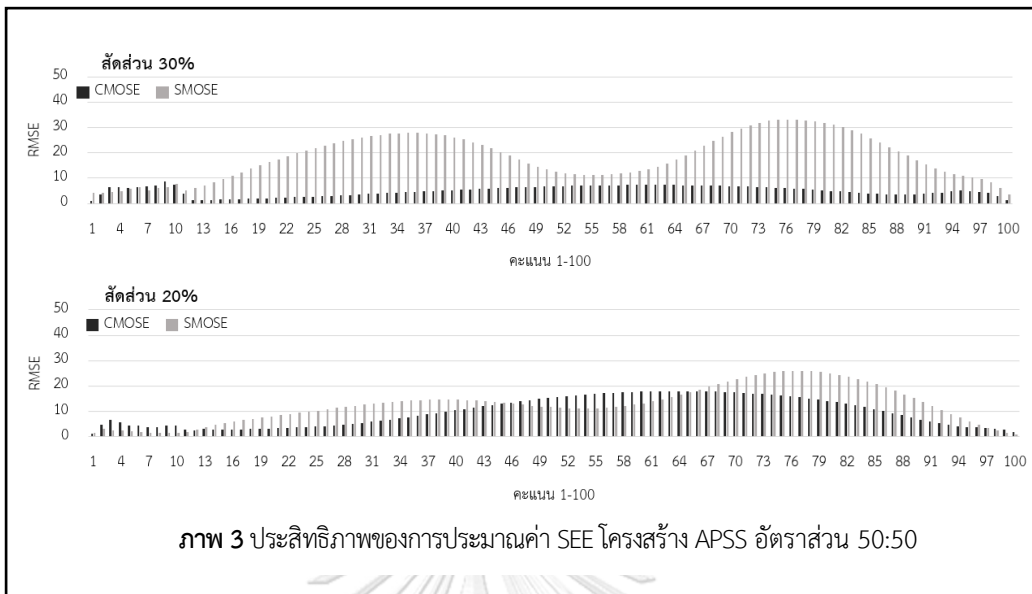
**เมื่อพิจารณากระบวนการ SMOSE โครงสร้าง CS สำหรับ (1) อัตราส่วนฯ 70:30** พบว่า ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบสำหรับสัดส่วนคะแนนข้อสอบรวมต่าง ๆ มีความแตกต่างกันในแต่ละช่วงคะแนน **(2) อัตราส่วนฯ 60:40** ช่วงคะแนน 20-60 พบว่า การเบี่ยงเบนในการประมาณค่า SE มากที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 10% ขณะที่ประสิทธิภาพในการประมาณค่า SE เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% และ 20% ใกล้เคียงกัน **(3) อัตราส่วนฯ 50:50** ช่วงคะแนน 20-45 พบว่า การเบี่ยงเบนในการประมาณค่า SE ต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 30% และ 20% ตามลำดับ ส่วนช่วงคะแนน 46-80 พบว่า การเบี่ยงเบนในการประมาณค่า SE ต่ำที่สุด เมื่อสัดส่วนคะแนนข้อสอบรวม คือ 20% และ 30% ตามลำดับ และ **(4) อัตราส่วนฯ 40:60** พบว่า แนวโน้มของประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบมีความแตกต่างกันในแต่ละช่วงคะแนน

จากผลการวิเคราะห์ข้างต้น สะท้อนให้เห็นว่า (1) ส่วนใหญ่ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบสำหรับกระบวนการ CMOSE ต่ำกว่ากระบวนการ SMOSE (2) ส่วนใหญ่ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบสำหรับกระบวนการ CMOSE ต่ำที่สุดเมื่อสัดส่วนคะแนนข้อสอบรวม 10% อย่างไรก็ตามประสิทธิภาพการประมาณค่า SE สำหรับสัดส่วนคะแนนข้อสอบรวมต่าง ๆ ใกล้เคียงกัน และ (3) ประสิทธิภาพการประมาณค่า SE ของคะแนนปรับเทียบสำหรับกระบวนการ SMOSE ในแต่ละสัดส่วนคะแนนข้อสอบรวมแต่ละแบบแตกต่างกันโดยแนวโน้มของประสิทธิภาพการประมาณค่า SE แตกต่างกันในแต่ละช่วงคะแนน

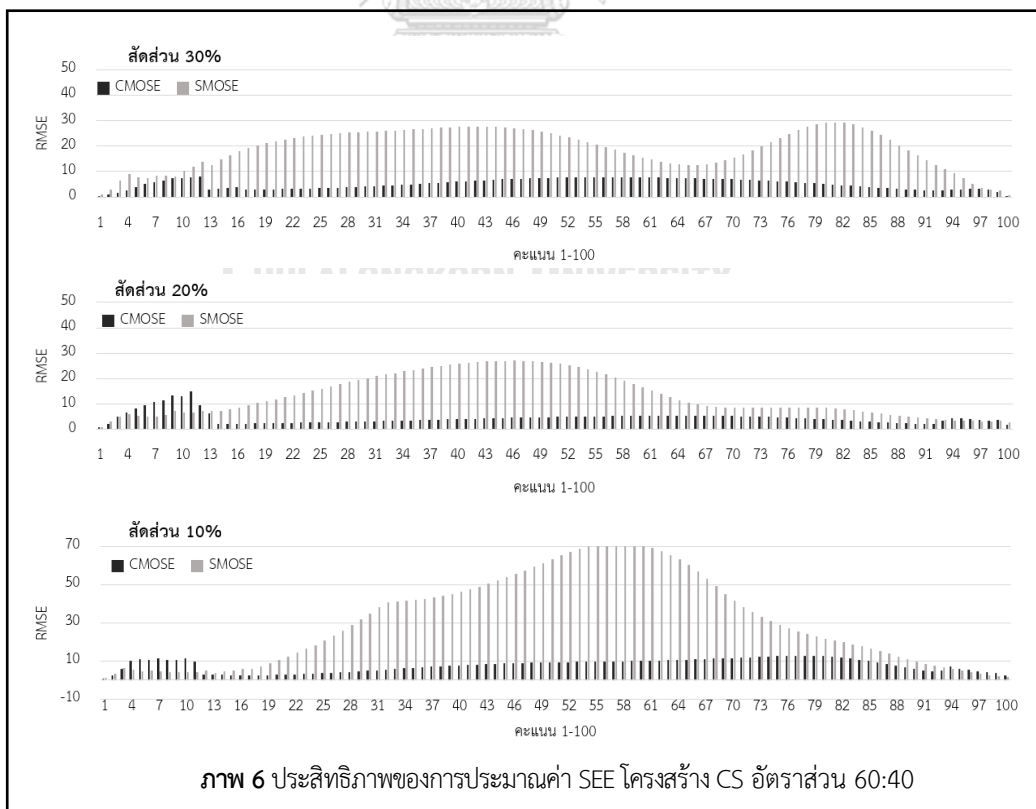
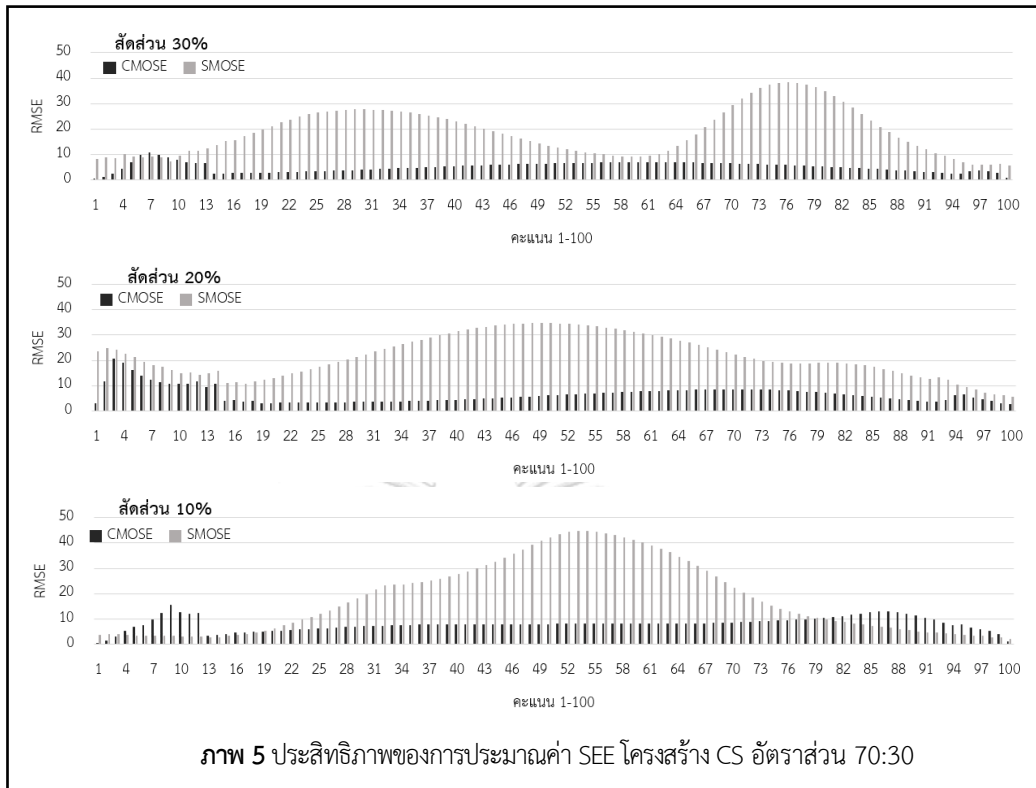
การเปรียบเทียบระหว่างกระบวนการ CMOSE และ SMOSE



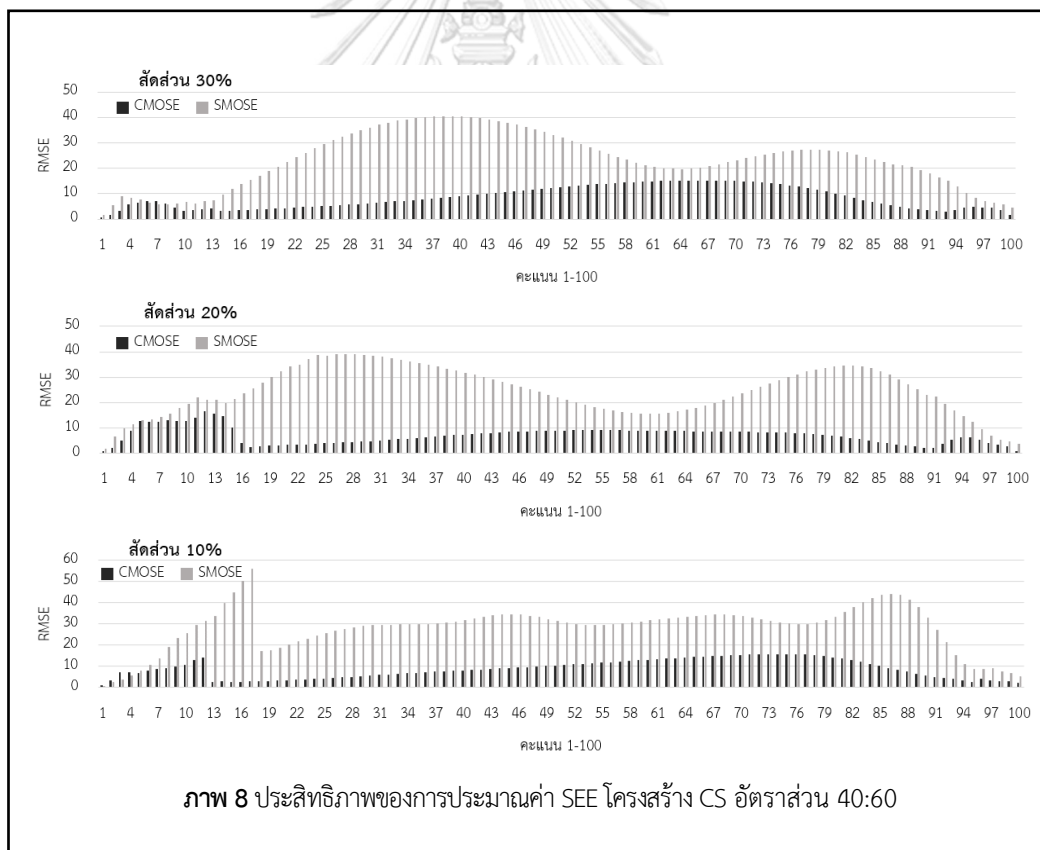
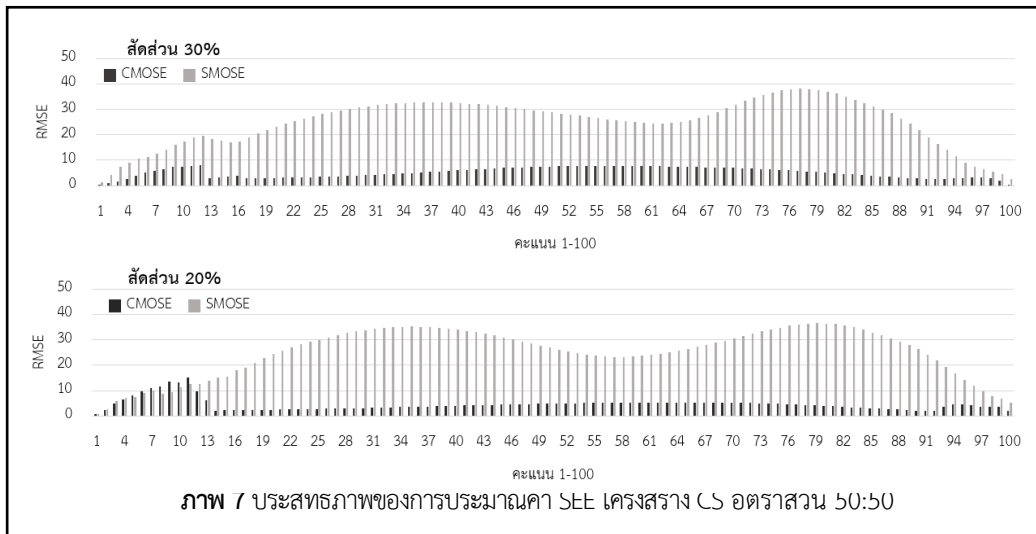
### การเปรียบเทียบระหว่างกระบวนการ CMOSE และ SMOSE



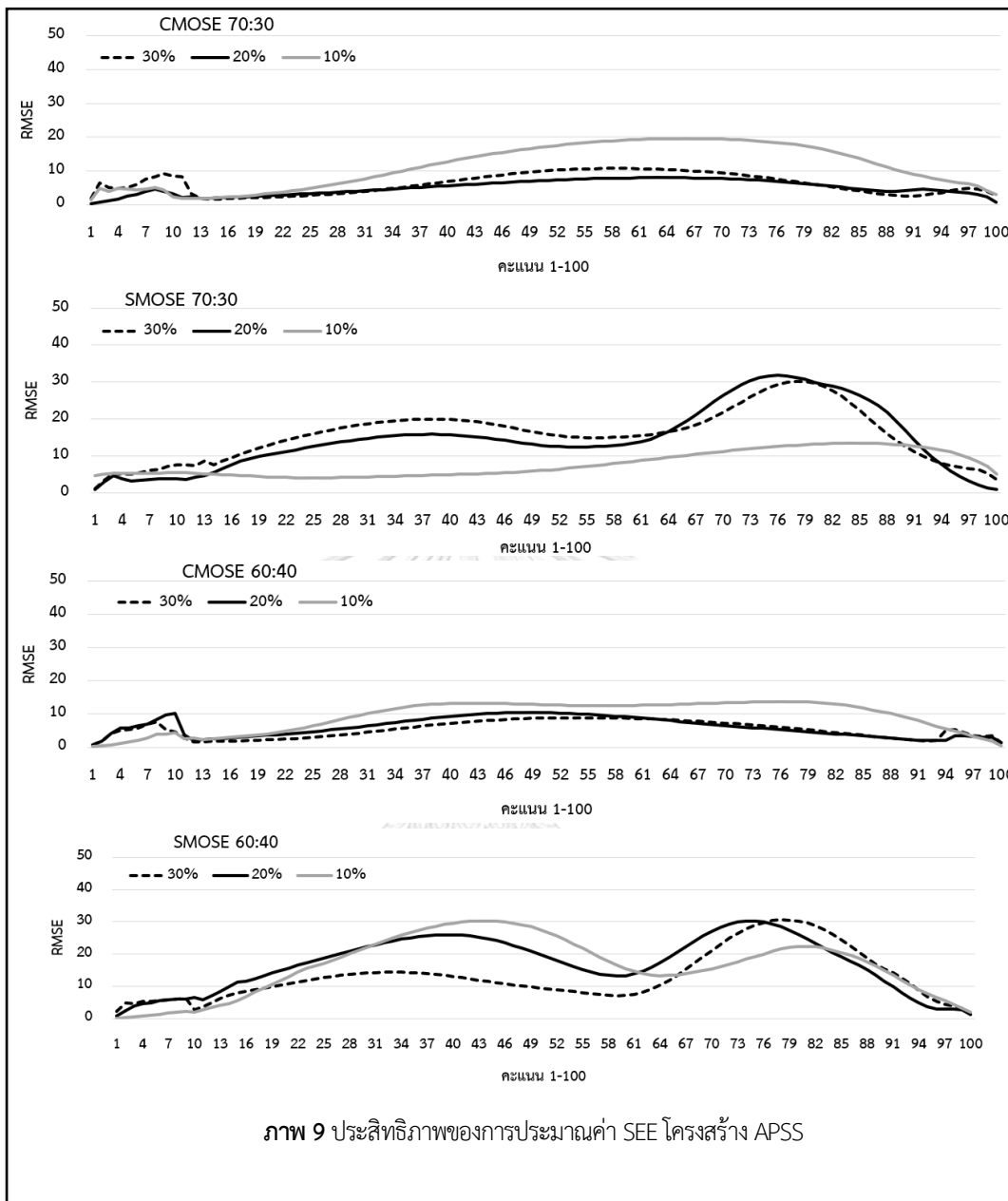
### การเปรียบเทียบระหว่างกระบวนการ CMOSE และ SMOSE



### การเปรียบเทียบระหว่างกระบวนการ CMOSE และ SMOSE

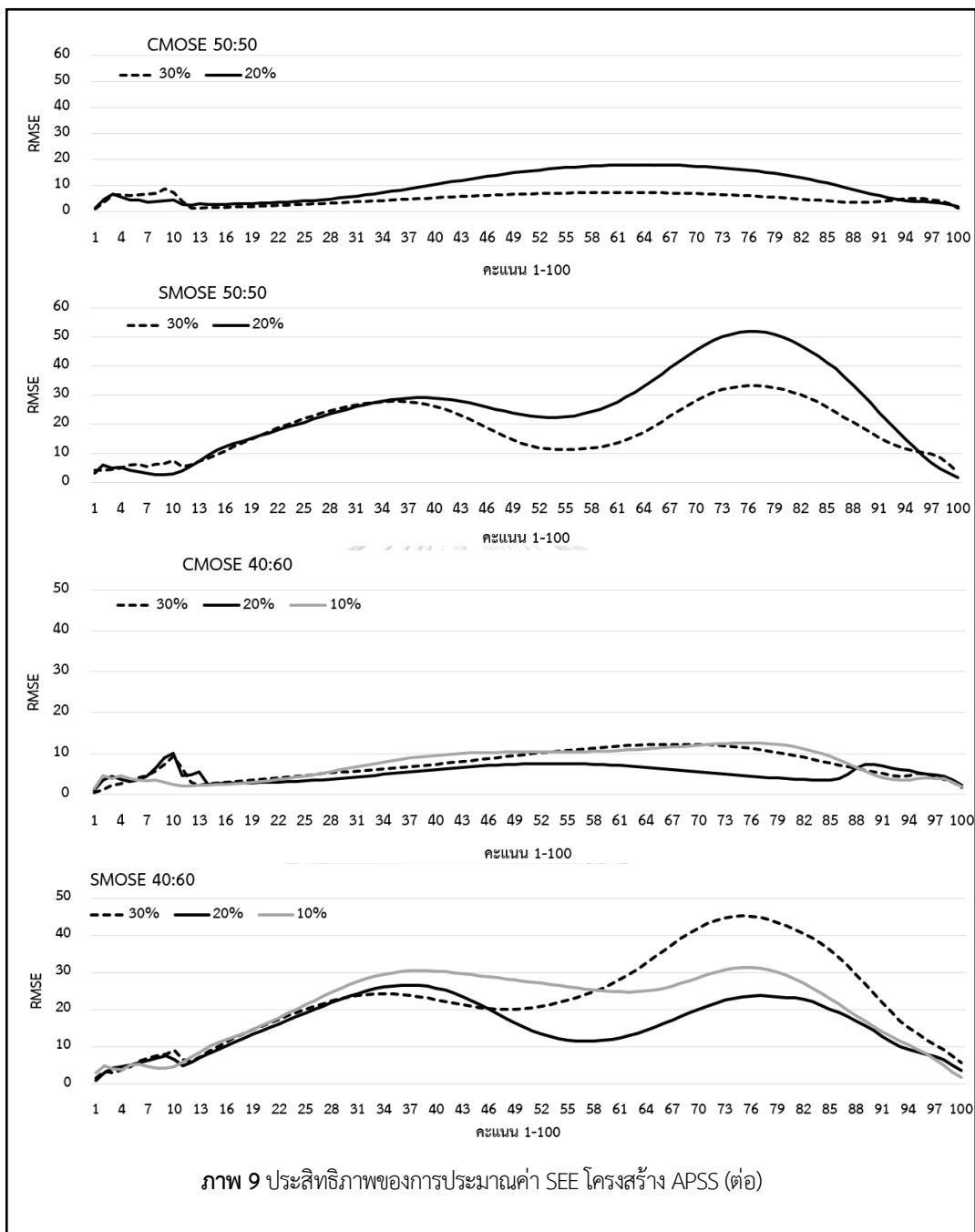


การเปรียบเทียบระหว่างสัดส่วนคะแนนข้อสอบรวม

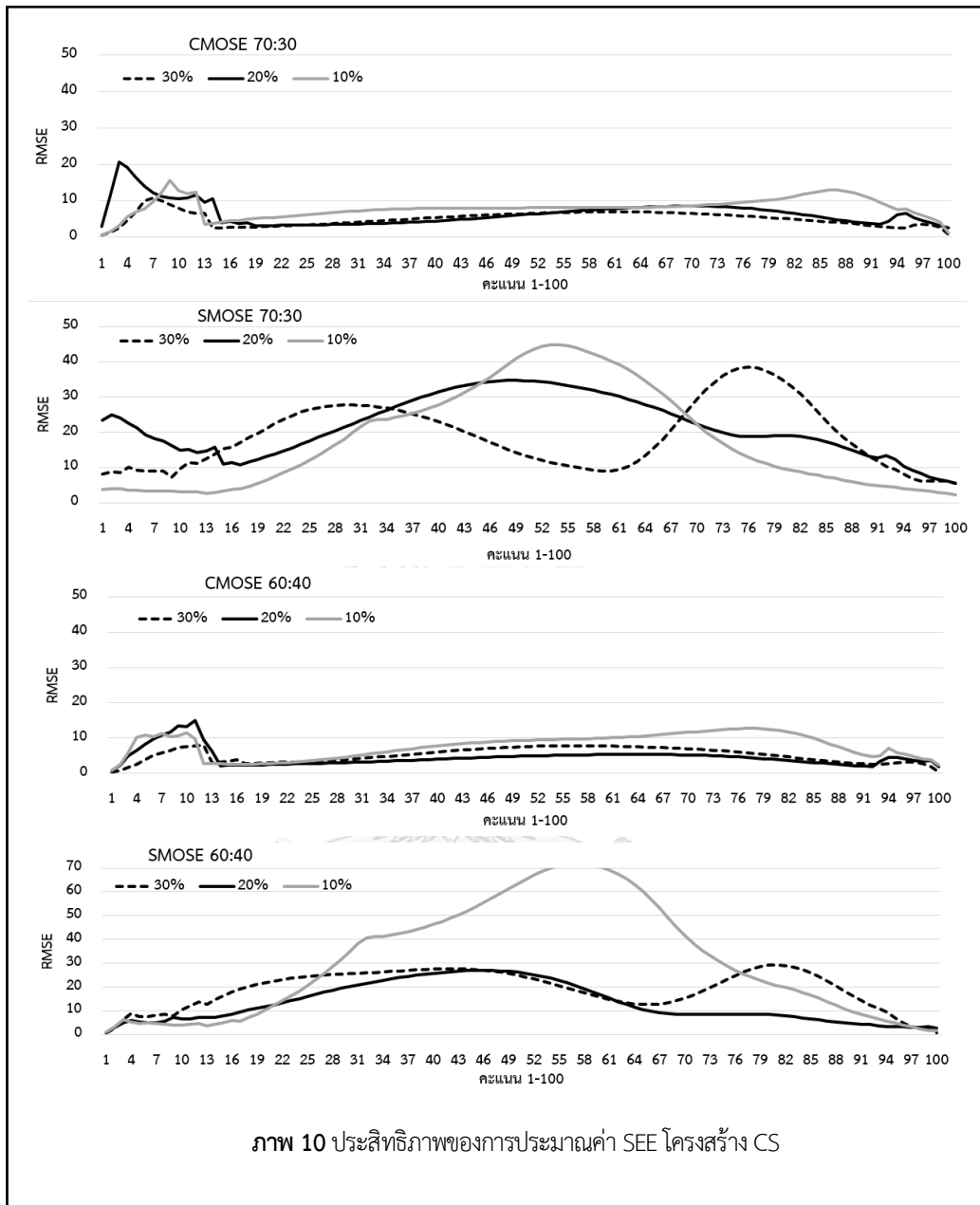




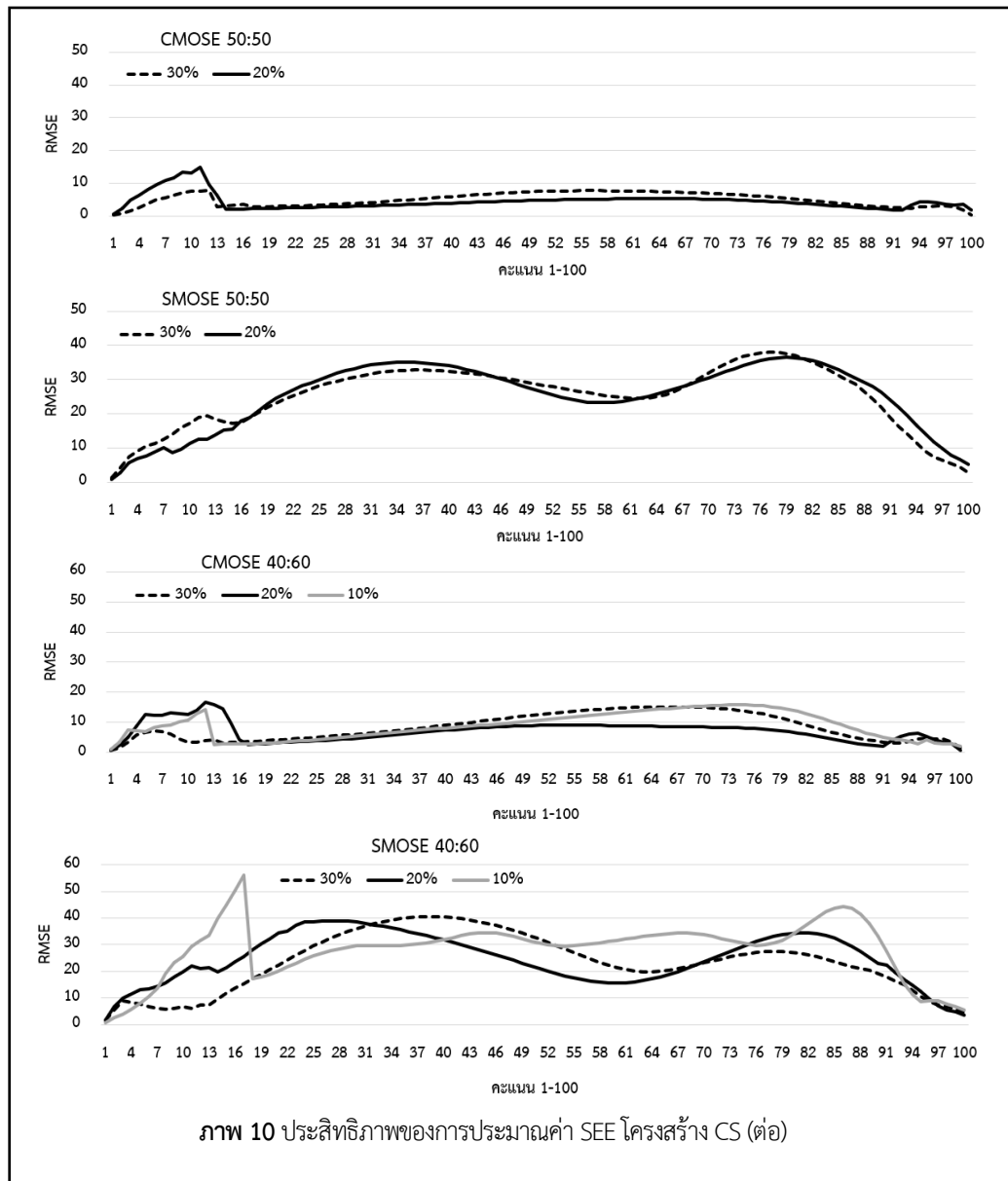
การเปรียบเทียบระหว่างสัดส่วนคะแนนข้อสอบรวม



### การเปรียบเทียบระหว่างสัดส่วนคะแนนข้อสอบรวม



### การเปรียบเทียบระหว่างสัดส่วนคะแนนข้อสอบรวม



### ประวัติผู้เขียนวิทยานิพนธ์

ชื่อ พนิดา พานิชวัฒน์ เกิดเมื่อวันที่ 22 มีนาคม พ.ศ. 2532 สำเร็จการศึกษาปริญญา  
ครุศาสตรบัณฑิต สาขามัธยมศึกษา-วิทยาศาสตร์ วิชาเอกฟิสิกส์ จุฬาลงกรณ์มหาวิทยาลัย ในปี  
การศึกษา 2555 และสำเร็จการศึกษาปริญญาครุศาสตรมหาบัณฑิต สาขาวิชาการวัดและ  
ประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์  
มหาวิทยาลัย ในปีการศึกษา 2557 และเข้าศึกษาต่อในหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต  
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยและจิตวิทยาการศึกษา คณะครุศาสตร์  
จุฬาลงกรณ์มหาวิทยาลัย



จุฬาลงกรณ์มหาวิทยาลัย  
CHULALONGKORN UNIVERSITY