การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสำหรับการประเมินพัฒนาการความสามารถของผู้สอบในแบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน

ปริญญานิพนธ์

ของ

สุนีย์ เงินยวง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
พฤษภาคม 2554

การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสำหรับการประเมินพัฒนาการความสามารถของผู้สอบในแบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน

ปริญญานิพนธ์

ของ

สุนีย์ เงินยวง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
พฤษภาคม 2554

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสำหรับการประเมินพัฒนาการความสามารถของผู้สอบในแบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน

บทคัดย่อ

ของ

สุนีย์ เงินยวง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
พฤษภาคม 2554

สุนีย์ เงินยวง. (2554). *การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสำหรับการประเมินพัฒนาการความสามารถของผู้สอบในแบบทดสอบผสมที่มีความยาว ความยากและการให้คะแนนแตกต่างกัน.* ปริญญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
คณะกรรมการควบคุม: รองศาสตราจารย์ ดร. องอาจ นัยพัฒน์,
รองศาสตราจารย์ ดร. ผจงจิต อินทสุวรรณ, รองศาสตราจารย์ ดร. สิริรัตน์ วิภาสศิลป์.

การศึกษาครั้งนี้มีจุดมุ่งหมายเพื่อตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสำหรับการประเมินพัฒนาการความสามารถของผู้สอบ ($\theta_2 - \theta_1$) ด้วยแบบทดสอบผสมที่ประกอบด้วยข้อสอบแบบเลือกตอบที่ตรวจให้คะแนนสองค่าและข้อสอบแบบเขียนตอบที่ตรวจให้คะแนนหลายค่า

การศึกษาครั้งนี้ใช้โมเดลโลจิสติกแบบสามพารามิเตอร์ (3PL) จำลองผลการตอบข้อสอบแบบเลือกตอบ และใช้โมเดลเจเนอรอลไรส์พาร์เชียลเครดิต (GPCM) จำลองผลการตอบข้อสอบแบบเขียนตอบของผู้สอบ 1,000 คนที่ทดสอบซ้ำสองครั้ง ตามตัวแปรที่ศึกษา 4 ตัว ได้แก่ ระดับพัฒนาการความสามารถ 9 ระดับ ความยาวของแบบทดสอบผสม 3 ขนาด ซึ่งกำหนดจากจำนวนข้อสอบแบบเลือกตอบต่อจำนวนข้อสอบแบบเขียนตอบ (30:10 24:8 และ 15:5) ความยากของแบบทดสอบผสมที่ใช้ในการทดสอบครั้งที่หนึ่ง 3 ระดับ และ การให้คะแนนของข้อสอบแบบเขียนตอบ 3 รูปแบบ รวมทั้งสิ้น 243 เงื่อนไข (9×3×3×3) การประเมินความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันพิจารณาจากค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าประมาณของพัฒนาการความสามารถ ($\hat{\theta}_2 - \hat{\theta}_1$)

ผลการศึกษาพบว่า

1. ค่าประมาณพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริงมีความสัมพันธ์ทางลบในระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

2. ในทุกเงื่อนไขที่ศึกษา เมื่อส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถของผู้สอบเป็น 0.80 และ 1.00 ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน มีค่าสูงกว่า เมื่อส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 อย่างไม่มีนัยสำคัญที่ระดับ .05 โดยพบว่า

2.1 เมื่อความยากและการให้คะแนนคงที่ แบบทดสอบผสมขนาด 24:8 มีความแม่นยำสูงกว่า แบบทดสอบผสมขนาด 15:5 และ ขนาด 30:10 อย่างมีนัยสำคัญที่ระดับ .05

2.2 เมื่อความยาวและการให้คะแนนคงที่ แบบทดสอบผสมที่ใช้ในการสอบครั้งที่หนึ่งที่มีระดับความยากเท่ากับระดับความสามารถของผู้สอบ (0.00) มีความแม่นยำต่ำกว่า แบบทดสอบผสม

ที่ใช้ในการสอบครั้งที่หนึ่งที่มีระดับความยากไม่เท่ากับระดับความสามารถของผู้สอบ (-0.50 และ 0.50) อย่างมีนัยสำคัญที่ระดับ .05

2.3 เมื่อความยาวและความยากคงที่ แบบทดสอบผสมที่ให้คะแนนข้อสอบแบบเขียนตอบ 3 ระดับ (0/1/2) มีความแม่นยำสูงกว่าแบบทดสอบผสมที่ให้คะแนนข้อสอบแบบเขียนตอบ 4 ระดับ (0/1/2/3) และ 5 ระดับ (0/1/2/3/4) อย่างมีนัยสำคัญที่ระดับ .05

AN ACCURACY INVESTIGATION OF CONCURRENT CALIBRATION FOR THE ASSESSMENT OF EXAMINEE'S GROWTH ABILITY IN MIXED-FORMAT TEST THAT HAS DIFFERENT TEST LENGTHS, ITEM DIFFICULTIES AND SCORING CATEGORIES

AN ABSTRACT

BY

SUNEE NGUENYUANG

Presented in Partial Fulfillment of the Requirements for the

Doctor of Education Degree in Testing and Measurement

at Srinakharinwirot University

May 2011

Sunee Nguenyuang. (2011). *An Accuracy Investigation of Concurrent Calibration for the Assessment of Examinee's Growth Ability in Mixed-Format Test that has Different Test Lengths, Item Difficulties, and Scoring Categories*. Dissertation, Ed.D. (Testing and Measurement). Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee: Assoc. Prof. Dr. Ong-Art Naiyapatana, Assoc. Prof. Dr. Pachongchit Intasuwan, Assoc. Prof. Dr. Sirirat Wipassilapa.

The main purpose of this study was to investigate the accuracy of concurrent calibration for the assessment of examinee's growth ability ($\theta_2 - \theta_1$) in mixed–format test consisting in terms of multiple choice (MC) items and constructed–response (CR) items where the MC was dichotomous response model and the CR was polytomous response model.

In order to fulfill this purpose, the 3PL/GPCM model combination was then used to simulate the item responses data of 1,000 examinees, in which four factors – growth ability, test lengths (the number of MC:CR, i.e. 30:10, 24:8, and 15:5), item difficulties and scoring categories of mixed–format test – were manipulated. In total, there were 243 conditions (9×3×3×3) with respect to the four variables. The accuracy of concurrent calibration was determined from the degrees of bias (BIAS) and the root mean square errors (RMSE) of the estimated growth ability ($\hat{\theta}_2 - \hat{\theta}_1$).

The results of the research were as follows:

1. Pearson's correlation coefficients between the estimated ability and the true ability were negatively high and statistically significant at the .01 level.

2. For all conditions, the accuracy of concurrent calibration when the standard deviations of growth ability were 0.80 and 1.00 was statistically insignificantly higher than when the standard deviation of growth ability was 1.2.

2.1 When the item difficulties and scoring categories were fixed, the accuracy of concurrent calibration of the 24:8 mixed–format test was statistically significantly higher, at the .05 level, than those of the 15:5 and the 30:10 mixed–format test.

2.2 When the test lengths and scoring categories were fixed, the accuracy of concurrent calibration of the 1st mixed–format test with the 0.00 difficulty was statistically

significantly higher, at the .05 level, than those of the 1st test mixed–format test with the -0.50 difficulty and the 0.50 difficulty.

2.3. When the test lengths and item difficulties were fixed, the accuracy of concurrent calibration of the mixed–format test with three–categories CR items (0/1/2) was statistically significantly higher, at the .05 level, than those of the mixed–format test with four–categories CR items (0/1/2/3) and with five–categories CR items (0/1/2/3/4).

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัย
จาก
มหาวิทยาลัยศรีนครินทรวิโรฒ
ประเภททุนงบประมาณแผ่นดิน ประจำปีงบประมาณ 2553

ปริญญานิพนธ์

เรื่อง

การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสำหรับการประเมินพัฒนาการความสามารถของผู้สอบในแบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน

ของ

สุนีย์ เงินยวง

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์ ดร.สมชาย สันติวัฒนกุล)

วันที่ ..... เดือน พฤษภาคม พ.ศ. 2554

คณะกรรมการควบคุมปริญญานิพนธ์	คณะกรรมการสอบปากเปล่า
..... ประธาน	..... ประธาน
(รองศาสตราจารย์ ดร.องอาจ นัยพัฒน์)	(อาจารย์ ดร.สุวพร เซ็มเฮง)
..... กรรมการ	..... กรรมการ
(รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ)	(รองศาสตราจารย์ ดร.องอาจ นัยพัฒน์)
..... กรรมการ	..... กรรมการ
(รองศาสตราจารย์ ดร.สิริรัตน์ วิภาสศิลป์)	(รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ)
	..... กรรมการ
	(รองศาสตราจารย์ ดร.สิริรัตน์ วิภาสศิลป์)
	..... กรรมการ
	(อาจารย์ ดร.สุนิสา จุ้ยม่วงศรี)

ประกาศคุณูประการ

ปริญญานิพนธ์นี้สำเร็จได้ด้วยความกรุณาอย่างสูงจากคณะกรรมการควบคุมปริญญานิพนธ์ รองศาสตราจารย์ ดร.องอาจ นัยพัฒน์ ประธานควบคุมปริญญานิพนธ์ รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ และรองศาสตราจารย์ ดร.สิริรัตน์ วิภาสศิลป์ กรรมการควบคุมปริญญานิพนธ์ ที่ได้ให้คำปรึกษา ข้อชี้แนะในการจัดทำงานวิจัยนี้ทุกขั้นตอน และแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่ตลอดจนให้กำลังใจและตักเตือนให้มี สติ สมาธิ และความบากบั่นพยายาม ทำงานให้สำเร็จ ผู้วิจัยขอกราบขอบพระคุณในความเมตตาของท่านทั้งสาม ด้วยความเคารพอย่างสูง

ขอกราบขอบพระคุณกรรมการที่แต่งตั้งเพิ่มเติมทั้งสองท่านคือ อาจารย์ ดร.สุวพร เซ็มเฮง และอาจารย์ ดร.สุนิสา จุ้ยม่วงศรี ที่ให้ความกรุณาเป็นกรรมการในการสอบปากเปล่าปริญญานิพนธ์ และให้ข้อเสนอแนะที่เป็นประโยชน์เพิ่มเติม ทำให้ปริญญานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ภาควิชาการวัดผล และวิจัยทางการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒทุกท่านที่ประสิทธิประสาทวิชาความรู้แก่ผู้วิจัย พร้อมทั้งให้คำปรึกษาและเอาใจใส่อย่างดียิ่งตลอดมา

ขอขอบพระคุณมหาวิทยาลัยเชียงใหม่ ที่สนับสนุนทุนการศึกษา “ทุนพัฒนาจารย์” ให้มีโอกาสศึกษาต่อในขั้นสูงสุด ขอขอบพระคุณคณบดีคณะศึกษาศาสตร์ คณาจารย์สาขาวิชาการประเมินผลและวิจัยการศึกษา คณาจารย์สาขาวิชาคณิตศาสตร์ ตลอดจนคณาจารย์ และเจ้าหน้าที่คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่ทุกท่านที่ให้ความช่วยเหลือ และเป็นกำลังใจด้วยดีเสมอมา

ขอขอบคุณ ดร. วัชราวุฒิ กฤตินธรรม ที่ได้ช่วยเขียนโปรแกรมรวมไฟล์ข้อมูลซึ่งสามารถลดขั้นตอนการทำงาน จนสามารถนำข้อมูลไปวิเคราะห์ได้โดยง่าย

ขอขอบคุณ คุณสุรัชน์ อินทสังข์ ที่เป็นกำลังใจ และให้ความช่วยเหลือทุกๆ ด้าน ขอขอบคุณน้องสาว น้องชาย เพื่อนและกัลยาณมิตรทุกท่าน ตลอดจนญาติพี่น้องทุกคน ที่มอบความหวังดีให้อย่างสม่ำเสมอ และขอขอบใจนักศึกษา นักเรียน ตลอดจนผู้เกี่ยวข้องที่ไม่ได้เอ่ยนามทุกคนที่มีส่วนร่วมให้ปริญญานิพนธ์สำเร็จด้วยดี ขอบคุณโอกาส เวลาและอุปสรรค ที่ให้คุณค่าของปริญญานิพนธ์มากกว่าวุฒิปริญญา

ขอกราบขอบพระคุณคุณพ่อและคุณแม่ที่สนับสนุน และเป็นแรงบันดาลใจที่ดีให้กับลูกเสมอ คุณค่าและคุณประโยชน์ของปริญญานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นเครื่องบูชา ทดแทนพระคุณ บิดา มารดา ครู อาจารย์ทุกท่านที่ได้อบรมสั่งสอนจนทำให้ผู้วิจัยประสบความสำเร็จในการศึกษา

สุนีย์ เงินยวง

สารบัญ

บทที่		หน้า
1	**บทนำ**	1
	ภูมิหลัง	1
	ความมุ่งหมายของการวิจัย	11
	ความสำคัญของการวิจัย	12
	ขอบเขตของการวิจัย	13
	กรอบแนวคิดในการวิจัย	15
	นิยามศัพท์เฉพาะ	21
	สมมติฐานในการวิจัย	24
2	**เอกสารและงานวิจัยที่เกี่ยวข้อง**	26
	แบบทดสอบผสม	26
	การคำนวณคะแนนการเปลี่ยนแปลงรายบุคคล	33
	การคำนวณคะแนนการเปลี่ยนแปลงรายบุคคลตามทฤษฎีการทดสอบแบบมาตรฐานเดิม	33
	การคำนวณคะแนนการเปลี่ยนแปลงรายบุคคลตามทฤษฎีการตอบข้อสอบ	42
	ทฤษฎีการตอบข้อสอบ	44
	การประมาณค่าพารามิเตอร์ด้วยโมเดลการตอบข้อสอบ	44
	การประมาณความสามารถ	62
	วิธีการปรับเทียบค่าพารามิเตอร์ตามทฤษฎีการตอบข้อสอบ	70
	ความคลุมเครือของมาตรวัดในทฤษฎีการตอบข้อสอบ	70
	วิธีการประมาณค่าพารามิเตอร์สำหรับสร้างมาตรวัดร่วม	71
	งานวิจัยที่เกี่ยวข้อง	77
3	**วิธีดำเนินการวิจัย**	110
	การกำหนดข้อมูลจำลอง	110

สารบัญ (ต่อ)

บทที่		หน้า
3	**(ต่อ)**	
	การจัดกระทำและการวิเคราะห์ข้อมูล.....	120
4	**ผลการวิเคราะห์ข้อมูล**.....	125
	ตอนที่ 1 การวิเคราะห์ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน.....	127
	ตอนที่ 2 การศึกษาและเปรียบเทียบความสัมพันธ์ของพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน.....	147
	ตอนที่ 3 การตรวจสอบความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน.....	180
	ตอนที่ 4 การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน....	227
5	**สรุป อภิปราย และข้อเสนอแนะ**.....	277
	สรุปผลการวิจัย.....	281
	อภิปรายผลการวิจัย.....	287
	ข้อเสนอแนะ.....	293
บรรณานุกรม.....		296
ภาคผนวก.....		318
ภาคผนวก ก.....		319
ภาคผนวก ข.....		329
ประวัติย่อผู้วิจัย.....		342

บัญชีตาราง

ตาราง **หน้า**

1 การเปรียบเทียบคุณลักษณะของข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบ 29

2 สรุปประเภทของโมเดลการตอบข้อสอบแบบให้คะแนนหลายค่าแบบเรียงอันดับ..... 59

3 สรุปลักษณะความสามารถของผู้สอบ 1,000 คนในการทดสอบซ้ำสองครั้ง ตามระดับพัฒนาการความสามารถที่เพิ่มขึ้น..... 111

4 สรุปลักษณะข้อสอบแบบทดสอบผสมที่ศึกษา เมื่อขนาดความยาว ระดับความยาก และระดับการให้คะแนนที่แตกต่างกัน..... 115

5 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (M1)..... 128

6 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (M2)..... 130

7 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (M3)..... 132

8 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 (M4)..... 134

9 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 (M5)..... 136

10 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 (M6)..... 138

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

11 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 (M7)..... 140

12 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 (M8)..... 142

13 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 (M9)..... 144

14 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_1$) กับค่าความสามารถจริงในการทดสอบครั้งที่ 1 (θ_1) เมื่อผู้สอบมีพัฒนาการความสามารถต่างกัน และลักษณะแบบทดสอบผสมต่างกัน..... 148

15 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2$) กับค่าความสามารถจริงในการทดสอบครั้งที่ 2 (θ_2) เมื่อผู้สอบมีพัฒนาการความสามารถต่างกัน และลักษณะแบบทดสอบผสมต่างกัน..... 154

16 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ยฟิชเชอร์ Z ของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันความสามารถจริง (θ_1) กรณีแบบทดสอบผสมมีความยาว ความยาก และ การให้คะแนนแตกต่างกัน..... 167

17 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน..... 169

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

18 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน..... 170

19 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ยฟิชเชอร์ Z ของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันความสามารถจริง (θ_2) ภายใต้เงื่อนไขแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ..... 172

20 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน..... 173

21 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน..... 175

22 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน..... 177

23 สรุปความสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 และครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าพารามิเตอร์ความสามารถจริง..... 179

24 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมีความยาว และ ความยากคงที่ จำแนกตามการให้คะแนนและระดับพัฒนาการ..... 182

25 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีความยาก และการให้คะแนนคงที่ จำแนกตามความยาวและระดับพัฒนาการ..... 187

26 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีการให้คะแนน และความยาวคงที่ จำแนกตามความยากและระดับพัฒนาการ..... 192

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

27 ค่าเฉลี่ยค่า RMSE ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมี ความยาว และความยากคงที่ จำแนกตามการให้คะแนน และระดับพัฒนาการ..... 197

28 ค่าเฉลี่ยของค่า RMSE ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมี ความยาก และการให้คะแนนคงที่ จำแนกตามความยาวและระดับพัฒนาการ..... 203

29 ค่าเฉลี่ยของค่า RMSE ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีการให้คะแนน และความยาวคงที่ จำแนกตามความยากและระดับพัฒนาการ..... 208

30 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ย BIAS กรณีแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ..... 213

31 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน..... 215

32 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน..... 216

33 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน..... 218

34 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ย RMSE กรณีแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ..... 219

35 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน..... 220

36 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน..... 222

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

37 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน..... 223

38 สรุปความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน..... 226

39 ค่าเฉลี่ยความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาว และ ความยากคงที่ จำแนกตามการให้คะแนนและระดับพัฒนาการ..... 229

40 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถรายบุคคล เมื่อแบบทดสอบผสมมีความยาก และการให้คะแนนคงที่ จำแนกตามความยาวและระดับพัฒนาการ..... 235

41 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีการให้คะแนน และความยาวคงที่ จำแนกตามระดับความยาก และระดับพัฒนาการ..... 239

42 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมี ความยาว ความยากคงที่ จำแนกตามระดับการให้คะแนนและระดับพัฒนาการ..... 245

43 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมี ความยาก และการให้คะแนนคงที่ จำแนกตามความยาวและระดับพัฒนาการ..... 250

44 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมี ความยาว และการให้คะแนนคงที่ จำแนกตามความยากและระดับพัฒนาการ..... 254

45 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกรณีแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ..... 260

46 การวิเคราะห์ความแตกต่างค่าเฉลี่ย BIASของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน..... 261

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

47 การวิเคราะห์ความแตกต่างค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน..... 263

48 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน 3 ระดับ จำแนกตามระดับ SD ของพัฒนาการความสามารถ..... 264

49 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกรณีแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ..... 267

50 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน..... 268

51 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน..... 270

52 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน..... 271

53 สรุปความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน..... 276

54 สรุปผลการตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีพัฒนาการความสามารถแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยากและการให้คะแนนของแบบทดสอบผสมแตกต่างกันโดยพิจารณาในภาพรวมและการเปรียบเทียบภายในกลุ่มและระหว่างกลุ่ม..... 286

55 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้เงื่อนไขความยาว 30:10 24:8และ15:5 ความยาก -0.50 และให้คะแนน 3 ระดับ..... 320

บัญชีตาราง (ต่อ)

ตาราง **หน้า**

56 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้ความยาว 30:10 24:8 และ15:5 ความยาก 0.00 และให้คะแนน 3 ระดับ 321

57 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้ความยาว 30:10 24:8 และ15:5 ความยากเฉลี่ย 0.50 และให้คะแนน 3 ระดับ 322

58 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก -0.50 และให้คะแนน 4 ระดับ..... 323

59 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก 0.00 และให้คะแนน 4 ระดับ..... 324

60 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก 0.50 และให้คะแนน 4 ระดับ..... 325

61 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก -0.50 และให้คะแนน 5 ระดับ..... 326

62 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก 0.00 และให้คะแนน 5 ระดับ..... 327

63 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก 0.50 และให้คะแนน 5 ระดับ..... 328

บัญชีภาพประกอบ

ภาพประกอบ **หน้า**

1 กรอบแนวคิดในการศึกษาความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันโดยโมเดลโลจิสติกแบบสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตของข้อมูลการทดสอบซ้ำด้วยแบบทดสอบผสมที่มีความยาว ความยาก และ การให้คะแนนข้อแตกต่างกัน..... 21

2 โค้งลักษณะข้อสอบ (ICC) ของตัวอย่างข้อสอบ 3 ข้อสำหรับโมเดลการตอบข้อสอบแบบ 1 พารามิเตอร์..... 45

3 ตัวอย่างโค้งลักษณะข้อสอบ (ICC) ของข้อสอบ 3 ข้อ สำหรับโมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์และ 3 พารามิเตอร์..... 47

4 โค้งคุณลักษณะของรายการคำตอบประจำข้อสอบในโมเดลการตอบข้อสอบแบบให้คะแนนหลายค่า..... 53

5 ขั้นตอนการกำหนดพารามิเตอร์ผู้สอบ..... 112

6 ขั้นตอนการกำหนดพารามิเตอร์ข้อสอบกรณีแบบทดสอบผสมมีความยาว 30:10 มีความยาก -0.50 และให้คะแนน 3 ระดับ..... 116

7 การกำหนดพารามิเตอร์ข้อสอบกรณีแบบทดสอบผสมมีความยาว 24:8 และ 15:5 มีความยาก -0.50 และให้คะแนน 3 ระดับ..... 116

8 ขั้นตอนการจัดกระทำและการวิเคราะห์ข้อมูล..... 124

9 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) เมื่อความยาวของแบบทดสอบผสมคงที่ 30:10 (9 ก) 24:8 (9 ข) และ 15:5 (9 ค)..... 149

10 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) เมื่อความยากของแบบทดสอบผสมคงที่ -0.50 (10 ก) 0.00 (10 ข) และ 0.50 (10 ค)..... 150

11 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) เมื่อแบบทดสอบผสมมีการให้คะแนนคงที่ 3 ระดับ (11 ก) 4 ระดับ (11 ข) และ 5 ระดับ (11 ค)..... 151

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

12 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยาวเป็น 30:10 และความยากของแบบทดสอบเป็น -0.50 (12 ก) 0.00(12 ข) 0.50(12 ค) 156

13 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยาวเป็น 24:8 และความยากของแบบทดสอบเป็น -0.50 (13 ก) 0.00 (13 ข) 0.50(13 ค)..... 157

14 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยาวเป็น 15:5 และความยากของแบบทดสอบเป็น -0.50 (14 ก) 0.00 (14 ข) 0.50(14 ค)..... 158

15 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยากของแบบทดสอบผสมเป็น -0.50 และให้คะแนน 3 ระดับ (15 ก) 4 ระดับ (15 ข) 5 ระดับ (15 ค)..... 159

16 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยากของแบบทดสอบผสมเป็น 0.00 และให้คะแนน 3 ระดับ (16 ก) 4ระดับ (16 ข) 5ระดับ (16 ค)..... 160

17 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยากของแบบทดสอบผสมเป็น 0.50 และให้คะแนน 3 ระดับ (17 ก) 4 ระดับ (17 ข) 5 ระดับ (17 ค)..... 161

18 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อให้คะแนน 3 ระดับ และความยาวเป็น 30:10 (18 ก) 24:8 (18 ข) 15:5 (18 ค)..... 162

19 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อให้คะแนน 4 ระดับ และความยาวเป็น 30:10 (19 ก) 24:8 (19 ข) 15:5 (19 ค)..... 163

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

20 แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อให้คะแนน 5 ระดับ และความยาวเป็น 30:10 (20 ก) 24:8 (20 ข) 15:5 (20 ค)..... 164

21 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 เมื่อแบบทดสอบมีความยาวแตกต่างกัน 3 ระดับ..... 168

22 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถครั้งที่ 1 เมื่อแบบทดสอบมีความยากแตกต่างกัน 3 ระดับ..... 169

23 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถครั้งที่ 1 เมื่อแบบทดสอบมีการให้คะแนนแตกต่างกัน 3 ระดับ..... 171

24 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบมีความยาวแตกต่างกัน 3 ขนาด..... 174

25 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีความยากแตกต่างกัน 3 ระดับ..... 176

26 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบมีการให้คะแนนแตกต่างกัน 3 ระดับ..... 177

27 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก -0.50 (27 ก) 0.00 (27 ข) 0.50 (27 ค)..... 183

28 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 (28 ก) 0.00 (28 ข) 0.50 (28 ค)..... 184

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

29 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก -0.50 (29 ก) 0.00 (29 ข) 0.50 (29 ค) 185

30 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 3 ระดับ (30 ก) 4 ระดับ (30 ข) 5 ระดับ (30 ค)..... 189

31 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.00 และให้คะแนน 3 ระดับ (31 ก) 4 ระดับ (31 ข) 5 ระดับ (31 ค)..... 190

32 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 3 ระดับ (32 ก) 4 ระดับ (32 ข) 5 ระดับ (32 ค)..... 191

33 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 3 ระดับ และความยาวของแบบทดสอบผสมเป็น 30:10 (33 ก) 24:8 (33 ข) 15:5 (33 ค)..... 194

34 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 4 ระดับและความยาวของแบบทดสอบผสมเป็น 30:10 (34 ก) 24:8 (34 ข) 15:5 (34 ค)..... 195

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

35 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 5 ระดับ และความยาวของแบบทดสอบผสมเป็น 30:10 (35 ก) 24:8 (35 ข) 15:5 (35 ค)..... 196

36 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก -0.50 (36 ก) 0.00 (36 ข) 0.50 (36 ค) 198

37 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 (37 ก) 0.00 (37 ข) 0.50 (37 ค) 199

38 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก -0.50 (38 ก) 0.00 (38 ข) 0.50 (38 ค) 200

39 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยากของแบบทดสอบผสมเป็น -0.50 และให้คะแนน 3 ระดับ (39 ก) 4 ระดับ (39 ข) 5 ระดับ (39 ค) 204

40 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยากของแบบทดสอบผสมเป็น 0.00 และให้คะแนน 3 ระดับ (40 ก) 4 ระดับ (40 ข) 5 ระดับ (40 ค)..... 205

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

41 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยากของแบบทดสอบผสมเป็น 0.50 และการให้คะแนน 3 ระดับ (41 ก) 4 ระดับ (41 ข) 5 ระดับ (41 ค)..... 206

42 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมยาว 30:10 (42 ก) 24:8 (42 ข) 15:5 (42 ค)..... 209

43 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อจำนวนรายการคำตอบเป็น 4 และความยาวเป็น 30:10 (43 ก) 24:8 (43 ข) 15:5 (43 ค) 211

44 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อจำนวนรายการคำตอบเป็น 5 และความยาวเป็น 30:10 (44 ก) 24:8 (44 ข) 15:5 (44 ค) 212

45 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน 3 ขนาด..... 215

46 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน 3 ระดับ..... 217

47 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน 3 ขนาด 218

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

48 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน 3 ขนาด..... 221

49 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน 3 ระดับ..... 222

50 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน 3 ระดับ..... 224

51 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก -0.50 (51 ก) 0.00 (51 ข) 0.50 (51 ค)..... 230

52 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 (52 ก) 0.00 (52 ข) 0.50 (52 ค) 231

53 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก -0.50 (53 ก) 0.00 (53 ข) 0.50 (53 ค) 232

54 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 3 ระดับ (54 ก) 4 ระดับ (54 ข) 5 ระดับ (50 ค)..... 236

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

55 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.00 และให้คะแนน 3 ระดับ (55 ก) 4 ระดับ (55 ข) 5 ระดับ (55 ค)..... 237

56 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 3 ระดับ (56 ก) 4 ระดับ (56 ข) 5 ระดับ (56 ค)..... 238

57 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมมีความยาว 30:10 (57 ก) 24:8 (57 ข) 15:5 (57 ค)..... 241

58 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 4 ระดับ และแบบทดสอบผสมมีความยาว 30:10 (58 ก) 24:8 (58 ข) 15:5 (58 ค)..... 242

59 แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 5 ระดับ และแบบทดสอบผสมมีความยาวเป็น 30:10 (59 ก) 24:8 (59 ข) 15:5 (59 ค)..... 243

60 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยาวเป็น 30:10 และความยากของแบบทดสอบเป็น -0.50 (60 ก) 0.00 (60 ข) 0.50 (60 ค)..... 246

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

61 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยาวเป็น 24:8 และความยากของแบบทดสอบเป็น -0.50 (61 ก) 0.00 (61 ข) 0.50 (61 ค)..... 247

62 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยาวเป็น 15:5 และความยากของแบบทดสอบเป็น -0.50 (62 ก) 0.00 (62 ข) 0.50 (62 ค)..... 248

63 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 3 ระดับ (63 ก) 4 ระดับ (63 ข) 5 ระดับ (63 ค)..... 251

64 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.00 และให้คะแนน 3 ระดับ (64 ก) 4 ระดับ (64ข) 5 ระดับ (64 ค)..... 252

65 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 3 ระดับ (65 ก) 4 ระดับ (65 ข) 5 ระดับ (65 ค)..... 253

66 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อมีการให้คะแนน 3 ระดับ และแบบทดสอบผสมยาว 30:10 (66 ก) 24:8 (66 ข) 15:5 (66 ค)..... 256

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

67 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อมีการให้คะแนน 4 ระดับ และแบบทดสอบผสมยาว 30:10 (67 ก) 24:8 (67 ข) 15:5 (67 ค)..... 257

68 แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อมีการให้คะแนน 5 ระดับ และแบบทดสอบผสมยาว 30:10 (68 ก) 24:8 (68 ข) 15:5 (68 ค) 258

69 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวต่างกัน..... 262

70 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากต่างกัน..... 263

71 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน..... 265

72 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อกลุ่มส่วนเบี่ยงเบนของพัฒนาการเป็น 1.00 0.80 และ 1.20..... 266

73 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน..... 269

74 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน..... 270

75 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน..... 272

บัญชีภาพประกอบ (ต่อ)

ภาพประกอบ **หน้า**

76 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อจำแนกตามกลุ่มส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการเป็น 1.00 0.80 และ 1.20..... 273

บทที่ 1
บทนำ

ภูมิหลัง

ปัจจุบันแนวโน้มการทดสอบในชั้นเรียนและการทดสอบระดับชาติ มีการใช้แบบทดสอบที่ผสมข้อสอบอย่างน้อยสองลักษณะต่างกันในฉบับเดียวมากขึ้น ดังผลการสำรวจของเลน (Cao. 2008: 14 citing Lane. 2005) พบว่าในสหรัฐอเมริกามีประมาณ 63% ของรัฐทั้งหมดใช้แบบทดสอบผสมในการวัดประเมินผล แบบทดสอบผสมสามารถวัดความสามารถได้ครอบคลุมกว่าการใช้แบบทดสอบที่มีข้อสอบลักษณะเดียว (Kim; & Lee. 2006: 53) การผสมข้อสอบเลือกตอบและข้อสอบเขียนตอบในแบบทดสอบฉบับเดียวกันเป็นการเพิ่มจุดแข็งและชดเชยข้อด้อยของข้อสอบแต่ละประเภทได้ โดยข้อสอบแบบเลือกตอบสามารถสุ่มเนื้อหาตัวอย่างที่ต้องการวัดได้กว้าง มีความเชื่อมั่นสูง ตรวจให้คะแนนเป็นปรนัย ในขณะที่ข้อสอบแบบเขียนอธิบายคำตอบสามารถวัดเนื้อหาและโครงสร้างได้อย่างเที่ยงตรง ยากต่อการเดา และวัดการรู้คิดที่ซับซ้อนได้ ดังนั้นหากนำแบบทดสอบผสมมาใช้ในการวัดประเมินผลซึ่งเป็นกระบวนการที่ดำเนินควบคู่ไปกับการจัดการเรียนการสอน ย่อมสามารถช่วยให้ได้ข้อมูลสำหรับวินิจฉัยความรู้พื้นฐานและทักษะที่จำเป็นสำหรับผู้เรียน ตัดสินความสำเร็จในการเรียนรู้ตามหลักสูตร และเป็นข้อมูลสารสนเทศในการวางแผนบริหารการจัดการศึกษา กำหนดนโยบายและพัฒนาหลักสูตรการศึกษาได้

โดยทั่วไปการตรวจสอบอย่างง่ายว่าผู้เรียนแต่ละคนมีพัฒนาการความสามารถเพิ่มขึ้นหลังจากการจัดการเรียนรู้ตามหลักสูตรหรือไม่จะพิจารณาจากคะแนนผลต่าง (Difference score) ของคะแนนก่อนและหลังเรียนของผู้สอบคนเดิมที่ทำแบบทดสอบฉบับเดิมหรือคู่ขนาน วิธีการคำนวณผลต่างของคะแนนดิบมีขั้นตอนการคำนวณไม่ยุ่งยากซับซ้อน จึงนิยมนำไปใช้กันอย่างกว้างขวาง แต่การใช้คะแนนผลต่างอย่างง่ายเป็นดัชนีที่แสดงถึงพัฒนาการเรียนรู้ของผู้เรียนที่มีข้อจำกัดหลายประการดังนี้ (Lord. 1958.; Cronbach; & Furby. 1970: 68 – 80; Linn; & Slinde. 1977: 121 – 150; May; & Nicewander. 1998: 882 – 897; Traub. 1994: 127 – 133; Willet; & Sayer. 1994. 363 – 365; Prieler; & Raven. 2002: 3 – 4.) ประการแรก ความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนและคะแนนผลต่างมีค่าเป็นลบ ประการที่สอง แม้ว่าค่าความเชื่อมั่นของแบบทดสอบก่อนเรียนและค่าความเชื่อมั่นของแบบทดสอบหลังเรียนมีค่าสูงแต่ความเชื่อมั่นของคะแนนผลต่างกลับมีค่าต่ำ ประการที่สาม คะแนนผลต่างอย่างง่ายขึ้นอยู่กับอิทธิพลคะแนนเพดาน (Ceiling effect) ซึ่งหมายถึง นักเรียนที่มีระดับความสามารถสูงมีโอกาสน้อยที่จะได้คะแนนผลต่างสูงกว่านักเรียนที่มีความสามารถต่ำ และประการที่สี่ การแปลความหมายของคะแนนผลต่างขึ้นอยู่กับอิทธิพลคะแนนพื้น (Floor effect) ซึ่งเป็น

ระดับความสามารถเริ่มต้นหรือคะแนนสอบครั้งแรก ดังเช่น กรณีที่ผู้สอบสองคนได้คะแนนผลต่างเท่ากัน ไม่สามารถสรุปได้แน่นอนว่าผู้สอบมีพัฒนาการความสามารถเท่ากัน นั่นคือการแปลความหมายของคะแนนผลต่างขึ้นอยู่กับคะแนนสอบครั้งแรกหรือระดับความสามารถเริ่มต้น

ในเวลาต่อมาได้มีการพัฒนาวิธีการคำนวณพัฒนาการขึ้นมาใหม่หลายวิธีเพื่อแก้ไขและหลีกเลี่ยงข้อจำกัดของคะแนนผลต่างอย่างง่าย แต่ล้วนเป็นวิธีการคำนวณที่พัฒนาขึ้นภายใต้ทฤษฎีการทดสอบแบบมาตรฐานเดิมซึ่งมีข้อจำกัดเกี่ยวกับเงื่อนไขของการทดสอบซ้ำหลายประการ (Fischer. 2003) ประการแรก คะแนนการเปลี่ยนแปลงของผู้สอบทุกคนถูกพิจารณาว่าต้องมีความคลาดเคลื่อนเท่ากัน ประการที่สอง ต้องทดสอบขนาดของกลุ่มตัวอย่างของผู้เข้าสอบทั้งสองครั้งเพื่อให้ได้พารามิเตอร์ที่ประมาณค่าเป็นอย่างดีซึ่งเป็นสิ่งที่ยากในการปฏิบัติจริง ประการที่สาม คะแนนที่คาดหวัง ความแปรปรวนของคะแนน ความสัมพันธ์ของคะแนน และ ความเชื่อมั่นของคะแนนเป็นค่าพารามิเตอร์ของประชากรที่ขึ้นอยู่กับการแจกแจงความสามารถของกลุ่มประชากรอ้างอิง ดังนั้นการวัดประเมินคะแนนการเปลี่ยนแปลงในแต่ละบุคคลย่อมขึ้นอยู่กับการแจกแจงนั้นด้วย และประการที่สี่ การวัดการเปลี่ยนแปลงและการวัดประเมินความแม่นยำของการวัดเป็นการนิยามภายใต้ทฤษฎีการทดสอบแบบมาตรฐานเดิมกำหนดว่าต้องมี “มาตรวัดที่ชัดเจน (Manifest scale)” ทำให้ถูกอิทธิพลของคะแนนเริ่มต้น และคะแนนเต็มสูงสุด บีบอัดมาตรวัดให้อยู่ในขอบเขตของคะแนนมวลความรู้

จากข้อจำกัดของวิธีคำนวณการเปลี่ยนแปลงตามทฤษฎีการทดสอบแบบมาตรฐานเดิมจึงมีการพัฒนาวิธีการคำนวณการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT) โดยเชื่อว่าโมเดลการตอบข้อสอบสามารถชดเชยข้อจำกัดของทฤษฎีการทดสอบแบบมาตรฐานเดิมได้หลายประการ (Fischer. 2003) ได้แก่ ความอิสระในการเลือกชุดของข้อสอบก่อนและหลังเรียน ค่าพารามิเตอร์ข้อสอบไม่มีการแปรเปลี่ยน (Invariance) เมื่อกลุ่มตัวอย่างเปลี่ยนไป ความแม่นยำของคะแนนการเปลี่ยนแปลงขึ้นอยู่กับระดับความยากของข้อสอบ ตลอดจนคะแนนที่คาดหวัง ความแปรปรวนของคะแนน ความสัมพันธ์ของคะแนน และ ความเชื่อมั่นของคะแนนซึ่งเป็นค่าพารามิเตอร์ของประชากรไม่ขึ้นอยู่กับการแจกแจงความสามารถของกลุ่มประชากรอ้างอิง แต่การคำนวณการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบ (IRT) มีเงื่อนไขว่าโมเดลการตอบข้อสอบที่ใช้ต้องมีความสอดคล้องกลมกลืนกับข้อมูลอย่างเพียงพอ ตัวอย่างของโมเดลการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบ (IRT) เช่น โมเดลคุณลักษณะแฝงโลจิสติกแบบเส้นตรงแบบผ่อนปรนข้อตกลงเบื้องต้น (Linear logistic latent trait model with relaxed assumptions: LLRA)(Fischer. 1976) โมเดลโค้งพัฒนาการ (Growth model) (Bock. 1976) โมเดลราสช์พหุมิติของแอนเดอร์สัน (Multidimensional Rasch Models) (Anderson. 1985) โมเดลราสช์พหุมิติเพื่อวัดการเปลี่ยนแปลงการเรียนรู้ของเอมเบรทสัน (Multidimensional Rasch Model for Learning and Change: MRMLC) (Embretson.

1991) การวัดการเปลี่ยนแปลงความสามารถ (Latent Change) (May; & Nicewander. 1998) และการวัดคะแนนความแตกต่างระหว่างมวลความรู้ตามทฤษฎีการตอบข้อสอบ (Heidenberg. 2000) เป็นต้น

หากพิจารณาขั้นตอนการคำนวณ และเงื่อนไขการใช้ของวิธีการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบดังที่กล่าวข้างต้น วิธีการคำนวณคะแนนการเปลี่ยนแปลงความสามารถ ($\theta_2 - \theta_1$) ของเมย์และไนส์วานเดอร์ (May; & Nicewander. 1998) เป็นวิธีการคำนวณเพื่อหาผลต่างระหว่างค่าประมาณความสามารถของผู้สอบแต่ละครั้ง จึงสามารถคำนวณได้ง่ายกว่าเมื่อเทียบกับโมเดลการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบอื่น การคำนวณวิธีนี้ช่วยขจัดปัญหาความคลาดเคลื่อนที่มักเกิดขึ้นจากคำนวณคะแนนผลต่างอย่างง่ายของคะแนนดิบได้ และผลต่างของค่าประมาณความสามารถมีความบิดเบือนของมาตรวัด (Scale distortion) น้อยกว่าผลต่างที่ได้จากการคำนวณคะแนนผลต่างอย่างง่ายที่คำนวณจากคะแนนรวมของข้อสอบทั้งฉบับ ซึ่งความบิดเบือนของมาตรวัดนั้นหมายถึงกรณีที่ผู้สอบมีระดับความสามารถสูงทำแบบทดสอบที่มีระดับความยากต่ำจึงทำให้ได้คะแนนพัฒนาการต่ำกว่าความเป็นจริงมากกว่ากลุ่มผู้มีความสามารถระดับปานกลางและต่ำในขณะที่ผู้สอบที่มีความสามารถระดับต่ำนั้นได้คะแนนพัฒนาการต่ำกว่าความเป็นจริงมากหากแบบทดสอบมีความยากระดับสูง

ดังที่กล่าวมาข้างต้นหากนำแบบทดสอบผสมมาใช้วัดพัฒนาการความสามารถของผู้เรียนโดยพิจารณาจากผลต่างของคะแนนดิบย่อมไม่สามารถสะท้อนความสามารถที่แท้จริงที่เกิดจากความแตกต่างของลักษณะข้อสอบซึ่งมีรูปแบบการให้คะแนนต่างกันได้ อีกทั้งคะแนนดิบที่ได้จากผลการทดสอบแต่ละครั้งมีความซับซ้อนมากเกินกว่าจะเปรียบเทียบกันได้โดยตรง การคำนวณพัฒนาการความสามารถของผู้เรียนด้วยโดยใช้ผลต่างคะแนนดิบที่ได้จากแบบทดสอบผสมจึงไม่เพียงพอสำหรับสะท้อนพัฒนาการการเรียนรู้ของผู้เรียน ดังนั้นวิธีการคำนวณคะแนนการเปลี่ยนแปลงความสามารถ ($\theta_2 - \theta_1$) อย่างง่ายจึงมีความเหมาะสมมากกว่าการคำนวณด้วยผลต่างคะแนนดิบ

อย่างไรก็ตามแม้คะแนนการเปลี่ยนแปลงความสามารถ ($\theta_2 - \theta_1$) ที่ได้จะมีความน่าเชื่อถือแต่ลักษณะของแบบทดสอบที่ใช้ส่งผลต่อการแปลความหมายของคะแนนพัฒนาการที่คำนวณได้เช่นกัน เช่น หากทดสอบซ้ำด้วยแบบทดสอบฉบับเดิม ความแตกต่างของคะแนนที่ได้อาจเป็นผลจากการจำข้อสอบเดิมได้มากกว่าเกิดจากการเรียนรู้ที่แท้จริง การแปลความหมายของผลต่างคะแนนจึงไม่สามารถสะท้อนถึงพัฒนาการการเรียนรู้ของผู้เรียนอย่างแท้จริงได้ทั้งหมด การทดสอบซ้ำโดยใช้แบบทดสอบคู่ขนานซึ่งมีการวัดเนื้อหาเดียวกันและระดับความยากของข้อสอบเท่ากันจึงถูกนำไปใช้ในการศึกษาพัฒนาการการเรียนรู้ของนักเรียนแทนการทดสอบซ้ำโดยใช้แบบทดสอบฉบับเดิม แต่การสร้างแบบทดสอบคู่ขนานสองฉบับที่วัดเนื้อหาเดียวกันและมีระดับความยากของทั้งสองฉบับเท่ากันนั้น

ปฏิบัติได้ยาก ดังนั้นผู้วิจัยจึงสนใจศึกษาการวัดพัฒนาการความสามารถที่ได้จากการทดสอบด้วยแบบทดสอบผสมที่มีระดับความยากต่างกัน

เมื่อพิจารณาการวัดพัฒนาการความสามารถโดยใช้แบบทดสอบผสมที่มีโครงสร้างการวัดเนื้อหาความสามารถเดียวกันแต่มีระดับความยากแตกต่างกัน พบว่าความสามารถของผู้สอบและลักษณะแบบทดสอบที่ใช้มีลักษณะสอดคล้องกับรูปแบบการเก็บรวบรวมข้อมูลสำหรับการปรับเทียบคะแนนสอบที่ได้จากการทดสอบผู้สอบสองกลุ่มประชากรที่มีระดับความสามารถไม่เท่าเทียมกันด้วยแบบทดสอบสองชุดที่มีความยากต่างกันแต่มีข้อสอบร่วมกัน (Common Item Non – Equivalent Group: CINEG) เนื่องจากการวัดพัฒนาการความสามารถเกิดขึ้นหลังการจัดกิจกรรมการเรียนรู้ตามหลักสูตรแล้วผู้เรียนย่อมมีความสามารถเพิ่มขึ้นหรือแตกต่างไปจากการทดสอบครั้งแรก จึงทำให้ระดับความสามารถของกลุ่มผู้สอบครั้งที่สองและครั้งที่หนึ่งย่อมแตกต่างกัน

การเปรียบเทียบคะแนนสอบของผู้สอบที่มีความสามารถไม่เท่ากัน และทำแบบทดสอบต่างฉบับซึ่งมีระดับความยากต่างกันไม่สามารถนำคะแนนเปรียบเทียบกันได้โดยตรง เนื่องจากความแตกต่างของคะแนนสอบอาจเกิดจากความแตกต่างของความสามารถของผู้สอบหรือระดับความยากของแบบทดสอบ จึงจำเป็นต้องแยกแหล่งความแปรปรวนทั้งสองออกจากกัน เพื่อให้สามารถตีความหมายได้ว่าผลต่างของคะแนนสอบเกิดจากความแตกต่างของความสามารถหรือจากแบบทดสอบ ดังนั้นหากต้องการเปรียบเทียบคะแนนสอบของผู้สอบที่มีความสามารถไม่เท่ากัน และทำแบบทดสอบต่างฉบับซึ่งมีระดับความยากต่างกัน จำเป็นต้องขจัดแหล่งความแปรปรวนเกี่ยวกับแบบทดสอบออกไป โดยกำหนดให้การทดสอบแต่ละครั้งมีข้อสอบร่วม (Linking items) ของแบบทดสอบทั้งสองฉบับหรือข้อสอบปรับเทียบ (Equating items) แล้วจึงแปลงค่าพารามิเตอร์ของข้อสอบหรือความสามารถของผู้สอบที่ได้มาจากแบบทดสอบที่มีระดับความยากแตกต่างกันให้อยู่บนมาตรวัดเดียวกันเพื่อให้สามารถนำมาเปรียบเทียบกันได้

การคำนวณคะแนนผลต่างค่าประมาณพารามิเตอร์ความสามารถ ($\theta_2 - \theta_1$) ตั้งอยู่บนฐานทฤษฎีการตอบข้อสอบ และวิธีการปรับเทียบคะแนนสอบตามทฤษฎีการตอบข้อสอบ (Item response theory: IRT) เป็นวิธีที่นิยมใช้เพื่อปรับให้ค่าประมาณพารามิเตอร์ของข้อสอบต่างฉบับอยู่บนมาตรวัดเดียวกัน (Paek; & Young. 2005: 199) ดังนั้นผู้วิจัยจึงสนใจนำวิธีการปรับเทียบคะแนนสอบตามทฤษฎีการตอบข้อสอบมาใช้ในการปรับค่าพารามิเตอร์ความสามารถให้อยู่บนมาตรวัดเดียวกันก่อนนำไปคำนวณพัฒนาการความสามารถ ($\theta_2 - \theta_1$)

การปรับเทียบคะแนนสอบตามทฤษฎีการตอบข้อสอบมี 3 ขั้นตอนหลัก (Kim. 2007:24) ได้แก่ ขั้นที่ 1 การเลือกและใช้โมเดลสำหรับประมาณค่าพารามิเตอร์ ขั้นที่ 2 การสร้างมาตรวัดร่วม และขั้นที่ 3 การแปลงคะแนนดิบให้อยู่บนมาตรวัดร่วมสำหรับการแปลความหมายและรายงานผล

คะแนน สำหรับวิธีการสร้างมาตรวัดร่วมที่นิยมใช้มี 3 วิธี ได้แก่ วิธีการประมาณค่าพารามิเตอร์แยกกัน (Separate calibration) วิธีการประมาณค่าพารามิเตอร์ที่กำหนดพารามิเตอร์ของข้อสอบร่วม (Fixed-item parameter calibration) และวิธีการประมาณค่าพารามิเตอร์พร้อมกัน (Concurrent calibration) (Kolen; & Brennan. 1995) ซึ่งแต่ละวิธีมีรายละเอียดโดยย่อดังนี้

การประมาณค่าพารามิเตอร์แยกกัน เป็นวิธีการประมาณค่าพารามิเตอร์ของข้อสอบแต่ละฉบับแยกครั้งกันตามกลุ่มผู้สอบ แล้วจึงนำค่าพารามิเตอร์ของข้อสอบร่วมทั้งสองฉบับคำนวณค่าสัมประสิทธิ์สำหรับสร้างมาตรวัดร่วมเพื่อใช้แปลงค่าประมาณพารามิเตอร์ของแบบทดสอบฉบับที่ต้องการให้อยู่ในมาตรวัดเดียวกับฉบับที่ใช้เป็นฐาน (Base group) โดยมีวิธีการคำนวณค่าสัมประสิทธิ์การสร้างมาตรวัดร่วมจำนวน 4 วิธี (Kim; & Lee. 2004) ได้แก่ วิธีคำนวณค่าเฉลี่ย/ค่าเฉลี่ย (The Mean/Mean Method) (Loyd & Hoover, 1980) วิธีคำนวณค่าเฉลี่ย/ซิกมา (The Mean/Sigma Method) (Marco. 1977) วิธีโค้งลักษณะของแบบทดสอบของแฮบารา (The Test-Characteristic Curve: TCC) (Haebara.1980) และวิธีโค้งลักษณะของแบบทดสอบ (TCC) ของสต็อกกิ้ง และลอร์ด (Stocking & Lord, 1983) เป็นต้น ส่วนวิธีการประมาณค่าพารามิเตอร์ที่กำหนดพารามิเตอร์ของข้อสอบร่วม เป็นการประมาณค่าพารามิเตอร์ข้อสอบร่วมในกลุ่มฐาน (Base group) ด้วยโมเดลการตอบข้อสอบก่อน แล้วจึงนำค่าพารามิเตอร์ข้อสอบร่วมที่ได้ไปกำหนดเป็นค่าพารามิเตอร์ข้อสอบร่วมในกลุ่มเป้าหมาย (Target group) แล้วจึงประมาณค่าพารามิเตอร์ของผลการทดสอบของกลุ่มเป้าหมาย และวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เป็นวิธีการประมาณค่าที่นำข้อมูลการตอบของทุกกลุ่มผู้สอบที่ทดสอบด้วยแบบทดสอบต่างฉบับมาประมาณค่าพารามิเตอร์พร้อมกันเพียงครั้งเดียว โดยใช้โปรแกรมคอมพิวเตอร์ที่ยอมให้มีการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบหลายกลุ่มได้ เช่น โปรแกรม PARSCALE (Muraki; &Bock.1999) และโปรแกรม MULTILOG (Thissen.1991) เป็นต้น

คุณภาพของวิธีการสร้างมาตรวัดร่วมทั้งสามวิธีข้างต้นแตกต่างกันเมื่อลักษณะแบบทดสอบ ลักษณะข้อสอบร่วม และระดับความเท่าเทียมกันของความสามารถกลุ่มผู้สอบแตกต่างกัน ดังเช่น ผลการศึกษาวิจัยของ คิม และ โคเฮน (Kim; & Cohen. 1998) ที่จำลองผลการตอบแบบทดสอบที่มีข้อสอบทั้งฉบับลักษณะเดียว พบว่าเมื่อจำนวนข้อสอบร่วมมาก วิธีการประมาณค่าพารามิเตอร์แยกกันแล้วคำนวณสัมประสิทธิ์การสร้างมาตรวัดร่วมแบบวิธีโค้งลักษณะของแบบทดสอบ (TCC) ให้ค่ารากที่สองของค่าเฉลี่ยผลต่างกำลังสอง (Root Mean Square Difference :RMSD) ของค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากต่ำกว่าวิธีการประมาณพารามิเตอร์พร้อมกันและวิธีการประมาณพารามิเตอร์พร้อมกันบนฐานการประมาณค่าแบบค่าประมาณสูงสุดปลายทาง (Marginal Maximum a Posteriori Estimation: MMAP) แต่ถ้าจำนวนข้อสอบร่วมน้อย

วิธีการสร้างมาตรวัดร่วมทั้งสามวิธีให้ค่าประมาณพารามิเตอร์ที่มีความคลาดเคลื่อนมาตรฐานไม่ต่างกัน

ต่อมาปี 2002 แฮนสัน และ บีกวิน (Hanson; & Béguin. 2002) ได้จำลองข้อมูลแบบทดสอบที่มีข้อสอบทั้งฉบับลักษณะเดียวเช่นกัน พบว่าการประมาณค่าพารามิเตอร์แบบพร้อมกันให้ค่าพารามิเตอร์ที่มีความคลาดเคลื่อนต่ำกว่าการประมาณค่าพารามิเตอร์แบบแยกกัน แต่ไม่ได้เกิดขึ้นทุกกรณีที่ศึกษา ทั้งนี้ แฮนสัน และ บีกวิน อธิบายว่าการประมาณค่าพารามิเตอร์พร้อมกันได้รวมกลุ่มตัวอย่างจากการทดสอบสองกลุ่มเข้าเป็นกลุ่มเดียวกัน จึงทำให้การประมาณค่าพารามิเตอร์แบบพร้อมกันมีค่าความคลาดเคลื่อนต่ำกว่าการประมาณค่าพารามิเตอร์แยกกัน ในปีเดียวกัน คิมและโคเฮน (Kim; & Cohen. 2002) ได้ข้อสรุปทำนองเดียวกันกับ แฮนสัน และ บีกวิน ว่าการประมาณค่าพารามิเตอร์พร้อมกันด้วยโมเดลเกรดเรสพอนส์จะได้ค่ารากที่สองของค่าเฉลี่ยผลต่างกำลังสอง (RMSD) ของพารามิเตอร์ความสามารถต่ำกว่าการประมาณพารามิเตอร์แยกกันแล้วคำนวณหาสัมประสิทธิ์การสร้างมาตรวัดร่วมแบบโค้งลักษณะแบบทดสอบ (TCC)

ผลการศึกษาเปรียบเทียบประสิทธิภาพของวิธีการสร้างมาตรวัดร่วมจากการจำลองข้อมูลดังกล่าวข้างต้นวิธีการประมาณค่าพามิเตอร์พร้อมกันให้ค่าประมาณพารามิเตอร์ที่มีความคลาดเคลื่อนต่ำกว่าวิธีการประมาณค่าพารามิเตอร์แบบแยกกัน แต่ผลการศึกษาที่ได้ไม่เพียงพอที่จะสรุปได้อย่างชัดเจนว่าวิธีการประมาณค่าแบบพร้อมกันให้ผลที่ดีกว่าอย่างสิ้นเชิง ต่อมาในปี 2003 โจเดียน และ คณะ (Jodoin; Keller & Swaminathan. 2003: 229 – 250) ได้ศึกษาระดับพัฒนาการความสามารถด้วยข้อมูลจริงจากแบบทดสอบที่ใช้มีการให้คะแนนข้อสอบหลายลักษณะในฉบับเดียวและประมาณค่าผลการตอบข้อสอบแยกตามลักษณะการให้คะแนนของข้อสอบ พบว่า วิธีการประมาณค่าพารามิเตอร์แยกกัน วิธีการประมาณพารามิเตอร์ที่กำหนดค่าพารามิเตอร์ของข้อสอบร่วมและวิธีการประมาณพารามิเตอร์พร้อมกัน ที่ใช้ตัวประมาณค่าพารามิเตอร์ความสามารถแบบ MLE (Maximum Likelihood Estimation) ให้ค่าประมาณพารามิเตอร์ความสามารถสูงกว่าตัวประมาณค่าแบบพารามิเตอร์ความสามารถแบบ EAP โดยวิธีการสร้างมาตรวัดร่วมแบบกำหนดพารามิเตอร์ของข้อสอบร่วม และวิธีการประมาณพารามิเตอร์พร้อมกันให้ค่าพัฒนาการความสามารถต่ำกว่าวิธีการประมาณพารามิเตอร์แยกกัน แต่ไม่สามารถตรวจสอบได้ว่าระดับพัฒนาการความสามารถที่ได้มีความคลาดเคลื่อนไปจากระดับพัฒนาการความสามารถจริงมากน้อยเพียงใด โจเดียน และ คณะ จึงเสนอให้มีการจำลองข้อมูลเพื่อตรวจสอบความคลาดเคลื่อนของพัฒนาการความสามารถที่ได้จากวิธีการสร้างมาตรวัดร่วมเมื่อทราบค่าความสามารถจริงต่อไป

ต่อมามีการศึกษาเปรียบเทียบคุณภาพวิธีการสร้างมาตรวัดร่วมของผลการตอบแบบทดสอบผสม โดยเริ่มจากคิม และ ลี (Kim; & Lee. 2004) ได้จำลองข้อมูลเพื่อศึกษาผลการปรับขยายวิธีการ

คำนวณสัมประสิทธิ์การสร้างมาตรวัดร่วมของวิธีจากการประมาณค่าพารามิเตอร์แยกกันทั้ง 4 วิธีจากแบบทดสอบที่ให้คะแนนแบบสองค่าเป็นแบบทดสอบผสม โดยใช้โมเดลโลจิสติกสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตประมาณค่าพารามิเตอร์ข้อสอบสองลักษณะในแบบทดสอบผสมพร้อมกัน พบว่ามาตรวัดร่วมที่ได้มีความคลาดเคลื่อนลดลงหากจำนวนกลุ่มตัวอย่างมากขึ้น จำนวนของข้อสอบร่วมของข้อสอบแต่ละลักษณะเพิ่มขึ้น หรือระดับความสามารถผู้สอบเท่าเทียมกัน ต่อมาปี 2005 เปค และยัง (Paek; & Young. 2005) ได้ตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์แบบกำหนดค่าพารามิเตอร์ข้อสอบให้คงที่พร้อมกับกำหนดการแจกแจงความสามารถก่อนการประมาณค่า (Fixed-person prior distribution) โดยจำลองข้อมูลแบบทดสอบผสม เมื่อผู้สอบมีค่าเฉลี่ยพัฒนาการความสามารถจากการสอบครั้งแรก 11 ระดับได้แก่ –0.5, –0.4, –0.3, –0.2, –0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5 และมีค่าส่วนเบี่ยงเบนมาตรฐานพัฒนาการความสามารถที่เป็นอัตราส่วนระหว่างส่วนเบี่ยงเบนมาตรฐานของความสามารถที่เปลี่ยนแปลงต่อส่วนเบี่ยงเบนมาตรฐานครั้งแรก 3 ระดับ ได้แก่ 1 0.8 และ 1.2 ผลการศึกษาพบว่า เมื่อผู้สอบมีพัฒนาการ ±0.1 ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ของความสามารถ (Mean Absoluted Error: MAE) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE) ต่ำกว่าผู้สอบที่มีพัฒนาการสูงกว่าหรือต่ำกว่า ±0.1 และ เมื่อพัฒนาการความสามารถของผู้สอบมีค่าส่วนเบี่ยงเบนมาตรฐานสูงกว่าส่วนเบี่ยงเบนมาตรฐานของความสามารถเริ่มต้น (1.2) จะทำให้ ค่า MAE และ ค่า RMSE ของค่าประมาณพารามิเตอร์ความสามารถมีค่าสูงกว่ากรณีที่ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่าหรือต่ำกว่าส่วนเบี่ยงเบนมาตรฐานของความสามารถเริ่มต้น (1 หรือ 0.8) อย่างไรก็ตามการศึกษาของเปค และยัง เปรียบเทียบเฉพาะวิธีการคำนวณสัมประสิทธิ์ของการประมาณค่าพารามิเตอร์แยกกัน แต่ไม่ได้ศึกษาหรือเปรียบเทียบประสิทธิภาพกับวิธีสร้างมาตรวัดร่วมอื่น

ในปี 2006 คิมและ โคเลน (Kim; & Kolen. 2006) ได้เปรียบเทียบวิธีการสร้างมาตรวัดร่วมจากแบบทดสอบผสม พบว่าค่าประมาณพารามิเตอร์ที่ได้การประมาณค่าพารามิเตอร์พร้อมกันมีความคลาดเคลื่อนต่ำกว่าและแข็งแกร่งต่อการเกิดผลการใช้แบบทดสอบผสมหรือความเป็นพหุมิติ (Multidimensional) ที่มาจากการผสมข้อสอบสองลักษณะในฉบับเดียวกันสูงกว่าวิธีการประมาณค่าพารามิเตอร์แยกกันที่ใช้วิธีการคำนวณสัมประสิทธิ์การสร้างมาตรวัดร่วมด้วยวิธีโค้งคุณลักษณะข้อสอบทั้งวิธีของสต็อกกิ้งและลอร์ด (Stocking; & Lord.1983) และ วิธีของแฮบารา (Haebara.1980) วิธีค่าเฉลี่ย/ค่าเฉลี่ย และ วิธีค่าเฉลี่ย/ซิกมา แต่มีข้อสังเกตว่าการประมาณค่าพารามิเตอร์พร้อมกันมีความคลาดเคลื่อนต่ำกว่าการประมาณค่าพารามิเตอร์แบบแยกกันที่ใช้วิธีการคำนวณสัมประสิทธิ์การปรับเทียบด้วยวิธีโค้งคุณลักษณะข้อสอบเพียงเล็กน้อยเท่านั้น

ปัจจัยสำคัญประการหนึ่งที่ทำให้วิธีการประมาณค่าพารามิเตอร์พร้อมกันให้ค่าพารามิเตอร์ที่มีค่าความลำเอียงต่ำและมีความคงที่ในการประมาณค่าสูงกว่าการประมาณค่าพารามิเตอร์แยกกัน คือ ข้อสอบทุกข้อทั้งที่เป็นข้อสอบร่วมและไม่เป็นข้อสอบร่วมต้องมีลักษณะตามข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ และโมเดลการตอบข้อสอบต้องมีความสอดคล้องกลมกลืนกับข้อมูล หากพบว่ามีการละเมิดข้อตกลงเบื้องต้นค่าประมาณพารามิเตอร์ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจะเกิดความคลาดเคลื่อนได้ (Hanson; & Béguin. 2002; Kim. 2004; Kim; & Cohen. 1998; Yao; & Mao. 2004) แต่ไม่อาจสรุปได้ว่าเมื่อข้อมูลขาดความเป็นเอกมิติวิธีการประมาณค่าแบบใดจะดีกว่ากัน (Béguin; & Hanson. 2001) ต่อมา เหยา และเหมา (Yao; & Mao. 2004) จึงนำข้อมูลที่มีความเป็นพหุมิติมาทดสอบ พบว่าการประมาณค่าพารามิเตอร์แบบแยกกันมีความแข็งแกร่งกว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ซึ่งต่างไปจากผลการศึกษาของคิม (Kim. 2004; Kim; & Kolen. 2006) ที่พบว่าการประมาณค่าพารามิเตอร์แบบพร้อมกันมีความแข็งแกร่งกว่าวิธีการประมาณค่าแยกกันแม้ว่าจะพบว่าข้อสอบมีความเป็นพหุมิติ นอกจากประเด็นการละเมิดข้อตกลงเบื้องต้นจะส่งผลต่อความคลาดเคลื่อนของวิธีการสร้างมาตรวัดร่วมแล้ว ฮู โรเจอรส์ และ วูคมิโรวิก (Hu; Rogers; & Vukmirovic. 2008) พบว่าหากค่าพารามิเตอร์ความยากของข้อสอบร่วมเกิดการแปรเปลี่ยน วิธีการปรับพารามิเตอร์พร้อมกันจะไม่ให้ผลการปรับเทียบค่าพารามิเตอร์ที่แม่นยำกว่าวิธีการปรับเทียบแบบประมาณค่าพารามิเตอร์แบบแยกกันแล้วใช้การแปลงเชิงเส้นแบบโค้งลักษณะข้อสอบ (TCC) และวิธีค่าเฉลี่ยและผลรวม (Mean/Sigma: M/S) หรือวิธีการประมาณค่าพารามิเตอร์แบบกำหนดให้พารามิเตอร์ข้อสอบร่วมมีค่าคงที่

แม้ว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันเป็นวิธีการสร้างมาตรวัดร่วมที่สามารถขจัดปัญหาความไม่เท่าเทียมกันได้ (Paek; & Young. 2005) ค่าประมาณพารามิเตอร์มีความคลาดเคลื่อนต่ำกว่าวิธีการประมาณค่าพารามิเตอร์แยกกัน อีกทั้งสามารถใช้ประมาณค่าผลการทดสอบด้วยแบบทดสอบผสมเมื่อผู้สอบมีความสามารถไม่เท่าเทียมกันได้แต่จากผลการศึกษาวิจัยที่ผ่านมานั้นไม่สามารถสรุปว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีคุณสมบัติที่ดีกว่าวิธีการสร้างมาตรวัดร่วมวิธีอื่นอย่างสิ้นเชิง ซึ่งแฮนสัน และ คณะ (Béguin; Hanson; & Glas. (2000); Béguin; & Hanson. (2002)) ได้ให้ข้อสังเกตว่าเมื่อผู้สอบสองกลุ่มมีความสามารถไม่เท่าเทียมกันทำแบบทดสอบที่วัดความสามารถแบบพหุมิติโดยที่ความสามารถนั้นมีค่าสหสัมพันธ์สูง วิธีประมาณค่าพารามิเตอร์พร้อมกันมีความแม่นยำต่ำกว่าวิธีสต็อกกิ้งและลอร์ด ทั้งนี้เนื่องจากมีโอกาสที่จะละเมิดข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ หรืออาจเป็นไปได้ว่าการประมาณค่าพารามิเตอร์พร้อมกันทำให้ได้ค่าพารามิเตอร์ข้อสอบร่วมเพียงชุดเดียวจึงไม่ได้มีการตรวจสอบซ้ำอีก นอกจากนี้วิธีการประมาณค่าพารามิเตอร์พร้อมกันไม่สามารถควบคุมได้เมื่อไม่มีข้อมูลระดับการตอบของข้อสอบจาก

แบบทดสอบเดิม ดังนั้นการตรวจสอบความแม่นยำของวิธีประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการหรือความสามารถผู้สอบสองกลุ่มหรือสองครั้งแตกต่างกันจึงมีจำเป็นอย่างยิ่ง เพื่อสามารถทำให้มั่นใจได้ว่าผลต่างค่าพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความแม่นยำเพียงพอสำหรับศึกษาทุกระดับพัฒนาการความสามารถที่ต่างกัน ในสถานการณ์ที่ลักษณะของแบบทดสอบที่ใช้มีคุณลักษณะของข้อสอบแตกต่างกัน

การศึกษาพัฒนาการความสามารถของผู้สอบกลุ่มเดิมด้วยแบบทดสอบผสมที่มีระดับความยากต่างกัน นอกจากความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันจะขึ้นอยู่กับการละเมิดข้อตกลงเบื้องต้นตามทฤษฎีการตอบข้อสอบแล้ว การเลือกโมเดลการตอบข้อสอบสำหรับประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถของผู้สอบ คุณลักษณะของข้อสอบ จำนวนข้อสอบร่วม ค่าพารามิเตอร์ของข้อสอบ ตลอดจนปัจจัยอื่นๆ อาจจะทำให้ค่าประมาณพารามิเตอร์ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเกิดความคลาดเคลื่อนที่ส่งผลกระทบต่อการคำนวณผลต่างความสามารถ ($\theta_2 - \theta_1$) ได้ จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวกับปัจจัยที่ส่งผลต่อความแม่นยำของค่าประมาณพารามิเตอร์ความสามารถ ผู้วิจัยกำหนดโมเดลสำหรับประมาณค่าความสามารถของผลการตอบแบบทดสอบผสม ขนาดความยาวของแบบทดสอบผสม และระดับการให้คะแนนของข้อสอบแบบให้คะแนนหลายค่า สำหรับใช้ศึกษาในครั้งนี้ ดังรายละเอียดต่อไปนี้

แบบทดสอบผสมสำหรับใช้ศึกษาในครั้งนี้หมายถึงแบบทดสอบที่ประกอบด้วยข้อสอบสองลักษณะ ได้แก่ข้อสอบแบบเลือกตอบที่ให้คะแนนแบบสองค่ากับข้อสอบเขียนอธิบายคำตอบซึ่งให้คะแนนแบบหลายค่า ดังนั้นจึงต้องการโมเดลการตอบข้อสอบสองโมเดลที่สามารถประมาณค่าพารามิเตอร์ของข้อสอบสองลักษณะไปพร้อมๆ กัน เพื่อให้ได้ค่าประมาณความสามารถ($\hat{\theta}$) บนมาตรวัดร่วมระหว่างผลการตอบข้อสอบสองลักษณะมีเพียงค่าเดียว ซึ่งในทางทฤษฎีแล้วมีความเป็นไปได้ที่จะรวมโมเดลการตอบข้อสอบสองโมเดลที่แตกต่างกันประมาณค่าความสามารถร่วมกัน (Yen; & Fitzpatrick. 2006: 118) ดังนั้นหากใช้ลักษณะการให้คะแนนของข้อสอบเป็นเกณฑ์ในการรวมโมเดลจึงมีทางเลือกในการผสมโมเดลอย่างหลากหลาย (Baker; & Kim. 2004) ซึ่งพบว่ามีหลายงานวิจัย (Li; Lissitz; & Yang.1999; Sykes; & Yen. 2000; Paek; & Young. 2005; Kim; & Lee. 2004; Kim; & Kolen.2006 ; Briggs; Weeks; & Wiley. 2008) เลือกใช้โมเดลโลจิสติกแบบสามพารามิเตอร์และโมเดลเจเนอรอลไรส์พาร์เชียลเครดิต สำหรับประมาณค่าพารามิเตอร์ของแบบทดสอบผสมข้อสอบแบบเลือกตอบและข้อสอบเขียนตอบ เนื่องจากโมเดลโลจิสติกแบบสามพารามิเตอร์และโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตเป็นโมเดลที่อยู่ในรูปทั่วไปของโมเดลอื่น

นอกจากวิธีการสร้างมาตรวัดร่วมและโมเดลการตอบข้อสอบสำหรับประมาณค่าพารามิเตอร์จะส่งผลต่อความแม่นยำของค่าประมาณพารามิเตอร์แล้ว ขนาดความยาวของแบบทดสอบ และ

ขนาดกลุ่มตัวอย่างส่งผลต่อความแม่นยำของค่าประมาณพารามิเตอร์ด้วยเช่นกัน โดยที่ความยาวของแบบทดสอบมีอิทธิพลต่อค่าประมาณความสามารถมากที่สุด ส่วนขนาดกลุ่มตัวอย่างมีอิทธิพลต่อค่าประมาณพารามิเตอร์ข้อสอบมากที่สุด (Yang. 2007) แบบทดสอบต้องมีขนาดยาวอย่างเพียงพอจึงจะทำให้ค่าประมาณพารามิเตอร์จากโมเดลการตอบข้อสอบแบบเอกมิติประมาณค่าพารามิเตอร์ให้ผลดีกว่าโมเดลการตอบข้อสอบแบบพหุมิติ แต่ถ้าแบบทดสอบมีขนาดสั้นจะทำให้ความแม่นยำของโมเดลการตอบข้อสอบแบบเอกมิติต่ำกว่าการใช้โมเดลการตอบข้อสอบแบบพหุมิติ (Zhang. 2004) ดังผลการศึกษาของ คิม และ ลี (Kim; & Lee. 2004,2006) ที่ได้ข้อสังเกตว่าแบบทดสอบผสมขนาด 30:2 มีค่าความลำเอียงสูงกว่า แบบทดสอบผสมขนาด 10:10 และ 15:5 เมื่อกำหนดระดับคะแนนข้อสอบแบบเขียนตอบทุกข้อให้มีคะแนน 5 ระดับ เท่ากัน ส่วนผลการศึกษาของคินเซย์ (Kinsey. 2003) พบว่าขนาดความยาวของแบบทดสอบผสมและสัดส่วนการให้คะแนนแบบสองค่าและการให้คะแนนแบบหลายค่าแตกต่างกันมีผลต่อค่าประมาณพารามิเตอร์ความสามารถ โดยแบบทดสอบที่มีขนาดยาวกว่าจะมีค่าสารสนเทศของแบบทดสอบสูงกว่าแบบทดสอบที่มีขนาดสั้นกว่า อย่างไรก็ตามที่ผ่านมาแม้ไม่พบข้อสรุปที่ชัดเจนว่าควรกำหนดความยาวที่เหมาะสมของแบบทดสอบแบบผสมไว้ที่เท่าใด แต่การศึกษาด้วยการจำลองข้อมูลแบบทดสอบผสมที่ผ่านมามักกำหนดความยาวจากจำนวนข้อสอบทั้งหมดในฉบับตามเวลาที่ใช้ในการทดสอบจริงแต่ละข้อ รวมทั้งคำนึงถึงระดับคะแนนของข้อสอบแบบเขียนตอบด้วย เช่น คามาตะ และ เทท (Kamata; &Tate. 2005) ได้กำหนดจำนวนข้อสอบแบบเลือกตอบต่อจำนวนข้อสอบแบบเขียนตอบในแบบทดสอบผสมแบบเดียวกับที่ เทท (Tate. 2003) ใช้เป็น 30:10 โดยให้ข้อสอบแบบเขียนตอบมีคะแนน 4 ระดับ (0/1/2/3) เพื่อทำให้จำนวนขั้นความน่าจะเป็นในการตอบข้อสอบแบบเขียนอธิบายตอบมี 30 ขั้น เท่ากับจำนวนขั้นความน่าจะเป็นในการตอบข้อสอบแบบเลือกตอบถูก ส่วนการศึกษาของ คิม(Kim.2004, Kim; & Kolen. 2006) กำหนดขนาดแบบทดสอบผสมตามจำนวนข้อสอบแบบเลือกตอบต่อจำนวนข้อสอบแบบเขียนตอบเป็น 36:15 โดยให้ข้อสอบแบบเขียนตอบมีคะแนน 3 ระดับ (0/1/2) จำนวน 9 ข้อและมีคะแนน 4 ระดับ (0/1/2/3) จำนวน 6 ข้อ เพื่อทำให้คะแนนรวมของข้อสอบทั้งสองลักษณะมี 36 คะแนนเท่ากัน ซึ่งจากผลการศึกษาของ เดอมาส์ (DeMars. 2004) พบว่าความคลาดเคลื่อนค่าประมาณพารามิเตอร์จากโมเดลนอมินอลเรสพอนส์มีความแปรปรวนลดลงหากจำนวนของรายการคำตอบ (Number of category) เพิ่มขึ้น เช่นเดียวกันกับ ปาร์ค (Park. 2000) พบว่าเมื่อจำนวนของรายการคำตอบมากขึ้นจะทำให้ความคลาดเคลื่อนของค่าประมาณพารามิเตอร์ของข้อสอบแบบให้คะแนนหลายค่าที่ประมาณค่าด้วยโมเดลเกรดเรสพอนส์และโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตลดลง

โดยสรุปการคำนวณหาผลต่างค่าประมาณความสามารถจากผลการตอบแบบทดสอบผสมที่มีระดับความยากต่างกัน มีปัจจัยเกี่ยวข้องที่อาจทำให้ผลต่างค่าประมาณความสามารถนั้นมีค่า

ใกล้เคียงหรือเทียบเท่ากับผลต่างความสามารถที่แท้จริง หรือเกิดความแม่นยำได้ ขึ้นอยู่กับโมเดลการตอบข้อสอบสำหรับประมาณค่าพารามิเตอร์ วิธีการสร้างมาตรวัดร่วม ขนาดความยาวของแบบทดสอบผสม ระดับคะแนนข้อสอบแบบเขียนตอบแต่ละข้อ และระดับความเท่าเทียมกันของความสามารถผู้สอบในการสอบแต่ละครั้ง โดยแต่ละครั้งที่สอบจำนวนกลุ่มตัวอย่างต้องมีขนาดใหญ่เพียงพอ แต่ด้วยความเป็นได้ของระดับพัฒนาการความสามารถเพิ่มขึ้นมีได้หลายลักษณะ ซึ่งความแตกต่างของระดับความสามารถในการทดสอบสองครั้งย่อมมีผลต่อค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ดังนั้นผู้วิจัยจึงสนใจจำลองข้อมูลผลการตอบแบบทดสอบผสมสองครั้ง โดยที่ระดับความยากของแบบทดสอบผสมสำหรับทดสอบครั้งที่ 2 สูงกว่าครั้งที่ 1แต่มีข้อสอบร่วมกันจำนวนหนึ่ง เพื่อตรวจสอบความแม่นยำของผลต่างค่าประมาณความสามารถที่ได้จากวิธีการสร้างมาตรวัดร่วมแบบประมาณค่าพารามิเตอร์พร้อมกันโดยใช้โมเดลการตอบข้อสอบแบบเอกมิติผสมระหว่างโมเดลโลจิสติกแบบสามพารามิเตอร์กับโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิต เพื่อศึกษาภายใต้เงื่อนไข ความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกัน โดยพิจารณาความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันจาก ความลำเอียงของค่าประมาณความสามารถ และรากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสองของค่าประมาณความสามารถเมื่อเทียบกับค่าประมาณความสามารถจริง

ความมุ่งหมายของการวิจัย

การศึกษาวิจัยครั้งนี้มีวัตถุประสงค์หลักเพื่อตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ด้วยแบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน โดยศึกษาภายใต้สถานการณ์จำลอง ซึ่งมีวัตถุประสงค์เฉพาะดังนี้

1. เพื่อศึกษาและเปรียบเทียบความสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง เมื่อผู้สอบมีพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยาก และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน

2. เพื่อศึกษาและเปรียบเทียบความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยาก และ การให้คะแนน ของแบบทดสอบผสมแตกต่างกัน

3. เพื่อศึกษาและเปรียบเทียบความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกัน

ความสำคัญของการวิจัย

การศึกษาวิจัยครั้งนี้เป็นการศึกษาด้วยข้อมูลจำลองเพื่อตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันในสถานการณ์ที่ผู้สอบมีระดับพัฒนาการความสามารถต่างกัน โดยจำลองและประมาณค่าข้อมูลแบบทดสอบผสมที่มีความยาว ความยาก และการคะแนนแตกต่างด้วยโมเดลโลจิสติกสามพารามิเตอร์ร่วมกับโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิต

ผลจากการศึกษาครั้งนี้ทำให้ได้ข้อสารสนเทศเกี่ยวกับความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถของผู้สอบที่คำนวณจากผลต่างความสามารถแตกต่างกันโดยใช้แบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน จึงช่วยให้มีแนวทางสำหรับนำวิธีการประมาณพารามิเตอร์พร้อมกันไปใช้ศึกษาพัฒนาการการเรียนรู้ของผู้เรียน อีกทั้งช่วยให้มีแนวทางในคำนวณพัฒนาการความสามารถของการทดสอบด้วยแบบทดสอบผสมที่มีความยากแตกต่างกัน ซึ่งจะช่วยให้ลดปัญหาการเกิดความคลาดเคลื่อนของคะแนนสอบอันเนื่องมาจากการจำข้อสอบ และปัญหาเรื่องความเท่าเทียมกันของมาตรวัดสำหรับเปรียบเทียบคะแนนสอบ

ผลการศึกษาครั้งนี้เป็นการศึกษาจากกลุ่มตัวอย่าง 1,000 คนจึงทำให้มีแนวทางสำหรับการศึกษาคะแนนพัฒนาการความสามารถในการทดสอบระดับเขตพื้นที่การศึกษา ระดับภาค หรือระดับชาติ ที่ทำให้ผลการตัดสินเกี่ยวกับคะแนนพัฒนาการมีความยุติธรรมน่าเชื่อถือ อีกทั้งยังได้แนวทางในการนำแบบทดสอบผสมมาใช้ในการศึกษาเปรียบเทียบระดับพัฒนาการความสามารถของผู้สอบ ซึ่งการใช้แบบทดสอบผสมจะช่วยให้สามารถวัดความรู้ผู้สอบได้ลึกและกว้างกว่าการศึกษาคะแนนพัฒนาการที่ใช้แบบทดสอบที่มีข้อสอบเพียงลักษณะเดียว นอกจากนี้ผลการศึกษาครั้งนี้จะช่วยให้ผู้เกี่ยวข้องได้มองเห็นความสำคัญและความจำเป็นในการปรับเทียบคะแนนสอบให้อยู่ในมาตรวัดเดียวกันก่อนนำคะแนนไปคำนวณพัฒนาการสามารถ การศึกษาในครั้งนี้จะช่วยยืนยันว่าการปรับเทียบคะแนนเป็นสิ่งจำเป็นสำหรับการคำนวณพัฒนาการความสามารถ เพื่อทำให้สามารถนำพัฒนาการความสามารถที่ได้แปลความหมายหรือนำไปตัดสินคุณภาพผู้เรียนหรือการศึกษาได้ อย่างยุติธรรมและเท่าเทียม

ขอบเขตของการวิจัย

การศึกษาวิจัยในครั้งนี้เป็นการศึกษาจากข้อมูลในสถานการณ์จำลองมีขอบเขตการศึกษาดังนี้

ข้อมูลที่ใช้ศึกษาด้วยสถานการณ์จำลอง

ข้อมูลที่ศึกษาเป็นข้อมูลจำลองที่โปรแกรม WINGEN2 (Han. 2007) จำลองแบบแผนการตอบแบบทดสอบผสมของผู้สอบจำนวน 1,000 คน ที่ทดสอบซ้ำสองครั้ง โดยใช้โมเดลโลจิสติกแบบสามพารามิเตอร์ในการจำลองผลการตอบข้อสอบแบบเลือกตอบที่ตรวจให้คะแนนสองค่า และใช้โมเดลเจเนอรอลไรส์พาร์เชียลเครดิตจำลองผลการตอบข้อสอบแบบเขียนตอบที่ตรวจให้คะแนนหลายค่า โดยกำหนดลักษณะข้อมูลสำหรับจำลองแบบแผนการตอบดังนี้

1. แบบทดสอบผสมที่ใช้มีลักษณะแตกต่างกันตามความยาว ความยาก และการให้คะแนนดังนี้ ความยาวมี 3 ขนาดตามจำนวนข้อสอบแบบเลือกตอบและจำนวนข้อสอบแบบเขียนตอบที่ใช้ทั้งฉบับได้แก่ 30:10 24:8 และ 15:5 โดยมีอัตราส่วนจำนวนข้อสอบแบบให้คะแนนสองค่าต่อจำนวนข้อสอบแบบให้คะแนนหลายค่าคงที่ 3:1 ความยากของแบบทดสอบมี 3 ระดับ ได้แก่ ระดับความยากเฉลี่ย -0.5 0.0 และ 0.5 และการให้คะแนนข้อสอบแบบเขียนตอบ 3 รูปแบบ ได้แก่ 3 ระดับ (0/1/2 คะแนน) 4 ระดับ (0/1/2/3 คะแนน) และ 5 ระดับ (0/1/2/3/4 คะแนน) รวมแบบทดสอบผสมที่ใช้ศึกษาทั้งสิ้น 27 ฉบับ (3×3×3)

2. การกำหนดการแจกแจงความสามารถของผู้สอบ ในการสอบครั้งที่หนึ่งให้ความสามารถของผู้สอบมีการแจกแจงปกติ ค่าเฉลี่ยเป็น 0.00 และ ส่วนเบี่ยงเบนมาตรฐานเป็น 1 และในการสอบครั้งที่สองผู้สอบมีพัฒนาการความสามารถเฉลี่ย 0.1 0.5 และ 1 หน่วย และมีส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 1.0 และ 1.2 ตามลำดับ รวมลักษณะความสามารถผู้สอบทั้งสิ้น 10 กลุ่ม [1+ (3×3) =10]

3. การจำลองผลแบบแผนการตอบของผู้สอบ 1,000 คน แยกตามระดับความสามารถและลักษณะของแบบทดสอบผสม รวมแบบแผนการตอบที่จำลองด้วยโปรแกรม WINGEN2 (Han. 2007) ทั้งสิ้น 243 ชุด ทำซ้ำเงื่อนไขละ 50 รอบ

4. แบบแผนการตอบแบบทดสอบผสมที่ได้ทุกคู่การทดสอบซ้ำ จะถูกประมาณค่าพารามิเตอร์ความสามารถผู้สอบโดยใช้วิธีการประมาณค่าพารามิเตอร์แบบพร้อมกัน (Concurrent Calibration) ด้วยโมเดลผสมระหว่างโมเดลโลจิสติกแบบสามพารามิเตอร์กับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิต รวมข้อมูลที่ใช้วิธีการประมาณค่าพารามิเตอร์แบบพร้อมทั้งหมด 243 คู่

5. การตรวจสอบความแม่นยำของค่าประมาณความสามารถ พิจารณาจากค่าความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าประมาณความสามารถของผู้สอบแต่ละคน เมื่อเทียบกับค่าความสามารถจริง และการตรวจสอบความแม่นยำ

ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันพิจารณาจากค่าความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าประมาณพัฒนาการความสามารถที่ได้แต่ละครั้งที่ทำซ้ำในแต่ละเงื่อนไขกับค่าเฉลี่ยพัฒนาการความสามารถจริง

ตัวแปรที่ศึกษา

1. **ตัวแปรอิสระ** มี 4 ตัวแปร ได้แก่

1.1 ระดับพัฒนาการความสามารถ 9 ระดับ

1.1.1 ค่าเฉลี่ยความสามารถเพิ่มขึ้น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 1

1.1.2 ค่าเฉลี่ยความสามารถเพิ่มขึ้น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1

1.1.3 ค่าเฉลี่ยความสามารถเพิ่มขึ้น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1

1.1.4 ค่าเฉลี่ยความสามารถเพิ่มขึ้น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8

1.1.5 ค่าเฉลี่ยความสามารถเพิ่มขึ้น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8

1.1.6 ค่าเฉลี่ยความสามารถเพิ่มขึ้น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8

1.1.7 ค่าเฉลี่ยความสามารถเพิ่มขึ้น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2

1.1.8 ค่าเฉลี่ยความสามารถเพิ่มขึ้น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2

1.1.9 ค่าเฉลี่ยความสามารถเพิ่มขึ้น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2

1.2 ความยาวของแบบทดสอบ

1.2.1 ขนาดยาว (แบบให้คะแนนสองค่า 30 ข้อ และ แบบให้คะแนนหลายค่า 10 ข้อ)

1.2.2 ขนาดปานกลาง (แบบให้คะแนนสองค่า 24 ข้อ และแบบให้คะแนนหลายค่า 8 ข้อ)

1.2.3 ขนาดสั้น (แบบให้คะแนนสองค่า 15 ข้อ และแบบให้คะแนนหลายค่า 5 ข้อ)

1.3 ระดับความยากของข้อสอบ

1.3.1 ต่ำกว่าระดับความสามารถของผู้สอบในการทดสอบครั้งที่ 1

1.3.2 เท่ากับระดับความสามารถของผู้สอบในการทดสอบครั้งที่ 1

1.3.3 สูงกว่าระดับความสามารถของผู้สอบในการทดสอบครั้งที่ 1

1.4 การให้คะแนนข้อสอบแบบให้คะแนนหลายค่า

1.4.1 3 ระดับ (0 1 และ 2 คะแนน)

1.4.2 4 ระดับ (0 1 2 และ 3 คะแนน)

1.4.3 5 ระดับ (0 1 2 3 และ 4 คะแนน)

ตัวแปรตาม

1. ความแม่นยำของค่าประมาณความสามารถ
2. ความแม่นยำของวิธีประมาณค่าพารามิเตอร์พร้อมกัน

กรอบแนวคิดในการวิจัย

การวิจัยครั้งนี้เป็นการตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันโดยใช้โมเดลโลจิสติกแบบสามพารามิเตอร์กับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิต เมื่อพัฒนาการความสามารถของนักเรียนเพิ่มขึ้นต่างกัน และแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน โดยศึกษาภายใต้สถานการณ์จำลอง ซึ่งผู้วิจัยได้ศึกษางานวิจัยและกำหนดกรอบแนวคิดไว้ดังนี้

วิธีการสร้างมาตรวัดร่วม

การศึกษาในครั้งนี้เป็นการตรวจสอบความแม่นยำของวิธีประมาณค่าพารามิเตอร์พร้อมกันเพื่อศึกษาพัฒนาการความสามารถของนักเรียน($\theta_2 - \theta_1$) ที่ทำการทดสอบซ้ำด้วยแบบทดสอบผสมที่มีระดับความยากแตกต่างกัน ซึ่งวิธีการสร้างมาตรวัดร่วมเพื่อให้สามารถเปรียบเทียบพารามิเตอร์ความสามารถได้นั้นมี 3 วิธีที่นิยมใช้ ได้แก่ วิธีการประมาณค่าพารามิเตอร์แยกกัน (Separate calibration) วิธีการประมาณค่าพารามิเตอร์ที่กำหนดพารามิเตอร์ของข้อสอบร่วม (Fixed-item parameter calibration) และวิธีการประมาณค่าพารามิเตอร์พร้อมกัน (Concurrent calibration) (Kolen & Brennan, 1995) ผลการศึกษาเปรียบเทียบคุณภาพของวิธีการปรับเทียบพารามิเตอร์จากงานวิจัยที่ผ่านมา (Hanson; & Béguin. 2002; Kim. 2004; Kim; & Cohen. 1998; Yao; & Mao. 2004) โดยส่วนใหญ่พบว่า วิธีการประมาณค่าพารามิเตอร์พร้อมกัน ให้ค่าประมาณพารามิเตอร์ที่มีความคลาดเคลื่อนต่ำกว่าวิธีการประมาณพารามิเตอร์แยกกัน และสามารถใช้ประมาณค่าพารามิเตอร์สำหรับข้อมูลแบบทดสอบผสมได้ แม้ว่าผู้สอบมีความสามารถไม่เท่าเทียมกัน (Nonequivalent group) แต่ผลการศึกษาไม่สามารถสรุปว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีคุณสมบัติที่ดีกว่าวิธีการปรับเทียบอื่นอย่างสิ้นเชิง อีกทั้งยังไม่มีการศึกษาเปรียบเทียบความแม่นยำของพัฒนาการความสามารถที่ใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับพัฒนาการความสามารถจริง เมื่อใช้แบบทดสอบผสมที่แตกต่างกันตามขนาดความยาว การแจกแจงความยากและ การให้คะแนน ดังนั้นผู้วิจัยจึงต้องการตรวจสอบว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันสามารถให้ค่าประมาณพารามิเตอร์ความสามารถที่แม่นยำทุกระดับพัฒนาการความสามารถของผู้สอบหรือไม่ เมื่อแบบทดสอบผสมมีความยาว ความยากและการให้คะแนนแตกต่างกัน ตามสถานการณ์จำลองที่กำหนด

โมเดลการประมาณค่าความสามารถของข้อมูลจากแบบทดสอบผสม

แบบทดสอบผสมที่ประกอบด้วยข้อสอบที่ให้คะแนนแบบสองค่าและหลายค่า เช่น แบบทดสอบวิชาคณิตศาสตร์ที่มีแบบทดสอบเลือกตอบ 25 ข้อ หรือแสดงวิธีทำ 3 ข้อซึ่งให้คะแนนตามเกณฑ์การให้คะแนนข้อละ 4 คะแนน (0/1/2/3/4) สามารถนำข้อมูลที่ได้ไปวิเคราะห์ค่าพารามิเตอร์

ผู้สอบและข้อสอบโดยแยกตามลักษณะข้อสอบจะได้ค่าประมาณความสามารถจากโมเดลการประมาณค่าตามสองโมเดลที่เลือกใช้แล้วจึงนำค่าพารามิเตอร์ที่ได้คำนวณบนมาตรวัดร่วมสำหรับแทนที่ค่าประมาณความสามารถจากทั้งสองโมเดลให้สามารถเปรียบเทียบกันได้ อย่างไรก็ตามสามารถวิเคราะห์ผลการตอบแบบทดสอบผสมโดยใช้โมเดลการตอบข้อสอบแบบให้คะแนนสองค่าและโมเดลการตอบข้อสอบแบบให้คะแนนหลายค่าร่วมกันประมาณค่าเพื่อให้ได้ค่าประมาณความสามารถซึ่งเป็นค่าร่วมระหว่างโมเดลทั้งสองได้ (Yen; & Fitzpatrick. 2006:118) ซึ่งโมเดลที่เป็นไปได้สำหรับประมาณค่าความสามารถของผลการตอบข้อสอบแบบเลือกตอบซึ่งให้คะแนน 0 หรือ 1 นั้นมีสามโมเดลได้แก่ โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (1PL) โมเดลโลจิสติกแบบสองพารามิเตอร์ (2PL) และโมเดลโลจิสติกแบบสามพารามิเตอร์ (3PL) และโมเดลสำหรับประมาณค่าผลการตอบข้อสอบแบบให้คะแนนหลายค่าแบบบางส่วนซึ่งให้คะแนนตามเกณฑ์ (Scoring Rubrics) มี 2 โมเดลได้แก่ โมเดลพาร์เชียลเครดิต (PCM) และโมเดลเจเนอรอลไรส์พาร์เชียลเครดิต (GPCM) แต่ด้วยลักษณะทั่วไปของโมเดลที่สอดคล้องกับลักษณะทั่วไปของข้อสอบแบบเลือกตอบที่ตรวจให้คะแนนสองค่ากับข้อสอบแบบเขียนตอบที่ตรวจให้คะแนนหลายค่า ทำให้โมเดลผสมที่มีการนำไปใช้ประมาณค่าพารามิเตอร์ของแบบทดสอบผสมมาก คือโมเดลโลจิสติกแบบสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิต (Li; Lissitz; & Yang.1999; Paek; & Young. 2005; Kim; & Lee. 2004; Kim; & Kolen.2006 ; Briggs; Weeks; & Wiley. 2008) ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงกำหนดโมเดลผสมสำหรับประมาณค่าผลการตอบแบบทดสอบผสม คือโมเดลโลจิสติกแบบสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิต

ระดับพัฒนาการความสามารถของผู้สอบ

ผลการศึกษาด้วยข้อมูลจำลองของ เปค และ ยัง (Paek; & Young. 2005) พบว่าในการทดสอบพัฒนาการความสามารถด้วยแบบทดสอบผสมเมื่อใช้วิธีการประมาณค่าพารามิเตอร์แบบกำหนดค่าพารามิเตอร์ข้อสอบให้คงที่พร้อมกับกำหนดการแจกแจงความสามารถผู้สอบก่อนการประมาณค่าจะได้ค่าประมาณพารามิเตอร์ที่มีความคลาดเคลื่อนไปจากค่าความสามารถจริง โดยพบว่าเมื่อผู้สอบมีพัฒนาการ ±0.1ค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absoluted Error: MAE) และ ค่า RMSE มีค่าต่ำกว่าผู้สอบมีพัฒนาการสูงกว่าหรือต่ำกว่า ±0.1และ เมื่อพัฒนาการความสามารถมีค่าส่วนเบี่ยงเบนมาตรฐานสูงกว่าส่วนเบี่ยงเบนมาตรฐานของความสามารถเริ่มต้น (1.2) ค่า MAE และ ค่า RMSE จะมีค่าสูงกว่าเมื่อส่วนเบี่ยงเบนมาตรฐานมีค่าเท่าหรือต่ำกว่า (1 หรือ 0.8) และระดับความสามารถเฉลี่ยสองกลุ่มแตกต่างกัน 0.5 ถือว่ามากพอจะทำให้เห็นความแตกต่าง (Li; & Lissitz, 2000) ซึ่งมักพบโดยทั่วไปในการศึกษาวิธีการปรับเทียบพารามิเตอร์หลายงานวิจัย (Lin, 2008; Davey et al., 1996; Kim, 2004; Li & Lissitz, 2000; Min, 2003; Oshima et al. 2000;

Skaggs & Lissitz, 1988) ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงกำหนดระดับพัฒนาการความสามารถของผู้สอบออกเป็น 9 กลุ่ม ตามระดับพัฒนาการความสามารถที่เพิ่มขึ้น และส่วนเบี่ยงเบนมาตรฐานของพัฒนาการความสามารถที่เพิ่มขึ้นดังนี้ (0.1, 1), (0.5, 1), (1, 1), (0.1, 0.8), (0.5, 0.8), (1, 0.8), (0.1, 1.2), (0.5, 1.2), (1, 1.2)

ความยาวของแบบทดสอบผสม

ลี ลิสสิท และ หยาง (Li; Lissitz; & Yang .1999) แนะนำว่า โมเดลตามทฤษฎีการตอบข้อสอบต้องการความเป็นเอกมิติและสอดคล้องกับข้อมูล และสัดส่วนของข้อสอบแบบให้คะแนนหลายค่าต่อข้อสอบแบบให้คะแนนหลายค่าในแบบทดสอบผสมส่งผลต่อความแม่นยำของสัมประสิทธิ์การปรับเทียบได้ ในการประมาณค่าพารามิเตอร์ ความยาวของแบบทดสอบมีอิทธิพลต่อค่าประมาณความสามารถมากที่สุด ในขณะที่ขนาดกลุ่มตัวอย่างมีอิทธิพลต่อการประมาณค่าพารามิเตอร์ข้อสอบมากที่สุด (Yang. 2007) เมื่อแบบทดสอบมีขนาดยาวอย่างเพียงพอทำให้การใช้โมเดลแบบเอกมิติให้ผลการประมาณค่าพารามิเตอร์ที่แม่นยำกว่าโมเดลพหุมิติ แต่ถ้าแบบทดสอบมีขนาดสั้น การใช้โมเดลแบบเอกมิติมีความแม่นยำต่ำกว่าการใช้โมเดลแบบพหุมิติ (Zhang. 2004) แต่จากการศึกษายังไม่พบว่าควรกำหนดความยาวที่เหมาะสมของแบบทดสอบแบบผสมไว้ที่เท่าใด อย่างไรก็ตามผลการศึกษาของ คินเซย์ (Kinsey. 2003) พบว่าแบบทดสอบผสมซึ่งกำหนดความยาวตามสัดส่วนจำนวนข้อสอบสัดส่วนของข้อสอบแบบให้คะแนนหลายค่าต่อข้อสอบแบบให้คะแนนหลายค่ามีผลต่อปริมาณค่าสารสนเทศของแบบทดสอบแบบผสมอย่างมีนัยสำคัญ ส่วนการศึกษาของ ไซมอน (Simon. 2008) ใช้แบบทดสอบที่มีข้อสอบแบบให้คะแนนสองค่าเพียงอย่างเดียวจำนวน 40 ข้อ และ 60 ข้อ กลับพบว่าความยาวของแบบทดสอบมีผลต่อคุณภาพวิธีการปรับเทียบค่าพารามิเตอร์ข้อสอบเพียงเล็กน้อยเท่านั้น เพราะว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันเป็นวิธีการประมาณค่าพารามิเตอร์ของทุกกลุ่มผู้สอบพร้อมกันจึงทำให้กลุ่มตัวอย่างมีขนาดใหญ่ขึ้น ค่าพารามิเตอร์ที่ประมาณได้จึงมีความคลาดเคลื่อนต่ำแม้ว่าจำนวนข้อสอบจะมีน้อยก็ตาม

ดังนั้นการศึกษาครั้งนี้ผู้วิจัยจึงได้กำหนดจำนวนข้อสอบแต่ละประเภทตามเวลาที่สามารถกระทำได้จริง โดยไม่ฝ่าฝืนข้อตกลงเบื้องต้นในการใช้ทฤษฎีการตอบข้อสอบเรื่องแบบทดสอบที่ใช้ต้องไม่เป็นแบบจำกัดความเร็ว (Speedy Test) ซึ่ง เดอมาส์ (DeMars. 2004:8) แนะนำว่าในสถานการณ์จริงที่ผู้สอบต้องใช้เวลาสอบข้อสอบแบบเขียนตอบอย่างน้อย 5 – 10 นาทีในการทำข้อสอบ 1 ข้อ และตามบริบทการจัดการทดสอบในชั้นเรียนของประเทศไทยที่ส่วนใหญ่กำหนดให้มีการทดสอบวิชาคณิตศาสตร์ประมาณ 2 - 3 ชั่วโมง ผู้วิจัยจึงกำหนดสัดส่วนของจำนวนข้อสอบแบบเลือกตอบต่อแบบแสดงวิธีทำเป็น 30:10 ส่วนแบบทดสอบขนาดกลาง และ ขนาดสั้นกำหนดตามสัดส่วนคงที่เท่ากันเป็น 24:8 และ 15 :5 ข้อ ตามลำดับ

การให้คะแนนข้อสอบเขียนตอบ

คินเซย์ (Kinsey. 2003) พบว่าสัดส่วนน้ำหนักคะแนนของประเภทข้อสอบในแบบทดสอบผสมส่งผลต่อค่าสารสนเทศของแบบทดสอบ โดยเฉพาะเมื่อให้แบบทดสอบผสมทุกข้อมีการให้คะแนนแบบหลายค่า (100 %) ค่าสารสนเทศของแบบทดสอบจะมีค่าสูงกว่ารูปแบบอื่นทั้งหมด และปาร์ก (Park. 2000) พบว่าจำนวนของรายการคำตอบมากขึ้นมีผลต่อการเพิ่มความแม่นยำค่าประมาณความสามารถที่ใช้โมเดลเจเนอรอลไรส์พาร์เชียลเครดิต ในขณะที่ โจเดียนและคณะ (Jodoin; et al. 2003) พบว่าค่าประมาณพารามิเตอร์ของข้อสอบแบบเลือกตอบและเติมคำที่ให้คะแนนแบบสองค่าอยู่สูงกว่าเส้นตรงการปรับเทียบในขณะที่ค่าประมาณพารามิเตอร์ของข้อสอบแบบแสดงวิธีทำที่ให้คะแนนแบบหลายค่าอยู่ต่ำกว่า แสดงว่ารูปแบบของแบบทดสอบที่เปลี่ยนไปตามน้ำหนักการให้คะแนนข้อสอบและลักษณะข้อสอบน่าจะมีผลต่อความแม่นยำของค่าประมาณพารามิเตอร์ได้ (Jodoin; et al. 2003: 248) และความคงเส้นคงวาของค่าประมาณพารามิเตอร์ของข้อสอบเขียนอธิบายคำตอบที่ได้จากการวิธีการปรับเทียบโดยใช้ความยากของข้อสอบเลือกตอบและความยากประจำขั้นของข้อสอบเขียนอธิบายคำตอบมีน้อยกว่า ค่าประมาณพารามิเตอร์ที่ได้จากการวิธีการปรับเทียบโดยใช้ความยากของข้อสอบเลือกตอบและความยากประจำข้อของข้อสอบเขียนอธิบายคำตอบ (ซึ่งถือว่าเป็นการกำหนดน้ำหนักของข้อสอบแต่ละข้อเท่ากัน) เทท (Tate. 2000) ได้แนะนำว่าจำนวนรายการคำตอบมากขึ้นความแปรปรวนของความคลาดเคลื่อนจากวิธีการสร้างมาตรวัดร่วมจะลดลง ซึ่ง เทท (Tate. 2003) ได้ศึกษาโดยใช้แบบทดสอบผสมที่มีจำนวนข้อสอบเขียนตอบ 30 ข้อ และข้อสอบเขียนอธิบายคำตอบ 10 ข้อที่มีคะแนน 4 ระดับ เพื่อทำให้ข้อสอบทั้งสองประเภทมีจำนวนขั้นการตอบ 30 ขั้นเท่ากัน ส่วน คิม (Kim. 2004, Kim; & Kolen. 2006) กำหนดขนาดแบบทดสอบผสมเป็น 36:15 โดยให้ข้อสอบแบบเขียนตอบ มีคะแนน 3 ระดับ (0/1/2) จำนวน 9 ข้อ และมีคะแนน 4 ระดับ (0/1/2/3) จำนวน 6 ข้อ เพื่อทำให้คะแนนรวมของข้อสอบทั้งสองลักษณะมี 36 คะแนนเท่ากัน

ดังนั้นเพื่อให้น้ำหนักคะแนนที่ได้จากข้อสอบให้คะแนนแบบสองค่าและข้อสอบแบบให้คะแนนแบบหลายค่ามีคะแนนรวมแตกต่างกันผู้วิจัยจึงได้กำหนดระดับคะแนนของข้อสอบเขียนอธิบายคำตอบเป็น 3 ระดับ (0/1/2) 4 ระดับ (0/1/2/3) และ 5 ระดับ(0/1/2/3/4) เพื่อให้สัดส่วนน้ำหนักคะแนนรวมของข้อสอบแบบให้คะแนนสองค่าสูงกว่าข้อสอบแบบให้คะแนนแบบหลายค่า (30:20, 24:16, 15:10) ข้อสอบแบบให้คะแนนสองค่าเท่ากับข้อสอบแบบให้คะแนนแบบหลายค่า (30:30, 24:24, 15:15) และ ข้อสอบแบบให้คะแนนสองค่าต่ำกว่าข้อสอบแบบให้คะแนนแบบหลายค่า (30:40, 24:32, 15:20) ตามลำดับ

ความยากของข้อสอบ

พารามิเตอร์ความยากของข้อสอบในการทดสอบครั้งแรกมีความลำเอียงต่อคะแนนพัฒนาการที่เรียกว่า “ความบิดเบือนของมาตรวัด (Scale distortion)” (May; & Nicewander. 1998) ทำให้คะแนนดิบครั้งที่ 2 เมื่อเทียบกับคะแนนดิบครั้งที่ 1 เพิ่มขึ้นแตกต่างกันตามลักษณะกลุ่มพัฒนาการที่ศึกษาที่มีคะแนนเริ่มต้นที่แตกต่างกัน ผลต่างของค่าประมาณความสามารถที่ประมาณค่าด้วยวิธี EAP ทำให้มาตรวัดบิดเบือนเพียงเล็กน้อย และ ลิน (Lin. 2008: 97) ศึกษาผลการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันด้วยโมเดลพหุมิติเมื่อระดับความยากเฉลี่ยของแบบทดสอบสองฉบับหลังสูงกว่าฉบับแรกเป็น 0.5 และมีค่าส่วนเบี่ยงเบนมาตรฐานเป็น 1 พบว่าค่าประมาณพารามิเตอร์ความสามารถมีความลำเอียงสูงกว่ากรณีความยากของแบบทดสอบสองฉบับเท่าเทียมกัน และผลการวิจัยยังปรากฏว่าการประมาณค่าโดยใช้โมเดลแบบเอกมิติยังมีความแข็งแกร่งต่อการละเมิดข้อตกลงเบื้องต้นด้วยเช่นกัน ซึ่งการกำหนดให้ค่าพารามิเตอร์ข้อสอบเพิ่มขึ้น 0.5 นั้นสอดคล้องการศึกษาของ เกา (Cao. 2008) ที่ต้องการให้ข้อสอบมีความเป็นพหุมิติสำหรับศึกษาผลการใช้โมเดลพหุมิติ แต่ได้ศึกษาผลการใช้โมเดลแบบเอกมิติในการประมาณค่าพารามิเตอร์พร้อมกันด้วย ซึ่งผลการศึกษาพบว่าการประมาณค่าพารามิเตอร์พร้อมกันโดยใช้โมเดลแบบเอกมิติ มีความแข็งแกร่งเพียงพอในการละเมิดข้อตกลงเบื้องต้นเรื่องความเป็นเอกมิติของข้อสอบได้ ดังนั้นผู้วิจัยจึงกำหนดให้ข้อสอบสำหรับทดสอบครั้งแรกมีลักษณะการแจกแจงค่าความยากของข้อสอบ 3 ลักษณะคือ ค่าเฉลี่ยความยากของแบบทดสอบเป็น -0.50 0.00 และ 0.50 และการทดสอบครั้งที่สองให้ข้อสอบเฉพาะสำหรับทดสอบครั้งที่ 2 มีความยากเพิ่มขึ้นจากเดิมข้อละ 0.50 สำหรับข้อสอบแบบเลือกตอบ และความยากเพิ่มขึ้นขั้นละ 0.50 สำหรับข้อสอบแบบเขียนตอบ

ขนาดกลุ่มตัวอย่าง

โปรแกรมที่สามารถประมาณค่าพารามิเตอร์ของแบบทดสอบผสมได้ เช่น PARSCALE (Muraki; & Bock, 1999) หรือ MULTILOG (Thissen,1991) ซึ่งในปัจจุบันพบว่า PARSCALE ยังขาดความสามารถในการวิเคราะห์กลุ่มหลายกลุ่มที่แตกต่างกัน แต่สำหรับ MULTILOG ยอมให้มีรูปแบบกลุ่มที่ต่างกันได้ โดยความแปรปรวนของประชากรของกลุ่มที่แตกต่างกันต้องมีค่าคงที่ อย่างไรก็ตามหากใช้โปรแกรม MULTILOG (Thissen. 1991) ในการประมาณค่าพารามิเตอร์ เมื่อประมาณค่าด้วยโมเดลเกรดเรสพอนส์กับข้อสอบแบบพหุวิภาค 25 ข้อที่มี 5 ระดับคำตอบ ค่าประมาณพารามิเตอร์ได้จะมีความแม่นยำเมื่อใช้กลุ่มตัวอย่างอย่างน้อย 500 คน และถ้าต้องการให้ความแม่นยำในการประมาณค่าสูงต้องใช้ผู้สอบ 1000 ถึง 2000 คน (Rise; & Yu. 1990) เมื่อประมาณค่าผลการตอบข้อสอบแบบทวิวิภาคด้วยโมเดลสองพารามิเตอร์ และสามพารามิเตอร์กับข้อสอบมีขนาด 30 ข้อ ถ้าต้องการให้ความแม่นยำในการประมาณค่าสูง ต้องใช้ผู้สอบจำนวน 500 คน และ1000 คน ตามลำดับ

(Hulin, et al. 1983) หากทราบค่าอำนาจจำแนกและค่าพารามิเตอร์การเดามาก่อนสามารถผ่อนปรนขนาดความยาวของข้อสอบและจำนวนผู้สอบได้ (Dragow, 1989; Gifford & Swaminathan, 1990; Harwell & Janosky, 1991; Seong, 1990; Skages & Stevenson, 1989; Swaminathan & Gifford, 1986) แม้ว่าขนาดของกลุ่มตัวอย่างไม่ได้เป็นตัวแปรที่สนใจศึกษาในครั้งนี้ แต่ขนาดกลุ่มตัวอย่างมีผลต่อความแม่นยำในการประมาณค่าพารามิเตอร์ ดังนั้นผู้วิจัยจึงกำหนดให้แต่ละเงื่อนไขศึกษามีจำนวนผู้สอบครั้งละ 1000 คน

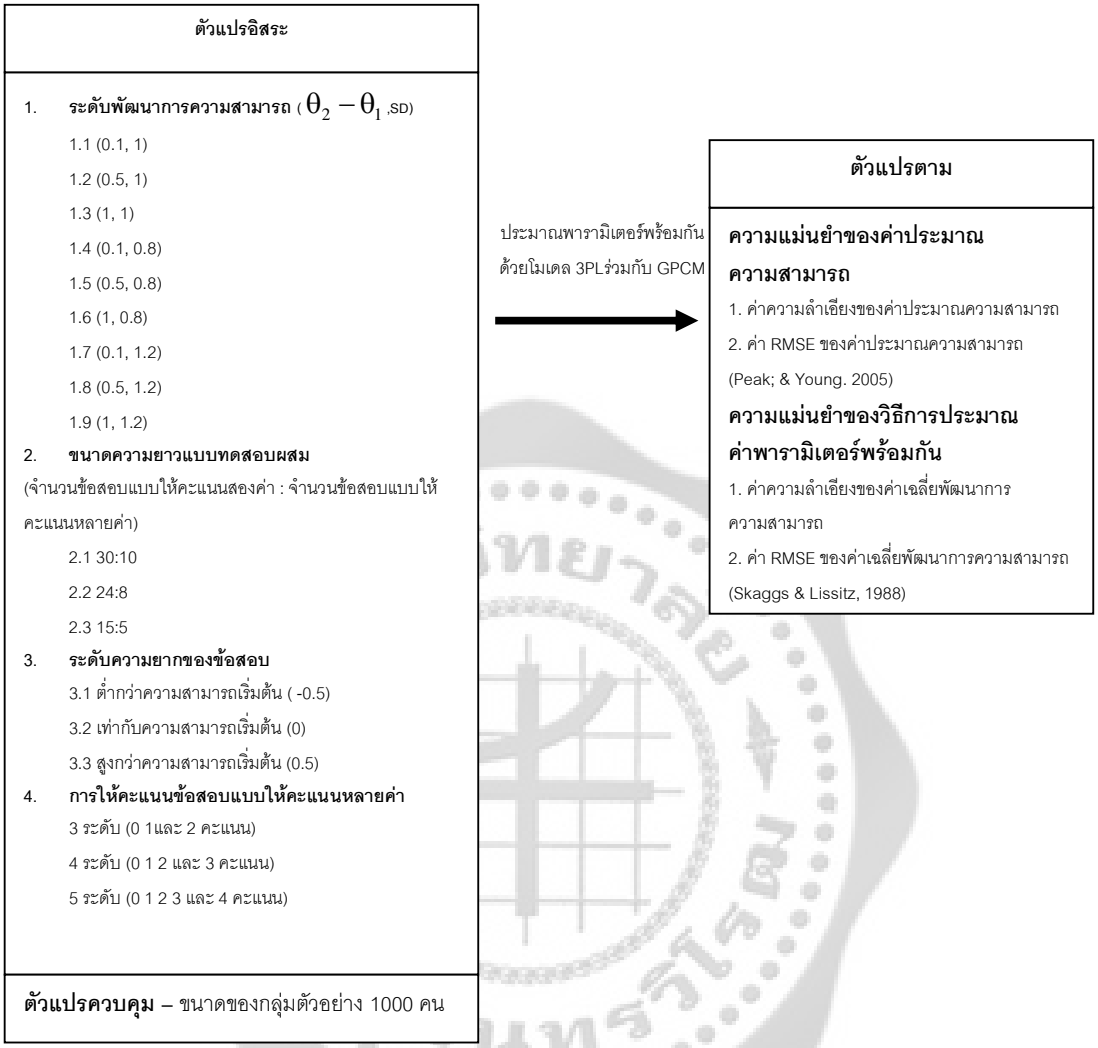
ความแม่นยำของค่าประมาณความสามารถ

ความแม่นยำของค่าประมาณความสามารถ คือ ความไม่แตกต่างกันระหว่างค่าประมาณพารามิเตอร์ความสามารถของการทดสอบแต่ละครั้งที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันเมื่อเทียบกับค่าความสามารถจริง ในแต่ละเงื่อนไขที่ศึกษา โดยพิจารณาจากความลำเอียง (Bias) และความคงที่ (Stability) ความลำเอียงคือค่าเฉลี่ยของค่าความแตกต่างระหว่างผลต่างค่าประมาณความสามารถกับผลต่างความสามารถจริงของผู้สอบแต่ละคน (Peak; & Young. 2005) ส่วนการประเมินความคงที่ (Stability) สามารถพิจารณาจากความเคลื่อนมาตรฐานในการประมาณค่า (Standard error of estimate) ด้วยการใช้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square Error: RMSE) (Peak; & Young. 2005)

ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน คือ ความไม่แตกต่างกันระหว่างค่าเฉลี่ยของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกับค่าเฉลี่ยพัฒนาการความสามารถจริง ในแต่ละเงื่อนไขที่ศึกษา เมื่อประมาณค่าพารามิเตอร์จากแบบแผนการตอบข้อสอบที่จำลองข้อมูลซ้ำหลายครั้ง โดยพิจารณาจากความลำเอียง (BIAS) และความคงที่ (Stability) ของพัฒนาการความสามารถ ความลำเอียง คือ ค่าเฉลี่ยของผลต่างระหว่างค่าเฉลี่ยพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกับค่าเฉลี่ยพัฒนาการความสามารถจริงที่ได้ในแต่ละครั้งที่ทำซ้ำ ส่วนการประเมินความคงที่ (Stability) สามารถพิจารณาจากความเคลื่อนมาตรฐานในการประมาณค่า (Standard error of estimate) ด้วยการใช้ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (Root Mean Square of Error: RMSE) ของพัฒนาการความสามารถซึ่งคำนวณจาก ค่ารากที่สองของค่าเฉลี่ยของผลต่างระหว่างค่าเฉลี่ยพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกับค่าเฉลี่ยพัฒนาการความสามารถจริงกำลังสอง ที่ได้ในแต่ละครั้งที่ทำซ้ำ (Skaggs & Lissitz. 1988)

จากที่กล่าวมาข้างต้นสรุปกรอบแนวคิดการวิจัยได้ดังภาพประกอบ 1



ภาพประกอบ 1 กรอบแนวคิดในการศึกษาความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันโดยโมเดลโลจิสติกแบบสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิต ของข้อมูลการทดสอบซ้ำด้วยแบบทดสอบผสมที่มีความยาว ความยาก และ การให้คะแนนข้อแตกต่างกัน

นิยามศัพท์เฉพาะ

1. **พัฒนาการความสามารถ** หมายถึง ผลต่างของระดับความสามารถของผู้สอบในการทดสอบครั้งที่สอง (θ_2) กับระดับความสามารถในการทดสอบครั้งที่หนึ่ง (θ_1)

2. **พัฒนาการความสามารถจริง หมายถึง** พัฒนาการความสามารถของผู้สอบที่กำหนดตามเงื่อนไขที่ศึกษา ซึ่งกำหนดจากผลต่างค่าประมาณความสามารถของการทดสอบสองครั้งด้วยสูตร $\theta_2 - \theta_1$ ในที่นี้กำหนดให้ในการสอบครั้งแรกผู้สอบมีค่าเฉลี่ยความสามารถจริงเป็น 0 และส่วน

เบี่ยงเบนมาตรฐานเป็น 1 และกำหนดระดับพัฒนาการความสามารถจริงของผู้สอบเป็น 9 ระดับตามค่าเฉลี่ยความสามารถที่เพิ่มขึ้นและค่าส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการ ได้แก่ (0.1, 1), (0.5, 1), (1, 1), (0.1, 0.8), (0.5, 0.8), (1, 0.8), (0.1, 1.2), (0.5, 1.2), (1, 1.2)

3. **ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน** หมายถึง ความไม่แตกต่างกันระหว่างค่าเฉลี่ยของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกับค่าเฉลี่ยพัฒนาการความสามารถจริงตามเงื่อนไขที่กำหนด เมื่อมีการประมาณค่าจากแบบแผนการตอบที่จำลองข้อมูลซ้ำหลายครั้ง โดยพิจารณาจากค่าความลำเอียง (Bias) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) มีค่าเข้าใกล้ศูนย์

4. **ความลำเอียง (BIAS) ของค่าเฉลี่ยของพัฒนาการความสามารถ** หมายถึง ค่าเฉลี่ยของผลต่างระหว่างค่าเฉลี่ยของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกับค่าเฉลี่ยพัฒนาการความสามารถจริง ตามเงื่อนไขที่กำหนด เมื่อมีการประมาณค่าจากแบบแผนการตอบที่จำลองข้อมูลซ้ำ 50 ครั้ง

5. **ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าเฉลี่ยของพัฒนาการความสามารถ** หมายถึง ค่ารากที่สองของค่าเฉลี่ยของผลต่างระหว่างค่าเฉลี่ยของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันกับค่าเฉลี่ยพัฒนาการความสามารถจริงตามเงื่อนไขที่กำหนด เมื่อมีการประมาณค่าจากแบบแผนการตอบที่จำลองข้อมูลซ้ำ 50 ครั้ง

6. **ความแม่นยำของค่าประมาณความสามารถ** หมายถึง ความไม่แตกต่างกันระหว่างค่าประมาณความสามารถของการทดสอบแต่ละครั้งที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันกับค่าความสามารถจริงของผู้สอบแต่ละคนตามเงื่อนไขที่กำหนด โดยพิจารณาจากค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าประมาณความสามารถ มีค่าเข้าใกล้ศูนย์

7. **ความลำเอียง (BIAS) ของค่าประมาณความสามารถ** หมายถึง ค่าเฉลี่ยของผลต่างระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันกับค่าความสามารถจริงในการทดสอบแต่ละครั้งของผู้สอบแต่ละคน ตามเงื่อนไขที่กำหนด

8. **ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าประมาณความสามารถ** หมายถึงค่ารากที่สองของค่าเฉลี่ยของผลต่างระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์แบบพร้อมกันกับค่าความสามารถจริงในการทดสอบแต่ละครั้งของผู้สอบแต่ละคนตามเงื่อนไขที่กำหนด

9. **แบบทดสอบผสม** หมายถึง แบบทดสอบซึ่งประกอบด้วยข้อสอบสองประเภทในฉบับเดียวกัน ได้แก่ ข้อสอบแบบเลือกตอบที่ให้คะแนนแบบตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน และข้อสอบเขียนตอบแบบแสดงวิธีทำที่ให้คะแนนลดหลั่นตามเกณฑ์การให้คะแนน ซึ่งในที่นี้เกณฑ์การให้คะแนนแบ่งเป็น 3 ระดับ (0 1 และ 2 คะแนน) 4 ระดับ (0 1 2 และ 3 คะแนน) และ 5 ระดับ (0 1 2 3 และ 4 คะแนน)

10. **วิธีการประมาณค่าพารามิเตอร์พร้อมกัน หมายถึง** วิธีการประมาณค่าพารามิเตอร์ที่นำคะแนนการตอบรายข้อในการทดสอบซ้ำด้วยแบบทดสอบผสมของผู้สอบแต่ละคนภายใต้เงื่อนไขที่ศึกษาประมาณค่าพารามิเตอร์ความสามารถโดยใช้โปรแกรมคอมพิวเตอร์ประมวลผลพร้อมกันครั้งเดียว

11. **โมเดลการตอบข้อสอบแบบเอกมิติของแบบทดสอบผสม หมายถึง** โมเดลการตอบข้อสอบแบบเอกมิติสำหรับประมาณค่าพารามิเตอร์ข้อสอบแบบตรวจให้คะแนนสองค่าพร้อมกับข้อสอบแบบตรวจให้คะแนนหลายค่าที่ทำให้ได้ค่าประมาณความสามารถเพียงค่าเดียว ในที่นี้ใช้โมเดลโลจิสติกแบบสามพารามิเตอร์ ร่วมกับโมเดลเจเนอรอลไวส์พาร์เชียลเครดิตในวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

12. **ความยาวของแบบทดสอบ หมายถึง** จำนวนข้อสอบแบบเลือกตอบที่ให้คะแนนสองค่าและจำนวนข้อสอบแบบเขียนตอบที่ให้คะแนนหลายค่าทั้งหมดในแบบทดสอบผสมทั้งฉบับ โดยเขียนในรูปอัตราส่วนจำนวนข้อสอบแบบเลือกตอบต่อจำนวนข้อสอบแบบเขียนตอบ ในการศึกษาครั้งนี้ผู้วิจัยกำหนดความยาวแบบทดสอบออกเป็นสามขนาดคือขนาด 30:10 ขนาด 24:8 และ ขนาด15:5 โดยแบบทดสอบผสมแต่ละขนาดมีระดับความยากเฉลี่ยเป็น 0.00 -0.50 และ 0.50 และมีระดับคะแนนข้อสอบเขียนตอบเป็น 3 ระดับ 4 ระดับ และ 5 ระดับ

13. **ความยากของข้อสอบ** หมายถึง การแจกแจงความยากของข้อสอบทั้งฉบับที่อยู่บนมาตรวัดความสามารถตามทฤษฎีการตอบข้อสอบของแบบทดสอบฉบับที่ 1 เมื่อเทียบกับระดับความสามารถเฉลี่ยของผู้สอบในการทดสอบครั้งที่ 1 ซึ่งมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 โดยในการศึกษาครั้งนี้แบ่งระดับความยากของข้อสอบเป็น 3 ระดับ ได้แก่ ค่าเฉลี่ยความยากของข้อสอบแบบทดสอบฉบับที่ 1 ต่ำกว่าระดับความสามารถเฉลี่ยของผู้สอบในการทดสอบครั้งที่ 1 (-0.5) ค่าเฉลี่ยความยากของข้อสอบแบบทดสอบฉบับที่ 1 เท่ากับระดับความสามารถเฉลี่ยของผู้สอบในการทดสอบครั้งที่ 1 (0.0) และค่าเฉลี่ยความยากของข้อสอบแบบทดสอบฉบับที่ 1 สูงกว่าระดับความสามารถเฉลี่ยของผู้สอบในการทดสอบครั้งที่ 1 (0.5)

14. **การให้คะแนน** หมายถึง การกำหนดระดับคะแนนของข้อสอบแบบให้คะแนนหลายค่าแต่ละข้อซึ่งมีเริ่มจากคะแนนเป็นศูนย์ (0) จนถึงคะแนนสูงสุดตามเกณฑ์การตรวจให้คะแนน ในการ

วิจัยครั้งนี้ กำหนดการให้คะแนนข้อสอบเขียนตอบแบ่งเป็น 3 ระดับ (0/1/2) 4 ระดับ (0/1/2/3) และ 5 ระดับ (0/1/2/3/4)

15. **ความสามารถเริ่มต้นของผู้สอบ** หมายถึง การแจกแจงความสามารถของผู้สอบบนเส้นความสามารถต่อเนื่องที่ใช้ในการทำแบบทดสอบฉบับที่ 1 ซึ่งในการศึกษาครั้งนี้ กำหนดให้ความสามารถของผู้สอบในการทดสอบครั้งที่ 1 มีค่าเฉลี่ยเป็น 0 และ ส่วนเบี่ยงเบนมาตรฐานเป็น 1

สมมติฐานของการวิจัย

ในการวิจัยภายใต้สถานการณ์จำลองครั้งนี้ผู้วิจัยใช้โมเดลโลจิสติกแบบสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตในวิธีการประมาณค่าพารามิเตอร์พร้อมกันด้วยแบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน เพื่อตรวจสอบความแม่นยำของวิธีการประมาณพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถแตกต่างกัน โดยเปรียบเทียบค่าประมาณพารามิเตอร์ความสามารถกับความสามารถจริงที่กำหนดตามระดับพัฒนาการความสามารถที่ศึกษา แยกตามความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสม ซึ่งผู้วิจัยได้ตั้งสมมติฐานการวิจัยไว้ดังนี้

1. ค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความสัมพันธ์ระดับสูงกับค่าความสามารถจริงที่กำหนดในทุกระดับพัฒนาการความสามารถ โดย

1.1 แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีความสัมพันธ์สูงกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ในทุกระดับการแจกแจงความยาก และ ระดับการให้คะแนนของแบบทดสอบผสม

1.2 แบบทดสอบผสมที่ระดับความยากของข้อสอบสอดคล้องกับระดับความสามารถผู้สอบ จะมีความสัมพันธ์สูงกว่าแบบทดสอบผสมที่ระดับความยากของข้อสอบไม่สอดคล้องกับระดับความสามารถผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนนของแบบทดสอบผสม

1.3 แบบทดสอบผสมที่มีระดับคะแนนข้อสอบแบบให้คะแนนหลายค่ามากกว่าจะมีความสัมพันธ์สูงกว่าแบบทดสอบผสมที่มีระดับคะแนนข้อสอบแบบให้คะแนนหลายค่าน้อยกว่า ในทุกขนาดความยาวและระดับความยากของแบบทดสอบผสม

2. ค่าประมาณพารามิเตอร์ความสามารถของผู้สอบที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) เข้าใกล้ศูนย์ ทุกระดับพัฒนาการความสามารถของผู้สอบ โดย

2.1 แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ทุกระดับความยาก และ ทุกระดับการให้คะแนนแบบทดสอบผสม

2.2 แบบทดสอบผสมที่มีระดับความยากสอดคล้องกับระดับความสามารถผู้สอบจะมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่ระดับความยากไม่สอดคล้องกับระดับความสามารถผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนนของแบบทดสอบผสม

2.3 แบบทดสอบผสมที่มีระดับการให้คะแนนมากกว่าจะมีค่าความลำเอียง (BIAS) และค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่ากว่าแบบทดสอบผสมที่มีระดับการให้คะแนนน้อยกว่า ในทุกขนาดความยาวและระดับความยากของข้อสอบของแบบทดสอบผสม

3. เมื่อระดับพัฒนาการความสามารถของผู้สอบมีส่วนเบี่ยงเบนมาตรฐานเท่ากับหรือต่ำกว่าส่วนเบี่ยงเบนมาตรฐานของความสามารถครั้งแรก (1หรือ 0.8) ค่าประมาณความสามารถจากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตมีความแม่นยำมากกว่าเมื่อส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถของผู้สอบสูงกว่าส่วนเบี่ยงเบนมาตรฐานของความสามารถครั้งแรก (1.2) โดย

3.1 แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีความแม่นยำมากกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ในทุกระดับความยาก และ ระดับการให้คะแนนของแบบทดสอบผสม

3.2 แบบทดสอบผสมที่มีระดับความยากสอดคล้องกับระดับความสามารถเริ่มต้นของผู้สอบ จะมีความแม่นยำมากกว่าแบบทดสอบผสมที่มีระดับความยากไม่สอดคล้องกับระดับความสามารถเริ่มต้นของผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนนของแบบทดสอบผสม

3.3 แบบทดสอบผสมที่มีระดับการให้คะแนนข้อสอบเขียนตอบสูงกว่า จะมีความแม่นยำมากกว่าแบบทดสอบผสมที่มีระดับการให้คะแนนข้อสอบเขียนตอบต่ำกว่า ในทุกขนาดความยาวและความยากของแบบทดสอบผสม

บทที่ 2
เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎีต่างๆ จากเอกสารและงานวิจัยที่เกี่ยวข้องจึงได้เรียบเรียงและนำเสนอตามหัวข้อต่อไปนี้

1. แบบทดสอบผสม
2. การคำนวณคะแนนการเปลี่ยนแปลงรายบุคคล
3. ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT)
4. วิธีการปรับเทียบพารามิเตอร์ตามทฤษฎีการตอบข้อสอบ
5. งานวิจัยที่เกี่ยวข้อง

แบบทดสอบผสม

ในการศึกษาครั้งนี้เป็นการตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสำหรับการประเมินพัฒนาการความสามารถของผู้สอบในแบบทดสอบผสมที่มีความยาว ความยากและการให้คะแนนแตกต่างกัน แบบทดสอบผสมในงานวิจัยครั้งนี้หมายถึงแบบทดสอบที่ประกอบด้วยข้อสอบแบบเลือกตอบที่ตรวจให้คะแนนแบบสองค่า (ตอบผิดได้ 0 ตอบถูกได้ 1 คะแนน) และข้อสอบแบบเขียนตอบที่ตรวจให้คะแนนแบบหลายค่า (ตอบผิดได้ 0 ตอบถูกได้คะแนน 1 2 3 หรือ 4 คะแนนตามเงื่อนไขที่กำหนด) สำหรับการวัดประเมินพัฒนาการการเรียนรู้ของนักเรียน

มีการศึกษาวิจัยเปรียบเทียบเพื่อศึกษาเกี่ยวกับความเหมือน ความต่าง จุดสำคัญหรือข้อด้อยของการใช้ข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบหลายงานวิจัยขึ้นอยู่กับว่าในการศึกษานั้นพิจารณาจากมุมมองทางการวัดทางจิตวิทยาหรือมุมมองทางด้านการรู้คิด ในมุมมองทางด้านการวัดทางจิตวิทยาการวิจัยส่วนใหญ่มักศึกษาเกี่ยวกับการเปรียบเทียบคุณลักษณะของข้อสอบทั้งสองรูปแบบในเรื่องของการวัดความรู้เนื้อหาได้ครอบคลุมเนื้อหาตามวัตถุประสงค์ ความเชื่อมั่นในการวัด ความเที่ยงตรง ความเป็นปรนัยในการตรวจให้คะแนน และประโยชน์หรือประสิทธิภาพในการใช้ เป็นต้น สำหรับมุมมองด้านการวัดการรู้คิดนั้นงานวิจัยก็มุ่งศึกษาเปรียบเทียบว่าข้อสอบลักษณะใดสามารถวัดความการรู้คิดได้ดีกว่าชัดกว่า ซึ่ง เกา (Cao. 2008) ได้สรุปความแตกต่างระหว่างข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบไว้ 7 ประการ (Cao. 2008: 10 – 13) ดังนี้

1. ภายใต้เวลาที่กำหนดข้อสอบแบบเลือกตอบวัดความรู้ความสามารถได้ครอบคลุมเนื้อหาที่ต้องการวัด แต่สำหรับข้อสอบแบบเขียนตอบนั้นผู้สอบต้องการเวลามากกว่าเพื่อสร้างหรือ

เขียนคำตอบด้วยตนเอง ดังนั้นโดยทั่วไปจึงพบว่าจำนวนข้อสอบแบบเขียนตอบในแบบทดสอบฉบับหนึ่งจะมีจำนวนไม่มาก ซึ่งการจำกัดจำนวนข้อสอบนี้ทำให้เกิดปัญหาการเป็นตัวแทนที่ดีของการวัดได้ครอบคลุมวัตถุประสงค์การเรียนรู้ได้ (Kolen. 1999 – 2000; Linn. 1995) เช่น กรณีที่แบบทดสอบฉบับหนึ่งมีข้อสอบแบบเขียนตอบเพียงจำนวนข้อ 10 ข้อหรือน้อยกว่า ลักษณะเนื้อหาที่นำมาออกข้อสอบจึงเป็นตัวกำหนดลักษณะของคะแนน หากข้อสอบข้อนั้นบังเอิญถามในส่วนที่ผู้สอบรู้คำตอบข้อนั้นเป็นอย่างดี แต่อาจเป็นส่วนที่รู้และเข้าในเนื้อหาเพียงบางส่วน ซึ่งตรงกับคำถามในข้อสอบพอดี จึงทำให้ได้คะแนนสอบสูงกว่าความเป็นจริง ในขณะเดียวกันหากผู้สอบขาดความรู้หรือความเข้าใจบางส่วนซึ่งเป็นส่วนที่ตรงกับคำถามในข้อสอบพอดี หรือข้อสอบมีลักษณะที่ผู้สอบไม่คุ้นเคยหรือไม่มีความรู้เลย ผู้สอบก็จะได้คะแนนต่ำกว่าความเป็นจริงมาก การที่คะแนนสอบที่ได้สูงหรือต่ำกว่าสิ่งที่ต้องการวัดจริงๆ จึงทำให้ไม่สามารถแปลความหมายได้ว่าเป็นคะแนนที่ได้มาจากความรู้ความสามารถทั้งหมดที่ต้องการวัด แบบทดสอบนั้นจึงมีความเชื่อมั่นลดลงซึ่งจะมีผลต่อการลงสรุปอ้างอิงไป

2. การตรวจให้คะแนนข้อสอบแบบเลือกตอบทำได้ง่าย ประหยัด และมีความเป็นปรนัยสูง ในขณะการให้ตรวจคะแนนที่ข้อสอบแบบเขียนตอบขึ้นอยู่กับกลุ่มผู้สอบ ผู้ตรวจ และรายละเอียดเกณฑ์ให้คะแนน (Scoring rubrics) จึงทำให้การตรวจใช้เวลามาก รวมถึงค่าใช้จ่ายสูง แม้ปัจจุบันจะมีการทดลองตรวจให้คะแนนโดยใช้คอมพิวเตอร์แต่ระบบการใช้ยังไม่ได้รับความความนิยมอย่างแพร่หลาย อย่างไรก็ตามแม้ว่าผู้ตรวจจะถูกฝึกการตรวจให้คะแนนเป็นอย่างดี เกณฑ์การให้คะแนนชัดเจน มีการควบคุมขั้นตอนการให้คะแนนอย่างเข้มงวด ข้อสอบแบบเขียนตอบข้อสอบยังคงมีความเป็นอัตนัยในการตรวจสูง และแตกต่างตามลักษณะของผู้ตรวจ ตลอดจนสถานการณ์ในการตรวจขณะนั้น

3. ข้อสอบแบบเขียนตอบมีความเที่ยงตรงด้านเนื้อหามากกว่าข้อสอบแบบเลือกตอบสามารถวัดเนื้อหาและทักษะได้ตรงตามวัตถุประสงค์มากกว่าข้อสอบแบบเลือกตอบ เช่นหากต้องการวัดความสามารถในการคิดได้หลากหลาย การนำเสนอคำตอบหรือการสร้างคำตอบจะเป็นตัวกำหนดโครงสร้างการวัดการคิดได้หลากหลาย ดังนั้นข้อสอบแบบเขียนตอบที่ให้อิสระในการตอบมากกว่าข้อสอบแบบเลือกตอบจึงสามารถวัดการคิดได้หลากหลายได้ตรงตามโครงสร้างมากกว่า นอกจากนั้นข้อสอบแบบเลือกตอบยังมีความไวหรืออ่อนไหวต่อวิธีการนำไปใช้ เช่นในการทดสอบวัดสติปัญญาข้อสอบแบบเลือกตอบจะมีความไวต่อการใช้กลยุทธ์กำจัดตัวเลือก (Burton. 2001) โดยที่ผู้สอบสามารถใช้ร่องรอยเพื่อไปหาคำตอบโดยปราศจากความรู้จริง จึงเป็นการเพิ่มโอกาสการเดาคำตอบถูก ถ้าคะแนนสอบที่ได้เป็นคะแนนได้มาจากความสามารถด้านการใช้กลยุทธ์นี้บางส่วนแล้วย่อมเกิดความแปรปรวนของคะแนนสูงและมีความเที่ยงตรงเชิงโครงสร้างลดลง

4. ข้อสอบแบบเขียนตอบต้องการให้ผู้สอบสร้างหรือเขียนคำตอบด้วยตนเอง ในขณะที่ข้อสอบแบบเลือกตอบจะต้องเลือกคำตอบตามตัวเลือกที่กำหนดให้ การให้ผู้สอบเขียนคำตอบเองจึงช่วยลดปัญหาการเดาสุ่ม ผู้สอบที่ไม่มีความรู้จริงจึงไม่สามารถสร้างเขียนคำตอบที่ถูกต้องได้ ในทางกลับกันข้อสอบแบบเลือกตอบเป็นข้อสอบที่ง่ายในการเดาคำตอบ ผลการเดาจึงทำให้ได้คะแนนสูงกว่าความสามารถจริง

5. ข้อสอบทั้งสองลักษณะสามารถวัดทักษะการรู้คิดได้ทั้งทักษะการรู้คิดระดับต่ำและระดับสูง ข้อสอบแบบเลือกตอบมักถูกใช้วัดทักษะขั้นต่ำ เช่น การวัดความรู้ ความจำ แต่มีหลายงานวิจัย (Hamilton; Nussbaum; & Snow, 1997; Wainer; & Thissen, 1993) ที่แสดงให้เห็นว่าข้อสอบแบบเลือกตอบสามารถวัดเนื้อหาระดับสูงเช่น ความรู้ความเข้าใจ การประเมิน และทักษะการแก้ปัญหาได้ ส่วนข้อสอบแบบเขียนตอบเป็นข้อสอบที่สามารถวัดความรู้ความสามารถได้ครบทุกทักษะอีกทั้งมีความหลากหลายตามลักษณะการตอบข้อสอบเช่น การเติมคำตอบ การสร้างประโยคที่สมบูรณ์ การตอบสั้น การเขียนบทความหลายหน้า การแก้ปัญหาคณิตศาสตร์หลายขั้นตอน ข้อสอบแบบเขียนตอบจึงดึงทักษะความสามารถทั้งหมดได้อย่างมาก

6. แม้ว่าข้อสอบแบบเลือกตอบจะสามารถวัดความรู้ที่ซับซ้อนได้ แต่ช่วงของความรู้ในข้อสอบแต่ละข้อนั้นข้อสอบเขียนตอบจะวัดได้กว้างกว่า เช่นกรณีการสอบวัดความสามารถทางด้านภาษาการฟัง การอ่าน หรือการเขียน ซึ่งไม่อาจจะวัดความสามารถด้านการพูดได้พร้อมกันได้โดยตรง ข้อจำกัดอีกประการของข้อสอบเลือกตอบคือเป็นการวัดความรู้ที่ต้องนำไปสู่คำตอบถูกเพียงข้อเดียวจากตัวเลือกหลายข้อดังนั้นการวัดความหลากหลายจึงถูกจำกัดโดยปริยาย การทดสอบโดยใช้ข้อสอบแบบเลือกตอบเพื่อวัดความคิดสร้างสรรค์ในการสร้างหรือประดิษฐ์สิ่งใหม่จึงเป็นไปได้ยาก ดังนั้นข้อสอบแบบเลือกตอบจึงไม่สามารถวัดทักษะการรู้คิดที่สูงได้อย่างสมบูรณ์เพียงพอเมื่อเปรียบเทียบกับข้อสอบแบบเขียนตอบ

7. ข้อสอบแบบเขียนตอบอาจจะช่วยให้ผู้เรียนเกิดการเรียนรู้และให้ข้อมูลที่เป็นประโยชน์ต่อการสอนของครูได้มากกว่าข้อสอบแบบเลือกตอบ ข้อสอบแบบเขียนตอบสามารถขยายความของคำตอบได้มากกว่าข้อสอบแบบเลือกตอบ จึงทำให้มีร่องรอยของขั้นตอนหาคำตอบของนักเรียนเพื่อให้ครูสามารถวินิจฉัยความคลาดเคลื่อนในการเรียนรู้และเข้าใจผิดของนักเรียนได้จากข้อความหรือร่องรอยในการเขียนคำตอบ (Lukhele; Thissen; & Wainer, 1994; Martinez, 1999)

สรุปความแตกต่างระหว่างข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบได้ดังตาราง 1

ตาราง 1 การเปรียบเทียบคุณลักษณะของข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบ

คุณลักษณะ	ข้อสอบแบบเลือกตอบ	ข้อสอบแบบเขียนตอบ
ด้านการวัดทางจิตวิทยา (Psychometric Perspective)		
การเป็นตัวแทนของเนื้อหาในหน่วยเวลาเดียวกัน	วัดได้ครอบคลุม	วัดได้บางช่วง
ความเชื่อมั่น (Reliability)	โดยทั่วไปมีค่าสูง	โดยทั่วไปมีค่าต่ำ
การให้คะแนน (Scoring)	เป็นปรนัย มีประสิทธิภาพ และเสียค่าใช้จ่ายไม่มาก	เป็นอัตนัย ใช้เวลาตรวจมาก และเสียค่าใช้จ่ายสูง
ความเที่ยงตรง (Validity)	โดยทั่วไปมีค่า ต่ำ	โดยทั่วไปมีค่า สูง
ความทนทานต่อการเดา(Robustness to guessing)	ต่ำ	สูง
ด้านความรู้ความคิด (Cognitive Perspective)		
ทักษะการรู้คิดCognitive skills	ระดับต่ำ หรือ ระดับสูง	ระดับต่ำ หรือ ระดับสูง
ช่วงพิสัยของการรู้คิด Cognitive range	แคบ	กว้าง
ประโยชน์ในการเรียนรู้และการสอน Facility in learning and teaching	โดยทั่วไปต่ำ	โดยทั่วไปสูง

ที่มา: เกา (Cao. 2008: 13 .online)

จากลักษณะเด่นและข้อจำกัดของข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบจะพบว่าการรวมข้อสอบที่มีรูปแบบการให้คะแนนต่างกันไว้ในฉบับเดียวกันนั้นเป็นการนำสิ่งดี การผสมข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบในแบบทดสอบฉบับเดียวกันเป็นการเพิ่มจุดแข็งและชดเชยข้อด้อยของข้อสอบแต่ละประเภทได้ โดยข้อสอบแบบเลือกตอบสามารถสุ่มเนื้อหาตัวอย่างที่ต้องการวัดได้กว้าง มีความเชื่อมั่นสูง ตรวจให้คะแนนเป็นปรนัย ในขณะที่ข้อสอบแบบเขียนอธิบายคำตอบสามารถวัดเนื้อหาและโครงสร้างได้อย่างเที่ยงตรง ยากต่อการเดา และวัดการรู้คิดที่ซับซ้อนได้ ดังนั้นหากนำแบบทดสอบผสมมาใช้ในการวัดประเมินผลซึ่งเป็นกระบวนการที่ดำเนินควบคู่ไปกับการจัดการเรียนการสอน ย่อมสามารถช่วยให้ได้ข้อมูลสำหรับวินิจฉัยความรู้พื้นฐานและทักษะที่จำเป็นสำหรับผู้เรียน ตัดสินความสำเร็จในการเรียนรู้ตามหลักสูตร และเป็นข้อมูลสารสนเทศในการวางแผนบริหารการจัดการศึกษา กำหนดนโยบายและพัฒนาหลักสูตรการศึกษาได้

ตัวอย่างการทดสอบที่ใช้แบบทดสอบผสม

การทดสอบเป็นวิธีการวัดและประเมินผลการเรียนรู้วิธีหนึ่งที่สำคัญที่จะช่วยประเมินความสามารถของผู้เรียน ตามแนวทางปฏิบัติในการวัดและประเมินผลการเรียนรู้ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน 2551 ซึ่งรูปแบบของข้อสอบที่ครูผู้สอนจะเลือกใช้ขึ้นอยู่กับ

วัตถุประสงค์ของการประเมิน ซึ่งหากใช้ข้อสอบที่มีรูปแบบที่หลากหลายย่อมส่งผลให้สามารถประเมินความสามารถของผู้เรียนได้ครอบคลุมตามวัตถุประสงค์มากขึ้น ดังนั้นในการทดสอบตั้งแต่ระดับชั้นเรียน ระดับโรงเรียน จนกระทั่งการทดสอบระดับชาติจึงได้มีการพัฒนาและนำแบบทดสอบที่ผสมที่ประกอบด้วยข้อสอบที่มีลักษณะต่างกันอย่างน้อยสองรูปแบบเพื่อทดสอบความรู้ความสามารถของผู้เรียน การทดสอบระดับชาติในประเทศเช่น การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) การทดสอบที่เกี่ยวข้องกับการศึกษาต่างประเทศเช่น The Test of English as a Foreign Language (TOEFL) Graduate Record Examination (GRE) Graduate Management Admission Test (GMAT) และ International English Language Testing System (IELTS) เป็นต้น สรุปลักษณะของการทดสอบแต่ลักษณะดังนี้

1. การทดสอบทางการศึกษาระดับชาติขั้นพื้นฐาน (O-NET) (สถาบันทดสอบทางการศึกษาแห่งชาติ. 2553)

O-NET เป็นการทดสอบเพื่อวัดความรู้และความคิดของนักเรียนระดับชั้นประถมศึกษาปีที่ 3 และ มัธยมศึกษาปีที่ 6 ตามมาตรฐานการเรียนรู้ในหลักสูตรการศึกษาขั้นพื้นฐาน พ.ศ. 2544 ครอบคลุมเนื้อหาวิชา 8 กลุ่มสาระการเรียนรู้ ได้แก่ ภาษาไทย สังคมศึกษา ศาสนาและวัฒนธรรม ภาษาอังกฤษคณิตศาสตร์วิทยาศาสตร์ สุขศึกษาและพลศึกษา การงานอาชีพและเทคโนโลยี และศิลปะ

ลักษณะของรูปแบบข้อสอบ O-NET

1) ปรนัย 4 ตัวเลือก/หลายตัวเลือก แบบเลือกคำตอบที่ถูกที่สุด 1 คำตอบ

2) ปรนัย 4 ตัวเลือก/หลายตัวเลือก แบบลือกคำตอบที่ถูกได้มากกว่า 1 คำตอบ

3) เลือกคำตอบจากกลุ่มสำพันธ์กัน

4) ระบายคำตอบที่เป็นค่าหรือตัวเลข

2. การทดสอบวัดความสามารถในการใช้ภาษาอังกฤษ (The Test of English as a Foreign Language: TOEFL)

TOEFL เป็นการทดสอบวัดความสามารถในการใช้ภาษาอังกฤษของผู้ที่ไม่ได้ใช้ภาษาอังกฤษเป็นภาษาประจำชาติ ข้อสอบที่ใช้เป็นข้อสอบแบบปรนัย เพื่อวัดความเข้าใจ ภาษาอังกฤษ (แบบอเมริกาเหนือ) ลักษณะคำถามส่วนใหญ่แล้วเป็นแบบปรนัย หรือ มีคำตอบให้เลือกตอบ แต่ปัจจุบันมีคำถามลักษณะใหม่ๆ ออกมาค่อนข้างมาก ไม่ว่าจะเป็น คำถามที่ให้เลือกภาพในการตอบ หรืออาจเป็นคำถามที่มีหลายคำตอบ หรือ ให้เรียงลำดับสิ่งของ หรือ จับคู่ให้เป็นหมวดหมู่ซึ่งปัจจุบันการสอบ TOEFL ได้เปลี่ยนมาใช้การทดสอบแบบ Computer - Based Testing แทนการทดสอบแบบ Paper test โดย เชื่อว่าวิธีนี้ จะสามารถวัดระดับ ความรู้ภาษาอังกฤษของนักเรียน ได้ถูกต้อง ใกล้เคียงความ

เป็นจริงมากกว่าการสอบแบบเดิม (Paper - Based Test) ซึ่งการทดสอบแบบ Computer - Based Testing จะแบ่งแบบทดสอบออกเป็น 4 ส่วน โดยแต่ละส่วนมีการกำหนดจำนวนคำถาม และระยะเวลาในการใช้ทำแบบทดสอบไว้ดังนี้

1) การฟัง (Listening) ทดสอบทักษะการฟังภาษาอังกฤษแบบอเมริกาเหนือ จำนวน 30-50 ข้อ ใช้เวลา40-60 นาที

2) ไวยากรณ์ (Structure) จำนวน 20-25 ข้อใช้เวลา15-20 นาที

3) การอ่าน (Reading) ทดสอบความเข้าใจในการอ่านเรื่องสั้นๆ จำนวน 44-55 ข้อ ใช้เวลา 70-90 นาที

4) การเขียน (Writing) ทดสอบความสามารถในการเขียนความเรียงในหัวข้อที่กำหนดในเวลา 30 นาที จำนวน 1 หัวข้อ

3. การทดสอบวัดเชาวน์ปัญญาทั่วไปสำหรับผู้ต้องการเรียนต่อในระดับสูงกว่าปริญญาตรี (Graduate Record Examination: GRE)

GRE เป็นการทดสอบวัดเชาวน์ปัญญาทั่วไป สำหรับผู้ต้องการเรียนต่อในระดับสูงกว่าปริญญาตรี ในประเทศสหรัฐอเมริกา ยกเว้น Business School ที่ต้องสอบ GMAT และ Law School ที่ต้องสอบ แบบทดสอบนั้นประกอบไปด้วยข้อสอบแบบเลือกตอบสามส่วนด้วยกันได้แก่ Quantitative Analytical และVerbal และอีกส่วนหนึ่งที่สอบแยกต่างหากคือ การเขียนเรียงความที่เรียกว่า Writing Assessmentในประเทศไทย ปัจจุบันได้เปลี่ยนรูปแบบการสอบ มาเป็นแบบ Computer Adaptive Test (CAT)

การสอบ GRE มีอยู่ 2 รูปแบบ คือ การสอบทั่วไป (General Test) และการสอบเฉพาะวิชาสาขา (Subject Test) ในวิชาต่างๆ 16 สาขา การสอบทั่วไป (General Test) เป็นการสอบ เพื่อวัดทักษะของผู้สอบที่มีอยู่ โดยวัดออกมาในรูปของคะแนนของความสามารถทางภาษา คำนวณ และความสามรถ ในเชิงวิเคราะห์ การสอบใช้เวลา 3 ชั่วโมง 30 นาที

4. การทดสอบสำหรับผู้ที่ต้องการสมัครเข้าศึกษาต่อในหลักสูตรด้านธุรกิจ (Graduate Management Admission Test: GMAT)

GMAT เป็นการทดสอบด้านคณิตศาสตร์และภาษาอังกฤษ เพื่อวัดความสามารถหรือความถนัดสำหรับการศึกษาต่อด้านนี้ สถาบันส่วนใหญ่จะพิจารณาผลสอบนี้เพื่อรับนักศึกษาเข้าเรียนต่อ ซึ่งการทดสอบสามารถทำได้โดยสอบที่ศูนย์สอบ ซึ่งมีกระจายอยู่ทั่วโลก การทดสอบจะทำผ่านคอมพิวเตอร์เท่านั้น โครงสร้างของข้อสอบ GMAT มีรูปแบบการสอบที่เรียกว่า Computer-Adaptive Test (CAT) ใช้เวลาสอบรวมทั้งหมดประมาณ 3.5 – 4 ชั่วโมง รวมเวลาพักแต่ละช่วง ลักษณะข้อสอบประกอบไปด้วย 3 ส่วน คือ

1. Analytical Writing Assessment เป็นการเขียน 2 essays แต่ละ essays ใช้เวลา 30 นาที ผลคะแนนเป็น 0 – 6

2. Quantitative section ประกอบด้วยคำถาม 37 ข้อ ใช้เวลาตอบ 75 นาที ลักษณะคำถามมี 2 แบบคือ problem solving และ data sufficiency ผลคะแนนเป็น 0 - 60

3. Verbal section: ประกอบด้วยคำถาม 41 ข้อ ใช้เวลาตอบ 75 นาที ลักษณะคำถามมี 3 รูปแบบคือ sentence correction, critical reasoning และ reading comprehension ผลคะแนนเป็น 0 - 60

5. การทดสอบทักษะด้านภาษาอังกฤษสำหรับผู้ที่จะศึกษาต่อหรือเพื่อการฝึกอบรม (International English Language Testing System: IELTS)

IELTS เป็นระบบทดสอบความรู้ภาษาอังกฤษ สำหรับผู้ที่ไม่ได้ใช้ ภาษาอังกฤษเป็นภาษาแม่ แต่มีความประสงค์ ที่จะใช้ภาษาอังกฤษในการศึกษาต่อ ลักษณะการสอบ IELTS จะทดสอบทักษะ 4 ส่วน ได้แก่ การอ่าน การเขียน การฟัง และการพูด โดยแบ่งประเภทการสอบเป็น 2 ประเภท คือ Academic และ General Training

1. การสอบแบบ Academic จะใช้ทดสอบผู้ที่จะเข้าศึกษาต่อในระดับอุดมศึกษาขึ้นไป

2. การสอบแบบ General Training ใช้ทดสอบผู้ที่จะเข้าศึกษา ต่อในสายอาชีพ หรือฝึกอบรมงาน วุฒิต่ำกว่าปริญญาตรี ระดับมัธยมศึกษา หลักสูตรระยะ สั้นๆ

ข้อแตกต่างของข้อสอบใน 2 ประเภทนี้คือ ข้อสอบส่วนของการอ่านและการเขียน ซึ่งจะแยกเป็นแบบ Academic และแบบ General Training สำหรับข้อสอบส่วนการฟังและการพูดจะเหมือนกัน

ดังตัวอย่างที่แสดงข้างต้นจะเห็นว่าการทดสอบระดับชาติ หรือการทดสอบเกี่ยวกับการศึกษาต่อต่างประเทศ ล้วนเป็นการทดสอบระดับมหภาคที่ได้ให้ความสนใจในการทดสอบวัดความรู้ความสามารถโดยเลือกใช้แบบทดสอบที่ประกอบด้วยข้อสอบที่มีหลายลักษณะในการสอบวัดครั้งเดียวกัน ด้วยลักษณะเด่นของแบบทดสอบผสมที่สามารถวัดและประเมินผลการเรียนรู้ได้อย่างกว้างขวาง ดังนั้นการศึกษาพัฒนาการความสามารถโดยใช้การทดสอบด้วยแบบทดสอบผสมจึงเป็นสิ่งที่ท้าทายและควรได้มีการศึกษาเพื่อหาข้อสรุป แนวทางการใช้ วิธีการวิเคราะห์ หรือการแปลความหมายของคะแนนสอบต่อไป

การคำนวณคะแนนการเปลี่ยนแปลงรายบุคคล

1. การคำนวณคะแนนการเปลี่ยนแปลงรายบุคคลตามทฤษฎีการทดสอบแบบมาตรฐานเดิม

1.1 การคำนวณผลต่างระหว่างคะแนนดิบ (Difference scores)

การวัดการเปลี่ยนแปลงวิธีแรกที่ได้รับการพัฒนานั้น เป็นการวัดอย่างง่ายโดยคำนวณผลต่างระหว่างคะแนนดิบ (Traub. 1994: 127 – 133; citing Lord 1956: 421 – 437) เป็นตัวดัชนีสำหรับแสดงถึงการเปลี่ยนแปลง โดยมีข้อตกลงเบื้องต้นในการใช้ว่า คะแนนดิบที่ได้จากการวัดประกอบด้วยคะแนนจริง (True score) และ คะแนนความคลาดเคลื่อนในการวัด (Measurement error) เมื่อคะแนนดิบการสอบครั้งที่ 1 ได้จากการวัดก่อนเรียน (X) และ คะแนนดิบครั้งที่ 2 ได้จากการวัดหลังเรียน (Y) ค่าความคลาดเคลื่อนของการวัดทั้งสองครั้งนั้นต้องมีค่าเท่ากัน สูตรการคำนวณเป็นดังสมการ

$$D_i = Y_i - X_i$$

เมื่อ D_i แทน คะแนนการเปลี่ยนแปลงของผู้สอบคนที่ i

X_i แทน คะแนนสอบครั้งแรกของผู้สอบคนที่ i

Y_i แทน คะแนนสอบครั้งที่สองของผู้สอบคนที่ i

ด้วยความสะดวกในการคำนวณ วิธีวัดการเปลี่ยนแปลงที่คำนวณจากผลต่างระหว่างคะแนนดิบจึงได้รับความนิยมและถูกใช้เป็นค่าดัชนีที่สะท้อนถึงการเปลี่ยนแปลงการเรียนของผู้เรียนมาจนกระทั่งปัจจุบัน แต่อย่างไรก็ตามการใช้ผลต่างของคะแนนดิบเพียงอย่างเดียวจากการวัดเพียงสองครั้ง เป็นการละเลยความสนใจต่อความคลาดเคลื่อนของการวัด รวมถึงความสามารถที่แท้จริงของผู้เข้าสอบ การใช้คะแนนผลต่างจึงมีความน่าเชื่อถือในการอ้างอิงต่ำ นอกจากนี้มีนักวิธีวิทยาหลายท่าน (Lord. 1958.; Cronbach; & Furby 1970: 68 – 80; Linn; & Slinde. 1977: 121 – 150; May; & Nicewander. 1998: 882 – 897; Traub. 1994: 127 – 133; Willet; & Sayer. 1994. 363 – 365; Prieler; & Raven. 2002: 3 – 4.) ได้ชี้ให้เห็นข้อด้อยของการนำคะแนนผลต่างอย่างง่ายไปแสดงถึงการเปลี่ยนแปลงความสามารถของผู้เรียนไว้ 4 ประการดังนี้

ข้อด้อยของวิธีการวัดการเปลี่ยนแปลงด้วยความแตกต่างระหว่างคะแนนดิบ

1. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบครั้งแรก กับ คะแนนการเปลี่ยนแปลงเป็นลบ หมายความว่า หากผู้สอบได้คะแนนสูงในการสอบครั้งแรกย่อมมีแนวโน้มที่คะแนนการเปลี่ยนแปลงมีค่าต่ำ ขณะที่ผู้สอบที่มีคะแนนสอบครั้งแรกต่ำกลับมีโอกาสที่คะแนนการเปลี่ยนแปลงมีค่าสูง จึงดูขัดแย้งกับสภาพจริงที่ควรจะเป็น กล่าวคือ โดยปกติผู้ที่มีความสามารถสูงย่อมสอบได้คะแนนสูงจึงมีแนวโน้มที่จะเรียนรู้ได้ดีกว่าผู้สอบได้คะแนนต่ำ ดังนั้นคะแนนการเปลี่ยนแปลงจึงน่าจะสูงกว่าผู้สอบมีความสามารถต่ำซึ่งสอบได้คะแนนต่ำ

2. ในขณะที่ความสัมพันธ์ระหว่างคะแนนการวัดครั้งแรกและคะแนนที่ได้จากครั้งที่สองมีค่าสูง ความเชื่อมั่นของคะแนนผลต่างระหว่างคะแนนดิบกลับมีค่าต่ำกว่าค่าความเชื่อมั่นของคะแนนที่ได้จากการสอบครั้งแรกและคะแนนที่ได้จากการสอบครั้งที่สอง

ค่าสัมประสิทธิ์ความเชื่อมั่นของคะแนนผลต่าง เป็นอัตราส่วนระหว่างความแปรปรวนของผลต่างของคะแนนจริง กับ ความแปรปรวนของผลต่างของคะแนนสังเกต ($\rho_D^2 = \frac{\sigma_{T_D}^2}{\sigma_{X_D}^2}$) โดยที่ความแปรปรวนของผลต่างของคะแนนจริงคำนวณจาก ผลต่างระหว่าง ผลรวมความแปรปรวนของคะแนนจริงที่ได้จากการสอบในแต่ละครั้ง กับ สองเท่าของความแปรปรวนร่วมของคะแนนจริงที่ได้จากการสอบทั้งสองครั้ง ($\sigma_{T_D}^2 = \sigma_{T_Y}^2 + \sigma_{T_X}^2 - 2\sigma_{T_{XY}}$) จะเห็นได้ว่า ค่าความแปรปรวนของผลต่างคะแนนจริงขึ้นอยู่กับ ความแปรปรวนของคะแนนจริงของการสอบครั้งแรก ความแปรปรวนของคะแนนจริงของการสอบครั้งที่สอง และความแปรปรวนร่วมระหว่างคะแนนจริงจากการสอบในครั้งแรกและครั้งที่สอง จึงทำให้ค่าความเชื่อมั่นของคะแนนการเปลี่ยนแปลงมีค่าต่ำลงเมื่อค่าความสัมพันธ์ระหว่างคะแนนสอบในครั้งแรกและครั้งที่สองมีค่าสูงขึ้น

3. คะแนนการเปลี่ยนแปลงขึ้นอยู่กับอิทธิพลของคะแนนสูงสุด (Ceiling Effect) หมายถึงโอกาสที่จะได้รับคะแนนการแปลงมากหรือน้อยนั้นขึ้นอยู่กับคะแนนเต็มในการสอบ เช่น เมื่อสอบได้คะแนนครั้งแรกสูงจนเกือบถึงคะแนนเต็มสูงสุด เมื่อใช้แบบทดสอบชุดเดิม ตรวจให้คะแนนเต็มสูงสุดที่ค่าเท่ากับการสอบครั้งแรก คะแนนการเปลี่ยนแปลงที่ได้จึงไม่อาจจะสูงเกินกว่าผลต่างของคะแนนสูงสุดกับคะแนนสอบครั้งแรกได้ แต่เมื่อเทียบกับผู้สอบอีกคนหนึ่งที่ได้คะแนนสอบในครั้งแรกต่ำกลับมีโอกาสสูงกว่าที่จะมีคะแนนการเปลี่ยนแปลงมากกว่า จึงเป็นเหตุให้คะแนนผลต่างอย่างง่ายไม่ได้สะท้อนความสามารถที่เปลี่ยนแปลงไปอย่างแท้จริง

4. การแปลความหมายของคะแนนการเปลี่ยนแปลงขึ้นอยู่กับคะแนนสอบครั้งแรก (Floor Effect) เมื่อผู้สอบได้คะแนนผลต่างเท่ากัน ไม่สามารถแปลความหมายได้ว่าผู้สอบทั้งสองคนนั้นมีความสามารถที่เปลี่ยนแปลงไปเท่ากันได้ ต้องพิจารณาถึงความสามารถหรือคะแนนสอบครั้งแรกด้วย โดยเฉพาะอย่างยิ่งกรณีที่ผู้สอบได้คะแนนสอบครั้งแรกต่ำมาก กับผู้สอบที่ได้คะแนนสอบครั้งแรกสูงมาก เช่น ผู้สอบคนแรกได้คะแนนในการสอบครั้งแรกและครั้งที่สองเป็น 10 คะแนน และ 40 คะแนน ตามลำดับ ในขณะที่ผู้สอบคนที่สองได้คะแนนในการสอบครั้งแรกและครั้งที่สองเป็น60 คะแนน และ 90 คะแนน ตามลำดับ เมื่อคำนวณผลต่างอย่างง่ายผู้สอบทั้งสองคนสองต่างได้ผลต่างของคะแนนสอบเป็น 30 ทั้งคู่ แต่ก็ไม่อาจจะแปลความได้ว่าผู้เรียนทั้งสองนั้นมีความสามารถที่พัฒนาขึ้นเท่าเทียมกัน

เมื่อการวัดการเปลี่ยนแปลงการเรียนรู้โดยใช้การคำนวณจากผลต่างระหว่างคะแนนดิบเพียงอย่างเดียวเกิดปัญหาการขาดความแม่นยำการทำนายความสามารถที่เปลี่ยนไป เกิดปัญหาในการแปลความหมาย จึงยากที่จะบอกถึงความสามารถที่เปลี่ยนแปลงของผู้สอบโดยแท้จริง นักวิธีวิทยารุ่นต่อมาจึงได้นำปัจจัยอื่นที่ส่งผลต่อการแปลความหมายของการวัดการเปลี่ยนแปลง เข้ามาร่วมคำนวณในสูตรเพื่อพยายามทำให้ผลการคำนวณที่ได้บ่งบอกได้ทั้งทิศทางปริมาณการเปลี่ยนแปลงและแปลความหมายการเปลี่ยนแปลงการเรียนรู้ได้อย่างแท้จริง

1.2 การคำนวณความแตกต่างของคะแนนส่วนที่เหลือ (Residual scores)

จากข้อจำกัดที่คะแนนการเปลี่ยนแปลงขึ้นอยู่กับอิทธิพลของคะแนนสูงสุด (Ceiling Effect) แมนนิ่ง และ ดูบอยส์ (จรรยา สิงห์ทอง. 2550: 27 อ้างอิงจาก Manning; & DuBois. 1958: 191 – 194; 1962: 287 – 361) จึงได้เสนอสูตรการคำนวณเพื่อแสดงผลการวัดการเปลี่ยนแปลง โดยพิจารณาจากคะแนนส่วนที่จะสามารถพัฒนาหรือเปลี่ยนแปลงได้อีกจนถึงคะแนนสูงสุด โดยเรียกส่วนที่จะพัฒนาได้มากที่สุดที่เป็นไปได้นี้ว่าคะแนนส่วนที่เหลือ

$$
\begin{aligned}
R_i &= Y_i - Y_i' \\
&= Y_i - [Y_i + B_{yx}(X_i - X)] \\
&= Y_i - [Y + (r_{xy} / S_x^2)(X_i - X)]
\end{aligned}
$$

เมื่อ R_i แทน คะแนนการเปลี่ยนแปลงส่วนที่เหลือผู้สอบคนที่ i

Y_i' แทน คะแนนส่วนที่เหลือผู้สอบคนที่ i

r_{xy} แทน สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบครั้งที่หนึ่งและครั้งที่สอง

ต่อมา ครอนบาค และ เฟอร์บี (Cronbach; & Furby. 1970: 74) ได้พัฒนาสูตรนี้ต่อด้วยหลักการเดิม แต่ตัดคะแนนสอบก่อนเรียน และหลังเรียนที่เบี่ยงเบนไปจากค่าเฉลี่ยของการสอบแต่ละครั้งออกไปจากสูตร และใช้การวิเคราะห์การถดถอยเชิงเส้นแทน โดยให้คะแนนสอบครั้งที่สองเป็นตัวแปรตาม ส่วนคะแนนสอบครั้งที่หนึ่งเป็นตัวแปรต้น การคำนวณเช่นนี้ให้ความสำคัญกับคะแนนสอบครั้งที่หนึ่งเพื่อเป็นฐานในการทำนายคะแนนสอบหลังเรียน จึงช่วยลดปัญหาเรื่องอิทธิพลความรู้เดิมได้

การคำนวณด้วยวิธีนี้จะมีค่าความเชื่อมั่นสูงกว่าการคำนวณด้วยผลต่างของคะแนนดิบ จึงทำให้ผู้สอบครั้งแรกได้คะแนนสูงมีคะแนนการเปลี่ยนแปลงสูงกว่าผู้มีความสามารถต่ำกว่า แต่อย่างไรก็ตามค่าความเชื่อมั่นของคะแนนการเปลี่ยนแปลงจะต่ำเมื่อคะแนนสอบทั้งสองครั้งมีความสัมพันธ์กันสูง จึงยากในการแปลความหมาย เนื่องจากค่าสัมประสิทธิ์การถดถอยแปรเปลี่ยนไปตามขนาดกลุ่มตัวอย่าง ค่าความคลาดเคลื่อนจึงยังตกค้างอยู่ คะแนนการเปลี่ยนแปลงที่คำนวณได้จากสูตรนี้จึงบอกได้เพียงว่าผู้สอบนั้นมีการเปลี่ยนแปลงสูงหรือต่ำกว่าการเปลี่ยนแปลงที่คาดหวังเท่านั้น

1.3 การคำนวณคะแนนผลต่างที่เป็นอิสระจากคะแนนสอบครั้งแรก (Base – free measurement of change)

เมื่อคะแนนในการสอบวัดครั้งแรกเป็นอุปสรรคต่อการแปลความหมายการวัดการเปลี่ยนแปลง ทัคเกอร์ ดามาริน และ เมสสิค (Tucker; Damarin; & Messick. 1966: 457 – 473) จึงได้เสนอสูตรที่ไม่นำคะแนนสอบครั้งแรกมาคำนวณ แต่พิจารณาเฉพาะผลต่างระหว่างคะแนนดิบจากการทดสอบครั้งที่สอง กับคะแนนทำนายผลการสอบครั้งที่สองด้วยคะแนนจริงจากการทดสอบครั้งแรก

$$\begin{aligned} B_i &= Y_i - Y_i' \\ &= Y_i - [Y + (B_{yx} / R_{xx'})(X_i - X)] \\ &= Y_i - [Y + (S_{T_x,T_Y} / S_{T_x}^2)(X_i - X)] \end{aligned}$$

เมื่อ B_i แทน คะแนนการเปลี่ยนแปลงที่เป็นอิสระจากการวัดครั้งแรกที่ใช้เป็นฐาน

$R_{xx'}$ แทน ค่าความเชื่อมั่นของเครื่องมือวัดก่อนเรียน

S_{T_x,T_Y} แทน ค่าความแปรปรวนร่วมระหว่างคะแนนจริงของการสอบครั้งแรกและคะแนนสอบครั้งที่สอง

$S_{T_x}^2$ แทน ค่าความแปรปรวนของคะแนนจริงในการสอบครั้งแรก

การวัดการเปลี่ยนแปลงที่คำนวณด้วยสูตรนี้ ช่วยขจัดปัญหาความสัมพันธ์ทางลบระหว่างคะแนนครั้งแรกกับคะแนนผลต่างได้ แต่ถ้าพิจารณาถึงความคลาดเคลื่อนแล้วกลับพบว่าวิธีนี้มีแนวโน้มที่จะให้ค่าความคลาดเคลื่อนสูงกว่า วิธีการคำนวณด้วยผลต่างอย่างง่าย และวิธีคะแนนส่วนที่เหลือ

1.4 การคำนวณจากคะแนนการเปลี่ยนแปลงสัมพัทธ์ (Relative Gain Score)

ศิริชัย กาญจนวาสี (2538. 144 – 149; อ้างอิงจาก Kanjanawasee. 1989: Alternative Strategies for policy analysis: An assessment of school effects on student's cognitive and affective mathematics outcomes in lower secondary school in Thailand. pp. 55 – 51) ได้พัฒนาสูตรการคำนวณที่สามารถแปลความหมายได้โดยพยายามขจัดอิทธิพลของคะแนนเต็ม (Ceiling Effect) การวัดการเปลี่ยนแปลงสัมพัทธ์เป็นการเปรียบเทียบระหว่างผลต่างของคะแนนที่ได้จากการทดสอบทั้งสองครั้งกับคะแนนการเปลี่ยนแปลงทั้งหมดที่เป็นไปได้ในรูปอัตราส่วน

$$S_i = \left[\frac{Y_i - X_i}{F - X_i} \right] \times 100$$

เมื่อ S_i แทน คะแนนการเปลี่ยนแปลงสัมพัทธ์

F แทน คะแนนเต็ม(คะแนนสูงสุดที่เป็นไปได้)

การคำนวณด้วยสูตรนี้ช่วยลดอิทธิพลคะแนนสูงสุด พร้อมกับคำนึงความสามารถครั้งแรก จึงทำให้หลายงานวิจัยที่เปรียบเทียบสูตรการคำนวณวิธีนี้กับวิธีการคำนวณอื่นแล้วพบว่ามีประสิทธิภาพสูงกว่าวิธีอื่น

ต่อมาปี 2548 สูตรนี้ถูกพัฒนาต่อโดย ธีระ อาชวเมธี; กมลวรรณ ตังธนกานนท์; และ สมหวัง พิธิยานุวัฒน์ (Teara Archawamety; Kamonwan Tangdhanakanond; & Somwung Pitiyanuwat. 2005: 213 – 227)

$$\text{Balance Relative Gain Score} = \left[\frac{Y_i - X_i}{F - (Y_i + X_i)/2}\right] \times 100$$

1.5 การคำนวณจากคะแนนเพิ่มที่แท้จริง (Estimated true gain score)

ลอร์ด (Traub. 1994. 135; citing Lord. 1956: 421 – 437) และ แม็ก เนมาร์ (McNamar. 1958: 47 – 55) ได้ใช้การวิเคราะห์การถดถอยหลายตัวแปร (Multiple Regression Analysis) เพื่อทำนายผลต่างระหว่างคะแนนจริงจากการทดสอบครั้งแรกกับครั้งที่สอง โดยใช้ตัวพยากรณ์เป็นคะแนนดิบจากการทดสอบทั้งสองครั้ง

$$V_i = K + W_x X_i + W_y Y_i$$
$$= Y + b_{vx.y}(X_i - X) + b_{vy.x}(Y_i - Y)$$

เมื่อ

V_i แทน คะแนนเพิ่มที่แท้จริง

K แทน ค่าคงที่

W_x แทน สัมประสิทธิ์การถดถอยพหุสำหรับการทำนายผลต่างระหว่างคะแนนจริงในการสอบครั้งที่สองกับครั้งแรก (Ty – Tx) ด้วย X

W_y แทน สัมประสิทธิ์การถดถอยพหุสำหรับการทำนายผลต่างระหว่างคะแนนจริงในการสอบครั้งที่สองกับครั้งแรก (Ty – Tx) ด้วย Y

$b_{vx.y}$ แทน สัมประสิทธิ์การถดถอยพหุสำหรับการทำนายผลต่างระหว่างคะแนนจริงในการสอบครั้งที่สองกับครั้งแรก (Ty – Tx) ด้วย X เมื่อควบคุม Y

$b_{vy.x}$ แทน สัมประสิทธิ์การถดถอยพหุสำหรับการทำนายผลต่างระหว่างคะแนนจริงในการสอบครั้งที่สองกับครั้งแรก (Ty – Tx) ด้วย X เมื่อควบคุม X

การใช้หลักการวิเคราะห์การถดถอยโดยใช้ตัวพยากรณ์เป็นคะแนนดิบจากการทดสอบทั้งสองครั้ง จัดว่าเป็นตัวประมาณค่าที่ดี เพราะคำนึงถึงพื้นความรู้เดิม การแปลความหมายของคะแนนการเปลี่ยนแปลงจึงถูกต้อง แต่อย่างไรก็ตามวิธีนี้ยังมีข้อจำกัดหลายประการเช่น มีความยุ่งยากในการ

คำนวณ ขาดการพิจารณาอิทธิพลคะแนนสูงสุด คะแนนขึ้นอยู่กับกลุ่มตัวอย่าง และต้องใช้กลุ่มตัวอย่างขนาดใหญ่

1.6 การวัดการเปลี่ยนแปลงการเรียนรู้ของ อรุณี อ่อนสวัสดิ์

อรุณี อ่อนสวัสดิ์ (2537: 70- 77; 2538: 1 – 15) พยายามแก้ปัญหาอิทธิพลของความรู้เดิม และอิทธิพลของคะแนนสูงสุด จึงได้พัฒนาสูตรโดยยึดทฤษฎีการเรียนรู้ของบลูมและอิทธิพลของคะแนนสูงสุด โดยมองว่าการเปลี่ยนแปลงการเรียนรู้เป็นฟังก์ชันของพื้นฐานความรู้เดิมร่วมกับ การเรียนการสอน และ อิทธิพลคะแนนสูงสุด ซึ่งการเปลี่ยนแปลงการเรียนรู้สามารถวัดด้วยค่าประมาณจริง หรือ วัดด้วยคะแนนดิบ เพื่อให้ครอบคลุมกับปัจจัยต่างๆ อรุณี อ่อนสวัสดิ์ จึงเสนอสูตรการคำนวณการวัดการเปลี่ยนแปลงการเรียนรู้ทั้งสิ้น 8 วิธี

1. $C1A_i = (1-W_2)\hat{T}_{Y_i} - \hat{T}_{X_i}$
2. $C1B_i = (1-W_2)Y_i - X_i$
3. $C2A_i = (1-W_2)\hat{T}_{Y_i} - (1-W_3)\hat{T}_{X_i}$
4. $C2B_i = (1-W_2)Y_i - (1-W_3)X_i$
5. $C3A_i = (1-W_{2.1})\hat{T}_{Y_i} - \hat{T}_{X_i}$
6. $C3B_i = (1-W_{2.1})Y_i - X_i$
7. $C4A_i = (1-W_2/\hat{T}_{X_i})\hat{T}_{Y_i} - \hat{T}_{X_i} + W_2F/\hat{T}_{X_i}$
8. $C4B_i = (1-W_2/X_i)Y_i - X_i + W_2F/X_i$

โดยที่ W_2 แทนสัมประสิทธิ์การถดถอย $T_y - T_x$ ลงบน $T_{(F-Y)}$ นั่นคือ $W_2 = \dfrac{-r_{DY}S_D}{R_{YY'}S_Y}$

W_3 แทนสัมประสิทธิ์การถดถอย $T_y - T_x$ ลงบน $T_{(F-X)}$ นั่นคือ $W_3 = \dfrac{-r_{DX}S_D}{R_{XX'}S_X}$

$W_{2.1}$ แทนสัมประสิทธิ์การถดถอย $T_y - T_x$ ลงบน $T_{(F-X)}$ มื่อควบคุม T_X

นั่นคือ $W_{2.1} = \dfrac{\left(S_X^2R_{XX'} + S_Y^2R_{YY'}2r_{XY}S_XS_Y\right)^{\frac{1}{2}}}{S_Y} \dfrac{\left(r_{DX}r_{DY} - R_{XX'}r_{DY}\right)}{\sqrt{R_{DD}}\left(R_{XX'}R_{YY'} - r_{XY}^2\right)}$

เมื่อเปรียบเทียบค่าความเชื่อมั่นและค่าความคลาดเคลื่อนที่ได้จากการคำนวณด้วยสูตรทั้ง 8 กับวิธีการการคำนวณด้วยการวัดคะแนนความแตกต่างอย่างง่าย และวิธีการวัดคะแนนเพิ่มที่แท้จริงของลอร์ด พบว่าไม่แตกต่างกัน แต่กลับมีความซับซ้อนในการคำนวณเพิ่มมากกว่า

ต่อมาในปี 2538 อรุณี อ่อนสวัสดิ์ จึงได้พัฒนาวิธีการวัดการเปลี่ยนแปลงขึ้นมาอีกแปดวิธีจากค่าประมาณคะแนนจริงและคะแนนดิบ โดยมีแนวความคิดหลักว่า คะแนนจริงมีความแตกต่างกันขึ้นอยู่กับ การเปลี่ยนแปลงพื้นความรู้เดิม การเปลี่ยนแปลงเมื่อเกิดการเรียนรู้จากการเรียนการสอน

และการเปลี่ยนแปลงเมื่อลบอิทธิพลของคะแนนสูงสุด มีรายละเอียดของวิธีการคำนวณทั้งสิ้น 4 แนวคิด มีรายละเอียดในแต่ละแนวคิดดังนี้

แนวคิดแรก อิทธิพลคะแนนสูงสุดขึ้นอยู่กับคะแนนจริงหลังเรียน และถือว่าผู้ที่ได้คะแนนหลังเรียนเท่ากันย่อมได้รับอิทธิพลสูงสุดเท่ากัน เมื่อใช้สัมประสิทธิ์การถดถอยอย่างง่าย ที่ได้จากคะแนนจริงซึ่งเกิดจากความแตกต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียน ถดถอยลงบนคะแนนจริงซึ่งเกิดจากความแตกต่างคะหว่างคะแนนเต็มและคะแนนสอบหลังเรียน จึงได้สูตรการคำนวณ 2 สูตร ซึ่งเป็นสูตรสำหรับค่าประมาณคะแนนจริง และ สำหรับคะแนนดิบ ตามลำดับดังนี้

9. $V1A_i = (1-W_2)\hat{T}_{Y_i} - (1+W_1)\hat{T}_{X_i}$

10. $V1B_i = (1-W_2)Y_i - (1+W_1)X_i$

แนวคิดที่สอง อิทธิพลคะแนนสูงสุดขึ้นอยู่กับคะแนนจริงหลังเรียนและคะแนนจริงก่อนเรียน เนื่องจากผู้ที่ได้คะแนนหลังเรียนเท่ากันอาจจะได้รับอิทธิพลคะแนนสูงสุดต่างกัน ดังนั้นตัวแปรที่มีอิทธิพลจึงมีสองตัว ได้แก่ 1) สัมประสิทธิ์การถดถอยอย่างง่าย ที่ได้จากคะแนนจริงซึ่งเกิดจากความแตกต่างของคะแนนจริงก่อนเรียนและหลังเรียน ถดถอยลงบน คะแนนจริงซึ่งเกิดจากความแตกต่างระหว่างคะแนนเต็มและคะแนนสอบหลังเรียน และ 2) สัมประสิทธิ์การถดถอยอย่างง่าย ที่ได้จากคะแนนจริงซึ่งเกิดจากความแตกต่างของคะแนนจริงก่อนเรียนและหลังเรียน ถดถอยลงบน คะแนนจริงซึ่งเกิดจากความแตกต่างคะหว่างคะแนนเต็มและคะแนนสอบก่อนเรียนจึงได้สูตรการคำนวณ 2 สูตร ซึ่งเป็นสูตรสำหรับค่าประมาณคะแนนจริง และ สำหรับคะแนนดิบ ตามลำดับดังนี้

11. $V2A_i = (1-W_2)\hat{T}_{Y_i} - (1+W_1-W_3)\hat{T}_{X_i}$

12. $V2B_i = (1-W_2)Y_i - (1+W_1-W_3)X_i$

แนวคิดที่สาม อิทธิพลคะแนนสูงสุดขึ้นอยู่กับคะแนนจริงหลังเรียน และคะแนนจริงก่อนเรียน เนื่องจาก การได้รับคะแนนสอบหลังเรียนเท่ากันอาจจะได้รับอิทธิพลคะแนนสูงสุดไม่เท่ากัน จึงใช้สัมประสิทธิ์ถดถอยหลายตัวแปรที่ได้จาก คะแนนจริงซึ่งเกิดจากความแตกต่างระหว่างคะแนนจริงหลังเรียนและก่อนเรียน ถดถอยลงบนคะแนนจริงซึ่งเกิดจากความแตกต่างของคะแนนเต็มและคะแนนหลังเรียน เมื่อควบคุมคะแนนจริงก่อนเรียน จึงได้สูตรการคำนวณ 2 สูตร ซึ่งเป็นสูตรสำหรับค่าประมาณคะแนนจริง และ สำหรับคะแนนดิบ ตามลำดับดังนี้

13. $V3A_i = (1-W_{2.1})\hat{T}_{Y_i} - (1+W_1)\hat{T}_{X_i}$

14. $V3B_i = (1-W_{2.1})Y_i - (1+W_1)X_i$

แนวคิดที่สี่ อิทธิพลคะแนนสูงสุดขึ้นอยู่กับคะแนนจริงหลังเรียน และคะแนนจริงก่อนเรียน เนื่องจาก การได้รับคะแนนสอบหลังเรียนเท่ากันอาจจะได้รับอิทธิพลคะแนนสูงสุดไม่เท่ากัน แต่อิทธิพลของคะแนนสูงสุดอาจจะแปรผันตามสัดส่วนของคะแนนจริงก่อนเรียน กล่าวคือ ถ้าคะแนนจริง

หลังเรียนเท่ากัน ผู้ที่ได้คะแนนจริงก่อนเรียนต่ำกว่าจะได้รับอิทธิพลคะแนนสูงสุดน้อยกว่าผู้ที่ได้คะแนนจริงก่อนเรียนสูงกว่า จึงใช้สัมประสิทธิ์ถดถอยอย่างงายที่ได้จาก คะแนนจริงซึ่งเกิดจากความแตกต่างระหว่างคะแนนจริงหลังเรียนและก่อนเรียน ถดถอยลงบนคะแนนจริงซึ่งเกิดจากความแตกต่างของคะแนนเต็มและคะแนนหลังเรียน หารด้วยคะแนนจริงก่อนเรียนจึงได้สูตรการคำนวณ 2 สูตร ซึ่งเป็นสูตรสำหรับค่าประมาณคะแนนจริง และ สำหรับคะแนนดิบ ตามลำดับดังนี้

15. $V4A_i = (1-W_2/\hat{T}_{X_i})\hat{T}_{Y_i} - (1+W_1)\hat{T}_{X_i} + W_2F/\hat{T}_{X_i}$

16. $V4B_i = (1-W_2/X_i)Y_i - (1+W_1)X_i + W_2F/X_i$

โดยที่ W_1 แทนสัมประสิทธิ์การถดถอย $T_y - T_x$ ลงบน T_X นั่นคือ $W_1 = \frac{r_{DX}S_D}{R_{XX'}S_X}$

1.7 การวัดคะแนนเพิ่มด้วยวิธีเทียบส่วนของ วินิจ เทือกทอง

วินิจ เทือกทอง (2537: 21 – 24) ได้ประยุกต์วิธีการคำนวณคะแนนผลต่างอย่างง่าย วิธีเศษเหลือ และวิธีของลอร์ด และวิธีของศิริชัย กาญจนวาสี มาผนวกเข้าเป็นสูตรการคำนวณใหม่โดยอิงโครงสร้างสูตรการคำนวณแบบการวัดการเปลี่ยนแปลงสัมพัทธ์ของศิริชัย กาญจนวาสี จำนวน 3 สูตร และ สูตรการคำนวณแบบวิธีเทียบส่วนร้อย จำนวน 3 สูตร รวมทั้งสิ้น 6 สูตร ดังนี้

1.7.1 สูตรการคำนวณคะแนนการเปลี่ยนแปลงแบบอิงโครงสร้างสูตรการคำนวณตามวิธีของศิริชัย กาญจนวาสีโดยประยุกต์จากวิธีการคำนวณผลต่างอย่างง่าย วิธีเศษเหลือ และวิธีของลอร์ด ประกอบด้วยสูตรคำนวณ สามสูตรดังนี้

1.7.1.1 คะแนนการเปลี่ยนแปลงดั้งเดิมเทียบกับศักยภาพผู้สอบ

$$SRG = \left[\frac{Y-X}{F-X}\right]\times 100$$

1.7.1.2 คะแนนการเปลี่ยนแปลงเศษเหลือเทียบกับศักยภาพผู้สอบ

$$SDG = \left[\frac{Y-b_wX}{F-X}\right]\times 100$$

1.7.1.3 คะแนนการเปลี่ยนแปลงของลอร์ดเทียบกับศักยภาพผู้สอบ

$$SLG = \left[\frac{\bar{G}+b_{GX.Y}(X-\bar{X})+b_{GY.X}(Y-\bar{Y})}{F-X}\right]\times 100$$

1.7.2 สูตรการคำนวณคะแนนการเปลี่ยนแปลงแบบอิงโครงสร้างสูตรการคำนวณตามวิธีเทียบส่วนร้อย โดยประยุกต์จากวิธีการคำนวณผลต่างอย่างง่าย วิธีเศษเหลือ และวิธีของลอร์ด ประกอบด้วยสูตรคำนวณสามสูตรประยุกต์ดังนี้

1.7.2.1 คะแนนการเปลี่ยนแปลงดั้งเดิมเทียบกับคะแนนสอบครั้งแรก

$$PRG = \left[\frac{Y-X}{X}\right]\times 100$$

1.7.2.2 คะแนนการเปลี่ยนแปลงเศษเหลือเทียบกับคะแนนสอบครั้งแรก

$$PDG = \left[\frac{Y - b_w X}{X} \right] \times 100$$

1.7.2.3 คะแนนการเปลี่ยนแปลงของลอร์ดเทียบกับคะแนนสอบครั้งแรก

$$PLG = \left[\frac{\bar{G} + b_{GX.Y}(X - \bar{X}) + b_{GY.X}(Y - \bar{Y})}{X} \right] \times 100$$

เมื่อ Y หมายถึง คะแนนดิบจากการทดสอบครั้งที่สอง

X หมายถึง คะแนนดิบจากการทดสอบครั้งที่หนึ่ง

$\bar{G}$ หมายถึง คะแนนเพิ่มเฉลี่ยของกลุ่มตัวอย่าง

b_w หมายถึง สัมประสิทธิ์การถดถอย

$b_{GX.Y}$ และ $b_{GY.X}$ หมายถึง สัมประสิทธิ์การถดถอยแบบบางส่วน

สูตรการคำนวณทั้งหกวิธีนี้เมื่อเปรียบเทียบความคลาดเคลื่อนระหว่างสูตรการคำนวณจากคะแนนผลต่างอย่างง่าย วิธีเศษเหลือ และ วิธีของลอร์ด พบว่าวิธีการวัดคะแนนความแตกต่างอย่างง่ายหรือความแตกต่างระหว่างคะแนนดิบ กับ วิธีของลอร์ด ให้ค่าความคลาดเคลื่อนในการวัดต่ำสุดเมื่อไม่คำนึงถึงอิทธิพลของคะแนนสูงสุด ในทางกลับกันถ้าคำนึงถึงอิทธิพลของคะแนนสูงสุด วิธีเศษเหลือเทียบกับศักยภาพของผู้สอบให้ค่าความคลาดเคลื่อนต่ำที่สุด

1.8 การวัดการเปลี่ยนแปลงของตำแหน่งเปอร์เซนต์ไทล์ (Change in Percentile Rank)

สูตรนี้ได้รับการเสนอโดยรัสเซล (Russel. 2001: Online) เป็นสูตรการคำนวณเฉพาะการศึกษาที่เน้นการเปลี่ยนแปลงทั้งแบบเพิ่มขึ้นหรือลดลงของตำแหน่งเปอร์เซนต์ไทล์เฉลี่ย (Mean percentile ranking) ด้วยวิธีการคำนวณที่ง่าย และไม่ซับซ้อนจึงทำให้วิธีการคำนวณนี้ได้รับความนิยม แต่อย่างไรก็ตามการใช้การเปลี่ยนตำแหน่งเปอร์เซนต์ไทล์เฉลี่ยนั้นมีข้อบกพร่องสองประการ ประการแรก การคำนวณค่าตำแหน่งเปอร์เซนไทล์เฉลี่ย (Mean percentile ranking) ขึ้นอยู่กับตำแหน่งเปอร์เซต์ไทล์รายบุคคล การประมาณค่าเฉลี่ยของกลุ่มจึงไม่ถูกต้อง และประการที่สองตำแหน่งเปอร์เซนต์ไทล์ที่แตกต่างกันในช่วงเวลาต่างกันย่อมบอกปริมาณการเปลี่ยนแปลงที่แท้จริงไม่ได้

1.9 การวัดคะแนนการเปลี่ยนแปลงเป็นร้อยละ (Percent Change)

ในปี 2001 นอกจากวิธีการวัดการเปลี่ยนแปลงของตำแหน่งเปอร์เซนต์ไทล์ แล้ว รัสเซล (Russell. 2001: 4 – 5)ได้เสนอวิธีการวัดคะแนนการเปลี่ยนแปลงเป็นร้อยละอีกสูตรหนึ่ง โดยอธิบายว่า คะแนนการเปลี่ยนแปลงเป็นคะแนนสอบครั้งที่สองที่เบี่ยงเบนไปจากคะแนนสอบครั้งแรก การคำนวณทำได้โดยนำคะแนนสอบครั้งที่สองหารด้วยคะแนนสอบครั้งแรกซึ่งจะดีกว่าวิธีการวัดการ

เปลี่ยนแปลงของตำแหน่งเปอร์เซนต์ไทล์ เงื่อนไขของการวัดคะแนนการเปลี่ยนแปลงเป็นร้อยละต้องเปลี่ยนคะแนนสอบให้เป็นมาตรฐาน หรือเทียบเท่ากับโค้งปกติ กลาส และฮอบคิน (Russel. 2001; Online; citing Glass; & Hopkins. 1984.) ได้กล่าวว่าการวัดคะแนนการเปลี่ยนแปลงเป็นร้อยละจัดว่าเป็นคะแนนที่อยู่ในมาตราวัดแบบอัตราส่วน (Ratio scale) และคะแนนมาตรฐานและคะแนนเทียบเท่าโค้งปกติต่างเป็นข้อมูลในระดับอัตรภาค (Interval scale) จึงสามารถอธิบายได้ชัดเจนกว่าการวัดการเปลี่ยนแปลงแบบเปอร์เซนต์ไทล์ ซึ่งมีคะแนนอยู่ในมาตราวัดแบบจัดอันดับ (Ordinal scale)

การวัดการเปลี่ยนแปลงทุกวิธีล้วนมีปัญหาเกี่ยวกับค่าความเชื่อมั่นของคะแนนการเปลี่ยนแปลงต่ำแทบทั้งสิ้น (Russel. 2001: Online; citing Willet. 1988.) แต่อย่างไรก็ตามการวัดการเปลี่ยนแปลงก็ยังมีความจำเป็น การศึกษาวิจัยเพื่อหาวิธีการวัดการเปลี่ยนแปลงที่เหมาะสมสำหรับสถานการณ์การใช้ หรือแต่ละเป้าหมายของการศึกษาที่แตกต่างกันไปจึงต้องดำเนินต่อไป อย่างไรก็ตามในการเลือกใช้วิธีการวัดการเปลี่ยนแปลงที่มีอยู่นั้น วิธีที่เลือกใช้ควรจะมีคุณสมบัติที่สำคัญสองประการ ประการแรกเป็นวิธีการวัดการเปลี่ยนแปลงที่มีมาตรฐานเพียงพอสำหรับการตีความหมายของทุกคะแนนสอบที่ได้เหมือนกัน และสามารถเปรียบเทียบกันได้ เมื่อมีการทดสอบด้วยแบบทดสอบต่างฉบับและนักเรียนต่างระดับชั้นกันได้ และประการที่สองเป็นวิธีการวัดการเปลี่ยนแปลงที่มีค่าความเชื่อมั่นพอประมาณ

2. การคำนวณคะแนนการเปลี่ยนแปลงรายบุคคลตามทฤษฎีการตอบข้อสอบ

2.1 การวัดการเปลี่ยนแปลงความสามารถแฝง (Latent change)

วิธีการวัดเปลี่ยนแปลงความสามารถแฝง (Latent change) (May; & Nicewander. 1998) เป็นวิธีคำนวณการเปลี่ยนแปลงความสามารถจากผลต่างของความสามารถแฝงที่ประมาณค่าตามโมเดลการตอบข้อสอบด้วยวิธี EAP

$$d_\theta = \theta_2 - \theta_1$$

เมื่อ θ_1 และ θ_2 คือ ความสามารถจริงจากการสอบครั้งที่หนึ่งและครั้งที่สองที่ประมาณค่าด้วยวิธี EAP

d_θ คือ ผลต่างของความสามารถที่แท้จริงที่ประมาณค่าตามโมเดลการตอบข้อสอบด้วยวิธี EAP

2.2 การวัดคะแนนความสามารถแท้จริงสัมพัทธ์

อวยพร เรืองตระกูล (อวยพร เรืองตระกูล. 2544:84 - 85) ได้พัฒนาสูตรการคำนวณสำหรับวัดการเปลี่ยนแปลงโดยใช้ความสามารถแท้จริงที่ได้จากการประมาณค่าด้วยโมเดลการตอบข้อสอบประยุกต์แบบอิงโครงสร้างสูตรการคำนวณตามทฤษฎีการทดสอบแบบมาตรฐานเดิมตามวิธีของศิริชัย และวิธีของวินิจ ดังนี้

2.2.1 สูตรการคำนวณคะแนนพัฒนาการความสามารถที่แท้จริงสัมพัทธ์เทียบกับศักยภาพการพัฒนา (Relative true ability of growth and potential ratio method)

$$S_\theta = \left[\frac{\theta_2 - \theta_1}{Max - \theta_1} \right]$$

2.2.2 สูตรการคำนวณคะแนนพัฒนาการความสามารถที่แท้จริงสัมพัทธ์เทียบกับความสามารถที่แท้จริงก่อนเรียน (Relative true ability of growth and initial ratio method)

$$V_\theta = \left[\frac{\theta_2 - \theta_1}{Max + \theta_1} \right]$$

เมื่อ θ_1 และ θ_2 คือ ความสามารถจริงจากการสอบครั้งที่หนึ่งและครั้งที่สองตามลำดับ

Max คือคะแนนความสามารถแท้จริงสูงสุด

2.3 การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้

ไฮเดนเบอร์ก (Hiedenberg) ได้พัฒนาวิธีการวัดการเปลี่ยนแปลงโดยใช้คะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ (Item response Theory Domain Score) เพื่อศึกษาความสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงกับคะแนนสอบครั้งแรก

$$D_\theta = \pi_2 - \pi_1$$

เมื่อ π_1 และ π_2 คือ คะแนนมวลความรู้ก่อนเรียนและหลังเรียนที่ประมาณค่าตามแนวคิดของบอค (Bock) ในทฤษฎีการตอบข้อสอบ

หากพิจารณาขั้นตอนการคำนวณ และเงื่อนไขการใช้ของวิธีการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบดังที่กล่าวข้างต้น วิธีการคำนวณคะแนนการเปลี่ยนแปลงความสามารถ ($\theta_2 - \theta_1$) ของ เมย์และไนส์วานเดอร์ (May; & Nicewander. 1998) เป็นวิธีที่สามารถคำนวณได้ง่ายกว่าเมื่อเทียบกับโมเดลการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบอื่น เป็นวิธีคำนวณที่ช่วยขจัดปัญหาความคลาดเคลื่อนที่มักเกิดขึ้นจากการคำนวณคะแนนผลต่างอย่างง่ายของคะแนนดิบได้ และผลต่างค่าประมาณความสามารถมีความบิดเบือนของมาตรวัด (Scale distortion) น้อยกว่าผลต่างที่ได้จากการคำนวณคะแนนผลต่างอย่างง่ายที่คำนวณจากคะแนนรวมของข้อสอบทั้งฉบับ ผลต่างของคะแนนดิบย่อมไม่สามารถสะท้อนความสามารถที่แท้จริงที่เกิดจากความแตกต่างของลักษณะข้อสอบซึ่งมีรูปแบบการให้คะแนนต่างกันได้ อีกทั้งคะแนนดิบที่ได้จากผลการทดสอบแต่ละครั้งมีความซับซ้อนมากเกินกว่าจะเปรียบเทียบกันได้โดยตรง ทำให้คำนวณพัฒนาการความสามารถของผู้เรียนโดยใช้ผลต่างคะแนนดิบที่ได้จากแบบทดสอบผสมโดยตรงจึงไม่เพียงพอสำหรับสะท้อนพัฒนาการการเรียนรู้ของผู้เรียน ดังนั้นวิธีการคำนวณคะแนนการเปลี่ยนแปลงความสามารถ($\theta_2 - \theta_1$) อย่างง่ายจึงมีความเหมาะสมมากกว่าการคำนวณด้วยผลต่างคะแนนดิบ

ทฤษฎีการตอบข้อสอบ (Item Response Theory: IRT)

ทฤษฎีการตอบข้อสอบ (IRT) เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างความสามารถที่มีอยู่ในตัวบุคคล (Latent trait or ability) กับผลการตอบข้อสอบหรือข้อคำถามโดยใช้โค้งลักษณะข้อสอบ (Item Characteristic Curve: ICC) ซึ่งมีการกำหนดลักษณะของข้อสอบด้วยพารามิเตอร์ความยาก (b) อำนาจจำแนก (a) และโอกาสการเดาข้อสอบถูก (c) (ศิริชัย กาญจนวาสี, 2545: 45 - 46) ทฤษฎีการตอบข้อสอบตั้งอยู่บนฐานความคิดที่สำคัญ 2 ประการ ประการแรก ผลการตอบข้อสอบสามารถอธิบายได้ด้วยความสามารถที่มีอยู่ภายในของผู้ตอบและประการที่สอง ความสัมพันธ์ระหว่างผลการตอบข้อสอบกับความสามารถที่มีอยู่ภายใน สามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อสอบ หรือโค้งลักษณะข้อสอบ (ICC) ซึ่งเป็นฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชันโลจิสติก (Logistic function) ซึ่งมีลักษณะใกล้เคียงกับฟังก์ชันปกติสะสม (Normal ogive function)

1. การประมาณค่าพารามิเตอร์ด้วยโมเดลการตอบข้อสอบ

โมเดลการตอบข้อสอบ เป็นโมเดลที่ขึ้นอยู่กับค่าประมาณความสามารถของผู้สอบจากคะแนนที่ได้จากการทดสอบ โมเดลการตอบข้อสอบเป็นโมเดลที่เขียนในรูปความสัมพันธ์ทางคณิตศาสตร์ เพื่อประมาณค่าความสามารถที่แตกต่างกันของแต่ละคนในการทำข้อสอบแต่ละข้อในการประมาณค่าความสามารถจากการทดสอบ การประมาณค่าคุณลักษณะข้อสอบด้วยโมเดลการตอบข้อสอบ เป็นอิสระจากสิ่งที่ขึ้นอยู่กับข้อสอบ และยอมให้มีการประมาณของความคลาดเคลื่อนในแต่ละข้อเป็นอันดับแรก นอกจากนี้โมเดลการตอบข้อสอบยอมให้มีค่าประมาณความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement) เป็นรายบุคคลด้วย แทนที่ข้อตกลงตามทฤษฎีมาตรฐานเดิมที่ระบุว่าความคลาดเคลื่อนมาตรฐานของการวัดของทุกคนมีค่าเท่ากัน ดังนั้นวัตถุประสงค์แรกของการใช้โมเดลการตอบข้อสอบคือการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ และพารามิเตอร์ของข้อสอบ ซึ่งค่าพารามิเตอร์ทั้งสองนั้นมีความสำคัญและลักษณะแตกต่างกันไป ขึ้นอยู่กับโมเดลการตอบข้อสอบที่ใช้

โมเดลการตอบข้อสอบแบบเอกมิติ (Unidimensional Model) เป็นการศึกษาพารามิเตอร์ความสามารถแฝงที่มีคุณลักษณะเด่นลักษณะเดียวเท่านั้น แต่อาจมีจำนวนพารามิเตอร์ของข้อสอบแตกต่างกันตามโมเดลการตอบข้อสอบที่ใช้ โมเดลการตอบข้อสอบสำหรับข้อมูลแบบทวิวิภาคเป็นโมเดลสำหรับประมาณค่าพารามิเตอร์ของข้อมูลที่ได้จากการทดสอบที่ให้คะแนนรายข้อสองค่า คือ ตอบถูกได้ 1 ตอบผิด ได้ 0 ได้แก่ โมเดลโลจิสติกแบบ 1 พารามิเตอร์ (1PL) โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (2PL) และโมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PL) แม้ว่าจะมีโมเดลโลจิสติก 4 พารามิเตอร์ (4PL) ก็ตาม แต่ยังเป็นเพียงความสนใจเฉพาะทางทฤษฎี ในทางปฏิบัติกลับไม่มีการนำไปใช้ประมาณค่าความสามารถแต่อย่างใด (Si. 2002: 12 – 13 Citing Barton; & Lord. 1981)

ค่าพารามิเตอร์ข้อสอบทั้งสามค่าในโมเดลการตอบข้อสอบซึ่งกำหนดคุณลักษณะข้อสอบสามประการได้แก่ ค่าความยาก (Difficulty) (หรือ พารามิเตอร์ที่ตั้ง (Location parameter) ปกติแทนด้วย “b”) ค่าอำนาจจำแนก (Discrimination) หรือ พารามิเตอร์ความชัน (Slope parameter ปกติแทนด้วย “a”) และ ค่าการเดา (Pseudo-chance) หรือพารามิเตอร์ต่ำกว่าแอทซิมโทด (Lower asymptote parameter), ปกติแทนด้วย “c”) นอกจากค่าพารามิเตอร์เหล่านี้จะถูกใช้สำหรับโมเดลการตอบแบบสองค่าแล้ว โมเดลการตอบข้อสอบแบบหลายค่าก็ใช้ค่าพารามิเตอร์ข้อสอบเหล่านี้ด้วยเช่นกัน ซึ่งจะได้กล่าวรายละเอียดต่อไป

1.1 โมเดลการตอบข้อสอบสำหรับประมาณค่าพารามิเตอร์แบบเอกมิติ

1.1.1 โมเดลการตอบข้อสอบสำหรับแบบทดสอบที่ให้คะแนนแบบสองค่า

1.1.1.1 โมเดลการตอบข้อสอบแบบหนึ่งพารามิเตอร์ (One-Parameter Logistic Model)

โค้งลักษณะข้อสอบของโมเดลการตอบข้อสอบแบบ 1 พารามิเตอร์ สามารถเขียนเป็นฟังก์ชันโลจิสติก ได้ดังสมการ

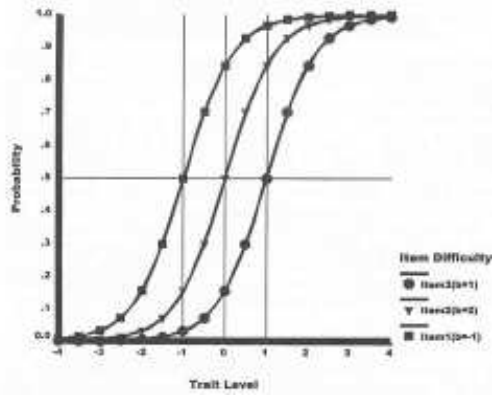
$$P_i(\theta) = \frac{1}{1+e^{-(\theta-b_i)}}$$

เมื่อ

$P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้เข้าสอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อ i ได้ถูกต้อง

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50

e แทน ค่าคงที่มีค่าเป็น 2.718



ภาพประกอบ 2 โค้งลักษณะข้อสอบ (ICC) ของตัวอย่างข้อสอบ 3 ข้อสำหรับโมเดลการตอบข้อสอบแบบ 1 พารามิเตอร์

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบข้อสอบแบบ 1 พารามิเตอร์มีลักษณะสำคัญคือ ค่าพารามิเตอร์ความยาก (b_i) จะแปรเปลี่ยนไปตามลักษณะของข้อสอบ โดยที่ค่าพารามิเตอร์อำนาจจำแนก (a_i) ของข้อสอบทุกข้อมีค่าเท่ากัน และ ค่าพารามิเตอร์การเดา (c_i)มีค่าเป็น 0

1.1.1.2 โมเดลการตอบข้อสอบแบบสองพารามิเตอร์(Two-Parameter Logistic Model)โค้งลักษณะข้อสอบของโมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์ สามารถเขียนเป็นฟังก์ชันโลจิสติก ได้ดังสมการ

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

เมื่อ

$P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้เข้าสอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อ i ได้ถูกต้อง

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก 0.50

a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นค่าความชันของโค้ง ICC ณ ตำแหน่ง b_i

e แทน ค่าคงที่มีค่าเป็น 2.718

D แทน ค่าคงที่มีค่าเป็น 1.70

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์มีลักษณะสำคัญคือ ค่าพารามิเตอร์ความยาก (b_i) และค่าพารามิเตอร์อำนาจจำแนก (a_i) จะแปรเปลี่ยนไปตามลักษณะของข้อสอบ โดยที่ ค่าพารามิเตอร์การเดา c_i มีค่าเป็น 0

1.1.1.3 โมเดลการตอบข้อสอบแบบสามพารามิเตอร์ (Three-Parameter Logistic Model)

โค้งลักษณะข้อสอบของโมเดลการตอบข้อสอบแบบ 3 พารามิเตอร์ สามารถเขียนเป็นฟังก์ชันโลจิสติกได้ดังสมการ

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

เมื่อ

$P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้เข้าสอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อ i ได้ถูกต้อง

b_i แทน ค่าพารามิเตอร์ความยากของข้อสอบข้อที่ i ซึ่งเป็นค่าที่แสดงตำแหน่งของ ICC ณ จุด θ ที่มีโอกาสตอบข้อสอบถูก $\frac{1+c_i}{2}$

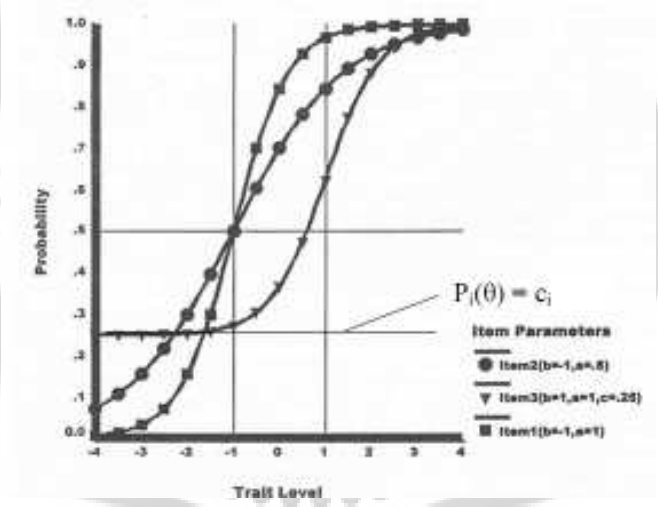
a_i แทน ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อที่ i ซึ่งเป็นค่าความชันของโค้ง ICC ณ ตำแหน่ง b_i

c_i แทน ค่าพารามิเตอร์การเดาข้อสอบถูก

e แทน ค่าคงที่มีค่าเป็น 2.718

D แทน ค่าคงที่มีค่าเป็น 1.70

โค้งลักษณะข้อสอบ (ICC) สำหรับโมเดลการตอบข้อสอบแบบ 3 พารามิเตอร์มีลักษณะสำคัญคือ ค่าพารามิเตอร์ความยาก (b_i) ค่าพารามิเตอร์อำนาจจำแนก (a_i) และค่าพารามิเตอร์การเดาข้อสอบถูก (c_i) จะแปรเปลี่ยนไปตามลักษณะของข้อสอบ



ภาพประกอบ 3 ตัวอย่างโค้งลักษณะข้อสอบ (ICC) ของข้อสอบ 3 ข้อ สำหรับโมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์และ 3 พารามิเตอร์

ที่มา: สี (Si, 2002:15. Online)

1.1.4 ฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ (Item and Test Information)

เป็นดัชนีสำหรับบ่งชี้คุณภาพของแบบทดสอบ (ศิริชัย กาญจนวาสี. 2534. อ้างอิงจาก Birnbaum. 1968) ซึ่งเป็นดัชนีที่ผสมระหว่างดัชนีคุณลักษณะของข้อสอบ ประกอบด้วยค่าพารามิเตอร์ความยาก ค่าพารามิเตอร์อำนาจจำแนก และความแปรปรวนของคะแนนรายข้อ

$$I_i(\theta) = \frac{\left[P_i^{'}(\theta)\right]^2}{P_i(\theta)\left(1-P_i(\theta)\right)}, i = 1, 2, ..., k$$

เมื่อ

$I_i(\theta)$ แทน ค่าฟังก์ชันสารสนเทศหรือค่าสารสนเทศที่ได้รับจากข้อสอบข้อที่ i สำหรับผู้สอบที่มีความสามารถ θ

$P_i^{'}(\theta)$ แทน ความชันของฟังก์ชันการตอบข้อสอบข้อที่ i ณ ตำแหน่งความสามารถ θ

$P_i(\theta)$ แทน ความน่าจะเป็นของผู้ตอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

จากสูตรของฟังก์ชันสารสนเทศของข้อสอบ เมื่อพิจารณาเทียบกับค่าพารามิเตอร์ a, b, และ c ของข้อสอบ สรุปได้ว่า

1. ถ้าผู้สอบมีความสามารถ θ ใกล้เคียงกับค่าพารามิเตอร์ b ของผู้สอบ ค่าสารสนเทศของข้อสอบจะมีค่าสูงขึ้น แต่ถ้าผู้สอบมีความสามารถ θ ห่างจากค่าพารามิเตอร์ b ของผู้สอบ ค่าสารสนเทศของข้อสอบจะมีค่าลดลง

2. ถ้าค่า พารามิเตอร์ a ของข้อสอบมีค่ามากขึ้น ค่าสารสนเทศของข้อสอบจะมีค่าสูงขึ้น

3. ถ้าค่า พารามิเตอร์ c ของข้อสอบมีค่าเข้าใกล้ศูนย์ ค่าสารสนเทศของข้อสอบจะมีค่าสูงขึ้น

4. ค่าสารสนเทศจะมีค่าสูงสุด ณ จุดที่ความสามารถ θ สูงสุด ซึ่งมีค่าเท่ากับพารามิเตอร์ความยาก b และ ค่า พารามิเตอร์ c เป็นศูนย์ แต่ ณ จุดที่ความสามารถ θ สูงสุด และค่าสูงกว่าพารามิเตอร์ความยาก b และ ค่าพารามิเตอร์ c มากกว่าศูนย์ค่าสารสนเทศจะมีค่าลดลง

1.1.2 ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information)

$$I(\theta) = \sum_{i=1}^{k} I_i(\theta), \quad i = 1, 2, ..., k$$

เมื่อ $I(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบที่ได้รับจากแบบทดสอบสำหรับผู้เข้าสอบที่มีความสามารถ θ

ค่าฟังก์ชันสารสนเทศแต่ละข้อมีอิสระจากกันในการประกอบเป็นค่าสารสนเทศของแบบทดสอบ ซึ่งต่างจากทฤษฎีการทดสอบแบบมาตรฐานเดิม ที่ค่าความยากและค่าอำนาจจำแนกของข้อสอบแต่ละข้อต่างส่งผลโดยตรงต่อค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับจึงไม่อาจจะคำนวณค่าแต่ละข้อโดยอิสระจากกันได้ คะแนนที่ได้จึงขึ้นอยู่กับลักษณะเฉพาะของกลุ่มผู้สอบและแบบทดสอบเฉพาะที่เลือกมาใช้

ค่าสารสนเทศมีค่าผกผันกับความคลาดเคลื่อนมาตรฐานของการประมาณค่า ดังนั้น การที่ค่าสารสนเทศมีค่าสูงในช่วงความสามารถ θ ใดก็จะมีความแม่นยำสูงในการประมาณค่าความสามารถของผู้ตอบในช่วง θ นั้น และมีค่าความคลาดเคลื่อนมาตรฐานต่ำ

1.1.5 ประสิทธิภาพสัมพัทธ์ (Relative Efficiency: RE (θ))

ฟังก์ชันสารสนเทศระหว่างแบบทดสอบที่วัดคุณลักษณะเดียวกันสามารถทำได้โดยการคำนวณประสิทธิภาพสัมพัทธ์ ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบทดสอบต่างฉบับ ณ ตำแหน่งความสามารถ θ เดียวกัน ดังนี้

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

เมื่อ

$I_A(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบฉบับ A ณ ตำแหน่งบนสเกลความสามารถร่วมกันระดับ θ

$I_B(\theta)$ แทน ค่าฟังก์ชันสารสนเทศของแบบทดสอบฉบับ B ณ ตำแหน่งบนสเกลความสามารถร่วมกันระดับ θ

การแปลความหมายของค่าดัชนีประสิทธิภาพสัมพัทธ์ สามารถนำไปช่วยคัดเลือกแบบทดสอบได้ดังนี้

RE(θ) เป็น 1 แสดงว่าแบบทดสอบทั้งสองฉบับมีประสิทธิภาพเท่ากัน สำหรับผู้ตอบที่มีความสามารถ θ

RE(θ) มากกว่า 1 แสดงว่าแบบทดสอบฉบับ A มีประสิทธิภาพสูงกว่าแบบทดสอบฉบับ B สำหรับผู้ตอบที่มีความสามารถ θ

RE(θ) น้อยกว่า 1 แสดงว่าแบบทดสอบฉบับ A มีประสิทธิภาพต่ำกว่าแบบทดสอบฉบับ B สำหรับผู้ตอบที่มีความสามารถ θ

เช่น ถ้า RE(θ) = 1.20 แสดงว่าแบบทดสอบฉบับ A มีประสิทธิภาพสูงกว่าแบบทดสอบฉบับ B สำหรับผู้ตอบที่มีความสามารถ θ โดยที่ แบบทดสอบ A เป็นแบบทดสอบที่ทำหน้าที่เสมือนว่ายาวกว่าแบบทดสอบฉบับ B 20% หรือ เราสามารถตัดแบบทดสอบฉบับ A ออก 20% แบบทดสอบฉบับ A ก็ยังคงมีความแม่นยำในการประมาณค่าความสามารถ ณ ตำแหน่ง θ ได้ดีเท่ากับแบบทดสอบฉบับ B

การเปรียบเทียบค่าฟังก์ชันสารสนเทศเฉลี่ยระหว่างแบบทดสอบที่วัดคุณลักษณะเดียวกัน สามารถทำได้โดยคำนวณประสิทธิภาพสัมพัทธ์เฉลี่ยซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบต่างฉบับ ณ ทุกตำแหน่ง θ

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

เมื่อ

$AI(\theta,X)$ แทน ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบฉบับ X ณ ทุกตำแหน่ง θ

$AI(\theta,Y)$ แทน ค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบฉบับY ณ ทุกตำแหน่ง θ

1.1.6 คะแนนความสามารถและความคลาดเคลื่อนของการประมาณค่า (Ability Score and Standard Error of Estimation)

1.1.6.1 คะแนนความสามารถของผู้ตอบ (Ability score)

คะแนนความสามารถของผู้ตอบจากการทำข้อสอบแบบให้คะแนน 0,1 จำนวน k ข้อ สัญลักษณ์การประมาณค่าเป็นดังนี้

ทฤษฎีการทดสอบแบบมาตรฐานเดิม X = T + E

เมื่อ X แทน คะแนนที่สังเกตได้ (Observe score) ซึ่งเป็นคะแนนรวมของข้อที่ตอบถูก มีค่าอยู่ระหว่าง 0 ถึง k

T แทน คะแนนจริง (True score) มีค่าอยู่ระหว่าง 0 ถึง k

π แทน คะแนนมวลความรู้เป็นสัดส่วนของการตอบถูกมีค่าอยู่ระหว่าง 0 – 100%

ทฤษฎีการตอบข้อสอบ (IRT)

θ แทน คะแนนความสามารถจริง (Ability score) นิยมคำนวณให้เป็นคะแนนมาตรฐาน (Standard score) ซึ่งมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเป็น 0 และ 1 ตามลำดับ θ มีค่าอยู่ระหว่าง $[-\infty,+\infty]$

$T=\sum_{i=1}^{k} P_i(\theta)$ แทน คะแนนจริง มีค่าอยู่ระหว่าง $[\sum c_i,k]$

$\pi=\frac{1}{k}\sum_{i=1}^{k} P_i(\theta)$ แทน คะแนนมวลความรู้ซึ่งคิดเป็นเปอร์เซนต์มีค่าอยู่ระหว่าง

$$\left[\frac{100}{k}\left(\sum c_i\right)\%,100\%\right]$$

1.1.6.2 ความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation)

ความคลาดเคลื่อนมาตรฐานของการประมาณค่า (Standard Error of Estimation) ความคลาดเคลื่อนมาตรฐานของการประมาณค่า $[SE(\theta)]$ เป็นค่าส่วนเบี่ยงเบนมาตรฐานของการแจกแจงความน่าจะเป็นของค่าประมาณค่าความสามารถที่แท้จริง (θ) ซึ่งเป็นค่า

สัดส่วนผกผันกับความถูกต้องแม่นยำของการประมาณค่าความสามารถ หรือค่าฟังก์ชันสารสนเทศของแบบทดสอบ ดังสูตร

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

เมื่อ

$SE(\hat{\theta})$ แทน ความคลาดเคลื่อนมาตรฐานของการประมาณค่าสำหรับผู้ตอบที่มีความสามารถ θ

$I(\theta)$ แทน สารสนเทศที่ได้จากแบบทดสอบสำหรับผู้ตอบที่มีความสามารถ θ

การสอบทุกครั้งย่อมมีความคลาดเคลื่อนในการประมาณค่าความสามารถ (θ) ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ มีค่าเท่ากับส่วนกลับของรากที่สองของค่าฟังก์ชันสารสนเทศของแบบทดสอบ ดังนั้นถ้าค่าฟังก์ชันสารสนเทศของแบบทดสอบมีค่าสูงแล้ว ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถย่อมมีค่าต่ำ นั่นคือ ถ้าฟังก์ชันสารสนเทศของแบบทดสอบมีค่าสูง ณ ผู้สอบระดับความสามารถ θ ใด (ฟังก์ชันการตอบข้อสอบแต่ละข้อมีความชันสูงและความแปรปรวนต่ำ ณ ระดับ θ นั้น) ความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถของผู้สอบ ณ ระดับ θ นั้นจะต่ำ หรือ มีความแม่นยำสูงในการประมาณค่าความสามารถของผู้สอบ (ศิริชัย กาญจนวาสี, 2534)

ค่า $SE(\theta)$ ใน IRT คล้ายกับ CTT แต่ต่างกันที่ $SE(\theta)$ มีค่าผันแปรไปตามตำแหน่ง θ ในขณะที่ความคลาดเคลื่อนมาตรฐานของการวัด(Standard Error of Measurement: SEM) เป็นค่าคงที่สำหรับผู้ตอบทุกคน

ลักษณะทั่วไปของ $SE(\theta)$

1. เมื่อแบบทดสอบมีจำนวนข้อมากขึ้นขนาดของ $SE(\theta)$ จะลดลง
2. แบบทดสอบที่มีค่าอำนาจจำแนกสูง และค่าการเดาถูกต่ำ ขนาดของ $SE(\theta)$จะลดลง
3. แบบทดสอบที่มีค่าพารามิเตอร์ความยากใกล้เคียงกับ θ ผู้สอบ(ไม่ว่าจะยากหรือง่ายเกินไป) ขนาดของ $SE(\theta)$ จะลดลง
4. ค่า $I(\theta)$ มีสูงกว่า 25 จะทำให้ขนาดของ $SE(\theta)$ ค่อนข้างคงที่และมีผลกระทบต่อความคลาดเคลื่อนในการประมาณค่า θ ต่ำ (Green; Yen & Burket,1989)

1.2 โมเดลการตอบข้อสอบสำหรับข้อสอบที่ให้คะแนนแบบหลายค่า

โมเดลการตอบข้อสอบของแบบทดสอบที่ให้คะแนนแบบหลายค่า (Polytomous IRT Models) เป็นโมเดลความสัมพันธ์ระหว่างความสามารถหรือคุณลักษณะของผู้ตอบ (θ) กับความน่าจะเป็นของการเลือกตอบแต่ละรายการคำตอบของข้อสอบหรือข้อคำถาม การประมาณค่าพารามิเตอร์สำหรับแต่ละรายการคำตอบของข้อสอบหรือคำถาม นำไปสู่การคำนวณค่าฟังก์ชัน

สารสนเทศของข้อสอบ เมื่อนำมารวมกัน ณ ตำแหน่ง θ เดียวกัน ทำให้ได้ค่าฟังก์ชันสารสนเทศของแบบทดสอบ และสามารถคำนวณค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่า θ ได้ในลักษณะเดียวกับโมเดลการตอบข้อสอบแบบตรวจให้คะแนนแบบทวิวิภาค (ศิริชัย กาญจนวาสี. 2545:101)

เมื่อแบบทดสอบมีจำนวนรายการคำตอบมากกว่าสองรายการคำตอบ จึงต้องใช้โมเดลการตอบข้อสอบพหุวิภาคเพื่อประมาณค่าพารามิเตอร์ โมเดลการตอบข้อสอบพหุวิภาคสามารถแบ่งได้สองกลุ่มตามประเภทของรายการคำตอบของข้อสอบ ได้แก่ แบบนามบัญญัติ (Nominal) และเรียงอันดับ (Ordinal) (Si, 2002) รายการคำตอบแบบนามบัญญัติตามธรรมชาติไม่ได้เรียงลำดับ เช่น คำถามที่ถามในแบบวัดบุคลิกภาพที่คำตอบสะท้อนให้เห็นว่าผู้ตอบมีลักษณะบุคลิกภาพตรงกับลักษณะใด ส่วนรายการคำตอบแบบเรียงอันดับนั้นเป็นการเรียงคุณลักษณะแฝงบนเส้นต่อเนื่องแสดงคุณลักษณะ เช่น จำนวนของขั้นตอนที่สมบูรณ์ในการแก้ปัญหาคณิตศาสตร์ หรือระดับรายการคำตอบในแบบวัดแบบลิเคิร์ท ผู้สอบที่มีคุณลักษณะสูงกว่าย่อมสามารถตอบในรายการคำตอบที่มีระดับคะแนนสูงกว่าได้ และผู้สอบที่มีความสามารถต่ำกว่า ย่อมมีความน่าจะเป็นมากกว่าที่จะตอบในรายการที่ต่ำกว่า บอค (Bock.1972) ได้พัฒนาโมเดลนอมินอลเรสพอนส์ (Nominal Response Model) ขึ้นสำหรับข้อคำถามที่มีรายการคำตอบอยู่ในระดับนามบัญญัติ และในขณะเดียวกันมีโมเดลการตอบแบบพหุวิภาคได้ถูกพัฒนาขึ้นสำหรับข้อสอบที่มีรายการคำตอบแบบเรียงอันดับ ซึ่งพัฒนาขึ้นตามประเภทของแบบทดสอบที่ใช้

1.2.1 โมเดลนอมินอลเรสพอนส์ (Nominal Response Model: NRM)

บอค (Bock.1972) ได้เสนอโมเดลในรูปทั่วไปของโมเดลการตอบข้อสอบแบบพหุวิภาคที่สามารถใช้ในการกำหนดความน่าจะเป็นของผลการตอบของผู้สอบในหนึ่งรายการคำตอบเมื่อเทียบกับรายการคำตอบอื่นว่าเป็นฟังก์ชันของความสามารถผู้สอบและคุณลักษณะรายการคำตอบโดยที่รายการคำตอบเหล่านี้ไม่จำเป็นต้องมีการเรียงลำดับ โมเดลพหุวิภาคนี้ต่างจากโมเดลทวิวิภาคที่มีฟังก์ชันการตอบข้อสอบ (Item Response Function) คือการมีฟังก์ชันการตอบในรายการคำตอบของข้อสอบ (Item Category Response Functions: ICRFs) ซึ่ง ICRFs ได้มาจากความน่าจะเป็นของการตอบในรายการคำตอบในแต่ละรายการที่ต่างกัน ความน่าจะเป็นของ ICRFs สำหรับโมเดลบอคเขียนได้ดังนี้

$$P_{ik}(\theta) = \frac{\exp(a_{ik}\theta + c_{ik})}{\sum_{h=1}^{m} \exp(a_{ih}\theta + c_{ih})}, k = 1, 2, ..., m$$

เมื่อ $P_{ik}(\theta)$ คือ ความน่าจะเป็นของผู้ตอบในการใช้ความสามารถ θ ตอบรายการคำตอบที่ k ในข้อสอบข้อ i

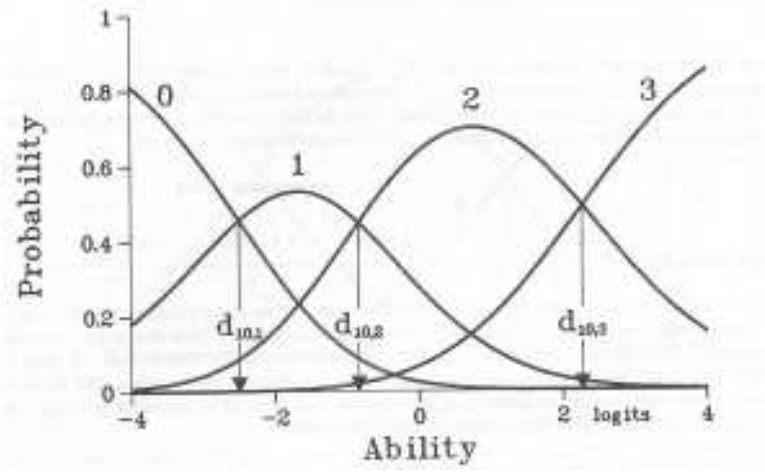
a_{ik} แทน ค่าพารามิเตอร์อำนาจจำแนกประจำรายการคำตอบที่ k

c_{ik} แทน ค่าพารามิเตอร์ความยากประจำรายการคำตอบที่ k

m แทน จำนวนของรายการคำตอบในข้อสอบข้อ i

แต่ละรายการคำตอบจึงมี ICRF หลายฟังก์ชันซึ่งจำนวนของ ICRFs ในข้อสอบแต่ละข้อเท่ากับจำนวนรายการคำตอบประจำข้อสอบนั้น โมเดลจึงถูกอธิบายไม่สมบูรณ์จำต้องถูกดึงออกมา บอคจึงกำหนดให้ผลรวมของพารามิเตอร์รายการคำตอบของแต่ละข้อมีค่าเป็นศูนย์ นั่นคือ $\sum_{k=1}^{m} a_{ik} = \sum_{h=1}^{m} c_{ih} = 0$ ดังนั้นการกำหนดเช่นนี้จึงทำให้แต่ละข้อสอบมี ICRFs เพียงชุดเดียวในข้อสอบหนึ่งข้อ

โค้งคุณลักษณะรายการคำตอบประจำข้อสอบ (Item Category Characteristic Curve :ICCC) ได้มาจาก การเขียนกราฟของ ICRF ทุกค่า θ ดังนั้น ICCCs ของข้อสอบที่มี 4 รายการคำตอบ จึงเป็นไปตามภาพ 1 โค้ง ICCCs ไม่เพิ่มขึ้นหรือลดลงทางเดียว ยกเว้นรายการคำตอบสูงสุดหรือต่ำสุดเท่านั้น โค้งต่ำสุด ICCC แสดงถึงรายการคำตอบสะท้อนระดับความสามารถที่ต่ำสุด ความน่าจะเป็นของการตอบในรายการคำตอบเพิ่มขึ้นตามเส้นบนเส้นต่อเนื่องความสามารถ (Ability continuum) โค้ง ICCC ของรายการคำตอบจึงลดลงทางเดียว ในทางตรงกันข้าม โค้ง ICCC สูงสุดแสดงความน่าจะเป็นของรายการคำตอบที่สะท้อนความสามารถระดับสูงสุด ความน่าจะเป็นของการตอบในรายการคำตอบนี้จึงเพิ่มขึ้นทางเดียวบนเส้นต่อเนื่องความสามารถ (Ability continuum)



ภาพประกอบ 4 โค้งคุณลักษณะของรายการคำตอบประจำข้อสอบในโมเดลการตอบข้อสอบแบบให้คะแนนหลายค่า

ที่มา: สี (Si. 2002: 19 citing in Wright and Masters, 1982: 188)

ในแต่ละระดับของความสามารถ θ ผลรวมของ ICRFs เท่ากับ 1 นั่นคือ $\sum_{k=1}^{m} P_{ik}(\theta/\theta=\theta_j)=1$ เพราะว่ารายการคำตอบแต่ละรายการต่างร่วมกัน

เมลเลนเบอร์ก (Mellenbergh. 1995) ได้แสดงให้เห็นว่า โมเดลของบอค สามารถเขียนในรูปใหม่ได้ในรูปของค่าลอการิทึมอัตราส่วนแต้มต่อ จำนวน m -1 ค่า ได้แยกตัวแปรการตอบนามบัญญัติเป็น m รายการ เพื่อทำให้เกิดตัวแปรการตอบแบบสองค่า m -1 ตัวแปร แต่ละตัวแปรการตอบแบบสองค่า ขึ้นอยู่กับตัวเลือกระหว่างรายการคำตอบ m กับ รายการคำตอบอ้างอิง เนื่องจากรายการคำตอบเป็นนามบัญญัติ จึงเลือกรายการคำตอบแรกเป็นรายการอ้างอิงและเพื่อความสะดวกจึงกำหนดพารามิเตอร์ให้มีค่าเป็นศูนย์ ลอการิทึมอัตราส่วนแต้มต่อ (log odds) ของการเลือกรายการคำตอบ k เมื่อเทียบกับรายการคำตอบแรกจึงเป็น

$$\ln\left(P_{ik}(\theta)/P_{i1}(\theta)\right)=\ln\left(\frac{\exp(a_{ik}\theta+c_{ik})}{\exp(a_{i1}\theta+c_{i1})}\right)=\left(a_{ik}-a_{i1}\right)\theta+\left(c_{ik}-c_{i1}\right),k=1,2,...,m$$

ส่วนโมเดลแบบให้คะแนนสองค่าภาคของอีก m -1 ตัวที่เหลือต่างก็เทียบเท่ากับ m ซึ่ง ICRFs ดังที่อธิบายตามโมเดลของบอค จะเห็นว่า โมเดลของเบิร์นบวมซึ่งเป็นโมเดลโลจิสติกแบบสองพารามิเตอร์ (2-PL) เป็นกรณีพิเศษของโมเดลบอค ที่มี m = 2 และความน่าจะเป็นในการตอบข้อสอบอย่างถูกต้อง $P_i(\theta)=P_2(\theta)$ และ ความน่าจะเป็นของการตอบข้อสอบไม่ถูกต้อง $Q_i(\theta)=P_{i1}(\theta)$ สิ่งที่ตามมาคือ

$$a_i(\theta-b_i)=\ln\left(\frac{P_i(\theta)}{1-P_i(\theta)}\right)=\ln\left(\frac{P_i(\theta)}{Q_i(\theta)}\right)=\ln\left(\frac{P_{i2}(\theta)}{Q_{i1}(\theta)}\right)=a_{i2}\theta+c_{i2}$$

และ $a_i=a_{i2}, b_i=-\frac{c_2}{a_i}$

ตามกรอบแนวคิดของ เมลเลนเบอร์กได้แสดงว่าโมเดลของบอค อธิบายการตอบข้อสอบแบบพหุวิภาคที่สามารถมองเป็นกลุ่มของโมเดลการตอบข้อสอบแบบทวิวิภาค ความน่าจะเป็นของการตอบในโมเดลทวิวิภาคเหล่านี้ยังคงได้มาจากฟังก์ชันการแจกแจงโลจิสติกเหมือนกับโมเดลโลจิสติกของเบิร์นบวม แต่ผลรวมของความน่าจะเป็นของสองรายการคำตอบในโมเดลทวิวิภาคไม่เท่ากับ 1 เหมือนกับโมเดลของเบิร์นบวม ต่อมา บอค (Bock. 1997) ได้พัฒนาแนวคิดต่ออีก โดยกล่าวว่า NRM เป็นแนวคิดเริ่มต้น สามารถเลือกเป็นโมเดลอย่างทางการได้ทั้งสองโมเดล ซึ่งเป็นการยืนยันแนวความคิดของเมลเลนเบิร์ก และต่อมาได้แสดงต่อไปว่า โมเดลของบอค ถูกนำไปสร้างโมเดลต่างๆ สำหรับข้อสอบที่มีรายการคำตอบเรียงลำดับ เพราะว่าโมเดลพหุวิภาคสามารถแยกออกเป็นกลุ่มของโมเดลทวิวิภาคได้ โมเดลสำหรับรายการคำตอบเรียงลำดับมีข้อจำกัดมากเพราะว่าลำดับของรายการคำตอบต้องผ่านการตรวจสอบว่า ลำดับนั้นแสดงในรูปของรายการคำตอบแบบต่อเนื่องหรือกลุ่มของรายการคำตอบที่มีจากการให้คะแนนสองค่า

1.2.2 โมเดลการตอบข้อสอบแบบพหุวิภาคที่รายการคำตอบเรียงลำดับ (Polytomous IRT Models with Ordinal Response Categories)

เนื่องจากโมเดลการตอบข้อสอบแบบพหุวิภาคมีหลายโมเดล อย่างไรก็ตามสามารถแบ่งกลุ่มของโมเดลพหุวิภาคได้เป็น 3 ประเภท ซึ่ง เฮมเคอร์ (Hemker. 2001) ศึกษาพบว่า โมเลนา (Molenaar. 1983) เป็นบุคคลแรกที่เปรียบเทียบโมเดลพหุวิภาคเรียงอันดับ ทิซเซน และ เสติร์นเบอร์ก (Thissen; & Steinberg. 1986) ได้จัดหมวดหมู่ของโมเดลการตอบข้อสอบไว้และได้จำแนกประเภทของโมเดลพหุวิภาคโดยใช้สมการทางคณิตศาสตร์ของฟังก์ชัน ICRFs. จึงแบ่งรายการคำตอบของโมเดลพหุวิภาคออกเป็นสองโมเดล

เมลเลนเบอร์ก (Mellenbergh. 1995) ได้ใช้โมเดลของบอคเป็นจุดเริ่มต้นและแบ่งระดับคำตอบออกเป็นสามประเภทเพื่อแยกรายการคำตอบไปสู่ระบบทวิวิภาค โดยเรียกชื่อว่า “โมเดลรายการคำตอบซึ่งอยู่ติดกัน (Adjacent-category models) โมเดลความน่าจะเป็นสะสม (Cumulative probability models) และโมเดลอัตราส่วนต่อเนื่อง (Continuation-ratio models)”

สำหรับโมเดลรายการคำตอบซึ่งอยู่ติดกัน (Adjacent-category models) เขาได้แบ่งระดับรายการคำตอบของข้อสอบไปเป็นคู่ของรายการคำตอบที่อยู่ติดกัน เช่น รายการที่ k และ k+1 เมื่อ k = 1, 2, …, m-1 และใช้โมเดลบอคเพื่อสร้างลอการิทึมอัตราส่วนแต้มต่อ (log odds)ในแต่ละคู่ ดังนี้

$$\ln\left(\frac{P_{i(k+1)}(\theta)}{1-P_{ik}(\theta)}\right)=\ln\left(\frac{\exp\left(a_{i(k+1)}\theta+c_{i(k+1)}\right)}{\exp\left(a_{ik}\theta+c_{ik}\right)}\right)=\left(a_{i(k+1)}-a_{ik}\right)\theta+\left(c_{i(k+1)}-c_{ik}\right)=a'_{ik}\theta+c'_{ik}$$

เมื่อ k = 1, 2, …, m-1 ลอการิทึมอัตราส่วนแต้มต่อ (log odds) m-1 คู่ เพื่ออธิบายโมเดลการตอบข้อสอบแบบให้คะแนนหลายค่าแบบจัดอันดับ ซึ่งอันดับของรายการคำตอบยังคงเป็นวิธีที่รายการคำตอบถูกแบ่งเป็นคู่ โมเดลการตอบข้อสอบแบบให้คะแนนหลายค่าแบบจัดอันดับในกลุ่มของโมเดลซึ่งมีรายการคำตอบติดกัน ได้แก่ โมเดลเจเนอรอลไรส์พาร์เชียลเครดิต (Generalized partial credit model: GPCM) ของ มูรากิ (Muraki,1992) (เมื่อ a_{ik} แทนอำนาจจำแนกประจำข้อทุกรายการk เป็น a_i ข้อ และ $-c_{ik}$ เป็นความยากประจำขั้น) โมเดลพาร์เชียลเครดิตของมาสเตอร์ (Partial credit model: PCM) (Masters, 1982) (เมื่อ a_i แทนอำนาจจำแนกประจำข้อทุกข้อมีค่าเป็น 1 a'_i ข้อ และ c'_{ik} เป็นความยากประจำขั้น) และโมเดลที่ปรับขยายจากโมเดล PCM อีก ได้แก่ โมเดลเรทติ้งสเกล (Rating scale model: RSM). ของแอนดริช (Andrich. 1978) ซึ่งเป็นการขยายความยากประจำขั้นให้สามารถเขียนให้อยู่ในรูปเชิงเส้นของพารามิเตอร์ความยากประจำข้อและรายการคำตอบได้

สำหรับโมเดลความน่าจะเป็นสะสม (Cumulative probability models) ระดับรายการคำตอบพหุวิภาคถูกแบ่งออกเป็นสองส่วน (ส่วนแรกเป็นรายการคำตอบที่ k ส่วนที่สองเป็น

รายการคำตอบที่ m – k เมื่อ k = 1, 2, …, m – 1) รายการคำตอบภายในแต่ละส่วนถูกยุบและความน่าจะเป็นสะสมถูกคำนวณในแต่ละส่วน ความน่าจะเป็นสะสมของรายการคำตอบ k รายการแรก คือ $P_{ik}^{*}(\theta) = P_{i1}(\theta) + ... + P_{ik}(\theta)$ เมื่อ k = 1, 2, …, m -1 และที่รายการสุดท้ายที่ m-k มีค่าเท่ากับ $1 - P_{ik}^{*}(\theta)$ แล้วจึงใช้โมเดลบอคสำหรับกำหนดลอการิทึมอัตราส่วนแต้มต่อของคู่ของความน่าจะเป็นสะสมดังนี้

$$\ln\left(\frac{1-P_{ik}^{*}(\theta)}{P_{ik}^{*}(\theta)}\right) = a_{ik}''\theta + c_{ik}'', k = 1,2,...,m-1$$

ลอการิทึมอัตราส่วนแต้มต่อ m-1 คู่ อธิบายประเภทของโมเดลพหุวิภาคแบบจัดอันดับอื่นอีก ระดับของรายการคำตอบถูกรักษาไว้โดยการใช้กลุ่มของรายการคำตอบแบบต่อเนื่อง เห็นได้อย่างชัดเจนว่า $P_{ik}^{*}(\theta)$ คือ ความน่าจะเป็นแบบเพิ่มขึ้นทางเดียว เมื่อค่า k เพิ่มขึ้น และลอการิทึมอัตราส่วนแต้มต่อของ $P_{ik}^{*}(\theta)$ ปกติแล้วมีค่ามากกว่าหรือเท่ากับลอการิทึมอัตราส่วนแต้มต่อของ $P_{i(k+1)}^{*}(\theta)$ สำหรับทุกค่า k ดังนั้นจึงได้ข้อสรุปสองประการ ประการแรก เส้นตรงที่แทนฟังก์ชันเชิงเส้นของ θ ในโมเดลจะไม่ตัดกันทุกค่าของ θ ซึ่งจะเป็นจริงเมื่อเส้นนั้นขนานกับเส้นอื่นเท่านั้น เส้นที่ขนานกันแสดงว่า พารามิเตอร์ความชัน a_{ik}'' เท่ากันทุกรายการ k ประการที่สอง เส้นตรงซึ่งที่ขึ้นอยู่กับค่าที่สูงกว่า k ปกติจะอยู่ทางขวาของเส้นที่มีค่าต่ำกว่า k ซึ่งหมายความว่า พารามิเตอร์ส่วนที่ตัดกัน มีการเปลี่ยนแปลงทางเดียวเมื่อค่า k เพิ่มขึ้น และขอบเขตของรายการคำตอบอยู่ในลำดับเดียวกันกับรายการคำตอบ ตัวอย่างของโมเดลในประเภทนี้ได้คือ โมเดลเกรดเรสพอนส์ (Graded response model: GRM) ของซามิไจมา (Samejima. 1969)

สำหรับโมเดลอัตราส่วนต่อเนื่องระดับของรายการคำตอบถูกแบ่งเป็นอัตราส่วนต่อเนื่องอัตราส่วนต่อเนื่องคือ อัตราส่วนระหว่างความน่าจะเป็นของรายการคำตอบ k ต่อ ความน่าจะเป็นสะสมของรายการที่อยู่สูงกว่า k เมื่อ k = 1, 2, …, m-1 ความน่าจะเป็นสะสมของรายการคำตอบ k มีค่าเท่ากับ $1 - P_{ik}^{*}(\theta)$ แล้วใช้โมเดลของบอคสำหรับกำหนดลอการิทึมอัตราส่วนแต้มต่อของอัตราส่วนต่อเนื่องจนได้ระดับของโมเดลการตอบข้อสอบแบบให้คะแนนหลายค่าดังสมการ

$$\ln\left(\frac{1-P_{ik}^{*}(\theta)}{P_{ik}(\theta)}\right) = a_{ik}'''\theta + c_{ik}''', \quad k = 1,2,...,m-1$$

ระดับของรายการคำตอบถูกรักษาไว้โดยการใช้รายการคำตอบต่อเนื่องหนึ่งรายการกับกลุ่มของรายการคำตอบ ตัวอย่างของโมเดลในกลุ่มโมเดลนี้ คือ โมเดลซีเควนท์เชียล (Sequential Model: SM) ของทูส (Tutz ,1990, 1997) (เมื่อพารามิเตอร์ความชันมีค่าเท่ากันทุกรายการคำตอบและทุกข้อ, $a_{ik}''' = a'''$ สำหรับทุก k และ i)

แม้ว่าโครงสร้างของโมเดลพหุวิภาคทั้งสามประเภทมีความคล้ายคลึงกัน แต่เมลเลนเบอร์กได้สรุปว่าการแปลความหมายของพารามิเตอร์ต่างกัน คุณลักษณะของข้อสอบและพัฒนาการรู้

คิดที่รวมอยู่ในข้อสอบแต่ละข้อจะเป็นตัวกำหนดว่าควรใช้โมเดลพหุวิภาคโมเดลใด ซึ่ง แวน เองเกเลนเบอร์ก (Si. 2002. citing Van Engelenburg.1997) ได้กล่าวในอีกทางหนึ่งว่า โมเดลการตอบข้อสอบควรจะสะท้อนคุณลักษณะของข้อสอบ โดยเขาได้ตั้งสมมติฐานว่ากระบวนการแก้ปัญหาของโมเดลพหุวิภาคพัฒนามาจากโมเดลทวิวิภาค และ คุณลักษณะของข้อสอบกำหนดวิธีการนำขั้นตอนแต่ละขั้นเข้ามาเกี่ยวข้อง คุณลักษณะของงานที่เขาได้จำแนกไว้ประกอบด้วย พัฒนาการประจำขั้น พร้อมกัน (simultaneous หรือ ต่อเนื่อง(sequential) กฎความต่อเนื่อง ดำเนินการทั้งหมด(try-all) หรือดำเนินการจนกระทั่งล้มเหลว และกลไกระดับ (กำหนด (fixed) หรือ ไม่กำหนด (not fixed)) การรวมคุณลักษณะของภาระงานเหล่านี้ไว้ด้วยกันจะกำหนดว่าควรจะใช้โมเดลการตอบข้อสอบพหุวิภาคประเภทใด ตัวอย่างเช่น ขั้นตอนของข้อสอบเป็นแบบต่อเนื่องกัน โดยที่กำหนดกลไกระดับไว้ ผู้สอบจะถูกยอมให้ทำแต่ละขั้นจนกระทั่งทำไม่ได้ โมเดลที่นำมาใช้คือ โมเดลอัตราส่วนต่อเนื่อง แอคเคอร์แมน (Akkermans. 1998) ได้ให้เหตุผลอีกทางหนึ่งว่า การเลือกใช้โมเดลการตอบข้อสอบควรสนใจคะแนนสอบมากกว่าข้อสอบและข้อสอบพหุวิภาคควรจะแบ่งตามเกณฑ์การให้คะแนนมากกว่า โมเดลการตอบข้อสอบพหุวิภาคที่ถูกเลือกควรจะสะท้อนกฎการให้คะแนนที่ใช้ โดยแอคเคอร์แมนแบ่งไว้ 3 ประเภทได้แก่ การให้คะแนนเป็นเกรด แบบคู่ขนาน และ แบบเรียงลำดับ การตัดสินทั้งหมดของผลการตอบข้อสอบของผู้สอบการให้คะแนนเป็นเกรด ให้คะแนนภายในมาตรวัด ส่วนการให้คะแนนแบบคู่ขนานให้ คะแนนแต่ละคุณลักษณะที่รวบรวมได้ซึ่งแสดงผลการตอบ คะแนนทั้งหมดของข้อสอบคือผลรวมของคะแนนที่ให้ คะแนนแบบเรียงลำดับ และการให้คะแนนตามการรวบรวมคุณลักษณะที่แสดงในผลการตอบด้วยการกำหนดระดับ คะแนนทั้งหมดเป็นการรวมคุณลักษณะที่ไม่แสดงออก และคุณลักษณะอื่นไม่ถูกพิจารณา จากนิยามกฎการให้คะแนนทั้งสามรูปแบบจึงได้โมเดลพหุวิภาคสามประเภทว่า โมเดลอัตราส่วนต่อเนื่องควรใช้สำหรับการให้คะแนนแบบเรียงลำดับ โมเดลความน่าจะเป็นสะสมเหมาะสมสำหรับการให้คะแนนแบบเกรด และโมเดลรายการคำตอบที่อยู่ติดกันเหมาะสำหรับการให้คะแนนแบบคู่ขนาน แอคเคอร์แมน ได้ให้เหตุผลทางทฤษฎีและทางปฏิบัติสำหรับความเชื่มโยงระหว่างโมเดลและกฎการให้คะแนน เช่น โมเดล GRM และ PCM ไม่ควรใช้กับการให้คะแนนแบบเรียงลำดับ แต่การใช้ SM เหมาะสมมากกว่า แอคเคอร์แมน ได้จำลองข้อมูลสองเวกเตอร์ ซึ่งใช้โมเดลแตกต่างกันแล้วเปรียบเทียบด้วยคอมพิวเตอร์ ผลการศึกษาพบว่า ขนาดตัวอย่างที่ต้องการใช้ในโปรแกรมคอมพิวเตอร์เพื่อให้มีอัตราความถูกต้อง 95% สำหรับเปรียบเทียบด้วยโมเดลที่มีรูปแบบการให้คะแนนเป็นสองเท่าของกลุ่มตัวอย่างที่เปรียบเทียบกับผลเมื่อโมเดลให้คะแนนแบบเดียวกัน ซึ่งชี้ให้เห็นว่ารูปแบบการให้คะแนนต่างกันส่งผลต่อค่าประมาณที่ได้

เฮมเคอร์ (Hemker. 2001) ได้สรุปงานวิจัยในการเปรียบเทียบโมเดลพหุวิภาคสามประเภทตามความหมายของขั้นของข้อสอบ คือ การให้คะแนนแบบสะสม การให้คะแนนแบบมีเงื่อนไข

และ การให้คะแนนแบบบางส่วน ขั้นข้อสอบเป็นแบบทวิวิภาคที่อธิบายโมเดลการตอบข้อสอบพหุวิภาคแบบจัดอันดับ นิยามกำหนดตามวิธีการให้คะแนนในขั้นของข้อสอบโดยเขาแบ่งประเภทของขั้นข้อสอบฟังก์ชันของขั้นการตอบข้อสอบ (Item step response functions: ISRFs) ซึ่งความน่าจะเป็นของการผ่านขั้นข้อสอบเป็นฟังก์ชันของความสามารถ θ รูปทั่วไปของ ISRF ของพารามิเตอร์ของโมเดลโลจิสติกสำหรับข้อสอบแบบให้คะแนนหลายค่าคือ

$$Y_{ik}(\theta) = \frac{\exp\left[a_i(\theta - b_{ik})\right]}{1+\exp\left[a_i(\theta - b_{ik})\right]},\ k = 1,2,...,m-1$$

เมื่อ a_i คือ พารามิเตอร์อำนาจจำแนก และ คือ b_{ik} พารามิเตอร์เฉพาะที่ การแปลความหมายของ b_{ik} ต่างกันตามนิยามของ ISRFs

ISRF ของ ขั้นประจำข้อสะสม k เขียนแทนด้วย $Y_{ik}(\theta) = C_{ik}(\theta) = P(X_i \geq k/\theta)$

เมื่อ X_i คือคะแนนของข้อสอบ i ขั้นประจำข้อเป็นแบบสะสมเพราะว่า ขั้นประจำข้อมีลำดับที่จำเพาะเจาะจง เมื่อผ่านขั้นหนึ่งแล้วถือว่า สามารถผ่านขั้นก่อนหน้าทั้งหมดแต่ถ้าไม่ผ่านขั้นต่อไปที่เหลือถือว่าไม่ผ่านด้วย รายการคำตอบถูกแบ่งเป็นสองส่วน คือส่วนที่ผ่านและส่วนที่ไม่ผ่าน ซึ่งมีความคล้ายคลึงกับ โมเดลความน่าจะเป็นสะสมของเมลเลนเบอร์ก และ วิธีการให้คะแนนคล้ายกับการให้คะแนนเป็นเกรด ของ แอคเคอร์แมน ตัวอย่างของข้อสอบประเภทนี้ได้แก่ การแก้ปัญหาคณิตศาสตร์หลายขั้นตอน เมื่อผู้สอบไม่ผ่านขั้นใดขั้นหนึ่งแล้วย่อมทำให้ขั้นที่เหลือไม่ถูกต้อง

ISRF ของ ขั้นที่ k ประจำข้อแบบมีเงื่อนไขกำหนดโดย

$$Y_{ik}(\theta) = M_{ik}(\theta) = \frac{P(X_i \geq k/\theta)}{P(X_i \geq k-1/\theta)}$$

ขั้นประจำข้อในโมเดลประเภทนี้มีลำดับที่เข้มงวด มีเงื่อนไขการใช้เพราะว่าขั้นประจำข้อจะไม่สามารถดำเนินการได้ถ้าขั้นก่อนหน้านั้นดำเนินการไม่สำเร็จ สำหรับขั้น k มีผู้สอบที่ได้คะแนนสูงกว่าหรือเท่ากับ k -1 เท่านั้นที่มีโอกาสทำได้ เพราะว่าผู้ที่ได้คะแนนต่ำกว่า k -1 ย่อมไม่มีโอกาสทำขั้น k-1 ได้ หรือ ได้ทำแต่ล้มเหลว ดังนั้น $X_i \geq k-1$ จึงเป็นเงื่อนไขของขั้น k โมเดลนี้จึงเทียบเท่ากับโมเดลอัตราส่วนต่อเนื่องของเมลเลนเบอร์ก ข้อสอบในโมเดลนี้ใช้เกณฑ์การให้คะแนนแบบเรียงลำดับตามแนวคิดของ แอคเคอร์แมน ซึ่งได้ให้ตัวอย่างประเภทของการทดสอบที่ให้คะแนนเช่นนนี้เป็นการทดสอบทักษะที่แสดงการกระทำจนกระทั่งเจอความสำเร็จในครั้งแรกหรือทำซ้ำๆจนพบการล้มเหลวครั้งแรก

ISRFของ ขั้น k ของข้อสอบให้คะแนนแบบบางส่วนกำหนดโดย

$$Y_{ik}(\theta) = A_{ik}(\theta) = \frac{P_{ik}(\theta)}{P_{ik}(\theta) + P_{i(k-1)}(\theta)}$$

โมเดลประเภทนี้เทียบเท่ากับ โมเดลรายการคำตอบติดกันของเมลเลนเบอร์ก โดยให้คะแนนแบบบางส่วนในแต่ละขั้นที่ทำได้ เพราะว่าขั้นข้อสอบแสดงผลรวมของงานย่อยหรือการแสดงลักษณะโดยการตอบข้อสอบ และแต่ละคำตอบจะถูกให้คะแนนบางส่วน โมเดลลักษณะนี้ให้คะแนนสอดคล้องกับเกณฑ์การให้คะแนนแบบขนานของแอคเคอร์แมน ตัวอย่างของข้อสอบประเภทนี้คือข้อสอบแบบหลายตัวเลือกและมีการให้คะแนนแบบบางส่วน เช่น ความเรียงที่ให้คะแนนแบบกำหนดเกณฑ์การให้คะแนน ได้ข้อสรุปลักษณะของโมเดลการตอบข้อสอบพหุวิภาคแบบเรียงอันดับตามตาราง 2

ตาราง 2 สรุปประเภทของโมเดลการตอบข้อสอบแบบให้คะแนนหลายค่าแบบเรียงอันดับ

ประเภทโมเดลตามแนวคิดเมลเลนเบอร์ก	เกณฑ์การให้คะแนนตามแนวคิดของแอคเคอร์แมน	ขั้นประจำข้อ ตามแนวคิดของ Hemker	รูปสมการที่แสดงแทนโมเดล
รายการคำตอบที่อยู่ติดกัน Adjacent category	คู่ขนาน Parallel	คะแนนบางส่วน (Partial credit)	partial credit model (PCM)
ความน่าจะเป็นสะสม Cumulative probability	เกรด Graded	คะแนนสะสม (Cumulative)	graded response model (GRM)
อัตราส่วนต่อเนื่อง Continuation ratio	เรียงลำดับ Sequential	คะแนนแบบมีเงื่อนไข (Conditional)	sequential model. (SM)

ที่มา: สี (Si, 2002: 27.Online)

1.2.2.1 โมเดลพาร์เชียลเครดิต (Partial credit model: PCM)

มาสเตอร์ (Masters,1982) ได้ขยายโมเดลทวิวิภาคของโมเดลราสช์เพื่อให้เป็นโมเดลพหุวิภาค โดยสมมติว่ารายการคำตอบถูกเรียงลำดับตามระดับความเชี่ยวชาญของผู้สอบที่มาสเตอร์ได้ให้แนวคิดต่อข้อสอบที่มีหลายขั้นตอนว่า ขั้นตอนการทำข้อสอบแต่ละขั้นแสดงถึงความสามารถที่ต่างกันระหว่างสองขั้นที่อยู่ติดกัน คะแนนแบบบางส่วนถูกให้แต่ละขั้นที่สมบูรณ์ผลของ PCM คือ โมเดลรายการคำตอบที่อยู่ติดกัน สมมติว่า m_i เป็นจำนวนขั้นของข้อสอบ รายการคำตอบของข้อสอบสามารถเขียนแทนด้วยคะแนนบางส่วนที่กำหนดเป็น 0 จนถึง m_i โมเดลถูกอธิบายโดย m_i ลอการิทึมอัตราส่วนแต้มต่อ $\ln\left(\frac{P_{ik}(\theta)}{P_{i(k-1)}(\theta)}\right) = \theta - b_{ik},\ k = 1,2,...,m_i$

ISRF ของแต่ละขั้น คือฟังก์ชันการตอบในโมเดลโลจิสติกแบบ 1 พารามิเตอร์ ค่าพารามิเตอร์ค่าเดียวที่ปรากฏในโมเดลคือ พารามิเตอร์ประจำรายการคำตอบ b_{ik}

$$\frac{P_{ik}(\theta)}{P_{i0}(\theta)} = \frac{P_{ik}(\theta)}{P_{i(k-1)}(\theta)} \cdot \frac{P_{i(k-1)}(\theta)}{P_{i(k-2)}(\theta)} \cdots \frac{P_{i1}(\theta)}{P_{i0}(\theta)} = \prod_{h=1}^{k} \exp(\theta - b_{ih}) = \exp \sum_{h=1}^{k} (\theta - b_{ih})$$

$$\sum_{k=1}^{m_i} \frac{P_{ik}(\theta)}{P_{i0}(\theta)} = \frac{\sum_{k=1}^{m_i} P_{ik}(\theta)}{P_{i0}(\theta)} = \sum_{k=1}^{m_i} \exp \sum_{h=1}^{k} (\theta - b_{ih})$$

เพราะว่า $\sum_{k=1}^{m_i} P_{ik}(\theta) = 1$ ซึ่งเขียนได้เป็น $\sum_{k=1}^{m_i} P_{ik}(\theta) = 1 - P_{i0}(\theta)$

ดังนั้น $$\frac{1 - P_{i0}(\theta)}{P_{i0}(\theta)} = \sum_{k=1}^{m_i} \exp \sum_{h=1}^{k} (\theta - b_{ih})$$

ซึ่งทำให้ $$P_{i0}(\theta) = \frac{1}{1 + \sum_{k=1}^{m_i} \exp \sum_{h=1}^{k} (\theta - b_{ih})}$$

$$P_{ik}(\theta) = P_{i0}(\theta) \cdot \exp \sum_{h=1}^{k} (\theta - b_{ih}) = \frac{\exp \sum_{h=1}^{k} (\theta - b_{ih})}{1 + \exp \sum_{h=1}^{k} (\theta - b_{ih})}$$

เพื่อความสะดวกกำหนดให้ $\sum_{h=0}^{0} (\theta - b_{ih}) \equiv 0$ ซึ่งหมายความว่า

$$\sum_{h=0}^{k} (\theta - b_{ih}) \equiv \sum_{h=1}^{k} (\theta - b_{ih}) \text{ และ } \exp \sum_{h=0}^{0} (\theta - b_{ih}) = 1$$

ส่งผลให้ PCM อยู่ในรูป

$$P_{ik}(\theta) = \frac{\exp \sum_{h=0}^{k} (\theta - b_{ih})}{\exp \sum_{h=0}^{0} (\theta - b_{ih}) + \sum_{k=1}^{m_i} \exp \sum_{h=1}^{k} (\theta - b_{ih})} = \frac{\exp \sum_{h=0}^{k} (\theta - b_{ih})}{\sum_{k=1}^{m_i} \exp \sum_{h=1}^{k} (\theta - b_{ih})}, \; k = 0, 1, 2, ..., m_i$$

สมการข้างต้นอธิบาย ICRF สำหรับการตอบในรายการคำตอบ k ของโมเดล PCM โดยตั้งอยู่บนข้อตกลงว่า ให้คะแนนหนึ่งคะแนนสำหรับการทำสำเร็จในแต่ละระดับ การตอบในรายการที่ k จะได้คะแนนบางส่วนของ k จากที่เป็นไปได้ทั้งหมด m_i รายการคำตอบถูกเรียงอันดับไว้ตามระดับความสามารถที่แสดงในการคำตอบหรือโดยลำดับต่อเนื่องของขั้นข้อสอบที่ต้องการเพื่อทำให้สำเร็จ เมื่อโมเดล PCM ถูกนำไปใช้กับข้อสอบที่มีหลายตัวเลือกสิ่งที่กล่าวมาข้างต้นจึงเป็นไปตามข้อตกลง

ค่าความยากเฉพาะที่ b_{ik} สามารถแบ่งออกไปบ่งชี้ค่าความยากประจำข้อและช่วงรายการคำตอบเช่น $b_{ik} = b_i + d_{ik}$. ผลต่างในระดับของความเชี่ยวชาญระหว่างรายการคำตอบk กับ k +1 เรียกว่าความยากประจำขั้น k หรือ เทรสโฮล d_{ik} ซึ่งเทรสโฮล dik มีค่าเท่ากับเส้นต่อเนื่องความสามารถ $(\theta - b_i)$ ที่เส้น ICCCs ของสองรายการติดกันตัดกัน ผู้สอบที่มีความสามารถ $\theta = b_{ik}$ จะมีความน่าจะเป็นในการเลือกสองรายการติดกันนั้นเท่ากันทั้งคู่ ดังนั้น เทรสโฮล dik จึงเหมือน

ขอบเขตระหว่างสองรายการคำตอบนี้ โมเดล PCM เทรสโฮลไม่จำเป็นต้องเรียงลำดับกัน กล่าวคือขั้นที่ยากกว่าสามารถอยู่หลังจากขั้นที่ง่ายกว่าได้ หรือ อยู่ก่อนขั้นที่ง่ายกว่าก็ได้ อย่างไรก็ตามโครงร่างของเทรสโฮลมีผลต่ออำนาจจำแนกของข้อสอบเพราะว่าการได้รับคะแนนเท่ากันในแต่ละขั้นซึ่งอาจจะมีความยากต่างกัน

1.2.2.2 **โมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตโมเดล** (Generalized partial credit model: GPCM)

มูรากิ (Muraki. 1992) ได้พัฒนาโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตโมเดล โดยการเพิ่มพารามิเตอร์ความชันเข้าไปในโมเดล พารามิเตอร์ความชันมีความคล้ายคลึงกับอำนาจจำแนกของโมเดลโลจิสติกแบบสองพารามิเตอร์ สำหรับโมเดล GPCM พารามิเตอร์ความชันมีค่าคงที่ในทุกรายการคำตอบของข้อสอบในข้อสอบข้อเดียวกัน แต่ สามารถแตกต่างกันได้เมื่อข้อสอบต่างกัน สำหรับโมเดลการตอบข้อสอบสำหรับข้อสอบแบบให้คะแนนสองค่า แบบสองพารามิเตอร์หรือสามพารามิเตอร์ พารามิเตอร์ความชันในการรวมกันของโครงร่างของเทรสโฮลซึ่งกำหนดค่าอำนาจจำแนกแต่ละข้อ ICRFs ของ GPCM จึงคล้ายกับ PCM ดังนี้

$$P_{ik}(\theta)=\frac{\exp\sum_{h=0}^{k}Da_i(\theta-b_i)}{\sum_{k=0}^{m_i}\exp\sum_{h=0}^{k}Da_i(\theta-b_i)},\quad k=0,1,2,...,m_i$$

โดยที่ a_i เป็นพารามิเตอร์ความชัน ซึ่งก็คือ อำนาจจำแนกข้อสอบ และมี D = 1.7 เป็นค่าคงที่ที่แปลงมาตรวัดความสามารถให้อยู่ระบบมาตรวัดเดียวกันโมเดลโค้งปกติสะสม ICRFs ของ GPCM มีที่มาคล้ายกับ PCM โดยเริ่มจาก ลอการิทึมอัตราส่วนแต้มต่อ

$$\ln\left(\frac{P_{ik}(\theta)}{P_{i(k-1)}(\theta)}\right)=a_i(\theta-b_{ik}),\ k=1,2,...,m_i$$

โดยไม่จำเป็นต้องมีการเรียงลำดับรายการสมการข้างซ้ายจึงมีค่าเป็น

$$\ln\left(\frac{P_{ik}(\theta)}{P_{i(k-1)}(\theta)}\right)=\ln\left(\frac{\exp(a_{ik}\theta+c_{ik})}{\exp(a_{i(k-1)}\theta+c_{i(k-1)})}\right)=(a_{ik}-a_{i(k-1)})\theta+(c_{ik}-c_{i(k-1)})$$

ดังนั้นโมเดลรายการคำตอบแบบนามบัญญัติจึงถูกควบคุมให้มีความเข้มงวดมากกว่าโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิต จึงได้ $(a_{ik}-a_{i(k-1)})\theta+(c_{ik}-c_{i(k-1)})=a_i(\theta-b_{ik})$ และความสัมพันธ์ระหว่างพารามิเตอร์ในโมเดลทั้งสองเขียนได้ว่า

$$a_i=(a_{ik}-a_{i(k-1)})\text{ และ }b_{ik}=-\frac{(c_{ik}-c_{i(k-1)})}{(a_{ik}-a_{i(k-1)})}$$

ซึ่งเป็นการยืนยันข้อสังเกตของ ทิซเซน และสเตียนเบอร์ก (Thissen; & Steinberg. 1984) และ บอค (Bock. 1997)

จากที่กล่าวมาข้างต้นโมเดลนอมินอลของบอคเป็นโมเดลในรูปทั่วไปของโมเดลพหุวิภาค ซึ่งมีความสัมพันธ์กับโมเดลทวิวิภาค ซึ่งจากการศึกษาของแอคเคอร์แมน (Akkerman, 1998) พบว่าโมเดลพหุวิภาคมีความแตกต่างตามทฤษฎีและการปฏิบัติ การให้คะแนนข้อสอบแบบใดเป็นตัวกำหนดว่าควรใช้โมเดลใดในการประมาณค่าพารามิเตอร์ความสามารถ หากใช้โมเดลผิดย่อมเกิดความคลาดเคลื่อนขึ้นและเกิดความลำเอียงในการใช้สำหรับประมาณค่าความสามารถผู้สอบ และเมลเลนเบอร์ก (Mellenbergh. 1995) ได้พบว่าโมเดลทั้งสามประเภทให้ผลไม่แตกต่างกันเมื่อมีการให้คะแนนแบบสองค่า

การเลือกใช้โมเดลการตอบข้อสอบแบบตรวจให้คะแนนแบบหลายค่า สิ่งที่ควรคำนึงถึงในทางปฏิบัติ (ศิริชัย กาญจนวาสี, 2545: 101) ได้แก่

1. ปรัชญาความเชื่อเกี่ยวกับโมเดล และจุดมุ่งหมายของการนำผลไปใช้ของผู้พัฒนาแบบทดสอบ
2. ความเป็นวิวิธพันธ์ของกลุ่มตัวอย่าง (Heterogeneous sample) และขนาดกลุ่มตัวอย่างต้องใหญ่เพียงพอที่จะทำให้ความคลาดเคลื่อนมาตรฐานของการประมาณค่าอยู่ในระดับที่ยอมรับได้ตามเป้าหมายของการนำผลไปใช้
3. เลือกใช้แบบแผนการตอบที่สะดวกและสามารถตรวจให้คะแนนได้ง่ายอย่างเป็นปรนัย
4. ข้อมูลที่นำมาวิเคราะห์จะต้องมีการตอบทุกรายการ

2. การประมาณค่าความสามารถ (Ability Estimation)

วิธีการประมาณค่าความสามารถตามทฤษฎีการตอบข้อสอบมีหลายวิธี พารามิเตอร์ความสามารถถูกประมาณค่าร่วมกับพารามิเตอร์ข้อสอบ หรือประมาณค่าเมื่อทราบพารามิเตอร์ข้อสอบแล้ว ถ้าค่าประมาณความสามารถไม่ถูกประมาณค่าร่วมกับพารามิเตอร์ข้อสอบ ค่าพารามิเตอร์ข้อสอบจะถูกประมาณค่าจากแบบแผนรายการคำตอบเป็นอันดับแรกโดยกำจัดอิทธิพลของค่าพารามิเตอร์ผู้สอบออกไป เทคนิคที่ใช้ในการประมาณค่าพารามิเตอร์ประกอบด้วย (Si; & Shumacker. 2004) วิธีความน่าจะเป็นสูงสุด (Maximum Likelihood method) (Baker, 1992) วิธีการถดถอยโลจิสติก (Logistic regression) (Reynolds; Perkins; & Brutten. 1994) วิธีไคสแควร์ต่ำสุด (Minimum chi-quadrant) (Zwinderman; & van der Wollenberg.1990) และ วิธีการประมาณค่าแบบเบเซียน (Bayesian modal estimation procedure) (Mislevy. 1986; Baker. 1992)

2.1 วิธีความน่าจะเป็นสูงสุด (Maximum Likelihood method)

ความน่าจะเป็น คือ ฟังก์ชันเกี่ยวกับความเป็นไปได้ของรูปแบบการตอบข้อสอบที่สังเกตได้ เวกเตอร์การตอบข้อสอบที่กำหนดให้เป็นกรณีหนึ่งของโมเดลการตอบข้อสอบ เมื่อข้อตกลงเรื่องความอิสระเฉพาะที่ยังคงอยู่ ฟังก์ชันความน่าจะเป็นสูงสุดคือผลคูณของความน่าจะเป็นที่เกี่ยวข้องกับเวกเตอร์การตอบรายบุคคล และเนื่องจากความน่าจะเป็นของการตอบข้อสอบเป็นฟังก์ชันของพารามิเตอร์ความสามารถและพารามิเตอร์ข้อสอบ ดังนั้นฟังก์ชันความน่าจะเป็นจึงจัดว่าเป็นฟังก์ชันของค่าพารามิเตอร์เหล่านั้นด้วยเช่นกัน เช่น ฟังก์ชันความน่าจะเป็นของผู้สอบคนที่ j ที่จะตอบข้อสอบ n ข้อโดยใช้โมเดลโลจิสติกแบบสองพารามิเตอร์ คือ

$$L(x/\theta_j, a, b) = \prod_{i=1}^{n} f\left(x_i/\theta_j, a_i, b_i\right)$$

โดยที่ $\mathbf{x} = (x_1, x_2, ..., x_n)$ เป็นเวกเตอร์คำตอบของผู้สอบคนที่ j ในการตอบข้อสอบ n ข้อ ($x_i = 1$ หรือ 0 สำหรับทุกข้อสอบ i) $\mathbf{a} = (a_1, a_2, ..., a_n)$ และ $\mathbf{b} = (b_1, b_2, ..., b_n)$ เป็นเวกเตอร์พารามิเตอร์ของข้อสอบ และ f เป็นฟังก์ชันการตอบข้อสอบของโมเดลโลจิสติกสองพารามิเตอร์ ดังนั้น L เป็นความน่าจะเป็นของการได้เวกเตอร์คำตอบ $\mathbf{x}$ เมื่อให้พารามิเตอร์ θ_j, $\mathbf{a}$, และ $\mathbf{b}$ การประมาณค่าความน่าจะเป็นสูงสุดถูกใช้เพื่อค้นหาค่าของพารามิเตอร์ที่ทำให้ค่าของ L มีค่ามากที่สุด เนื่องจากฟังก์ชันความน่าจะเป็น L ประกอบด้วยผลคูณของฟังก์ชัน การคำนวณสามารถทำได้ง่ายโดยใช้ลอการิทึมของฟังก์ชัน L แทน ฟังก์ชัน L เดิม ฟังก์ชันลอการิทึม คือ ฟังก์ชันทางเดียวซึ่ง ln L มีค่าสูงสุดเมื่อ L เข้าใกล้ค่าสูงสุด ลอการิทึมของฟังก์ชันความน่าจะเป็นจึงถูกใช้เพื่อหาคำตอบ ในทางปฏิบัติตัวประมาณค่าของพารามิเตอร์ได้มาจากคำตอบของสมการอนุพันธ์อันดับหนึ่ง $\frac{\partial \ln L}{\partial \theta} = 0$ (สมการความน่าจะเป็น) ซึ่งเป็นเงื่อนไขสำหรับค่าสูงสุดเฉพาะที่ที่เกิดขึ้น ซึ่งถ้ามีค่าเฉพาะที่ที่มากกว่าหนึ่งค่า จะเลือกค่าที่มากที่สุด สำหรับเซตคำตอบของผู้สอบ n คนที่ทำแบบทดสอบที่ให้คะแนนสองค่า ฟังก์ชันความน่าจะเป็น กำหนดไว้ว่า

$$L(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N/\theta_1, \theta_2, ..., \theta_N) = \prod_{j=1}^{N} L(x_j/\theta_j) = \prod_{j=1}^{N}\prod_{i=1}^{n} L(x_{ij}/\theta_j) = \prod_{j=1}^{N}\prod_{i=1}^{n} P_{ij}^{x_{ij}} Q_{ij}^{1-x_{ij}}$$

จากนั้นใส่ Log ทั้งสองข้างของสมการ

$$\ln L(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N/\theta_1, \theta_2, ..., \theta_N) = \sum_{j=1}^{N}\sum_{i=1}^{n}[x_{ij} \ln P_{ij} + (1-x_{ij}) \ln(1-P_{ij})]$$

ค่าประมาณที่มีความน่าจะเป็นสูงสุดของพารามิเตอร์ความสามารถ θ ก็จะได้มาพร้อมกันจากการแก้สมการ

$$\frac{\partial}{\partial \theta_j} \ln L(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N/\theta_1, \theta_2, ..., \theta_N) = 0 \text{ สำหรับ } j = 1, 2, ..., N$$

คำตอบของสมการความน่าจะเป็นสามารถแก้ได้ด้วยการใช้ ขั้นตอนวิธีของนิวตัน – ราฟสัน (Newton-Raphson) เมื่อ ฟังก์ชันความน่าจะเป็นสามารถหาอนุพันธ์อันดับสองได้ เช่น อนุพันธ์อันดับสองฟังก์ชันความน่าจะเป็นสามารถหาค่าได้ด้วยวิธีนิวตัน – ราฟสัน โดยเริ่มจาก ค่าเริ่มต้นสำหรับค่าประมาณของพารามิเตอร์ในโมเดล จำนวนของข้อสอบที่ตอบถูก ถูกนำไปประมาณค่าความสามารถ และใช้ค่าสถิติรายข้อตามทฤษฎีการทดสอบแบบมาตรฐานเดิมเช่น สัดส่วนการตอบถูกและสหสัมพันธ์ไบซีเรียลประมาณค่าพารามิเตอร์ข้อสอบ ในแต่ละรอบของการทำซ้ำ ค่าประมาณใหม่สำหรับพารามิเตอร์ถูกกำหนดบนฐานของค่าประมาณจากการดำเนินการก่อนหน้านั้น ตัวอย่างเช่น เมื่อกำหนดให้ $\left[\hat{\theta}_j\right]$ เป็นค่าประมาณความสามารถของผู้สอบคนที่ j ที่ระดับการทำซ้ำครั้งที่ t^{th} ดังนั้นค่าประมาณความสามารถสำหรับการทำซ้ำครั้งที่ $(t+1)^{th}$ คือ

$$\left[\hat{\theta}_j\right]_{t+1} = \left[\hat{\theta}_j\right]_t - \left[\frac{\partial \ln L(\mathbf{x}/\theta)}{\partial \theta_j}\right]_t \Big/ \left[\frac{\partial^2 \ln L(\mathbf{x}/\theta)}{\partial \theta_j^{\,2}}\right]_t$$

ผลต่างระหว่างค่าประมาณตัวเก่าและใหม่ $\left(\left[\hat{\theta}_j\right]_{t+1} - \left[\hat{\theta}_j\right]_t\right)$ ถูกคำนวณในแต่ละครั้งหมุนเวียนไปจนกระทั่ง ผลต่างมีค่าต่ำกว่าค่าต่ำสุดที่กำหนดไว้ก่อนหน้า ดังนั้น ค่าประมาณจึงค่อยๆ มาบรรจบกัน และเป็นค่าประมาณพารามิเตอร์สูงที่สุดที่เป็นไปได้ เบเคอร์ (Baker, 1992) ได้ให้รายละเอียดการหาอนุพันธ์สำหรับประมาณค่าด้วยขั้นตอนวิธีนิวตัน-ราฟสันไว้เพื่อใช้กับโมเดลการตอบข้อสอบแบบทวิวิภาคต่างๆซึ่งพบว่าวิธีการประมาณค่าแบบความน่าจะเป็นสูงสุดที่นิยมใช้กับโมเดลการตอบข้อสอบมี สามวิธี ได้แก่ วิธีความน่าจะเป็นสูงสุดร่วมกัน (Joint Maximum Likelihood: JML) วิธีความน่าจะเป็นสูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood: CML) และ วิธีความน่าจะเป็นสูงสุดแบบมีขอบเขต (Marginal Maximum Likelihood: MML)

2.1.1 วิธีความน่าจะเป็นสูงสุดร่วมกัน (Joint Maximum Likelihood: JML)

วิธีการประมาณค่าแบบJML พัฒนาโดยเบิร์นบวม (Birnbaum. 1968) เขาได้การประมาณค่าซ้ำ โดยที่แต่ละการทำซ้ำมีสองขั้นตอนสำหรับการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถร่วมกัน ขั้นตอนแรกเป็นการประมาณค่าพารามิเตอร์ด้วยค่าพารามิเตอร์เริ่มต้นของพารามิเตอร์ข้อสอบที่ทราบค่าแล้ว ขั้นตอนที่สองเป็นนำค่าสุดท้ายในการประมาณค่า ไปกำหนดให้เป็นพารามิเตอร์ความสามารถที่ทราบค่าเพื่อประมาณพารามิเตอร์ผู้สอบในรอบต่อไป ขั้นตอนทั้งสองขั้นนี้ถูกกระทำซ้ำกันไปเรื่อยๆ จนกระทั่งทั้งสองค่าประมาณพารามิเตอร์ข้อสอบและค่าประมาณพารามิเตอร์ความสามารถมาบรรจบกัน วิธี JML จึงเป็นวิธีการดำเนินการโดยตรงไปตรงมา แต่อย่างไรก็ตามวิธี JML มีข้อจำกัดในการใช้ ประการแรก ค่าประมาณพารามิเตอร์ใน JML มีความไม่สอดคล้องกัน ค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถถูกประมาณค่าร่วมกัน ค่าพารามิเตอร์ข้อสอบเป็นพารามิเตอร์เชิงโครงสร้างที่ถูกกำหนดตามความยาวของแบบทดสอบ และพารามิเตอร์

ความสามารถเป็นพารามิเตอร์ที่เกิดขึ้นโดยบังเอิญเนื่องจากจำนวนของพารามิเตอร์ความสามารถเพิ่มขึ้นในขณะที่ขนาดกลุ่มตัวอย่างเล็กลง เนย์แมน และ สกอต (Neyman and Scott , 1948, citing Si, 2002) ได้แสดงให้เห็นว่าปริมาณค่าพารามิเตอร์ที่เกิดขึ้นโดยบังเอิญจำนวนมากส่งผลต่อความสอดคล้องของการประมาณค่าของพารามิเตอร์เกี่ยวกับโครงสร้างซึ่งหมายความว่า ค่าประมาณพารามิเตอร์ข้อสอบจะไม่ลู่เข้าสู่ค่าจริงเมื่อจำนวนผู้สอบมีขนาดเพิ่มขึ้น นอกจากนั้น พบว่าค่าประมาณพารามิเตอร์ในวิธี JML มีความลำเอียง เดอ กรูจเตอร์ (De Gruijter. 1990. citing Si. 2002) พบว่าความลำเอียงของค่าประมาณพารามิเตอร์ ซึ่งเป็นความแตกต่างระหว่างค่าพารามิเตอร์จริงและค่าประมาณพารามิเตอร์ นั้นขึ้นอยู่กับ การแจกแจงคะแนนรวมที่คาดหวัง ขณะที่แบบทดสอบมีความยาวสูงขึ้น ความลำเอียงมีความสำคัญลดลง เนื่องจาก พารามิเตอร์ความสามารถสามารถประมาณค่าได้อย่างถูกต้องมากกว่า จึงต้องใช้กลุ่มตัวอย่างขนาดใหญ่ และแบบทดสอบขนาดยาวเพื่อลดความลำเอียงเมื่อใช้การประมาณค่าวิธี JML ด้วยเหตุนี้วิธี JML จึงมีความนิยมใช้ไม่มากโดยเฉพาะอย่างยิ่ง เมื่อศึกษาด้วยเครื่องมือวัดขนาดเล็ก นอกจากนี้ วิธี JML ไม่สามารถใช้ประมาณค่ากรณีที่ผู้สอบได้คะแนนเต็ม หรือ ศูนย์คะแนน หรือข้อสอบข้อนั้นมีผู้ตอบถูกทุกคน หรือไม่มีผู้ตอบถูกเลย ผู้สอบที่ได้คะแนนศูนย์จะมีค่าประมาณความสามารถเป็น $-\infty$ ในขณะที่ ผู้สอบได้คะแนนเต็มจะมีค่าประมาณความสามารถเป็น $+\infty$ และในทำนองเดียวกัน ข้อสอบที่ไม่มีผู้ตอบถูกจะมีค่าประมาณความยากเป็น $+\infty$ ในขณะที่ ข้อสอบที่ตอบถูกทุกคนจะมีค่าประมาณความยาก $-\infty$

ข้อจำกัด (Hambleton; Swaminathan; & Rogers. 1991)

1. ไม่สามารถประมาณค่า θ สำหรับผู้ตอบที่ได้คะแนนเต็มหรือศูนย์
2. ไม่สามารถประมาณค่าพารามิเตอร์ของข้อสอบสำหรับข้อสอบที่มีผู้ตอบถูกหมดทุกคนหรือตอบผิดหมดทุกคน (ก่อนการวิเคราะห์ต้องกำจัดปัญหาตามข้อที่1, 2 ออกให้หมด)
3. ค่าพารามิเตอร์ที่ประมาณค่าได้สำหรับโมเดล 2 และ 3 พารามิเตอร์ จะมีค่าคงเส้นคงวาก็ต่อเมื่อ มีจำนวนข้อสอบมากและกลุ่มผู้เข้าสอบมีขนาดใหญ่
4. การประมาณค่าพารามิเตอร์สำหรับโมเดล 3 พารามิเตอร์ อาจมีปัญหาถ้าไม่จำกัดค่า θ และพารามิเตอร์ของข้อสอบ ทางแก้หนึ่ง คือ ประมาณค่าพารามิเตอร์ด้วยวิธีของเบส์ (Bayesian estimates) โดยมีการกำหนดการแจกแจงเบื้องต้นของค่าพารามิเตอร์

2.1.1.1 **วิธีความน่าจะเป็นสูงสุดแบบมีเงื่อนไข** (Conditional Maximum Likelihood: CML)

ค่าประมาณพารามิเตอร์วิธี JML ได้ค่าประมาณที่อาจเกิดความไม่สอดคล้องและมีความลำเอียงได้ เพราะว่าค่าพารามิเตอร์เหล่านี้ถูกประมาณค่าร่วมกับพารามิเตอร์ความสามารถ แอนเดอร์สัน (Andersen. 1972) จึงได้พัฒนาวิธี CML ขึ้นเพื่อให้ค่าประมาณพารามิเตอร์เกิดความสอดคล้องและมีประสิทธิภาพ โดยแยกตัวประกอบที่ไม่ทราบค่าพารามิเตอร์ความสามารถออกจากสมการความน่าจะเป็นก่อน สิ่งที่ต้องการคือ ค่าสถิติเพียงพอ (Sufficient statistics) สำหรับพารามิเตอร์ความสามารถและผู้สอบ ซึ่งสามารถพบได้ใน โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ (1-PL) เท่านั้น จำนวนของข้อสอบที่ตอบถูกเป็นสถิติเพียงพอสำหรับพารามิเตอร์ความสามารถ และจำนวนของคำตอบที่ถูกต้องเป็นสถิติเพียงพอสำหรับพารามิเตอร์ความยากข้อสอบ

วิธี CML แทนที่ฟังก์ชันความน่าจะเป็น $L(\mathbf{x}/\boldsymbol{\theta})$ ด้วย $L(\mathbf{x}/\mathbf{r})$ เมื่อ $\mathbf{x}$ แทนเวกเตอร์คำตอบของแต่ละคนในกลุ่มตัวอย่าง $\mathbf{r}$ แทน เวกเตอร์ที่ประกอบด้วยจำนวนการตอบถูกของผู้สอบแต่ละคน แสดงความสพัมพันธ์ได้ว่า $L(\mathbf{x}/\mathbf{r}) = L(\mathbf{x}/\boldsymbol{\theta})/L(\mathbf{r}/\boldsymbol{\theta})$ ซึ่งเป็นความอิสระของ θ เพราะว่าเทอมในจำนวนที่เป็นเศษ และจำนวนในตัวส่วน ถูกขจัดออกในแต่ละส่วนออกจากกันแม้ว่าวิธีประมาณค่า CML จะมีข้อดีในการแยกประมาณค่าพารามิเตอร์ความสามารถและพารามิเตอร์ข้อสอบแต่มีข้อจำกัด ประการแรก ไม่สามารถประมาณค่าพารามิเตอร์เมื่อผู้สอบและข้อสอบมีคะแนนเป็นศูนย์ หรือคะแนนเต็ม ประการที่สอง แม้ว่าผู้สอบมีจำนวนข้อสอบที่ตอบถูกเท่ากันแต่ตอบถูกในข้อสอบที่ต่างกัน เมื่อประมาณค่าด้วย CML กลับถูกประมาณค่าให้มีความสามารถเท่ากัน และประการที่สาม วิธี CML มักประสบปัญหาเมื่อแบบทดสอบมีขนาดยาว การสูญหายของข้อมูลที่ซับซ้อน และ จำนวนรายการคำตอบของข้อสอบพหุวิภาคมีจำนวนมาก

2.1.1.2 **วิธีความน่าจะเป็นสูงสุดแบบมีขอบเขต** (Marginal Maximum Likelihood: MML)

วิธีประมาณค่า CML ยอมให้ ค่าประมาณพารามิเตอร์ข้อสอบอิสระจากเงื่อนไขของพารามิเตอร์ความสามารถ ซึ่งต้องการค่าสถิติเพียงพอสำหรับประมาณค่าพารามิเตอร์ความสามารถ ความจำกัดของเงื่อนไขกับการใช้เฉพาะโมเดล 1-PL เท่านั้น บอค และ ไลเบอร์แมน (Bock; & Lieberman.1970) ได้ แนะนำวิธี การจัดการพารามิเตอร์ความสามารถที่ไม่ทราบค่า โดยใช้ฟังก์ชันความน่าจะเป็นแบบมีเงื่อนไขแทนที่ขึ้นอยู่กับความสามารถผู้สอบคนที่ i แทน เช่น $L(\mathbf{x}/\boldsymbol{\theta})$ บอค และ ไลเบอร์แมน ได้แนะนำฟังก์ชันความน่าจะเป็นแบบไม่มีเงื่อนไข, $L(\mathbf{x})$ ซึ่งเป็น ความน่าจะเป็นของแบบรูปการตอบที่สังเกตได้ $\mathbf{x}$ จากกลุ่มผู้สอบที่ไม่ทราบค่าพารามิเตอร์ที่ถูกสุ่มเลือกมาจากประชากร ข้อมูลการตอบที่สังเกตได้ถูกจัดให้เป็นตัวอย่างที่สุ่มจากประชากร เมื่อประชากรมีการแจก

แจงความสามารถที่อธิบายโดยฟังก์ชันความหนาแน่นต่อเนื่อง $g(\theta)$,ฟังก์ชันความน่าจะเป็นแบบไม่มีเงื่อนไขถูกกำหนดโดยการหาปฏิยานุพันธ์แบบจำกัดเขต

$$L(\mathbf{x}) = \int_{-\infty}^{+\infty} L(\mathbf{x}/\theta) \cdot g(\theta) d\theta$$

เมื่อ $L(\mathbf{x})$ เป็น ฟังก์ชันของพารามิเตอร์ข้อสอบเท่านั้น เพราะว่าพารามิเตอร์ความสามารถถูกขจัดออกไป การหาปฏิยานุพันธ์แบบจำกัดเขตในรูปทั่วไปไม่สามารถแสดงในรูปปิดได้ดังนั้น จึงใช้ขั้นตอนที่เรียกว่า Gaussian quadrature เพื่อค้นหาค่าของหาปฏิยานุพันธ์เนื่องจากฟังก์ชันความน่าจะเป็นแบบไม่มีเงื่อนไขเป็นค่าที่คาดหวังบนกลุ่มประชากรแทนค่าที่สังเกตได้ ขั้นตอนการประมาณค่า จึงต่างไปจากการใช้ในวิธี JML วิธีการเริ่มแรกที่ Bock and Lieberman (1970) ได้แนะนำไว้ยากที่จะดำเนินการด้วยคอมพิวเตอร์ได้ และไม่สามารถปฏิบัติได้กับแบบทดสอบขนาดยาว ต่อมาจึงได้นำวิธี MML มาใช้อีกครั้ง บอค และ อิทคิน (Bock; &Aitken.1981) ได้ใช้ขั้นตอนวิธี EM ซึ่งเป็นวิธีการสำหรับการประมาณค่าแบบ MML ขั้นตอนวิธี EM มีสองขั้น คือ ขั้นการคาดหวัง และขั้นการทำให้มากที่สุด ในขั้นแรก ค่าของความถี่ที่คาดหวังที่จุดการสร้างสี่เหลี่ยมจัตุรัส และความถี่ที่คาดหวังของผู้สอบที่สอบผ่านถูกคำนวณ ค่าคาดหวังเหล่านี้ถูกนำมาคำนวณในสมการการประมาณค่าสำหรับการประมาณค่าความน่าจะเป็นสูงสุดในขั้นการทำให้มากที่สุด ทั้งสองขั้น จะถูกทำซ้ำไปจนกระทั่งบรรจบกัน แล้วใช้ วิธี Newton-Gauss แก้สมการความน่าจะเป็นสูงสุด

วิธี MML เหนือกว่าวิธีความน่าจะเป็นสูงสุดเหมาะสมสำหรับใช้กับทุกโมเดลการตอบข้อสอบ และเหมาะสมกับแบบทดสอบทุกขนาด สามารถให้ค่าประมาณเมื่อผู้สอบและข้อสอบได้คะแนนเต็มหรือศูนย์คะแนน จึงทำให้ไม่เกิดการสูญเสียสารสนเทศของผู้สอบและข้อสอบ ค่าประมาณของความคลาดเคลื่อนมาตรฐานของข้อสอบเป็นค่าประมาณที่เหมาะสมของความแปรปรวนของกลุ่มตัวอย่างของค่าประมาณที่คาดหวัง อย่างไรก็ตามวิธี MML มีข้อจำกัดเล็กน้อย ประการแรก MML เป็นการคำนวณที่ซับซ้อนและต้องใช้ประสบการณ์ ประการที่สอง การแจกแจงความสามารถไม่ทราบมาก่อนมันจึงถูกสมมติให้มีการแจกแจงปกติ อย่างไรก็ตาม ผลที่ได้จึงได้จากการแจกแจงปกติมีค่าเพียงเล็กน้อยและการแจกแจงความสามารถของประชากรโดยแท้จริงสามารถถูกประมาณค่าจากข้อมูล ประการที่สาม เบเกอร์ (Baker. 1992) ชี้ว่า MML ใช้ขั้นตอนวิธี EM แก้ปัญหาความไม่สอดคล้องของพารามิเตอร์ข้อสอบในวิธี JML แต่ยังคงมีปัญหาค่าประมาณความสามารถที่แปลกๆในข้อมูลที่หลงเหลือมาจากการแก้ปัญหาไม่ได้ จึงไม่มีวิธีที่ง่ายในการประมาณค่าพารามิเตอร์ที่มีรูปแบบการตอบที่เบี่ยงเบนไปจากปกติ ดังนั้นด้วยเหตุนี้จึงมีการพัฒนา วิธีการประมาณค่าแบบเบเชียน ใน MML (Bock; & Aitkin. 1981; Mislevy. 1986)

2.2 วิธีการประมาณค่าแบบเบเซียน (Bayesian estimation)

วิธีการประมาณค่าแบบเบเซียนพัฒนามาจากทฤษฎีของเบส์ (Bayes) ซึ่งเน้นความสัมพันธ์ระหว่างความน่าเป็นแบบมีเงื่อนและไม่มีเงื่อนไขของการเกิดขึ้นของเหตการณ์สามารถเขียนเป็น $P(A/B) \propto P(B/A) \cdot P(A)$ เมื่อ P(A|B) และ P(B|A) เป็นความน่าจะเป็นแบบมีเงื่อนไข และ P(A) เป็นความน่าจะเป็นแบบไม่มีเงื่อนไขของการเกิดขึ้นของเหตุการณ์ ฟังก์ชันความน่าจะเป็น $L(\mathbf{x}/\boldsymbol{\theta})$ เป็นความน่าจะเป็นแบบมีเงื่อนไขผลที่ตามมาจากทฤษฎีเบส์คือ

$$P(\theta/\mathbf{x}) \propto L(\mathbf{x}/\theta) \cdot P(\theta)$$

เมื่อ $P(\theta/\mathbf{x})$ เป็นความน่าจะเป็นของความสามารถ θ ภายใต้เงื่อนไขของเวกเตอร์การตอบ $\mathbf{x}$ ซึ่งเป็นที่เข้าใจว่าเป็นการแจกแจงของค่าประมาณความสามารถที่สอดคล้องกับการตอบข้อสอบ

$P(\theta)$ คือ การแจกแจงของความสามารถก่อนการประมาณค่า เช่น ฟังก์ชันความหนาแน่นของการการแจกแจงความสามารถก่อน การแจกแจงความสามารถหลังคือ สัดส่วนของผลคูณของฟังก์ชันความน่าจะเป็นและการแจกแจงความสามารถก่อน

วิธีการประมาณค่าแบบ Bayesian estimation ใช้ความสัมพันธ์นี้และสารสนเทศจากการแจกแจงความสามารถก่อนการประมาณค่าในการประมาณค่าพารามิเตอร์ความสามารถ การแจกแจงความสามารถก่อนการประมาณค่าจะมีสาระสำคัญเมื่อความแปรปรวนของการแจกแจงมีค่าเล็กน้อย อย่างไรก็ตามถ้าหากความแปรปรวนมีค่ามากจะทำให้การแจกแจงความสามารถก่อนการประมาณค่าไม่ให้ค่าสารสนเทศ ในอีกทางหนึ่งค่าสารสนเทศของการแจกแจงความสามารถก่อนการประมาณค่าจะลดค่าประมาณความสามารถลงไปจนถึงค่าเฉลี่ยของการแจกแจงก่อนการประมาณค่าและป้องกันค่าประมาณความสามารถแตกต่างกันไปจนเป็นค่าที่ไม่สามารถเป็นเหตุผลได้ในขั้นตอนการประมาณค่า คุณสมบัติของวิธีเบเซียนช่วยแก้ปัญหาในการประมาณค่าความน่าจะเป็นสูงสุด และให้ค่าเฉลี่ยสำหรับการประมาณค่าที่มีรูปแบบคำตอบที่เบี่ยงเบนไปจากปกติ เช่น ศูนย์คะแนน หรือคะแนนเต็ม วิธีการประมาณค่าสองวิธีถูกใช้ร่วมกันในการประมาณค่าแบบ MML ได้แก่ ค่าประมาณสูงสุดหลังการประมาณค่า (Maximum a posteriori: MAP) และค่าประมาณหลังการประมาณค่าที่คาดหวัง(Expected a posteriori: EAP)

2.2.1 ค่าประมาณสูงสุดหลังการประมาณค่า (Maximum a posteriori: MAP)

MAP คือ ค่าประมาณความสามารถ θ ที่สูงสุดของลอการิทึมของฟังก์ชันความหนาแน่นของการแจกแจงความสามารถหลังการประมาณค่า

$$\ln P(\theta/\mathbf{x}) = \ln\left[L(\mathbf{x}/\theta) \cdot P\left(\theta\right)\right] = \ln L(\mathbf{x}/\theta) + \ln P\left(\theta\right)$$

ค่าประมาณ MAP estimate มักถูกเรียกว่าเป็น ค่าประมาณเกี่ยวกับ Bayesian ซึ่งได้มาจากการแก้ปัญหาของสมการความน่าจะเป็นโดยใช้กระบวนการ Newton-Raphson

$$\frac{\partial \ln P(\theta/\mathbf{x})}{\partial\theta} = \frac{\partial \ln L(\mathbf{x}/\theta)}{\partial\theta} + \frac{\partial \ln P(\theta)}{\partial\theta} = 0$$

เนื่องจากขอบเขตการประมาณค่าแบบเบเซียน ถูกพิจารณาว่าเป็นส่วนขยายของวิธี MML กระบวนการประมาณค่าจึงใช้ประโยชน์ของขั้นตอนวิธี EM ตามที่อธิบายไว้แล้ว สองขั้นตอนถูกกระทำซ้ำจนกระทั่งเป็นไปตามเกณฑ์การบรรจบกัน เกณฑ์การบรรจบกันกำหนดตามจำนวนของวงจร EM วิธีการประมาณค่าด้วย Bayesian ซึ่สามารถหาได้แม้ว่าผู้สอบจะได้คะแนนเต็ม หรือ ศูนย์คะแนน

2.2.2 **ค่าประมาณที่คาดหวังหลังการประมาณค่า** (Expected a posteriori: EAP)

EAP เป็นค่าเฉลี่ยของการแจกแจงความสามารถหลังการประมาณค่าของผู้สอบ เช่น ค่าที่คาดหวังของ θ ในการแจกแจงความสามารถหลังการประมาณค่า ที่กำหนดในรูป

$$E(\theta/\mathbf{x}) = \frac{\int_{-\infty}^{\infty} L(\mathbf{x}/\theta)P(\theta)\theta d\theta}{\int_{-\infty}^{\infty} L(\mathbf{x}/\theta)P(\theta)\theta d\theta}$$

สมการข้างต้นเป็นการหาปฏิยานุพันธ์ ในทางปฏิบัติ ปฏิยานุพันธ์สามารถคำนวณโดยใช้ Gauss-Hermite quadrature. ซึ่งคำนวณจาก

$$E(\theta/\mathbf{x}) = \frac{\sum_{k=1}^{q} L(\mathbf{x}/X_k)A(X_k)X_k}{\sum_{k=1}^{q} L(\mathbf{x}/X_k)A(X_k)}$$

สมการข้างต้นไม่มีการทำซ้ำและได้ค่าประมาณ EAP โดยตรงดังนั้น บอคและ อิทคิน (Bock; & Aitkin.1981) ได้แนะนำ EAP เป็นส่วนหนึ่งของกระบวนการประมาณค่า MML

ในการศึกษาครั้งนี้ เป็นการศึกษาผลต่างของค่าประมาณความสามารถแฝงที่ได้จากการทดสอบซ้ำสองครั้งด้วยแบบทดสอบผสมซึ่งได้มาจากโมเดลการตอบข้อสอบที่แตกต่างกัน ดังนั้นจึงจำเป็นต้องใช้วิธีการประมาณค่าแบบเดียวกันเพื่อให้ได้ค่าประมาณความสามารถที่ดีที่สุดมาเปรียบเทียบกันภายใต้เงื่อนไขเดียวกัน จากการศึกษาพบว่าวิธีประมาณค่าพารามิเตอร์ MML เป็นวิธีที่ได้รับการแนะนำมากที่สุดในกลุ่มความน่าจะเป็นมากที่สุด ดังนั้นในการศึกษาครั้งนี้จึงเลือกโปรแกรมคอมพิวเตอร์ MULTILOG (Thissen.1991) ซึ่งเป็นโปรแกรมที่รองรับการประมาณค่าแบบ MML ที่มีการรวมวิธีการค่าประมาณแบบ Bayesianทั้ง MAP และ EAP เพื่อใช้สำหรับประมาณค่าพารามิเตอร์ในทุกโมเดลซึ่ง MULTILOG สามารถวิเคราะห์ด้วยโมเดลที่ผสมโมเดลการตอบข้อสอบแบบให้คะแนนแบบสองค่าและโมเดลการตอบข้อสอบให้คะแนนหลายค่าในการประมาณค่าร่วมกันโดยโปรแกรม MULTILOG ถือว่าโมเดลการตอบข้อสอบให้คะแนนแบบสองค่าเป็นกรณีย่อยของโมเดลการตอบข้อสอบให้คะแนนแบบหลายค่า

วิธีการปรับเทียบค่าพารามิเตอร์ตามทฤษฎีการตอบข้อสอบ

ในการศึกษาครั้งนี้เป็นการตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันในการสร้างมาตรวัดร่วมเมื่อผู้สอบพัฒนาการการเรียนรู้แตกต่างกันในเงื่อนไขต่างๆ ผู้วิจัยจึงได้สรุปลักษณะธรรมชาติของมาตรวัดตามทฤษฎีการตอบข้อสอบ และ วิธีการสร้างมาตรวัดร่วม เมื่อผู้สอบมีความสามารถไม่เท่าเทียมกัน โดยเริ่มจาก ความคลุมเครือของมาตรวัดในทฤษฎีการตอบข้อสอบ วิธีการประมาณค่าสำหรับปรับเทียบค่าพารามิเตอร์ ดังรายละเอียดต่อไปนี้

1. **ความคลุมเครือของมาตรวัดในทฤษฎีการตอบข้อสอบ** (Scale indeterminacy in IRT)

ความคลุมเครือของมาตรวัดในทฤษฎีการตอบข้อสอบ คือ ความอิสระในการกำหนดจุดกำเนิด (Origin) และ หน่วย (Unit) ของมาตรวัดความสามารถได้ตามต้องการ ตัวอย่างเช่น กำหนดให้มาตรวัด K และ L เป็นมาตรวัดสำหรับโมเดลโลจิสติกสามพารามิเตอร์ (3PL) ที่แตกต่างกันในรูปสมการเชิงเส้น โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (3PL) คือ ความน่าจะเป็นของการตอบข้อสอบข้อ j ถูกของผู้สอบที่มีความสามารถ θ_i เขียนเป็นสมการได้ว่า

$$p_{ij} = p_{ij}(\theta_i)/a_j,b_j,c_j) = c_j + (1-c_j)\frac{1}{1+\exp\left[-1.7a_j(\theta_i - b_j)\right]}$$

เมื่อ

a_j แทน พารามิเตอร์อำนาจจำแนก

b_j แทน พารามิเตอร์ความยาก

c_j แทน พารามิเตอร์การเดา

ให้ θ เป็นตัวแปรแฝง ซึ่งหมายถึง สมรรถภาพหรือความสามารถบนมาตรวัด K สามารถแปลงให้อยู่ในรูปเชิงเส้นบนมาตรวัด L ได้ดังสมการข้างล่าง

$$\theta_{Li} = A\theta_{Ki} + B$$

เมื่อ A และ B เป็นค่าคงที่ ในสมการการแปลงเชิงเส้น (หรือ ค่าความชันและจุดตัดบนแกนของเส้นตรง) โดยที่ θ_{Li} และ θ_{Ki} แทนค่าความสามารถของผู้สอบ คนที่ i บนมาตรวัด L และ K ตามลำดับ

และให้ ค่าพารามิเตอร์ข้อสอบบนมาตรวัด K ถูกแปลงให้อยู่บนมาตรวัด L ดังสมการ

$$a_{Lj} = \frac{a_{Kj}}{A}$$

$$b_{Li} = Ab_{Ki} + B$$

$$C_{Li} = C_{Ki}$$

เมื่อ a_{Li}, b_{Li}, c_{Li} แทน ค่าพารามิเตอร์ของข้อสอบ j บนมาตรวัด L และ a_{Ki}, b_{Ki}, c_{Ki} แทนค่าพารามิเตอร์ของข้อสอบ j บนมาตรวัด K โดยที่ค่าพารามิเตอร์ c เป็นอิสระจากการแปลง

การแทนที่ θ_{Li} ด้วย θ_{Ki} การแทนที่ a_{Li} ด้วย a_{Ki} การแทนที่ b_{Li}ด้วย b_{Ki} และการแทนที่c_{Li} ด้วย c_{Ki} จะไม่เปลี่ยนโอกาสในการตอบข้อสอบ j ถูกนั่นคือ

$$p_{ij}(\theta_{Li})/a_{Lj},b_{Lj},c_{Lj}) = p_{ij}(A\theta_{Ki}+B)/\frac{a_{Kj}}{A}, Ab_{Kj}+B, c_{Kj}) = p_{ij}(\theta_{Ki})/a_{Kj},b_{Kj},c_{Kj})$$

ดังนั้นถ้าโมเดลการตอบข้อสอบมีความสอดคล้องกลมกลืนกับข้อมูล เมื่อแปลงมาตรวัดความสามารถ θ ให้อยู่ในรูปเชิงเส้น โมเดลการตอบข้อสอบนั้นย่อมมีความสอดคล้องกลมกลืนกับข้อมูล พร้อมกับได้ค่าพารามิเตอร์ข้อสอบที่ถูกแปลงค่าด้วย (Kolen; & Brennan. 2004)

ในทางปฏิบัติ ความคลุมเครือของมาตรวัดถูกขจัดได้โดยการกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความสามารถ θ ให้มีค่าเป็น 0 และ 1 ตามลำดับ การกำหนดความสามารถเช่นนี้นำไปสู่การกำหนดของพารามิเตอร์ข้อสอบด้วย มาตรวัดที่ได้มักถูกเรียกว่า มาตรวัด “0, 1” (Mislevy & Bock, 1990) โปรแกรมคอมพิวเตอร์ที่สามารถใช้ได้ ได้แก่ BILOG-MG (Zimowski, Muraki, Mislevey, & Bock, 1996), MULTILOG (Thissen, 1991) และ PARSCALE (Muraki & Bock, 1997)

2. วิธีการประมาณค่าพารามิเตอร์สำหรับสร้างมาตรวัดร่วม (Calibration methods)

รูปแบบการเก็บรวบข้อมูลด้วยแบบทดสอบต่างฉบับที่มีข้อสอบร่วมกับกลุ่มความสามารถไม่เท่าเทียมกัน (Common-item nonequivalent groups design) แบบทดสอบต่างฉบับที่ใช้ต้องมีข้อสอบร่วมกันจำนวนหนึ่งเพื่อนำไปทดสอบกับกลุ่มตัวอย่างที่มีจากกลุ่มประชากรที่มีการแจกแจงความสามารถแตกต่างกัน เนื่องจากผู้สอบแต่ละกลุ่มไม่ได้ถูกพิจารณาว่ามีความสามารถเท่าเทียมกัน วิธีการแทนที่ค่าประมาณพารามิเตอร์จากผลการทดสอบที่ได้ในแต่ละกลุ่มลงบนมาตรวัดเดียวกันสามารถทำได้โดยประมาณค่าพารามิเตอร์ของแต่ละกลุ่มไปพร้อมกันในครั้งเดียว หรือแยกกันประมาณค่าครั้งละหนึ่งกลุ่มแล้วจึงแปลงค่าพารามิเตอร์ของข้อสอบโดยใช้ค่าพารามิเตอร์ของข้อสอบร่วม เนื่องจากมีวิธีการประมาณค่าพารามิเตอร์เพื่อสร้างมาตรวัดร่วมตามทฤษฎีการตอบข้อสอบหลายวิธีจึงมีการศึกษาวิจัยเพื่อเปรียบเทียบคุณภาพของวิธีการประมาณพารามิเตอร์สำหรับการปรับเทียบพารามิเตอร์หลายงานวิจัย วิธีการประมาณค่าพารามิเตอร์สำหรับสร้างมาตรวัดร่วมมีรายละเอียดดังนี้

2.1 การประมาณพารามิเตอร์แบบแยกกัน (Separate calibration)

การประมาณพารามิเตอร์แบบแยกกัน (Separate calibration) เป็นวิธีการประมาณค่าที่ได้จากการใช้โปรแกรมคอมพิวเตอร์ประมาณค่าพารามิเตอร์ของแบบทดสอบครั้งละหนึ่งฉบับจากผู้สอบกลุ่มเดียว มาตรวัด (Scale) สำหรับใช้ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละฉบับมาจาก

กลุ่มผู้สอบต่างๆที่มีความสามารถไม่เท่าเทียมกันต่างประมาณค่าด้วยมาตรวัด “0,1” แต่เมื่อผู้สอบมีความสามารถไม่เท่าเทียมกันแม้จะมีการใช้มาตรวัด “0,1” แต่ผลที่ได้ย่อมเป็นอิสระจากกัน ทั้งนี้เนื่องจากมาตรวัดในโมเดลการตอบข้อสอบมีคลุมเครือ (Scale indeterminacy) หรืออาจจะกล่าวได้ว่า ค่าพารามิเตอร์แต่ละกลุ่มไม่ได้อยู่บนมาตรวัดเดียวกัน นั่นเองอย่างไรก็ตามแม้ว่ามาตรวัด “0,1” ที่ใช้ประมาณค่าพารามิเตอร์จากกลุ่มผู้สอบที่มีความสามารถไม่เท่าเทียมกันแต่มาตรวัดเหล่านี้ต่างมีความสัมพันธ์เชิงเส้น (Kim. 2007) ดังนั้นในการสร้างมาตรวัดร่วมที่จะทำให้ค่าพารามิเตอร์ที่ได้จากการใช้มาตรวัด “0,1” ประมาณค่าพารามิเตอร์แยกตามกลุ่มผู้สอบสามารถแปลงค่าพารามิเตอร์ให้อยู่บนมาตรวัด “0,1” ของกลุ่มแรกหรือกลุ่มฐานได้ จึงต้องการสมการการแปลงเชิงเส้น (Linear transformation)

ในการประมาณค่าพารามิเตอร์แบบแยกกันจะได้ค่าพารามิเตอร์ของข้อสอบร่วมระหว่างกลุ่มสองชุด ซึ่งค่าพารามิเตอร์เหล่านี้ถูกนำไปคำนวณหาพารามิเตอร์การแปลง A และ B เพื่อแปลงค่าพารามิเตอร์ของข้อสอบที่ไม่เป็นข้อสอบร่วมให้สามารถแทนค่าลงบนมาตรวัดที่ต้องการเปรียบเทียบได้ ดังนั้นค่าสัมประสิทธิ์ A และ B จึงมีความสำคัญต่อการแปลงค่าประมาณพารามิเตอร์ของกลุ่มเป้าหมายไปยังมาตรวัดของกลุ่มฐานหรือกลุ่มเปรียบเทียบ

วิธีการแปลงมาตรวัด (Scale transformation)

การแปลงมาตรวัดสำหรับโมเดลการตอบข้อสอบแบบให้คะแนนสองค่าสามารถทำได้ 4 วิธี ได้แก่ วิธีโมเมนต์ (Moments methods) วิธีโค้งคุณลักษณะ (Characteristic curve method) วิธีค่าไคสแคว์ต่ำสุด (Minimum chi-square methods) และวิธีกำลังสองน้อยสุด (Least squares method) (Kim. 2004)

2.1.1 **การแปลงมาตรวัดด้วยวิธีโมเมนต์** ประกอบด้วย วิธีค่าเฉลี่ยและค่าเฉลี่ย (Mean/mean method) (Loyd & Hoover, 1980) และวิธีค่าเฉลี่ยและผลรวม (Mean/sigma method) (Marco, 1977) การแปลงมาตรวัดด้วยวิธีโค้งลักษณะประกอบด้วย วิธีโค้งลักษณะของแฮบาระ (Haebara’s characteristic curve method) (Haebara, 1980) และวิธีโค้งลักษณะของสต็อกกิ้งและลอร์ด (Stocking and Lord’s test characteristic curve method) (Stocking & Lord, 1983) ส่วนการแปลงมาตรวัดด้วยวิธี ไคว์สแควต่ำน้อยสุดนำเสนอโดย ดิฟจี้ (Divgi.1985) วิธีกำลังสองน้อยสุดนำเสนอโดย โอกาซาวาระ (Ogasawara .2001) วิธีการแปลงมาตรวัดที่นิยมใช้กันโดยทั่วไปคือ วิธีโมเมนต์ และวิธีโค้งคุณลักษณะ ดังมีสาระสำคัญดังนี้

2.1.1.1 **วิธีค่าเฉลี่ยและผลรวม (The mean/sigma method)**

วิธีค่าเฉลี่ยและผลรวม ใช้ค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานจากข้อสอบร่วมที่อยู่ในแบบทดสอบ K และ L คำนวณหาค่า A และ B (Marco, 1977) ดังนี้

$$A = \frac{\sigma(b_L)}{\sigma(b_K)}, \qquad B = \mu(b_L) - A\mu(b_K)$$

เมื่อ

$\sigma(b_L)$ เป็น ค่าส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก b ของข้อสอบร่วมจากแบบทดสอบฉบับ L

$\sigma(b_K)$ เป็น ค่าส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก b ของข้อสอบร่วมจากแบบทดสอบฉบับ K

$\mu(b_L)$ เป็น ค่าเฉลี่ยของพารามิเตอร์ความยาก b ของข้อสอบร่วมจากแบบทดสอบฉบับ L

$\mu(b_K)$ เป็น ค่าเฉลี่ยของพารามิเตอร์ความยาก b ของข้อสอบร่วมจากแบบทดสอบฉบับ K

2.1.1.2 **วิธีค่าเฉลี่ยและค่าเฉลี่ย** (The mean/mean method)

วิธีค่าเฉลี่ยและค่าเฉลี่ยใช้ ค่าเฉลี่ยของค่าประมาณพารามิเตอร์อำนาจจำแนก a และ พารามิเตอร์ความยาก b จากข้อสอบร่วมคำนวณหาค่า A และ B (Loyd & Hoover, 1980)

$$A = \frac{\mu(a_K)}{\mu(a_L)}, \qquad B = \mu(b_L) - A\mu(b_K)$$

เมื่อ

$\mu(a_K)$ เป็น ค่าเฉลี่ยของพารามิเตอร์อำนาจจำแนก a ของข้อสอบร่วมจากแบบทดสอบฉบับ K

$\mu(a_L)$ เป็น ค่าเฉลี่ยของพารามิเตอร์อำนาจจำแนก a ของข้อสอบร่วมจากแบบทดสอบฉบับและ L

$\mu(b_L)$ เป็น ค่าเฉลี่ยของพารามิเตอร์ความยาก b ของข้อสอบร่วมจากแบบทดสอบฉบับ L

$\mu(b_K)$ เป็นค่าเฉลี่ยของพารามิเตอร์ความยาก b ของข้อสอบร่วมจากแบบทดสอบฉบับ K

ปัญหาที่สำคัญของวิธีโมเมนต์เกิดขึ้นเมื่อโค้งคุณลักษณะของข้อสอบ(Item characteristic curves: ICC) ทั้งหมด เกิดจากการรวมของค่าพารามิเตอร์ a, b, และ c ที่ประมาณค่ามาจากช่วงระดับความสามารถที่ต่างกัน ข้อสอบที่มีระดับความยากที่แตกต่างกันมาก (Kolen & Brennan, 2004) แฮบาระ (Haebara ,1980) และ สต็อกกิ้งและลอร์ด (Stocking; & Lord:1983) จึงได้พัฒนาวิธีที่พิจารณาพารามิเตอร์ของข้อสอบไปพร้อมกัน วิธีนี้จึงเรียกว่า วิธีโค้งคุณลักษณะ (The Characteristic Curve methods)

2.1.2 วิธีโค้งคุณลักษณะ (The Characteristic Curve methods)

2.1.2.1 วิธีโค้งคุณลักษณะของแฮบารา (Haebara: 1980)

วิธีโค้งคุณลักษณะของแฮบาระใช้ผลรวมของกำลังสองของผลต่างระหว่างโค้งคุณลักษณะข้อสอบ(ICC) แต่ละข้อสำหรับผู้สอบที่มีความสามารถตามที่กำหนด) ของผู้สอบ

$$Hdiff(\theta_i) = \sum_{j:V}\left[p_{ij}(\theta_{Li} / \hat{a}_{Lj}, \hat{b}_{Lj}, \hat{c}_{Lj}) - p_{ij}(\theta_{Li} / \frac{\hat{a}_{Kj}}{A}), A\hat{b}_{Kj} + B, \hat{c}_{Kj} \right]^2$$

สมการข้างต้นแทน ผลรวมกำลังสองผลต่างระหว่างโค้งคุณลักษณะของข้อสอบแต่ละข้อทุกข้อที่เป็นข้อสอบร่วม ดังนั้น $Hdiff$ จึงคำนวณจากผู้สอบแต่ละคนเพื่อหาค่าคงที่สำหรับการแปลง A และ B ที่น้อยที่สุดจากเกณฑ์ $HCrit = \sum_i Hdiff(\theta_i)$

2.1.2.2 วิธีโค้งคุณลักษณะของสต็อกกิ้งและลอร์ด (Stocking; &Lord. 1983)

วิธีของสต็อกกิ้งและลอร์ด ตรงข้ามกับวิธีของแฮบาระ โดยใช้ค่ากำลังสองของผลต่างระหว่างโค้งลักษณะของแบบทดสอบ (Test Characteristic Curves: TCC) ที่มีคุณลักษณะ θ_i และ $\sum_{j:V} p_{ij}(\theta_i)$ แทน ผลรวมของโค้งคุณลักษณะข้อสอบที่เป็นข้อสอบร่วมทุกข้อ (j: V) สมการผลต่างกำลังสองคือ

$$SLdiff(\theta_i) = \left[\sum_{j:V} p_{ij}(\theta_{Li} / \hat{a}_{Lj}, \hat{b}_{Lj}, \hat{c}_{Lj}) - \sum_{j:V} p_{ij}(\theta_{Li} / \frac{\hat{a}_{Kj}}{A}), A\hat{b}_{Kj} + B, \hat{c}_{Kj} \right]^2$$

สมการนี้ แต่ละโค้ง ICC ของข้อสอบร่วมทุกข้อถูกรวมเข้าด้วยกันเพื่อคำนวณโค้งคุณลักษณะแบบทดสอบ (TCC) แล้วยกกำลังของผลต่างระหว่าง TCC ของแบบทดสอบสองฉบับ ดังนั้น $SLdiff$ จึงเป็นการคำนวณจากกลุ่มผู้สอบเพื่อหาค่าคงที่การแปลง A และ B ที่ต่ำที่สุดโดยใช้เกณฑ์

$$SLCrit = \sum_i SLdiff(\theta_i)$$

และจากการศึกษาวิจัยเพื่อเปรียบเทียบวิธีการคำนวณค่าคงที่การแปลงหลังจากการประมาณพารามิเตอร์แยกกันโดยใช้โมเดลการตอบข้อสอบสำหรับข้อมูลที่ให้คะแนนสองค่า พบว่าวิธีของแฮบาระ (Haebara.1980) และ วิธีสต็อกกิ้งและลอร์ด (Stocking; & Lord. 1983) ให้ผลที่คงเส้นคงวามากกว่าวิธีโมเมนต์ (Baker & Al-Karni. 1991; Hanson; & Beguin.2002; Kim; & Cohen, 1992)

2.2 การประมาณค่าพารามิเตอร์โดยการกำหนดพารามิเตอร์ข้อสอบร่วม (Fixed calibration)

วิธีการประมาณค่าพารามิเตอร์ด้วยการกำหนดพารามิเตอร์ข้อสอบร่วมสำหรับหามาตรวัดร่วมนั้นทำได้โดยการกำหนด หรือแปลงค่าพารามิเตอร์ข้อสอบร่วมของแบบทดสอบสองฉบับที่มีระดับใกล้เคียงกันด้วยโมเดลโลจิสติกสามพารามิเตอร์ โดยที่มีวิธีการกำหนดสองวิธี ได้แก่ วิธีการกำหนดให้พารามิเตอร์การเดามีค่าคงที่ (FC) และ วิธีการกำหนดให้พารามิเตอร์ อำนาจจำแนก ความยาก และการเดา มีค่าคงที่ (FABC)

2.2.1 วิธีการกำหนดให้พารามิเตอร์การเดามีค่าคงที่ (FC)

วิธี FC ค่าพารามิเตอร์การเดาจากกลุ่มผู้สอบอ้างอิงจะถูกนำไปใช้เป็นค่าตั้งต้นสำหรับกลุ่มประชากรเป้าหมาย โดยที่ไม่ถูกนำมาประมาณค่าใหม่อีก แต่สำหรับค่าพารามิเตอร์ความยากและพารามิเตอร์อำนาจจำแนก จะเป็นอิสระในการประมาณค่า หลังจากประมาณค่าพารามิเตอร์ข้อสอบแล้วขั้นตอนในการคำนวณหา A และ B และใช้ค่าเหล่านี้ สำหรับสร้างสมการการแปลงเชิงเส้นของข้อสอบที่ไม่ใช่ข้อสอบร่วมเป็นเช่นเดียวกับการประมาณค่าพารามิเตอร์แบบแยกกันอย่างไรก็ตาม ลี และ ฟิตซ์แพทริก (Lee ; &Fitzpatrick .2002) ได้ให้เหตุผลแย้งว่า การกำหนดพารามิเตอร์การเดาของข้อสอบร่วมนั้นสามารถแสดงถึงความจริงที่ว่าค่าพารามิเตอร์การเดาสำหรับข้อสอบร่วมไม่ได้ส่งผลต่อการแปลงเชิงเส้นใดใด และทำให้วิธี FC สามารถเลี่ยงความยากของการได้ค่า c ที่ถูกต้องจากข้อสอบร่วมนั้นขึ้นอยู่กับการลักษณะการนำโมเดลโลจิสติกสามพารามิเตอร์ไปใช้

2.2.2 วิธีการกำหนดให้พารามิเตอร์ อำนาจจำแนก ความยาก และการเดา มีค่าคงที่ (FABC)

วิธีการกำหนดให้พารามิเตอร์ อำนาจจำแนก ความยาก และการเดา มีค่าคงที่ (FABC) เป็นวิธีที่กำหนดค่าพารามิเตอร์ a, b, และ c ของข้อสอบร่วม โดยกำหนดค่าพารามิเตอร์ a, b, และ c ของข้อสอบร่วมการประมาณค่าจากกลุ่มผู้สอบอ้างอิงแล้วจึงประมาณค่าข้อสอบอื่นที่ไม่ใช่ข้อสอบร่วมด้วยมาตรวัดเดียวกับข้อสอบร่วมที่ข้อสอบเหล่านี้ตั้งอยู่ ซึ่งอยู่บนมาตรวัดกลุ่มอ้างอิงนั่นเอง

วิธี FABC มักใช้ในบริบทการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized adaptive testing) (Ban, Hanson, Wang, & Harris, 2001; Wainer & Mislevy, 1990) และในการสร้างธนาคารแบบทดสอบ (Li et al., 1997) ลี และ คณะ (Li et al. 1997) กล่าวว่าทั้งวิธี FABC และ วิธีการประมาณค่าพารามิเตอร์พร้อมกัน ต่างประมาณค่าพารามิเตอร์บนฐานการรวมกลุ่มผู้สอบจากการทดสอบหลายฉบับเข้าด้วยกัน ดังนั้นวิธีทั้งสองจึงมีจุดเด่นตรงที่กลุ่มตัวอย่างมีขนาดใหญ่ ลักษณะเช่นนี้จึงทำให้ลดปัญหาการประมาณค่า b และ c ไม่ถูกต้องได้ ซึ่งปัญหานี้มักเกิดขึ้นได้เมื่อมีกลุ่มความสามารถต่ำจำนวนมากจึงไม่สามารถประมาณค่าพารามิเตอร์การเดาของ

ข้อสอบง่ายได้ ดังนั้นวิธี FABC จึงมีคุณสมบัติของวิธีการประมาณพารามิเตอร์พร้อมกันและให้ความคงเส้นคงวามากกว่าการแปลงเชิงเส้นเนื่องจากขจัดความคลาดเคลื่อนของการปรับเทียบและนำพารามิเตอร์การเดามาคำนวณสำหรับการแปลงมาตรวัดด้วย ต่อมา ปี 2006 คิม (Kim. 2006) ได้พัฒนาวิธี FABC ตามจำนวนของรอบการประมาณค่าใหม่ (Number of prior updates) ได้แก่ วิธีไม่ปรับให้เป็นข้อมูลใหม่ก่อน (the no prior update: NPU) ซึ่งเป็นการสมมติว่าการแจงแจกแความสามารถก่อนการประมาณมีการแจกแจงปกติ N(0,1) และไม่ปรับให้เป็นข้อมูลใหม่ระหว่างวงจรการประมาณค่าแบบค่าคาดหวังสูงสุด Expectation-Maximization (EM) cycles

สำหรับ FABC วิธีต่อมาเป็นวิธีการปรับให้ข้อมูลใหม่ก่อนการประมาณค่าอย่างง่าย (the simple prior Update: SPU) เป็นวิธีที่ในขั้นต้นทำการปรับข้อมูลของความสามารถเริ่มต้นก่อนการประมาณค่าตามที่กำหนดแล้วนำผลการแจกแจงภายหลังค่าที่ประมาณได้เป็นข้อมูลใหม่สำหรับการคำนวณรอบที่สองของ FABC ในการปรับข้อมูลซ้ำใหม่ การแจกแจงความสามารถก่อนการประมาณค่าถูกทำซ้ำใหม่หลายรอบจนกระทั่งสำเร็จตามวิธี FABC หรือกล่าวอีกทางหนึ่งว่า การแจกแจงความสามารถภายหลังการประมาณค่าชุดใหม่ในขั้นตอนการคำนวณค่าประมาณพารามิเตอร์ถูกนำไปเป็นความสามารถก่อนการประมาณในรอบถัดไปจนกระทั่งผลต่างระหว่างการแจกแจงก่อนและหลังมีค่าน้อยที่สุด

2.3 การประมาณค่าพารามิเตอร์พร้อมกัน (Concurrent calibration)

การประมาณค่าพารามิเตอร์พร้อมกัน เป็นการประมาณค่าพารามิเตอร์ของข้อสอบทุกข้อทุกฉบับทุกกลุ่มที่สอบนำมาประมาณค่าในครั้งเดียวกัน จึงทำให้มั่นใจได้ว่า ค่าพารามิเตอร์อยู่บนมาตรวัดเดียวกัน นั่นคือ อยู่บนมาตรวัด “0,1” เหมือนกัน หรือ ใช้มาตรวัดร่วมกันนั่นเอง สิ่งสำคัญเมื่อใช้การประมาณค่าพารามิเตอร์พร้อมกันคือ ต้องใช้โปรแกรมคอมพิวเตอร์ที่ยอมให้สามารถประมาณค่ากลุ่มตัวอย่างหลายกลุ่มพร้อมกันได้ (Kolen & Brennan, 2004) โปรแกรมที่สามารถใช้ได้ ได้แก่ BILOG-MG และMULTILOG ทั้ง BILOG-MG และ MULTILOG ใช้วิธีการประมาณค่าแบบความน่าจะเป็นสูงสุดปลายทาง (Marginal Maximum Likelihood: MML) เพื่อการประมาณค่าพารามิเตอร์ด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ วิธีการประมาณค่าพารามิเตอร์พร้อมกัน นำข้อมูลผลการสอบทั้งหมดทุกกลุ่มมาคำนวณ โดยข้อสอบเฉพาะกลุ่มที่นักเรียนไม่ได้ทำถูกกำหนดให้เป็นข้อสอบ “not reached” นั่นคือผลการตอบของผู้สอบมีข้อมูลการตอบข้อสอบสำหรับข้อสอบที่ทำและข้อสอบที่ไม่ทำที่ลงรหัสเป็น “not reached” ค่าประมาณทั้งหมดจึงตั้งอยู่บนมาตรวัดเดียวกันโดยใช้การคำนวณเพียงครั้งเดียว การศึกษาจากการจำลองข้อมูลเพื่อเปรียบเทียบผลการวัดตามแนวนอนระหว่างวิธีการปรับพารามิเตอร์พร้อมกันและแยกกันพบว่า วิธีประมาณพารามิเตอร์พร้อมกันมีความถูกต้องมากกว่าเมื่อข้อมูลมีความสอดคล้องกลมกลืนกับโมเดลการตอบข้อสอบ (Kim & Cohen, 1998; Hanson &

Béguin, 2002) แต่แข็งแกร่งน้อยกว่าต่อการละเมิดข้อตกลงทฤษฎีการตอบข้อสอบเมื่อข้อสอบมีความเป็นพหุมิติ (Béguin, Hanson, & Glas, 2000; Béguin & Hanson, 2001) ต่อมามีการปรับปรุงวิธีการประมาณค่าพารามิเตอร์พร้อมกันเพื่อขจัดปัญหาเกี่ยวกับความเป็นพหุมิติเช่น วิธีการประมาณพารามิเตอร์พร้อมกันเป็นรายคู่ (Pair-wise concurrent calibration) (Karkee et al., 2003) วิธีประมาณค่าพารามิเตอร์กึ่งพร้อมกัน (Semi-concurrent calibration (Meng et al., 2006) และ วิธีการประมาณค่าพารามิเตอร์พร้อมกันด้วยการแจกแจงเดียวกันทุกระดับชั้น (Concurrent calibration with a single distribution across all grades) (Karkee et al., 2006)

แม้ว่าคุณภาพของวิธีการสร้างมาตรวัดร่วมดังวิธีที่กล่าวข้างต้นมีความแม่นยำและเงื่อนไขการใช้แตกต่างกันเมื่อเงื่อนไขแตกต่างกัน แต่วิธีการประมาณค่าพารามิเตอร์พร้อมกันเป็นการประมาณค่าพารามิเตอร์ของข้อสอบทุกข้อทุกฉบับทุกกลุ่มที่สอบนำมาประมาณค่าในครั้งเดียวกัน จึงทำให้มั่นใจได้ว่า ค่าพารามิเตอร์อยู่บนมาตรวัดเดียวกัน ผู้วิจัยจึงสนใจนำวิธีการประมาณค่าพารามิเตอร์พร้อมกันมาใช้ร่วมกับการคำนวณพัฒนาการความสามารถ

งานวิจัยที่เกี่ยวข้อง

1. งานวิจัยเกี่ยวกับการวัดการเปลี่ยนแปลง

งานวิจัยในประเทศเกี่ยวกับการวัดการเปลี่ยนแปลง

วินิจ เทือกทอง (2537) เปรียบเทียบประสิทธิภาพของการคำนวณคะแนนพัฒนาการ 5 วิธี คือ วิธีดั้งเดิม วิธีเศษเหลือ วิธีของลอร์ด วิธีของศิริชัย และวิธีเทียบส่วนร้อย ตามลักษณะการแจกแจงของคะแนนการวัดก่อนเรียนและคะแนนการวัดหลังเรียน ซึ่งมีรูปแบบการแจกแจงที่แตกต่างกัน 5 เงื่อนไขได้แก่ ค่าคะแนนพัฒนาการ ค่าสหสัมพันธ์ระหว่างคะแนนการวัดก่อนเรียนกับคะแนนพัฒนาการ ขนาดกลุ่มตัวอย่าง ความเบ้ของข้อมูล และความโด่งของข้อมูล โดยใช้การจำลองสถานการณ์ด้วยเทคนิคมอนติคาร์โล คำนวณค่าคะแนนพัฒนาการของแต่ละวิธีการตามแต่ละสถานการณ์จำนวน 1000 ครั้ง พิจารณาค่าความคลาดเคลื่อนที่น้อยที่สุดเป็นเกณฑ์ประสิทธิภาพ ผลการวิจัยพบว่าเมื่อไม่ได้คำนึงถึงเพดานของคะแนนการสอบวัดแล้ว วิธีการคำนวณค่าคะแนนพัฒนาการด้วยวิธีดั้งเดิม หรือวิธีของลอร์ด มีค่าความคลาดเคลื่อนน้อยที่สุด แต่เมื่อมีการปรับแก้อิทธิพลคะแนนสูงสุดของคะแนนการสอบวัด วิธีการคำนวณค่าคะแนนเพิ่มด้วยวิธีเศษเหลือเทียบกับศักยภาพของผู้เรียนมีค่าความคลาดเคลื่อนน้อยที่สุด

ประสิทธิ์ ไชยกาล (2539) เปรียบเทียบประสิทธิภาพของโมเดลลิสเรล 3 แบบ ที่ใช้ในการศึกษาตัวแปรที่สัมพันธ์กับการเปลี่ยนแปลงในระยะยาวของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ได้แก่ โมเดลลิสเรลที่มีการวัดการเปลี่ยนแปลงในรูปโมเดลของการวิเคราะห์องค์ประกอบระยะยาวที่วัด

ด้วยตัวบ่งชี้ตัวเดียว และโมเดลลิสเรลที่มีการวัดการเปลี่ยนแปลงการเรียนรู้ในรูปโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้หลายตัว กลุ่มตัวอย่างที่ใช้ในการวิจัย คือนักเรียนชั้นประถมศึกษาปีที่ 6 ของโรงเรียนในสังกัดกรุงเทพมหานคร จำนวน 606 คน ดำเนินการเก็บข้อมูลเป็นลักษณะการเก็บข้อมูลระยะยาว โดยวัดตัวแปรด้านผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ และเจตคติต่อพฤติกรรมการเรียนคณิตศาสตร์ 3 ครั้ง และวัดตัวแปรด้านความถนัดทางคณิตศาสตร์ 1 ครั้ง เมื่อพิจารณาจากความคลาดเคลื่อนต่ำที่สุดและความสัมพันธ์ของตัวแปรที่สัมพันธ์กับการเปลี่ยนแปลงมีความไม่แปรเปลี่ยนเป็นเกณฑ์ ผลการวิจัยสรุปได้ว่า โมเดลลิสเรลที่มีการวัดการเปลี่ยนแปลงในโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้หลายตัว มีประสิทธิภาพสูงกว่าโมเดลลิสเรลที่มีการวัดการเปลี่ยนแปลงในรูปโมเดลพื้นฐานการวิเคราะห์องค์ประกอบระยะยาวที่วัดตัวบ่งชี้เดียว สำหรับโมเดลการวัดการเปลี่ยนแปลงระยะยาว พบว่า โมเดลการวัดการเปลี่ยนแปลงในรูปโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้หลายตัว มีประสิทธิภาพสูงที่สุด เพราะโมเดลสามารถประมาณค่าพารามิเตอร์ที่บ่งชี้การเปลี่ยนแปลงที่เกิดขึ้นตลอดช่วงเวลาได้ และโมเดลมีความคลาดเคลื่อนต่ำที่สุดด้วย รองลงไปคือ โมเดลการวัดการเปลี่ยนแปลงในรูปโมเดลพื้นฐานการวิเคราะห์องค์ประกอบระยะยาวที่วัดตามลำดับ

สมถวิล วิจิตรวรรณา (2543) เปรียบเทียบประสิทธิภาพของโมเดลการวัดการเปลี่ยนแปลงระยะยาว ชนิดตัวแปรเดี่ยวและชนิดตัวแปรพหุ ระหว่างโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง โมเดลพหุระดับ และโมเดลกึ่งซิมเพลกซ์ที่มีตัวแปรแฝงพัฒนาการ ด้วยดัชนีความสอดคล้องของโมเดลกับข้อมูล คือ สถิติไค-สแควร์ ดัชนี GFI และ CFI ดัชนีรากกำลังสองเฉลี่ยของเศษเหลือ (RMR) และดัชนีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ (RMSEA) การวิจัยครั้งนี้ได้นำเสนอโมเดลการวัดการเปลี่ยนแปลงโมเดลใหม่ คือ โมเดลกึ่งซิมเพลกซ์ที่มีตัวแปรแฝงพัฒนาการ ที่จะสามารถอธิบายคะแนนเริ่มต้นและอัตราพัฒนาการที่แท้จริง ข้อมูลการวิจัยครั้งนี้ได้จากประชากรนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 469 คน ในโรงเรียนขยายโอกาสทางการศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดสมุทรสงคราม ภาคเรียนที่ 1 ปีการศึกษา 2542 โดยวัดนักเรียนคนเดิม 5 ครั้งในช่วงเวลาต่างกัน เครื่องมือที่สร้างขึ้นในการวิจัยครั้งนี้คือ แบบสอบคู่ขนานวิชาคณิตศาสตร์ 5 ฉบับที่วัด 2 คุณลักษณะ คือ การคิดคำนวณและการแก้โจทย์ปัญหา กระบวนการพัฒนาแบบสอบที่สำคัญ คือ การสร้างลักษณะเฉพาะของข้อสอบและการปรับเทียบคะแนน ระหว่าง บางฉบับที่ไม่เป็นคู่ขนาน การวิเคราะห์ข้อมูลโมเดลพหุระดับใช้โปรแกรม HLM ส่วนโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง และโมเดลกึ่งซิมเพลกซ์ที่มีตัวแปรแฝงพัฒนาการ ใช้โปรแกรมการวิเคราะห์ความแปรปรวนร่วม EQS ผลการศึกษาครั้งนี้ พบว่า โมเดลโค้งพัฒนาการที่มีตัวแปรแฝงเป็นโมเดล ที่ใช้อธิบายการวัดการเปลี่ยนแปลงระยะ

ยาวที่มีประสิทธิภาพสูงสุด ทั้งชนิดตัวแปรเดี่ยวและตัวแปรพหุ รองลงมาคือ โมเดลกึ่งซิมเพลกซ์ที่มีตัวแปรแฝงพัฒนาการและโมเดลพหุระดับ

อวยพร เรืองตระกูล (2544) ได้พัฒนาและวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ 9 วิธี ประกอบด้วย วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธีได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ 2) วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน 3) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ 4) วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ และ 5) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพ วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบข้อสอบ 2 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง และ 2) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติ สำหรับการเรียนรู้และการเปลี่ยนแปลง และวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา 2 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์ เมื่อเทียบกับศักยภาพการพัฒนา และ 2) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์ เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 2 กลุ่มๆ ละ 698 คน และ 637 คน จากโรงเรียนสังกัดกรมสามัญศึกษา เขตกรุงเทพมหานคร 12 โรงเรียน เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับที่คู่ขนานกัน ข้อมูลที่ใช้ในการวิจัยมี 3 ชุด สองชุดแรกเป็นข้อมูลปฐมภูมิที่ได้จากการเก็บรวบรวมข้อมูลด้วยแบบทดสอบคณิตศาสตร์ที่วัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบฉบับเดิม และแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน ส่วนข้อมูลชุดที่สาม เป็นข้อมูลทุติยภูมิ ที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 ที่มีการวัดซ้ำ 5 ครั้ง มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) การวิเคราะห์ข้อมูลเป็นการประมาณค่าคะแนนพัฒนาการ และการวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ ในด้านความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อน ด้วยการวิเคราะห์โมเดลเชิงเส้นตรงระดับลดหลั่น และสถิติทดสอบ Hotelling

ผลการวิจัยสรุปได้ดังนี้ 1) เมื่อเปรียบเทียบผลการประมาณค่าคะแนนพัฒนาการระหว่าง 3 กลุ่มวิธี พบว่า กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีตอบสนองข้อสอบ ที่ผู้วิจัยพัฒนามีคุณภาพไม่แตกต่างกัน 2) เมื่อเปรียบเทียบระหว่างวิธีภายใน กลุ่มทฤษฎีการทดสอบแบบดั้งเดิม พบว่า วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ และวิธีการวัดคะแนนพัฒนาการสัมพัทธ์ที่คุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูลสองชุดแรก และวิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูลชุดที่สาม 3) เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการ

ตอบสนองข้อสอบ พบว่า วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลงมีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง 4) เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการตอบข้อสอบที่ผู้วิจัยพัฒนา พบว่า วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียนมีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค

ศศิวิมล อมตชีวิน (2546) ศึกษาผลการวิเคราะห์เหลื่อมลำดับในโมเดลโค้งพัฒนาการที่มีช่วงเวลาการวัดแตกต่างกัน โดยทำการเปรียบเทียบระหว่างโมเดลที่มีช่วงเวลาการวัด 3 - 9 ช่วงเวลากับโมเดลที่มีช่วงเวลาการวัดครบสมบูรณ์ซึ่งมีช่วงเวลาการวัด 10 ช่วงเวลา และเพื่อเปรียบเทียบประสิทธิภาพระหว่างโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง 2 รูปแบบ คือ โมเดลพัฒนาการเชิงเส้นตรงและโมเดลพัฒนาการพารามิเตอร์อิสระ ในการศึกษาการเปลี่ยนแปลงระยะยาวของพัฒนาการทางกายภาพด้านน้ำหนัก และส่วนสูงของนักเรียนในระดับประถมศึกษา กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักเรียนระดับชั้นประถมศึกษาปีที่ 1 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษากรุงเทพมหานคร จำนวน 840 คน ผู้วิจัยวิเคราะห์ข้อมูลโดยใช้การวิเคราะห์โมเดลลิสเรลแบบมีตัวแปรแฝง ผลการวิจัย สรุปได้ว่า 1) โมเดลที่มีช่วงเวลาการวัด 5 ช่วงเวลามีความสอดคล้องกับโมเดลที่มีช่วงเวลาการวัด 10 ช่วงเวลามากที่สุด สำหรับการวิเคราะห์เหลื่อมลำดับด้วยรูปแบบโมเดลพัฒนาการเชิงเส้นตรง 2) โมเดลที่มีช่วงเวลาการวัด 5 และ 6 ช่วงเวลามีความสอดคล้องกับโมเดลที่มีช่วงเวลาการวัด 10 ช่วงเวลามากที่สุด สำหรับการวิเคราะห์เหลื่อมลำดับด้วยรูปแบบโมเดลพัฒนาการพารามิเตอร์อิสระ 3) โมเดลพัฒนาการพารามิเตอร์อิสระมีประสิทธิภาพในการศึกษาพัฒนาการทางกายภาพด้านน้ำหนักและส่วนสูงของนักเรียนระดับประถมศึกษามากกว่าโมเดลพัฒนาการเชิงเส้นตรง

รณชิต พฤษกรรม (2547) ศึกษาอิทธิพลของปัจจัยที่มีต่อพัฒนาการของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ที่ระดับพัฒนาการต่างกัน ปัจจัยที่ศึกษา ได้แก่ ตัวแปรเจตคติต่อวิชาคณิตศาสตร์ ตัวแปรแรงจูงใจใฝ่สัมฤทธิ์ ตัวแปรความรู้พื้นฐานเดิม และตัวแปรเพศของนักเรียน เมื่อกำหนดเกณฑ์เป็นคะแนนพัฒนาการที่เพิ่มขึ้น 5% 10% 15% 20% และ 25% ตามลำดับ และเพื่อศึกษาฟังก์ชันการอยู่รอด มัธยฐานระยะเวลาการอยู่รอด และโอกาสที่จะมีพัฒนาการผ่านเกณฑ์สูงสุด ของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนที่รับพัฒนาการต่างกัน เมื่อกำหนดเกณฑ์เป็นคะแนนพัฒนาการที่เพิ่มขึ้น 5% 10% 15% 20% และ 25% ตามลำดับ กลุ่มตัวอย่างที่ใช้คือ นักเรียนชั้นมัธยมศึกษาปีที่ 2 สังกัดกรมสามัญศึกษา เขตกรุงเทพมหานคร จำนวน 452 คน การวิเคราะห์ข้อมูลใช้วิธีการการวิเคราะห์อิทธิพลผสม กลุ่มเวลาการอยู่รอดด้วยโปรแกรม MIXOR และวิเคราะห์ตารางชีพด้วยโปรแกรม SPSS for window version 11.50 ผลการวิจัยสรุปได้ว่า เมื่อกำหนดเกณฑ์เป็นคะแนน

พัฒนาการที่เพิ่มขึ้น 5% 10% 15% 20% และ 25% ตามลำดับ นักเรียนชั้นมัธยมศึกษาปีที่ 2 มีโอกาสที่จะมีพัฒนาการผ่านเกณฑ์สูงสุดในช่วงระยะเวลาที่ 1 1 4 4 4 และ 4 ตามลำดับ โดยมีโอกาสที่จะมีพัฒนาการผ่านเกณฑ์ในแต่ละระดับพัฒนาการเท่ากับ 0.8744 0.6667 0.7333 0.7426 และ 0.7020 ตามลำดับ นักเรียนอยู่รอดได้นานกว่าช่วงระยะเวลาดังกล่าว เท่ากับ 39.16% 50.00% 15.38% 20.49% และ 25.85% ตามลำดับ แต่ละเกณฑ์คะแนนพัฒนาการมีมัธยฐานระยะเวลาการอยู่รอดเท่ากับ 0.82 1.00 2.42 2.87 และ 3.60 ตามลำดับ ตัวแปรทำนายที่มีอิทธิพลต่อโอกาสที่จะมีพัฒนาการผ่านเกณฑ์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 คือ แรงจูงใจใฝ่สัมฤทธิ์มีอิทธิพลต่อโอกาสที่จะมีพัฒนาการผ่านเกณฑ์ทั้ง 5 เกณฑ์ เจตคติต่อวิชาคณิตศาสตร์มีอิทธิพลต่อโอกาสที่จะมีพัฒนาการผ่านเกณฑ์ คะแนนพัฒนาการที่เพิ่มขึ้น 10% 15% และ 20% ตามลำดับ ส่วนความรู้พื้นฐานเดิมมีอิทธิพลต่อโอกาสที่จะมีพัฒนาการผ่านเกณฑ์เพียงเกณฑ์เดียวคือ คะแนนพัฒนาการที่เพิ่มขึ้น 15%

ศุภลักษณ์ ใจแสวงทรัพย์ (2547) ศึกษาความสัมพันธ์เชิงสาเหตุของปัจจัยที่ส่งผลต่อพัฒนาการทางการเรียนคณิตศาสตร์ของนักเรียน ทดสอบความไม่แปรเปลี่ยนของโมเดลสมการโครงสร้างที่สร้างขึ้นระหว่างนักเรียนเพศชายและเพศหญิง และ ทดสอบความไม่แปรเปลี่ยนของโมเดลสมการโครงสร้างที่สร้างขึ้นระหว่างคะแนนพัฒนาการที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และคะแนนพัฒนาการที่มีการให้คะแนนความรู้บางส่วน (Partial credit) ตัวแปรที่ใช้ในการวิจัย ประกอบด้วย ตัวแปรแฝง 7 ตัวแปร และตัวแปรสังเกตได้ 12 ตัวแปร กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัด กรมสามัญศึกษา กรุงเทพมหานคร จำนวน 334 คน และครู จำนวน 11 คน ข้อมูลในส่วนของคะแนนพัฒนาการวิชาคณิตศาสตร์เป็นข้อมูลทุติยภูมิที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ ที่มีการเก็บข้อมูล จากการสอบซ้ำ 3 ครั้ง ห่างกันครั้งละ 4 สัปดาห์ โดยใช้แบบทดสอบฉบับเดียวกัน วิเคราะห์ข้อมูลด้วยสถิติเชิงบรรยาย การวิเคราะห์สหสัมพันธ์แบบเพียร์สัน และการวิเคราะห์ด้วยสถิติขั้นสูงโดยการวิเคราะห์เส้นทางและการวิเคราะห์กลุ่มพหุด้วยโปรแกรมลิสเรล

ผลการวิจัยสรุปได้ดังนี้ 1) โมเดลปัจจัยที่ส่งผลต่อคะแนนพัฒนาการวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 มีความสอดคล้องกับข้อมูลเชิงประจักษ์ มีค่า χ^2 = 2.295, df = 5, p = 0.807, GFI = 0.999, AGFI = 0.982 และ RMR = 0.017 ตัวแปรในโมเดลสามารถอธิบายความแปรปรวนของพัฒนาการวิชาคณิตศาสตร์ได้ร้อยละ 58.9 2) ตัวแปรพัฒนาการของการเรียนคณิตศาสตร์ ได้รับอิทธิพลทางตรงจากฐานะทางเศรษฐกิจของผู้ปกครองสูงสุด รองลงมาคือ ผลสัมฤทธิ์ทางการเรียนเดิม การศึกษาของครูผู้สอน และเจตคติต่อวิชาคณิตศาสตร์ของนักเรียน และได้รับอิทธิพลทางอ้อมจากการศึกษาของผู้ปกครองสูงสุด โดยส่งผ่านทางฐานะทางเศรษฐกิจของ

ผู้ปกครองและเจตคติต่อวิชาคณิตศาสตร์ รองลงมา คือ ฐานะทางเศรษฐกิจของผู้ปกครอง โดยส่งผ่านทางผลสัมฤทธิ์ทางการเรียนเดิม และการศึกษาของครูผู้สอน โดยส่งผ่านทางเจตคติต่อวิชาคณิตศาสตร์ 3) โมเดลปัจจัยที่ส่งผลต่อคะแนนพัฒนาการวิชาคณิตศาสตร์ของนักเรียนเพศชายและเพศหญิง ชั้นมัธยมศึกษา ปีที่ 3 มีความไม่แปรเปลี่ยนในด้านรูปแบบของโมเดล และในด้านพารามิเตอร์ LX LY และ GA 4) โมเดลปัจจัยที่ส่งผลต่อคะแนนพัฒนาการวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ได้จากการตรวจให้ค่าคะแนนแบบทวิวิภาคและได้จากการตรวจให้คะแนนความรู้บางส่วน มีความไม่แปรเปลี่ยนในด้านรูปแบบของโมเดลและในด้านพารามิเตอร์ LX LY GA BE และ PS

พัชรี จันทร์เพ็ง (2547) ศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนโดยใช้วิธีการปรับเทียบแนวตั้ง ตามวิธีทฤษฎีตอบสนองข้อสอบร่วมกับวิธีการปรับเทียบเชิงเส้นตรงเพื่อศึกษาประสิทธิภาพของการปรับเทียบแนวตั้งตามวิธีทฤษฎีตอบสนองข้อสอบแบบโลจิสติก 3 พารามิเตอร์ร่วมกับวิธีการปรับเทียบเชิงเส้นตรงที่ใช้แบบสอบร่วมภายใน และศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 2 และ 3 ในแต่ละระดับชั้น โดยภาพรวมและจำแนกตามกลุ่มความสามารถ คือ กลุ่มต่ำ กลุ่มปานกลาง และกลุ่มสูง กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 2 และ 3 ของโรงเรียนในเขตพื้นที่การศึกษาสกลนคร เขต 1 จำนวน 1,941 คน โดยสุ่มกลุ่มตัวอย่างด้วยวิธีการสุ่มแบบแบ่งชั้น กลุ่มตัวอย่างแบ่งเป็น 2 กลุ่ม คือ กลุ่มปรับเทียบคะแนน เป็นนักเรียนระดับชั้น ม.1 ม.2 และ ม.3 จำนวน 543 534 และ 504 คน ตามลำดับ และกลุ่มสอบทานผล เป็นนักเรียนระดับชั้น ม.1 ม.2 และ ม.3 จำนวน 113 120 และ 127 คน ตามลำดับ เครื่องมือที่ใช้ในการวิจัยเครื่องนี้เป็นแบบวัดความสามารถทางคณิตศาสตร์ เรื่อง "จำนวนและการดำเนินการ" สำหรับระดับชั้นมัธยมศึกษาปีที่ 1 2 และ 3 รวม 3 ฉบับ ฉบับละ 30 ข้อ ที่มีค่าความเที่ยงของแบบสอบเท่ากับ 0.803, 0.912 และ 0.857 ตามลำดับ โดยมีข้อสอบร่วมระหว่างระดับชั้น ม.1 และ ม.2 จำนวน 6 ข้อ ข้อสอบร่วมระหว่างระดับชั้น ม.2 และ ม.3 จำนวน 6 ข้อ โดยวิเคราะห์ข้อมูลเพื่อประมาณค่าความสามารถของนักเรียน ด้วยโปรแกรม BILOG 3.04 คำนวณหาค่าคงที่เพื่อสร้างสมการปรับเทียบคะแนนด้วยโปรแกรม EQUATE 2.0 ปรับเทียบคะแนนเชิงเส้น ด้วยวิธี "กำลังสองน้อยสุด" ด้วยโปรแกรม SPSS for Windows Version 11.0 สรุปผลการวิจัยได้ดังนี้ 1) ความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนตามรูปแบบที่ศึกษาสำหรับผู้สอบที่มีความสามารถแตกต่างกัน 3 ระดับ พบว่ามีค่าเท่ากับ 0.063 ซึ่งเป็นความคลาดเคลื่อนมาตรฐานที่ต่ำและความเพียงพอของความคลาดเคลื่อนของการปรับเทียบคะแนน พบว่า ค่าดัชนีความแตกต่างของการปรับเทียบคะแนนอยู่ในระดับน่าพอใจ (C = 0.225) 2) ผลการศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนโดยภาพรวม พบว่า ค่าเฉลี่ยของคะแนนความสามารถทั้ง 3 ระดับชั้น

แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยความสามารถทางคณิตศาสตร์ของนักเรียนเพิ่มขึ้นตามระดับชั้นที่สูงขึ้น นั่นคือ ระดับชั้น ม.1 เมื่อขึ้น ม.2 มีพัฒนาการสูงขึ้น 0.167 คะแนน นักเรียนระดับชั้น ม.2 เมื่อขึ้น ม.3 มีพัฒนาการสูงขึ้น 0.174 คะแนน และนักเรียนระดับชั้น ม.1 เมื่อขึ้น ม.3 มีพัฒนาการมากที่สุด เท่ากับ 0.312 คะแนน ส่วนผลการศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนจำแนกตามกลุ่มความสามารถ คือ กลุ่มต่ำ กลุ่มปานกลาง และกลุ่มสูง พบว่า นักเรียนในแต่ละกลุ่มสามารถ มีคะแนนพัฒนาการเพิ่มขึ้นตามระดับชั้นที่สูงขึ้น โดยพิจารณาจากอัตราพัฒนาการ และนักเรียนในกลุ่มความสามารถสูงมีอัตราพัฒนาการมากที่สุด รองลงมา คือ กลุ่มปานกลาง และกลุ่มต่ำ ตามลำดับ

จรรยา สิงห์ทอง (2550) ศึกษาผลของการวัดการเปลี่ยนแปลงแบบวัดซ้ำสองครั้งจำนวน 5 วิธี ได้แก่ 1) การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ 2) การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ 3) การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง 4) การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และ 5) การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ โดยเทียบผลของการวัดการเปลี่ยนแปลงจากค่าความสัมพันธ์ของคะแนนการเปลี่ยนแปลงและประสิทธิภาพของการวัดคะแนนการเปลี่ยนแปลงกับวิธีการประมาณค่าที่ใช้เป็นเกณฑ์ คือ การวัดการเปลี่ยนแปลงด้วยวิธีโมเดลพหุระดับ ภายใต้เงื่อนไข ความยาวของแบบทดสอบ 3 ขนาด คือ 40 30 และ 20 ข้อ และขนาดกลุ่มตัวอย่าง 4 ขนาด คือ 1000 500 300 และ 100 คน ผลการวิจัยสรุปว่า การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์มีความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงไม่เป็นลบเข้าใกล้ศูนย์ ส่วนคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์กับวิธีโมเดลพหุระดับมีความสัมพันธ์กันทางบวกอยู่ในระดับที่สูงกว่าวิธีอื่น ในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง การวัดคะแนนการเปลี่ยนแปลงทุกวิธีมีความเที่ยงตรงระดับเดียวกันในการประมาณค่าคะแนนการเปลี่ยนแปลง การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์มีความเชื่อมั่นสูงกว่าวิธีอื่นในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง

ธีรยุทธ ภูเขา (2550) ศึกษาประสิทธิภาพของโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง โมเดลออโตรีเกรซซีพที่มีตัวแปรแฝงพัฒนาการ และโมเดลกึ่งซิมเพล็กซ์ที่มีตัวแปรแฝงพัฒนาการ ในการศึกษาปัจจัยด้านเชาวน์ปัญญาและเชาวน์อารมณ์ที่ส่งผลต่ออัตราพัฒนาการวิชาคณิตศาสตร์ และมีจุดมุ่งหมายเฉพาะเพื่อ ศึกษาปัจจัยด้านเชาวน์ปัญญาและเชาวน์อารมณ์ที่ส่งผลต่ออัตราพัฒนาการวิชาคณิตศาสตร์ และศึกษาปัจจัยด้านเชาวน์อารมณ์และเชาวน์ปัญญาที่ส่งผลต่อผลสัมฤทธิ์วิชาคณิตศาสตร์ กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 2 ปีการศึกษา 2549 จำนวน 460 คน ในสังกัดสำนักงานเขตพื้นที่การศึกษาชัยนาท เครื่องมือในการวิจัยคือ

แบบทดสอบวัดเชาวน์ปัญญา แบบทดสอบวัดเชาวน์อารมณ์ และแบบทดสอบคู่ขนานวิชาคณิตศาสตร์ 4 ฉบับ ข้อมูลสำหรับการวิจัยเป็นข้อมูลระยะยาว 4 ครั้ง ในระยะเวลา 4 เดือน โดยมีค่าความเชื่อมั่นในการวัดทั้ง 4 ครั้งที่คำนวณจากสูตร KR20 เท่ากับ 0.778 0.896 0. 913 และ0.915 ตามลำดับ ความเชื่อมั่นที่คำนวณจากสูตรr_B เท่ากับ 0.785 0.899 0.915 และ 0.918 ตามลำดับ วิเคราะห์โมเดลสมการโครงสร้างโดยใช้โปรแกรมลิสเรล

ผลการวิจัยพบว่า โมเดลพัฒนาการที่มีตัวแปรแฝงมีประสิทธิภาพสูงสุดในการศึกษาปัจจัยด้านเชาวน์ปัญญา และเชาวน์อารมณ์ที่ส่งผลต่ออัตราพัฒนาการทางคณิตศาสตร์ (χ^2 =191.76, df = 64, p= 0.00, RMSEA=0.066, RMR=0.080, GFI= 0.94, AGFI=0.92, NFI=0.96, IFI=0.97) รองลงมาคือ โมเดลออโตรีเกรสซีพที่มีตัวแปรแฝงพัฒนาการและโมเดลกึ่งซิมเพล็กซ์ที่มีตัวแปรแฝงพัฒนาการตามลำดับ ค่าน้ำหนักความสำคัญของเชาวน์ปัญญาที่ส่งผลต่ออัตราพัฒนาการวิชาคณิตศาสตร์และผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์เท่ากับ 0.47 และ 2.53 ตามลำดับ ส่วนเชาวน์อารมณ์ไม่ส่งผลต่ออัตราพัฒนาการทางคณิตศาสตร์และผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์

การศึกษาเกี่ยวกับการวัดการเปลี่ยนแปลงในประเทศไทย เป็นการศึกษาเพื่อหาวิธีการวัดหรือโมเดลที่สามารถให้คะแนนพัฒนาการที่มีความน่าเชื่อถือและแม่นยำมากที่สุด รวมทั้งมีการศึกษาปัจจัยที่เกี่ยวข้องกับระดับพัฒนาการ และเปรียบเทียบระดับพัฒนาการเมื่อสถานการณ์ต่างกันเพื่อใช้เป็นข้อมูลสำหรับประเมินผลการจัดการเรียนการสอนตามหลักสูตร จากงานวิจัยในประเทศเกี่ยวกับการวัดการเปลี่ยนแปลงหรือการวัดพัฒนาการที่มีจำนวนมากและดำเนินมาอย่างต่อเนื่องย่อมแสดงให้เห็นว่าการวัดพัฒนาการยังคงมีประโยชน์สำหรับการประเมินผลการจัดการเรียนการสอน แต่อย่างไรก็ตามการเลือกวิธีการวัดการเปลี่ยนแปลงหรือวิธีการคำนวณพัฒนาการไปใช้นั้นจักต้องคำนึงถึงปัจจัยที่เกี่ยวข้องรอบด้าน และเงื่อนไขที่กำหนดไว้แต่ละวิธีเพื่อให้คะแนนพัฒนาการที่คำนวณได้เป็นสิ่งที่สะท้อนความสามารถที่แท้จริงของผู้สอบมากที่สุด

งานวิจัยต่างประเทศเกี่ยวกับการวัดการเปลี่ยนแปลง

เอมเบรทสัน (Embretson.1991) ได้เสนอโมเดลราสช์พหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (Multidimensional Rasch Model for Learning and Change: MRMLC) เพื่อใช้วัดคะแนนพัฒนาการ โดยพัฒนาจากโมเดลราสช์แบบพหุมิติด้วยการเพิ่มโครงสร้าง Wiener simplex เพื่อหาคะแนนพัฒนาการรายบุคคล ซึ่งอยู่ในเทอมของความสามารถของผู้สอบที่เพิ่มขึ้นระหว่างช่วงเวลา เมื่อค่าพารามิเตอร์ของข้อสอบมีค่าคงที่ แต่ MRMLC ไม่ต้องใช้แบบทดสอบฉบับเดิมจึงเลี่ยงปัญหาการจำข้อสอบได้ อันส่งผลให้เกิดความลำเอียงในการตอบซึ่งจัดว่าเป็นการละเมิดข้อตกลงเบื้องต้นของความอิสระเฉพาะที่ (Local Independence) ระหว่างข้อสอบ เอมเบรทสันได้ทดลองใช้

โมเดล MRMLC กับข้อมูลจริงซึ่งได้จากแบบทดสอบวัดความสามารถในการเรียนรู้มิติสัมพันธ์ (SLAT) ประกอบด้วยข้อสอบจำนวน 76 เป็นแบบทดสอบก่อนเรียน 24 ข้อ แบบทดสอบหลังเรียน อีก 2 ชุด ชุดละ 24 ข้อ รวมทั้งสิ้น 72 ข้อ แต่เพื่อให้เกิดความเชื่อมโยงระหว่างกลุ่มจึงเพิ่มข้อสอบร่วมในแบบทดสอบก่อน อีก 4 ข้อ เมื่อหาสหสัมพันธ์ระหว่าง MRMLC กับคะแนนมาตรฐานที่ได้จากการวัดการเปลี่ยนแปลงแบบมาตรฐานเดิมในการทดสอบก่อนเรียน คะแนนเพิ่มครั้งที่หนึ่ง คะแนนเพิ่มครั้งที่สอง พบว่ามีความสัมพันธ์กันในระดับสูงทุกค่า (.994, .953 และ .949) ในขณะที่เปรียบเทียบกับโมเดลของ Anderson ปรากฏว่าค่าสหสัมพันธ์กับคะแนนมาตรฐานนั้น ค่าสหสัมพันธ์ของทั้งสองวิธีในการสอบก่อนเรียนคะแนนเพิ่มครั้งที่หนึ่งและครั้งที่สองมีค่าต่ำ (.994, .598, และ .389)

ฟิสเชอร์ และ พาร์เซอร์ (Fischer; & Parzer. 1991) ได้เสนอวิธีการใช้โมเดลเรทติ้งสเกลเชิงเส้น (Linear Rating Scale Model: LRSM) ซึ่งให้คะแนนโดยกำหนดระดับคะแนนที่มีช่วงคะแนนเท่ากันทุกช่วงในรูปเชิงเส้นตรง เพื่อศึกษาการเปลี่ยนแปลง ฟิสเชอร์ และ พาร์เซอร์ ประมาณค่าพารามิเตอร์โดยใช้วิธีความน่าจะเป็นสูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood: CML) และทดสอบสมมติฐาน H_0: ไม่มีการเปลี่ยนแปลง โดยใช้อัตราส่วนความน่าจะเป็น (Likelihood ratio)ฟิสเชอร์ และ พาร์เซอร์ ทดสอบแนวคิดโดยใช้ข้อมูลจริงจาก 2 กลุ่มคนไข้ที่ภายนอกคลีนิกของมหาวิทยาลัยเวียนนาที่มีความผิดปกติทางจิต กลุ่มแรกเป็นกลุ่มทดลองจำนวน 37 คนที่ได้รับการรักษาด้วยยาและเข้าฝึกอบรมผ่อนคลายด้วยการฝึกการสนทนากลุ่มเป็นเวลา 7 วัน ส่วนกลุ่มควบคุมมีจำนวน 50 คน ได้รับยาเพียงอย่างเดียว คนไข้ทั้งหมดประเมินตนเองโดยทำแบบวัดอาการด้านต่างๆ ซึ่งเป็นมาตรวัดประมาณค่า 4 ช่วง ประเมินก่อนรักษา และหลังการรักษาเมื่อเวลาผ่านไป 1 ปี ผลการวิจัยไม่พบการเปลี่ยนแปลงเกิดขึ้นจากการทดสอบทั้งสองครั้ง และมีปฏิสัมพันธ์ระหว่างลักษณะของคนไข้กับการจัดกระทำ (treatment) ที่ให้ ส่วนผลการศึกษาจากการใช้ข้อมูลจริง ฟิสเชอร์ และ พาร์เซอร์พบว่าข้อมูลไม่เป็นไปตามข้อตกลงเรื่องความเป็นเอกมิติของแบบวัด เนื่องจากข้อมูลนั้นใช้แบบวัดที่เป็นพหุมิติ และใช้ข้อคำถามกับข้อมูลเพียงหนึ่งข้อ หรือ หนึ่งลักษณะอาการที่ตอบในการวัดหลายครั้ง แต่ข้อมูลจริงมีข้อคำถามหลายลักษณะอาการ จึงยากที่จะทดสอบสมมติฐานได้

ต่อมาฟิสเชอร์ และ โพนอคนี (Fischer; & Ponocny. 1994) ได้เสนอโมเดลพาร์เชียลเครดิตเชิงเส้น (Linear Partial Credit Model: LPCM) ซึ่งเป็นโมเดลที่พัฒนาจากโมเดลราสช์แบบเอกมิติสำหรับให้คะแนนแบบพหุวิภาคสำหรับวัดพัฒนาการ แทนการใช้โมเดลเรทติ้งสเกลเชิงเส้น (LRSM) ที่พบว่าเมื่อใช้กับข้อมูลจริงในการวัดคะแนนการเปลี่ยนแปลงของอาการคนไข้ทางจิตมีข้อจำกัดเกี่ยวกับค่าพารามิเตอร์ของข้อคำถาม (β_i) พารามิเตอร์ของช่วงคะแนน (ω_h) และพารามิเตอรีของผู้สอบ (θ_u) ที่ทำให้ข้อมูลที่ได้เหมาะกับโมเดลในการวัดพัฒนาการ (Fischer; & Parzer. 1991) LPCM มีความยืดหยุ่นกว่า LRSM และ LRSM เป็นกรณีหนึ่งของ LPCM แต่เนื่องจากการวิเคราะห์ LPCM มี

ค่าประมาณสูงกว่า LRSM ดังนั้น ฟิสเชอร์ และ โพนอคนี จึงพัฒนาวิธีการประมาณค่า CML ที่ใช้ใน LRSMให้มีประสิทธิภาพมากขึ้น โดยใช้ Broyden-Fletcher Goldfarb Shanno (quasi-Newton) ในการประมาณค่าพารามิเตอร์ แทนการประมาณค่าด้วย Newton Raphson algorithm ที่ใช้ใน LRSM

ในการทดลองโมเดลกับข้อมูลมาฟิสเชอร์ และ โพนอคนี ได้ใช้กลุ่มตัวอย่างคนไข้ที่ผิดปกติทางจิตชุดเดิมที่วิเคราะห์ด้วย LRSM ในการทดลองของ LPCM นี้ได้ทดลองกับคนไข้กลุ่มทดลองจำนวน 37 คน ที่ได้รับยาร่วมกับการฝึกอบรมการผ่อนคลายความเครียดด้วยการมีส่วนร่วมในกลุ่มและกลุ่มควบคุม 50 คน ที่รักษาด้วยยาอย่างเดียว ทั้งสอบกลุ่มต้องประเมินอาการตนเองด้วยแบบประเมินแบบมาตรประมาณค่า 4 ช่วง จำนวน 24 ข้อก่อนทำการทดลอง หลังเข้ารับการรักษา 1 ปี ทั้งสองกลุ่มประเมินตนเองใหม่ด้วยข้อคำถาม 24 ข้อเดิมที่ปรับใหม่โดยให้คะแนนแบบบางส่วนตามโมเดล LPCM ผลการวิจัยพบว่า กลุ่มทดลองไม่มีการเปลี่ยนแปลงในอาการป่วย แสดงว่าการเข้าฝึกอบรมไม่ได้มีผลให้อาการป่วยทางจิตบรรเทาลลง ซึ่งสอดคล้องกับผลการวิจัยจาก LRSM และยังให้ผลการวิจัยอีกว่าการใช้ยามีผลทำให้เกิดการเปลี่ยนแปลง

จากงานวิจัยนี้คณะผู้วิจัยได้นำเสนอโปรแกรมสำเร็จรูป LPCM สำหรับใช้วิเคราะห์ข้อมูลสำหรับการวัดซ้ำที่มีประสิทธิภาพ มีความถูกต้องในการประมาณค่าอิทธิพลของตัวแปรมากกว่าการประมาณค่าพารามิเตอร์ของข้อสอบ มีประโยชน์และง่ายสำหรับการวิเคราะห์ข้อมูลที่เป็นการตอบแบบพหุวิภาค (Polytomous responses) และสามารถทดสอบสมมติฐานการเปลี่ยนแปลง (Testing hypotheses on change)

เอมเบรทสัน (Embretson, 1995) ได้พัฒนาโมเดลราสช์พหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (MRMLC) โดยนำกระบวนการและโครงสร้างของความรู้เข้าใส่ไปในโมเดลราสช์โดยตั้งชื่อโมเดลใหม่ว่า MRMLC+ โดยที่โมเดลสามารถหาพัฒนาการรายบุคคล โดยการพิจารณาความสามารถที่เพิ่มขึ้นระหว่างช่วงเวลาและ จากพัฒนาการในกระบวนการเรียนและความรู้ที่เพิ่มขึ้นซึ่งอยู่ในรูปของพารามิเตอร์ของข้อสอบ โดย เอมเบรทสัน นำโมเดลไปใช้กับพนักงานใหม่ของ Air Force จำนวน 578 คน ที่ได้รับการอบรมขั้นพื้นฐานจำนวน 3 สัปดาห์ โดยใช้แบบทดสอบการแก้ปัญหาทางคณิตศาสตร์ทำการทดสอบ 3 ครั้ง ด้วยแบบทดสอบที่ไม่ซ้ำกันคือ แบบทดสอบก่อนเรียน ระหว่างเรียน และหลังเรียน โดยแบบทดสอบก่อนเรียนและแบบทดสอบหลังเรียนมีจำนวนฉบับละ 10 ข้อ ประกอบด้วยปัญหาเรื่องอัตราส่วนจำนวน 7 ข้อ และ เรื่องของอัตราส่วนที่ขยายตัวแปรและจำนวนสมการมากขึ้น จำนวน 3 ข้อ ส่วนแบบทดสอบระหว่างเรียนจะเป็นแบบทดสอบที่มีเรื่องอัตราส่วนจำนวน 7 ข้อ และข้อสอบร่วมจำนวน 3 ข้อ จัดการสอบแบบจัตุรัสลาติน (Latin Square) ผลการวิจัยพบว่า จำนวนคำในโจทย์ปัญหา ระดับความสามารถในการอ่าน จำนวนตัวแปร มีค่าสหสัมพันธ์สูงกับค่าความยากที่ได้จากโมเดลราสช์ พารามิเตอร์ของโมเดล MRMLC+ ประมาณค่าด้วยโปรแกรม

LINLOG ความสามารถเริ่มต้นและความสามารถที่เพิ่มขึ้น (Initial Competency Level and Modifiability) ของแต่ละคนถูกประมาณค่าด้วยวิธี Newton-Raphson การแปลความหมายของระดับความสามารถเริ่มต้นและความสามารถที่เพิ่มขึ้นได้จากการนำระดับความสามารถของแต่ละคนมาใส่ในมาตรของค่าความยากของข้อสอบที่แยกให้อยู่ในรูปของชนิดของจำนวนคำในโจทย์ปัญหา ระดับความสามารถในการอ่าน จำนวนตัวแปร ซึ่งจากผลการวิจัยพบว่าโมเดล MRMLC+ มีประโยชน์ในการทำความเข้าใจลักษณะและระดับความสามารถของพารามิเตอร์ของคนสามารถประยุกต์ใช้ได้กับข้อสอบที่มีความซับซ้อนในด้านของการวัดความรู้ที่มีอยู่

เมย์ และไนส์วานเดอร์ (May & Nicewander, 1998) ได้ศึกษาวิธีการวัดคะแนนความแตกต่างของค่าประมาณความสามารถ (Difference score) ที่จะช่วยป้องกันหรือลดขนาดของอิทธิพลที่เกิดจากความยากของแบบทดสอบในครั้งแรกที่มีความลำเอียงต่อคะแนนการเปลี่ยนแปลงของผู้สอบที่มีความสามารถแตกต่างกัน เมย์ และไนส์วานเดอร์ ได้ศึกษาด้วยการจำลองข้อมูล (Simulation) แบ่งผู้สอบเป็น 3 กลุ่ม ตามระดับความสามารถจริง (θ) คือ กลุ่มเก่ง ปานกลาง และอ่อน และสุ่มข้อสอบมา 4 ประเภทประเภทละ 50 ข้อ ได้แก่ ข้อสอบที่มีความยากใกล้เคียงกัน (Uniform) ข้อสอบที่มีระดับความยากกระจายเป็นโค้งปกติ(Symmetric) ข้อสอบที่ง่ายและข้อสอบที่ยาก และใช้โมเดลของการเปลี่ยนแปลงแฝง (Latent Change) 3 โมเดล คือ โมเดลที่ 1 พัฒนาการคงที่ ($\theta_2 = \theta_2 + 1$)โมเดลที่ 2 พัฒนาการเพิ่มขึ้น($\theta_2 = 1.2\theta_2 + 1$) และ โมเดลที่ 3 พัฒนาการลดลง($\theta_2 = 0.8\theta_2 + 1$) และใช้วิธีวัดหาคะแนนพัฒนาการ 2 แบบคือ คะแนนพัฒนาการที่ได้จากความแตกต่างของคะแนนดิบที่ได้จากการวัดครั้งที่ 2 และคะแนนดิบจากการวัดครั้งแรก และคะแนนพัฒนาการที่ได้จากผลต่างของการประมาณค่าด้วยโมเดลการตอบข้อสอบแบบ 3 พารามิเตอร์ ในการวัดครั้งที่ 2 และครั้งที่ 1 (θ_2-θ_1) โดยใช้วิธีการประมาณค่าแบบ EAP (θ)(Expected a posteriori estimates) ผลการวิจัยพบว่าคะแนนพัฒนาการที่มาจากผลต่างของคะแนนดิบตอบถูกในการวัดครั้งที่ 1 และ 2 นั้นในทุกประเภทของแบบทดสอบและกลุ่มความสามารถของผู้สอบ และคะแนนพัฒนาการทุกคะแนนบิดเบือนจากคะแนนพัฒนาการที่แท้จริง โดยเฉพาะอย่างยิ่งในกลุ่มคนเก่ง จะให้คะแนนพัฒนาการต่ำกว่าความเป็นจริงมากกว่ากลุ่มอ่อนและปานกลาง สำหรับกลุ่มอ่อนนั้นจะได้คะแนนพัฒนาการต่ำกว่าความเป็นจริงมากในกรณีข้อสอบยาก สำหรับคะแนนพัฒนาการที่ได้จากคะแนนความแตกต่างที่วิเคราะห์ด้วยโมเดลการตอบข้อสอบแบบบ EAP(θ) พบว่าคะแนนพัฒนาการที่ได้มีค่าใกล้เคียงกับคะแนนพัฒนาการที่แท้จริงเกือบทุกกรณี ยกเว้นในกรณีคนเก่งที่ใช้แบบทดสอบง่าย และคนอ่อนที่ใช้แบบทดสอบยาก ที่คะแนนความแตกต่างที่ได้จะต่ำกว่าความเป็นจริงเล็กน้อย ซึ่งมีค่าน้อยเมื่อเทียบกับคะแนนพัฒนาการที่ได้จากผลต่างของคะแนนดิบ สำหรับแบบทดสอบที่ใช้แบบทดสอบปรับเหมาะ คะแนนพัฒนาการที่ได้จากวิธี EAP (θ) จะมีค่าใกล้เคียงกับค่าคะแนนพัฒนาการที่แท้จริงทุกกรณี

โมเดลการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบมีหลายวิธี และใช้โมเดลประมาณค่าแตกต่างกันไปตามจำนวนพารามิเตอร์ที่สอดคล้องกับข้อมูล วิธีการได้มาซึ่งค่าพารามิเตอร์ข้อสอบหรือผู้สอบนั้นมีหลายขั้นตอน ต้องใช้เวลา และเครื่องมือสำหรับช่วยคำนวณค่าพารามิเตอร์ ตลอดจนข้อมูลสำหรับศึกษาพัฒนาการความสามารถนั้นต้องมีความสอดคล้องกลมกลืนกับโมเดลการวัดการเปลี่ยนแปลงที่เลือกใช้ด้วย ดังนั้นหากพิจารณาวิธีการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบแล้วจะพบว่าวิธีการวัดการเปลี่ยนแปลงความสามารถซึ่งนำค่าพารามิเตอร์ความสามารถในการสอบครั้งสองลบด้วยค่าพารามิเตอร์ความสามารถในการสอบครั้งที่หนึ่ง เป็นวิธีการคำนวณการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบที่มีขั้นตอนไม่มากและสะดวกในการคำนวณ ใกล้เคียงกับวิธีการคำนวณผลต่างคะแนนดิบซึ่งสามารถคำนวณได้ง่าย และนิยมใช้เป็นดัชนีบ่งชี้ถึงความสามารถที่เพิ่มขึ้นของผู้เรียน

2. งานวิจัยเกี่ยวกับการประมาณค่าพารามิเตอร์และปัจจัยที่เกี่ยวข้อง

โยส์ (Yoes. 1994) ได้เปรียบเทียบวิธีการประมาณค่า 3 วิธีคือ วิธี JML วิธี Bayesian และวิธี MML โดยตรวจสอบความเที่ยงตรงของกระบวนการประมาณค่าพารามิเตอร์ข้อสอบด้วยโปรแกรม Logist, BILOG และ ASCAL ด้วยการจำลองข้อมูลแบบมอนติคาร์โล ภายใต้เงื่อนไขขนาดกลุ่มตัวอย่าง (250 500 1000 และ 2000) ความยาวแบบทดสอบ (15 25 50 75 และ 100) การแจกแจงความสามารถ (แบบปกติ แบบเท่ากัน แบบเบ้ทางลบ แบบเบ้ทางบวก) และคุณลักษณะแบบทดสอบประกอบด้วยค่าอำนาจจำแนกปานกลาง กับสูง และ การแจกแจงค่าความยากเท่ากัน กับแบบปกติ โดยประมาณค่าพารามิเตอร์ 240 ครั้ง ในแต่ละวิธีประมาณค่า ผลการศึกษาดูจากค่าสหสัมพันธ์ ค่าความลำเอียง และ RMSE เมื่อใช้วิธีประมาณแบบ MML ค่าจากแปรแกรม BILOG ในการกำหนดค่าทั่วๆไป ได้ผลดีกับข้อสอบที่มีจำนวนข้อน้อยและกลุ่มตัวอย่างขนาดเล็กให้ค่าประมาณพารามิเตอร์อำนาจจำแนกเกินกว่าความเป็นจริง ค่าการเดาไม่ดีนัก แต่ค่าประมาณความยากดีพอๆกัน และพบว่าขนาดกลุ่มตัวอย่างและความยาวแบบทดสอบเป็นปัจจัยที่สำคัญในการประมาณค่าพารามิเตอร์ ส่วนการแจกแจงความสามารถไม่มีผลต่อกระบวนการในการประมาณค่า

เคลเดอร์แมน (Kelderman. 1996: 155 – 168) ได้ศึกษาผลการใช้โมเดลราชส์สำหรับการให้คะแนนแบบบางส่วน เมื่อโมเดลที่อยู่ในรูปพหุมิติ โมเดลอาจจะกำหนดขึ้นเพื่อเกิดการตอบข้อสอบบนเส้นความสามารถต่อเนื่อง โมเดลพาร์เชียลเครดิตแบบพหุมิติ การตอบแตกต่างกันอาจจะอธิบายได้ว่ามีคุณลักษณะที่ต่างกัน ในการวิเคราะห์ใช้ข้อมูลของ van Kuyk (1988) จากแบบทดสอบเกี่ยวกับมโนมมติเกี่ยวกับขนาดและ การใช้แบบทดสอบเมตริกซ์โพรเกรสซีพของราเวน (Raven Progressive Metrices) ทำการศึกษาโดยใช้การประมาณค่าความน่าจะเป็นสูงสุดรวมถึงการทดสอบ ความ

สอดคล้องกลมกลืน และทิศทางการนำไปใช้ ผลการทดสอบ ความสอดคล้องกลมกลืนของแบบทดสอบทั้งสองฉบับพบว่า โมเดลพาร์เชียลเครดิตแบบพหุมิติมีความเหมาะสมมากกว่าโมเดลพาร์เชียลเครดิตแบบมิติเดียว

คุก (Cook. 1996) ได้เปรียบเทียบโมเดลพาร์เชียลเครดิต โมเดลเจอเนอรอลไรส์พาร์เชียลเครดิต และโมเดลเกรดเรสพอนส์ในบริบทของแบบทดสอบย่อย (Testlet) ใช้ข้อมูลจากการตอบของผู้สอบจำนวน 2548 คน ในการตอบวัดความรู้ทางภาษาในเดือนพฤศจิกายน 1994 ของการสอบ SAT I แล้วจึงจำลองข้อมูลเพิ่มมาอีกสองชุดโดยการใช้วิธีการวิเคราะห์องค์ประกอบเชิงเส้น ศึกษาค่าสหสัมพันธ์เพียร์สันและรากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSEs) หลังจากที่ทำการปรับเทียบคะแนนแล้ว เปรียบเทียบค่าสารสนเทศของแบบทดสอบและชุดข้อสอบย่อย และเปรียบเทียบความสอดคล้องกลมกลืนโมเดล พบว่า โมเดลทั้งสามให้ค่าประมาณความสามารถที่ดี ค่าสัมประสิทธิ์สหสัมพันธ์อยู่ระหว่าง .9747 ถึง .9978 และทั้งสามโมเดลได้ผลดีในการประมาณค่าเมื่อทราบพารามิเตอร์ผู้สอบ ในส่วนของข้อมูลที่จำลองขึ้น ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างองค์ประกอบแรก คะแนนมาตรฐาน Z และค่าประมาณความสามารถอยู่ในช่วง .9349 ถึง .9462 ความแตกต่างระหว่างฟังก์ชันสารสนเทศของโมเดลพาร์เชียลเครดิตและโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตสะท้อนให้เห็นค่าประมาณพารามิเตอร์อำนาจจำแนกสำหรับโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตสำหรับข้อมูลที่ค่าพารามิเตอร์อำนาจจำแนกของโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตสูงกว่า1.0 โมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตให้ค่าสาระสนเทศสูงกว่าอีกสองโมเดลบนมาตรวัดความสามารถเดียวกัน สำหรับข้อมูลที่ค่าพารามิเตอร์อำนาจจำแนกของโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตต่ำกว่า1.0 โมเดลพาร์เชียลเครดิตให้ค่าสาระสนเทศสูงที่สุดบนมาตรวัดความสามารถเดียวกัน ส่วนโมเดลเกรดเรสพอนส์ให้ค่าสารสนเทศสูงกว่าอีกสองโมเดลเฉพาะกลุ่มที่มีค่าความสามารถในระดับต่ำ

แต-ลิม และ ฮาร์เวล (Tay- Lim & Harwell. 1997) ได้สังเคราะห์งานวิจัยที่สามารถใช้เพื่อเป็นข้อมูลในการใช้โมเดลการตอบข้อสอบ ศึกษาศักยภาพของการสังเคราะห์งานวิจัยและความสัมพันธ์ระหว่างความแม่นยำของการประมาณค่าพารามิเตอร์ และจำนวนของข้อสอบและผู้สอบ ผลการสังเคราะห์ พบว่ามีงานวิจัยประมาณ 63 เปอร์เซนต์ที่ศึกษาของดัชนีความแม่นยำในการประมาณค่าจากจำนวนข้อสอบและผู้สอบ เมื่องานวิจัยส่วนมากให้ความสนใจจำนวนข้อสอบและจำนวนผู้สอบ ที่ส่งผลต่อความแม่นยำในการวัดมากเกินไป ผลการสังเคราะห์งานวิจัยจึงแนะนำให้ศึกษาปัจจัยอื่น เช่นวิธีการที่ใช้ในการประมาณค่าพารามิเตอร์ การแสดงบทบาทความสำคัญของความแม่นยำในการวัด เช่น การอธิบายความสำคัญของปริมาณของความแปรปรวน

ปาร์ก (Park.1997) ได้ศึกษาความแม่นยำของการประมาณค่าพารามิเตอร์ด้วยโมเดล GPCM โดยแบ่งการศึกษาออกเป็นสองส่วน ส่วนแรกเป็นการศึกษาปัจจัยที่ส่งผลต่อความแม่นยำ ความแปรปรวน และความลำเอียงต่อวิธีประมาณค่า ซึ่งใช้ในโมเดล GPCM ผลการศึกษาพบว่าตัวประมาณค่า MML ของพารามิเตอร์ของโมเดล GPCM ที่ได้จากโปรแกรม PARSCALE ให้ผลที่ยอมรับได้ในแต่ละเงื่อนไข แต่เมื่อพิจารณาความลำเอียง ค่าเฉลี่ยความลำเอียงไม่ลดลงเมื่อขนาดตัวอย่างและความยาวข้อสอบเพิ่มขึ้น ความลำเอียงได้ส่งผลต่อค่า RMSE ที่โตกว่าในการประมาณค่าพารามิเตอร์รายการคำตอบ จึงแนะให้ศึกษาผลของความลำเอียงที่มีต่อค่าประมาณความสามารถ การพัฒนาคลังข้อสอบ และการทดสอบแบบปรับเหมาะด้วยโมเดลการตอบข้อสอบแบบพหุส่วนที่สอง เป็นการศึกษาประสิทธิภาพของวิธีการประมาณค่าแบบเบส์ เมื่อประมาณค่าด้วยโมเดล GPCM ผลการศึกษาพบว่า ให้ความแม่นยำสูงในการประมาณค่าพารามิเตอร์เมื่อข้อมูลมีขนาดเล็กโดยเฉพาะค่าพารามิเตอร์ความชัน ขณะที่มีผลเพียงเล็กน้อยต่อความแม่นยำในการประมาณค่าพารามิเตอร์ความยากประจำข้อ

ลี ลิสซี และ หยาง (Li; Lissitz; & Yang. 1999: 1 – 31) ได้ศึกษาวิธีการเทียบคะแนนโดยใช้โค้งลักษณะข้อสอบกับแบบทดสอบผสม ระหว่างข้อสอบแบบทวิวิภาคและพหุวิภาค โดยใช้โมเดลเจเนอรอลไรซ์ พาร์เชียลเครดิต ด้วยการจำลองข้อมูลจากพารามิเตอร์ของแบบทดสอบการอ่านและการเขียนของนักเรียนเกรด 4 ในโรงเรียนระดับตำบล ปัจจัยที่ศึกษาได้แก่ สัดส่วนของข้อสอบร่วมที่ให้คะแนนแบบทวิวิภาคและพหุวิภาค โดยมีข้อสอบพหุวิภาค 5 ข้อ มี 4 รายการคำตอบภายใต้เงื่อนไขข้อสอบร่วม 10 15 และ 20 ข้อ ซึ่งอ้างจากงานวิจัยของ โดโนฮิว (Donohuge, 1994) ขนาดของกลุ่มตัวอย่างเป็น 1000 2000 3000 คน สถานการณ์ที่ศึกษาการกลับสู่สภาพเดิมของพารามิเตอร์ (Parameter recovery) เมื่อใช้เฉพาะความคลาดเคลื่อนของค่าประมาณพารามิเตอร์ในแบบทดสอบเทียบคะแนน กับ สถานการณ์เทียบคะแนนในกลุ่มข้อมูลจริง เมื่อความคลาดเคลื่อนของค่าประมาณพารามิเตอร์ที่มีอยู่ทั้งแบบทดสอบเทียบคะแนนและแบบทดสอบร่วม ใช้ประเภทการเทียบคะแนนตามแนวนอนและแนวตั้ง ดังนั้นเงื่อนไขการศึกษาจึงมีทั้งสิ้น 36 เงื่อนไข (3× 3 ×2 ×2) ประเมินความแม่นยำของสัมประสิทธิ์การเปรียบเทียบคะแนนโดยใช้ค่าความลำเอียง (BIAS) คือ ค่าเฉลี่ยความแตกต่างของค่าพารามิเตอร์จริงและค่าประมาณที่สอดคล้องกัน และค่ารากที่สองของความคลาดเคลื่อนเฉลี่ย (RMSE) ใช้การวิเคราะห์การถดถอยพหุคูณในการสรุปองค์ประกอบต่างๆที่ใช้ในการจำลองข้อมูล ว่าส่งผลต่อความแม่นยำในสัมประสิทธิ์การเทียบคะแนนหรือไม่ เมื่อพิจารณาให้องค์ประกอบอื่นๆ มีค่าคงที่ ผลจากการศึกษาพบว่าวิธีการเทียบคะแนนโดยใช้โค้งลักษณะข้อสอบ ในโมเดล GPCM สามารถประยุกต์ใช้ได้กับแบบทดสอบผสม ภายใต้เงื่อนไขที่กำหนดได้ ความลำเอียงของสัมประสิทธิ์การเทียบคะแนนมีค่าเข้าใกล้ศูนย์และสอดคล้องกับค่า RMSE ซึ่งมีค่าต่ำ โดยผู้วิจัยได้

อภิปรายไว้ว่า ค่าทั้งสองอาจจะเพิ่มมากขึ้นในการทดสอบจริง และ รูปร่างการแจกแจงการสุ่มเชิงประจักษ์ของสัมประสิทธิ์การเทียบคะแนนที่ได้รับการประมาณค่าตามเงื่อนไขต่างๆ มีรูปร่างเป็นโค้งปกติมีความแปรปรวนต่ำ และไม่มี outlier นอกจากนี้ในงานวิจัยยังปรากฏว่า การที่ข้อสอบตรวจให้คะแนนหลายค่าเป็นข้อสอบร่วมร่วมกับข้อสอบที่ตรวจให้คะแนนสองค่าที่มีจำนวนข้อมากกว่า ไม่ให้ค่าลำเอียงและค่า RMSE ของสัมประสิทธิการเทียบคะแนนที่ต่ำกว่า คาดว่าเกิดจากผลกระทบของการกำหนดน้ำหนักสัดส่วนของรูปแบบข้อสอบที่แตกต่างกัน ที่อาจจะส่งผลต่อความแม่นยำของสัมประสิทธิ์การเทียบคะแนนได้ จึงควรมีการศึกษาเงื่อนไขนี้ให้ชัดเจนต่อไป นั่นแสดงว่าลักษณะพารามิเตอร์ข้อสอบกับกลุ่มข้อสอบร่วมมีความสำคัญต่อความแม่นยำของสัมประสิทธิการเทียบคะแนนมากกว่าองค์ประกอบของจำนวนข้อสอบร่วม และผลการวิเคราะห์การถดถอยของ log [RMSE] ในค่าประมาณ A หรือ B จากองค์ประกอบต่างๆที่จำลองขึ้นมา พบว่ามีความไวต่อการผันแปรและความแม่นยำของการประมาณค่าในแต่ละการประมาณค่าพารามิเตอร์ของการแปลง แสดงว่าองค์ประกอบต่างๆ ที่จำลองขึ้นมาส่งผลต่อการประมาณค่าสัมประสิทธิ์การเทียบคะแนน

บาสตารี (Bastari. 2000. บทคัดย่อ)เพื่อศึกษาผลขององค์ประกอบต่างๆที่มีต่อความแม่นยำในการประมาณค่าพารามิเตอร์ข้อสอบร่วมกัน เข้าสู่ในมาตรวัดร่วม โดยใช้โมเดลสามพารามิเตอร์ร่วมกับโมเดลเกรดเรสพอนด์สำหรับการเขียนตอบแบบมีโครงสร้าง เป็นการศึกษาด้วยรูปแบบแบบทดสอบร่วม ปัจจัยที่ศึกษาได้แก่ ขนาดความยาวแบบทดสอบ สัดส่วนของข้อสอบในแต่ละชนิดความยาวของข้อสอบร่วม ขนาดกลุ่มตัวอย่าง การแจกแจงความสามารถ และ วิธีการปรับเทียบ ข้อมูลแบบสองค่าและหลายค่าทั้งแบบฉบับเดียงและข้อสอบร่วมได้มาจากการจำลองข้อมูลแตก่างกันตามฟังก์ชันของปัจจัยที่ศึกษา ผลการศึกษาพบว่า ข้อสอบแบบเขียนตอบแบบมีโครงสร้างมีอิทธิพลอย่างมากต่อการประมาณค่าพารามิเตอร์สำหรับชนิดของรูปแบบข้อสอบ โดยทั่วไป ค่าพารามิเตอร์ความชันถูกประมาณค่าอย่างลำเอียงเล็กน้อยแต่มีค่าความแปรปรวนสูง ส่วนค่าพารามิเตอร์ความยากนั้นถูกประมาณค่าด้วยความลำเอียงเล็กน้อยแต่มีค่าความแปรปรวนสูงเมื่อข้อสอบเป็นแบบตอบแบบมีโครงสร้าง แต่สำหรับข้อสอบแบบเลือกตอบกลับให้ผลตรงกันข้าม สำหรับค่าประมาณพารามิเตอร์การเดาผลการศึกษาค่อนข้างน่าพอใจ สัมประสิทธิ์การแปลงเป็นค่าประมาณสัมพัทธ์ที่ดีโดยสรุปแล้วการศึกษามีแนวโน้มว่าค่าความแม่นยำในการวัดจะมีค่าสูงเมื่อ แบบทดสอบทีมีขนาดยาว มีจำนวนข้อสอบแบบเลือกตอบมากกว่า มีข้อสอบร่วมขนาดยาว จำนวนผู้สอบมาก และใช้วิธีการประมาณค่าพร้อมกัน

ปาร์ก (Park. 2000) ได้ศึกษาผลของการใช้จำนวนของรายการคำตอบที่แตกต่างกันที่มีต่อการประมาณค่าความสามารถของโมเดลเกรดเรสพอนส์และโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตและดูอิทธิพลของความยาวแบบทดสอบ ระดับความต้องการลดหรือเพิ่มรายการคำตอบ และความ

แตกต่างของค่าอำนาจจำแนก การศึกษาเป็นการจำลองข้อมูลแบบมอนติคาร์โลแบบตัวแปรระหว่างกลุ่มสองตัวคือ โมเดลการตอบข้อสอบ (โมเดลเกรดเรสพอนส์และโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิต) และ ขนาดความยาว (5 และ 15) ตัวแปรภายในกลุ่มสามตัวคือ จำนวนรายการคำตอบ (3 4 5 6 และ 7) ความต้องการลดหรือเพิ่มรายการคำตอบ (ต่ำปานกลาง และ สูง) และค่าอำนาจจำแนก (.6 .9 และ 1.4) ใช้การตอบแบบ 5 รายการคำตอบเป็นฐาน ส่วนจำนวนรายการคำตอบ 3 และ 4 จะได้มาจากการลดรายการคำตอบที่ต่ำกว่าของข้อสอบแบบ 5 รายการคำตอบ ส่วนจำนวนรายการคำตอบ 6 และ 7 ได้มาจากการเพิ่มรายการคำตอบที่สูงกว่าของข้อสอบแบบ 5 รายการคำตอบ ความต้องการเพิ่มหรือลดรายการคำตอบตัดสินจากค่าความยากของข้อสอบ ข้อสอบที่มีความยากต่ำเช่น -1.0 เช่นมีการตอบจำนวนน้อยในรายการคำตอบระดับต่ำกว่าแสดงให้เห็นว่าต้องตัดรายการนั้นออก ส่วนกรณีที่มีผู้สอบตอบรายการที่สูงกว่ามากแสดงว่าจำเป็นต้องเพิ่มรายการคำตอบ ผลการศึกษาได้ข้อค้นพบ 5 ประการ ประการแรก โมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตให้ค่าประมาณความสามารถแม่นยำกว่าโมเดลเกรดเรสพอนส์ ประการที่สอง แบบทดสอบที่มีขนาด 15 ข้อให้ค่าประมาณที่ดีกว่าแบบทดสอบที่มี 5 ข้อ ประการที่สามโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตมีความไวต่อจำนวนรายการคำตอบมากกว่าโมเดลเกรดเรสพอนส์ในการประมาณค่าความสามารถ ประการที่สี่ ความต้องการเพิ่มหรือลดจำนวนรายการคำตอบมีผลต่อโมเดลเจอเนอรอลไรส์พาร์เชียลเครดิตมาก และประการที่ห้า ความแตกต่างระหว่างความแม่นยำของการประมาณค่าของทั้งสองโมเดลแตกต่างมากเมื่อค่าอำนาจจำแนกมีค่าต่ำ

ไซเคส และ เยน (Sykes; & Yen. 2000) ได้เปรียบเทียบความสอดคล้องกลมกลืนของข้อมูลที่ประมาณค่าพารามิเตอร์ด้วยโมเดล 1PL รวมกับ โมเดล 1PPC และ โมเดล 3PL และ โมเดล 2PPC จากแบบทดสอบผสมซึ่งเป็นแบบทดสอบย่อย 6 ฉบับ ที่มีอัตราส่วนของจำนวนข้อสอบเขียนตอบแบบมีโครงสร้างต่อจำนวนข้อสอบแบบเลือกตอบ และ ค่าความยากโดยรวมแตกต่างกัน คะแนนของแบบทดสอบแบบเขียนตอบขึ้นอยู่กับระดับของคะแนนตามเกณฑ์การให้คะแนน (Scoring Rubric) ผลการศึกษาพบว่า ค่าประมาณพารามิเตอร์จากการรวมโมเดล1PL รวมกับ โมเดล 1PPC ให้ผลที่ด้อยกว่า โมเดล 3PL รวมกับ โมเดล 2PPC เนื่องจากมีข้อจำกัดเรื่องการไม่มีการเดา และค่าอำนาจจำแนกในแต่ละข้อต้องเท่ากัน และได้ศึกษาโดยจำลองข้อมูลให้มีค่าอำนาจจำแนกแตกต่างกัน โมเดล1PL รวมกับ โมเดล 1PPC ให้ค่าประมาณสารสนเทศของข้อสอบเกินกว่าความเป็นจริงสำหรับข้อสอบแบบเขียนตอบแบบมีโครงสร้างที่มีระดับการให้คะแนนสูงกว่า 3 ระดับ นอกจากนี้เป็นที่สังเกตว่ามีข้อสอบในบางข้อมีค่าสารสนเทศซึ่งรวมคะแนนให้ค่าประมาณเกินกว่าความเป็นจริงถึง 300% โดยเฉพาะอย่างยิ่งเมื่อใช้โมเดล1PL รวมกับ โมเดล 1PPC ค่าสารสนเทศที่เพิ่มมากขึ้นนี้ทำให้เกิดค่าความคลาดเคลื่อนของความสามารถผู้สอบที่ต่ำกว่าความเป็นจริง แต่อย่างไรก็ตามผู้วิจัยได้แนะนำว่า

โมเดล1PL รวมกับ โมเดล 1PPC สามารถนำไปใช้ได้แต่ต้องดำเนินการภายใต้ข้อจำกัดและเงื่อนไขที่กำหนด

เวลล์ สับโกเวียก และเซอร์ลิน (Wells; Subkoviak; & Serlin. 2002) ศึกษาผลของการแปรเปลี่ยนค่าพารามิเตอร์ข้อสอบที่ส่งผลต่อค่าประมาณความสามารถด้วยโมเดล 2 พารามิเตอร์ ศึกษาจากการจำลองข้อมูลและวิเคราะห์จากแบบทดสอบที่มีความยาวสองขนาด กลุ่มตัวอย่างสองขนาด สามประเภทการแปรเปลี่ยนค่าพารามิเตอร์ และ ปริมาณร้อยละของข้อสอบที่เกิดการแปรเปลี่ยนค่าพารามิเตอร์ ผลต่างระหว่างค่าประมาณความสามารถที่แม้จริงคำนวณจากค่าประมาณพารามิเตอร์ที่มีทั้งข้อสอบที่มีการแปรเปลี่ยนค่าพารามิเตอร์และไม่มีการแปรเปลี่ยนค่าพารามิเตอร์ ผลการศึกษาพบว่าจากผลจำลองข้อมูลมีผลต่อค่าประมาณความสามารถแฝงเพียงเล็กน้อยตามเงื่อนไขที่ศึกษา อย่างไรก็ตามจากการศึกษาในครั้งนี้ตั้งอยู่บนเงื่อนไขที่ว่าเป็นการวัดซ้ำเมื่อเวลาผ่านไป 1 ปี ดังนั้นอาจจะเป็นไปได้ว่าร้อยละของการแปรเปลี่ยนและขนาดของการแปรเปลี่ยนอาจจะเหมือนกันกับการที่มีการทดสอบมากกว่า 1 ปี และพบว่าขนาดของกลุ่มตัวอย่างและร้อยละของข้อสอบที่แปรเปลี่ยนนั้นมีผลต่อความสัมพันธ์ระหว่างการแปรเปลี่ยนและค่าประมาณความสามารถแฝง การแปรเปลี่ยนค่าพารามิเตอร์ส่งผลต่อค่าประมาณความสามารถแฝงมากขึ้นเมื่อกลุ่มตัวอย่างเป็น 1000 คน พร้อมด้วยการแปรเปลี่ยนค่าพารามิเตอร์อำนาจจำแนก และ และการแปรเปลี่ยนของค่าพารามิเตอร์อำนาจจำแนกร่วมกับค่าความยาก กับแบบทดสอบที่มีเปอร์เซนต์จำนวนข้อสอบที่แปรเปลี่ยนสูงกว่าค่าพารามิเตอร์ข้อสอบแปรเปลี่ยนมีผลต่อค่าประมาณความสามารถอย่างแตกต่างกันขึ้นอยู่กับประเภทของการแปรเปลี่ยน เมื่อค่าอำนาจจำแนกแปรเปลี่ยน ค่าอำนาจจำแนกร่วมกับค่าความยากแปรเปลี่ยน ค่าประมาณความสามารถแฝงจะถูกส่งผลก็ต่อเมื่อค่าประมาณความสามารถนั้นตรงกับการแจกแจงค่าความสามารถแฝง ในขณะที่ ค่าความยากแปรเปลี่ยน นั้นส่งผลเมื่อ ค่าประมาณความสามารถแฝงมีการแจกแจงแบบเอกรูป

สี (Si. 2002) ได้ศึกษาผลการจำลองข้อมูลเพื่อดูผลของรูปแบบการให้คะแนน การประมาณค่าพารามิเตอร์ข้อสอบ ลักษณะของเทรชโฮล (Threshold) และการแจกแจงความสามารถก่อนเรียนที่มีต่อความแม่นยำของการประมาณค่าพารามิเตอร์ความสามารถผู้สอบด้วยโมเดลการตอบข้อสอบทั้งสิ้น 7 โมเดล ได้แก่ โมเดลโลจิสติก 1 พารามิเตอร์ โมเดลโลจิสติก 2 พารามิเตอร์ โมเดลโลจิสติก 3 พารามิเตอร์ โมเดลนอมินอล โมเดลพาร์เชียลเครดิต โมเดลเจเนอรอลไรส์พาร์เชียลเครดิต และโมเดลแบบเลือกตอบ ผู้วิจัยได้จำลองข้อมูลโดยกำหนดให้ผู้สอบ 1,000 คน ทำแบบทดสอบ 30 ข้อ ซึ่งทราบค่าทั้งพารามิเตอร์ข้อสอบและค่าประมาณความสามารถ ส่วนข้อมูลการตอบข้อสอบนั้นถูกกำหนดให้สามารถประมาณค่าประมาณความสามารถได้ทั้งโมเดลแบบทวิวิภาคและพหุวิภาคทั้งหมด 7 โมเดล ความแม่นยำของการประมาณค่าความสามารถพิจารณาจาก อัตราความครอบคลุม และรากที่สอง

ของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง ผลการศึกษาพบว่า เมื่อนำพิจารณาทุกเงื่อนไขรวมกันทั้งหมด โมเดลการตอบข้อสอบแบบพหุวิภาคให้ค่าประมาณความสามารถที่แม่นยำกว่าโมเดลทวิวิภาค พร้อมกับให้ค่าอัตราความถูกต้องสูงกว่าและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่า โมเดลการประมาณค่าพารามิเตอร์ข้อสอบแบบหนึ่งพารามิเตอร์ ให้ค่าที่แตกต่างไปจากการประมาณค่าด้วยโมเดลสอง และสามพารามิเตอร์ในทุกเงื่อนไข ส่วนโมเดลการตอบข้อสอบแบบพหุวิภาค โมเดลพาร์เชียลเครดิตให้ค่าประมาณที่แม่นยำมากกว่าโมเดลอื่นอีกสามโมเดล ส่วนโมเดลนอมินอลแสดงให้เห็นว่ามีค่าความถูกต้องต่ำสุดเมื่อเทียบกับโมเดลเจเนอรอลไรซ์พาร์เชียลเครดิตโมเดล และ โมเดลแบบเลือกตอบ ผลที่ได้แสดงให้เห็นว่า การแจกแจงความสามารถเริ่มต้นที่แน่นอนมีผลต่อความแม่นยำในการประมาณค่าความสามารถ แต่อย่างไรก็ตาม ไม่สามารถสรุปได้อย่างชัดเจนว่าลำดับของความแม่นยำในแต่ละกลุ่มความสามารถทั้งสี่กลุ่มนั้นเป็นผลมาจากปฏิสัมพันธ์ระหว่างการแจกแจงความสามารถและลักษณะของเทรซโฮล อัตราการกลับสู่ปกติ มีค่าต่ำกว่าเมื่อข้อสอบมีรายการคำตอบที่มีเทรซโฮลที่ไม่เท่ากัน ซึ่งใกล้กับจุดปลายของเส้นต่อเนื่องของความสามารถ หรือของค่าความยากของข้อสอบและถูกจัดให้เป็นตัวแทนกลุ่มประชากรที่ตรงกันกับลักษณะการเบ้ตรงปลายเส้นต่อเนื่องของความสามารถ

คินเซย์ (Kinsey. 2003) ได้ศึกษาผลของความยาวของแบบทดสอบ(10 20 และ 30 ข้อ) รูปแบบการให้คะแนน (สัดส่วนของการให้คะแนนแบบสองค่าต่อการให้คะแนนแบบหลายค่า) และโมเดลการวิเคราะห์ข้อสอบ (IRT และ Rasch) ที่มีต่อค่าประมาณความสามารถ ระดับสารสนเทศของแบบทดสอบ และเกณฑ์ความเหมาะสมของแบบทดสอบที่มีรูปแบบผสม ใช้การจำลองข้อมูลเพื่อศึกษาผลการทำแบบทดสอบที่ให้คะแนนแบบหลายค่า จำนวน 30 ข้อของกลุ่มตัวอย่างจำนวน 1,000 คนโดยใช้โมเดล GPCM และ โปรแกรม SAS แล้วแปลงข้อมูลเป็นคะแนนแบบสองค่า ทั้งหมด 11 รูปแบบตามสัดส่วน จึงได้แบบทดสอบผสมจำนวน 33 ชุด การศึกษาครั้งนี้ใช้โปรแกรม MULTILOG สำหรับการคำนวณค่าประมาณความสามารถ ค่าความคลาดเคลื่อนมาตรฐาน ค่าสารสนเทศของข้อสอบและแบบทดสอบ ค่าความเชื่อมั่น และดัชนีความกลมกลืน (fit) ในการเปรียบเทียบกระบวนการวิเคราะห์ของ IRT และ RASCH นั้นใช้โปรแกรม SPSS เพื่อศึกษา ค่าประมาณความสามารถและความคลาดเคลื่อนของค่าประมาณความสามารถโดยจัดเป็นการวิเคราะห์ความแปรปรวนแบบ 3×11×2 แฟคทอเรียล รายงานค่าขนาดของผลและอำนาจการทดสอบของแต่ละกระบวนการวิเคราะห์ พร้อมทั้งใช้วิธีเปรียบเทียบภายหลังของเชฟเฟ่เพื่อดูความแตกต่างรายคู่ในแต่ละองค์ประกอบ ค่าสารสนเทศของแบบทดสอบถูกวิเคราะห์และเปรียบเทียบในแต่ละช่วงความสามารถรวมทั้งสิ้น 66 รูปแบบ ผลการศึกษาพบว่า ขนาดความยาวของแบบทดสอบและสัดส่วนการให้คะแนนแบบหลายค่ามีผลต่อปริมาณค่าสารสนเทศของแบบทดสอบแบบผสมอย่างมีนัยสำคัญ โดยส่วนใหญ่

แบบทดสอบที่มีการให้คะแนนแบบหลายค่า 100 % นั้นจะให้ค่าสารสนเทศของแบบทดสอบสูงกว่ารูปแบบอื่นทั้งหมด ซึ่งดูจะเป็นจริงพิเศษสำหรับผู้เข้าสอบที่มีค่าประมาณความสามารถที่ต่ำกว่า การเปรียบเทียบความเหมาะสมของแบบทดสอบโดยใช้กระบวนการวิเคราะห์ของ IRT และ Rasch โดยพิจารณาค่าความคลาดเคลื่อนมาตรฐานสำหรับค่าประมาณความสามารถ ความเชื่อมั่นแบบมีขอบเขต (Marginal Reliability) และ ดัชนีความกลมกลืน (-2LL) พบว่ามีเฉพาะค่าความคลาดเคลื่อนมาตรฐานเท่านั้นที่แตกต่างกันอย่างมีนัยสำคัญ ผลที่เกิดขึ้นนี้แสดงให้เห็นถึงความชัดเจนที่ว่าขนาดของผลที่ได้ม่สำคัญนัก ส่วนความเหมาะสมของแบบทดสอบแบบผสมจะมีค่าสูงเมื่อแบบทดสอบมีความยาวและสัดส่วนการให้คะแนนแบบหลายค่ามีค่ามากขึ้น ในดัชนีบางตัวของ IRT ชี้ให้เห็นผลของค่าสารสนเทศของแบบทดสอบที่พัฒนาขึ้นเพียงเล็กน้อย ส่วนการศึกษาความแตกต่างโดยรวมพบว่ากระบวนการวิเคราะห์ข้อสอบทั้ง IRT และ Rasch ไม่มีความแตกต่างกัน และแนะนำว่างานวิจัยครั้งต่อไปควรจะนำบริบทเกี่ยวกับความยากของแบบทดสอบ คะแนนของผู้เข้าสอบ และการให้คะแนนแบบบางส่วนอัตโนมัติมาศึกษาร่วมกับวิธีการวัดแบบดั้งเดิมเพื่อดูผลที่เกิดขึ้นในการวิเคราะห์แบบทดสอบผสม

ยอน (Yon. 2006) ได้ศึกษาขั้นตอนการปรับเทียบตามแนวตั้งสำหรับมาตรวัดแบบพหุมิติด้วย ฟังก์ชันโค้งลักษณะแบบทดสอบ TCF กับ วิธีไม่ตั้งฉาก NOP ใช้ทั้งข้อมูลจริงและข้อมูลจากการจำลอง โดยใช้การร่วมกันสองระดับคือเกรด 7 และ 8 กับ เกรด 6 และ 7 ข้อมูลได้ถูกจำลองสำหรับประมาณค่าด้วยโครงสร้างแบบทดสอบเดี่ยวและผสม รวมถึงการแปลงค่าประมาณอำนาจจำแนกและค่าพารามิเตอร์ความยากของแบบทดสอบเทียบเท่ากับรูปแบบที่เป็นฐาน สำหรับแต่ละคู่ของระดับชั้น เปรียบเทียบความสามารถของวิธีการทั้งสองโดยพิจารณาจากความคลาดเคลื่อนร่วมทั้งค่าอำนาจจำแนกและค่าความยาก โดยเรียกว่าเป็นความลำเอียง ค่าสหสัมพันธ์ระหว่างค่าประมาณจากฐานและแบบที่ปรับเทียบแล้ว และ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ในส่วนของการใช้ข้อมูลจริง จากการวิเคราะห์ความแปรปรวนแบบการวัดซ้ำ (Repeated measure ANOVA) พบว่าโดยทั่วไปข้อมูลจากแบบทดสอบเดี่ยวที่ใช้ร่วมกันให้ความแม่นยำดีกว่า และ วิธีตั้งฉากมีแนวโน้มให้ค่าประมาณพารามิเตอร์อำนาจจำแนกแม่นยำมากกว่าเล็กน้อยตามเกณฑ์ของความสัมพันธ์ และค่า RMSE ส่วนกรณีของค่าความยาก วิธีตั้งฉากให้ค่าประมาณที่ชัดเจนว่ามีความลำเอียงมากกว่าแต่ให้ค่าสหสัมพันธ์สูง แต่ค่า RMSE ต่ำ กรณีที่ใช้ข้อมูลจริง การประมาณค่า NOP มีความคลาดเคลื่อนน้อยกว่า TCF โดยเฉพาะ ค่าประมาณความยากที่แปลงจากการใช้ TCF แสดงให้เห็นว่ามีค่าเฉลี่ยความแตกต่างเป็นค่าบวกสูงกว่าค่าประมาณจากฐานเดิม ทั้งหมด จึงสรุปได้ว่ามาตรวัดแนวตั้งตามทฤษฎีการทดสอบหลายมิติมีความแม่นยำน้อยเมื่อข้อมูลแบบผสมเมื่อเทียบกับข้อมูลเดี่ยว แต่สามารถบรรเทาปัญหาได้โดยใช้วิธีไม่ตั้งฉากแทนวิธี TCF

เหยา และ โบตัน (Yao; & Boughton. 2007) ได้ศึกษาผลการใช้โมเดลการตอบข้อสอบแบบพหุมิติสำหรับแบบทดสอบซึ่งประกอบด้วยข้อสอบแบบทวิวิภาคและพหุวิภาค เพื่อศึกษาการประมาณค่าตามคะแนนสอบย่อยตามวัตถุประสงค์ ด้วยการจำลองข้อมูลใช้วิธีการประมาณค่าแบบ Markov chainMonte Carlo (MCMC) สำหรับข้อมูลแบบพหุมิภาค ตลอดจนการประมาณค่าความสามารถ ในแต่ละวัตถุประสงค์ย่อย และอัตราการจัดประเภท ใช้การจำลองข้อมูลใช้ข้อมูลจริงจากการทดสอบขนาดใหญ่ระดับรัฐ ที่มีคะแนนสอบย่อยภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่างต่างกัน และ ค่าความสัมพันธ์ระหว่างข้อสอบย่อยแต่ละชุดเป็น .0, .1, .3, .5, .7, .9 ตามลำดับ ผลการศึกษาพบว่าการความถูกต้องของการรายงานข้อมูลวินิจฉัยในแต่ละความสามารถย่อย แบบทดสอบย่อยต้องมีค่าความสัมพันธ์สูง

ชอน ลี และ แอนสเลย์ (Chon; Lee; & Ansley. 2007) ได้ศึกษาผลการวิเคราะห์แบบทดสอบผสมด้วยโมเดลการตอบข้อสอบโดยใช้ข้อมูลจริงซึ่งประกอบด้วย ข้อสอบแบบทวิวิภาคและข้อสอบแบบพหุวิภาคมีโมเดลที่สามารถวิเคราะห์ได้ 9 โมเดล และวิธีการประมาณค่า 2 วิธี เปรียบเทียบจากค่าสถิติความสอดคล้องกลมกลืนแบบดั้งเดิม และแบบทางเลือก โมเดลแบบทวิวิภาคสามโมเดล และโมเดล พหุวิภาค 3 โมเดล นำมารวมกันเพื่อวิเคราะห์แบบทดสอบผสมทั้งแบบวิเคราะห์พร้อมกันและแยกจากกันการประเมินความสอดคล้องกลมกลืนใช้ PARSCALE's G^2 และดัชนีสองตัวคือ S - X^2 and S - G^2สถิติซึ่งเสนอโดยออลันโด และทิสเซน (Orlando and Thissen. 2000) ที่สามารถใช้ได้กับแบบทดสอบผสม ผลการศึกษาพบว่า โมเดลโลจิสติกแบบสามพารามิเตอร์รวมกับโมเดลเตเนอรอลไรซ์พาร์เชียลเครดิต ให้ค่าความสอดคล้องกลมกลืนสูงกว่าการรวมด้วยโมเดลอื่น ในขณะที่โมเดลโลจิสติกแบบหนึ่งพารามิเตอร์ให้ค่าความผิดพลาด(Misfit) สูงกว่าโมเดลอื่น และเมื่อเปรียบเทียบระหว่างค่าความสอดคล้องกลมกลืน ผลยังไม่สอดคล้องกันว่าค่าดัชนีตัวใดความสอดคล้องกลมกลืนใดดีที่สุดแต่พบว่าค่าดัชนีความสอดคล้องกลมกลืนแบบใหม่นั้นใช้พิจารณาความสอดคล้องกลมกลืนได้ดีกว่าดัชนีตัวเดิม

ซอง (Seong.1990) ได้ศึกษาความไวของวิธีประมาณค่าพารามิเตอร์ข้อสอบและผู้สอบแบบความน่าจะเป็นสูงสุดเมื่อการแจกแจงความสามารถเดิมของผู้สอบไม่ตรงกับการแจกแจงความสามารถจริงที่ซ่อนอยู่ โดยใช้ข้อสอบ 45 ข้อจำนวน 30 ชุดจากข้อมูลที่กำหนดให้มีการแจกแจงความสามารถจริงสามประเภท วิเคราะห์ด้วยโปรแกรม BILOG การกำหนดความสามารถเริ่มต้นอย่างเหมาะสม เพิ่มความแม่นยำของการประมาณค่าพารามิเตอร์ข้อสอบและผู้สอบเมื่อกลุ่มตัวอย่างมีขนาดใหญ่ แต่ถ้ากลุ่มตัวอย่างมีขนาดเล็ก การกำหนดความสามารถเริ่มต้นอย่างเหมาะสม เพิ่มความแม่นยำของการประมาณค่าพารามิเตอร์ผู้สอบแต่ไม่มีผลต่อการประมาณค่าพารามิเตอร์ผู้สอบมีเพียงข้อมูลขนาดใหญ่ และ สอดคล้องกับความสามารถแฝง และการแจงค่าความสามารถเดิมที่เพิ่มจำนวน

เป็นกำลังสองของความแม่นยำในการประมาณค่าพารามิเตอร์ อย่างไรก็ตาม ความแม่นยำของค่าประมาณ ถูกทำให้เพิ่มขึ้นโดยการเพิ่มจำนวนของจุดกำลังสองโดยไม่ใส่ใจขนาดกลุ่มตัวอย่าง และความเหมาะสมของการแจกแจงความสามารถเดิมจำนวนของผู้สอบมีผลสำคัญต่อความแม่นยำในการประมาณค่าพารามิเตอร์

หยาง (Yang. 2007) ได้ศึกษาความแข็งแกร่งต่อละเมิดความเป็นเอกมิติในการประมาณค่าพารามิเตอร์และความสอดคล้องกลมกลืนของโมเดล ด้วยการใช้โมเดลราชส์แบบเอกมิติและพหุมิติภายใต้เงื่อนไข ระดับความเป็นพหุมิติ ขนาดกลุ่มตัวอย่าง และความยาวแบบทดสอบต่างกัน โดยพิจารณาจากความแม่นยำของการวัด ซึ่งตรวจสอบจากค่า RMSE ความลำเอียง และค่าสหสัมพันธ์แบบเพียร์สัน พิจารณาจากความสอดคล้องกลมกลืนของข้อมูลดูจากดัชนีสองค่าคือ ค่าเฉลี่ยกำลังสองของ Infit และ Outfit และค่าสถิติ t รวมทั้งอัตราความคลาดเคลื่อนประเภทที่หนึ่ง ผลการศึกษาพบว่าความยาวของแบบทดสอบมีอิทธิพลต่อค่าประมาณความสามารถมากที่สุดในขณะที่ขนาดกลุ่มตัวอย่างมีอิทธิพลต่อการประมาณค่าพารามิเตอร์ข้อสอบมากที่สุด ส่วนระดับความเป็นพหุมิตินั้นพบว่า มีรูปแบบของค่าที่ศึกษาที่น่าสนใจโดยเฉพาะเมื่อระดับความเป็นพหุมิติอยู่ในระดับปานกลางหรือระดับสูง โมเดลราชส์แบบพหุมิติจะให้ค่าประมาณความสามารถที่แม่นยำสูงกว่าเมื่อเปรียบเทียบกับโมเดลราสซ์แบบเอกมิติ ในขณะที่แบบทดสอบที่มีระดับความเป็นพหุมิติต่ำ โมเดลเอกมิติได้แสดงว่ามีความแข็งแกร่งต่อการละเมิดข้อตกลงเรื่องความเป็นเอกมิติอย่างไรก็ตาม ในการประมาณค่าความยากของข้อสอบ ระดับความเป็นพหุมิติไม่ได้ส่งผลต่อความแม่นยำในการประมาณค่า ส่วนผลการศึกษาความสอดคล้องกลมกลืนของข้อมูลพบว่าขนาดกลุ่มตัวอย่างส่งผลต่อค่าเฉลี่ยกำลังของความสอดคล้องกลมกลืน แต่ไม่ส่งผลต่อค่าสถิติ t ความยาวแบบทดสอบส่งผลทั้งค่าสถิติทั้งสองในการประมาณค่าความสามารถ โดยทั่วไปค่าเฉลี่ย Infit และค่าสถิติ t มีความคงเส้นคงวากว่าค่า Outfit ส่วนค่าความคลาดเคลื่อนประเภทที่ 1 ของโมเดลราสซ์แบบเอกมิติและพหุมิติให้ผลต่างกันเล็กน้อยเมื่อระดับความเป็นพหุมิติเพิ่มขึ้น อย่างไรก็ตาม ค่าอัตราความคลาดเคลื่อนประเภท 1 ของค่าเฉลี่ยกำลังสองของความสอดคล้องกลมกลืนของการประมาณค่าพารามิเตอร์ข้อสอบมีแนวโน้มลดลงเมื่อเพิ่มขนาดตัวอย่าง และของพารามิเตอร์ผู้สอบลดลงเมื่อเพิ่มขนาดความยาวแบบทดสอบ อัตราความคลาดเคลื่อนของค่าสถิติ t ไม่ได้ขึ้นอยู่กับขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ

โมเดลการตอบข้อสอบสำหรับแบบทดสอบผสมซึ่งมีข้อสอบตั้งแต่สองลักษณะแตกต่างกันในฉบับเดียว สามารถใช้โมเดลการตอบข้อสอบสำหรับข้อสอบทวิวิภาคร่วมกับโมเดลพหุวิภาคในการประมาณค่าพารามิเตอร์ข้อสอบและผู้สอบให้อยู่บนมาตรวัดเดียวกันได้ แต่โมเดลการตอบข้อสอบแบบทวิวภาค และ โมเดลการตอบข้อสอบแบบพหุวิภาคมีหลายโมเดล ดังนั้นการเลือกโมเดลสำหรับประมาณค่าแบบทดสอบผสมจึงต้องพิจารณาตามลักษณะของข้อสอบในแบบทดสอบผสม และความ

สอดคล้องกลมกลืนกับข้อมูลด้วย การประมาณค่าพารามิเตอร์ด้วยโมเดลต่างกันย่อมให้ค่าประมาณพารามิเตอร์แตกต่างกัน และตัวประมาณค่าต่างกันก็ส่งผลให้ค่าประมาณพารามิเตอร์ที่ใช้โมเดลการตอบข้อสอบเหมือนกันแตกต่างกันได้ นอกจากนั้นยังมีปัจจัยที่ส่งผลต่อความแม่นยำในการประมาณค่าพารามิเตอร์คือ ความยาวของแบบทดสอบ ขนาดกลุ่มตัวอย่าง และจำนวนพารามิเตอร์ในโมเดลโลจิสติก นอกจากนี้ยังพบว่าปัจจัยสำคัญที่ส่งผลต่อความแม่นยำในการวัดคะแนนการเปลี่ยนแปลงได้แก่ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนและคะแนนการเปลี่ยนแปลง การประมาณค่าพารามิเตอร์ ความยาวของแบบทดสอบ และขนาดกลุ่มตัวอย่างเป็นต้น ดังนั้นผู้วิจัยจึงสนใจตรวจสอบผลการประมาณค่าพารามิเตอร์พร้อมกันโดยใช้โมเดลผสมที่ประกอบด้วย 3PL และ GPCM ร่วมกัน เมื่อความยาวของแบบทดสอบ ระดับความยากของข้อสอบ และระดับคะแนนของข้อสอบแบบเขียนตอบแตกต่างกัน

3. งานวิจัยเกี่ยวกับการปรับเทียบค่าพารามิเตอร์

อัญชลี ศรีกลชาญ (2552) ศึกษาผลของการปรับเทียบคะแนนตามแนวทฤษฎีการตอบสนองข้อสอบ สำหรับแบบทดสอบที่มีรูปแบบผสมระหว่างการตรวจให้คะแนนแบบสองค่าตามโมเดล IRT โลจิสติกแบบ 3 PL และการตรวจให้คะแนนแบบหลายค่าตามโมเดล GPC โดยมีวัตถุประสงค์ เพื่อวิเคราะห์และเปรียบเทียบคุณภาพการปรับเทียบคะแนนด้วยวิธีโค้งคุณลักษณะ (Characteristic Curve) และวิธีการประมาณค่าพารามิเตอร์พร้อมกัน (Concurrent Calibration) เมื่อแบบแผนการเก็บรวบรวมข้อมูล และสัดส่วนของข้อสอบทั้ง 2 ชนิดในแบบทดสอบรูปแบบผสม และสัดส่วนของข้อสอบทั้ง 2 ชนิดในข้อสอบร่วมของแบบทดสอบรูปแบบผสมต่างกัน ผลการศึกษาพบว่าคุณภาพของการปรับเทียบคะแนนด้วยวิธีโค้งคุณลักษณะและการประมาณค่าพารามิเตอร์พร้อมกัน มีความคลาดเคลื่อนต่ำ และให้ผลที่ใกล้เคียงกันเมื่อมีรูปแบบการเก็บรวบรวมข้อมูลกับกลุ่มผู้สอบที่มีความสามารถเท่าเทียมกัน สำหรับกลุ่มผู้สอบที่มีความสามารถไม่เท่าเทียมกัน วิธีการปรับเทียบด้วยการประมาณค่าพารามิเตอร์พร้อมกันจะมีความคลาดเคลื่อนในการปรับเทียบต่ำกว่าวิธีโค้งคุณลักษณะ และความคลาดเคลื่อนต่ำมีค่าลดลงเมื่อสัดส่วนของข้อสอบทั้งสองชนิดในแบบทดสอบมีจำนวนข้อสอบที่ตรวจให้คะแนนสองค่าเพิ่มขึ้น และเมื่อสัดส่วนของข้อสอบทั้งสองชนิดในข้อสอบร่วมมีจำนวนข้อสอบที่ตรวจให้คะแนนสองค่าลดลง และพบว่าค่าเฉลี่ย MSE ที่ได้จากการปรับเทียบคะแนนสำหรับแบบทดสอบทั้ง 4 สัดส่วนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และค่าเฉลี่ย MSE ที่ได้จากการปรับเทียบคะแนนสำหรับข้อสอบร่วมทั้ง 6 สัดส่วนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกจากนี้ยังพบว่า การปรับเทียบด้วยวิธีโค้งคุณลักษณะที่เก็บรวบรวมข้อมูลแบบกลุ่มไม่เท่าเทียม กับการปรับเทียบด้วยการประมาณค่าพารามิเตอร์พร้อมกันกับกลุ่มผู้สอบไม่เท่าเทียมมีค่าเฉลี่ย MSE สูงกว่ากรณีอื่น

ไวท์แมน และ แชมเพลน (Wightman; & Champlain. 1993) ได้ศึกษาเปรียบเทียบสองวิธีการปรับเทียบตามทฤษฎีการตอบข้อสอบโดยใช้โมเดลโลจิสติกแบบสามพารามิเตอร์กับผลการทดสอบก่อนการทดสอบ GMAT (Graduate Management Admissions Testing Program: GMAT) และการทดสอบ GMAT วิธีแรกเป็นประมาณค่าพารามิเตอร์แบบพร้อมกัน (Concurrent Calibration) ซึ่งประมาณค่าพารามิเตอร์ของการทดสอบทั้งสองครั้งพร้อมกันด้วยโปรแกรม LOGIST วิธีที่สองเป็นการประมาณค่าพารามิเตอร์การทดสอบแบบแยกกันโดยประมาณค่าพารามิเตอร์แต่ละการทดสอบแยกกันโดยกำหนดให้ความสามารถผู้สอบคงที่โดยประมาณค่าผลการทดสอบ GMAT ก่อน แล้วประมาณค่าผลการทดสอบ ผลการศึกษาพบว่า ค่าเฉลี่ยของพารามิเตอร์ความยากที่ได้จากการประมาณค่าแต่ละวิธีมีค่าคล้ายคลึงกันมาก และพบว่าการประมาณค่าแบบแยก ในกรณีที่ค่าพารามิเตอร์ความยากของข้อสอบมีค่าสูง จะให้ค่าสูงกว่าความเป็นจริง (Overestimate)เล็กน้อย และกรณีที่ค่าพารามิเตอร์ความยากของข้อสอบต่ำ จะมีค่าต่ำกว่าความเป็นจริงเล็กน้อย (Underestimated) ส่วนค่าพารามิเตอร์อำนาจจำแนกมีค่าต่ำกว่ามีค่าต่ำกว่าความเป็นจริง (Underestimated) อย่างสม่ำเสมอ สุดท้าย ความชันของสมการถดถอยของความสามารถและคุณลักษณะข้อสอบด้วยการใช้การปรับพารามิเตอร์พร้อมกันมีความชันมากกว่า ผลที่เกิดขึ้นนี้สอดคล้องกับงานวิจัยที่มีมาก่อนของ สต็อกกิ้ง และไอเนอร์ (Stocking; & Eigner. 1986) และ แนะนำว่า ข้อสอบใหม่ที่นำมาใช้สอบร่วมกับข้อสอบที่ผ่านการหาคุณภาพแล้ว ควรจะนำมาปรับเทียบค่าพารามิเตอร์โดยใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับข้อสอบที่ผ่านการทดสอบหาคุณภาพแล้ว

ตัง และ ไอเนอ (Tang; & Eignor. 1997) ศึกษาผลการรวมโมเดลการตอบข้อสอบสำหรับข้อมูลทวิวิภาคและโมเดลพหุวิภาคร่วมกันประมาณค่าผลการทดสอบ TOEFL ข้อสอบTOEFL ทุกด้าน ได้แก่ คำศัพท์ (Vocabulary) และ ความเข้าใจในการอ่าน (Reading Comprehension) และ การทดสอบการเขียนภาษาอังกฤษ (Test of Written English : TWE) และ ข้อสอบ TOEFL ความเข้าใจในการฟัง (Listening Comprehension) และ ข้อสอบการทดสอบการพูดภาษาอังกฤษ (Test of Spoken English: TSE®) ประมาณค่าพารามิเตอร์พร้อมกัน โดยใช้โมเดลเจเนอรอลไรซ์พาร์เชียลเครดิตร่วมกับโมเดลโลจิสติกแบบสามพารามิเตอร์ เปรียบเทียบกับการใช้โมเดลเกรดเรสพอนส์ร่วมกับโมเดลโลจิสติกแบบสามพารามิเตอร์ ผลการศึกษาพบว่า ข้อมูลจาก การรวมการอ่านและการเขียนที่ทำให้เป็นข้อสอบส่วนคำศัพท์และความเข้าใจในการอ่านของข้อสอบ TOEFL และข้อสอบTWE นั้น มีความสอดคล้องกลมกลืนกับโมเดลเจเนอรอลไรซ์พาร์เชียลเครดิตร่วมกับโมเดลโลจิสติกแบบสามพารามิเตอร์ และโมเดลเกรดเรสพอนส์ร่วมกับโมเดลโลจิสติกแบบสามพารามิเตอร์ ในทำนองเดียวกันข้อมูลที่รวมการฟังและการพูดเป็นส่วนความเข้าใจในการฟังของข้อสอบ TOEFL และข้อสอบ TSE นั้น

ก็สอดคล้องกลมกลืนกับโมเดลที่ศึกษาทั้งคู่เช่นกัน และเมื่อเปรียบเทียบระหว่างโมเดลเจเนอรอลไรซ์พาร์เชียลเครดิตกับโมเดลเกรดเรสพอนส์แล้วพบว่าเมื่อใช้ PARSCALE โมเดลเจเนอรอลไรซ์พาร์เชียลเครดิต ให้ผลที่ดีกว่า

ลี กริฟฟิท และ แตม (Li; Griffith; &Tam. 1997) ศึกษาค่าที่เกี่ยวข้องในการใช้รูปแบบการปรับเทียบตามทฤษฎีการตอบข้อสอบ โดยการใช้ข้อสอบร่วมเพียงชุดเดียวในการปรับเทียบค่าพารามิเตอร์แบบกำหนดพารามิเตอร์ข้อสอบร่วม (Fixed common item parameters: FCIP) โดยศึกษาข้อมูลเชิงประจักษ์เกี่ยวกับความเหมาะสมของรูปแบบการปรับเทียบโดยใช้นักเรียน 6 กลุ่ม สำหรับการเก็บรวบรวมกลุ่มตัวอย่าง 8,357 คน เพื่อตรวจสอบการกลับสู่สภาพเดิมของพารามิเตอร์ (Parameter recovery) ที่จะสะท้อนความแข็งแกร่งของวิธี FCIP ภายใต้เงื่อนไขของความคลาดเคลื่อนมาตรฐานของค่าความยากและค่าการเดาขนาดใหญ่ โดยใช้วิธีการปรับเทียบแบบประมาณค่าพารามิเตอร์แยกกันแล้วใช้วิธีโค้งลักษณะ (Characteristics Curve Method : CCM) ในการคำนวณสัมประสิทธิ์การปรับเทียบ ผลการศึกษาพบว่าค่าประมาณความสามารถที่ถูกปรับพารามิเตอร์ด้วยวิธีการนี้มีความสอดคล้องกันมาก ยกเว้นเมื่อนักเรียนมีระดับความสามารถต่ำมาก สำหรับการใช้วิธี CCM ทำให้ค่าพารามิเตอร์ข้อสอบจากการปรับพารามิเตอร์มีความสอดคล้องกันมากยกเว้นค่าพารามิเตอร์การเดา การศึกษาภายใต้วิธี CCM ที่ศึกษาด้วยข้อมูลจำลองนี้พบว่าค่าประมาณพารามิเตอร์มีความแม่นยำและคงเส้นคงวามาก

คิม และ โคเฮน (Kim; & Cohen. 1998) ได้เปรียบเทียบวิธีการปรับเทียบตามทฤษฎีการตอบข้อสอบสามวิธีได้แก่ วิธีการปรับพารามิเตอร์แยกกัน (Separate Calibration) ที่ใช้วิธีการแปลงเชิงเส้นแบบโค้งลักษณะข้อสอบ(Test Characteristic Curve: TCC) วิธีการปรับพารามิเตอร์พร้อมกัน (Concurrent Calibration)บนฐานการประมาณค่าแบบค่าประมาณสูงสุดปลายทาง (Marginal Maximum a Posteriori Estimation: MMAP) และวิธีการปรับพารามิเตอร์พร้อมกัน (Concurrent Calibration)บนฐานการประมาณค่าแบบความน่าจะเป็นสูงสุดปลายทาง (Marginal Maximum Likelihood Estimation: MMAP) ข้อมูลที่ใช้ศึกษาได้มาจากการจำลองข้อมูลของผู้สอบ 500 คน ทำแบบทดสอบ 50 ข้อ จากนั้นจึงสุ่มกลุ่มที่ศึกษากลุ่มละ 50 คนโดยให้มีความสามารถเฉลี่ยเป็น 0 และค่าส่วนเบี่ยงเบนมาตรฐานเป็น 1 จำนวนสองกลุ่ม และกลุ่มผู้สอบมีความสามารถเฉลี่ยเป็น 1 และค่าส่วนเบี่ยงเบนมาตรฐานเป็น 1 และเปรียบเทียบผลตามจำนวนข้อสอบร่วม 5 10 25 และ 50 ข้อ ผลการศึกษาพบว่าเมื่อแบบทดสอบมีข้อสอบร่วมจำนวนน้อย วิธีการปรับพารามิเตอร์แยกกันและใช้การแปลงเชิงเส้นแบบ TCC มี ค่ารากที่สองของค่าเฉลี่ยผลต่างกำลังสอง (Root Mean Square Difference :RMSD) ของค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากต่ำสุด ส่วนเมื่อแบบทดสอบมีข้อสอบร่วมจำนวนมากขึ้น วิธีการปรับเทียบทั้งสามให้ผลไม่ต่างกัน

ลี ลิสสิทซ์ และ หยาง (Li; Lissitz; & Yang.1999) ได้ศึกษาผลการใช้วิธีการประมาณค่าพารามิเตอร์แบบแยกกันแล้วใช้วิธีการแปลงค่าสัมประสิทธิ์การปรับเทียบแบบโค้งลักษณะข้อสอบ (Characteristic Curves Method: CCM) เพื่อ สร้างมาตรวัดร่วม (Common Scaling) ของรูปแบบการปรับเทียบแบบใช้แบบทดสอบร่วมซึ่งเป็นแบบทดสอบผสมระหว่างข้อสอบแบบทวิวิภาคและข้อสอบแบบพหุวิภาค กับกลุ่มผู้สอบที่มีความสามารถแตกต่างกัน ผลการศึกษาพบว่า ในทุกเงื่อนไขที่ศึกษา ค่าความลำเอียงของการกลับสู่สภาพปกติของสัมประสิทธิ์การปรับเทียบ มีค่าเข้าใกล้ศูนย์และค่า root mean squared error (RMSE) RMSE ต่ำ เมื่อนำสัมประสิทธิ์การปรับเทียบไปเขียนกราฟแสดงลักษณะการแจกแจงพบว่ากราฟการแจกแจงเป็นรูปสมมาตรโค้งระฆังคว่ำ ค่าความแปรปรวนต่ำและมีจำนวน Outlier จึงสรุปได้ว่า วิธีThe CCM มีความเหมาะสมสำหรับคำนวณสัมประสิทธิ์การปรับเทียบเมื่อใช้กับแบบทดสอบผสมเช่นเดียวกับการใช้กับแบบทดสอบแบบทวิวิภาคเพียงอย่างเดียวหรือแบบทดสอบพหุวิภาคเพียงอย่างเดียว

เบเกอร์และ อัลคานิ (Baker; & Al-Karni.1991) ได้ศึกษาเพื่อเปรียบเทียบวิธีการคำนวณหาสัมประสิทธิ์การปรับเทียบด้วยวิธีของ ลอย และ ฮูเวอร์ (Loyd: &Hoover. 1980) และวิธีของสต็อกกิ้งและ ลอร์ด (Stocking; &Lord. 1983) วิธีแรกตั้งอยู่บนค่าสถิติผลรวมของค่าพารามิเตอร์แบบทดสอบส่วนวิธีหลังตั้งอยู่บนความเหมาะสมกับโค้งคุณลักษณะข้อสอบโดยใช้ฟังก์ชันกำลังสองของการสูญหายน้อยที่สุด สถานการณ์ที่ทำการศึกษามีสามสถานการณ์ได้แก่ การวัดตามแนวนอน การวัดตามแนวตั้ง และ สถานการณ์ที่มีเป็นปกติในการศึกษาการผลการกลับสู่สภาพเดิม (recovery) ของพารามิเตอร์ตามทฤษฎีการตอบข้อสอบ ผลการศึกษาแสดงว่า วิธีการคำนวณค่าสัมประสิทธิ์การปรับเทียบทั้งสองให้ค่าสัมประสิทธิ์ที่คล้ายคลึงกันทั้งสามสถานการณ์ นอกจากนี้ เมื่อใช้วิธีทั้งสองกระทำกับข้อมูลข้อสอบ SAT สองชุดพบความแตกต่างเล็กน้อย คณะผู้วิจัยแนะนำว่าโดยทั่วไป วิธีของลอย และ ฮูเวอร์ (Loyd: &Hoover. 1980) จะยังคงให้ค่าสัมประสิทธิ์การปรับเทียบที่เหมาะสม ส่วนวิธีของสต็อกกิ้ง และ ลอร์ด (Stocking; &Lord. 1983) ให้ค่าสัมประสิทธิ์สูงกว่า ลอย และ ฮูเวอร์ (Loyd: &Hoover. 1980) และมีความไวต่อคุณลักษณะแบบทดสอบแบบไม่ปกตินิ้อยกว่า เมื่อผู้ใช้มีเหตุผลที่ทำให้เชื่อว่าพารามิเตอร์ของแบบทดสอบอาจจะมีปัญหาต่อการประมาณค่าพารามิเตอร์ ผู้วิจัยแนะนำว่าการใช้วิธีของ สต็อกกิ้ง และ ลอร์ด (Stocking; &Lord. 1983) เหมาะสมกว่า

แฮนสัน และ บีกวิน (Hanson; & Béguin. 2002) ได้ศึกษาเปรียบเทียบคุณสมบัติของการประมาณค่าพารามิเตอร์พร้อมกัน กับการประมาณพารามิเตอร์แยกกันแล้วใช้วิธีคำนวณสัมประสิทธิ์การปรับเทียบแบบวิธี mean/sigma วิธี mean/mean วิธี SL และ วิธี Haebara ของรูปแบบการเก็บรวมรวมข้อมูลแบบใช้แบบทดสอบร่วมกับกลุ่มที่มีความสามารถไม่เท่าเทียมกัน (Common-item nonequivalent) โดยมีเงื่อนไขที่ศึกษา 4 เงื่อนไขได้แก่ (1) โปรแกรมสำหรับใช้ประมาณค่า ได้แก่

โปรแกรม MULTILOG และ โปรแกรม BILOG-MG (2) ขนาดของกลุ่มตัวอย่าง ได้แก่ 3,000 คน และ 1,000คน (3) จำนวนของข้อสอบร่วม ได้แก่ 20 ข้อ และ 10 ข้อ และ(4) กลุ่มความสามารถเท่าเทียมกัน และกลุ่มความสามารถไม่เท่าเทียมกัน ได้แก่ ไม่มีความแตกต่างของค่าเฉลี่ย และความแตกต่างของค่าเฉลี่ยที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 1 การศึกษาครั้งนี้ใช้แบบทดสอบคณิตศาสตร์ ของ ACT (1997) ฉบับ A และ ฉบับ Z แล้วสุ่มกลุ่มผู้สอบให้มีความสามารถเท่าเทียมกันจำนวน 2696 และ 2670 คน แบบทดสอบแต่ละฉบับแบ่งออกเป็นสามส่วนส่วนละ 20 ข้อ โดยที่เนื้อหาและค่าสถิติของข้อสอบแต่ละส่วนใกล้เคียงกันเท่าที่เป็นไปได้ จากนั้นจึงนำส่วนแรกของแบบทดสอบฉบับ A 10 ข้อ และสองส่วนสุดท้ายของแบบทดสอบฉบับ Z 20 ข้อ มาสร้างแบบทดสอบฉบับ B นำค่าประมาณพารามิเตอร์ของข้อสอบของแบบทดสอบฉบับ B ลงบนมาตรวัดของแบบทดสอบฉบับ A โดยใช้ข้อสอบร่วม 20 ข้อ ผลการศึกษาพบว่าเมื่อขนาดกลุ่มตัวอย่างมีขนาดใหญ่ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย (Mean Squared Error: MSE) มีค่าน้อยกว่าขนาดกลุ่มตัวอย่างมีขนาดเล็ก และ เมื่อจำนวนข้อสอบร่วมเป็น 20 ข้อ มีค่า MSE น้อยกว่ามีข้อสอบร่วม 10 ข้อ ส่วนผลการศึกษาโดยทั่วไปพบว่าโปรแกรม MULTILOG และ BILOG-MGให้ผลใกล้เคียงกัน ภายใต้การประมาณค่าพารามิเตอร์แยกกันแล้วใช้การเชื่อมพารามิเตอร์ด้วยวิธี โค้งลักษณะของแบบทดสอบทั้งวิธี ของ SL และ วิธีของ Haebara มีความคลาดเคลื่อนต่ำกว่า วิธีโมเมนต์ (mean/mean และ mean/sigma) นอกจากนี้ พบว่าการประมาณค่าพารามิเตอร์พร้อมกัน มีค่าความคลาดเคลื่อนต่ำกว่าวิธีการประมาณพารามิเตอร์แบบแยก ยกเว้นกรณีที่ใช้โปรแกรม MULTILOG กับกลุ่มความสามารถไม่เท่าเทียมกัน เมื่อศึกษาต่อไปพบว่า ความลำเอียงกำลังสองมีแนวโน้มสูงขึ้นเมื่อระดับคะแนนต่ำลงโดยเฉพาะกลุ่มผู้สอบที่มีความสามารถไม่เท่าเทียมกันคิมและโคเฮน (Kim; & Cohen. 2002) ได้เปรียบเทียบวิธีการปรับเทียบเพื่อพัฒนามาตรวัดร่วมสำหรับโมเดลเกรดเรสพอนส์ตามทฤษฎีการตอบข้อสอบ 2 วิธี วิธีแรก วิธีการปรับพารามิเตอร์แยกกัน (Seperate Calibration) ที่ใช้วิธีการแปลงเชิงเส้นแบบ โค้งลักษณะข้อสอบ (Test Characteristic Curve: TCC) และ วิธีการปรับพารามิเตอร์พร้อมกัน (Concurrent Calibration) ผลการศึกษาพบว่าวิธีการปรับพารามิเตอร์ของการนำข้อมูลกลุ่มเปรียบเทียบและกลุ่มเป้าหมายรวมกัน มีค่ารากที่สองของผลต่างกำลังสองเฉลี่ย (Root Mean Square Difference :RMSD)ของค่าพารามิเตอร์อำนาจจำแนก และค่าพารามิเตอร์ความยากต่ำกว่าเล็กน้อย ในทำนองเดียวกัน วิธีการปรับพารามิเตอร์พร้อมกัน (Concurrent Calibration) ยังคงมีค่ารากที่สองของค่าเฉลี่ยผลต่างกำลังสอง (Root Mean Square Difference :RMSD)ของความสามารถต่ำกว่าการวิธีการปรับพารามิเตอร์แยกกัน (Separate Calibration) ที่ใช้วิธีการแปลงเชิงเส้นแบบ โค้งลักษณะข้อสอบ (Test Characteristic Curve: TCC)

คาร์กี และคณะ (Karkee et al. 2003) ได้นำวิธีการประมาณค่าพารามิเตอร์พร้อมกันแบบรายคู่เพื่อศึกษาความเหมาะสมกับข้อมูลที่เกิดปัญหาความเป็นพหุมิติ การศึกษาใช้ผลการประเมินความสามารถคณิตศาสตร์ CSAP 2002 ของนักเรียนเกรด 5 ถึง 10 ซึ่งมีข้อสอบร่วมอย่างน้อย 20 ข้อของเกรดที่อยู่ติดกัน แล้วสุ่มกลุ่มตัวอย่างนักเรียนเกรดละ 10,000 คน การใช้การประมาณค่าพารามิเตอร์พร้อมกันรายคู่ประมาณค่าพารามิเตอร์พร้อมกันสามลักษณะ คือ จากคู่ของเกรดที่อยู่ชั้นติดกันได้แก่ เกรด 5 กับ 6 เกรด 7 กับ 8 และ เกรด 9 กับ 10 จากนั้นจึงใช้การแปลงเชิงเส้นตามวิธีของสต็อกกิ้งและลอร์ด เพื่อเชื่อมโยงผลการประมาณค่าพารามิเตอร์พร้อมกันรายคู่ทั้งสามเข้าด้วยกันพร้อมกับการประมาณพารามิเตอร์พร้อมกันโดยใช้ เกรด 7 กับ 8 เป็นกลุ่มอ้างอิง นั่นคือ ข้อสอบร่วมเกรด 8 กับ 9 ถูกนำไปใช้เพื่อเชื่อมการประมาณค่าระหว่าง เกรด 9 กับ 10 และ เกรด 7 กับ 8 ในทำนองเดียวกัน ข้อสอบร่วมเกรด 6 และ 7 ใช้เพื่อเชื่อมโยงข้อมูลระหว่างเกรด 5 กับ 6 และ เกรด 7 กับ 8 แล้วศึกษาผลที่ได้จากพัฒนาการระหว่างเกรดต่อเกรดซึ่งวัดจากค่าเฉลี่ยของผลต่างแต่ละเกรด ผลการศึกษาพบว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกัน วิธีประมาณค่าพารามิเตอร์แยกกัน และวิธีการประมาณค่าพารามิเตอร์พร้อมกันเป็นรายคู่ ให้ผลเหมือนกันทั้งหมด ส่วนค่าส่วนเบี่ยงเบนมาตรฐานภายในเกรดเดียวกันมีค่าคงที่อยู่ระหว่าง 64ถึง 74 แต่สำหรับค่า ค่าส่วนเบียงเบนมาตรฐานในเกรด 5 และ 6 ที่ใช้การประมาณค่าพร้อมกันเป็นรายคู่ กลับให้ค่าสูงกว่า คือ 74 และ 72 ตามลำดับเมื่อเปรียบเทียบกับเกรดอื่นที่มีค่าอยู่ระหว่าง 64 และ 68. เมื่อตรวจสอบลักษณะการการแจกแจงแต่ละเกรดแยกกันโดยเปรียบเทียบการแจกแจงความถี่สะสมสำหรับวิธีการประมาณพารามิเตอร์ทั้งสามวิธีทั้ง 6 ระดับ พร้อมกับการคำนวณเปอร์เซนต์ไทล์ที่ 10 25 50 75 และ 90 พบว่ามีการแจกแจงความถี่สะสมความคล้ายคลึงกันทุกวิธี โดยที่คะแนนบนมาตรวัดของแต่ละตำแหน่งเปอร์เซนต์ไทล์มีค่าเหมือนกันทั้งหมด ยกเว้นที่ตำแหน่งเปอร์เซนไทล์อันดับสูง (เปอร์เซนต์ไทล์ที่ 75 และ 90)ของเกรด 9 และ 10 ที่คะแนนบนมาตรจากการประมาณพารามิเตอร์พร้อมกันให้ค่าค่อนข้างแตกต่างออกไป แต่วิธีประมาณค่าพารามิเตอร์พร้อมกับรายคู่ ให้ค่าคะแนนต่ำกว่า มัธยฐานแต่สูงกว่ามัธยฐานในเกรด 5 และ 6 เมื่อเทียบกับวิธีอื่น ผลของคะแนนพัฒนาการระหว่างเกรดและความแปรปรวน รวมถึงการแจกแจงของแต่ละเกรดแยกกัน ยังไม่ได้สนับสนุนว่าการประมาณพารามิเตอร์พร้อมกันเป็นคู่นั้นมีความเหมาะสมกว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันและแยกกันแต่อย่างใด

เปค และ ยัง (Paek; & Young. 2005) ได้ศึกษาผลจำลองข้อมูลแบบทดสอบผสมเพื่อตรวจสอบการกลับคืนสภาพของพัฒนาการความสามารถของนักเรียนโดยใช้วิธีการปรับเทียบพารามิเตอร์แบบกำหนดค่าพารามิเตอร์ข้อสอบร่วมพร้อมกับกำหนดการแจกแจงความสามารถผลสอบก่อนการประมาณค่า (Fixed-Prior Distribution)แบบประมาณค่าซ้ำและไม่ซ้ำ โดยพิจารณาจากค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute error: MAE) ค่ารากที่สองของค่าเฉลี่ยความ

คลาดเคลื่อนกำลังสอง (RMSE) และ ค่าผลรวมของความคลาดเคลื่อนสัมบูรณ์ (Sum of Absolute Error :SAE) ของจำนวนผู้สอบแยกตามระดับความสามารถ (PL) เมื่อกำหนดจุดตัดระดับความสามารถผู้สอบเป็น 4 กลุ่ม ตามตำแหน่งควอไทล์ที่หนึ่ง ตำแหน่งมัธยฐานและ ตำแหน่งควอไทล์ที่สาม ตัวแปรที่ เปค และ ยังศึกษาได้แก่ ระดับพัฒนาการความสามารถที่แตกต่างกันตาม ค่าเฉลี่ยความสามารถที่เพิ่มขึ้นจากการทดสอบครั้งแรก ทั้งหมด 11 ระดับ ได้แก่ -0.5, -0.4, -0.3, -0.2, -0.1, 0, 0.1, 0.2, 0.3, 0.4, 0.5 และค่าส่วนเบี่ยงเบนมาตรฐานของพัฒนาการความสามารถสามระดับคือ 0.8, 1, 1.2 ผลการศึกษาพบว่า การใช้วิธีการปรับเทียบพารามิเตอร์แบบกำหนดค่าพารามิเตอร์ข้อสอบร่วมพร้อมกับกำหนดการแจกแจงความสามารถผลสอบก่อนการประมาณค่า (Fixed-Prior Distribution) ส่งผลต่อการกลับคืนสภาพของพัฒนาการความสามารถเมื่อผู้สอบมีพัฒนาการ โดยพบว่าการปรับการแจกแจงความสามารถใหม่ซ้ำทำให้การกลับคืนสภาพของพัฒนาการความสามารถดีกว่าการประมาณค่าพารามิเตอร์แบบกำหนดพารามิเตอร์ข้อสอบร่วมแบบไม่ทำซ้ำ และเมื่อผู้สอบมีพัฒนาการความสามารถที่มีค่าสัมบูรณ์มากขึ้นการกลับคืนสภาพของพัฒนาการความสามารถมีค่าลดลง พบว่าหากประมาณค่าความสามารถก่อนการประมาณค่าซ้ำส่งผลต่อการกลับคืนสภาพพัฒนาการ แต่หากไม่ประมาณค่าซ้ำ เมื่อส่วนเบี่ยงเบนมาตรฐานของพัฒนาการเป็น 1.2 ทำให้ค่าประมาณความสามารถมีค่าต่ำกว่าความเป็นจริง และเมื่อส่วนเบี่ยงเบนมาตรฐานของพัฒนาการเป็น 0.8 ทำให้ค่าประมาณความสามารถมีค่าสูงกว่าความเป็นจริง และพบว่าค่า MAE และRMSE จากการประมาณค่าความสามารถซ้ำและไม่ซ้ำเมื่อผู้สอบมีพัฒนาการเฉลี่ยเพิ่มขึ้นหรือลดลงมากกว่า 0.1 ส่วนเบี่ยงเบนมาตรฐานเป็น 1 ไม่แตกต่างกัน แต่เมื่อ ส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 หรือ 1.2 การประมาณค่าไม่ซ้ำให้ค่า MAE และRMSE สูงกว่าการประมาณค่าซ้ำทุกระดับพัฒนาการความสามารถ และมีแนวโน้มว่าค่า MAE และRMSE มีค่าสูงขึ้นเมื่อค่าสัมบูรณ์ของพัฒนาการเฉลี่ยมีค่ามากขึ้นจนทำให้กราฟเป็นรูปพาราโบลาหงาย และมีค่าสูงสุดเมื่อส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 ส่วนวิธีการประมาณค่าความสามารถซ้ำใหม่ค่า MAE และRMSE สูงสุดเมื่อส่วนเบี่ยงเบนมาตรฐานของพัฒนาการเป็น 1.2 รองลงมาคือ 1 และ 0.8 ตามลำดับ เพิ่มขึ้นหรือลดลง 0.1 หน่วย น้อยกว่าเมื่อผู้สอบมีพัฒนาการเฉลี่ยเพิ่มขึ้นหรือลดลงมากกว่า 0.1 หน่วย และเมื่อค่าเฉลี่ยความสามารถที่เพิ่มขึ้นมีค่าส่วนเบี่ยงเบนมาตรฐานสูงกว่าส่วนเบี่ยงมาตรฐานของความสามารถเริ่มต้น (1.2) MAE, RMSE สูงกว่า เมื่อค่าเฉลี่ยความสามารถที่เพิ่มขึ้นมีค่าส่วนเบี่ยงเบนมาตรฐานน้อยกว่าหรือเท่ากับส่วนเบี่ยงมาตรฐานของความสามารถเริ่มต้น (0.8, 1) ดังนั้นเมื่อผู้สอบมีพัฒนาการความสามารถ ± 0.1 จำเป็นต้องพิจารณาเพื่อตัดสินใจในการนำพัฒนาการความสามารถจากการใช้วิธีการปรับเทียบพารามิเตอร์แบบกำหนดค่าพารามิเตอร์ข้อสอบร่วมพร้อมกับกำหนดการแจกแจงความสามารถผลสอบก่อนการประมาณค่าไปใช้ต่อไป

คิมและ โคเลน (Kim; & Kolen. 2006) ได้เปรียบเทียบวิธีการปรับเทียบตามทฤษฎีการตอบข้อสอบ 2 วิธี ได้แก่ วิธีการประมาณค่าพารามิเตอร์แยกกันแล้วคำนวณค่าสัมประสิทธิ์การปรับเทียบด้วยวิธีโค้งคุณลักษณะข้อสอบ (characteristic curve methods) ทั้งวิธีของสต็อกกิ้ง - ลอร์ด (Stocking-Lord) และ วิธีของ แฮบารา (Haebara) วิธี Mean/Mean และ วิธี Mean/Sigma กับ วิธีการปรับเทียบพารามิเตอร์แบบพร้อมกัน โดยเน้นการศึกษาไปที่ระดับความแข็งแกร่งต่ออิทธิพลของรูปแบบข้อสอบ (Degree of robustness to format effects: FEs) เมื่อใช้วิธีการปรับเทียบดังกล่าวกับข้อมูลแบบพหุมิติที่เป็นผลจากอิทธิพลแบบทดสอบซึ่งเป็นแบบทดสอบผสม ปริมาณของอิทธิพลของรูปแบบ (FEs) คือค่าสหสัมพันธ์ระหว่างคุณลักษณะเด่นสองลักษณะที่ปรากฏในแบบทดสอบซึ่งเป็นแบบเลือกตอบและเขียนตอบแบบมีโครงสร้างโดยตั้งอยู่บนสมมติฐานในความเป็นไปได้ในการปฏิบัติโดยสมมติให้มี FEs เมื่อแบบทดสอบผสมมีมิติสัมพันธ์กันตามลักษณะข้อสอบ ผลการศึกษาโดยทั่วไปพบว่า การปรับเทียบแบบประมาณค่าพารามิเตอร์พร้อมกันให้ผลดีกว่า การปรับเทียบแบบประมาณค่าพารามิเตอร์แยกกันที่ใช้วิธีการคำนวณสัมประสิทธิการปรับเทียบทั้ง 4 วิธี และมีความแข็งแกร่งต่อการอิทธิพลการเกิด Fes อย่างไรก็ตาม การปรับเทียบแบบประมาณค่าพารามิเตอร์พร้อมกันนี้ให้ผลดีกว่าการเชื่อมต่อมาตรวัดด้วยวิธีโค้งคุณลักษณะข้อสอบเพียงเล็กน้อยเท่านั้น แม้ว่าการปรับพารามิเตอร์พร้อมกันและวิธีการเชื่อมต่อมาตรวัดร่วมแบบโค้งคุณลักษณะข้อสอบจะแสดงถึงความแข็งแกร่งต่อการอิทธิพล Fes ที่ระดับความสัมพันธ์ 0.5 ก็ตาม แต่หลักฐานที่แสดงยังไม่เพียงพอให้สรุปผลในทุกประเภทของแบบทดสอบ

คาร์กี และคณะ (Karkee et al. 2006) ได้ตรวจสอบการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันแบบกำหนดให้ทุกกลุ่มมีการแจกแจงเท่ากันทุกระดับซึ่งอาจจะเกิดปัญหาได้เมื่อมิติของแบบทดสอบเปลี่ยนแปลงตามระดับชั้น ข้อมูลได้มาจากการทดสอบภาษาอังกฤษ 5 ระดับ ซึ่งเป็นแบบทดสอบเลือกตอบ 35 – 38 ข้อ โดยใช้แบบทดสอบภาษาอังกฤษระดับ 3 เป็นกลุ่มอ้างอิง กลุ่มของผู้สอบเกรดอนุบาล ถึง 12 ทำการทดสอบทั้งระดับสูงและต่ำกว่าเกรดของตน ผลการศึกษาแสดงให้เห็นว่า การประมาณพารามิเตอร์พร้อมกันด้วยการแจกแจงแบบเดียวกันให้ค่าเฉลี่ยพารามิเตอร์ b และค่าประมาณความสามารถ θ สูงกว่าสำหรับแบบทดสอบระดับ 1 และให้ค่าเฉลี่ยพารามิเตอร์ b และค่าประมาณความสามารถ θ สูงกว่าสำหรับแบบทดสอบระดับ 5 ผลการใช้การประมาณพารามิเตอร์แยกกันได้ผลเหมือนกันแต่มีรูปแบบที่สุดขั้วมากกว่า ในขณะที่การประมาณพารามิเตอร์พร้อมกันโดยใช้การแจกแจงแยกกันกลับให้ผลตรงกันข้ามคือค่าเฉลี่ยพารามิเตอร์ b และ ค่าประมาณความสามารถ θ ต่ำกว่าสำหรับแบบทดสอบระดับ 5 และให้ค่าเฉลี่ยพารามิเตอร์ b และ ค่าประมาณความสามารถ θ สูงกว่าสำหรับแบบทดสอบระดับ 1 ซึ่งแสดงว่า การประมาณพารามิเตอร์แบบแยกกันและการประมาณพารามิเตอร์พร้อมกันด้วยใช้การแจกแจงแบบเดียวให้พัฒนาการระหว่างเกรด

ต่ำกว่า ดังที่พบได้ว่าค่าเฉลี่ยของผลต่างของระดับที่ติดกัน คณะผู้วิจัยได้อธิบายว่าเนื่องมาจากความเป็นพหุมิติ และได้วิเคราะห์ด้วยข้อมูลที่เหมือนกันและได้รายงานว่ามิติหลายมิติเกิดขึ้นระหว่างเกรดที่ติดกันเท่านั้นไม่ได้เกิดขึ้นภายในระดับเดียวกัน อย่างไรก็ตามไม่มีมีการนำเสนอผลการวิเคราะห์ความเป็นมิติของข้อมูลไว้ ส่วนการศึกษาความสัมพันธ์ระหว่างวิธีการประมาณพารามิเตอร์และความเป็นพหุมิติ การวิเคราะห์องค์ประกอบและการใช้หลายเนื้อหาที่มีมิติต่างกัน มีค่าส่วนเบี่ยงเบนมาตรฐานภายในระดับแสดงให้เห็นแบบรูปการลดลงจากระดับ 1 ไปยังระดับ 3 แต่เพิ่มขึ้นจากระดับ 3 ไประดับ 5 ทุกวิธีการประมาณค่า แต่ ความชันของการลดลงและเพิ่มขึ้นของการประมาณพารามิเตอร์พร้อมกันโดยใช้การแจกแจงแยกกันมีมากกว่า การแจกแจงของแต่ละเกรดวัดจากคะแนนในแต่ละตำแหน่งเปอร์เซนต์ไทล์ที่ต่างกันได้แก่ ตำแหน่งที่ 10 25 50 75 และ 90 พบว่า วิธีการแยกกันประมาณค่าคะแนนมีค่าเพิ่มขึ้นเมื่อระดับเพิ่มขึ้นในทุกตำแหน่งเปอร์เซนต์ไทล์

เมง และคณะ (Meng; et al. 2006) ได้เปรียบเทียบวิธีการประมาณค่าพารามิเตอร์พร้อมกัน วิธีการประมาณค่าพารามิเตอร์แยกกัน และวิธีการประมาณค่าพารามิเตอร์กึ่งพร้อมกัน ภายใต้เงื่อนไขเนื้อหาแตกต่างกัน (ภาษา, การวัดปริมาณ และไม่ใช้ภาษา) โมเดลการตอบข้อสอบ (1PL and 3PL) และ ตัวประมาณค่าความสามารถ (MLE, EAP, และQD) เป็นการศึกษาโดยใช้แบบทดสอบวัดความรู้คิด (Cognitive Abilities Tests: CogAT) 6 ฉบับ (Lohman & Hagen, 2003) โดยใช้รูปแบบการทดสอบแบบใช้ข้อสอบร่วม การประมาณพารามิเตอร์กึ่งพร้อมกัน ประมาณผลการทดสอบ 4 ระดับไปพร้อมกัน ได้แก่ ระดับ A B C และ D ประมาณพารามิเตอร์พร้อมกันเมื่อเทียบกับมาตรวัดของ D และระดับ D E F และ G ประมาณพารามิเตอร์พร้อมกันเมื่อเทียบกับมาตรวัดของ D คณะผู้วิจัยใช้การประมาณพารามิเตอร์กึ่งพร้อมกันเพื่อแก้ไขปัญหาความเป็นพหุมิติ ขณะที่ยังคงใช้ข้อมูลที่เป็นไปได้มากที่สุดจากข้อสอบร่วม อย่างไรก็ตามแม้ว่าจะใช้แบบวัดที่มีเนื้อหาแตกต่างกันคุณลักษณะของแบบทดสอบที่คล้ายกับ CogAT ถูกพิจารณาแล้วว่ามีความเป็นเอกมิติมากกว่าการทดสอบข้ามเกรดหรือข้ามระดับแบบทดสอบมากกว่าแบบทดสอบวัดผลสัมฤทธิ์ ภายใต้ความแปรปรวนภายในระดับเดียวกัน พบว่าโดยทั่วไปการประมาณค่าพารามิเตอร์แบบแยกกัน มีความแปรปรวนที่คงเส้นคงวา ขณะที่การประมาณค่าพารามิเตอร์พร้อมกันและกึ่งพร้อมกันมีรูปแบบลดลงเล็กน้อย เมื่อพิจารณาผลต่างค่าเฉลี่ยและขนาดของผล วิธีการประมาณค่าพารามิเตอร์กึ่งพร้อมกัน ให้ผลที่เกิดจากการรวมกันระหว่างการประมาณพารามิเตอร์แบบแยกและพร้อมกัน เมื่อพิจารณา ระยะห่างตามแนวนอนโดยการวัดความแตกต่างความสามารถที่เปอร์เซนต์ไทล์ต่างกันตำแหน่งที่ 5 10 25 50 75 90 และ 95 สรุปอย่างคร่าวๆได้ว่า วิธีการประมาณค่าพารามิเตอร์แบบแยก ให้ค่าประมาณและรูปแบบที่คล้ายกันในทุกตำแหน่งเปอร์เซนต์ไทล์ ในขณะที่การประมาณค่าพารามิเตอร์พร้อมกันและกึ่งพร้อมกันให้ค่าประมาณและรูปแบบที่แตกต่างในทุกตำแหน่ง

คิม (Kim. 2006) เปรียบเทียบวิธีการปรับเทียบแบบกำหนดค่าพารามิเตอร์ข้อสอบร่วมสามลักษณะได้แก่ (NPU, SPU, and IPU) กับ วิธีการประมาณค่าพารามิเตอร์แบบแยกกัน และวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ในการปรับเทียบตามแนวนอนแบบทดสอบที่ใช้เป็นแบบทดสอบระดับชาติแบบเลือกตอบจำนวน 50 ข้อ สองชุดซึ่งใช้กับผู้สอบสองกลุ่มที่ถูกสุ่มมาให้มีความสามารถใกล้เคียงกัน หรือเท่าเทียมกัน จากนั้นจึงแบ่งข้อสอบออกเป็น 5 ส่วน ส่วนละ 10 ข้อ โดยที่เนื้อหาของคุณลักษณะทางสถิติทั้งห้าส่วนนั้นใกล้เคียงกันเท่าที่จะเป็นไปได้ แบบทดสอบแรก เป็นแบบทดสอบฉบับ Y ซึ่งถูกใช้ทั้งฉบับ ส่วนฉบับ X ถูกสร้างมาใหม่โดยเกิดจากการรวมสองส่วนแรกของแบบทดสอบฉบับ Y กับ สามส่วนหลังของแบบทดสอบ X เดิม เป็น นั่นคือ แบบทดสอบฉบับ X และแบบทดสอบฉบับ Y มีข้อสอบร่วมจำนวน 20 ข้อ จากนั้นจึงสร้างข้อมูลผลการตอบด้วยการกำหนดการแจกแจงของแบบทดสอบ X เป็น N(0,1), N(1,1), และ N(-1,1) ระดับของการกลับสู่สภาพปกติของพารามิเตอร์ประชากรสำหรับข้อสอบฉบับ X ถูกประเมินจาก ค่าความลำเอียง และ ค่า RMSE ของพารามิเตอร์ความยาก บนฐานการทำซ้ำ 100 ครั้ง เนื่องจาก ค่าประมาณพารามิเตอร์ความยากของข้อสอบอยู่บนมาตรวัดเดียวกับพารามิเตอร์ความสามารถจึงให้สารสนเทศโดยตรงเกี่ยวกับการกลับสู่มาตรวัดและพารามิเตอร์ ผลการศึกษาพบว่า วิธีการปรับเทียบแบบกำหนดพารามิเตอร์ข้อสอบสามวิธี เป็นวิธีที่เหมาะสมกับข้อมูลที่มีการแจกแจง N(0,1) และพบว่า SPU และ IPU เหมาะสมกับข้อมูลที่มีการแจกแจง N(-1,1) และN(1,1) ส่วนวิธี The NPU มีแนวโน้มได้ค่าประมาณพารามิเตอร์ความยากสูงกว่าความเป็นจริง (Overestimate) เมื่อข้อมูลมีการแจกแจงN(-1,1) และ ค่าประมาณพารามิเตอร์ความยากต่ำกว่าความเป็นจริง (Underestimate) เมื่อข้อมูลมีการแจกแจงN(1,1) วิธีThe SPU and IPU เป็นวิธีที่เหมาะสมที่สุด ที่ให้ค่าความคลาดเคลื่อนต่ำในการกลับสู่สภาพปกติของค่าพารามิเตอร์ความยากสำหรับข้อสอบที่ไม่ใช่ข้อสอบร่วมให้ผลเช่นเดียวกันกับวิธีการประมาณพารามิเตอร์แบบแยกและพร้อมกัน สำหรับการกลับสู่สภาพปกติของพารามิเตอร์อำนาจจำแนกทั้งสามวิธีมีความเหมาะสม แต่วิธีที่ศึกษามีความคลาดเคลื่อนสูงกว่าวิธีการประมาณพารามิเตอร์แบบแยกและพร้อมกัน

คิม (Kim. 2007) ศึกษาพัฒนาการของนักเรียนระหว่างเกรด ความแปรปรวนระหว่างเกรดและ การแจกแจงความสามารถของแต่ละเกรด โดยใช้วิธีการประมาณค่าพารามิเตอร์ 3 วิธี ได้แก่ วิธีการประมาณค่าพารามิเตอร์พร้อมกัน วิธีการประมาณค่าพารามิเตอร์แบบแยกกัน และวิธีการประมาณค่าพารามิเตอร์แบบกำหนดค่าพารามิเตอร์สำหรับข้อสอบร่วมในกลุ่มข้อมูลใหม่ และเลือกใช้วิธีประมาณค่าพารามิเตอร์ความสามารถ 5 วิธีได้แก่ ตัวประมาณค่าความน่าจะเป็นสูงสุด (Maximum Likelihood Estimator (MLE) จากแบบแผนคะแนนแต่ละข้อ ค่าที่คาดหวังภายหลัง (Expected A Posteriori (EAP) จากแบบแผนคะแนนแต่ละข้อ ตัวประมาณค่าความน่าจะเป็นสูงสุดเทียม (pseudo-MLE) จากคะแนนรวม ค่าที่คาดหวังภายหลังเทียม (Pseudo-EAP) จากคะแนนรวมและ การแจกแจง

กำลังสอง (Quadrature Distribution: QD) การศึกษาใช้ข้อมูลจากการทดสอบนักเรียนเกรด 3 ถึง 8 ด้วยแบบทดสอบทักษะเบื้องต้นของไอโอวา (the Iowa Tests of Basic Skills: ITBS) ที่ประกอบด้วย แบบทดสอบคำศัพท์ (Vocabulary) แบบทดสอบการแก้ปัญหาคณิตศาสตร์และการแปลความหมายข้อมูล (Math Problem Solving and Data Interpretation: MPD) แบบทดสอบความสามารถในการอ่าน (Reading Comprehension: RC) และ แบบทดสอบวิทยาศาสตร์ ผลการศึกษาพบว่า แบบทดสอบความสามารถในการอ่าน (RC) และ แบบทดสอบวิทยาศาสตร์ วิธีการประมาณค่าพารามิเตอร์พร้อมกันทำให้ได้พัฒนาการน้อยกว่าและพัฒนาการลดลงอย่างช้าๆตามระดับชั้นที่ต่ำลง ค่าความแปรปรวนน้อยกว่าในทุกระดับ และในระดับชั้นต่ำลงเกิดการแบ่งแยกน้อยกว่า เมื่อเปรียบเทียบกับวิธีการประมาณค่าพารามิเตอร์แบบแยกกัน และวิธีการประมาณค่าพารามิเตอร์แบบกำหนดค่าพารามิเตอร์สำหรับข้อสอบร่วมในกลุ่มข้อมูลใหม่ ส่วนแบบทดสอบคำศัพท์ (Vocabulary) และแบบทดสอบการแก้ปัญหาคณิตศาสตร์และการแปลความหมายข้อมูล (MPD) พัฒนาการระหว่างเกรด ความแปรปรวนระหว่างเกรด และ การแบ่งแยกการแจกแจงมีความแตกต่างเล็กน้อย

เมื่อเปรียบเทียบตัวประมาณค่าพารามิเตอร์ความสามารถ ในทุกเนื้อหาแบบทดสอบ มีแนวโน้มว่า ค่า SD ของแต่ละเกรด ตัวประมาณค่า pseudo-MLE $\geq$ MLE > QD > EAP $\geq$ pseudo-EAP ส่วน ขนาดของผล (Effect size) พบว่าตัวประมาณค่า pseudo-EAP $\geq$ EAP > QD > MLE $\geq$ pseudo-MLE อย่างไรก็ตาม ทุกตัวประมาณค่าความสามารถ ระดับของการลดลงของความแปรปรวนในแต่ละเกรดมีความคล้ายคลึงกัน เมื่อเปรียบเทียบเนื้อหา แบบทดสอบ Vocabulary และ MPD เปรียบเทียบกับแบบทดสอบความสามารถในการอ่าน และ แบบทดสอบวิทยาศาสตร์ แล้วพบว่ามีพัฒนาการต่ำกว่าแต่มั่นคงและความแปรปรวนในแต่ละเกรดมีค่าน้อยกว่า ส่วนการแบ่งแยกระดับการแจกแจงพบว่าพัฒนาการสูงกว่า โดยพบว่ามีความแตกต่างค่าเฉลี่ยสูงกว่า สำหรับแบบทดสอบความสามารถในการอ่าน และ แบบทดสอบวิทยาศาสตร์เมื่อใช้ขนาดของผลเป็นผลต่างมาตรฐาน จะได้พัฒนาการที่ลดลง

ฮู โรเจอร์ส และ วูคมิโรวิก (Hu; Rogers; & Vukmirovic. 2008) เปรียบเทียบความสามารถของวิธีการปรับเทียบตามทฤษฎีการตอบข้อสอบ 4 วิธีที่ข้อสอบร่วมมีการแปรเปลี่ยนพารามิเตอร์ความยาก 10 ค่า วิธีการปรับเทียบที่ศึกษาได้แก่วิธีการปรับพารามิเตอร์พร้อมกัน (Concurrent Calibration) การปรับพารามิเตอร์แยกกัน (Seperate Calibration) ที่ใช้วิธีการแปลงเชิงเส้นแบบ โค้งลักษณะข้อสอบ(Test Characteristic Curve: TCC) และวิธี ค่าเฉลี่ยหารผลรวม (Mean/Sigma: M/S) และการปรับพารามิเตอร์แบบกำหนดให้พารามิเตอร์ข้อสอบร่วมมีค่าคงที่ (Fixed Common Item Parameter: FCIP) โดยที่ไม่นำข้อสอบที่เป็น Outlier มาพิจารณาด้วย การศึกษาใช้ข้อมูลจำลอง ผลการศึกษาพบว่าเมื่อไม่มี Outlier วิธีการปรับเทียบที่ใช้โค้งลักษณะข้อสอบ และวิธี ค่าเฉลี่ยหารผลรวมให้ผลดี

ที่สุด แต่เมื่อพบ Outlier วิธีการปรับเทียบที่ศึกษาทุกวิธีที่ถ่วงน้ำหนัก Outlier (ยกเว้นวิธีที่ใช้การแปลงแบบ M/S ที่ถ่วงน้ำหนัก Outlier) ให้ผลที่ดีกว่าวิธีการปรับเทียบโดยไม่นำค่า Outlier มาพิจารณา

ไซมอน (Simon. 2008) ศึกษาจำลองข้อมูลโดยใช้โมเดลการตอบข้อสอบแบบพหุมิติประมาณค่าพารามิเตอร์ข้อสอบเพื่อเปรียบเทียบคุณสมบัติของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน กับ วิธีประมาณค่าพารามิเตอร์แยกกันที่ใช้วิธีการคำนวณสัมประสิทธิ์การปรับเทียบที่ต่างกัน 4 วิธี โดยศึกษาภายใต้เงื่อนไข ขนาดของกลุ่มตัวอย่าง (500 1000 และ 3000 คน) ความยาวของแบบทดสอบแบบให้คะแนนสองค่า (40 ข้อ และ 60 ข้อ) ความเท่าเทียมกันของความสามารถ (ความสามารถเท่ากัน และความสามารถต่างกัน .5 และ 1โดยที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 1) และระดับความสัมพันธ์ระหว่างความสามารถ 2 มิติ พิจารณาคุณภาพการปรับเทียบจากค่า RMSE และค่าความลำเอียง (Bias) ผลการศึกษาโดยทั่วไปพบว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีคุณสมบัติดีกว่าวิธีการประมาณค่าพารามิเตอร์แยกกัน แม้ว่าผู้สอบมีระดับความสามารถเฉลี่ยแตกต่างกัน 0.5 และระดับความสัมพันธ์ของความสามารถ 2 มิติมีค่าสูงวิธีการประมาณค่าพร้อมกันมีคุณสมบัติที่ดีกว่าเนื่องจากมีกลุ่มตัวอย่างขนาดใหญ่ โดยเฉพาะเมื่อแบบทดสอบมีขนาด 40 ข้อ เมื่อพิจารณาเฉพาะวิธีการคำนวณสัมประสิทธิ์การปรับเทียบ วิธีโค้งลักษณะข้อสอบ มีแนวโน้มดีกว่า วิธีอื่น เมื่อกลุ่มตัวอย่างมีความสามารถไม่เท่าเทียมกัน โดยที่วิธี Min's method ไม่ดีเท่ากับวิธีอื่น แต่เมื่อกลุ่มผู้สอบมีความสามารถเท่าเทียมกัน วิธีการคำนวณสัมประสิทธิ์การปรับเทียบเชิงเส้นไม่ต่างกัน

จากการศึกษางานวิจัยที่เกี่ยวกับการปรับเทียบค่าพารามิเตอร์จะเห็นว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ให้ค่าประมาณพารามิเตอร์ที่มีความคลาดเคลื่อนต่ำกว่าวิธีการประมาณพารามิเตอร์แยกกัน และสามารถใช้ประมาณค่าพารามิเตอร์สำหรับข้อมูลแบบทดสอบผสมเมื่อผู้สอบมีความสามารถไม่เท่าเทียมกัน (Nonequivalent group) ได้ วิธีการประมาณค่าพารามิเตอร์พร้อมกันเป็นวิธีการปรับเทียบที่สามารถขจัดปัญหาความไม่เท่าเทียมกัน เนื่องจากข้อมูลทั้งหมดถูกนำมาใช้ประโยชน์ ในขณะที่วิธีการปรับค่าพารามิเตอร์แยกกันไม่สามารถทำได้ ซึ่งผลการศึกษาวิจัยที่ผ่านมาแม้ว่าส่วนใหญ่วิธีการประมาณค่าพารามิเตอร์พร้อมกันจะมีคุณสมบัติที่ดีกว่าวิธีการปรับเทียบอื่น แต่ในกรณีที่ผู้สอบสองกลุ่มมีความสามารถสามารถไม่เท่าเทียมกันพบว่าวิธีการประมาณค่าพารามิเตอร์มีความคลาดเคลื่อนสูง ซึ่งหากจะนำวิธีการประมาณค่าพารามิเตอร์พร้อมกันไปใช้ในการวัดพัฒนาการความสามารถที่ความสามารถของผู้สอบในการสอบสองครั้งไม่เท่าเทียมกัน การศึกษาความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันจึงเป็นสิ่งจำเป็น ดังนั้นผู้วิจัยจึงสนใจตรวจสอบความแม่นยำวิธีการประมาณค่าพารามิเตอร์พร้อมกันสำหรับใช้ร่วมในการวัดพัฒนาการความสามารถของผู้สอบเมื่อลักษณะของแบบทดสอบผสมมีขนาดความยาว ระดับความยาก และการให้คะแนนแตกต่างกัน

บทที่ 3
วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นศึกษาภายใต้สถานการณ์จำลอง (Simulation Data) มีจุดมุ่งหมายเพื่อตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถแตกต่างกัน โดยใช้โมเดลโลจิสติกสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตในการประมาณค่าพารามิเตอร์ความสามารถผลการตอบแบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน โดยมีขั้นตอนดังนี้

1. การกำหนดข้อมูลจำลอง
2. การจัดกระทำและการวิเคราะห์ข้อมูล

การกำหนดข้อมูลจำลอง

ข้อมูลที่ศึกษาเป็นข้อมูลจำลองแบบแผนการตอบข้อสอบในแบบทดสอบผสมด้วยโปรแกรม WINGEN2 (Han. 2007) ของผู้สอบจำนวน 1,000 คน ที่ได้ทดสอบซ้ำสองครั้ง โดยใช้โมเดลโลจิสติกแบบสามพารามิเตอร์ในการจำลองผลการตอบข้อสอบแบบเลือกตอบที่ตรวจให้คะแนนแบบสองค่า และใช้โมเดลเจเนอรอลไรส์พาร์เชียลเครดิตจำลองผลการตอบข้อสอบแบบเขียนตอบที่ตรวจให้คะแนนแบบหลายค่า มีรายละเอียดการจำลองแบบแผนการตอบ และการกำหนดค่าพารามิเตอร์ของผู้สอบและข้อสอบ ดังนี้

1. การกำหนดระดับพัฒนาการความสามารถของผู้สอบ

ผลต่างของค่าเฉลี่ยระดับความสามารถระหว่างกลุ่ม 0.50 จัดว่ามีความแตกต่างมากพอจะเป็นผลให้เกิดความแตกต่าง (Li; & Lissitz. 2000) และพบในหลายงานวิจัยที่ศึกษาเกี่ยวกับวิธีการสร้างมาตรวัดร่วมด้วยข้อมูลจำลอง (Davey et al.1996; Kim. 2004; Li; & Lissitz, 2000; Min. 2003; Oshima et al. 2000; Skaggs & Lissitz. 1988) ในการศึกษาครั้งนี้จึงกำหนดความแตกต่างของระดับพัฒนาการความสามารถของการสอบโดยพิจารณาจากค่าเฉลี่ยความสามารถของผู้สอบที่เพิ่มขึ้นเป็น 3 ระดับ คือ พัฒนาการความสามารถเพิ่มขึ้นเล็กน้อย (0.1) พัฒนาการความสามารถเพิ่มขึ้นมาก (0.5) และพัฒนาการความสามารถเพิ่มขึ้นมากที่สุด (1.0) โดยที่ค่าเฉลี่ยความสามารถมีลักษณะการกระจายเป็น 3 ระดับได้แก่ เพิ่มขึ้นเท่ากันทุกคน (SD = 1) ผู้มีระดับความสามารถต่ำมีพัฒนาการสูงกว่าผู้มีระดับความสามารถสูง (SD = 0.8) ผู้มีระดับความสามารถสูงมีพัฒนาการสูงกว่าผู้มีระดับความสามารถต่ำ (SD = 1.2) ผู้วิจัยได้กำหนดระดับความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และครั้งที่ 2 ดังนี้

1.1 การทดสอบครั้งที่ 1 กำหนดให้ความสามารถ (θ_1) ของผู้สอบจำนวน 1,000 คน มีการแจกแจงปกติ โดยมีค่าเฉลี่ยความสามารถเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1

1.2 การทดสอบครั้งที่ 2 กำหนดให้ความสามารถ (θ_2) ของผู้สอบจำนวน 1,000 คน มีความสามารถเฉลี่ยเพิ่มขึ้นจากความสามารถในการทดสอบครั้งแรกเป็น 0.1 0.5 และ 1 ตามลำดับ และมีค่าส่วนเบี่ยงเบนมาตรฐานเป็น 0.8, 1, และ 1.2 ตามลำดับ โดยที่ความสามารถในการทดสอบครั้งที่สอง (θ_2) เขียนให้อยู่ในรูปสมการเชิงเส้นของค่าความสามารถในการทดสอบครั้งที่ 1(θ_1) ดังนั้นในการศึกษาครั้งนี้จึงประกอบด้วยกลุ่มผู้สอบซึ่งมีลักษณะความสามารถแตกต่างกันทั้งหมด 9 กลุ่ม ดังนี้

รูปแบบที่ 1: $\theta_2 = \theta_1 + 0.1$

รูปแบบที่ 2: $\theta_2 = \theta_1 + 0.5$

รูปแบบที่ 3: $\theta_2 = \theta_1 + 1.0$

รูปแบบที่ 4: $\theta_2 = 0.8\theta_1 + 0.1$

รูปแบบที่ 5: $\theta_2 = 0.8\theta_1 + 0.5$

รูปแบบที่ 6: $\theta_2 = 0.8\theta_1 + 1.0$

รูปแบบที่ 7: $\theta_2 = 1.2\theta_1 + 0.1$

รูปแบบที่ 8: $\theta_2 = 1.2\theta_1 + 0.5$

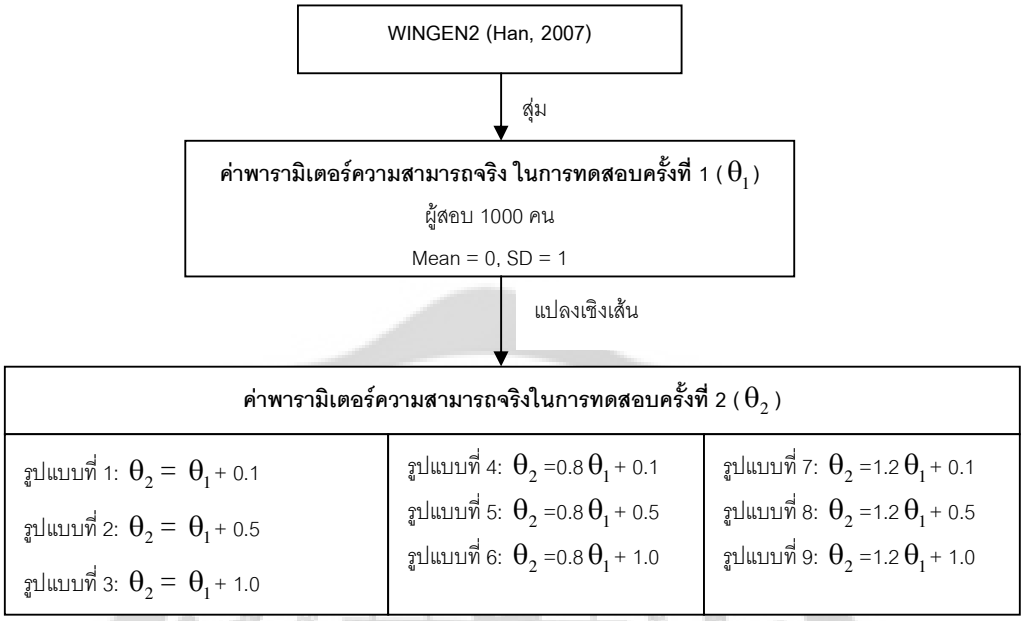
รูปแบบที่ 9: $\theta_2 = 1.2\theta_1 + 1.0$

สรุปลักษณะผู้สอบแยกตามระดับพัฒนาการความสามารถที่ใช้ศึกษาดังตาราง 3 และแสดงแผนภาพขั้นตอนการกำหนดพารามิเตอร์ความสามารถผู้สอบในแต่ละครั้งที่สอบได้ดังภาพประกอบ 5

ตาราง 3 สรุปลักษณะความสามารถของผู้สอบ 1,000 คนในการทดสอบซ้ำสองครั้ง ตามระดับพัฒนาการความสามารถที่เพิ่มขึ้น

ระดับพัฒนาการความสามารถเพิ่มขึ้น	ความสามารถในการทดสอบครั้งที่ 1		ความสามารถในการทดสอบครั้งที่ 2	
	Mean	SD	Mean	SD
น้อย (0.1)	0.00	1.00	0.10	0.80
	0.00	1.00	0.10	1.00
	0.00	1.00	0.10	1.20
มาก (0.5)	0.00	1.00	0.50	0.80
	0.00	1.00	0.50	1.00
	0.00	1.00	0.50	1.20
มากที่สุด (1.0)	0.00	1.00	1.00	0.80
	0.00	1.00	1.00	1.00
	0.00	1.00	1.00	1.20

การกำหนดพารามิเตอร์ผู้สอบ



ภาพประกอบ 5 ขั้นตอนการกำหนดพารามิเตอร์ผู้สอบ

2. **การกำหนดค่าพารามิเตอร์ของข้อสอบในแบบทดสอบผสม**

ในการศึกษาครั้งนี้เป็นการศึกษาพัฒนาการความสามารถของผู้สอบจำนวน 1,000 คนที่ทำการทดสอบซ้ำสองครั้งด้วยแบบทดสอบผสมซึ่งประกอบด้วยข้อสอบแบบเลือกตอบที่ให้คะแนนสองค่าและข้อสอบแบบเขียนตอบที่ให้คะแนนหลายค่า โดยแบบทดสอบผสมที่ใช้ทดสอบครั้งที่สองมีความยากสูงกว่าแบบทดสอบผสมที่ใช้ทดสอบครั้งที่สอง แต่มีความยาว และระดับการให้คะแนนเท่ากัน

การวิจัยครั้งนี้ศึกษาเฉพาะพัฒนาการความสามารถที่เกิดจากการทดสอบซ้ำสองครั้งโดยใช้แบบทดสอบผสมที่มีความยาว และระดับการให้คะแนนเท่ากัน แต่ความยากของแบบทดสอบในการทดสอบครั้งที่ 2 สูงกว่า ความยากของแบบทดสอบในการทดสอบครั้งที่ 1 ค่าพารามิเตอร์ความยากของข้อสอบทั้งสองรูปแบบในแบบทดสอบผสมทั้งสองฉบับมีการแจกแจงแบบปกติ และแบบทดสอบผสมทั้งสองฉบับมีข้อสอบร่วมที่ประกอบด้วยข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบประมาณ 33% ของจำนวนข้อสอบทั้งฉบับ ซึ่งคิมและโคเลน(Kim; & Kolen. 2004) แนะนำว่าจำนวนข้อสอบร่วมที่เหมาะสมควรมีประมาณ 20% ของจำนวนข้อสอบทั้งหมดโดยที่ค่าพารามิเตอร์ความยากของข้อสอบร่วมของแบบทดสอบทั้งสองฉบับอยู่ระหว่างค่าพารามิเตอร์ความยากของแบบทดสอบฉบับ

ที่ 1 และฉบับที่ 2 รายละเอียดการกำหนดค่าพารามิเตอร์ข้อสอบของแบบทดสอบผสมฉบับที่ 1 และฉบับที่ 2 ทั้งหมด 27 เงื่อนไขเป็นดังนี้

2.1 การกำหนดพารามิเตอร์ข้อสอบในแบบทดสอบผสมฉบับที่ 1 และฉบับที่ 2 มีความยากเฉลี่ยต่างกัน 0.50 และมีระดับการให้คะแนนข้อสอบเขียนตอบ 3 ระดับ

2.1.1 กรณีแบบทดสอบผสมฉบับแรกมีระดับความยากเฉลี่ยเป็น -0.50

ผู้วิจัยสุ่มค่าพารามิเตอร์ของข้อสอบแบบให้คะแนนสองค่าด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์จำนวน 50 ข้อ โดยกำหนดให้ข้อสอบแต่ละข้อมีค่าอำนาจจำแนกอยู่ระหว่าง 0.80 ถึง 1.20 และมีความยากเฉลี่ยเป็น -0.50 ที่แจกแจงแบบปกติ และค่าพารามิเตอร์การเดาของข้อสอบมีค่าระหว่าง 0.01 – 0.25 และสุ่มค่าพารามิเตอร์ของข้อสอบแบบให้คะแนนหลายค่าด้วยโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตที่มีคะแนน 3 ระดับ จำนวน 16 ข้อ โดยกำหนดให้แต่ละข้อมีค่าอำนาจจำแนกอยู่ระหว่าง 0.80 ถึง 1.20 และมีความยากเฉลี่ยเป็น -0.50 ที่แจกแจงแบบปกติ

2.1.1.1 การสร้างแบบทดสอบผสมขนาด 30:10

ในการคัดเลือกข้อสอบแต่ละประเภทสำหรับสร้างแบบทดสอบผสมขนาด 30:10 จำนวน 2 ฉบับ ผู้วิจัยดำเนินการดังนี้

การคัดเลือกข้อสอบแบบเลือกตอบ

ผู้วิจัยเรียงข้อสอบ 50 ข้อตามระดับความยาก แล้วแบ่งข้อสอบเป็น 2 ชุดที่มีค่าเฉลี่ยความยากเท่ากัน แล้วเลือกข้อสอบสำหรับใช้เป็นข้อสอบร่วม ชุดละ 5 ข้อ รวมเป็น 10 ข้อ ดังนั้น ข้อสอบ 50 ข้อจึงถูกแบ่งเป็น 3 กลุ่มได้แก่ ข้อสอบกลุ่ม A จำนวน 20 ข้อ สำหรับเป็นข้อสอบแบบเลือกตอบเฉพาะในการทดสอบครั้งที่ 1 ข้อสอบกลุ่ม B จำนวน 10 ข้อ สำหรับเป็นข้อสอบแบบเลือกตอบที่ใช้เป็นข้อสอบร่วมในการทดสอบทั้งครั้งที่ 1 และครั้งที่ 2 และข้อสอบที่เหลืออีก 20 ข้อเป็นข้อสอบกลุ่ม C สำหรับใช้เป็นข้อสอบเฉพาะในการทดสอบครั้งที่ 2 โดยข้อสอบกลุ่ม C นี้ในแต่ละข้อจะถูกเพิ่มความยากจากเดิมอีกข้อละ 0.50

ดังนั้นแบบทดสอบผสมในการทดสอบครั้งที่ 1 จึงประกอบด้วยข้อสอบเลือกตอบกลุ่ม A จำนวน 20 ข้อ และกลุ่ม B จำนวน10 ข้อและแบบทดสอบผสมในการทดสอบครั้งที่ 2 จึงประกอบด้วยข้อสอบเลือกตอบกลุ่ม C จำนวน 20 ข้อ และกลุ่ม B จำนวน 10 ข้อ

การคัดเลือกข้อสอบแบบเขียนตอบ

ผู้วิจัยเรียงข้อสอบ 16 ข้อตามระดับความยาก แล้วแบ่งข้อสอบเป็น 2 ชุดที่มีค่าเฉลี่ยความยากเท่ากัน แล้วเลือกข้อสอบสำหรับใช้เป็นข้อสอบชุดละ 4 ข้อ ดังนั้นข้อสอบแบบเขียนตอบ 16 ข้อจึงถูกแบ่งเป็น 3 กลุ่มได้แก่ ข้อสอบกลุ่ม A จำนวน 6 ข้อ สำหรับเป็นข้อสอบแบบเขียนตอบเฉพาะในการทดสอบครั้งที่ 1 ข้อสอบกลุ่ม B จำนวน 4 ข้อ สำหรับเป็นข้อสอบแบบเขียนตอบที่ใช้เป็นข้อสอบร่วมในการทดสอบทั้งครั้งที่ 1 และครั้งที่ 2 และข้อสอบที่เหลืออีก 6 ข้อเป็นข้อสอบกลุ่ม C

สำหรับใช้เป็นข้อสอบเฉพาะในการทดสอบครั้งที่ 2 โดยข้อสอบกลุ่ม C นี้ในแต่ละขั้นรายการคำตอบจะเพิ่มความยากจากเดิมขั้นละ 0.5

ดังนั้นแบบทดสอบผสมในการทดสอบครั้งที่ 1 จึงประกอบด้วยข้อสอบเขียนตอบกลุ่ม A จำนวน 6 ข้อ และกลุ่ม B จำนวน 4 ข้อและแบบทดสอบผสมในการทดสอบครั้งที่ 2 จึงประกอบด้วยข้อสอบแบบเขียนตอบกลุ่ม C จำนวน 6 ข้อ และกลุ่ม B จำนวน 4 ข้อ

การรวมข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบเป็นแบบทดสอบผสม

แบบทดสอบผสมสำหรับการทดสอบครั้งที่ 1 ประกอบด้วยข้อสอบแบบเลือกตอบเฉพาะครั้งที่ 1จำนวน 20 ข้อ และข้อสอบแบบเลือกตอบที่เป็นข้อสอบร่วมจำนวน 10 ข้อ และข้อสอบแบบเขียนตอบเฉพาะครั้งที่ 1จำนวน 6 ข้อ และข้อสอบแบบเขียนตอบที่เป็นข้อสอบร่วม 4 ข้อ

แบบทดสอบผสมสำหรับการทดสอบครั้งที่ 2 ประกอบด้วยข้อสอบแบบเลือกตอบเฉพาะครั้งที่ 1 จำนวน 20 ข้อ และข้อสอบแบบเลือกตอบที่เป็นข้อสอบร่วม 10 ข้อ และข้อสอบแบบเขียนตอบเฉพาะครั้งที่ 1 จำนวน 6 ข้อ และข้อสอบแบบเขียนตอบที่เป็นข้อสอบร่วม 4 ข้อ

ดังนั้นจะได้ค่าพารามิเตอร์ของข้อสอบแบบทดสอบผสมสำหรับการทดสอบสองครั้งสำหรับศึกษา ภายใต้เงื่อนไขความยาว 30:10 ความยาก -0.5 และมีการให้คะแนน 3 ระดับ

2.1.1.2 **การสร้างแบบทดสอบผสมขนาด 24: 8**

แบบทดสอบผสมขนาด 24:8 คือแบบทดสอบผสมที่ประกอบด้วยข้อสอบที่มีข้อสอบแบบเลือกตอบที่ตรวจให้คะแนนแบบสองค่า 24 ข้อ และข้อสอบแบบเขียนตอบที่ตรวจให้คะแนนแบบหลายค่า 8 ข้อ โดยที่ข้อสอบแบบเลือกตอบ 24 ข้อ ได้มาจากการสุ่มข้อสอบแบบเลือกตอบที่เป็นข้อสอบเฉพาะจากแบบทดสอบผสมขนาด 30:10 ที่ใช้ทดสอบครั้งที่ 1 จำนวน 16 ข้อ และ จากข้อสอบร่วมจำนวน 8 ข้อ ส่วนข้อสอบแบบเขียนตอบ 8 ข้อ ได้มาจากการสุ่มข้อสอบแบบเขียนตอบที่เป็นข้อสอบเฉพาะจากแบบทดสอบผสมขนาด 30:10 ที่ใช้ทดสอบครั้งที่ 2 จำนวน 5 ข้อ และข้อสอบร่วมจำนวน 3 ข้อ โดยแบบทดสอบผสมฉบับใหม่ต้องมีความยากเฉลี่ยเท่ากับหรือใกล้เคียงกับความยากเฉลี่ยของแบบทดสอบผสมขนาด 30:10

2.1.1.3 **การสร้างแบบทดสอบผสมขนาด 15: 5**

แบบทดสอบผสมขนาด 15:5 คือแบบทดสอบผสมที่ประกอบด้วยข้อสอบที่มีข้อสอบแบบเลือกตอบที่ตรวจให้คะแนนแบบสองค่า 15 ข้อ และข้อสอบแบบเขียนตอบที่ตรวจให้คะแนนหลายค่า 5 ข้อ โดยที่ข้อสอบแบบเลือกตอบ 15 ข้อ ได้มาจากการสุ่มข้อสอบข้อสอบแบบเลือกตอบที่เป็นข้อสอบเฉพาะจากแบบทดสอบผสมขนาด 24:8 ที่ใช้ทดสอบครั้งที่ 1 จำนวน 10 ข้อ และ ข้อสอบร่วมจำนวน 5 ข้อ ส่วนข้อสอบแบบเขียนตอบ 5 ข้อ ได้มาจากการสุ่มข้อสอบแบบเขียนตอบที่เป็นข้อสอบเฉพาะจากแบบทดสอบผสมขนาด 24:8 จำนวน 3 ข้อ และข้อสอบร่วมจำนวน 2 ข้อ

2.1.2 กรณีแบบทดสอบฉบับแรกมีระดับความยากเฉลี่ยเป็น 0 และ 0.5

ดำเนินการเช่นเดียวกันกับข้อ 2.1.1 โดยกำหนดให้ความยากเฉลี่ยเป็น 0 และ 0.5 ตามลำดับ

2.2 แบบทดสอบผสมฉบับที่ 1 และฉบับที่ 2 มีความยากเฉลี่ยต่างกัน 0.5 มีระดับการให้คะแนนเป็น 4 ระดับ และ 5 ระดับ

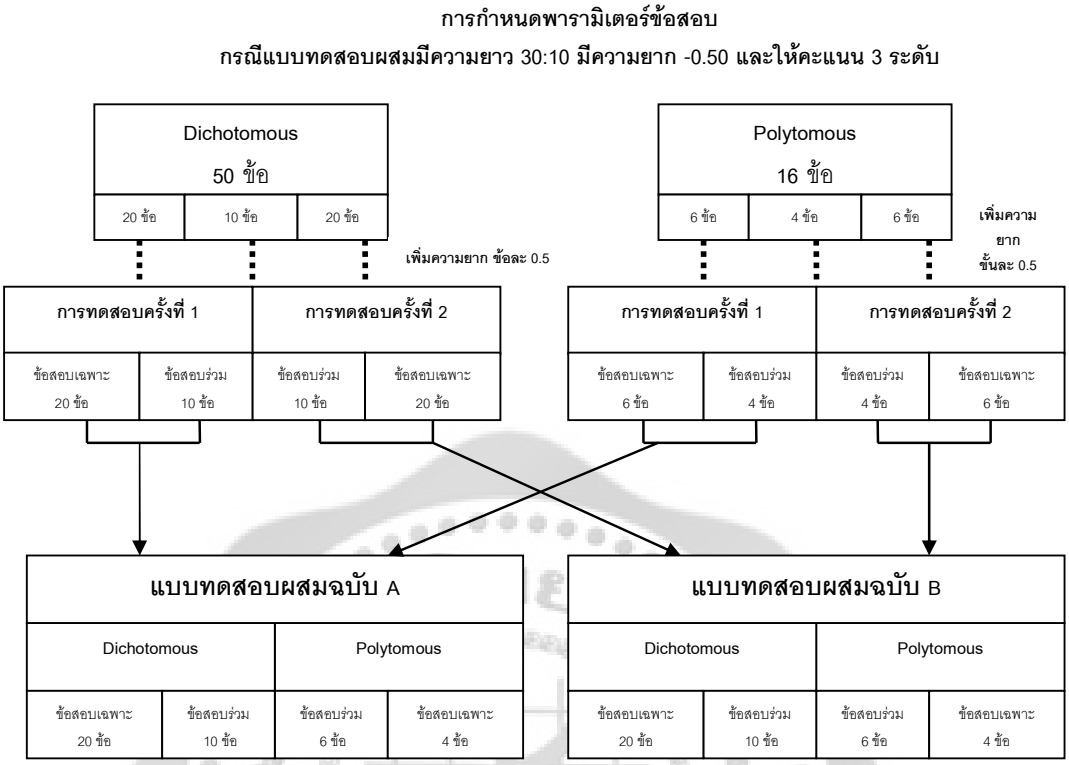
ดำเนินการเช่นเดียวกันกับข้อ 2.1

ดังนั้นแบบทดสอบผสมตามเงื่อนไขที่ต้องการศึกษาที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน มีทั้งสิ้น 27 ฉบับ (3×3×3 = 27) แสดงดังตาราง 4

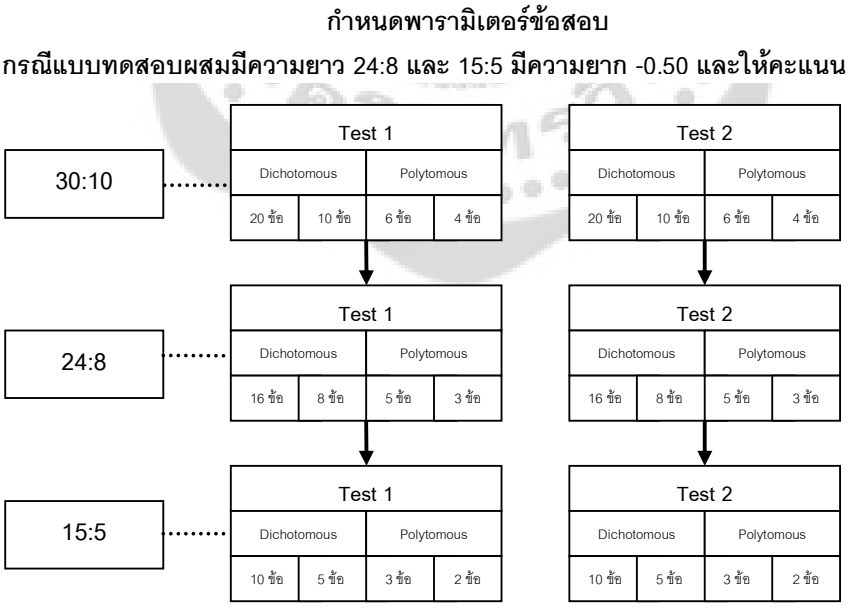
ตาราง 4 สรุปลักษณะข้อสอบแบบทดสอบผสมที่ศึกษา เมื่อขนาดความยาว ระดับความยาก และระดับการให้คะแนนแตกต่างกัน

ระดับการให้คะแนน	ระดับความยากข้อสอบเมื่อเทียบกับระดับความสามารถเริ่มต้น	แบบทดสอบฉบับที่ 1			แบบทดสอบฉบับที่ 2		
		30:10	24:8	15:5	30:10	24:8	15:5
3 ระดับ (0 - 2)	น้อยกว่า	-0.50	-0.50	-0.50	0.00	0.00	0.00
	เท่ากับ	0.00	0.00	0.00	0.50	0.50	0.50
	มากกว่า	0.50	0.50	0.50	1.00	1.00	1.00
4 ระดับ (0 - 3)	น้อยกว่า	-0.5	-0.5	-0.5	0.00	0.00	0.00
	เท่ากับ	0.00	0.00	0.00	0.50	0.50	0.50
	มากกว่า	0.50	0.50	0.50	1.00	1.00	1.00
5 ระดับ (0 - 4)	น้อยกว่า	-0.50	-0.50	-0.50	0.00	0.00	0.00
	เท่ากับ	0.00	0.00	0.00	0.50	0.50	0.50
	มากกว่า	0.50	0.50	0.50	1.00	1.00	1.00

ขั้นตอนการกำหนดพารามิเตอร์ข้อสอบกรณีแบบทดสอบผสมมีความยาว 30:10 24:8 และ 15:5 มีความยาก -0.50 และให้คะแนน 3 ระดับ แสดงดังภาพประกอบ 6 และ 7



ภาพประกอบ 6 ขั้นตอนการกำหนดพารามิเตอร์ข้อสอบกรณีแบบทดสอบผสมมีความยาว 30:10 มีความยาก -0.50 และให้คะแนน 3 ระดับ



ภาพประกอบ 7 การกำหนดพารามิเตอร์ข้อสอบกรณีแบบทดสอบผสมมีความยาว 24:8 และ 15:5 มีความยาก -0.50 และให้คะแนน 3 ระดับ

3. การจำลองแบบแผนการตอบ

การจำลองแบบแผนการตอบเป็นการจำลองผลการทำข้อสอบผสมด้วยโปรแกรม WINGEN2 (Han. 2007) โดยต้องกำหนดระดับความสามารถของผู้สอบ และค่าพารามิเตอร์ของข้อสอบเพื่อให้โปรแกรมจำลองผลการตอบข้อสอบ ดังนั้นการจำลองแบบแผนการตอบข้อสอบสำหรับจำลองข้อมลในการศึกษาครั้งนี้จึงกำหนดค่าพารามิเตอร์ความสามารถผู้สอบตามที่กำหนดไว้ในข้อ 1 และค่าพารามิเตอร์ของข้อสอบตามที่กำหนดไว้ข้อ 2 มีรายละเอียดการกำหนดค่าพารามิเตอร์เพื่อจำลองแบบแผนการตอบด้วยโปรแกรม WINGEN2 (Han. 2007) ดังนี้

3.1 การจำลองแบบแผนการตอบภายใต้เงื่อนไขแบบทดสอบผสมมีความยาว 30:10 มีความยาก -0.50 และให้คะแนน 3 ระดับ

3.3.1 การจำลองแบบแผนการตอบข้อสอบแบบทดสอบผสมฉบับที่ 1

กำหนดค่าพารามิเตอร์ความสามารถของผู้สอบในการทำข้อสอบให้มีค่าเฉลี่ยเป็น 0 ส่วนเบี่ยงเบนมาตรฐานเป็น 1 และกำหนดค่าพารามิเตอร์ของข้อสอบแบบเลือกตอบ 30 ข้อและข้อสอบแบบเขียนตอบ 10 ข้อตามที่กำหนดไว้ในข้อ 2 ของการกำหนดค่าพารามิเตอร์ข้อสอบ (ดูรายละเอียดในภาคผนวก ก) แล้วจึงสั่งให้โปรแกรม WINGEN2 (Han, 2007) จำลองผลการทำข้อสอบทั้งสองลักษณะพร้อมกัน จะได้ผลการคะแนนการทำข้อสอบแบบเลือกตอบ 30 ข้อ (ได้ 0 หรือ 1 คะแนน) และผลการทำข้อสอบแบบเขียนตอบที่ให้คะแนน 3 ระดับ (ได้ 0 1 หรือ 2 คะแนน)

3.3.2 การจำลองแบบแผนการตอบข้อสอบแบบทดสอบผสมฉบับที่ 2

การจำลองแบบแผนการทำข้อสอบในครั้งที่ 2 เป็นจำลองผลการทำข้อสอบตามค่าพารามิเตอร์ความสามารถของผู้สอบที่กำหนดซึ่งมีทั้งหมด 9 กลุ่ม ตามระดับพัฒนาการที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่แตกต่างกัน ดังนี้

กรณีผู้สอบมีพัฒนาการเพิ่มขึ้น 0.1 และมีส่วนเบี่ยงเบนมาตรฐานเป็น 1

กำหนดค่าพารามิเตอร์ความสามารถของผู้สอบในการทำข้อสอบให้มีค่าเฉลี่ยเป็น 0.1 ส่วนเบี่ยงเบนมาตรฐานเป็น 1 (ตามที่กำหนดไว้ในข้อ 1 ของการกำหนดค่าพารามิเตอร์ความสามารถของผู้สอบ) และกำหนดค่าพารามิเตอร์ของข้อสอบแบบเลือกตอบ 30 ข้อและข้อสอบแบบเขียนตอบ 10 ข้อสำหรับใช้ทดสอบในครั้งที่ 2 ตามที่กำหนดไว้ในข้อ 2 (การกำหนดค่าพารามิเตอร์ข้อสอบดูรายละเอียดที่ภาคผนวก ก) แล้วจึงสั่งให้โปรแกรม WINGEN2 (Han, 2007) จำลองผลการทำข้อสอบทั้งสองลักษณะพร้อมกัน จะได้ผลการคะแนนการทำแบบทดสอบผสมขนาด 30:10 ที่ประกอบด้วยข้อสอบแบบเลือกตอบจำนวน 30 ข้อ (ได้ 0 หรือ 1 คะแนน) และผลการทำข้อสอบแบบเขียนตอบที่ให้คะแนน 3 ระดับ (ได้ 0 1 หรือ 2 คะแนน) จำนวน 10 ข้อ

กรณีผู้สอบมีพัฒนาการเพิ่มขึ้นรูปแบบอื่นที่เหลืออีก 8 กลุ่มดำเนินการเช่นเดียวกันจนได้ผลการตอบข้อสอบแบบทดสอบผสมในการทดสอบครั้งที่ 2 จำนวน 9×27 = 243 ชุด โดยกำหนดให้โปรแกรมจำลองข้อมูลแบบแผนการตอบข้อสอบซ้ำตามเงื่อนไขที่ศึกษา เงื่อนไขละ 50 ชุด

3.3.3 การรวมแบบแผนการตอบข้อสอบแบบทดสอบผสมฉบับที่ 1 และแบบทดสอบผสมฉบับที่ 2

ในการประมาณค่าพารามิเตอร์พร้อมกันจะประมาณค่าความสามารถในการทำแบบทดสอบผสมครั้งที่ 1 และครั้งที่ 2 พร้อมกันด้วยการประมวลผลด้วยโปรแกรมครั้งเดียว ดังนั้นเมื่อได้ผลการทำข้อสอบครั้งที่ 1 และครั้งที่ 2 ภายใต้เงื่อนไขเดียวกัน และตรวจสอบความถูกต้องของข้อมูลที่จำลองได้แล้ว จึงนำแบบแผนการตอบข้อสอบที่เป็นผลคะแนนการทำแบบทดสอบผสมสองครั้งรวมเข้าเป็นผลการตอบข้อสอบชุดเดียวกัน โดยเพิ่มจำนวนข้อสอบจากเดิมที่มี เท่ากับจำนวนข้อสอบเฉพาะในแบบทดสอบผสมแต่ละฉบับ เช่น ในการสอบใช้แบบทดสอบผสมที่มีความยาว 30:10 ซึ่งมีข้อสอบเฉพาะที่ใช้ในแต่ละฉบับเป็นข้อสอบแบบเลือกตอบ 20 ข้อ และข้อสอบแบบเขียนตอบ 6 ข้อ ดังนั้นเมื่อรวมผลการตอบข้อสอบสองฉบับเข้าด้วยกันจะได้เป็นผลการทำข้อสอบชุดใหม่ทั้งหมด 66 ข้อ โดยที่กำหนดให้คะแนนข้อสอบที่ผู้สอบไม่ได้ทำได้คะแนนเป็น 0 ทุกข้อ ตัวอย่างการรวมแบบแผนการตอบข้อสอบเมื่อแบบทดสอบผสมมีความยาว 30:10 เป็นดังนี้

ตัวอย่างการรวมแบบแผนการตอบข้อสอบเมื่อแบบทดสอบผสมมีความยาว 30:10

แบบทดสอบผสมมีความยาว 30:10 คือแบบทดสอบที่มีข้อสอบทั้งหมด 40 ข้อ ที่ประกอบด้วยข้อสอบแบบเลือกตอบ 30 ข้อ และข้อสอบแบบเขียนตอบ 10 ข้อ การรวมแบบแผนการตอบแบบทดสอบผสมขนาด 30:10 ในการทดสอบครั้งที่ 1 ของผู้สอบ 1,000 คน และแบบแผนการตอบแบบทดสอบผสมขนาด 30:10 ในการทดสอบครั้งที่ 2 ของผู้สอบ 1,000 คนเดิม เข้าเป็นผลการตอบข้อสอบชุดเดียวกันจึงเป็นการแสดงผลการทำข้อสอบ 66 ข้อของผู้สอบ 2,000 คน ข้อสอบ 66 ข้อ ซึ่งประกอบด้วยข้อสอบเฉพาะสำหรับการทดสอบแต่ละครั้งเป็นข้อสอบเลือกตอบ 20 ข้อ และ ข้อสอบแบบเขียนตอบ 6 ข้อ และมีข้อสอบร่วมที่ผู้สอบ 2,000 ทำครบทุกคนเป็นข้อสอบเลือกตอบ 10 ข้อ และข้อสอบแบบเขียนตอบ 4 ข้อ

ผลการตอบข้อสอบ 66 ข้อมีรายละเอียดดังนี้

ผลการตอบของผู้สอบ 1,000 คนแรก

ข้อ 1 – 20 เป็นข้อสอบแบบเลือกตอบและข้อ 21 – 26 เป็นข้อสอบแบบเขียนตอบ ที่ได้มาจากผลการตอบข้อสอบเฉพาะในการทดสอบครั้งที่ 1

ข้อ 27 – 36 เป็นข้อสอบเลือกตอบ และ ข้อ 37 – 42 เป็นข้อสอบเขียนตอบที่ได้จากผลการตอบข้อสอบร่วมในการทดสอบครั้งที่ 1

ข้อ 43 – 62 เป็นข้อสอบเลือกตอบ และ ข้อ 63 – 66 เป็นข้อสอบเขียนตอบ ซึ่งหมายถึงผลการทำข้อสอบเฉพาะในการทดสอบครั้งที่ 2 แต่ผู้สอบ 1000 คนแรกไม่ได้ทำ ดังนั้นจึงต้องใส่คะแนนข้อ 43 – 66 เป็น 0

ผลการตอบข้อสอบของผู้สอบ 1000 คนหลัง (IDตั้งแต่ 1001 – 2000)

ข้อ 1 - 20 เป็นข้อสอบแบบเลือกตอบและข้อ 21- 26 เป็นข้อสอบแบบเขียนตอบ ที่ได้มาจากผลการตอบข้อสอบเฉพาะในการทดสอบครั้งที่ 1 ซึ่งผู้สอบ 1000 คนหลังไม่ได้ทำ ดังนั้นจึงต้องใส่คะแนนข้อ 1 – 26 เป็น 0

ข้อ 27 – 36 เป็นข้อสอบเลือกตอบ และ ข้อ 37 – 42 เป็นข้อสอบเขียนตอบที่ได้จากผลการตอบข้อสอบร่วมในการทดสอบครั้งที่ 2

ข้อ 43 – 62 เป็นข้อสอบเลือกตอบ และ ข้อ 63 – 66 เป็นข้อสอบเขียนตอบ ได้มาจากผลการทำข้อสอบเฉพาะในการทดสอบครั้งที่ 2

กรณีการรวมแบบแผนการตอบข้อสอบเมื่อแบบทดสอบผสมมีความยาว 24:8 และ 15:5

ดำเนินการในทำนองเดียวกับกรณีการรวมแบบแผนการตอบข้อสอบเมื่อแบบทดสอบผสมมีความยาว 30:10 โดยที่

แบบทดสอบผสมความยาว 24:8 สองฉบับเมื่อรวมเป็นผลการตอบชุดเดียวกันมีข้อสอบทั้งหมด 53 ข้อ โดยมีข้อสอบร่วมที่เป็นข้อสอบเลือกตอบ 8 ข้อและข้อสอบเขียนตอบ 3 ข้อ และข้อสอบเฉพาะสำหรับสอบแต่ละครั้งเป็นข้อสอบเลือกตอบ 16 ข้อและข้อสอบเขียนตอบ 5 ข้อ

แบบทดสอบผสมความยาว 15:5 สองฉบับเมื่อรวมเป็นผลการตอบชุดเดียวกันมีข้อสอบทั้งหมด 33 ข้อ โดยมีข้อสอบร่วมที่เป็นข้อสอบเลือกตอบ 5 ข้อและข้อสอบเขียนตอบ 2 ข้อ และข้อสอบเฉพาะสำหรับสอบแต่ละครั้งเป็นข้อสอบเลือกตอบ 10 ข้อและข้อสอบเขียนตอบ 3 ข้อ

สรุปจำนวนแบบแผนการตอบสำหรับประมาณค่าพารามิเตอร์พร้อมกัน

เนื่องจากการศึกษาในครั้งนี้กำหนดให้ระดับความสามารถในการทำแบบทดสอบผสมครั้งที่ 1 ของผู้สอบ 1,000 คน มีการแจกแจงปกติ โดยมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 แบบทดสอบผสมฉบับที่ 1 ที่ใช้ทดสอบมีลักษณะแตกต่างกันตามเงื่อนไขที่ศึกษาจำนวน 27 ฉบับ ดังนั้นได้แบบแผนการตอบข้อสอบฉบับที่ 1 ทั้งสิ้น 27 ชุด (27×1=27)

สำหรับการทดสอบครั้งที่ 2 นั้นกำหนดให้ผู้สอบมีระดับความสามารถในการทำแบบทดสอบเพิ่มขึ้นจากการสอบครั้งที่ 1 แตกต่างกันจำนวน 9 ระดับ เมื่อแต่ละกลุ่มทำแบบทดสอบ

ผสมที่มีลักษณะต่างกันตามเงื่อนไขที่ศึกษาจำนวน 27 ฉบับ จึงได้แบบแผนการตอบแบบทดสอบผสมฉบับที่ 2 ทั้งสิ้น 243 ชุด (27×9= 243)

ดังนั้นจึงเมื่อรวมแบบแผนการตอบข้อสอบแบบทดสอบผสมทั้งสองครั้งเข้าเป็นผลการตอบชุดเดียวกันจึงมีแบบแผนการตอบแบบทดสอบจำนวนทั้งสิ้น 243 กรณีสำหรับนำไปประมาณค่าพารามิเตอร์ความสามารถผู้สอบโดยใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกัน (Concurrent Calibration) ด้วยโมเดลผสมระหว่างโมเดลโลจิสติกแบบสามพารามิเตอร์กับโมเดลเจเนอรอลไรส์-พาร์เชียลเครดิต

การจัดกระทำและการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลในการวิจัย

เมื่อได้แบบแผนการตอบข้อสอบรายข้อของผู้สอบ 1,000 คนจากแบบทดสอบผสมทั้งสองฉบับที่มีลักษณะแตกต่างกันตามเงื่อนไขที่ศึกษาแล้ว ผู้วิจัยตรวจสอบค่าพารามิเตอร์ของข้อสอบและผู้สอบที่ได้จากโปรแกรม WINGEN2 (Han. 2007) อีกครั้งด้วยโปรแกรม MULTILOGเพื่อยืนยันว่าแบบแผนการตอบที่ได้มีค่าพารามิเตอร์ความสามารถและพารามิเตอร์ข้อสอบเป็นไปตามเงื่อนไขที่กำหนดแล้วจึงนำแบบแผนการตอบข้อสอบที่ได้มาวิเคราะห์ผลการศึกษา ดังรายละเอียดต่อไปนี้

1. **การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ**

การวิเคราะห์สถิติพื้นฐานของข้อมูลจากสถานการณ์จำลองมีขั้นตอนดังนี้

1.1ผู้วิจัยนำข้อมูลรายการคำตอบรายข้อของผู้สอบแต่ละคนจากสถานการณ์จำลองซึ่งเป็นผลการตอบข้อสอบซ้ำสองครั้งของผู้สอบ 1000 คน ที่มีพัฒนาการความสามารถเพิ่มขึ้น 9 ระดับ ซึ่งได้จากการทำแบบทดสอบแบบผสมที่มีความยาว 3 ขนาด ซึ่งกำหนดตามสัดส่วนข้อสอบแบบเลือกตอบ และจำนวนข้อสอบแบบเขียนตอบเป็น 30:10 24:8 และ 15:5 โดยระดับค่าเฉลี่ยความยากของแบบทดสอบผสม 3 ลักษณะได้แก่ ต่ำกว่าเท่ากับ และสูงกว่า ระดับความสามารถเริ่มต้น คือมีค่าเฉลี่ยเป็น -0.50 0.00 และ 0.50 ตามลำดับ ทั้งนี้กำหนดให้การให้คะแนนข้อสอบแบบเขียนตอบมี 3 รูปแบบ คือให้คะแนน 3 ระดับ (0/1/2 คะแนน) ให้คะแนน 4 ระดับ (0/1/2/3 คะแนน) และ ให้คะแนน 5 ระดับ (0/1/2/3/4 คะแนน) รวมทั้งสิ้น 243 กรณี วิเคราะห์สถิติพื้นฐานของคะแนนดิบประกอบด้วย คะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด คะแนนต่ำสุด พิสัย ความคลาดเคลื่อนมาตรฐาน และลักษณะการแจกแจงข้อมูล

1.2 ตรวจสอบความความถูกต้องของข้อมูลจากสถานการณ์จำลอง โดยนำแบบแผนการตอบข้อสอบแต่ละข้อจากการตอบของผู้สอบแต่ละคนที่ได้จากการทดสอบแต่ละครั้งไปวิเคราะห์โดยใช้โปรแกรม MULTILOG พิจารณาค่าพารามิเตอร์ข้อสอบ ค่าอำนาจจำแนก (a) ค่าความ

ยาก (b) และค่าความน่าจะเป็นในการเดาคำตอบถูก (c) ในโมเดลโลจิสติกแบบสามพารามิเตอร์ และค่าพารามิเตอร์ความชันของข้อคำถาม (α) ในโมเดลเจเนอรอลไรซ์พาร์เชียลเครดิต (GPCM) แล้วเปรียบเทียบกับพารามิเตอร์ข้อสอบจริงที่กำหนดในโปรแกรม WINGEN2 (Han. 2007) สำหรับการจำลองแบบแผนการตอบ จากนั้นจึงทดสอบความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ข้อสอบที่ได้จากโปรแกรมโดยใช้ t – test สำหรับทดสอบกลุ่มตัวอย่าง 2 กลุ่มที่ไม่อิสระจากกัน (Two dependent samples t-test)

1.3 การประมาณค่าพารามิเตอร์ผู้สอบด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ผู้วิจัยใช้โมเดลโลจิสติกแบบสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรซ์พาร์เชียลเครดิตประมาณค่าพารามิเตอร์ความสามารถจากแบบแผนการตอบในแต่ละเงื่อนไขที่รวมผลการตอบสองครั้งเข้าด้วยกันแล้วโดยใช้โปรแกรม MULTILOG จากนั้นจึงนำค่าประมาณพารามิเตอร์ความสามารถที่ประมาณค่าได้วิเคราะห์สถิติพื้นฐานประกอบด้วย ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด ค่าต่ำสุด พิสัย ความคลาดเคลื่อนมาตรฐาน และลักษณะการแจกแจงข้อมูล

2. การศึกษาความสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถ กับความสามารถจริง

การศึกษาความสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถ กับความสามารถจริง เป็นการศึกษาความสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถในการทดสอบครั้ง 1 กับความสามารถจริงครั้งที่ 1 (ค่าที่กำหนดในการจำลองข้อมูล) และความสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถในการทดสอบครั้ง 2 กับความสามารถจริงครั้งที่ 2 (ค่าที่กำหนดในการจำลองข้อมูล) โดยมีลำดับดังนี้

2.1 การศึกษาความสัมพันธ์ระหว่างความสามารถจริง กับค่าประมาณความสามารถที่ประมาณค่าจากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ภายใต้เงื่อนไขเดียวกันทุกเงื่อนไข เป็นการการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

2.2 การเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างความสามารถจริง กับค่าประมาณความสามารถที่ประมาณค่าจากวิธีการประมาณค่าพารามิเตอร์พร้อมกันภายใต้เงื่อนไขเดียวกัน เป็นการศึกษาเปรียบเทียบโดยใช้คะแนนมาตรฐาน (Z) ของฟิชเชอร์ โดยนำค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันมาแปลงเป็นคะแนนมาตรฐาน (Z) ของฟิชเชอร์ แล้วเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ที่ได้ตามเงื่อนไขโดยใช้การวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (Repeated Measure)

3. การศึกษาความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

การศึกษาความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ในการทดสอบด้วยแบบทดสอบผสมที่มีขนาดความยาว ระดับความยากและระดับการให้คะแนนแตกต่างกัน พิจารณาจากความลำเอียงและความเชื่อมั่นของค่าประมาณพารามิเตอร์ความสามารถที่ประมาณค่าจากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อเทียบกับค่าความสามารถจริงที่กำหนด โดยแบ่งเป็นการศึกษาความแม่นยำของค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแต่ละคน และความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

3.1 การศึกษาความแม่นยำของค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแต่ละคน เป็นการศึกษาความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณพัฒนาการความสามารถของผู้สอบแต่ละคนที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

3.1.1 การศึกษาความลำเอียงของค่าประมาณพัฒนาการความสามารถจากวิธีการประมาณค่าพารามิเตอร์พร้อมกันที่ได้ในแต่ละเงื่อนไข เป็นการคำนวณค่าเฉลี่ยของผลต่างระหว่างค่าประมาณพัฒนาการความสามารถกับค่าพัฒนาการความสามารถจริงรายคน โดยการคำนวณค่าความลำเอียงซึ่งเป็นการคำนวณแบบคิดเครื่องหมาย (Average Signed Bias: ASB) เพื่อดูทิศทางของความลำเอียง หากค่า BIAS เป็นลบ หมายความว่า ค่าประมาณความสามารถที่ได้เป็นค่าประมาณที่ต่ำกว่าความสามารถจริง แต่ในทางกลับกัน หากค่า BIAS เป็นบวก หมายความว่า ค่าประมาณความสามารถที่ได้เป็นค่าประมาณที่สูงกว่าความสามารถจริง โดยมีสูตรการหาค่าความลำเอียงซึ่ง*ปรับ*จากสูตรการหาค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absoluted Error: MAE) ของ เปค และ ยัง (Peak; & Young. 2005) ดังนี้

$$\text{MAE} = \frac{\sum_{i}^{N}\left|\hat{\theta}_i - \theta_i\right|}{N}$$, (Peak; & Young. 2005) ปรับใหม่เป็น $$\text{BIAS} = \frac{\sum_{i}^{N}[(\hat{\theta}_2 - \hat{\theta}_1)_i - (\theta_2 - \theta_1)_i]}{N}$$

เมื่อ $\hat{\theta}_i$ แทน ค่าประมาณความสามารถของผู้สอบคนที่ i

θ_i แทน ความสามารถจริงของผู้สอบคนที่ i

N แทน จำนวนผู้สอบทั้งหมด

$(\hat{\theta}_2 - \hat{\theta}_1)_i$ แทน ค่าประมาณพัฒนาการความสามารถจากวิธีการประมาณค่าพารามิเตอร์พร้อมกันของผู้สอบคนที่ i

$(\theta_2 - \theta_1)_i$ แทน ค่าพัฒนาการความสามารถจริงของผู้สอบคนที่ i

3.1.2 การศึกษาค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า เป็นการศึกษาจากค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error: RMSE) ของพัฒนาการความสามารถซึ่งค่า RMSE เป็นวิธีการประเมินความคงที่ (Stability) ในการประมาณค่าพารามิเตอร์ที่บอกถึงขนาดของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ความสามารถ ถ้าค่า RMSE มีค่าเข้าใกล้ศูนย์ หมายความว่า ค่าประมาณความสามารถที่ได้มีความแม่นยำในการวัดสูง โดยมีสูตรดังนี้ (Peak; & Young. 2005)

$$RMSE = \sqrt{\frac{\sum_{i}^{N}[(\hat{\theta}_2 - \hat{\theta}_1)_i - (\theta_2 - \theta_1)_i]^2}{N}}$$

เมื่อ N แทน จำนวนผู้สอบทั้งหมด

$(\hat{\theta}_2 - \hat{\theta}_1)_i$ แทน ค่าประมาณพัฒนาการความสามารถจากวิธีการประมาณค่าพารามิเตอร์พร้อมกันของผู้สอบคนที่ i

$(\theta_2 - \theta_1)_i$ แทน ค่าพัฒนาการความสามารถจริงของผู้สอบคนที่ i

3.2 การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันในแต่ละเงื่อนไข เป็นการศึกษาความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าเฉลี่ยพัฒนาการความสามารถจากผลการประมาณค่าแบบแผนการตอบซ้ำ 50 ครั้งเมื่อเปรียบเทียบกับพัฒนาการความสามารถจริง

3.2.1 การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันพิจารณาจากค่าความลำเอียง (BIAS) และ รากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าเฉลี่ยพัฒนาการความสามารถที่ทำซ้ำ 50 ครั้ง เมื่อเปรียบเทียบกับค่าเฉลี่ยความสามารถจริง ซึ่งคำนวณจากสูตรดังนี้ (Skaggs; & Lissitz. 1988)

$$BIAS = \frac{\sum_{i}^{p}(\hat{\theta}_i - \theta_i)}{p}, \qquad RMSE = \sqrt{\frac{\sum_{i}^{p}(\hat{\theta}_i - \theta_i)^2}{p}}$$

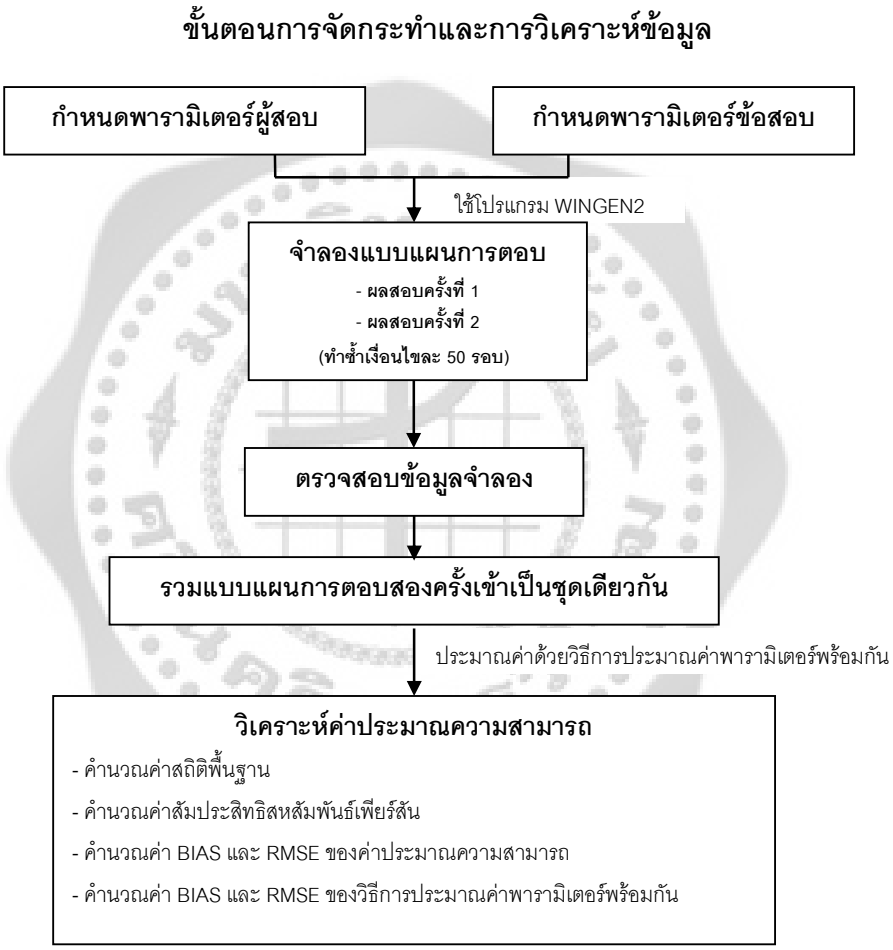
เมื่อ θ_i คือ ค่าพารามิเตอร์ความสามารถจริงของข้อมูลที่ทำซ้ำครั้งที่ i

$\hat{\theta}_i$ คือ ค่าเฉลี่ยพัฒนาการความสามารถจากวิธีการประมาณค่าพารามิเตอร์พร้อมกันของข้อมูลที่ทำซ้ำครั้งที่ i

p คือ จำนวนครั้งทั้งหมดของการทำซ้ำ

3.3 เปรียบเทียบค่าสถิติบรรยายของ BIAS และ RMSE ของค่าเฉลี่ยพัฒนาการความสามารถที่ได้ตามเงื่อนไขโดยใช้การวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (Repeated Measure) เพื่อตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันภายใต้เงื่อนไขที่ศึกษาแล้วเขียนกราฟแสดงความสัมพันธ์ระหว่างค่าสถิติ BIAS หรือ RMSE ในแต่ละเงื่อนไขที่ศึกษา

สรุปขั้นตอนการการจัดกระทำและการวิเคราะห์ข้อมูลดังภาพประกอบ 8



ภาพประกอบ 8 ขั้นตอนการจัดกระทำและการวิเคราะห์ข้อมูล

บทที่ 4
ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้เป็นการศึกษาภายใต้สถานการณ์จำลอง (Simulation Data) โดยมีจุดมุ่งหมายเพื่อตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันโดยใช้โมเดลโลจิสติกสามพารามิเตอร์ (3PL) ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิต (GPCM) ประมาณค่าความสามารถของผู้สอบ 1,000 คนที่ทดสอบซ้ำสองครั้งด้วยแบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน และผู้สอบมีระดับพัฒนาการความสามารถแตกต่างกัน นำเสนอผลการวิเคราะห์ข้อมูลเป็น 4 ตอน ดังนี้

ตอนที่ 1 การวิเคราะห์ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ตอนที่ 2 การศึกษาความสัมพันธ์ของพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ตอนที่ 3 การตรวจสอบความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ตอนที่ 4 การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

สัญลักษณ์ที่ใช้ในการนำเสนอผลการวิเคราะห์ข้อมูล

L แทน สัดส่วนจำนวนข้อสอบแบบให้คะแนนสองค่าต่อจำนวนข้อสอบแบบให้คะแนนหลายค่า

B แทน ค่าเฉลี่ยความยากของแบบทดสอบผสมสำหรับการทดสอบครั้งที่ 1

NC แทน ระดับคะแนนของข้อสอบแบบให้คะแนนหลายค่าในแบบทดสอบผสม

M1 แทน พัฒนาการความสามารถของผู้สอบที่อยู่ในรูปแบบ $\theta_2 = \theta_1 + 0.1$

M2 แทน พัฒนาการความสามารถของผู้สอบที่อยู่ในรูปแบบ $\theta_2 = \theta_1 + 0.5$

M3 แทน พัฒนาการความสามารถของผู้สอบที่อยู่ในรูปแบบ $\theta_2 = \theta_1 + 1.0$

M4 แทน พัฒนาการความสามารถของผู้สอบที่อยู่ในรูปแบบ $\theta_2 = 0.8\theta_1 + 0.1$

M5 แทน พัฒนาการความสามารถของผู้สอบที่อยู่ในรูปแบบ $\theta_2 = 0.8\theta_1 + 0.5$

M6 แทน พัฒนาการความสามารถของผู้สอบที่อยู่ในรูปแบบ $\theta_2 = 0.8\theta_1 + 1.0$

M7 แทน พัฒนาการความสามารถของผู้สอบที่อยู่ในรูปแบบ $\theta_2 = 1.2\theta_1 + 0.1$

M8 แทน พัฒนาการความสามารถของผู้สอบที่อยู่ในรูปแบบ $\theta_2 = 1.2\theta_1 + 0.5$

M9 แทน พัฒนาการความสามารถของผู้สอบที่อยู่ในรูปแบบ $\theta_2 = 1.2\theta_1 + 1.0$

θ แทน ค่าพารามิเตอร์ความสามารถจริงของผู้สอบ

θ_1 แทน ค่าพารามิเตอร์ความสามารถจริงของผู้สอบในการทดสอบครั้งที่ 1

θ_2 แทน ค่าพารามิเตอร์ความสามารถจริงของผู้สอบในการทดสอบครั้งที่ 2

$\theta_2 - \theta_1$ แทน พัฒนาการความสามารถจริงของผู้สอบที่คำนวณจากผลต่างของความสามารถผู้สอบครั้งที่ 2 กับครั้งที่ 1

$\hat{\theta}_1$ แทน ค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งแรกที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

$\hat{\theta}_2$ แทน ค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

$\hat{\theta}_2 - \hat{\theta}_1$ แทน พัฒนาการความสามารถของผู้สอบที่คำนวณจากผลต่างของความสามารถผู้สอบในการทดสอบครั้งที่ 2 กับครั้งที่ 1 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

Mean แทน ค่าเฉลี่ยเลขคณิต

SD แทน ส่วนเบี่ยงเบนมาตรฐาน

BIAS แทน ค่าความลำเอียง

RMSE แทน ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง

DS แทน ข้อสอบที่มีการตรวจให้คะแนนแบบสองค่า

PS แทน ข้อสอบที่มีการตรวจให้คะแนนแบบหลายค่า

Test A แทน แบบทดสอบที่ใช้ในการทดสอบครั้งที่ 1

Test B แทน แบบทดสอบที่ใช้ในการทดสอบครั้งที่ 2

SK แทน ความเบ้

KU แทน ความโด่ง

ตอนที่ 1 การวิเคราะห์ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ผู้วิจัยได้ตรวจสอบลักษณะโดยทั่วไปของข้อมูลพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ได้แก่ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง โดยจำแนกข้อมูลตามเงื่อนไขลักษณะของแบบทดสอบผสมที่ศึกษา ได้แก่ ความยาว ความยาก และการให้คะแนน นำเสนอผลการวิเคราะห์ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันแยกตามกลุ่มรูปแบบพัฒนาการความสามารถจริง 9 รูปแบบ โดยเรียงตามลำดับดังนี้

1.1 พัฒนาการความสามารถจริงเป็น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (M1)

1.2 พัฒนาการความสามารถจริงเป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (M2)

1.3 พัฒนาการความสามารถจริงเป็น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (M3)

1.4 พัฒนาการความสามารถจริงเป็น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 (M4)

1.5 พัฒนาการความสามารถจริงเป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 (M5)

1.6 พัฒนาการความสามารถจริงเป็น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 (M6)

1.7 พัฒนาการความสามารถจริงเป็น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 .2 (M7)

1.8 พัฒนาการความสามารถจริงเป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 (M8)

1.9 พัฒนาการความสามารถจริงเป็น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 (M9)

ผลการวิเคราะห์ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ปรากฏดังตาราง 5 – 13

ตาราง 5 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (M1)

L	B	NC	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_2 - \hat{\theta}_1$			
			Mean	SD	Mean	SD	Mean	SD	SK	KU
	θ		-0.007	0.987	0.093	0.987	0.100	0.000	-	-
		3	-0.545	0.383	-0.402	0.261	0.144	0.124	1.439	1.823
	-0.50	4	-0.471	0.337	-0.426	0.244	0.045	0.097	2.109	5.021
		5	-0.434	0.295	-0.420	0.190	0.014	0.108	2.002	3.717
		3	-0.446	0.353	-0.385	0.317	0.061	0.047	0.932	0.900
30:10	0.00	4	-0.398	0.317	-0.392	0.312	0.006	0.034	0.476	2.219
		5	-0.375	0.278	-0.363	0.171	0.012	0.109	2.027	3.872
		3	-0.339	0.295	-0.220	0.187	0.120	0.110	1.832	3.658
	0.50	4	-0.315	0.272	-0.270	0.183	0.046	0.091	2.162	5.435
		5	-0.308	0.243	-0.293	0.151	0.014	0.095	2.130	4.771
		3	-0.551	0.371	-0.479	0.350	0.072	0.041	0.491	-0.020
	-0.50	4	-0.502	0.347	-0.426	0.220	0.076	0.129	1.564	1.910
		5	-0.433	0.274	-0.415	0.165	0.018	0.111	1.925	3.281
		3	-0.441	0.332	-0.292	0.200	0.150	0.135	1.360	1.533
24:8	0.00	4	-0.416	0.315	-0.346	0.188	0.070	0.129	1.669	2.469
		5	-0.370	0.255	-0.352	0.147	0.018	0.111	1.960	3.625
		3	-0.347	0.277	-0.210	0.169	0.137	0.111	1.611	2.753
	0.50	4	-0.340	0.268	-0.274	0.162	0.066	0.109	1.854	3.496
		5	-0.311	0.221	-0.293	0.130	0.018	0.094	2.096	4.623
		3	-0.523	0.313	-0.390	0.225	0.133	0.093	0.660	0.010
	-0.50	4	-0.427	0.283	-0.399	0.210	0.028	0.078	1.598	2.317
		5	-0.393	0.222	-0.407	0.157	-0.014	0.074	1.524	2.359
		3	-0.448	0.300	-0.381	0.274	0.066	0.045	0.242	-0.140
15:5	0.00	4	-0.372	0.273	-0.338	0.188	0.034	0.089	1.514	1.912
		5	-0.349	0.218	-0.358	0.145	-0.010	0.081	1.448	1.992
		3	-0.334	0.269	-0.215	0.180	0.119	0.094	1.193	1.618
	0.50	4	-0.288	0.251	-0.252	0.172	0.036	0.083	1.855	3.388
		5	-0.280	0.202	-0.285	0.134	-0.005	0.075	1.643	2.820

จากตาราง 5 สรุปผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถจริงของผู้สอบอยู่ในรูปแบบ M1 คือ $\theta_2 = \theta_1 + 0.1$ ได้ว่า ค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และครั้งที่ 2 ทุกเงื่อนไขมีค่าเป็นลบ โดยค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 มีค่าตั้งแต่ -0.551 ถึง -0.280 ส่วนค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 มีค่าตั้งแต่ -0.479 ถึง -0.210 และค่าเฉลี่ยพัฒนาการความสามารถของผู้สอบมีค่าตั้งแต่ -0.014 ถึง 0.150 โดยส่วนใหญ่พบว่าค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการสอบครั้งที่ 2 สูงกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1 แต่ยกเว้นสำหรับกรณีที่แบบทดสอบผสมมีความยาว 15:5 และมีการให้คะแนน 5 ระดับที่พบว่าค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 มีค่าต่ำกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่งพบว่าพัฒนาการความสามารถในทุกเงื่อนไขมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติ โดยเงื่อนไขที่มีความเบ้ขวามากที่สุดคือ กรณีแบบทดสอบผสมยาว 30:10 มีความยาก 0.50 และการให้คะแนนมี 4 ระดับ รองลงมาคือ กรณีแบบทดสอบผสมยาว 30:10 มีความยาก 0.50 และการให้คะแนนมี 5 ระดับ และกรณีแบบทดสอบยาว 30:10 มีความยาก -0.50 และการให้คะแนนมี 4 ระดับ โดยมีค่าเท่ากับ 2.162 2.130 และ 2.109 ตามลำดับ ส่วนเงื่อนไขที่มีความโด่งมากที่สุด 3 อันดับแรกคือ กรณีแบบทดสอบยาว 30:10 มีความยาก 0.50 และการให้คะแนนมี 4 ระดับ กรณีแบบทดสอบยาว 30:10 มีความยาก -0.50 และการให้คะแนนมี 4 ระดับ และกรณีแบบทดสอบยาว 30:10 มีความยาก 0.50 และการให้คะแนนมี 5 ระดับ โดยมีค่าเท่ากับ 5.435 5.021 และ 4.771 ตามลำดับ

ตาราง 6 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (M2)

L	B	NC	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_2 - \hat{\theta}_1$			
			Mean	SD	Mean	SD	Mean	SD	SK	KU
	θ		-0.007	0.987	0.493	0.987	0.500	0.000	-	-
		3	-0.545	0.383	-0.506	0.297	0.039	0.103	2.729	7.385
	-0.50	4	-0.471	0.337	-0.519	0.293	-0.048	0.086	2.938	11.539
		5	-0.434	0.295	-0.493	0.239	-0.058	0.084	2.207	6.622
		3	-0.446	0.353	-0.512	0.389	-0.066	0.053	-1.029	0.919
30:10	0.00	4	-0.398	0.317	-0.508	0.399	-0.110	0.100	-0.832	-0.123
		5	-0.375	0.278	-0.426	0.225	-0.051	0.074	2.505	8.722
		3	-0.339	0.295	-0.295	0.235	0.044	0.071	3.373	12.561
	0.50	4	-0.315	0.272	-0.338	0.237	-0.023	0.063	3.276	14.639
		5	-0.308	0.243	-0.344	0.200	-0.037	0.061	2.434	9.674
		3	-0.551	0.371	-0.620	0.411	-0.070	0.058	-0.955	0.433
	-0.50	4	-0.502	0.347	-0.512	0.263	-0.010	0.106	2.839	8.375
		5	-0.433	0.274	-0.480	0.213	-0.047	0.082	2.414	7.152
		3	-0.441	0.332	-0.373	0.242	0.068	0.097	2.551	6.972
24:8	0.00	4	-0.416	0.315	-0.418	0.239	-0.002	0.092	3.222	11.154
		5	-0.370	0.254	-0.405	0.196	-0.035	0.072	2.774	9.920
		3	-0.347	0.277	-0.277	0.210	0.070	0.075	2.963	9.705
	0.50	4	-0.340	0.268	-0.333	0.210	0.007	0.072	3.258	12.212
		5	-0.311	0.220	-0.337	0.175	-0.026	0.060	2.736	10.89
		3	-0.523	0.313	-0.483	0.258	0.040	0.069	2.055	5.060
	-0.50	4	-0.427	0.283	-0.480	0.257	-0.053	0.069	1.856	5.509
		5	-0.393	0.222	-0.465	0.203	-0.073	0.050	1.097	4.002
		3	-0.448	0.300	-0.493	0.324	-0.045	0.045	-1.089	1.199
15:5	0.00	4	-0.372	0.273	-0.412	0.239	-0.040	0.061	2.358	8.385
		5	-0.349	0.218	-0.409	0.192	-0.060	0.047	1.938	8.744
		3	-0.334	0.269	-0.287	0.221	0.047	0.058	2.507	7.681
	0.50	4	-0.288	0.251	-0.316	0.223	-0.028	0.056	2.170	7.535
		5	-0.280	0.202	-0.331	0.184	-0.050	0.040	1.304	5.825

จากตาราง 6 ผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถจริงของผู้สอบอยู่ในรูปแบบ M2 คือ $\theta_2 = \theta_1 + 0.5$ พบว่า ค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และครั้งที่ 2 ทุกเงื่อนไขมีค่าเป็นลบ โดยค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 มีค่าตั้งแต่ -0.551 ถึง -0.280 ค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 มีค่าตั้งแต่ -0.620 ถึง -0.277 และค่าเฉลี่ยพัฒนาการความสามารถมีค่าตั้งแต่ -0.110 ถึง 0.070 โดยส่วนใหญ่พบว่าค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 ต่ำกว่าค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 ยกเว้นกรณีแบบทดสอบผสมมีความยาว 30:10 มีความยาก -0.50 และ 0.50 ที่ให้คะแนน 3 ระดับ กรณีแบบทดสอบผสมมีความยาว 24:8 มีความยาก 0.00 ที่ให้คะแนน 3 ระดับ กรณีแบบทดสอบผสมมีความยาว 24:8 มีความยาก 0.50 ให้คะแนน 3 และ 4 ระดับ และกรณีแบบทดสอบผสมมีความยาว 15:5 มีความยาก -0.50 และ 0.50 ให้คะแนน 3 ระดับ ที่พบว่าค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 มีค่าต่ำกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่ง พบว่าพัฒนาการความสามารถในทุกเงื่อนไขมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติ โดย กรณีแบบทดสอบผสมยาว 30:10 มีความยาก 0.50 และให้คะแนน 3 ระดับ เป็นเงื่อนไขที่มีความเบ้ขวามากที่สุดโดยมีค่าเท่ากับ 3.373 ส่วนกรณีแบบทดสอบผสมยาว 30:10 มีความยาก 0.50 และให้คะแนน 4 ระดับ มีความโด่งสูงกว่าโค้งปกติมากที่สุดโดยมีค่าเท่ากับ 14.639 และพบว่ามี 4 กรณีที่มีการแจกแจงแบบเบ้ซ้ายคือ กรณีแบบทดสอบผสมยาว 15:5 มีความยาก 0.00 และให้คะแนน 4 ระดับ กรณีแบบทดสอบผสมยาว 30:10 มีความยาก 0.00 และให้คะแนน 3 ระดับ กรณีแบบทดสอบผสมยาว 24:8 มีความยาก -0.50 และให้คะแนน 3 ระดับ และกรณีแบบทดสอบผสมยาว 30:10 มีความยาก 0.00 และให้คะแนน 4 ระดับ โดยมีความเบ้เป็น -1.089 -1.029 -0.955 และ -0.832 ตามลำดับ

ตาราง 7 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 (M3)

L	B	NC	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_2 - \hat{\theta}_1$			
			Mean	SD	Mean	SD	Mean	SD	SK	KU
	θ		-0.007	0.987	0.993	0.987	1.000	0.000	-	-
30:10	-0.50	3	-0.545	0.383	-0.657	0.323	-0.112	0.131	2.369	6.423
		4	-0.471	0.337	-0.667	0.335	-0.196	0.154	0.991	1.339
		5	-0.434	0.295	-0.612	0.286	-0.178	0.128	0.857	0.463
	0.00	3	-0.446	0.353	-0.714	0.466	-0.268	0.155	-0.255	-1.022
		4	-0.398	0.317	-0.709	0.491	-0.311	0.230	-0.172	-1.187
		5	-0.375	0.278	-0.537	0.279	-0.162	0.118	0.815	0.371
	0.50	3	-0.339	0.295	-0.416	0.284	-0.077	0.081	1.894	6.985
		4	-0.315	0.272	-0.457	0.295	-0.142	0.118	0.617	-0.171
		5	-0.308	0.243	-0.444	0.259	-0.136	0.105	0.682	-0.491
24:8	-0.50	3	-0.551	0.371	-0.830	0.459	-0.279	0.139	-0.021	-1.110
		4	-0.502	0.347	-0.644	0.301	-0.142	0.148	1.600	3.373
		5	-0.433	0.274	-0.586	0.254	-0.153	0.114	1.015	1.413
	0.00	3	-0.441	0.332	-0.498	0.282	-0.056	0.101	2.947	10.358
		4	-0.416	0.315	-0.535	0.288	-0.119	0.121	1.582	3.943
		5	-0.370	0.254	-0.497	0.247	-0.128	0.096	0.969	1.433
	0.50	3	-0.347	0.277	-0.386	0.257	-0.039	0.073	2.295	8.844
		4	-0.340	0.268	-0.440	0.263	-0.100	0.103	0.976	1.343
		5	-0.311	0.220	-0.423	0.230	-0.112	0.092	0.698	-0.037
15:5	-0.50	3	-0.523	0.314	-0.612	0.281	-0.088	0.086	2.176	6.518
		4	-0.427	0.283	-0.607	0.296	-0.180	0.134	0.538	-0.153
		5	-0.393	0.222	-0.568	0.253	-0.175	0.105	0.170	-0.991
	0.00	3	-0.447	0.300	-0.661	0.373	-0.214	0.106	-0.361	-0.836
		4	-0.372	0.273	-0.531	0.286	-0.159	0.114	0.605	0.108
		5	-0.349	0.218	-0.503	0.246	-0.154	0.092	0.198	-0.809
	0.50	3	-0.334	0.269	-0.402	0.267	-0.068	0.067	1.206	4.555
		4	-0.288	0.251	-0.428	0.277	-0.140	0.112	0.404	-0.708
		5	-0.280	0.202	-0.420	0.241	-0.140	0.096	0.204	-1.133

จากตาราง 7 ผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถจริงของผู้สอบอยู่ในรูปแบบ M3 คือ $\theta_2 = \theta_1 + 1.0$ พบว่า ค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และครั้งที่ 2 ทุกเงื่อนไขมีค่าเป็นลบโดยค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 มีค่าตั้งแต่ -0.551 ถึง -0.280 ส่วนค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 มีค่าตั้งแต่ -0.830 ถึง -0.402 และพัฒนาการความสามารถมีค่าตั้งแต่ -0.279 ถึง -0.039 โดยพบว่าค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 ต่ำกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1 ทุกเงื่อนไข

เมื่อพิจารณาแต่ละเงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม พบว่ากรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน กรณีให้คะแนน 5 ระดับ จะมีค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 เข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.0) มากที่สุด รองลงมาคือกรณีให้คะแนน 4 ระดับ และ 3 ระดับตามลำดับ แต่สำหรับค่าประมาณความสามารถในการทดสอบครั้งที่สองกลับให้ผลตรงกันข้าม คือ กรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน กรณีให้คะแนน 3 ระดับ ส่วนใหญ่มีค่าเฉลี่ยของค่าประมาณความสามารถเข้าใกล้ค่าเฉลี่ยความสามารถจริง (1.0) มากที่สุด รองลงมาคือ กรณีให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ ยกเว้นกรณีแบบทดสอบผสมยาว 30:10 มีความยาก 0.00 และให้คะแนน 5 ระดับที่มีค่าเฉลี่ยเข้าใกล้ค่าเฉลี่ยความสามารถจริง (1.0) มากที่สุด

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่งพบว่า พัฒนาการความสามารถส่วนใหญ่มีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติ โดย กรณีแบบทดสอบยาว 24:8 มีความยาก 0.00 และให้คะแนน 3 ระดับ เป็นเงื่อนไขที่มีความเบ้ขวามากที่สุดโดยมีค่า 2.947 และส่วนกรณีแบบทดสอบยาว 30:10 มีความยาก 0.50 และให้คะแนน 4 ระดับมีความโด่งสูงกว่าโค้งปกติมากที่สุด โดยมีค่า 13.564 นอกจากนี้พบว่า มี 2 กรณีที่มีการแจกแจงแบบเบ้ซ้ายและโด่งต่ำกว่าโค้งปกติคือ กรณีแบบทดสอบยาว 30:10 มีความยาก 0.00 และให้คะแนน 3 และ 4 ระดับ โดยมีความเบ้เป็น -0.271 และ -0.300 และความโด่งเป็น -0.766 และ -0.964 ตามลำดับ

ตาราง 8 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 (M4)

L	B	NC	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_2 - \hat{\theta}_1$			
			Mean	SD	Mean	SD	Mean	SD	SK	KU
	θ		-0.007	0.987	0.095	0.79	0.101	0.197	-0.060	-0.245
		3	-0.546	0.383	-0.378	0.211	0.167	0.173	1.331	1.082
	-0.50	4	-0.471	0.337	-0.393	0.194	0.078	0.145	1.695	2.239
		5	-0.434	0.295	-0.390	0.147	0.044	0.150	1.960	3.349
		3	-0.446	0.353	-0.348	0.240	0.098	0.117	1.484	1.562
30:10	0.00	4	-0.398	0.317	-0.382	0.201	0.016	0.129	2.577	6.480
		5	-0.375	0.278	-0.333	0.129	0.042	0.151	2.091	4.069
		3	-0.339	0.295	-0.195	0.142	0.144	0.154	1.923	3.707
	0.50	4	-0.315	0.272	-0.238	0.138	0.077	0.135	2.119	4.511
		5	-0.308	0.243	-0.263	0.110	0.044	0.134	2.227	4.984
		3	-0.551	0.371	-0.444	0.278	0.107	0.101	0.877	-0.151
	-0.50	4	-0.502	0.347	-0.398	0.175	0.104	0.174	1.474	1.261
		5	-0.433	0.274	-0.389	0.128	0.044	0.148	1.890	2.975
		3	-0.441	0.332	-0.269	0.154	0.172	0.180	1.470	1.658
24:8	0.00	4	-0.416	0.315	-0.317	0.143	0.099	0.174	1.714	2.272
		5	-0.370	0.254	-0.325	0.108	0.045	0.149	2.072	3.972
		3	-0.347	0.277	-0.189	0.129	0.158	0.150	1.738	2.930
	0.50	4	-0.340	0.268	-0.246	0.120	0.094	0.150	1.911	3.363
		5	-0.311	0.22	-0.267	0.096	0.044	0.127	2.248	5.257
		3	-0.523	0.313	-0.372	0.184	0.151	0.133	0.790	-0.105
	-0.50	4	-0.427	0.283	-0.371	0.164	0.056	0.123	1.560	1.594
		5	-0.393	0.222	-0.381	0.119	0.012	0.108	1.673	2.320
		3	-0.448	0.300	-0.354	0.213	0.093	0.095	0.962	0.358
15:5	0.00	4	-0.372	0.273	-0.310	0.144	0.062	0.132	1.543	1.601
		5	-0.349	0.218	-0.331	0.106	0.018	0.118	1.738	2.617
		3	-0.334	0.269	-0.193	0.136	0.141	0.135	1.516	2.153
	0.50	4	-0.288	0.251	-0.223	0.128	0.065	0.126	1.889	3.173
		5	-0.280	0.202	-0.257	0.096	0.023	0.111	2.023	4.109

จากตาราง 8 สรุปผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถจริงของผู้สอบอยู่ในรูปแบบ M4 คือ $\theta_2 = 0.8\,\theta_1 + 0.1$ ได้ว่า ค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และครั้งที่ 2 ทุกเงื่อนไขมีค่าเป็นลบ โดยพบว่าค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 มีค่าตั้งแต่ -0.551 ถึง -0.280 ส่วนค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 มีค่าตั้งแต่ -0.444 ถึง -0.189 และค่าเฉลี่ยของพัฒนาการความสามารถมีค่าตั้งแต่ 0.012 ถึง 0.172 แสดงว่าค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 สูงกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1 ทุกเงื่อนไข

เมื่อพิจารณาแต่ละเงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม พบว่ากรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน เมื่อให้คะแนน 5 ระดับ จะมีค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 เข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.0) มากที่สุด รองลงมาคือ การให้คะแนน 4 และ 3 ระดับ ตามลำดับ แต่สำหรับค่าประมาณความสามารถในการสอบครั้งที่ 2 กลับให้ผลตรงกันข้าม คือ กรณีแบบทดสอบผสมมีความยาว และความยากเท่ากัน การให้คะแนน 3 ระดับ มีค่าเฉลี่ยของค่าประมาณความสามารถเข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.10) มากที่สุด รองลงมาคือ การให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่งพบว่าพัฒนาการความสามารถส่วนใหญ่มีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติ โดยกรณีแบบทดสอบยาว 30:10 มีความยาก 0.00 และให้คะแนน 4 ระดับ เป็นเงื่อนไขที่มีความเบ้ขวามากที่สุดโดยมีค่า 2.577 และมีความโด่งสูงกว่าโค้งปกติมากที่สุดเช่นกัน โดยมีค่า 6.480 นอกจากนี้พบว่ามี 2 กรณีที่มีการแจกแจงแบบโด่งต่ำกว่าโค้งปกติ ได้แก่ กรณีแบบทดสอบยาว 24:8 มีความยาก -0.50 ให้คะแนน 3 ระดับ และกรณี แบบทดสอบยาว 15:5 มีความยาก -0.50 ให้คะแนน 3 ระดับ โดยมีค่าเท่ากับ -0.151 และ -0.105 ตามลำดับ

ตาราง 9 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 (M5)

L	B	NC	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_2 - \hat{\theta}_1$			
			Mean	SD	Mean	SD	Mean	SD	SK	KU
	θ		-0.007	0.987	0.495	0.79	0.501	0.197	-0.060	-0.245
		3	-0.545	0.383	-0.488	0.251	0.058	0.140	1.989	3.478
	-0.50	4	-0.471	0.337	-0.492	0.251	-0.021	0.106	2.642	7.927
		5	-0.434	0.295	-0.465	0.198	-0.030	0.116	2.158	4.919
		3	-0.446	0.353	-0.475	0.314	-0.029	0.053	2.093	5.066
30:10	0.00	4	-0.398	0.317	-0.460	0.325	-0.062	0.049	0.313	-0.394
		5	-0.375	0.278	-0.396	0.180	-0.020	0.112	2.338	5.949
		3	-0.339	0.295	-0.271	0.187	0.069	0.114	2.566	6.782
	0.50	4	-0.315	0.272	-0.307	0.192	0.009	0.094	2.935	9.965
		5	-0.308	0.243	-0.315	0.157	-0.007	0.098	2.461	7.352
		3	-0.551	0.371	-0.592	0.342	-0.041	0.045	1.617	3.599
	-0.50	4	-0.502	0.347	-0.488	0.227	0.014	0.130	2.129	4.263
		5	-0.433	0.274	-0.454	0.174	-0.021	0.114	2.266	5.094
		3	-0.441	0.332	-0.351	0.197	0.090	0.139	1.988	3.736
24:8	0.00	4	-0.416	0.315	-0.390	0.196	0.026	0.126	2.426	5.865
		5	-0.370	0.254	-0.377	0.154	-0.007	0.109	2.462	6.307
		3	-0.347	0.277	-0.256	0.168	0.091	0.113	2.366	5.829
	0.50	4	-0.340	0.268	-0.306	0.170	0.034	0.106	2.572	7.171
		5	-0.311	0.220	-0.310	0.137	0.000	0.093	2.558	7.644
		3	-0.523	0.314	-0.468	0.218	0.055	0.102	1.398	1.707
	-0.50	4	-0.427	0.283	-0.454	0.215	-0.027	0.087	1.938	4.317
		5	-0.393	0.222	-0.438	0.166	-0.045	0.073	2.111	5.089
		3	-0.448	0.300	-0.467	0.263	-0.019	0.051	1.305	1.895
15:5	0.00	4	-0.372	0.273	-0.384	0.196	-0.011	0.088	2.055	4.793
		5	-0.349	0.218	-0.380	0.153	-0.031	0.076	2.233	5.373
		3	-0.334	0.269	-0.266	0.178	0.068	0.096	2.141	4.800
	0.50	4	-0.288	0.251	-0.287	0.178	0.002	0.087	2.239	5.936
		5	-0.280	0.202	-0.301	0.144	-0.021	0.070	2.338	6.588

จากตาราง 9 สรุปผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถจริงของผู้สอบอยู่ในรูปแบบ M5 คือ $\theta_2 = 0.8\,\theta_1 + 0.5$ ได้ว่าค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และครั้งที่ 2 ทุกเงื่อนไขมีค่าเป็นลบโดยค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 มีค่าตั้งแต่ -0.551 ถึง -0.280 ส่วนค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 มีค่าตั้งแต่ -0.592 ถึง -0.256 และพัฒนาการความสามารถมีค่าตั้งแต่ -0.045 ถึง 0.091 โดยส่วนใหญ่พบว่าค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 สูงกว่าค่าประมาณความสามารถในการทดสอบครั้งที่ 1 แต่สำหรับกรณีที่ให้คะแนน 5 ระดับพบว่าค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 มีค่าต่ำกว่าค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ทุกเงื่อนไข

เมื่อพิจารณาแต่ละเงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม พบว่ากรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน เมื่อให้คะแนน 5 ระดับ ค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 เข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.00) มากที่สุด รองลงมาคือ การให้คะแนน 4 ระดับ และ 3 ระดับ ตามลำดับ แต่สำหรับค่าประมาณความสามารถในการสอบครั้งที่ 2 กลับให้ผลตรงกันข้าม คือ กรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน เมื่อให้คะแนน 3 ระดับส่วนใหญ่มีค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 เข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.50) มากที่สุด รองลงมาคือ การให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ แต่สำหรับกรณีแบบทดสอบมีความยาว 30:10 มีความยาก 0.00 และกรณีแบบทดสอบยาว 15:5 มีความยาก 0.00 พบว่าการให้คะแนน 5 ระดับ มีค่าเฉลี่ยมากที่สุด

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่งพบว่าพัฒนาการความสามารถในทุกเงื่อนไขมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติ โดยกรณีแบบทดสอบยาว 30:10 มีความยาก 0.50 ที่ให้คะแนน 4 ระดับ มีความเบ้ขวาและโด่งสูงกว่าโค้งปกติมากที่สุด โดยมีค่าเท่ากับ 2.935 และ 9.965 ตามลำดับ รองลงมาคือ กรณีแบบทดสอบยาว 30:10 มีความยาก -0.50 ที่ให้คะแนน 4 ระดับ โดยมีความเบ้ขวาเท่ากับ 2.642 และความโด่งเท่ากับ 7.927

ตาราง 10 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 0.8 (M6)

L	B	NC	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_2 - \hat{\theta}_1$			
			Mean	SD	Mean	SD	Mean	SD	SK	KU
	θ		-0.007	0.987	0.995	0.790	1.001	0.197	-0.060	-0.245
		3	-0.545	0.383	-0.652	0.281	-0.106	0.143	2.445	5.970
	-0.50	4	-0.471	0.337	-0.654	0.297	-0.183	0.144	1.574	3.215
		5	-0.434	0.295	-0.595	0.246	-0.161	0.133	1.196	1.477
		3	-0.448	0.353	-0.681	0.393	-0.233	0.092	0.264	-0.440
30:10	0.00	4	-0.398	0.317	-0.672	0.425	-0.274	0.169	0.195	-0.952
		5	-0.375	0.278	-0.515	0.237	-0.139	0.119	1.289	2.037
		3	-0.339	0.295	-0.398	0.237	-0.058	0.093	2.919	10.206
	0.50	4	-0.315	0.272	-0.434	0.250	-0.118	0.111	1.422	2.584
		5	-0.308	0.243	-0.419	0.216	-0.111	0.106	1.084	0.888
		3	-0.551	0.371	-0.815	0.397	-0.265	0.092	0.326	-0.338
	-0.50	4	-0.502	0.347	-0.632	0.265	-0.130	0.149	1.985	4.473
		5	-0.433	0.274	-0.570	0.218	-0.136	0.120	1.448	2.707
		3	-0.441	0.332	-0.482	0.241	-0.041	0.117	2.938	8.750
24:8	0.00	4	-0.416	0.315	-0.520	0.252	-0.104	0.125	2.126	5.699
		5	-0.370	0.254	-0.480	0.210	-0.110	0.105	1.465	3.189
		3	-0.347	0.277	-0.370	0.215	-0.023	0.090	3.062	10.682
	0.50	4	-0.340	0.268	-0.418	0.224	-0.078	0.106	1.743	4.118
		5	-0.311	0.220	-0.398	0.191	-0.087	0.093	1.161	1.531
		3	-0.523	0.314	-0.608	0.244	-0.086	0.096	2.402	5.959
	-0.50	4	-0.427	0.283	-0.594	0.261	-0.167	0.119	1.090	1.231
		5	-0.393	0.222	-0.547	0.218	-0.154	0.094	0.500	-0.196
		3	-0.447	0.300	-0.643	0.317	-0.195	0.063	0.070	0.332
15:5	0.00	4	-0.372	0.273	-0.513	0.248	-0.141	0.104	1.208	1.869
		5	-0.349	0.218	-0.482	0.208	-0.133	0.084	0.660	0.349
		3	-0.334	0.269	-0.385	0.224	-0.051	0.074	2.715	9.127
	0.50	4	-0.288	0.251	-0.406	0.237	-0.118	0.101	1.035	0.830
		5	-0.280	0.202	-0.396	0.202	-0.115	0.086	0.605	-0.218

จากตาราง 10 สรุปผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถจริงของผู้สอบอยู่ในรูปแบบ M6 คือ $\theta_2 = 0.8\,\theta_1 + 1.0$ ได้ว่า ค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และครั้งที่ 2 ทุกเงื่อนไขมีค่าเป็นลบโดยค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 มีค่าตั้งแต่ -0.551 ถึง -0.280 ส่วนค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 มีค่าตั้งแต่ -0.815 ถึง -0.370 และค่าเฉลี่ยพัฒนาการความสามารถมีค่าตั้งแต่ -0.274 ถึง -0.041 โดยพบว่าค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 มีค่าต่ำกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1 ทุกเงื่อนไข

เมื่อพิจารณาแต่ละเงื่อนไขความยาว ความยาก และรูปแบบการให้คะแนนของแบบทดสอบผสมพบว่ากรณีที่ความยาว และความยากเท่ากัน เมื่อให้คะแนน 5 ระดับจะมีค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 เข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.00) มากที่สุด รองลงมาคือรูปแบบการให้คะแนน 4 ระดับ และ 3 ระดับ ตามลำดับ ส่วนค่าประมาณความสามารถในการสอบครั้งสองกลับให้ผลตรงกันข้ามในทุกกรณีที่แบบทดสอบผสมมียากเป็น 0.50 เมื่อให้คะแนน 3 ระดับมีค่าเฉลี่ยของค่าประมาณความสามารถเข้าใกล้ค่าเฉลี่ยความสามารถจริง (1.0) มากที่สุด

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่งพบว่าพัฒนาการความสามารถในทุกเงื่อนไขมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติ โดยเงื่อนไขที่มีความเบ้ขวาและโด่งสูงกว่าโค้งปกติมากที่สุดคือ กรณีแบบทดสอบยาว 30:10 มีความยาก 0.50 และให้คะแนน 3 ระดับ โดยมีค่าเท่ากับ –3.062 และ 10.682 ตามลำดับ และพบว่ามี 5 กรณีที่พัฒนาการความสามารถมีลักษณะโด่งต่ำกว่าโค้งปกติ คือ กรณีแบบทดสอบมีความยาว 30:10 มีความยาก 0.00 และให้คะแนน 3 ระดับ และ 4 ระดับ กรณีแบบทดสอบมีความยาว 24:8 มีความยาก -0.50 และให้คะแนน 3 ระดับ และกรณีแบบทดสอบมีความยาว 15:5 มีความยาก -0.50 และ 0.50 และให้คะแนน 5 ระดับ

ตาราง 11 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.1 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 (M7)

L	B	NC	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_2 - \hat{\theta}_1$			
			Mean	SD	Mean	SD	Mean	SD	SK	KU
	θ		-0.007	0.987	0.092	1.185	0.099	0.197	0.060	-0.245
		3	-0.545	0.383	-0.424	0.303	0.121	0.088	1.979	4.862
	-0.50	4	-0.471	0.337	-0.455	0.283	0.016	0.068	3.535	15.637
		5	-0.434	0.295	-0.450	0.229	-0.016	0.072	2.427	7.064
		3	-0.446	0.353	-0.425	0.387	0.021	0.049	-1.686	3.418
30:10	0.00	4	-0.398	0.317	-0.441	0.382	-0.043	0.077	-1.947	3.307
		5	-0.375	0.278	-0.393	0.212	-0.018	0.072	2.272	6.737
		3	-0.339	0.295	-0.245	0.232	0.094	0.069	1.785	5.058
	0.50	4	-0.315	0.272	-0.302	0.225	0.014	0.057	2.735	11.393
		5	-0.308	0.243	-0.419	0.04	-0.112	0.217	2.385	6.270
		3	-0.551	0.371	-0.514	0.412	0.037	0.058	-1.466	2.000
	-0.50	4	-0.502	0.347	-0.455	0.258	0.047	0.096	2.187	5.432
		5	-0.433	0.274	-0.441	0.203	-0.008	0.075	2.167	5.125
		3	-0.441	0.332	-0.315	0.241	0.126	0.096	1.409	2.782
24:8	0.00	4	-0.416	0.315	-0.375	0.228	0.041	0.093	2.015	5.193
		5	-0.370	0.254	-0.379	0.184	-0.010	0.076	2.097	5.458
		3	-0.347	0.277	-0.233	0.209	0.114	0.074	1.456	3.437
	0.50	4	-0.340	0.268	-0.302	0.199	0.038	0.075	1.977	5.239
		5	-0.311	0.220	-0.319	0.166	-0.009	0.061	1.940	5.226
		3	-0.523	0.313	-0.409	0.265	0.114	0.061	0.609	1.137
	-0.50	4	-0.427	0.283	-0.425	0.247	0.002	0.050	2.709	9.909
		5	-0.393	0.222	-0.433	0.193	-0.040	0.046	1.185	2.515
		3	-0.448	0.300	-0.409	0.327	0.038	0.051	-1.000	1.121
15:5	0.00	4	-0.372	0.273	-0.368	0.229	0.005	0.054	2.032	6.119
		5	-0.349	0.218	-0.386	0.183	-0.037	0.051	0.875	1.484
		3	-0.334	0.269	-0.238	0.219	0.096	0.061	0.573	1.256
	0.50	4	-0.288	0.251	-0.281	0.211	0.008	0.049	2.067	5.814
		5	-0.280	0.202	-0.313	0.173	-0.033	0.045	0.766	1.397

จากตาราง 11 สรุปผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถจริงของผู้สอบอยู่ในรูปแบบ M7 คือ $\theta_2 = 1.2\theta_1 + 0.1$ ได้ค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และครั้งที่ 2 ทุกเงื่อนไขมีค่าเป็นลบโดยค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 1 มีค่าตั้งแต่ -0.551 ถึง -0.288 ส่วนค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 มีค่าตั้งแต่ -0.514 ถึง -0.233 และพัฒนาการความสามารถมีค่าตั้งแต่ -0.112 ถึง 0.126 โดยส่วนใหญ่พบว่าค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 สูงกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1 ยกเว้นกรณีที่แบบทดสอบผสมมีการให้คะแนน 5 ระดับทุกเงื่อนไข พบว่าค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 มีค่าต่ำกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1

เมื่อพิจารณาแต่ละเงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม พบว่ากรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน การให้คะแนน 5 ระดับจะมีค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 เข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.0) มากที่สุด รองลงมาคือ การให้คะแนน 4 ระดับ และ 3 ระดับ ตามลำดับ แต่สำหรับค่าประมาณความสามารถในการสอบครั้งสองกลับให้ผลตรงกันข้าม คือ กรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน ส่วนใหญ่เมื่อให้คะแนน 3 ระดับค่าเฉลี่ยของค่าประมาณความสามารถเข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.1) มากที่สุด ยกเว้นกรณีแบบทดสอบมีความยาก -0.50 ที่ให้คะแนน 5 ระดับ ให้ค่าเฉลี่ยสูงที่สุด

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่งพบว่าพัฒนาการความสามารถในทุกเงื่อนไขมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติ โดยเงื่อนไขที่มีความเบ้ขวาและโด่งสูงกว่าโค้งปกติมากที่สุดคือกรณีแบบทดสอบผสมยาว 30:10 มีความยาก -0.50 และรูปแบบการให้คะแนนมี 4 ระดับ มีค่าเท่ากับ 3.535 และ 15.637 ตามลำดับ และพบว่ามี 4 กรณีที่พัฒนาการความสามารถมีลักษณะการแจกแจงแบบเบ้ซ้าย คือ กรณีแบบทดสอบมีความยาว 30:10 มีความยาก 0.00 และให้คะแนน 3 และ 4 ระดับ โดยมีค่าเท่ากับ -1.686 และ -1.947 ตามลำดับ กรณีแบบทดสอบมีความยาว 24:8 มีความยาก -0.50 และให้คะแนน 3 ระดับ มีค่าเท่ากับ -1.466 และ กรณีแบบทดสอบมีความยาว15:5 มีความยาก 0.00 และให้คะแนน 3 ระดับมีค่าเท่ากับ -1.000

ตาราง 12 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 0.5 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 (M8)

L	B	NC	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_2 - \hat{\theta}_1$			
			Mean	SD	Mean	SD	Mean	SD	SK	KU
	θ		-0.007	0.987	0.492	1.185	0.499	0.197	0.060	-0.245
		3	-0.545	0.383	-0.523	0.335	0.022	0.087	3.157	11.181
	-0.50	4	-0.471	0.337	-0.545	0.329	-0.074	0.094	1.680	6.755
		5	-0.434	0.295	-0.519	0.273	-0.085	0.072	1.789	7.104
		3	-0.446	0.353	-0.549	0.454	-0.102	0.117	-0.999	-0.143
30:10	0.00	4	-0.398	0.317	-0.557	0.462	-0.158	0.166	-0.993	-0.265
		5	-0.375	0.278	-0.455	0.261	-0.080	0.060	2.345	12.941
		3	-0.339	0.295	-0.320	0.278	0.019	0.049	3.443	19.387
	0.50	4	-0.315	0.272	-0.368	0.277	-0.053	0.061	1.141	6.093
		5	-0.308	0.243	-0.376	0.240	-0.068	0.046	1.337	7.458
		3	-0.551	0.371	-0.648	0.465	-0.098	0.112	-0.778	-0.600
	-0.50	4	-0.502	0.347	-0.533	0.294	-0.031	0.097	2.968	10.326
		5	-0.433	0.274	-0.503	0.242	-0.070	0.069	2.240	8.753
		3	-0.441	0.332	-0.393	0.279	0.048	0.075	3.401	13.303
24:8	0.00	4	-0.416	0.315	-0.446	0.274	-0.029	0.078	3.535	15.528
		5	-0.370	0.254	-0.433	0.231	-0.063	0.054	3.034	16.809
		3	-0.347	0.277	-0.300	0.250	0.046	0.049	3.476	17.056
	0.50	4	-0.340	0.268	-0.361	0.246	-0.021	0.056	2.951	14.654
		5	-0.311	0.220	-0.365	0.211	-0.054	0.043	2.191	13.058
		3	-0.523	0.313	-0.496	0.291	0.027	0.058	1.977	6.705
	-0.50	4	-0.427	0.283	-0.502	0.286	-0.075	0.079	0.755	2.353
		5	-0.393	0.222	-0.491	0.236	-0.099	0.056	-0.152	0.844
		3	-0.448	0.300	-0.517	0.374	-0.069	0.089	-0.983	-0.083
15:5	0.00	4	-0.372	0.273	-0.437	0.273	-0.065	0.063	1.043	5.258
		5	-0.349	0.218	-0.437	0.227	-0.088	0.046	0.209	3.420
		3	-0.334	0.269	-0.310	0.261	0.024	0.044	1.395	7.125
	0.50	4	-0.288	0.251	-0.345	0.260	-0.057	0.059	0.383	2.402
		5	-0.280	0.202	-0.360	0.221	-0.080	0.046	-0.324	0.850

จากตาราง 12 ผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถจริงของผู้สอบอยู่ในรูปแบบ M8 คือ $\theta_2 = 1.2\theta_1 + 0.5$ พบว่า ค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และการทดสอบครั้งที่ 2 ทุกเงื่อนไขมีค่าเป็นลบโดยค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 มีค่าตั้งแต่ -0.551 ถึง -0.280 ส่วนค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 มีค่าตั้งแต่ -0.648 ถึง -0.300 และค่าเฉลี่ยของพัฒนาการความสามารถมีค่าตั้งแต่ -0.158 ถึง 0.048 โดยพบว่าค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 สูงกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1 เฉพาะกรณีแบบทดสอบผสมยาว 30:10 มีความยาก -0.50 และ 0.50 ให้คะแนน 3 ระดับแบบทดสอบยาว 24:8 มีความยาก 0.00 และ 0.50 ให้คะแนน 3 ระดับและกรณีแบบทดสอบผสมยาว 15:5 มีความยาก -0.50 และ 0.50 ให้คะแนน 3 ระดับ ส่วนกรณีอื่นที่เหลือค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ต่ำกว่าค่าประมาณความสามารถในการทดสอบครั้งที่ 1

เมื่อพิจารณาในแต่ละเงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม พบว่ากรณีที่แบบทดสอบผสมมีความยาวและความยากเท่ากันเมื่อให้คะแนน 5 ระดับ ค่าเฉลี่ยของค่าประมาณความสามารถเข้าใกล้ค่าเฉลี่ยความสามารถจริงมากที่สุดในทุกเงื่อนไขของการทดสอบครั้งที่ 1 และเงื่อนไขส่วนใหญ่ของการทดสอบครั้งที่ 2 ยกเว้นบางกรณีได้แก่ กรณีที่แบบทดสอบผสมมีความยาว 30:10 มีความยาก 0.50 กรณีแบบทดสอบผสมมีความยาว 24:8 มีความยาก 0.00 และ 0.50 และกรณีแบบทดสอบมีความยาว 15:5 มีความยาก 0.50 ที่พบว่าค่าเฉลี่ยของค่าประมาณความสามารถเมื่อให้คะแนน 3 ระดับเข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.5) มากที่สุด

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่งพบว่าพัฒนาการความสามารถในทุกเงื่อนไขมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติ โดยเงื่อนไขที่มีความเบ้ขวามากที่สุดคือ กรณีแบบทดสอบยาว 24:8 มีความยาก 0.00 และให้คะแนน 4 ระดับ โดยมีค่าเท่ากับ 3.535 เงื่อนไขที่มีความโด่งสูงกว่าโค้งปกติมากที่สุดคือ กรณีแบบทดสอบมีความยาว 30:10 มีความยาก 0.50 และให้คะแนน 3 ระดับ มีค่าเท่ากับ 19.387 และพบว่ามี 6 กรณีที่พัฒนาการความสามารถมีลักษณะการแจกแจงแบบเบ้ซ้าย โดยกรณีแบบทดสอบมีความยาว 30:10 มีความยาก 0.00 และให้คะแนน 3 เบ้ซ้ายมากที่สุดมีค่าเท่ากับ -0.999 และ พบว่ามี 4 กรณีที่พัฒนาการความสามารถมีความโด่งต่ำกว่าโค้งปกติ คือ กรณีแบบทดสอบมีความยาว 24:8 มีความยาก -0.50 และให้คะแนน 3 ระดับ โดยมีค่าเท่ากับ -0.600

ตาราง 13 ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีพัฒนาการความสามารถจริงเป็น 1.0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 (M9)

L	B	NC	$\hat{\theta}_1$		$\hat{\theta}_2$		$\hat{\theta}_2 - \hat{\theta}_1$			
			Mean	SD	Mean	SD	Mean	SD	SK	KU
	θ		-0.007	0.987	0.992	1.185	0.999	0.197	0.060	-0.245
30:10	-0.50	3	-0.545	0.383	-0.660	0.357	-0.114	0.134	1.715	4.215
		4	-0.471	0.337	-0.680	0.365	-0.208	0.173	0.525	0.117
		5	-0.434	0.296	-0.630	0.316	-0.195	0.139	0.524	-0.365
	0.00	3	-0.446	0.353	-0.740	0.523	-0.294	0.215	-0.408	-1.137
		4	-0.398	0.317	-0.742	0.541	-0.343	0.285	-0.342	-1.257
		5	-0.375	0.278	-0.559	0.312	-0.184	0.131	0.381	-0.604
	0.50	3	-0.339	0.295	-0.436	0.325	-0.096	0.097	0.442	0.942
		4	-0.315	0.272	-0.480	0.331	-0.165	0.139	0.132	-1.005
		5	-0.308	0.243	-0.470	0.296	-0.162	0.121	0.284	-1.095
24:8	-0.50	3	-0.551	0.371	-0.840	0.507	-0.289	0.185	-0.156	-1.257
		4	-0.502	0.347	-0.653	0.326	-0.152	0.157	1.171	2.120
		5	-0.433	0.274	-0.601	0.282	-0.168	0.121	0.615	0.265
	0.00	3	-0.441	0.332	-0.509	0.315	-0.068	0.106	1.886	6.431
		4	-0.416	0.315	-0.555	0.317	-0.138	0.135	0.907	1.921
		5	-0.370	0.255	-0.524	0.279	-0.154	0.114	0.424	0.144
	0.50	3	-0.347	0.277	-0.405	0.292	-0.058	0.084	0.722	2.419
		4	-0.340	0.268	-0.461	0.294	-0.121	0.115	0.360	-0.261
		5	-0.311	0.220	-0.448	0.264	-0.137	0.107	0.177	-1.034
15:5	-0.50	3	-0.523	0.314	-0.617	0.31	-0.094	0.095	1.180	3.015
		4	-0.427	0.283	-0.618	0.322	-0.191	0.152	0.185	-0.764
		5	-0.393	0.222	-0.585	0.278	-0.192	0.121	-0.041	-1.274
	0.00	3	-0.448	0.300	-0.674	0.418	-0.226	0.150	-0.410	-1.063
		4	-0.372	0.273	-0.546	0.314	-0.174	0.133	0.121	-0.640
		5	-0.349	0.218	-0.527	0.275	-0.178	0.112	-0.132	-1.130
	0.50	3	-0.334	0.269	-0.418	0.302	-0.084	0.085	-0.042	0.090
		4	-0.288	0.251	-0.449	0.310	-0.161	0.133	0.030	-1.106
		5	-0.280	0.202	-0.445	0.274	-0.165	0.119	-0.170	-1.283

จากตาราง 13 สรุปผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถจริงของผู้สอบอยู่ในรูปแบบ M9 คือ $\theta_2 = 1.2\theta_1 + 1.0$ ได้ว่า ค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบทุกเงื่อนไขในการทดสอบครั้งที่ 1 และการทดสอบครั้งที่ 2 มีค่าเป็นลบโดยค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 มีค่าตั้งแต่ -0.551 ถึง -0.280 ส่วนค่าเฉลี่ยของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 2 มีค่าตั้งแต่ -0.840 ถึง -0.405 และค่าเฉลี่ยของพัฒนาการความสามารถมีค่าตั้งแต่ -0.343 ถึง -0.058 โดยพบว่าค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 สูงกว่าค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ทุกเงื่อนไข

เมื่อพิจารณาแต่ละเงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม พบว่ากรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน การให้คะแนน 5 ระดับมีค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 เข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.00) มากที่สุด รองลงมาคือ การให้คะแนน 4 ระดับ และ 3 ระดับ ตามลำดับ แต่สำหรับค่าประมาณความสามารถในการสอบครั้งที่ 2 พบว่าทุกกรณีที่แบบทดสอบผสมมีความยากเป็น 0.50 ค่าเฉลี่ยของค่าประมาณความสามารถเมื่อให้คะแนน 3 ระดับในทุกขนาดความยาวเข้าใกล้ค่าเฉลี่ยความสามารถจริง (1.00) มากที่สุด ส่วนกรณีอื่นให้ผลในทำนองเดียวกันกับการทดสอบครั้งที่ 1

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่งพบว่าพัฒนาการความสามารถในทุกเงื่อนไขมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติ โดยเงื่อนไขที่มีความเบ้ขวามากที่สุดคือ กรณีแบบทดสอบยาว 30:10 มีความยาก 0.50 และรูปแบบการให้คะแนนมี 3 ระดับ มีค่าเท่ากับ 1.886 และเงื่อนไขที่มีความโด่งสูงกว่าโค้งปกติมากที่สุดคือกรณีแบบทดสอบยาว 24:8 มีความยาก 0.50 และรูปแบบการให้คะแนนมี 3 ระดับ มีค่าเท่ากับ 16.996 ตามลำดับ และพบว่ามี 2 กรณีที่พัฒนาการความสามารถมีลักษณะการแจกแจงแบบเบ้ซ้ายและโด่งต่ำกว่าโค้งปกติ คือ กรณีแบบทดสอบมีความยาว 30:10 มีความยาก 0.00 และให้คะแนน 3 ระดับ และ 4 ระดับ โดยมีค่าเท่ากับ -0.453 และ -0.928 กับ และ -0.534 และ -0.927 ตามลำดับ

สรุปผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ผลการวิเคราะห์ค่าสถิติพื้นฐานของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันตามระดับพัฒนาการความสามารถจริงทั้งหมด 9 รูปแบบ สรุปได้ว่า ค่าเฉลี่ยของค่าประมาณความสามารถของการสอบทั้งสองครั้งทุกเงื่อนไขมีค่าเป็นลบและระดับพัฒนาการความสามารถมีทั้งค่าที่เป็นบวกและค่าที่เป็นลบ โดยส่วนใหญ่ค่าเฉลี่ยของค่าประมาณความสามารถในการสอบครั้งที่ 2 สูงกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1 ยกเว้นกรณีกลุ่มผู้สอบมีพัฒนาการความสามารถเพิ่มขึ้น 0.5 (M2 และ M7) และ เพิ่มขึ้น 1.0 (M3 และ M9) ค่าประมาณความสามารถในการสอบครั้งที่ 2 มีค่าเฉลี่ยต่ำกว่าค่าประมาณความสามารถในการสอบครั้งที่ 1

เมื่อพิจารณาแต่ละเงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม พบว่ากรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน การให้คะแนน 5 ระดับส่วนใหญ่มีค่าเฉลี่ยของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 จะเข้าใกล้ค่าเฉลี่ยความสามารถจริง (0.00) มากที่สุด รองลงมาคือ การให้คะแนน 4 ระดับ และ 3 ระดับ ตามลำดับ แต่สำหรับค่าประมาณความสามารถในการสอบครั้งที่ 2 กลับให้ผลตรงกันข้าม คือ กรณีที่แบบทดสอบผสมมีความยาว และความยากเท่ากัน การให้คะแนน 3 ระดับส่วนใหญ่ ค่าเฉลี่ยของค่าประมาณความสามารถจะเข้าใกล้ค่าเฉลี่ยความสามารถจริงมากที่สุด ยกเว้นสำหรับกรณีแบบทดสอบยาว 30:10 มีความยาก 0.00 ให้คะแนน 5 ระดับ ค่าเฉลี่ยของค่าประมาณความสามารถจะเข้าใกล้ค่าเฉลี่ยความสามารถจริงมากที่สุด

เมื่อพิจารณาการแจกแจงของพัฒนาการความสามารถจากค่าความเบ้และความโด่งพบว่าพัฒนาการความสามารถในทุกเงื่อนไขมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา และโด่งสูงกว่าโค้งปกติแต่มีบางกรณีที่พัฒนาการความสามารถมีลักษณะการแจกแจงแบบเบ้ซ้ายและโด่งต่ำกว่าโค้งปกติ

ตอนที่ 2 การศึกษาและเปรียบเทียบความสัมพันธ์ของพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ผู้วิจัยได้ศึกษาความสัมพันธ์ระหว่างพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าพารามิเตอร์ความสามารถจริงโดยคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันแล้วแปลงค่าสัมประสิทธิ์สหสัมพันธ์เป็นคะแนนมาตรฐาน Z ของฟิชเชอร์ (Fischer's Z) เพื่อเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์เมื่อเงื่อนไขพัฒนาการความสามารถผู้สอบและลักษณะของแบบทดสอบผสมแตกต่างกัน ผลการศึกษาและเปรียบเทียบนำเสนอเป็นสองส่วน ดังนี้

2.1 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ทุกระดับพัฒนาการความสามารถ และทุกเงื่อนไขของแบบทดสอบผสมที่มีความยาว ความยาก และรูปแบบการให้คะแนนแตกต่างกัน โดยแบ่งเป็น

2.1.1 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_1$) กับค่าความสามารถจริงในการทดสอบครั้งที่ 1 (θ_1)

2.1.2 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2$) กับค่าความสามารถจริงในการทดสอบครั้งที่ 2 (θ_2)

2.2 ผลการเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อพัฒนาการความสามารถผู้สอบแตกต่างกันและลักษณะของแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน

2.2.1 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ความสามารถจริง (θ_1)

2.2.2 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_2$) ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ความสามารถจริง (θ_2)

ผลศึกษาและเปรียบเทียบความสัมพันธ์ของพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าพารามิเตอร์ความสามารถจริง นำเสนอในรูปตาราง และใช้แผนภาพประกอบเพื่อให้มองเห็นแบบแผน และแนวโน้มความสัมพันธ์ ตามลำดับดังนี้

2.1 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ผู้วิจัยศึกษาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับพารามิเตอร์ความสามารถจริงในการทดสอบครั้งที่ 1 และการ

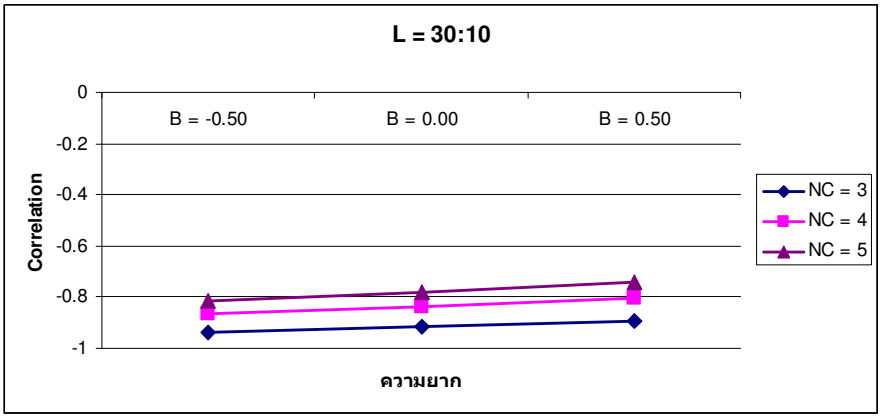
ทดสอบครั้งที่ 2 ทุกระดับพัฒนาการความสามารถ และทุกเงื่อนไขของแบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกัน ผลปรากฏดังนี้

2.1.1 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_1$) กับค่าความสามารถจริงในการทดสอบครั้งที่ 1 (θ_1)

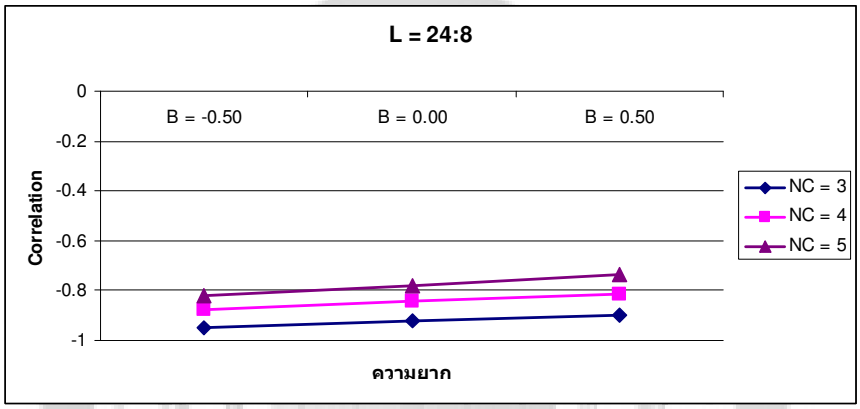
ตาราง 14 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_1$) กับค่าความสามารถจริงในการทดสอบครั้งที่ 1 (θ_1) เมื่อผู้สอบมีพัฒนาการความสามารถต่างกัน และลักษณะแบบทดสอบผสมต่างกัน

L	B	NC	M1	M2	M3	M4	M5	M6	M7	M8	M9
30:10	-0.50	3	-0.939**	-0.939**	-0.939**	-0.940**	-0.939**	-0.939**	-0.939**	-0.939**	-0.939**
		4	-0.867**	-0.867**	-0.867**	-0.867**	-0.867**	-0.867**	-0.867**	-0.867**	-0.867**
		5	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**
	0.00	3	-0.915**	-0.915**	-0.915**	-0.915**	-0.915**	-0.915**	-0.915**	-0.915**	-0.915**
		4	-0.838**	-0.838**	-0.838**	-0.838**	-0.837**	-0.838**	-0.838**	-0.838**	-0.838**
		5	-0.781**	-0.781**	-0.781**	-0.781**	-0.781**	-0.781**	-0.781**	-0.781**	-0.781**
	0.50	3	-0.893**	-0.893**	-0.893**	-0.893**	-0.893**	-0.893**	-0.893**	-0.893**	-0.893**
		4	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**
		5	-0.740**	-0.740**	-0.740**	-0.740**	-0.740**	-0.740**	-0.740**	-0.740**	-0.740**
24:8	-0.50	3	-0.948**	-0.948**	-0.948**	-0.948**	-0.948**	-0.948**	-0.948**	-0.948**	-0.948**
		4	-0.877**	-0.877**	-0.877**	-0.877**	-0.877**	-0.877**	-0.877**	-0.877**	-0.877**
		5	-0.819**	-0.819**	-0.819**	-0.819**	-0.819**	-0.819**	-0.819**	-0.819**	-0.819**
	0.00	3	-0.921**	-0.921**	-0.921**	-0.921**	-0.921**	-0.921**	-0.921**	-0.921**	-0.921**
		4	-0.845**	-0.845**	-0.845**	-0.845**	-0.845**	-0.845**	-0.845**	-0.845**	-0.845**
		5	-0.780**	-0.780**	-0.780**	-0.78**	-0.780**	-0.780**	-0.78**	-0.780**	-0.779**
	0.50	3	-0.899**	-0.899**	-0.899**	-0.899**	-0.899**	-0.899**	-0.899**	-0.899**	-0.899**
		4	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**	-0.812**
		5	-0.736**	-0.735**	-0.735**	-0.735**	-0.735**	-0.735**	-0.735**	-0.735**	-0.735**
15:5	-0.50	3	-0.962**	-0.962**	-0.961**	-0.961**	-0.961**	-0.962**	-0.962**	-0.962**	-0.962**
		4	-0.850**	-0.850**	-0.850**	-0.850**	-0.850**	-0.850**	-0.850**	-0.850**	-0.850**
		5	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**	-0.805**
	0.00	3	-0.951**	-0.951**	-0.950**	-0.950**	-0.950**	-0.951**	-0.951**	-0.951**	-0.951**
		4	-0.842**	-0.842**	-0.842**	-0.842**	-0.842**	-0.842**	-0.842**	-0.842**	-0.842**
		5	-0.792**	-0.792**	-0.792**	-0.792**	-0.792**	-0.792**	-0.792**	-0.792**	-0.792**
	0.50	3	-0.919**	-0.919**	-0.919**	-0.919**	-0.919**	-0.919**	-0.919**	-0.919**	-0.919**
		4	-0.798**	-0.798**	-0.798**	-0.798**	-0.798**	-0.798**	-0.798**	-0.798**	-0.798**
		5	-0.742**	-0.742**	-0.742**	-0.741**	-0.742**	-0.742**	-0.742**	-0.742**	-0.742**

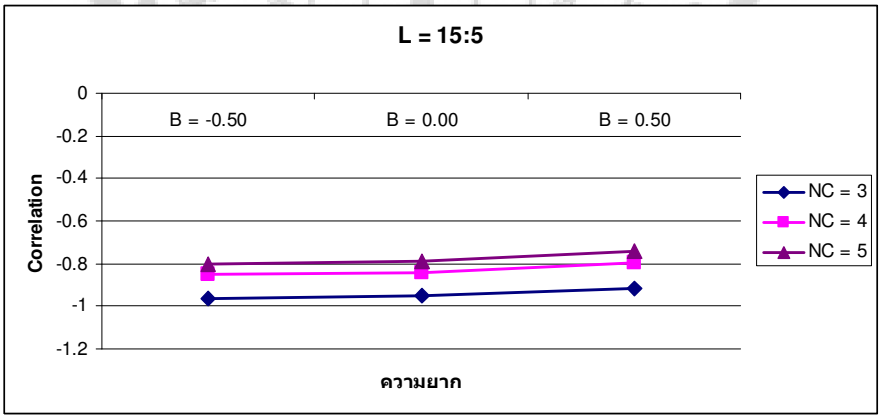
นำผลการวิเคราะห์ข้อมูลตารางมาเขียนเป็นแผนภูมิประกอบได้ดังภาพประกอบ 9 (ก – ค) ภาพประกอบ 10 (ก – ค) และภาพประกอบ 11 (ก – ค)



9 ก

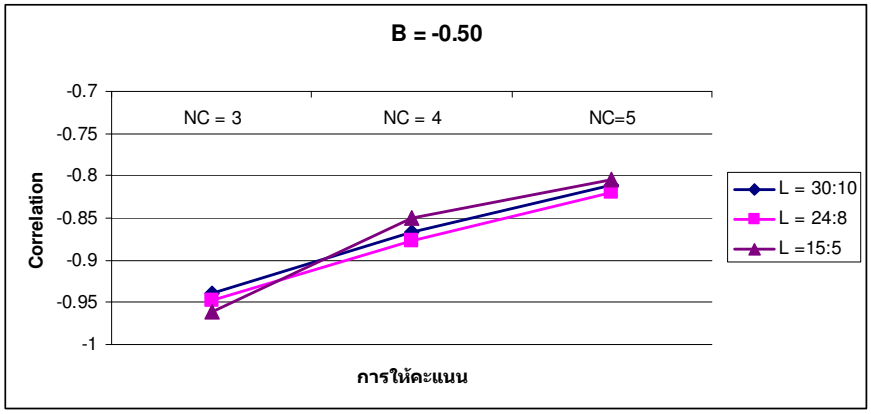


9 ข

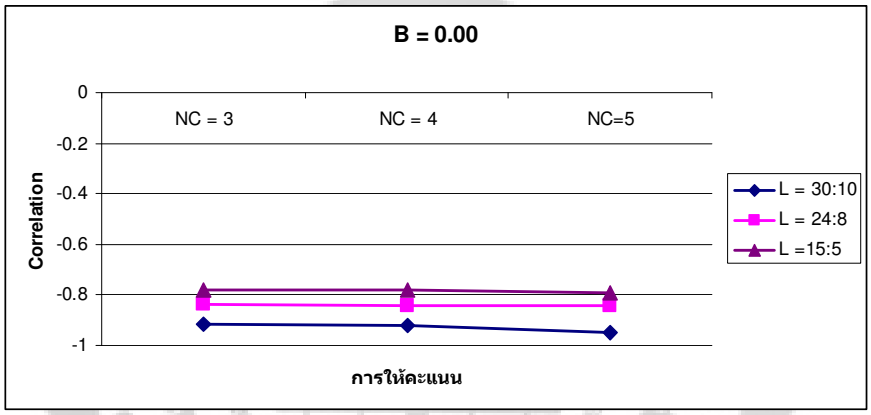


9 ค

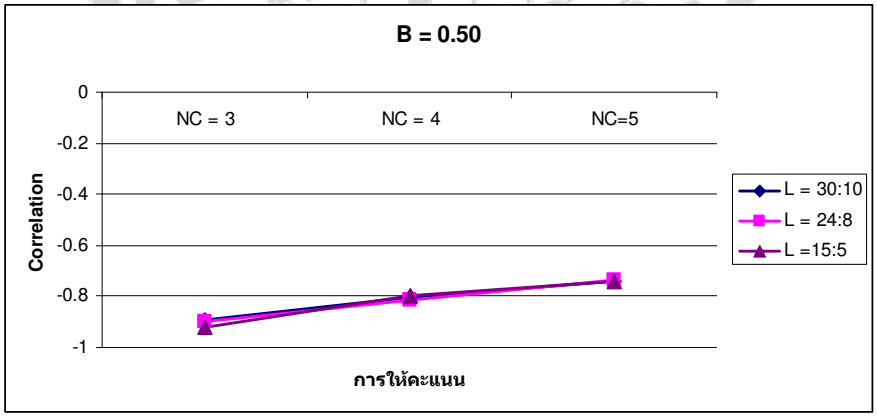
ภาพประกอบ 9 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) เมื่อความยาวของแบบทดสอบผสมคงที่ 30:10 (9 ก) 24:8 (9 ข) และ 15:5 (9 ค)



10 ก

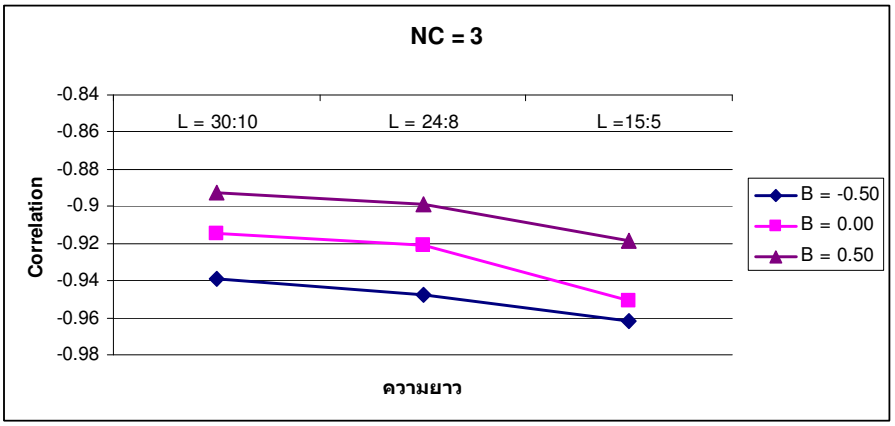


10 ข

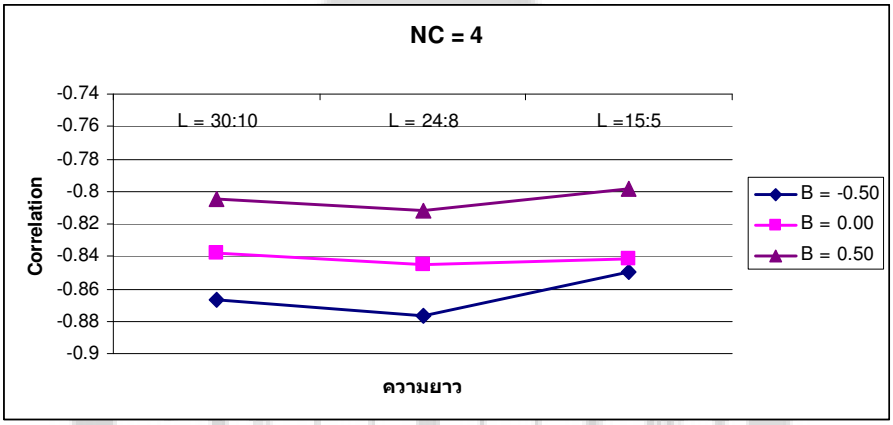


10 ค

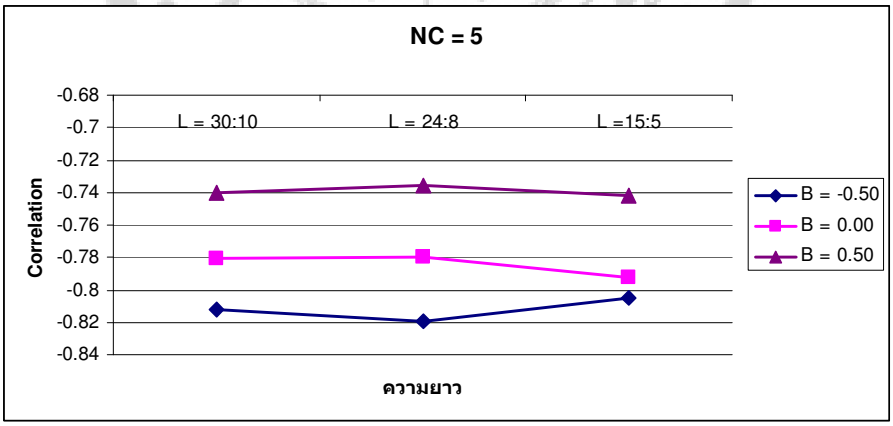
ภาพประกอบ 10 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) เมื่อความยากของแบบทดสอบผสมคงที่ -0.50 (10 ก) 0.00 (10 ข) และ 0.50 (10 ค)



11 ก



11 ข



11 ค

ภาพประกอบ 11 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) เมื่อแบบทดสอบผสมมีการให้คะแนนคงที่ 3 ระดับ (11 ก) 4 ระดับ (11 ข) และ 5 ระดับ (11 ค)

กรณีความยาวของแบบทดสอบผสมคงที่

จากตาราง 14 และภาพประกอบ 9 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาวคงที่ ค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) มีความสัมพันธ์ทางลบในระดับสูงอย่างมีนัยสำคัญที่ระดับ .01 โดยพบว่าภายใต้เงื่อนไขเดียวกันทุกกลุ่มพัฒนาการความสามารถ (M1 – M9) มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากัน ดังนี้

แบบทดสอบผสมที่มีความยาว 30:10 มีค่าสัมประสิทธิ์สหสัมพันธ์ตั้งแต่ –0.944 ถึง –0.724 แบบทดสอบผสมที่มีความยาว 24:8 มีค่าสัมประสิทธิ์สหสัมพันธ์ตั้งแต่ –0.949 ถึง –0.718 และแบบทดสอบผสมความยาว 15:5 มีค่าสัมประสิทธิ์สหสัมพันธ์ตั้งแต่ – 0.960 ถึง – 0.725

ภายใต้เงื่อนไขความยาวและความยากของแบบทดสอบผสมคงที่พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) ของแบบทดสอบผสมที่ให้คะแนน 3 ระดับ สูงกว่า สัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) ของแบบทดสอบผสมให้คะแนน 4 ระดับ และ 5 ระดับ

กรณีความยากของแบบทดสอบผสมคงที่

จากตาราง 14 และ จากภาพประกอบ 10 (ก – ค) เมื่อแบบทดสอบมีความยากคงที่ -0.50 0.00 และ 0.50 พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) มีแนวโน้มไม่แตกต่างกันตามขนาดความยาวของแบบทดสอบ โดยพบว่า

เมื่อแบบทดสอบผสมมีความยาก -0.50 พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) ของแบบทดสอบผสมที่มีความยาว 24:8 และให้คะแนน 5 ระดับ มีค่าต่ำกว่าแบบทดสอบผสมที่มีความยาว 30:10 และ 15:5 เล็กน้อย ส่วนกรณีอื่นไม่แตกต่างกัน

เมื่อแบบทดสอบผสมมีความยาก 0.00 พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 1 (θ_1) กับค่าความสามารถจริง (θ_1) มีแนวโน้มไม่แตกต่างกันตามขนาดความยาวของแบบทดสอบ แต่ค่าสัมประสิทธิ์สหสัมพันธ์ลดลงเมื่อระดับการให้คะแนนเพิ่มขึ้น

เมื่อแบบทดสอบผสมมีความยาก 0.50 พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 1 (θ_1) กับค่าความสามารถจริง (θ_1) ของแบบทดสอบที่มีความยาว 30:10 สูงกว่าแบบทดสอบขนาด 24:8 และ15:5 ทุกระดับการให้คะแนน

กรณีการให้คะแนนของแบบทดสอบผสมคงที่

จากตาราง 14 และจากภาพประกอบ 11 (ก – ค) เมื่อแบบทดสอบมีรูปแบบการให้คะแนนคงที่ 3 4 และ 5 ระดับ พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) แตกต่างกันตามขนาดความยาวและความยากของแบบทดสอบดังนี้

กรณีที่ให้คะแนน 3 ระดับ 4 ระดับ และ 5 ระดับ แบบทดสอบผสมที่มีค่าความยาก -0.50 มีแนวโน้มที่สัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการสอบครั้งที่ 1 ที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าความสามารถจริงสูงกว่าแบบทดสอบที่มีความยากเป็น 0.00 และ 0.50 ตามลำดับในทุกระดับความยาว

กรณีที่ให้คะแนน 3 ระดับของทุกระดับเงื่อนไขความยาก เมื่อแบบทดสอบมีความยาวลดลงสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ที่ได้จากการประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าความสามารถจริงมีค่าสูงขึ้น แต่สำหรับทุกระดับเงื่อนไขความยากกรณีที่มีการให้คะแนน 4 ระดับและ 5 ระดับ เมื่อแบบทดสอบมีความยาวลดลงค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ที่ได้จากการประมาณค่าด้วยวิธีประมาณค่าพารามิเตอร์พร้อมกันกับค่าความสามารถจริงลดลง

สรุปความสัมพันธ์ของพารามิเตอร์ความสามารถในการสอบครั้งที่ 1 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ค่าประมาณความสามารถในการสอบครั้งที่ 1 ($\hat{\theta}_1$) ที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าความสามารถจริง (θ_1) มีความสัมพันธ์ทางลบในระดับสูงอย่างนัยสำคัญทางสถิติที่ระดับ .01 ทุกเงื่อนไข (r มีค่าตั้งแต่ -0.939 ถึง -0.740) และพบว่าทุกกลุ่มระดับพัฒนาการความสามารถมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากันภายใต้เงื่อนไขแบบทดสอบผสมที่ความยาว ความยาก และระดับการให้คะแนนเท่ากัน แต่พบว่าเมื่อเงื่อนไขของแบบทดสอบผสมมีความยาว ความยากและการให้คะแนนแตกต่างกันค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้มีค่าแตกต่างกัน เมื่อระดับการให้คะแนนเพิ่มขึ้นค่าสัมประสิทธิ์สหสัมพันธ์มีค่าลดลง เมื่อความยากเพิ่มขึ้นสัมประสิทธิ์สหสัมพันธ์มีค่าลดลง แต่ค่าสัมประสิทธิ์สหสัมพันธ์ไม่เพิ่มขึ้นหรือลดลงตามขนาดความยาวของแบบทดสอบผสม

2.1.2 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2$) กับค่าความสามารถจริงในการทดสอบครั้งที่ 2 (θ_2)

ตาราง 15 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2$) กับค่าความสามารถจริงในการทดสอบครั้งที่ 2 (θ_2) เมื่อผู้สอบมีพัฒนาการความสามารถต่างกัน และลักษณะแบบทดสอบผสมต่างกัน

L	B	NC	M1	M2	M3	M4	M5	M6	M7	M8	M9
		3	-0.941**	-0.967**	-0.981**	-0.948**	-0.972**	-0.988**	-0.937**	-0.960**	-0.974**
	-0.50	4	-0.866**	-0.925**	-0.963**	-0.869**	-0.933**	-0.975**	-0.866**	-0.917**	-0.951**
		5	-0.810**	-0.897**	-0.957**	-0.820**	-0.913**	-0.971**	-0.817**	-0.888**	-0.945**
		3	-0.895**	-0.931**	-0.963**	-0.912**	-0.938**	-0.968**	-0.891**	-0.926**	-0.957**
30:10	0.00	4	-0.804**	-0.883**	-0.944**	-0.914**	-0.893**	-0.955**	-0.809**	-0.880**	-0.934**
		5	-0.752**	-0.862**	-0.942**	-0.755**	-0.884**	-0.957**	-0.758**	-0.852**	-0.927**
		3	-0.881**	-0.925**	-0.963**	-0.900**	-0.938**	-0.971**	-0.871**	-0.917**	-0.956**
	0.50	4	-0.771**	-0.874**	-0.945**	-0.779**	-0.890**	-0.960**	-0.776**	-0.863**	-0.932**
		5	-0.687**	-0.829**	-0.926**	-0.686**	-0.857**	-0.946**	-0.978**	-0.815**	-0.910**
		3	-0.929**	-0.956**	-0.976**	-0.937**	-0.961**	-0.981**	-0.925**	-0.952**	-0.970**
	-0.50	4	-0.872**	-0.927**	-0.963**	-0.876**	-0.933**	-0.974**	-0.874**	-0.921**	-0.952**
		5	-0.817**	-0.895**	-0.955**	-0.820**	-0.911**	-0.968**	-0.818**	-0.889**	-0.944**
		3	-0.907**	-0.941**	-0.970**	-0.918**	-0.947**	-0.976**	-0.900**	-0.936**	-0.965**
24:8	0.00	4	-0.811**	-0.892**	-0.950**	-0.816**	-0.902**	-0.962**	-0.815**	-0.886**	-0.939**
		5	-0.731**	-0.849**	-0.929**	-0.736**	-0.869**	-0.948**	-0.742**	-0.839**	-0.919**
		3	-0.878**	-0.923**	-0.961**	-0.892**	-0.933**	-0.970**	-0.870**	-0.914**	-0.954**
	0.50	4	-0.765**	-0.868**	-0.943**	-0.770**	-0.883**	-0.955**	-0.772**	-0.859**	-0.930**
		5	-0.668**	-0.817**	-0.919**	-0.669**	-0.845**	-0.937**	-0.680**	-0.806**	-0.903**
		3	-0.949**	-0.967**	-0.980**	-0.951**	-0.970**	-0.984**	-0.945**	-0.963**	-0.974**
	-0.50	4	-0.858**	-0.917**	-0.958**	-0.864**	-0.925**	-0.969**	-0.862**	-0.913**	-0.947**
		5	-0.765**	-0.864**	-0.940**	-0.752**	-0.874**	-0.953**	-0.783**	-0.863**	-0.928**
		3	-0.926**	-0.953**	-0.974**	-0.940**	-0.960**	-0.978**	-0.920**	-0.947**	-0.969**
15:5	0.00	4	-0.824**	-0.898**	-0.953**	-0.834**	-0.908**	-0.962**	-0.826**	-0.892**	-0.941**
		5	-0.716**	-0.835**	-0.926**	-0.695**	-0.850**	-0.942**	-0.731**	-0.830**	-0.914**
		3	-0.891**	-0.931**	-0.965**	-0.908**	-0.941**	-0.970**	-0.881**	-0.922**	-0.958**
	0.50	4	-0.778**	-0.873**	-0.944**	-0.790**	-0.889**	-0.954**	-0.776**	-0.864**	-0.931**
		5	-0.649**	-0.802**	-0.912**	-0.625**	-0.822**	-0.932**	-0.670**	-0.797**	-0.896**

** มีนัยสำคัญที่ระดับ .01

และนำผลการวิเคราะห์ข้อมูลตารางมาเขียนเป็นแผนภูมิประกอบได้ดังภาพประกอบ 12 (ก – ค) ภาพประกอบ 13 (ก – ค) และภาพประกอบ 14 (ก – ค)

จากตาราง 15 พบว่าค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) มีความสัมพันธ์ทางลบในระดับสูงอย่างมีนัยสำคัญที่ระดับ .01 โดยภายใต้เงื่อนไขเดียวกันค่าสัมประสิทธิ์สหสัมพันธ์มีความแตกต่างกันตามกลุ่มระดับพัฒนาการ ค่าสัมประสิทธิ์สหสัมพันธ์สูงสุดในแต่ละกลุ่มพัฒนาการ M1 M2 M3 M4 M5 M6 M7 M8 และ M9 มีค่าเท่ากับ -0.941 -0.967-0.981 -0.948 -0.972 -0.988 -0.937 -0.960 และ -0.974 ตามลำดับ และค่าสัมประสิทธิ์สหสัมพันธ์ต่ำสุดในแต่ละกลุ่มพัฒนาการ M1 M2 M3 M4 M5 M6 M7 M8 และ M9 มีค่าเท่ากับ-0.649 -0.802 -0.912 -0.625 -0.822 -0.932 -0.670 -0.797 และ -0.896 ตามลำดับ

กรณีความยากของแบบทดสอบผสมคงที่

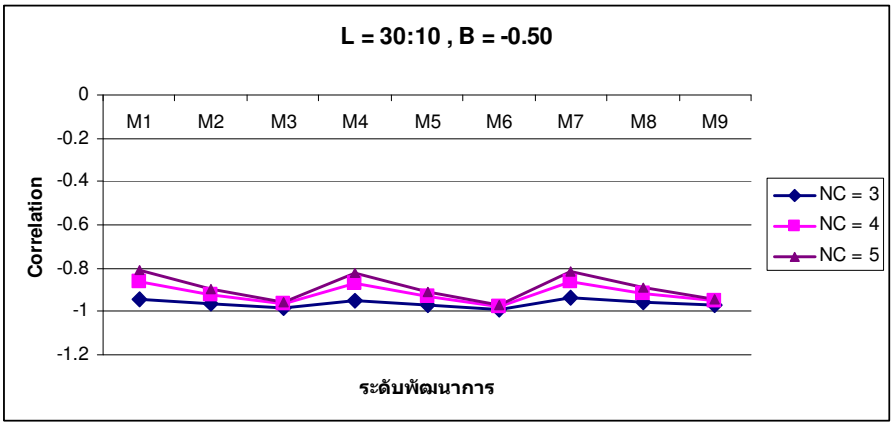
จากภาพประกอบ 12 (ก – ค) ภาพประกอบ 13 (ก – ค) และภาพประกอบ 14 (ก – ค) พบว่า เมื่อแบบทดสอบผสมมีความยาวและความยากคงที่ กรณีให้คะแนน 3 ระดับ มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ในทางลบสูงกว่ากรณีให้คะแนน 4 ระดับ และ 5 ระดับ ทุกกลุ่มพัฒนาการความสามารถ

กรณีความยากของแบบทดสอบผสมคงที่

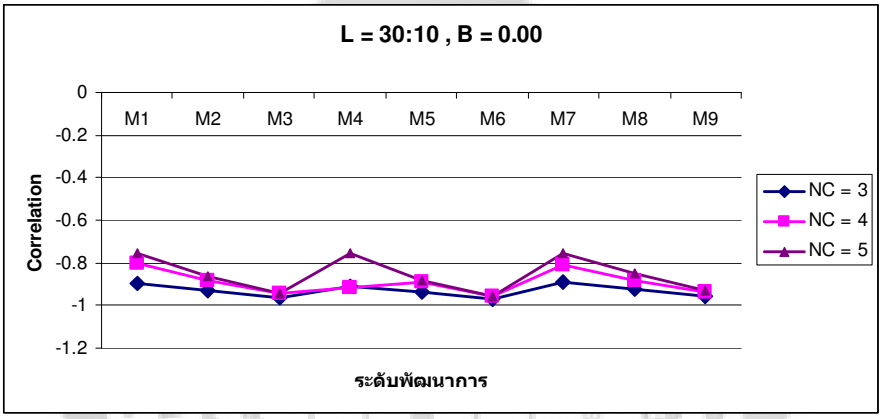
จากภาพประกอบ 15 (ก – ค) พบว่า เมื่อความยากเป็น -0.50 และมีการให้คะแนน 3 ระดับ แบบทดสอบผสมที่มีความยาว 24:8 มีค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ในทางลบต่ำกว่าแบบทดสอบผสมที่มีความยาว 30:10 และ 15:5 ทุกระดับพัฒนาการ แต่พบว่าเมื่อความยากเป็น -0.50 และให้คะแนน 4 ระดับ และ 5 ระดับ แบบทดสอบผสมที่มีความยาว 24:8 มีค่าสูงกว่าแบบทดสอบผสมมีความยาว 30:10 และ 15:5 ทุกระดับพัฒนาการ

จากภาพประกอบ 16 (ก – ค) พบว่า เมื่อความยากเป็น 0.00 และมีการให้คะแนน 3 ระดับ และ 4 ระดับ แบบทดสอบผสมที่มีความยาว 15:5 ค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2)ในทางลบสูงกว่า แบบทดสอบผสมที่มีความยาว 24:8 และ 30:10 ทุกกลุ่มพัฒนาการความสามารถ แต่พบว่าเมื่อความยากเป็น 0.00 และมีการให้คะแนน 5 ระดับ แบบทดสอบผสมทุกขนาดความยาว มีค่าใกล้เคียงกันทุกระดับพัฒนาการ

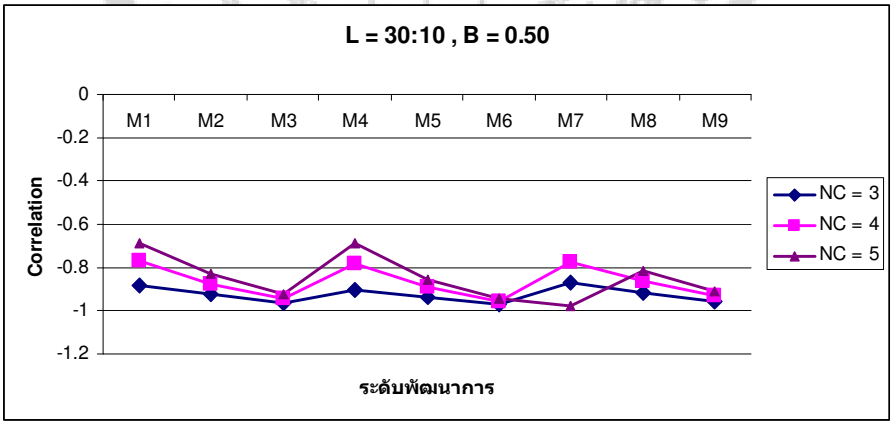
จากภาพประกอบ 17 (ก – ค) พบว่า เมื่อความยาก 0.50 และทุกระดับการให้คะแนน แบบทดสอบผสมทุกขนาดความยาวมีค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ไม่แตกต่างกันทุกกลุ่มพัฒนาการความสามารถ



12 ก

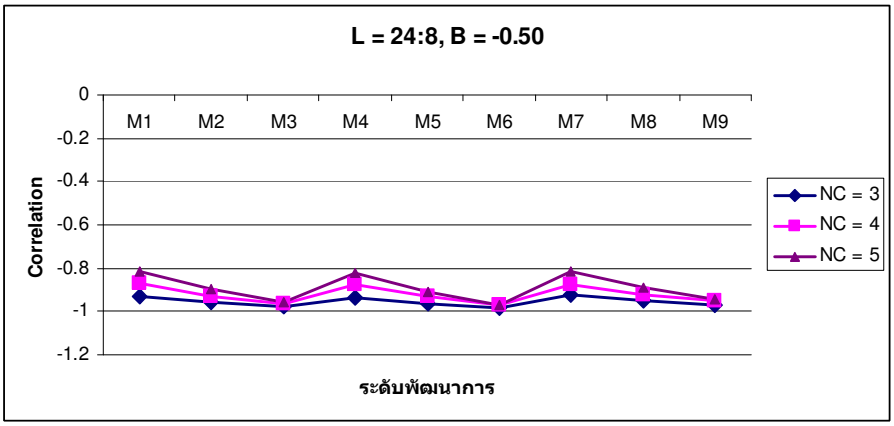


12 ข

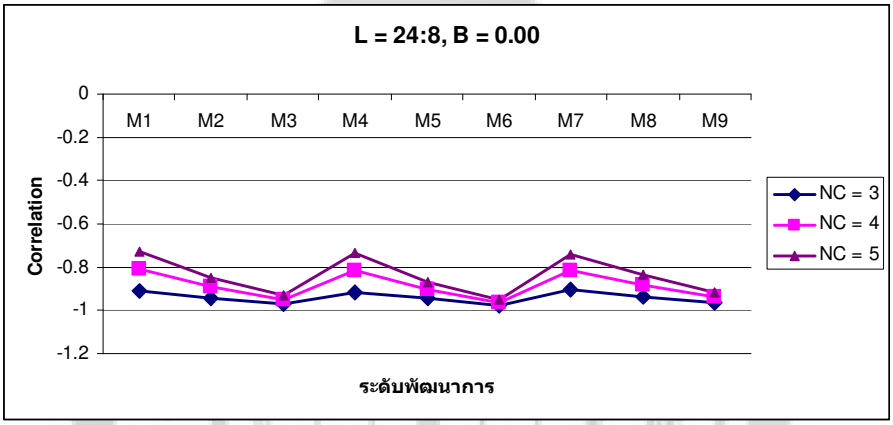


12 ค

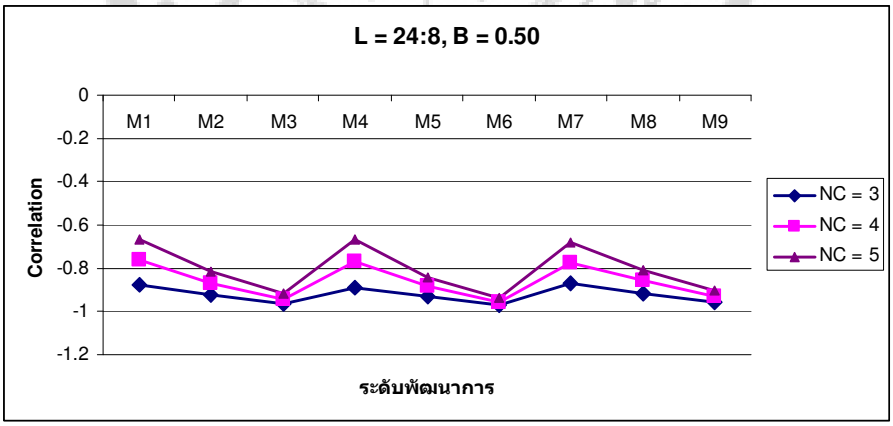
ภาพประกอบ 12 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยาวเป็น 30:10 และความยากของแบบทดสอบเป็น -0.50 (12 ก) 0.00(12 ข) 0.50(12 ค)



13 ก

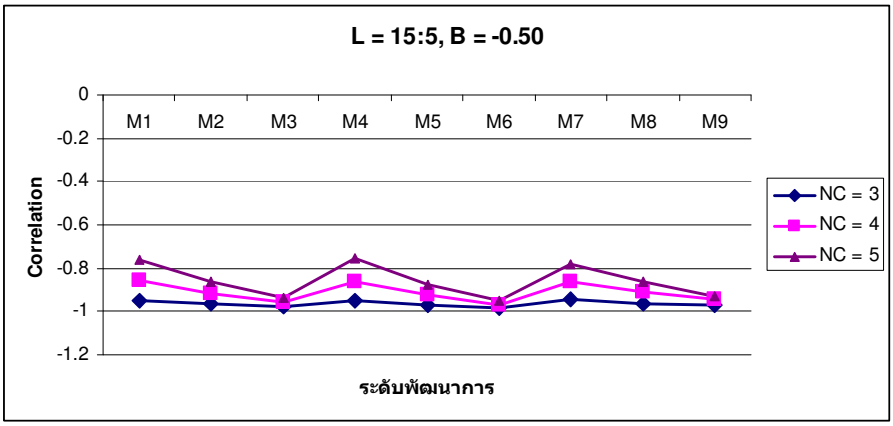


13 ข

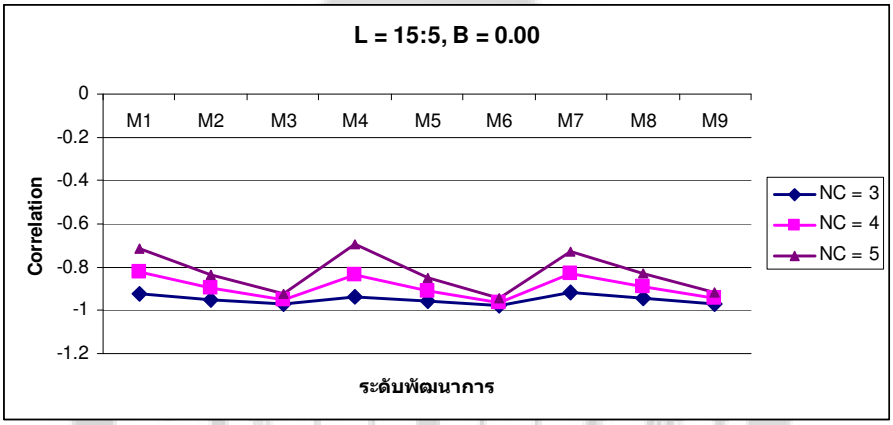


13 ค

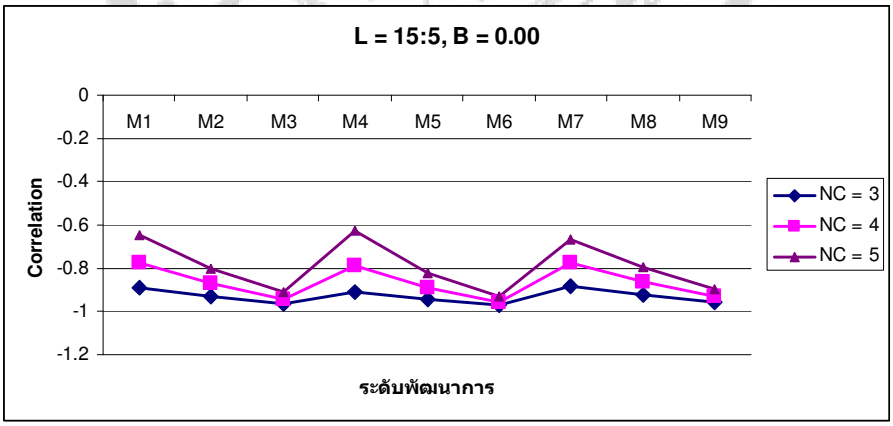
ภาพประกอบ 13 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยาวเป็น 24:8 และความยากของแบบทดสอบเป็น -0.50 (13 ก) 0.00 (13 ข) 0.50(13 ค)



14 ก

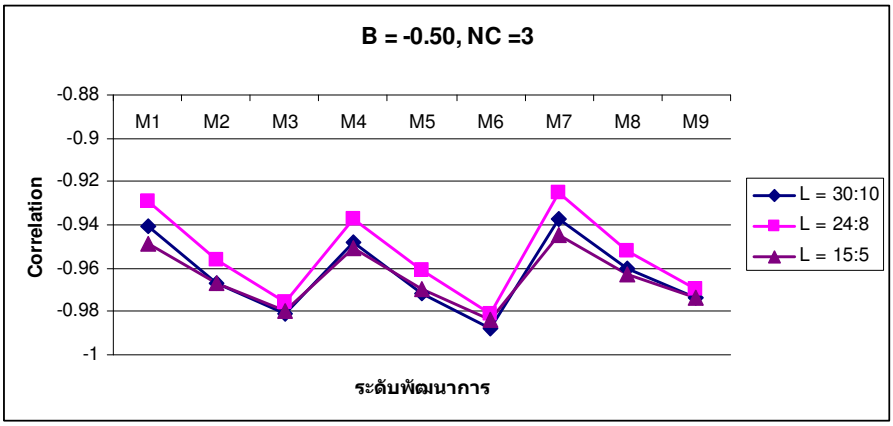


14 ข

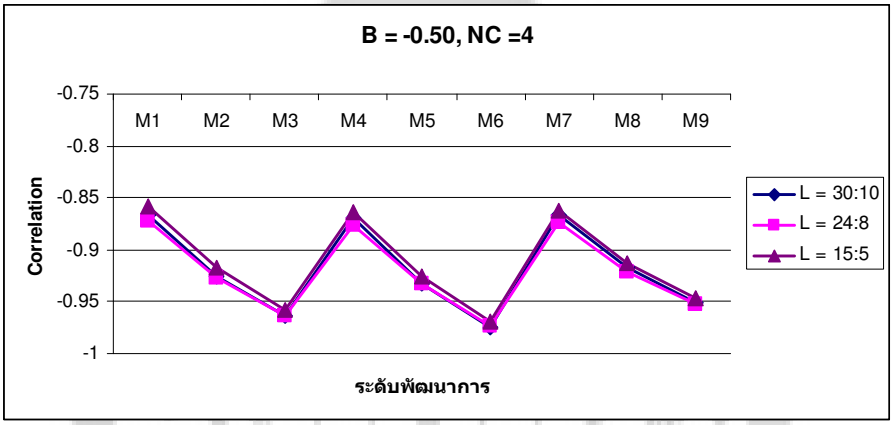


14 ค

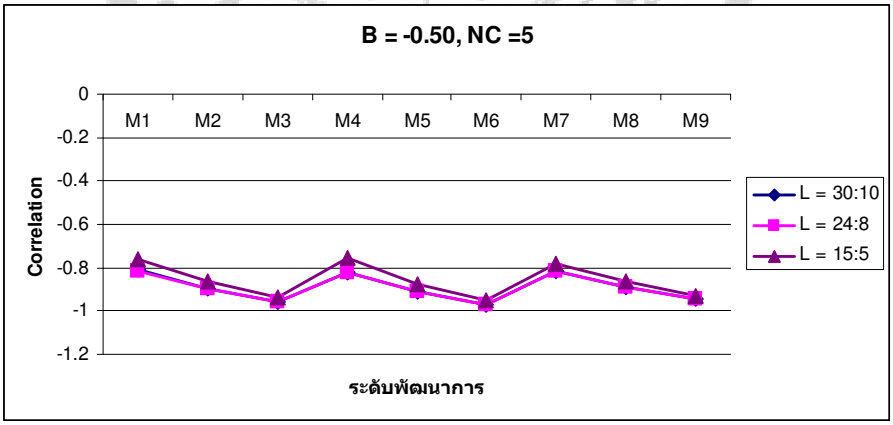
ภาพประกอบ 14 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยาวเป็น 15:5 และความยากของแบบทดสอบเป็น -0.50 (14 ก) 0.00 (14 ข) 0.50 (14 ค)



15 ก

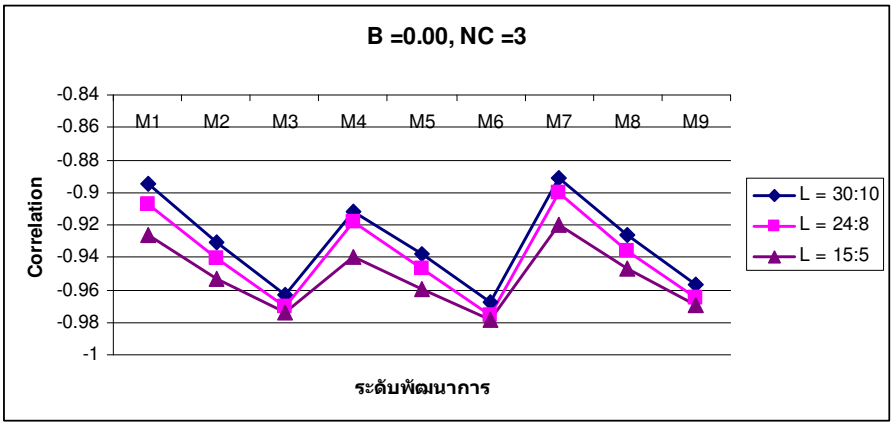


15 ข

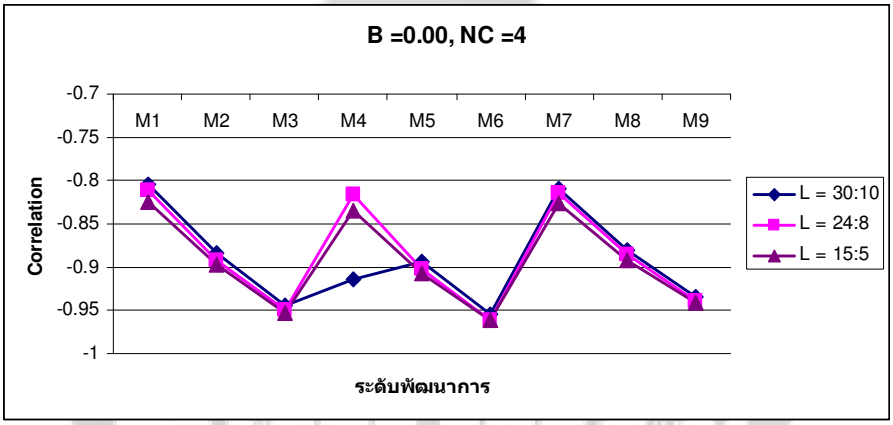


15 ค

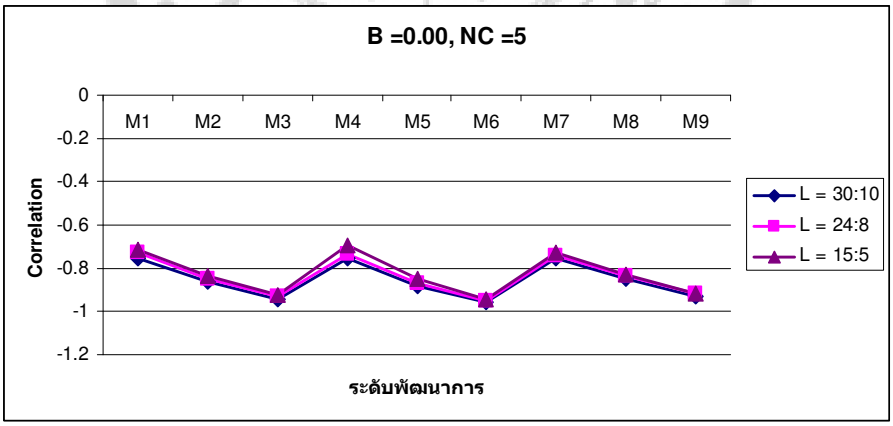
ภาพประกอบ 15 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยากของแบบทดสอบผสมเป็น -0.50 และให้คะแนน 3 ระดับ (15 ก) 4 ระดับ (15 ข) 5 ระดับ (15 ค)



16 ก

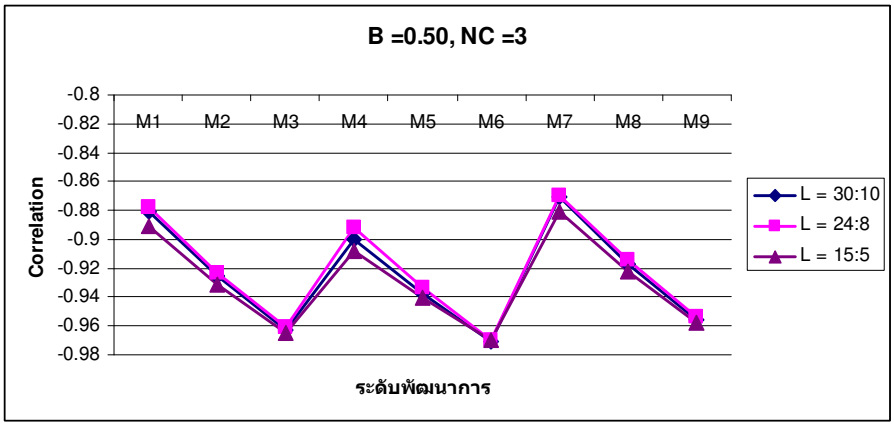


16 ข

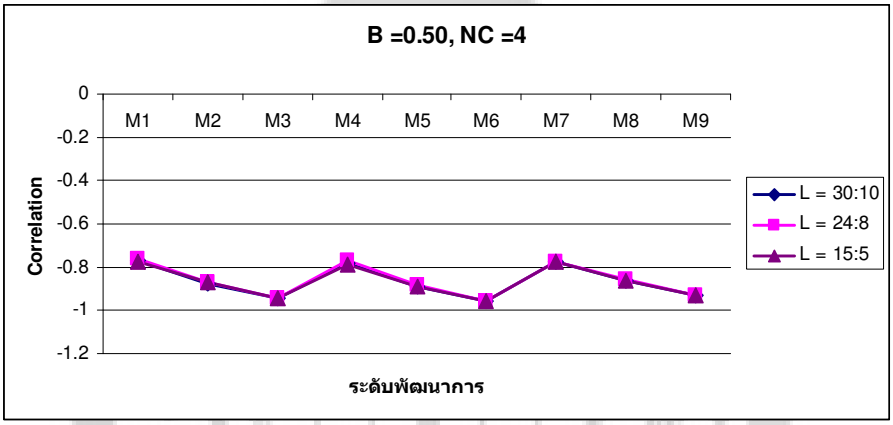


16 ค

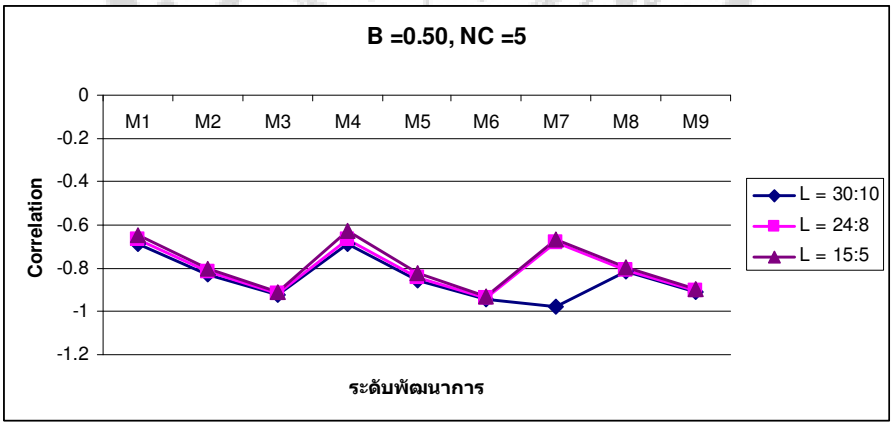
ภาพประกอบ 16 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยากของแบบทดสอบผสมเป็น 0.00 และให้คะแนน 3 ระดับ (16 ก) 4ระดับ (16 ข) 5ระดับ (16 ค)



17 ก

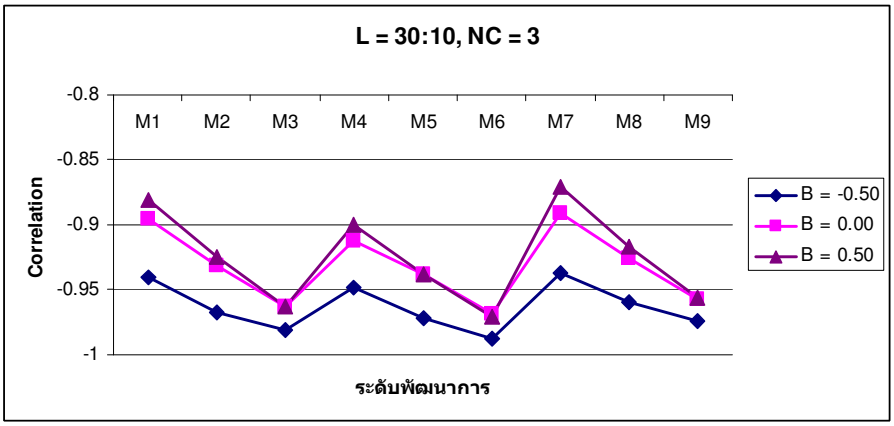


17 ข

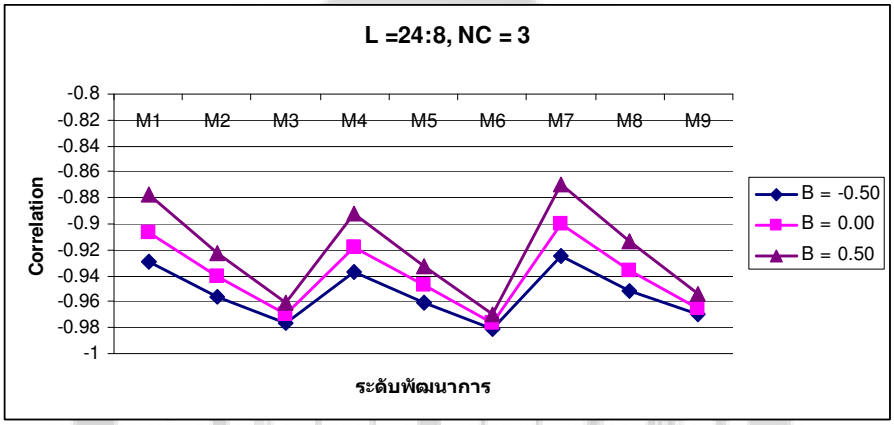


17 ค

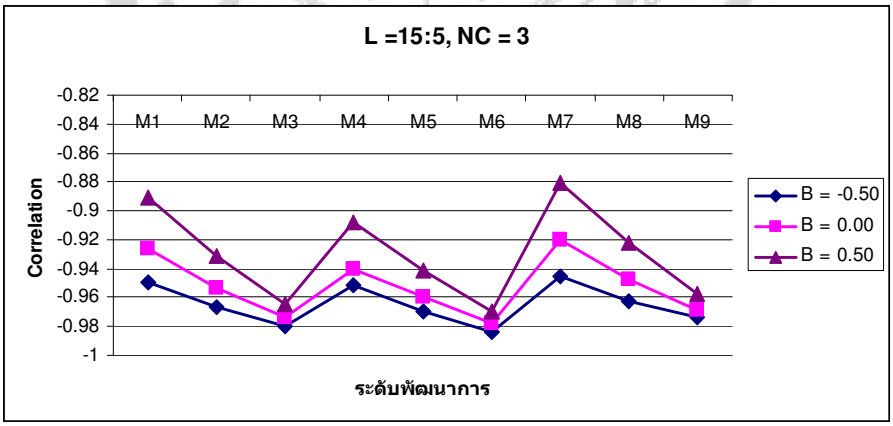
ภาพประกอบ 17 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อความยากของแบบทดสอบผสมเป็น 0.50 และให้คะแนน 3 ระดับ (17 ก) 4 ระดับ (17 ข) 5 ระดับ (17 ค)



18 ก

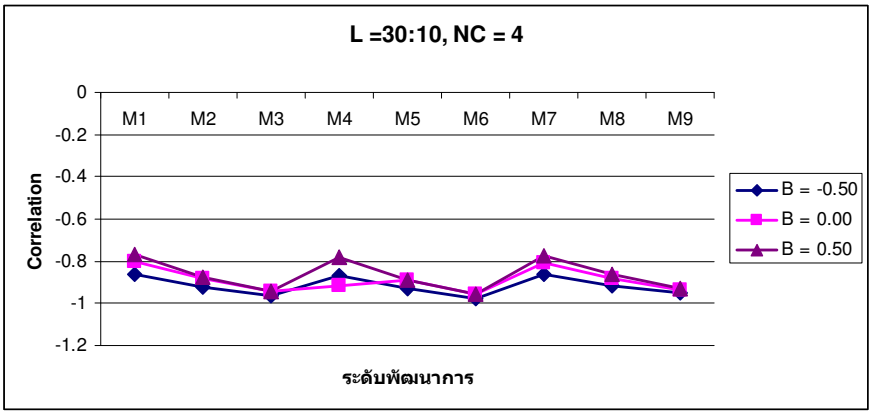


18 ข

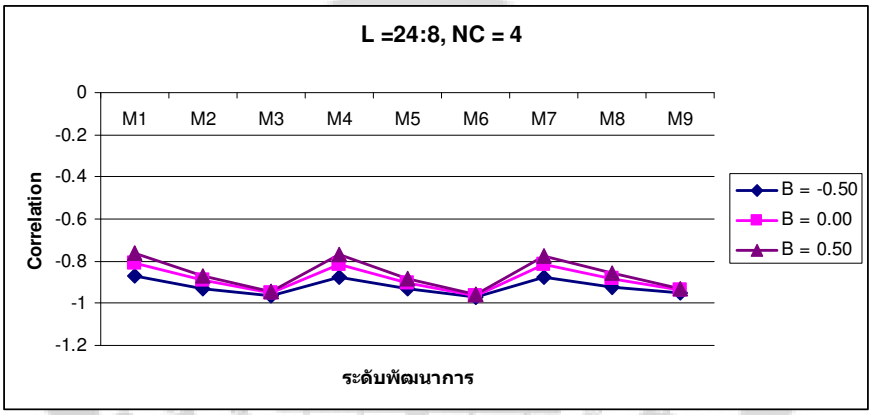


18 ค

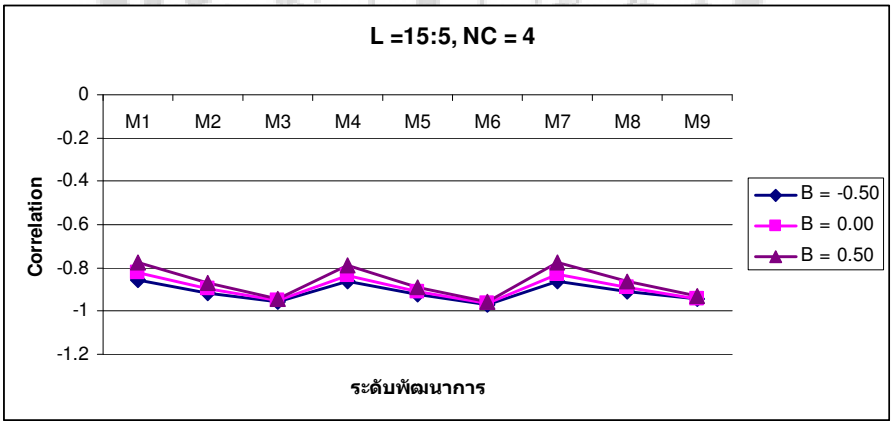
ภาพประกอบ 18 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อให้คะแนน 3 ระดับ และความยาวเป็น 30:10 (18 ก) 24:8 (18 ข) 15:5 (18 ค)



19 ก

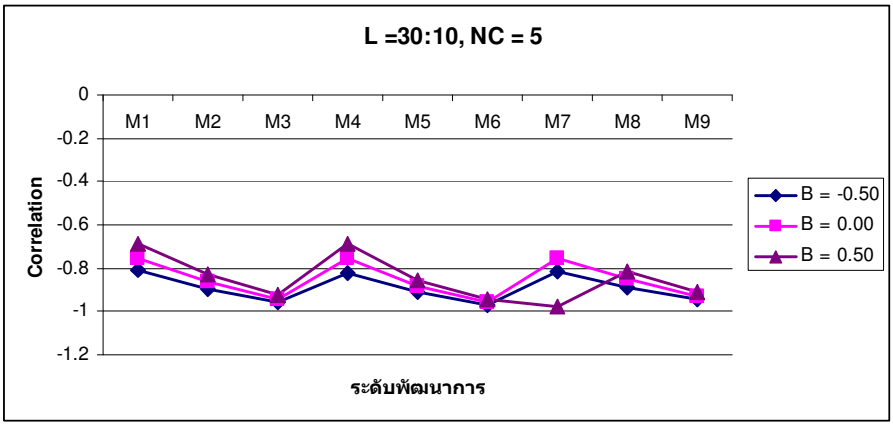


19 ข

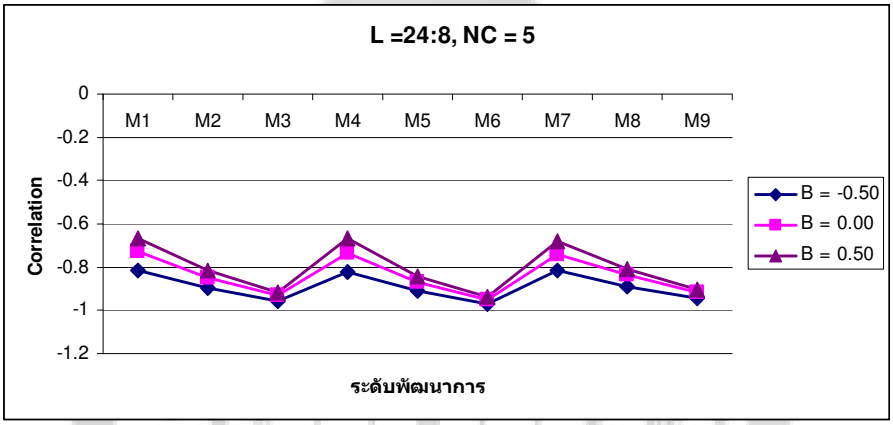


19 ค

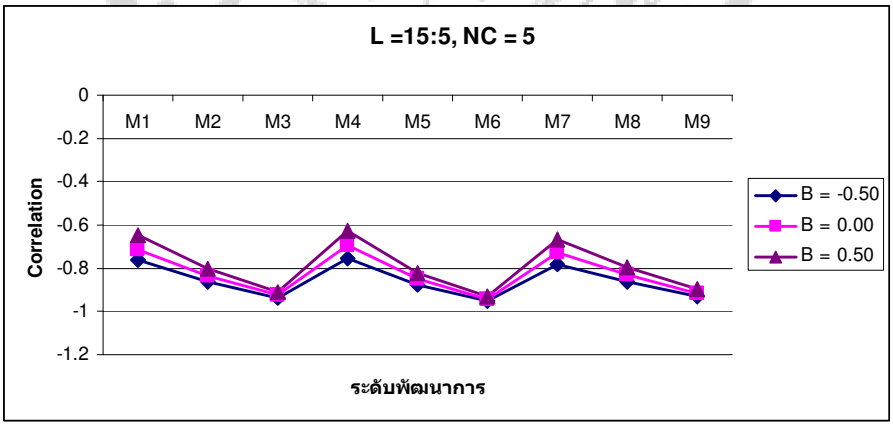
ภาพประกอบ 19 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อให้คะแนน 4 ระดับ และความยาวเป็น 30:10 (19 ก) 24:8 (19 ข) 15:5 (19 ค)



20 ก



20 ข



20 ค

ภาพประกอบ 20 (ก – ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) เมื่อให้คะแนน 5 ระดับ และความยาวเป็น 30:10 (20 ก) 24:8 (20 ข) 15:5 (20 ค)

กรณีการให้คะแนนของแบบทดสอบผสมคงที่

จากภาพประกอบ 18 (ก – ค) ภาพประกอบ 19 (ก – ค) และภาพประกอบ 20 (ก – ค) พบว่า ทุกระดับการให้คะแนน และทุกขนาดความยาวแบบทดสอบ แบบทดสอบผสมที่มีความยาก -0.50 มีค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2)ในทางลบสูงกว่า แบบทดสอบผสมที่มีความยาก 0.50 และ 0.00 ทุกกลุ่มพัฒนาการความสามารถ

สรุปความสัมพันธ์ของพารามิเตอร์ความสามารถในการสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) มีความสัมพันธ์ทางลบในระดับสูงอย่างมีนัยสำคัญที่ระดับ .01 ทุกเงื่อนไข และพบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ในแต่ละกลุ่มพัฒนาการภายใต้เงื่อนไขเดียวกัน มีค่าไม่เท่ากัน โดยพบว่า

แบบทดสอบผสมที่มีความยาว 30:10 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ในทางลบสูงกว่า แบบทดสอบผสมมีความยาว 24:8 และ 15:5 ในทิศทางลบทุกกลุ่มพัฒนาการความสามารถ

เมื่อแบบทดสอบผสมที่มีความยาก -0.50 มีค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ในทางลบสูงกว่า แบบทดสอบผสมที่มีความยาก 0.00 และ 0.50 ในทิศทางลบทุกกลุ่มพัฒนาการความสามารถ

แบบทดสอบผสมให้คะแนน 3 ระดับ มีค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ในทางลบสูงกว่า แบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 5 ระดับ ในทิศทางลบทุกกลุ่มพัฒนาการความสามารถ

2.2 ผลการเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ตามเงื่อนไขพัฒนาการความสามารถผู้สอบและลักษณะของแบบทดสอบผสมที่มีความยาว ความยาก และรูปแบบการให้คะแนนแตกต่างกัน

2.2.3 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันความสามารถจริง (θ_1)

ผู้วิจัยได้เปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 1 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ภายใต้เงื่อนไขพัฒนาการความสามารถผู้สอบ และลักษณะของแบบทดสอบผสมที่มีความยาว ความยาก และรูปแบบการให้คะแนนแตกต่างกัน เพื่อตรวจสอบสมมติฐานข้อ 2 ที่ตั้งไว้ว่า *ค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความสัมพันธ์ระดับสูงกับค่าความสามารถจริงที่กำหนดในทุกระดับพัฒนาการความสามารถ*

การทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันเริ่มจากการแปลงค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันเป็นคะแนนมาตรฐานของฟิชเชอร์ Z ทุกเงื่อนไขที่ศึกษา แต่เนื่องจากค่าฟิชเชอร์ Z ในแต่ละกลุ่มพัฒนาการมีค่าเท่ากันภายใต้เงื่อนไขเดียวกัน ผู้วิจัยจึงทดสอบโดยวิเคราะห์ความแปรปรวนทางเดียววัดซ้ำ (Repeated Measure Analysis of Variance) แยกตามเงื่อนไขความยาว ความยาก และการให้คะแนนแบบทดสอบผสม ปรากฏผลดังตาราง 16

ผลการวิเคราะห์ตามตาราง 16 ผลการทดสอบตัวแปรภายในกลุ่มพบว่าเมื่อแบบทดสอบผสมมีความยากต่างกัน และการให้คะแนนแตกต่างกันค่าเฉลี่ยฟิชเชอร์ Z ของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่พบว่าเมื่อความยาวของแบบทดสอบผสมต่างกันค่าเฉลี่ยฟิชเชอร์ Z ของพัฒนาการความสามารถแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 และผลการทดสอบพบว่าไม่มีปฏิสัมพันธ์ระหว่างความยาว กับ ความยากของแบบทดสอบ ความยาวและการให้คะแนนของแบบทดสอบ ความยาก และการให้คะแนนของแบบทดสอบ และ ความยาว ความยากและการให้คะแนนของแบบทดสอบ ที่ระดับนัยสำคัญทางสถิติ .05 แสดงว่าค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถของผู้สอบแตกต่างกันเมื่อแบบทดสอบผสมมีความยาก และการให้คะแนนแตกต่างกัน

ตาราง 16 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ยฟิชเชอร์ Z ของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 1 ($\hat{\theta}_1$) ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันความสามารถจริง (θ_1) กรณีแบบทดสอบผสมมีความยาว ความยาก และ การให้คะแนนแตกต่างกัน

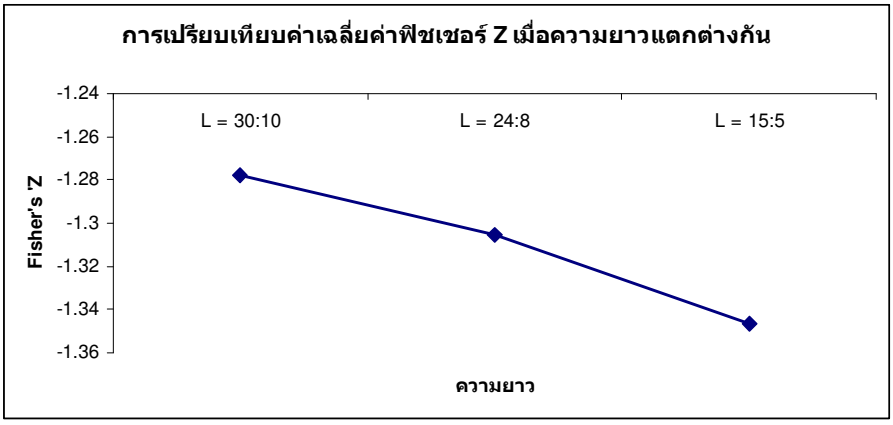
แหล่งความแปรปรวน	SS	df	MS	F	ระดับนัยสำคัญ	ขนาดของผล
ทดสอบผลตัวแปรภายในกลุ่ม (Within)						
ความยาว	0.021	1.056	0.020	2.215	0.173	0.217
ความคลาดเคลื่อน	0.077	8.453	0.009			
ความยาก	0.264	2	0.132	61.931*	0.000	0.886
ความคลาดเคลื่อน	0.034	16	0.002			
การให้คะแนน	1.844	1.040	1.772	158.186*	0.000	0.951
ความคลาดเคลื่อน	0.093	8.322	0.011			

* มีนัยสำคัญทางสถิติที่ระดับ .05

เมื่อผลการทดสอบพบว่าแบบทดสอบผสมมีความยาก การให้คะแนนแตกต่างกัน ส่งผลให้ค่าเฉลี่ยฟิชเชอร์ Z แตกต่างกัน ผู้วิจัยจึงเปรียบเทียบความแตกต่างของค่าฟิชเชอร์ Z ของค่าประมาณพารามิเตอร์ ในแต่ละคู่ระดับของเงื่อนไขความยาก และแต่ละคู่ระดับการให้คะแนน พร้อมทั้งนำเสนอภาพประกอบเพื่อให้เห็นแนวโน้มของค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละเงื่อนไข รวมทั้งกรณีที่ความยาวของแบบทดสอบผสมแตกต่างกันด้วย ผลการเปรียบเทียบรายคู่จึงนำเสนอเป็น 3 ตอนเพื่อทดสอบสมมติฐานย่อยของสมมติฐานการวิจัยข้อ 1 จำนวน 3 ข้อ ดังนี้

2.2.1.1 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 เมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

เนื่องจากค่าเฉลี่ยฟิชเชอร์ Z ของพัฒนาการความสามารถเมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 เพื่อให้เห็นแนวโน้มของค่าเฉลี่ยฟิชเชอร์ Z ผู้วิจัยจึงเสนอภาพประกอบ 21



ภาพประกอบ 21 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถของผู้สอบในการทดสอบครั้งที่ 1 เมื่อแบบทดสอบมีความยาวแตกต่างกัน 3 ระดับ

จากภาพประกอบ 21 พบว่าค่าเฉลี่ยฟิชเชอร์ Z ของแบบทดสอบผสมทุกขนาดมีค่าเป็นลบโดยที่แบบทดสอบผสมที่มีความยาว 15:5 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือแบบทดสอบผสมที่มีความยาว 24:8 และ 30:10 ตามลำดับ

2.2.1.2 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 เมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถเพื่อทดสอบสมมติฐานข้อ 1.2 ที่ตั้งไว้ว่า *แบบทดสอบผสมที่ระดับความยากของข้อสอบสอดคล้องกับระดับความสามารถผู้สอบ จะมีค่าความสัมพันธ์สูงกว่าแบบทดสอบผสมที่ระดับการแจกแจงความยากของข้อสอบไม่สอดคล้องกับระดับความสามารถผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนนของแบบทดสอบผสม* ปรากฏผลการวิเคราะห์ดังตาราง 17

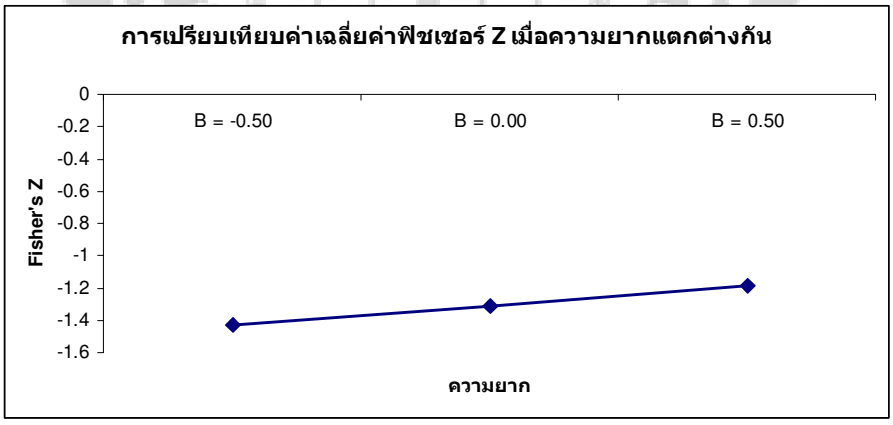
จากตาราง 17 พบว่า เมื่อแบบทดสอบมีความยาวและระดับการให้คะแนนคงที่ ค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถครั้งที่ 1 เมื่อแบบทดสอบผสมมีความยากต่างกันมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือ แบบทดสอบที่มีความยาก 0.00 และ 0.50 ตามลำดับ เมื่อเปรียบเทียบค่าเฉลี่ยรายคู่ พบว่าทุกคู่มีค่าเฉลี่ยแตกต่างกัน โดยแบบทดสอบที่มีความยาก -0.50 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากกว่ากว่าแบบทดสอบที่มีความยาก 0.00 และ 0.50 และแบบทดสอบที่ให้มีความยาก 0.00 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากกว่าแบบทดสอบที่มีความยาก 0.50 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 17 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาก	Mean	B = -0.50	B = 0.00	B = 0.50
	B = -0.50	-1.428	-	0.112*	0.242*
	B = 0.00	-1.316		-	0.130*
	B = 0.50	-1.186			-

* มีนัยสำคัญทางสถิติที่ระดับ 0.05

เพื่อให้เห็นความแตกต่างของค่าเฉลี่ยฟิชเชอร์ Z เมื่อแบบทดสอบมีความยากแตกต่างกัน 3 ระดับผู้วิจัยจึงนำเสนอเป็นแผนภาพดังภาพประกอบ 18



ภาพประกอบ 22 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถครั้งที่ 1 เมื่อแบบทดสอบมีความยากแตกต่างกัน 3 ระดับ

จากภาพประกอบ 22 พบว่าค่าเฉลี่ยฟิชเชอร์ Z ของแบบทดสอบผสมทุกระดับความยากมีค่าเป็นลบโดยที่แบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือแบบทดสอบผสมที่มีความยาก 0.00 และ 0.50 ตามลำดับ

2.2.1.3 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถ การทดสอบครั้งที่ 1 เมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถ เพื่อทดสอบสมมติฐานข้อ 1.3 ที่ตั้งไว้ว่า *แบบทดสอบผสมที่มีระดับคะแนนข้อสอบแบบให้คะแนนหลายค่ามากกว่าจะมีค่าความสัมพันธ์สูงกว่าแบบทดสอบผสมที่มีระดับคะแนนข้อสอบแบบให้คะแนนหลายค่าน้อยกว่า ในทุกขนาดความยาวและระดับความยากของแบบทดสอบผสม*ปรากฏผลการวิเคราะห์ดังตาราง 18

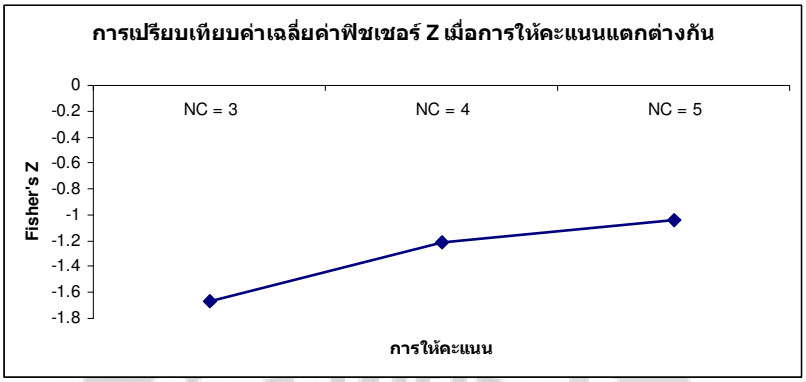
จากตาราง 18 พบว่า เมื่อแบบทดสอบมีความยาวและความยากคงที่ ค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถจากแบบทดสอบผสมที่มีการให้คะแนนต่างกันมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบทดสอบผสมที่มีการให้คะแนน 3 ระดับมีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือ แบบทดสอบที่มีการให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ เมื่อเปรียบเทียบค่าเฉลี่ยรายคู่ พบว่าทุกคู่แตกต่างกัน โดยแบบทดสอบผสมที่ให้คะแนน 3 ระดับมีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากกว่าแบบทดสอบที่ให้คะแนน 4 ระดับ และ 5 ระดับ และแบบทดสอบผสมที่ให้คะแนน 4 ระดับมีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากกว่าแบบทดสอบที่ให้คะแนน 5 ระดับอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 18 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	การให้คะแนน	Mean	NC=3	NC=4	NC=5
	NC = 3	-1.666	-	0.448*	0.620*
	NC = 4	-1.218		-	0.171*
	NC = 5	-1.046			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

เพื่อให้เห็นความแตกต่างของค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถ เมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน 3 ระดับ ผู้วิจัยจึงนำเสนอเป็นแผนภาพ ดังภาพประกอบ 23



ภาพประกอบ 23 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถครั้งที่ 1 เมื่อแบบทดสอบมีการให้คะแนนแตกต่างกัน 3 ระดับ

จากภาพประกอบ 23 พบว่าค่าเฉลี่ยฟิชเชอร์ Z ของแบบทดสอบผสมทุกระดับการให้คะแนนมีค่าเป็นลบโดยที่แบบทดสอบผสมที่ให้คะแนน 3 ระดับ มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือ แบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ

2.2.2 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันความสามารถจริง (θ_2)

ผู้วิจัยได้เปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ตามเงื่อนไขพัฒนาการความสามารถผู้สอบ และลักษณะของแบบทดสอบผสมที่มีความยาว ความยาก และรูปแบบการให้คะแนนแตกต่างกันเพื่อตรวจสอบสมมติฐานข้อ 2 ที่ตั้งไว้ว่า *ค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความสัมพันธ์ระดับสูงกับค่าความสามารถจริงที่กำหนดในทุกระดับพัฒนาการความสามารถ*

การทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันคำนวณโดยแปลงค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันเป็นคะแนนมาตรฐานของฟิชเชอร์ Z ตามกลุ่มระดับพัฒนาการ 9 กลุ่มภายใต้เงื่อนไขความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกันแล้วนำมา

ทดสอบโดยวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (Repeated Measure Analysis of Variance) ปรากฏผลดังตาราง 19

ตาราง 19 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ยฟิชเชอร์ Z ของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถครั้งที่ 2 ($\hat{\theta}_2$) ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันความสามารถจริง (θ_2) ภายใต้เงื่อนไขแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ

แหล่งความแปรปรวน	SS	df	MS	F	ระดับนัยสำคัญ	ขนาดของผล
ทดสอบผลตัวแปรภายในกลุ่ม (Within)						
ความยาว	2.556	2.000	1.278	95.448*	0.000	0.923
ความยาก	9.702	1.138	8.524	105.371*	0.000	0.929
การให้คะแนน	0.132	1.132	0.117	3.13571	0.108	0.282
ความยาว × ความยาก	0.058	1.902	0.030	0.550	0.579	0.064
ความยาว × การให้คะแนน	0.058	1.902	0.030	0.550	0.579	0.064
ความยาก × การให้คะแนน	0.266	1.895	0.141	2.670	0.104	0.250
ความยาว × ความยาก × การให้คะแนน	0.217	2.190	0.099	1.605	0.229	0.167

* มีนัยสำคัญทางสถิติที่ระดับ .05

ผลการวิเคราะห์ตามตาราง 19 พบว่าผลการทดสอบหลักเมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด ความยากของแบบทดสอบผสม 3 ระดับ ค่าเฉลี่ยฟิชเชอร์ Z ของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่การให้คะแนนของแบบทดสอบผสม 3 ขนาด มีค่าเฉลี่ยฟิชเชอร์ Z ของพัฒนาการความสามารถแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 และพบว่าไม่มีปฏิสัมพันธ์ระหว่างความยาว กับ ความยาก ความยาวและการให้คะแนน ความยาก และการให้คะแนน และ ความยาว ความยากและการให้คะแนน ที่ระดับนัยสำคัญทางสถิติ .05 แสดงว่าค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถแตกต่างกันเมื่อแบบทดสอบผสมมีความยาว และความยากแตกต่างกัน

เมื่อผลการทดสอบพบว่าแบบทดสอบผสมมีความยาว และความยาก แตกต่างกัน ส่งผลให้ค่าเฉลี่ยฟิชเชอร์ Z แตกต่างกัน ผู้วิจัยจึงเปรียบเทียบความแตกต่างของค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณพารามิเตอร์ความสามารถ ในแต่ละคู่ระดับของ 2 เงื่อนไข ได้แก่ ความยาว และความยาก

และเพื่อให้เห็นแนวโน้มของค่าเฉลี่ยฟิชเชอร์ Z ครบทุกเงื่อนไขจึงได้เสนอผลการเปรียบเทียบ เมื่อแบบทดสอบผสมให้คะแนนต่างกันด้วย ผลการเปรียบเทียบรายคู่จึงนำเสนอเป็น 3 ตอนเพื่อทดสอบสมมติฐานย่อยของสมมติฐานการวิจัยข้อ 1 จำนวน 3 ข้อดังนี้

2.2.2.1 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 เมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถ เพื่อทดสอบสมมติฐานข้อ 1.1 ที่ตั้งไว้ว่า *แบบทดสอบผสมที่มีขนาดยาวกว่า จะมีค่าความสัมพันธ์สูงกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ในทุกระดับการแจกแจงความยาก และ ระดับการให้คะแนนของแบบทดสอบผสม* ปรากฏผลการวิเคราะห์ดังตาราง 20

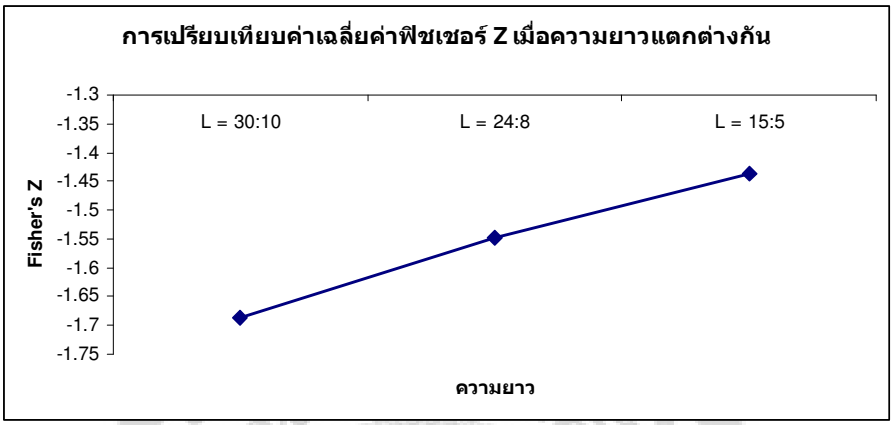
จากตาราง 20 พบว่า เมื่อแบบทดสอบผสมมีความยากและระดับการให้คะแนนคงที่ ค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถจากแบบทดสอบผสมที่มีความยาวแตกต่างกันแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบทดสอบผสมที่มีความยาว 30:10 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือ แบบทดสอบที่มีความยาว 24:8 และ ความยาว 15:5 ตามลำดับ เมื่อเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าทุกคู่ความยาวของแบบทดสอบผสมมีค่าเฉลี่ยฟิชเชอร์ Z แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบทดสอบผสมขนาดความยาว 30:10 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากกว่าแบบทดสอบที่มีความยาว 24:8 และ 15:5 และ แบบทดสอบขนาดความยาว 24:8 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากกว่า แบบทดสอบที่มีความยาว 15:5

ตาราง 20 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาว	Mean	L = 30:10	L = 24:8	L = 15:5
	L = 30:10	-1.688	-	0.140*	0.251*
	L = 24:8	-1.548		-	0.111*
	L = 15:5	-1.437			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

เพื่อแสดงให้เห็นแนวโน้มและทิศทางความแตกต่างของค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถเมื่อแบบทดสอบผสมมีขนาดความยาวแตกต่างกัน ผู้วิจัยจึงนำเสนอผลการเปรียบเทียบค่าเฉลี่ยดังภาพประกอบ 24



ภาพประกอบ 24 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบมีความยาวแตกต่างกัน 3 ขนาด

จากภาพประกอบ 24 พบว่าค่าเฉลี่ยฟิชเชอร์ Z ของแบบทดสอบผสมทุกระดับความยาวมีค่าเป็นลบโดยที่แบบทดสอบผสมที่มีความยาว 30:10 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือแบบทดสอบผสมที่มีความยาว 24:8 และ 15:5 ตามลำดับ

2.2.2.2 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 เมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถเพื่อทดสอบสมมติฐานข้อ 1.2 ที่ตั้งไว้ว่า *"แบบทดสอบผสมที่ระดับความยากของข้อสอบสอดคล้องกับระดับความสามารถผู้สอบ จะมีความสัมพันธ์สูงกว่าแบบทดสอบผสมที่ระดับความยากของข้อสอบไม่สอดคล้องกับระดับความสามารถผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนนของแบบทดสอบผสม"* ปรากฏผลการวิเคราะห์ดังตาราง 21

ตาราง 21 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

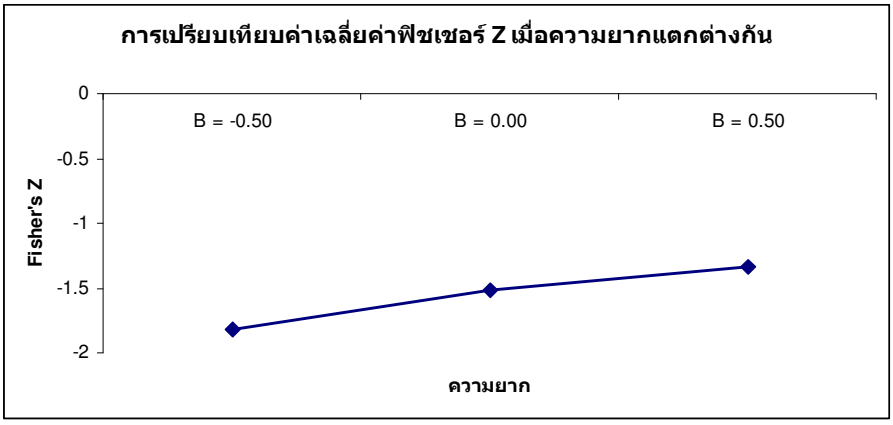
องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาก	Mean	B = -0.50	B = 0.00	B = 0.50
	B = -0.50	-1.820	-	0.303*	0.484*
	B = 0.00	-1.517		-	0.182*
	B = 0.50	-1.336			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 21 พบว่า เมื่อแบบทดสอบมีความยาวและระดับการให้คะแนนคงที่ ค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถจากแบบทดสอบผสมที่มีความยากต่างกันมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือ แบบทดสอบที่มีความยาก 0.00 และ 0.50 ตามลำดับ เมื่อเปรียบเทียบค่าเฉลี่ยรายคู่ พบว่าทุกคู่มีค่าเฉลี่ยฟิชเชอร์ Z แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากกว่าแบบทดสอบที่มีความยาก 0.00 และ 0.50 และ แบบทดสอบผสมที่มีความยาก 0.00 มีค่าเฉลี่ยฟิชเชอร์ Z ป็นลบมากกว่าแบบทดสอบผสมที่มีความยาก 0.50

เพื่อให้เห็นความแตกต่างของค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถ เมื่อแบบทดสอบมีความยากแตกต่างกัน 3 ขนาด ผู้วิจัยจึงนำเสนอเป็นแผนภาพดังภาพประกอบ 25

จากภาพประกอบ 25 พบว่าค่าเฉลี่ยฟิชเชอร์ Z ของแบบทดสอบผสมทุกระดับความยากมีค่าเป็นลบโดยที่แบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือแบบทดสอบผสมที่มีความยาก 0.00 และ 0.50 ตามลำดับ



ภาพประกอบ 25 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีความยากแตกต่างกัน 3 ระดับ

2.2.2.3การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 เมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

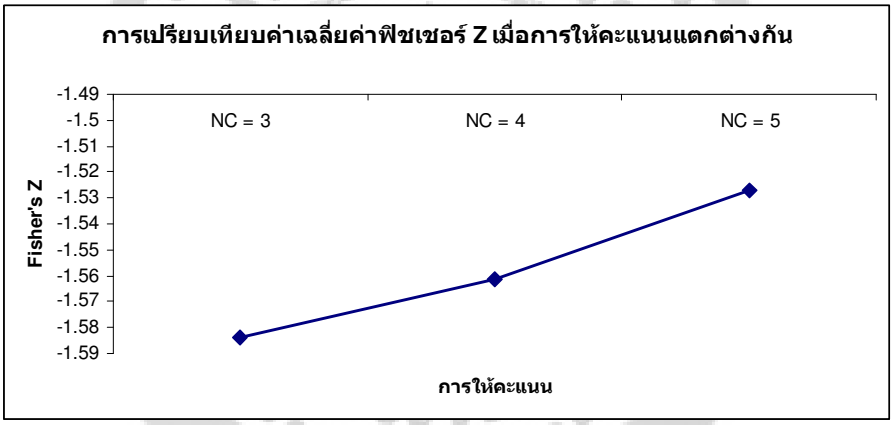
ผู้วิจัยเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถเพื่อทดสอบสมมติฐานข้อ 1.3 ที่ตั้งไว้ว่า *"แบบทดสอบผสมที่มีระดับคะแนนข้อสอบแบบให้คะแนนหลายค่ามากกว่าจะมีความสัมพันธ์สูงกว่าแบบทดสอบผสมที่มีระดับคะแนนข้อสอบแบบให้คะแนนหลายค่าน้อยกว่า ในทุกขนาดความยาวและระดับความยากของแบบทดสอบผสม"* ปรากฏผลการวิเคราะห์ดังตาราง 22

จากตาราง 22 พบว่าเมื่อแบบทดสอบมีความยาวและความยากคงที่ ค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 จากแบบทดสอบผสมที่มีความการให้คะแนนต่างกันมีความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบทดสอบผสมที่มีการให้คะแนน 3 ระดับมีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือ แบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ

เพื่อให้เห็นแนวโน้มความแตกต่างของค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถ เมื่อแบบทดสอบมีการให้คะแนนแตกต่างกัน 3 ระดับผู้วิจัยจึงนำเสนอเป็นแผนภาพดังภาพประกอบ 26

ตาราง 22 การวิเคราะห์ความแตกต่างค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	การให้คะแนน	Mean	NC=3	NC=4	NC=5
	NC = 3	-1.584	-	0.023	0.057
	NC = 4	-1.561		-	0.034
	NC = 5	-1.527			-



ภาพประกอบ 26 การเปรียบเทียบค่าเฉลี่ยฟิชเชอร์ Z ของค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบมีการให้คะแนนแตกต่างกัน 3 ระดับ

จากภาพประกอบ 26 พบว่าค่าเฉลี่ยฟิชเชอร์ Z ของแบบทดสอบผสมทุกระดับการให้คะแนนมีค่าเป็นลบโดยแบบทดสอบผสมที่มีการให้คะแนน 3 ระดับ มีค่าเฉลี่ยฟิชเชอร์ Z เป็นลบมากที่สุด รองลงมาคือ แบบทดสอบผสมที่มีการให้คะแนน 4 ระดับ และ 5 ระดับ

สรุปความสัมพันธ์ระหว่างค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าพารามิเตอร์ความสามารถจริง

1. ค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1)

ค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) มีความสัมพันธ์ทางลบในระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกเงื่อนไข และภายใต้เงื่อนไขเดียวกัน ทุกกลุ่มพัฒนาการมีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากัน มีรายละเอียดดังนี้

เมื่อแบบทดสอบผสมมีความยาก และการให้คะแนนแตกต่างกัน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่เมื่อแบบทดสอบผสมมีความยาวต่างกัน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 โดยพบว่า

แบบทดสอบผสมที่มีความยาวขนาด 15:5 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1) สูงกว่าแบบทดสอบผสมที่มีขนาด 24:8 และขนาด 30:10 ในทิศทางลบ

แบบทดสอบผสมที่มีความยาก -0.50 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1)สูงกว่าแบบทดสอบผสมมีความยาก 0.00 และ 0.50 ในทิศทางลบ

แบบทดสอบผสมที่มีการให้คะแนน 3 ระดับ มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 ($\hat{\theta}_1$) กับค่าความสามารถจริง (θ_1)สูงกว่ากรณีที่แบบทดสอบผสมมีการให้คะแนน 4 ระดับ และ 5 ระดับ ในทิศทางลบ

2. ค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2)

ค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) มีความสัมพันธ์ทางลบในระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกเงื่อนไข โดยพบว่าภายใต้เงื่อนไข ความยาว ความยาก และการให้คะแนนของแบบทดสอบผสมเท่ากันแต่ละกลุ่มพัฒนาการมีค่าสัมประสิทธิ์สหสัมพันธ์ ไม่เท่ากัน มีรายละเอียดดังนี้

เมื่อแบบทดสอบผสมมีความยาวและความยากแตกต่างกันค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่สำหรับกรณีแบบทดสอบผสมมีการให้คะแนนแตกต่างกันค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) แตกต่างกันอย่างไม่มีนัยสำคัญที่ระดับ .05 โดยพบว่า

แบบทดสอบผสมที่มีขนาดความยาว 30:10 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ในทิศทางลบสูงกว่าแบบทดสอบที่มีความยาวขนาด 24:8 และ ขนาด 15:5

แบบทดสอบผสมที่มีความยาก -0.50 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ในทิศทางลบสูงกว่าแบบทดสอบผสมมีความยาก 0.00 และ 0.50

แบบทดสอบผสมที่มีการให้คะแนน 3 ระดับ มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 2 ($\hat{\theta}_2$) กับค่าความสามารถจริง (θ_2) ในทิศทางลบสูงกว่าแบบทดสอบผสมที่มีการให้คะแนน 4 ระดับ และ 5 ระดับ

สรุปความสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1และ ครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าพารามิเตอร์ความสามารถจริงได้ดังตาราง 23

ตาราง 23 สรุปความสัมพันธ์ระหว่างค่าประมาณความสามารถในการทดสอบครั้งที่ 1 และ ครั้งที่ 2 ที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าพารามิเตอร์ความสามารถจริง

ค่าประมาณความสามารถ	**ลักษณะความสัมพันธ์**		**การเปรียบเทียบความแตกต่าง**	
	ทิศทาง	ปริมาณ (ภายใต้เงื่อนไขเดียวกัน)	แตกต่างอย่างมีนัยสำคัญ .05	แตกต่างอย่างไม่มีนัยสำคัญ .05
การทดสอบครั้งที่ 1	**ลบ**	**มีค่าเท่ากันทุกกลุ่มพัฒนาการ**	**ความยาก** - ยาก -0.50 มีค่า r สูงสุดทางลบ **การให้คะแนน** - 3 ระดับ มีค่า r สูงสุดทางลบ	**ความยาว** - ขนาด 15:5 มีค่า r สูงสุดทางลบ
การทดสอบครั้งที่ 2	**ลบ**	**แตกต่างกันตามกลุ่มพัฒนาการ**	**ความยาว** - ขนาด 30:10 มีค่า r สูงสุดทางลบ ความยาก - ยาก -0.50 มี r ค่าสูงสุดทางลบ	**การให้คะแนน** - 3 ระดับ มีค่า r สูงสุดทางลบ

หมายเหตุ: ยาว 30:10 และ ยาว 15:5 หมายถึงแบบทดสอบผสมที่มีความยาว 30:10 และ 15:5 ตามลำดับ

ยาก -0.50 หมายถึงแบบทดสอบผสมที่ ใช้ในการทดสอบครั้งที่ 1 ความยากเฉลี่ยเป็น - 0.50

3 ระดับ หมายถึง ระดับคะแนนของข้อสอบแบบเขียนตอบในแบบทดสอบผสมมี 3 ระดับได้แก่ 0 1 และ 2 คะแนน

ตอนที่ 3 การตรวจสอบความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ในการศึกษาตอนนี้ผู้วิจัยคำนวณผลต่างระหว่างพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับพัฒนาการความสามารถจริงของผู้สอบ 1,000 คน เพื่อศึกษาค่าเฉลี่ยของความลำเอียง (BIAS) และค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของค่าประมาณความสามารถกำลังสอง (RMSE) ของพัฒนาการความสามารถรายบุคคลในแต่ละเงื่อนไขตามวัตถุประสงค์การวิจัยข้อ 2 ที่ระบุว่า *"เพื่อศึกษาและเปรียบเทียบความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไขความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกัน"* นำเสนอผลการตรวจสอบความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เป็นสองส่วน ดังนี้

3.1 การศึกษาความลำเอียง (BIAS) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของพัฒนาการความสามารถกำลังสอง (RMSE) โดยแยกตามเงื่อนไขที่ใช้ศึกษา ดังนี้

3.1.1 การศึกษาความลำเอียง (BIAS) ของพัฒนาการความสามารถ เมื่อ

3.1.1.1 ความยาวคงที่

3.1.1.2 ความยากคงที่

3.1.1.3 การให้คะแนนคงที่

3.1.2 การศึกษาค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถ เมื่อ

3.1.2.1 ความยาวคงที่

3.1.2.2 ความยากคงที่

3.1.2.3 การให้คะแนนคงที่

3.2 การเปรียบเทียบความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถ

3.2.1 การเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ

3.2.2 การเปรียบเทียบค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถ

3.1 การศึกษาความลำเอียง (BIAS) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถ

3.1.1 การศึกษาความลำเอียง (BIAS) ของพัฒนาการความสามารถ

ในตอนนี้เสนอผลตามเงื่อนไขของแบบทดสอบผสมที่ต้องการศึกษา เมื่อความยาวคงที่ ความยากคงที่ และ การให้คะแนนคงที่ปรากฏผลดังนี้

3.1.1.1 ความยาวคงที่

กรณีแบบทดสอบผสมยาว 30:10

จากตาราง 24 และภาพประกอบ 23 (ก – ค) เมื่อแบบทดสอบผสมมีความยาว เป็น 30:10 และมีความยาก -0.50 และ 0.50 พบว่ากรณีที่ให้คะแนน 3 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด รองลงมาคือ กรณีให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ แต่สำหรับกรณีแบบทดสอบผสมมีความยาว 30:10 และมีความยากเป็น 0.00 นั้นพบว่ากลุ่มผู้สอบที่มีพัฒนาการเพิ่มขึ้น 0.1 (M1, M4 และ M7) กรณีให้คะแนน 3 ระดับจะมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำกว่ากรณีให้คะแนน 4 ระดับ และ 5 ระดับ ส่วนกลุ่มผู้สอบที่มีพัฒนาการรูปแบบอื่นที่เหลือ กรณีที่ให้คะแนน 5 ระดับจะมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำกว่า กรณีให้คะแนน 3 ระดับ และ 4 ระดับ

กรณีแบบทดสอบผสมยาว 24:8

จากตาราง 24 และภาพประกอบ 24 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 กรณีให้คะแนน 4 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด รองลงมาคือ กรณีให้คะแนน 3 ระดับ และ 5 ระดับ ตามลำดับ แต่สำหรับกรณีแบบทดสอบผสมมีความยาว 24:8 และมีความยาก 0.00 และ 0.50 กรณีให้คะแนน 4 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเฉพาะในกลุ่ม M1 และ M4 เท่านั้น ส่วนกลุ่ม M2 กรณีให้คะแนน 5 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่ม M5 M6 M7 M8 และM9 กรณีให้คะแนน 3 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

กรณีแบบทดสอบผสมยาว 15:5

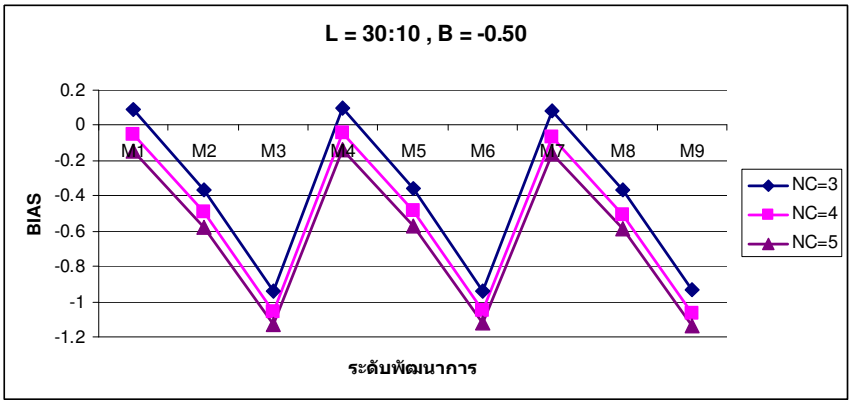
จากตาราง 24 และภาพประกอบ 25 (ก – ค) เมื่อแบบทดสอบผสมมีความยาวเป็น 15:5 มีความยาก -0.50 และ 0.50 กลุ่มพัฒนาการส่วนใหญ่ มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อให้คะแนน 3 ระดับ ยกเว้นกลุ่มพัฒนาการ M2 และ M4 พบว่ากรณีให้คะแนน 4 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถมีค่าต่ำที่สุด ส่วนกรณีแบบทดสอบผสมมีความยาวเป็น 15:5 และมีความยากเป็น 0.00 นั้นเฉพาะกลุ่ม M1 M4 และ M7 เท่านั้นที่พบว่าค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อให้คะแนน 3 ระดับ

ส่วนกลุ่ม M2 M5 M8 และM9 พบว่าค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อให้คะแนน 4 ระดับ และกลุ่ม M3 และ M6 พบว่าค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อให้คะแนน 5 ระดับ

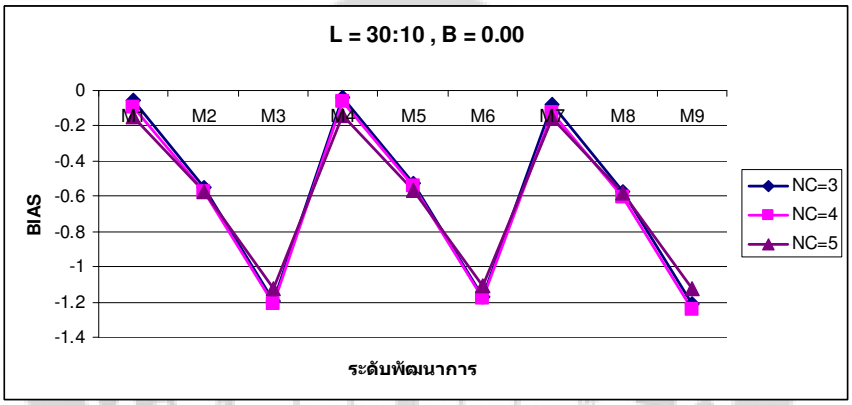
ตาราง 24 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมีความยาว และ ความยากคงที่ จำแนกตามการให้คะแนนและระดับพัฒนาการ

L	B	NC	M1	M2	M3	M4	M5	M6	M7	M8	M9
		3	0.044	-0.461	-1.112	0.066	-0.444	-1.108	0.023	-0.476	-1.113
	-0.50	4	-0.055	-0.548	-1.196	-0.023	-0.522	-1.184	-0.082	-0.572	-1.207
		5	-0.086	-0.558	-1.178	-0.057	-0.532	-1.162	-0.114	-0.583	-1.194
		3	-0.039	-0.566	-1.268	-0.003	-0.530	-1.234	-0.078	-0.601	-1.292
30:10	0.00	4	-0.094	-0.610	-1.311	-0.085	-0.563	-1.275	-0.141	-0.657	-1.342
		5	-0.088	-0.551	-1.162	-0.059	-0.522	-1.141	-0.117	-0.579	-1.183
		3	0.020	-0.456	-1.077	0.043	-0.432	-1.060	-0.004	-0.480	-1.095
	0.50	4	-0.054	-0.523	-1.142	-0.025	-0.493	-1.120	-0.085	-0.480	-1.163
		5	-0.086	-0.537	-1.136	-0.057	-0.508	-1.113	-0.210	-0.567	-1.161
		3	-0.028	-0.570	-1.279	0.006	-0.543	-1.266	-0.062	-0.596	-1.288
	-0.50	4	-0.024	-0.510	-1.142	0.003	-0.487	-1.132	-0.051	-0.530	-1.150
		5	-0.082	-0.547	-1.153	-0.058	-0.523	-1.138	-0.107	-0.568	-1.167
		3	0.050	-0.566	-1.056	0.071	-0.411	-1.042	0.028	-0.450	-1.067
24:8	0.00	4	-0.028	-0.610	-1.119	0.000	-0.475	-1.105	-0.057	-0.528	-1.137
		5	-0.082	-0.551	-1.128	-0.056	-0.508	-1.111	-0.108	-0.562	-1.153
		3	0.037	-0.566	-1.039	0.057	-0.411	-1.024	0.015	-0.452	-1.057
	0.50	4	-0.034	-0.610	-1.100	-0.007	-0.467	-1.080	-0.061	-0.520	-1.120
		5	-0.082	-0.551	-1.112	-0.058	-0.500	-1.089	-0.107	-0.553	-1.136
		3	0.033	-0.570	-1.088	0.050	-0.446	-1.087	0.015	-0.472	-1.093
	-0.50	4	-0.072	-0.510	-1.180	-0.045	-0.528	-1.169	-0.097	-0.574	-1.189
		5	-0.114	-0.547	-1.175	-0.089	-0.546	-1.155	-0.139	-0.597	-1.190
		3	-0.034	-0.545	-1.214	-0.008	-0.521	-1.197	-0.060	-0.568	-1.225
15:5	0.00	4	-0.066	-0.540	-1.159	-0.040	-0.513	-1.142	-0.094	-0.563	-1.173
		5	-0.110	-0.560	-1.154	-0.083	-0.533	-1.135	-0.135	-0.587	-1.177
		3	0.019	-0.545	-1.068	0.040	-0.434	-1.053	-0.003	-0.475	-1.083
	0.50	4	-0.064	-0.540	-1.140	-0.036	-0.500	-1.120	-0.091	-0.555	-1.159
		5	-0.105	-0.560	-1.140	-0.079	-0.522	-1.117	-0.132	-0.579	-1.163

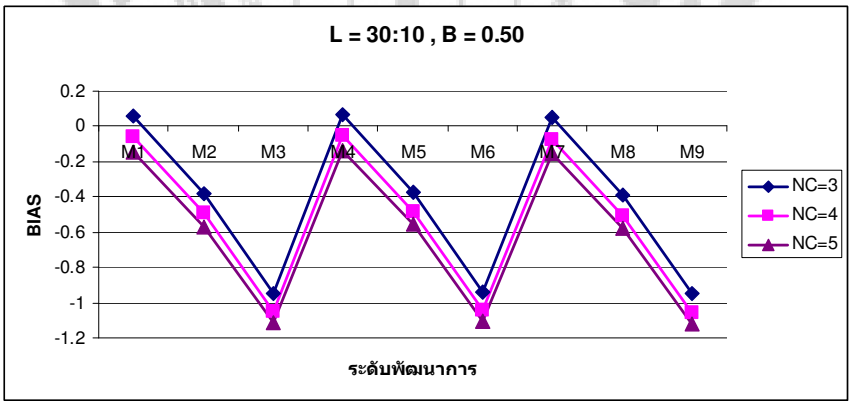
*___ หมายถึง มีค่า BIAS ต่ำที่สุด ภายใต้เงื่อนไขความยาว และ ความยากเท่ากัน



27 ก

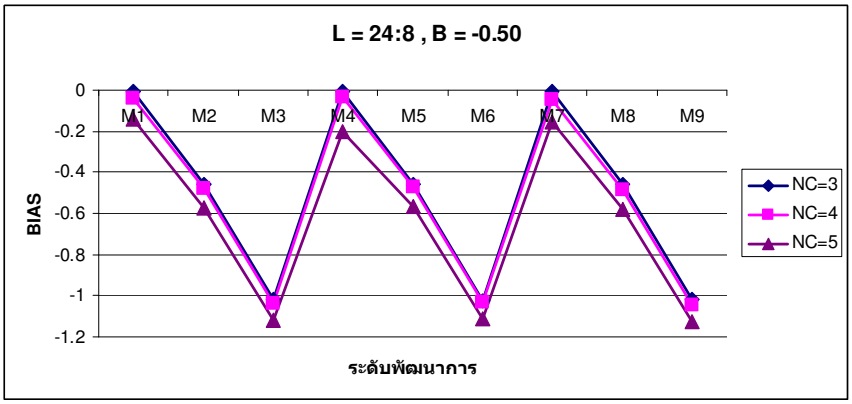


27 ข

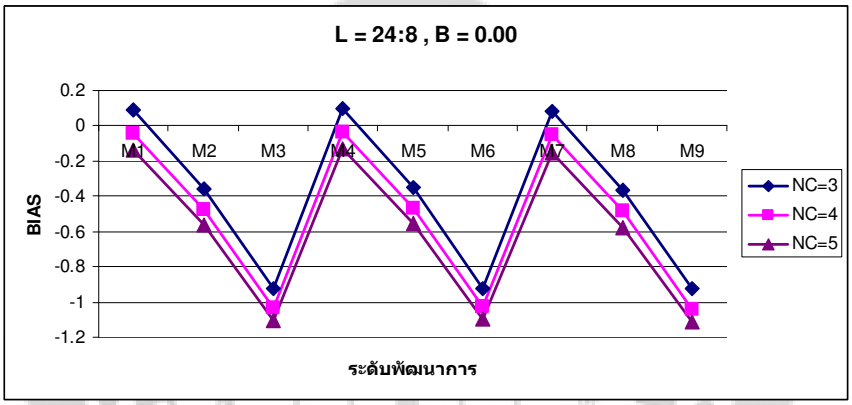


27 ค

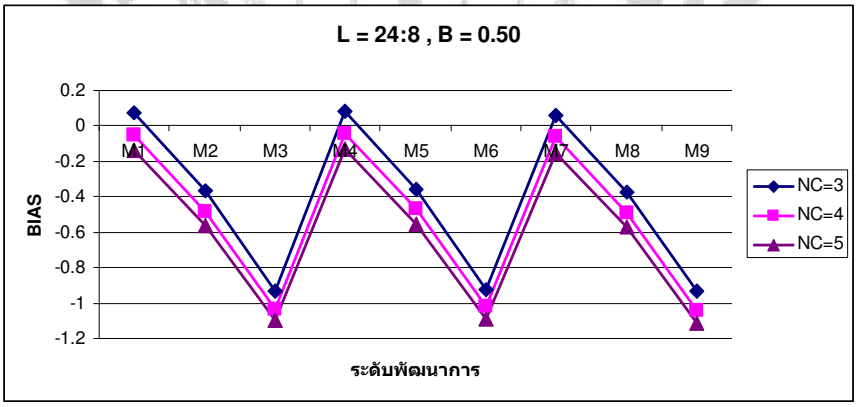
ภาพประกอบ 27 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก -0.50 (27 ก) 0.00 (27 ข) 0.50 (27 ค)



28 ก

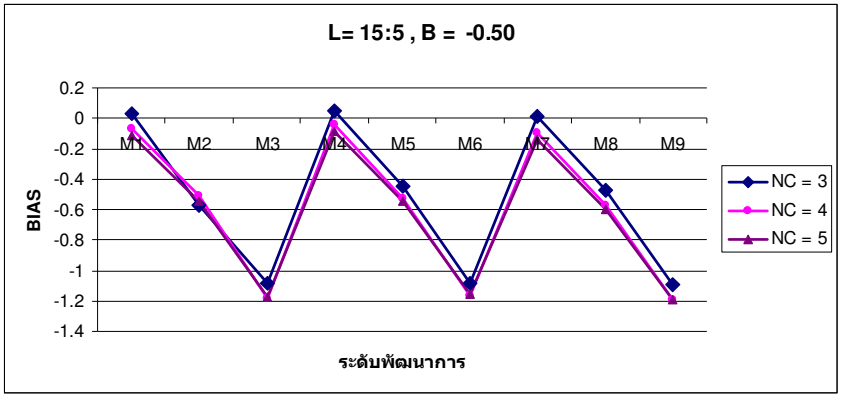


28 ข

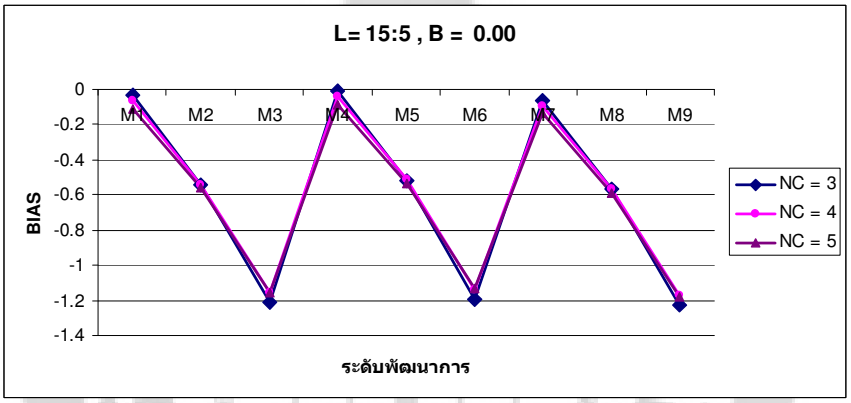


28 ค

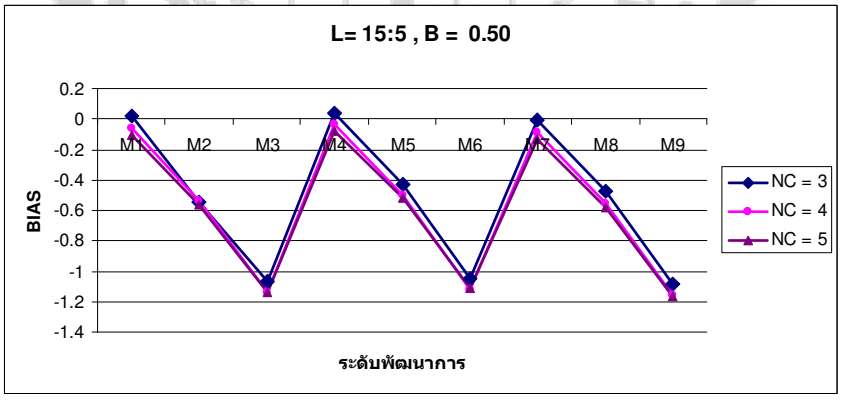
ภาพประกอบ 28 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 (28 ก) 0.00 (28 ข) 0.50 (28 ค)



29 ก



29 ข



29 ค

ภาพประกอบ 29 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก -0.50 (29 ก) 0.00 (29 ข) 0.50 (29 ค)

3.1.1.2 ความยากคงที่

กรณีความยากเป็น -0.50

จากตาราง 25 และภาพประกอบ 30 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยากเป็น -0.50 และให้คะแนน 3 ระดับ ในกลุ่มพัฒนาการ M1 และ M4 พบว่าแบบทดสอบผสมขนาด 24:8 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำสุด ในกลุ่ม M2 และ M5 แบบทดสอบผสมขนาด 30:10 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำสุด ส่วนกลุ่มที่เหลือแบบทดสอบผสมขนาด 15:5 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำสุด

เมื่อแบบทดสอบผสมมีความยากเป็น -0.50 และให้คะแนน 4 ระดับ พบว่าในทุกกลุ่มพัฒนาการแบบทดสอบผสมขนาด 24:8 มีค่าความลำเอียง (BIAS) ต่ำที่สุด ส่วนในกลุ่มพัฒนาการ M2 แบบทดสอบผสมขนาด 15:5 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถเท่ากับแบบทดสอบผสมขนาด 24:8

เมื่อแบบทดสอบผสมมีความยากเป็น -0.50 และให้คะแนน 5 ระดับ พบว่าในทุกกลุ่มพัฒนาการยกเว้น M4 แบบทดสอบผสมขนาด 24:8 มีค่าความลำเอียง (BIAS) ต่ำที่สุด และสำหรับกลุ่ม M2 พบว่าแบบทดสอบผสมขนาด 15:5 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถเท่ากับแบบทดสอบผสมขนาด 24:8 กลุ่ม M7 แบบทดสอบผสมขนาด 30:10 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถเท่ากับแบบทดสอบผสมขนาด 24:8 ส่วนกลุ่ม M4 นั้นพบว่าแบบทดสอบผสมขนาด 30:10 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

กรณีความยากเป็น 0.00

จากตาราง 25 และภาพประกอบ 31 (ก – ค) พบว่าเมื่อความยากแบบทดสอบผสมเป็น 0.00 และให้คะแนน 3 ระดับ กลุ่มพัฒนาการ M1 และ M2 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถมีค่าต่ำที่สุดเมื่อแบบทดสอบผสมยาว 15:5 สำหรับกลุ่มพัฒนาการ M4 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมยาว 30:10 ส่วนกลุ่มพัฒนาการที่เหลือความลำเอียง (BIAS) ของพัฒนาการความสามารถมีค่าต่ำที่สุดในกรณีแบบทดสอบผสมมีขนาด 24:8

เมื่อความยากแบบทดสอบผสมเป็น 0.00 และให้คะแนน 4 ระดับ พบว่าในทุกกลุ่มพัฒนาการยกเว้นกลุ่มพัฒนาการ M2 ความลำเอียง (BIAS) ของพัฒนาการความสามารถมีค่าต่ำที่สุดเมื่อแบบทดสอบผสมยาว 24:8 แต่สำหรับกลุ่มพัฒนาการ M2 นั้นมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมยาว 15:5

เมื่อความยากแบบทดสอบผสมเป็น 0.00 และให้คะแนน 5 ระดับ พบว่าทุกกลุ่มพัฒนาการ มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมยาว

24:8 และในกรณีแบบทดสอบผสมยาว 24:8 และ 30:10 พบว่ากลุ่ม M2 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถเท่ากัน

ตาราง 25 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีความยาก และการให้คะแนนคงที่ จำแนกตามความยาวและระดับพัฒนาการ

B	NC	L	M1	M2	M3	M4	M5	M6	M7	M8	M9
-0.50	3	30:10	0.044	-0.461	-1.112	0.066	-0.444	-1.108	0.023	-0.476	-1.113
		24:8	-0.028	-0.570	-1.279	0.006	-0.543	-1.266	-0.060	-0.596	-1.288
		15:5	0.033	-0.570	-1.088	0.050	-0.446	-1.087	0.015	-0.472	-1.093
	4	30:10	-0.055	-0.548	-1.196	-0.023	-0.522	-1.184	-0.080	-0.572	-1.207
		24:8	-0.024	-0.510	-1.142	0.003	-0.487	-1.132	-0.050	-0.530	-1.150
		15:5	-0.072	-0.510	-1.180	-0.045	-0.528	-1.169	-0.100	-0.574	-1.189
	5	30:10	-0.086	-0.558	-1.178	-0.057	-0.532	-1.162	-0.110	-0.583	-1.194
		24:8	-0.082	-0.547	-1.153	-0.058	-0.523	-1.138	-0.110	-0.568	-1.167
		15:5	-0.114	-0.547	-1.175	-0.089	-0.546	-1.155	-0.140	-0.597	-1.190
0.00	3	30:10	-0.039	-0.566	-1.268	-0.003	-0.530	-1.234	-0.080	-0.601	-1.292
		24:8	0.050	-0.566	-1.056	0.071	-0.411	-1.042	0.028	-0.450	-1.067
		15:5	-0.034	-0.545	-1.214	-0.008	-0.521	-1.197	-0.060	-0.568	-1.225
	4	30:10	-0.094	-0.610	-1.311	-0.085	-0.563	-1.275	-0.140	-0.657	-1.342
		24:8	-0.028	-0.610	-1.119	0.000	-0.475	-1.105	-0.060	-0.528	-1.137
		15:5	-0.066	-0.540	-1.159	-0.040	-0.513	-1.142	-0.090	-0.563	-1.173
	5	30:10	-0.088	-0.551	-1.162	-0.059	-0.522	-1.141	-0.120	-0.579	-1.183
		24:8	-0.082	-0.551	-1.128	-0.056	-0.508	-1.111	-0.110	-0.562	-1.153
		15:5	-0.110	-0.560	-1.154	-0.083	-0.533	-1.135	-0.140	-0.587	-1.177
0.50	3	30:10	0.020	-0.456	-1.077	0.043	-0.432	-1.060	0.000	-0.480	-1.095
		24:8	0.037	-0.566	-1.039	0.057	-0.411	-1.024	0.015	-0.452	-1.057
		15:5	0.019	-0.545	-1.068	0.040	-0.434	-1.053	0.000	-0.475	-1.083
	4	30:10	-0.054	-0.523	-1.142	-0.025	-0.493	-1.120	-0.080	-0.480	-1.163
		24:8	-0.034	-0.610	-1.100	-0.007	-0.467	-1.080	-0.060	-0.520	-1.120
		15:5	-0.064	-0.540	-1.140	-0.036	-0.500	-1.120	-0.090	-0.555	-1.159
	5	30:10	-0.086	-0.537	-1.136	-0.057	-0.508	-1.113	-0.210	-0.567	-1.161
		24:8	-0.082	-0.551	-1.112	-0.058	-0.500	-1.089	-0.110	-0.553	-1.136
		15:5	-0.105	-0.560	-1.140	-0.079	-0.522	-1.117	-0.130	-0.579	-1.163

*___ หมายถึง มีค่า BIAS ต่ำที่สุด ภายใต้เงื่อนไขความยาก และ ระดับการให้คะแนนเท่ากัน

กรณีความยากเป็น 0.50

จากตาราง 25 และภาพประกอบ 32 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยากเป็น 0.50 กรณีให้คะแนน 3 ระดับ ในกลุ่ม M1 M4 และ M7 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมยาว 15:5 กลุ่ม M2 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมยาว 30:10 ส่วนกลุ่มที่เหลือมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมยาวแบบทดสอบผสมขนาด 24:8

เมื่อความยากแบบทดสอบผสมเป็น 0.00 และให้คะแนน 4 ระดับ ในกลุ่ม M2 และ M8 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมยาว 30:10 ส่วนกลุ่มที่เหลือมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมยาว 24:8 มีค่าความลำเอียง (BIAS)

เมื่อความยากแบบทดสอบผสมเป็น 0.00 และให้คะแนน 5 ระดับ ในกลุ่ม M2 และ M4 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมยาว 30:10 ส่วนกลุ่มที่เหลือมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมยาว 24:8

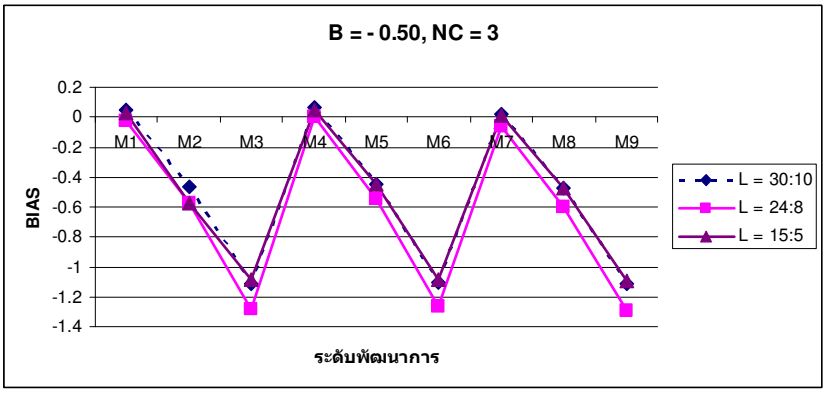
3.1.1.3 การให้คะแนนคงที่

กรณีให้คะแนน 3 ระดับ

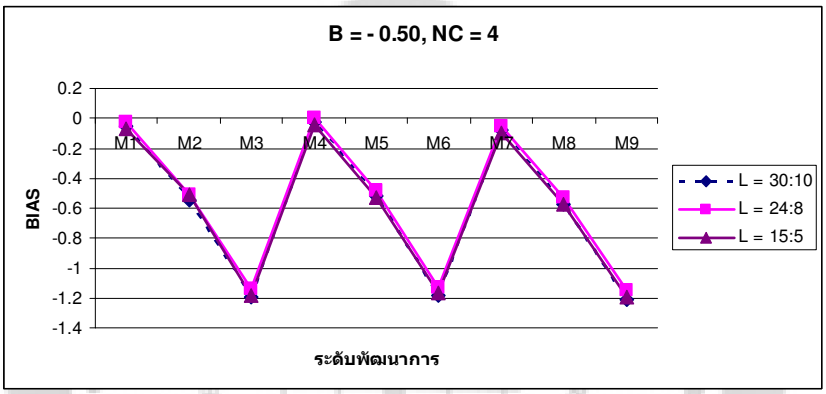
จากตาราง 26 และภาพประกอบ 33 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 30:10 และให้คะแนน 3 ระดับ ในกลุ่ม M4 พบว่าแบบทดสอบผสมที่มีความยาก 0.00 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถในกลุ่ม M8 ต่ำที่สุดกรณีแบบทดสอบผสมมีความยาก -0.50 แต่สำหรับกลุ่มที่เหลือความลำเอียง (BIAS) ของพัฒนาการความสามารถมีค่าต่ำที่สุดเมื่อแบบทดสอบผสมมีความยาก 0.50

กรณีแบบทดสอบผสมมีความยาว 24:8 และให้คะแนน 3 ระดับ ในกลุ่ม M1 และ M4 พบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด กลุ่ม M2 และ M5 แบบทดสอบที่มีความยาก 0.00 และ 0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่มที่เหลือแบบทดสอบผสมที่มีความยาก 0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

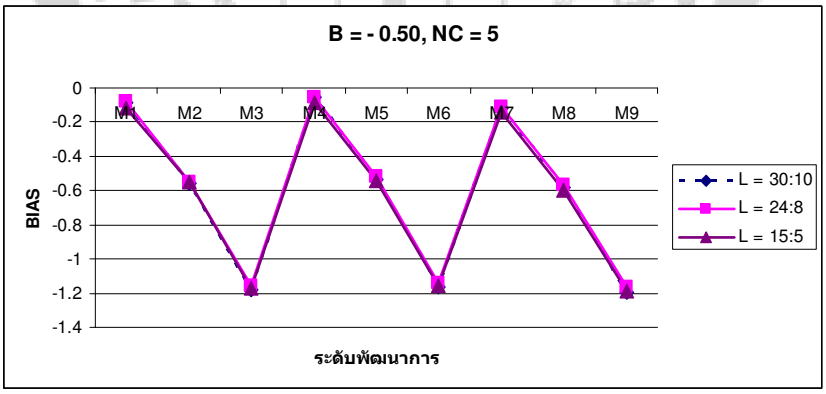
กรณีแบบทดสอบผสมมีความยาว 15:5 และให้คะแนน 3 ระดับ ค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถของกลุ่ม M4 ต่ำที่สุดเมื่อแบบทดสอบที่มีความยาก 0.00 กลุ่ม M8 ต่ำที่สุดเมื่อแบบทดสอบที่มีความยาก -0.50 ส่วนกลุ่มที่เหลือแบบทดสอบที่มีความยาก 0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด



30 ก

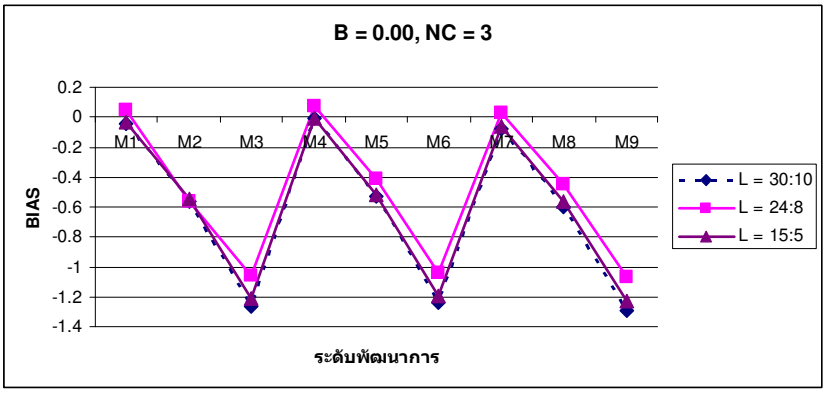


30 ข

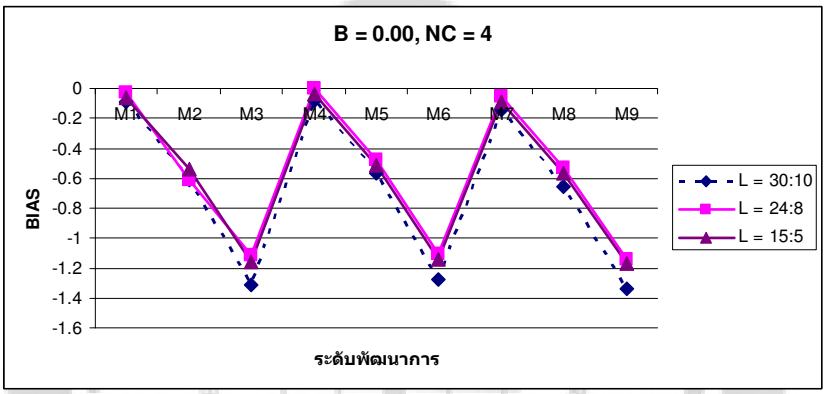


30 ค

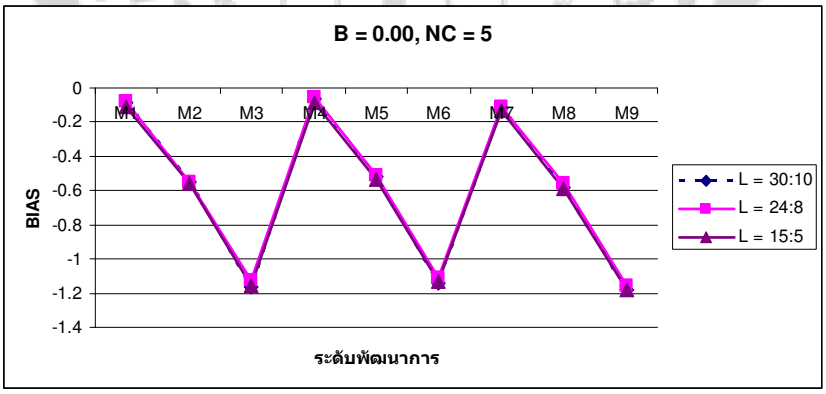
ภาพประกอบ 30 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 3 ระดับ (30 ก) 4 ระดับ (30 ข) 5 ระดับ (30 ค)



31 ก

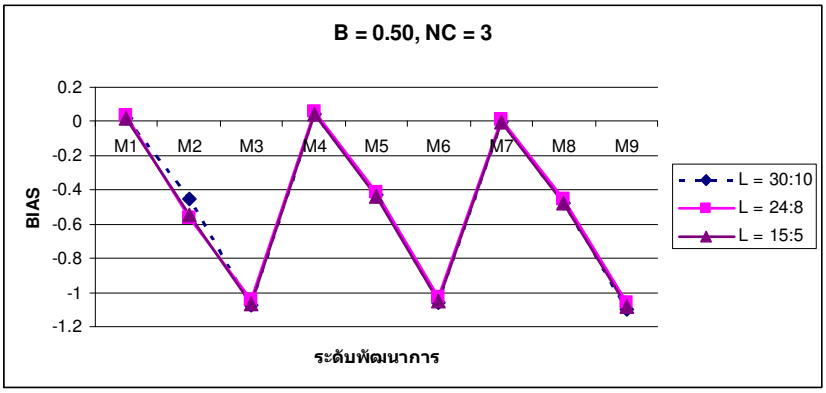


31 ข

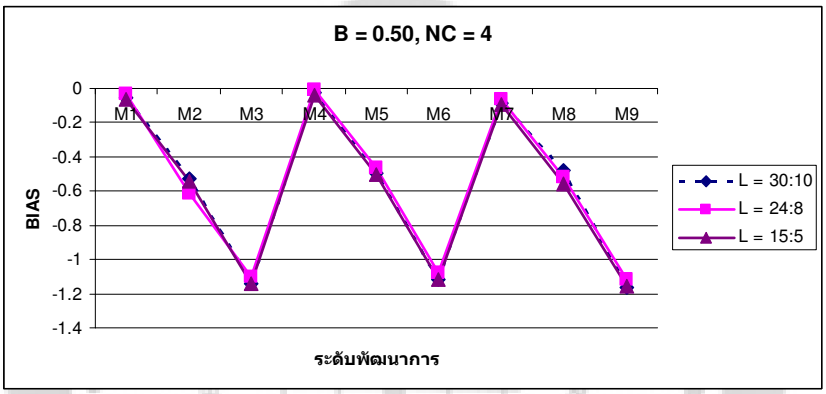


31 ค

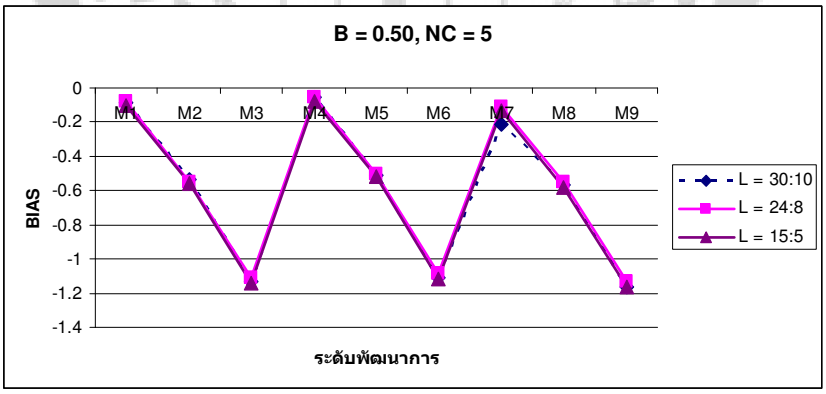
ภาพประกอบ 31 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.00 และให้คะแนน 3 ระดับ (31 ก) 4 ระดับ (31 ข) 5 ระดับ (31 ค)



32 ก



32 ข



32 ค

ภาพประกอบ 32 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 3 ระดับ (32 ก) 4 ระดับ (32 ข) 5 ระดับ (32 ค)

ตาราง 26 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีการให้คะแนน และความยาวคงที่ จำแนกตามความยากและระดับพัฒนาการ

NC	L	B	M1	M2	M3	M4	M5	M6	M7	M8	M9
		-0.50	0.044	-0.461	-1.112	0.066	-0.444	-1.108	0.023	-0.476	-1.113
	30:10	0.00	-0.039	-0.566	-1.268	-0.003	-0.530	-1.234	-0.078	-0.601	-1.292
		0.50	0.020	-0.456	-1.077	0.043	-0.432	-1.060	-0.004	-0.480	-1.095
		-0.50	-0.028	-0.570	-1.279	0.006	-0.543	-1.266	-0.062	-0.596	-1.288
3	24:8	0.00	0.050	-0.566	-1.056	0.071	-0.411	-1.042	0.028	-0.450	-1.067
		0.50	0.037	-0.566	-1.039	0.057	-0.411	-1.024	0.015	-0.452	-1.057
		-0.50	0.033	-0.570	-1.088	0.050	-0.446	-1.087	0.015	-0.472	-1.093
	15:5	0.00	-0.034	-0.545	-1.214	-0.008	-0.521	-1.197	-0.060	-0.568	-1.225
		0.50	0.019	-0.545	-1.068	0.040	-0.434	-1.053	-0.003	-0.475	-1.083
		-0.50	-0.055	-0.548	-1.196	-0.023	-0.522	-1.184	-0.082	-0.572	-1.207
	30:10	0.00	-0.094	-0.610	-1.311	-0.085	-0.563	-1.275	-0.141	-0.657	-1.342
		0.50	-0.054	-0.523	-1.142	-0.025	-0.493	-1.120	-0.085	-0.480	-1.163
		-0.50	-0.024	-0.510	-1.142	0.003	-0.487	-1.132	-0.051	-0.530	-1.150
4	24:8	0.00	-0.028	-0.610	-1.119	0.000	-0.475	-1.105	-0.057	-0.528	-1.137
		0.50	-0.034	-0.610	-1.100	-0.007	-0.467	-1.080	-0.061	-0.520	-1.120
		-0.50	-0.072	-0.510	-1.180	-0.045	-0.528	-1.169	-0.097	-0.574	-1.189
	15:5	0.00	-0.066	-0.540	-1.159	-0.040	-0.513	-1.142	-0.094	-0.563	-1.173
		0.50	-0.064	-0.540	-1.140	-0.036	-0.500	-1.120	-0.091	-0.555	-1.159
		-0.50	-0.086	-0.558	-1.178	-0.057	-0.532	-1.162	-0.114	-0.583	-1.194
	30:10	0.00	-0.088	-0.551	-1.162	-0.059	-0.522	-1.141	-0.117	-0.579	-1.183
		0.50	-0.086	-0.537	-1.136	-0.057	-0.508	-1.113	-0.210	-0.567	-1.161
		-0.50	-0.082	-0.547	-1.153	-0.058	-0.523	-1.138	-0.107	-0.568	-1.167
5	24:8	0.00	-0.082	-0.551	-1.128	-0.056	-0.508	-1.111	-0.108	-0.562	-1.153
		0.50	-0.082	-0.551	-1.112	-0.058	-0.500	-1.089	-0.107	-0.553	-1.136
		-0.50	-0.114	-0.547	-1.175	-0.089	-0.546	-1.155	-0.139	-0.597	-1.190
	15:5	0.00	-0.110	-0.560	-1.154	-0.083	-0.533	-1.135	-0.135	-0.587	-1.177
		0.50	-0.105	-0.560	-1.140	-0.079	-0.522	-1.117	-0.132	-0.579	-1.163

*___ หมายถึง มีค่า BIAS ต่ำที่สุด ภายใต้เงื่อนไขความยาว และ ระดับการให้คะแนนเท่ากัน

กรณีให้คะแนน 4 ระดับ

จากตาราง 26 และภาพประกอบ 34 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 30:10 และให้คะแนน 4 ระดับ ในกลุ่ม M4 และ M7 พบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่มที่เหลือพบว่าแบบทดสอบผสมที่มีความยาก 0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

กรณีแบบทดสอบผสมมีความยาว 24:8 และให้คะแนน 4 ระดับ ในกลุ่ม M1 M2 และ M7 พบว่าค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมมีความยาก -0.50 ส่วนกลุ่ม M4 พบว่าแบบทดสอบผสมที่มีความยาก 0.00 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่มที่เหลือส่วนใหญ่พบว่าแบบทดสอบผสมที่มีความยาก 0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

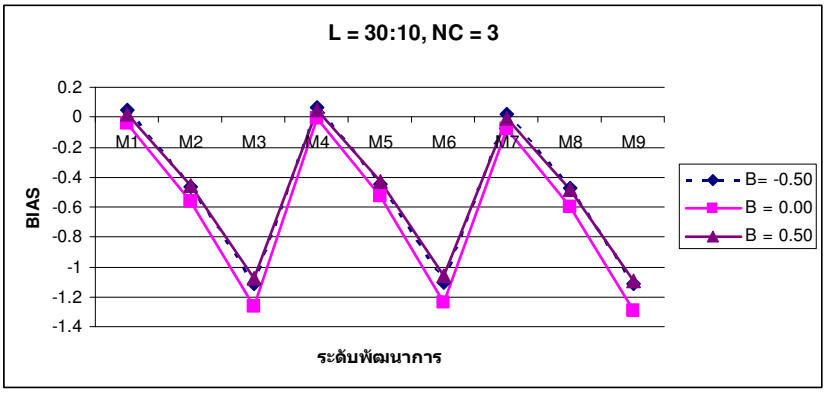
กรณีแบบทดสอบผสมมีความยาว 15:5 และให้คะแนน 4 ระดับ ในกลุ่ม M2 พบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่มที่เหลือส่วนใหญ่พบว่าแบบทดสอบผสมที่มีความยาก 0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

กรณีให้คะแนน 5 ระดับ

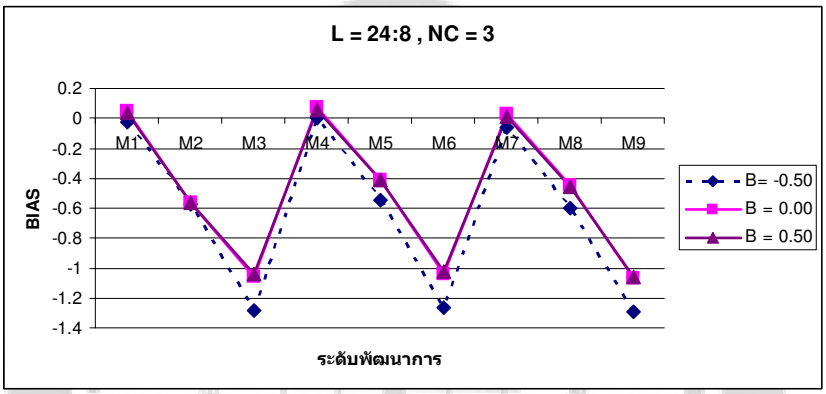
จากตาราง 26 และภาพประกอบ 35 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 30:10 และให้คะแนน 5 ระดับ ในกลุ่ม M1 M4 และ M7 พบว่าค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดเมื่อแบบทดสอบผสมที่มีความยาก -0.50 แต่กลุ่มที่เหลือพบว่าค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดกรณีแบบทดสอบผสมมีความยาก 0.50

กรณีแบบทดสอบผสมมีความยาว 24:8 และให้คะแนน 5 ระดับ ในกลุ่ม M1ค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถเท่ากันทุกระดับความยาก กลุ่ม M2 และ M7 พบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด กลุ่ม M4 พบว่าแบบทดสอบผสมที่มีความยาก 0.00 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่มที่เหลือส่วนใหญ่พบว่าแบบทดสอบผสมที่มีความยาก 0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

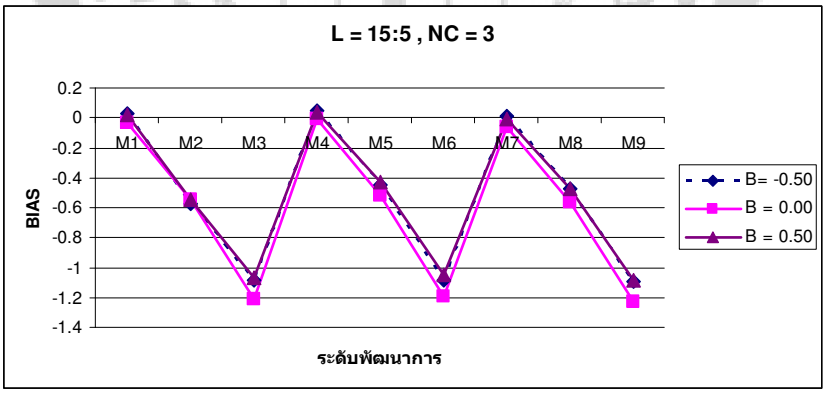
กรณีแบบทดสอบผสมมีความยาว 15:5 และให้คะแนน 5 ระดับ ในกลุ่ม M2 พบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่มที่เหลือส่วนใหญ่พบว่าแบบทดสอบผสมที่มีความยาก 0.50 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด



33 ก

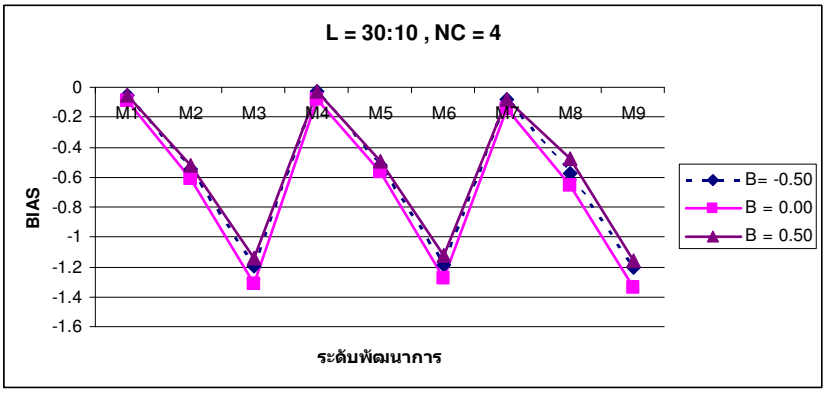


33 ข

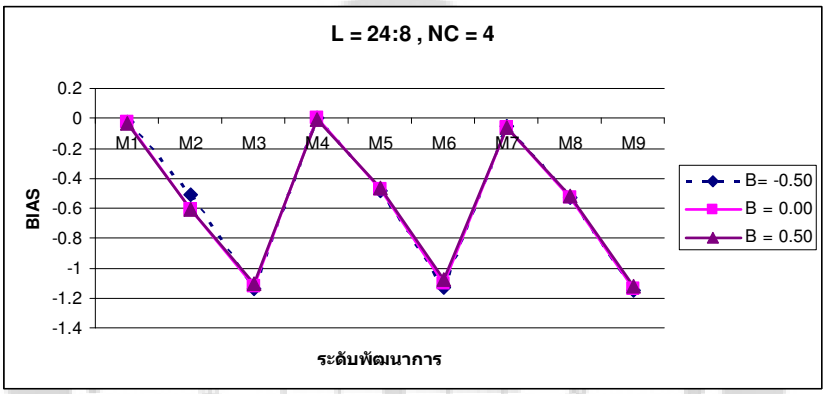


33 ค

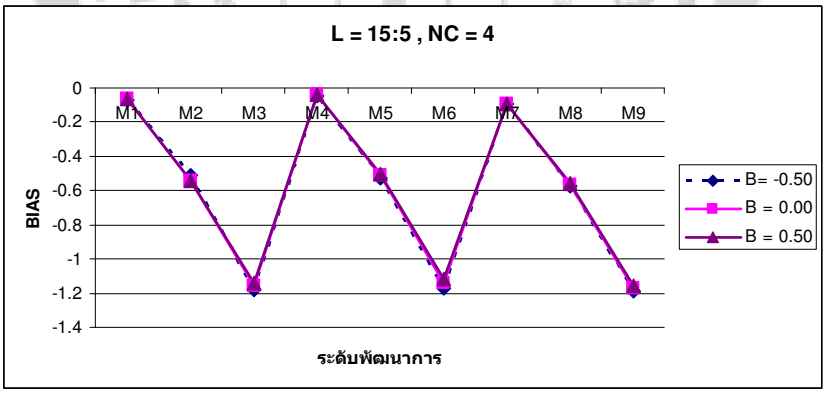
ภาพประกอบ 33 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 3 ระดับ และความยาวของแบบทดสอบผสมเป็น 30:10 (33 ก) 24:8 (33 ข) 15:5 (33 ค)



34 ก

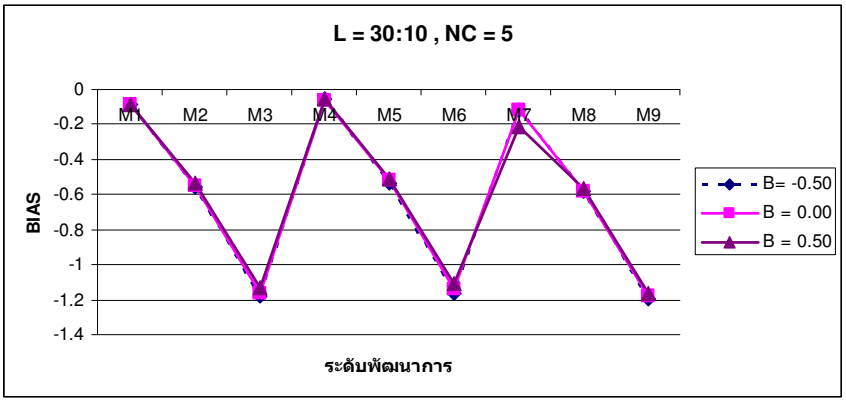


34 ข

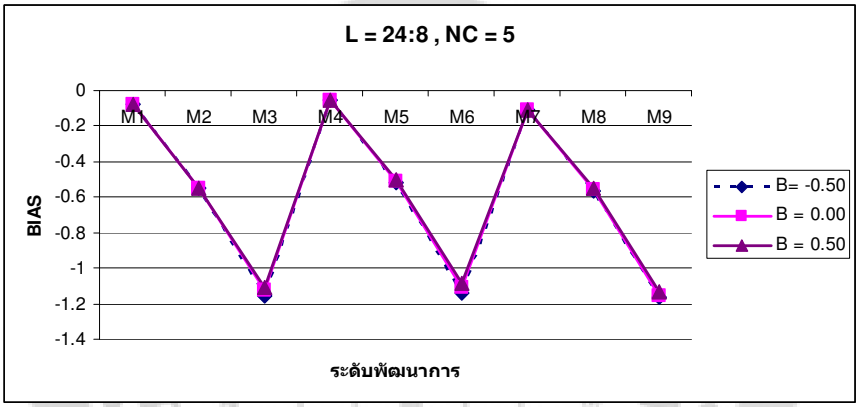


34 ค

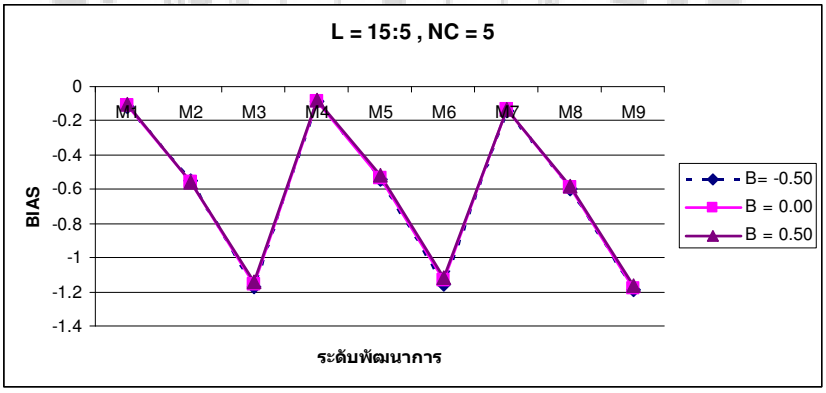
ภาพประกอบ 34 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 4 ระดับและความยาวของแบบทดสอบผสมเป็น 30:10 (34 ก) 24:8 (34 ข) 15:5 (34 ค)



35 ก



35 ข



35 ค

ภาพประกอบ 31 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 5 ระดับ และความยาวของแบบทดสอบผสมเป็น 30:10 (35 ก) 24:8 (35 ข) 15:5 (35 ค)

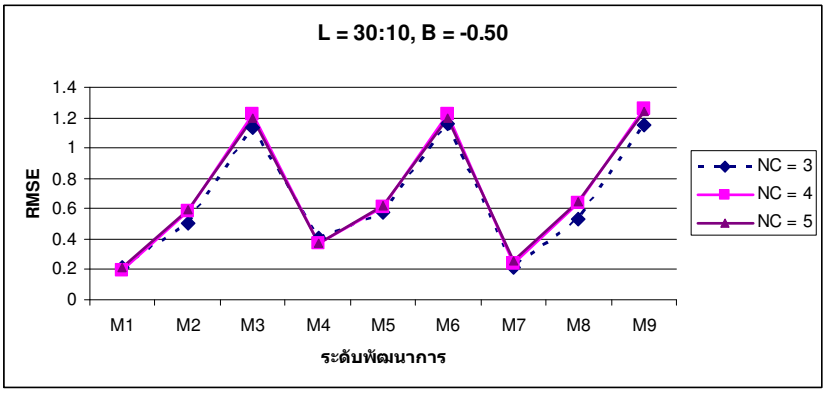
3.1.2 การศึกษาค่ารากที่สองของความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถ

3.1.2.1 ความยาวคงที่

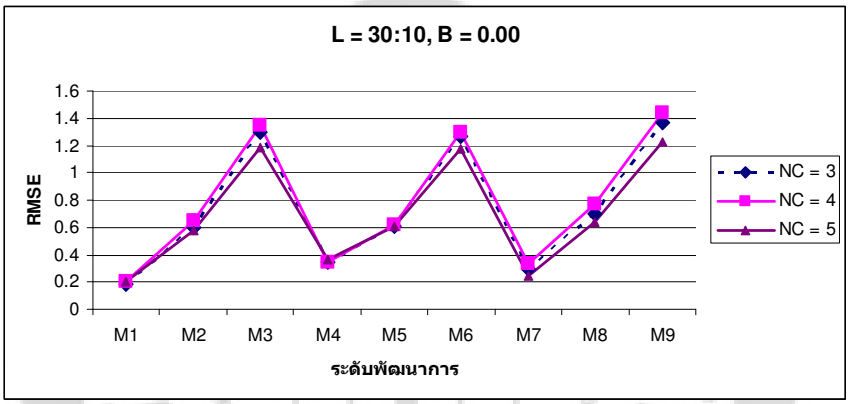
ตาราง 27 ค่าเฉลี่ยค่า RMSE ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมี ความยาว และความยากคงที่ จำแนกตามการให้คะแนน และระดับพัฒนาการ

L	B	NC	M1	M2	M3	M4	M5	M6	M7	M8	M9
		3	0.210	0.502	1.133	0.404	0.575	1.162	0.210	0.535	1.149
	-0.50	4	0.199	0.581	1.219	0.368	0.613	1.223	0.243	0.639	1.256
		5	0.215	0.590	1.199	0.371	0.622	1.201	0.253	0.644	1.241
		3	0.181	0.598	1.294	0.347	0.606	1.262	0.293	0.698	1.366
30:10	0.00	4	0.200	0.647	1.349	0.344	0.619	1.299	0.338	0.769	1.436
		5	0.206	0.578	1.180	0.367	0.609	1.175	0.243	0.637	1.230
		3	0.171	0.481	1.089	0.367	0.540	1.097	0.192	0.535	1.134
	0.50	4	0.168	0.544	1.157	0.344	0.573	1.148	0.222	0.535	1.212
		5	0.182	0.557	1.151	0.343	0.586	1.140	0.298	0.623	1.208
		3	0.202	0.611	1.307	0.349	0.624	1.300	0.311	0.699	1.356
	-0.50	4	0.221	0.552	1.167	0.400	0.602	1.182	0.225	0.591	1.193
		5	0.219	0.581	1.174	0.373	0.618	1.178	0.249	0.628	1.210
		3	0.212	0.471	1.074	0.407	0.548	1.093	0.196	0.506	1.102
24:8	0.00	4	0.202	0.538	1.138	0.392	0.586	1.147	0.213	0.585	1.180
		5	0.208	0.564	1.144	0.367	0.600	1.147	0.239	0.618	1.197
		3	0.181	0.460	1.052	0.370	0.526	1.065	0.193	0.508	1.094
	0.50	4	0.182	0.520	1.116	0.361	0.562	1.114	0.208	0.576	1.162
		5	0.184	0.548	1.126	0.338	0.580	1.118	0.232	0.609	1.181
		3	0.220	0.508	1.111	0.384	0.571	1.140	0.248	0.545	1.134
	-0.50	4	0.219	0.592	1.206	0.363	0.618	1.208	0.277	0.650	1.243
		5	0.233	0.609	1.199	0.359	0.629	1.191	0.294	0.670	1.244
		3	0.216	0.591	1.241	0.351	0.611	1.234	0.308	0.667	1.290
15:5	0.00	4	0.214	0.576	1.182	0.366	0.605	1.180	0.263	0.634	1.224
		5	0.229	0.594	1.176	0.363	0.618	1.169	0.281	0.655	1.228
		3	0.191	0.489	1.086	0.367	0.546	1.095	0.224	0.543	1.126
	0.50	4	0.194	0.558	1.160	0.348	0.584	1.152	0.249	0.624	1.211
		5	0.208	0.578	1.159	0.344	0.598	1.145	0.269	0.645	1.215

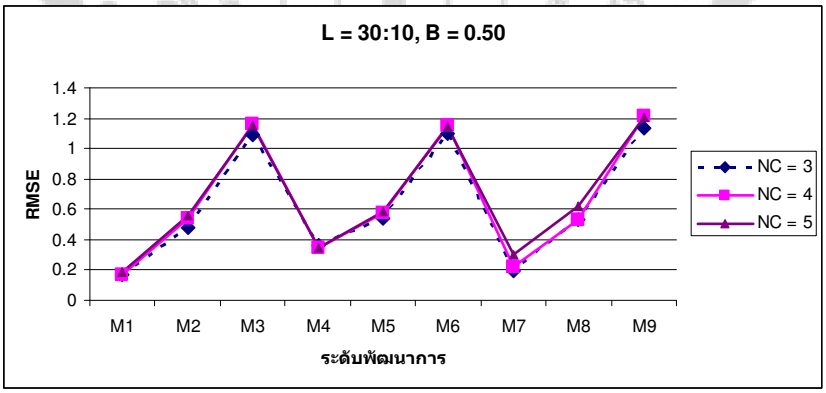
*___ หมายถึง มีค่า RMSE ต่ำที่สุด ภายใต้เงื่อนไขความยาว และ ความยากเท่ากัน



36 ก

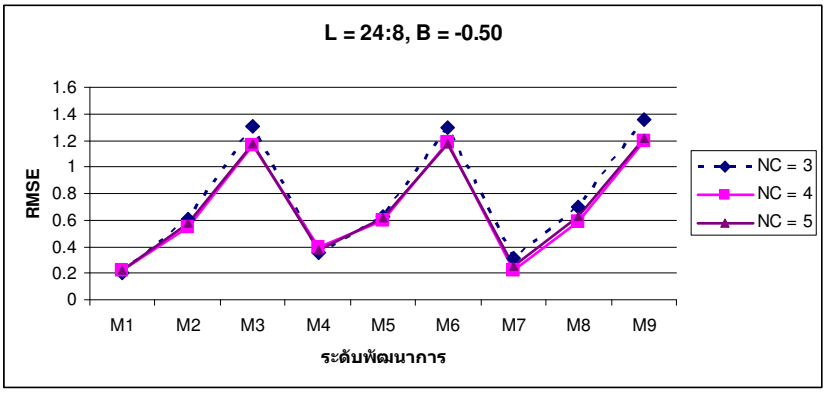


36 ข

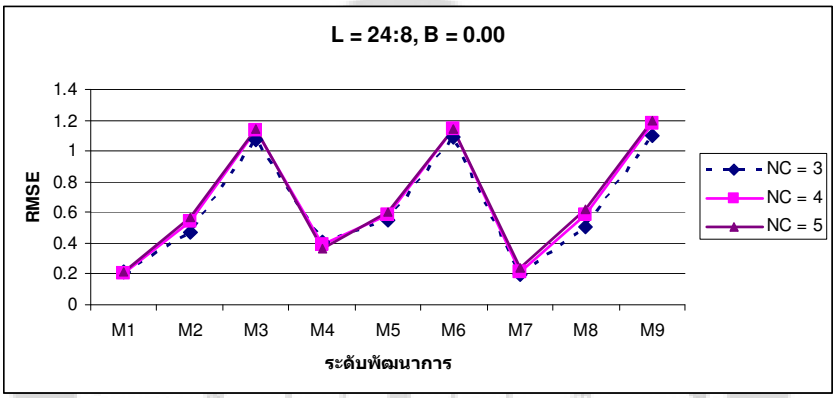


36 ค

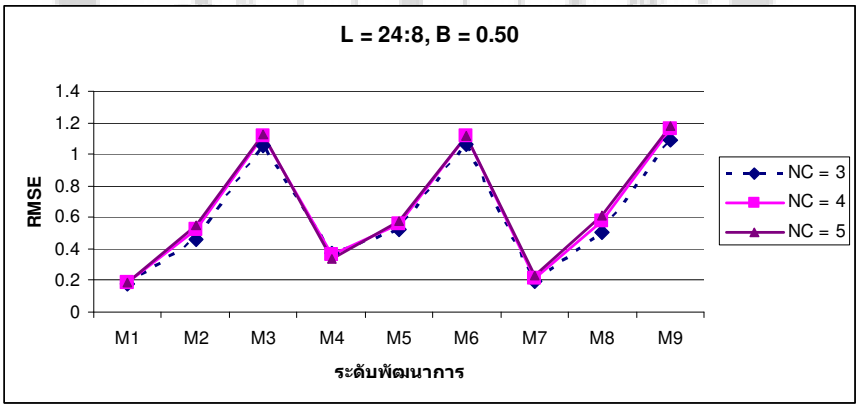
ภาพประกอบ 36 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก -0.50 (36 ก) 0.00 (36 ข) 0.50 (36 ค)



37 ก

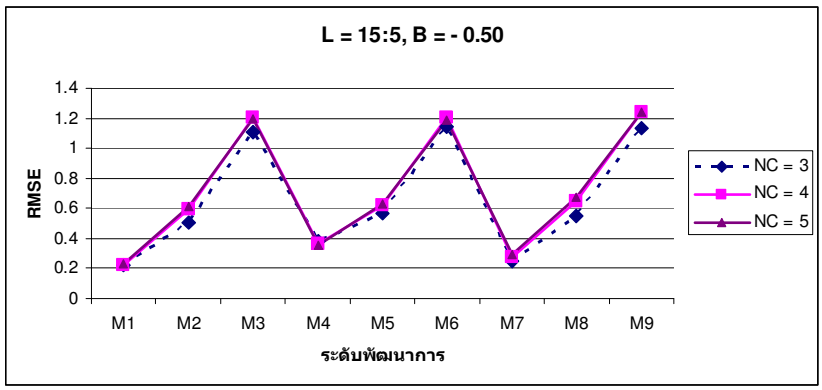


37 ข

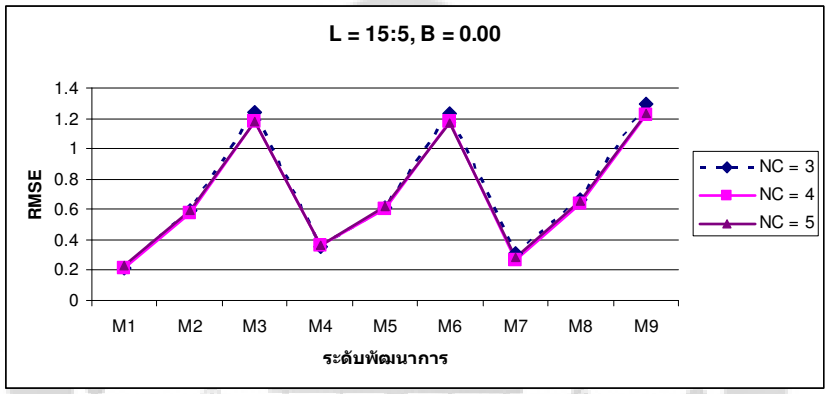


37 ค

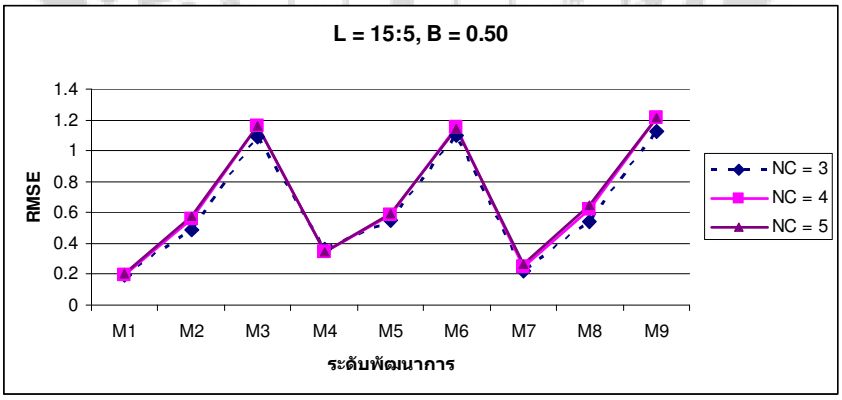
ภาพประกอบ 37 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 (37 ก) 0.00 (37 ข) 0.50 (37 ค)



38 ก



38 ข



38 ค

ภาพประกอบ 38 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก -0.50 (38 ก) 0.00 (38 ข) 0.50 (38 ค)

กรณีแบบทดสอบผสมมีความยาว 30:10

จากตาราง 27 และภาพประกอบ 36 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก -0.50 กลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 3 ระดับ ยกเว้นกลุ่ม M1 และ กลุ่ม M4 ที่พบว่าค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 4 ระดับ

เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก 0.00 พบว่ากลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 5 ระดับ ส่วนกลุ่ม M1 และ M5 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 3 ระดับ และกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 4 ระดับ

เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก 0.50 พบว่ากลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 3 ระดับ ส่วนในกลุ่ม M1 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดกรณีให้คะแนน 4 ระดับ และกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดกรณีให้คะแนน 5 ระดับ

กรณีแบบทดสอบผสมมีความยาว 24:8

จากตาราง 27 และภาพประกอบ 37 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 พบว่ากลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 4 ระดับ ส่วนกลุ่ม M1 และ M6 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 5 ระดับ และกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 3 ระดับ

เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก 0.00 พบว่ากลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 3 ระดับ ส่วนกลุ่ม M1 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 4 ระดับ และกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 5 ระดับ

เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก 0.50 พบว่ากลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 3 ระดับ ยกเว้นกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 5 ระดับ

กรณีแบบทดสอบผสมมีความยาว 15:5

จากตาราง 27 และภาพประกอบ 38 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก -0.50 พบว่ากลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดที่ให้คะแนน 3 ระดับ ส่วนกลุ่ม M1 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้

คะแนน 4 ระดับ และกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 5 ระดับ

เมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก 0.00 พบว่ากลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 4 ระดับ ส่วนกลุ่ม M3 และ M6 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 5 ระดับมี และกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 3 ระดับ

เมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก 0.50 พบว่ากลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 3 ระดับ ยกเว้นกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 5 ระดับ

3.1.2.2 ความยากคงที่

กรณีความยากแบบทดสอบผสมเป็น -0.50

จากตาราง 28 และภาพประกอบ 39 (ก – ค) เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 3 ระดับ พบว่ากลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 15:5 ส่วนกลุ่ม M1 และ M4 ค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 24:8 และกลุ่ม M2 M7 และ M8 ค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 30:10

เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 4 ระดับ พบว่ากลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 24:8 ยกเว้นกลุ่ม M1 และ M4 ค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 30:10

เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 5 ระดับ พบว่ากลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 24:8 ยกเว้นกลุ่ม M1 ที่ค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 30:10 และ กลุ่ม M4 ที่ค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 15:5

กรณีความยากแบบทดสอบผสม เป็น 0.00

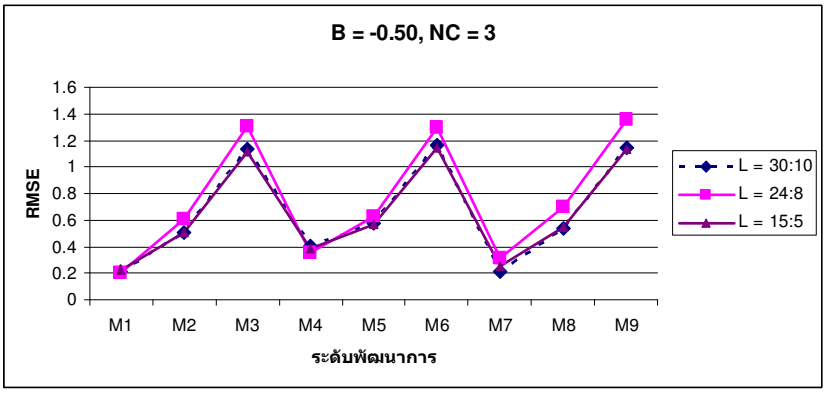
จากตาราง 28 และภาพประกอบ 40 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาก 0.00 และให้คะแนน 3 ระดับ หรือ 4 ระดับ กลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 24:8 ส่วนกลุ่ม M1 และ M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 30:10

เมื่อแบบทดสอบผสมมีความยาก 0.00 และให้คะแนน 5 ระดับ กลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 24:8 ยกเว้นกลุ่ม M1 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 30:10 และกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 15:5

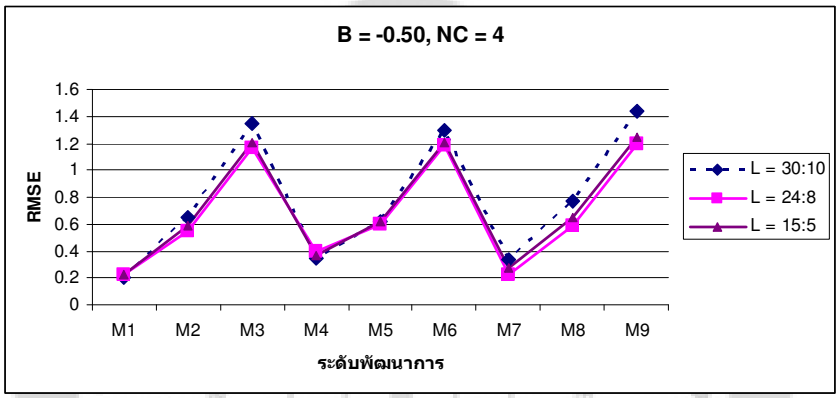
ตาราง 28 ค่าเฉลี่ยของค่า RMSE ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมี ความยาก และการให้คะแนนคงที่ จำแนกตามความยาวและระดับพัฒนาการ

B	NC	L	M1	M2	M3	M4	M5	M6	M7	M8	M9
-0.50	3	30:10	0.210	0.502	1.133	0.404	0.575	1.162	0.210	0.535	1.149
		24:8	0.202	0.611	1.307	0.349	0.624	1.300	0.311	0.699	1.356
		15:5	0.220	0.508	1.111	0.384	0.571	1.140	0.248	0.545	1.134
	4	30:10	0.200	0.647	1.349	0.344	0.619	1.299	0.338	0.769	1.436
		24:8	0.221	0.552	1.167	0.400	0.602	1.182	0.225	0.591	1.193
		15:5	0.219	0.592	1.206	0.363	0.618	1.208	0.277	0.650	1.243
	5	30:10	0.215	0.590	1.199	0.371	0.622	1.201	0.253	0.644	1.241
		24:8	0.219	0.581	1.174	0.373	0.618	1.178	0.249	0.628	1.210
		15:5	0.233	0.609	1.199	0.359	0.629	1.191	0.294	0.670	1.244
0.00	3	30:10	0.181	0.598	1.294	0.347	0.606	1.262	0.293	0.698	1.366
		24:8	0.212	0.471	1.074	0.407	0.548	1.093	0.196	0.506	1.102
		15:5	0.216	0.591	1.241	0.351	0.611	1.234	0.308	0.667	1.290
	4	30:10	0.200	0.647	1.349	0.344	0.619	1.299	0.338	0.769	1.436
		24:8	0.202	0.538	1.139	0.392	0.586	1.147	0.213	0.585	1.180
		15:5	0.214	0.576	1.182	0.366	0.605	1.180	0.263	0.634	1.224
	5	30:10	0.206	0.578	1.180	0.367	0.609	1.175	0.243	0.637	1.230
		24:8	0.208	0.564	1.145	0.367	0.600	1.147	0.239	0.618	1.197
		15:5	0.229	0.594	1.176	0.363	0.618	1.169	0.281	0.655	1.228
0.50	3	30:10	0.171	0.481	1.089	0.367	0.540	1.097	0.192	0.535	1.134
		24:8	0.181	0.460	1.053	0.370	0.526	1.065	0.193	0.508	1.094
		15:5	0.191	0.489	1.086	0.367	0.546	1.095	0.224	0.543	1.126
	4	30:10	0.168	0.544	1.157	0.344	0.573	1.148	0.222	0.535	1.212
		24:8	0.182	0.520	1.116	0.361	0.562	1.114	0.208	0.576	1.162
		15:5	0.194	0.558	1.160	0.348	0.584	1.152	0.249	0.624	1.211
	5	30:10	0.182	0.557	1.151	0.343	0.586	1.140	0.298	0.623	1.208
		24:8	0.184	0.548	1.126	0.338	0.580	1.118	0.232	0.609	1.181
		15:5	0.208	0.578	1.159	0.344	0.598	1.145	0.269	0.645	1.215

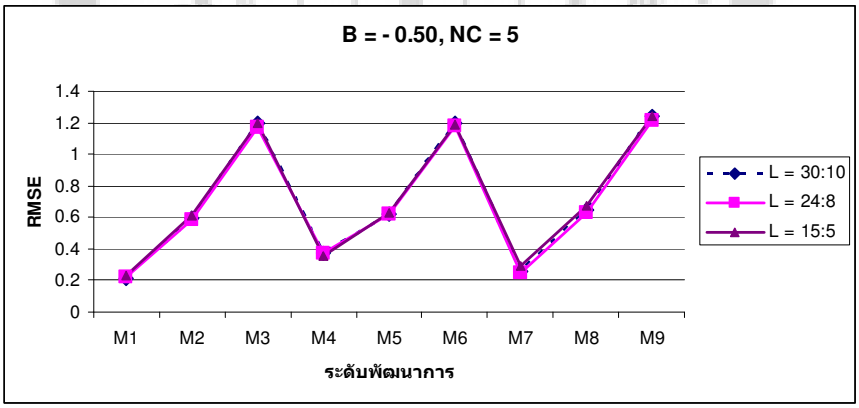
*___ หมายถึง มีค่า RMSE ต่ำที่สุด ภายใต้เงื่อนไขความยาก และ ระดับการให้คะแนนเท่ากัน



39 ก

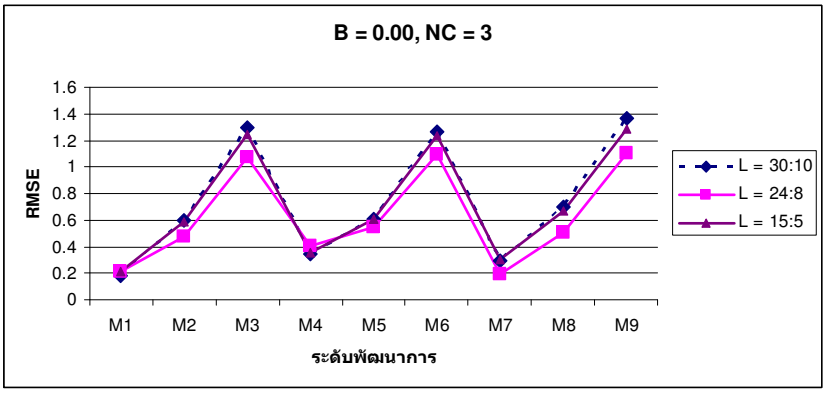


39 ข

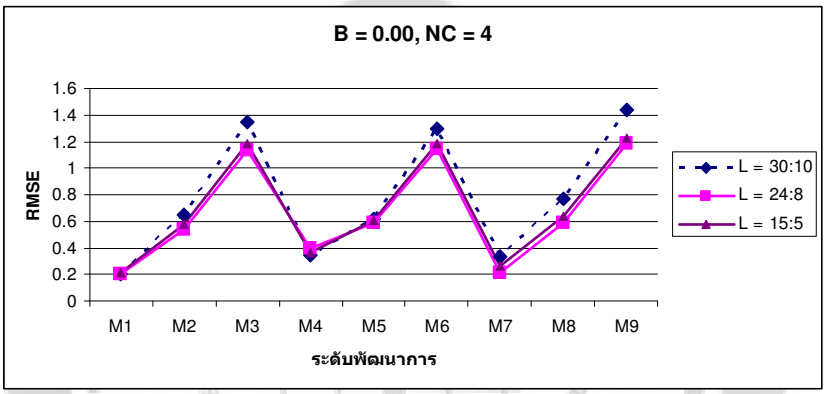


39 ค

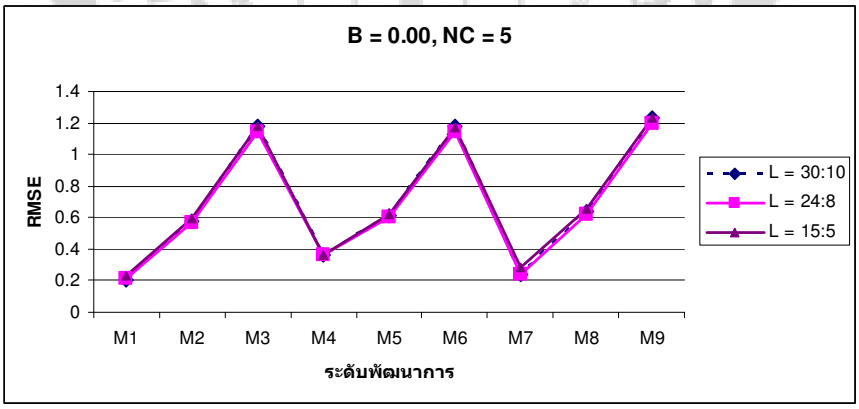
ภาพประกอบ 39 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยากของแบบทดสอบผสมเป็น -0.50 และให้คะแนน 3 ระดับ (39 ก) 4 ระดับ (39 ข) 5 ระดับ (39 ค)



40 ก

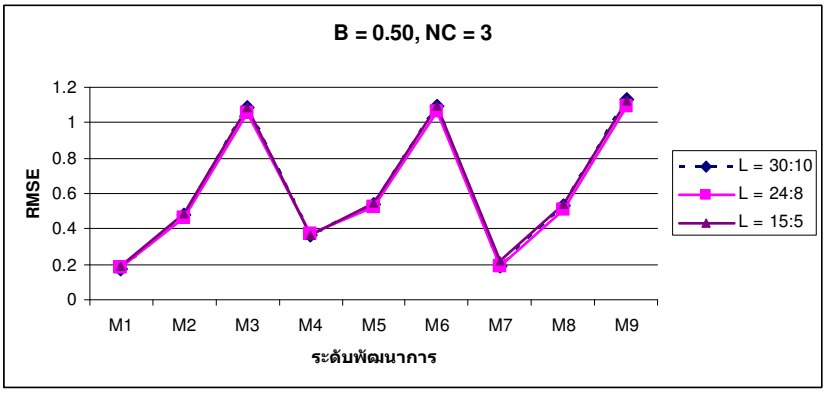


40 ข

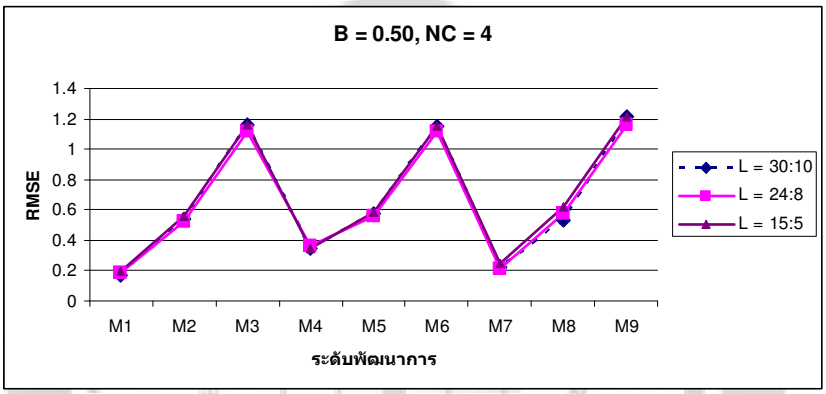


40 ค

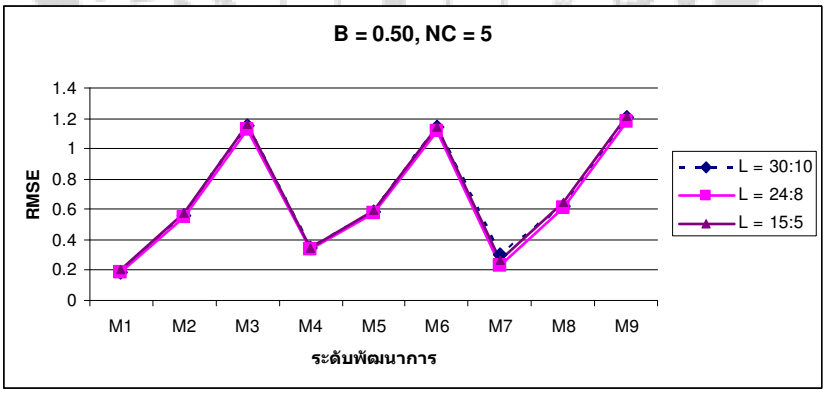
ภาพประกอบ 40 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยากของแบบทดสอบผสมเป็น 0.00 และให้คะแนน 3 ระดับ (40 ก) 4 ระดับ (40 ข) 5 ระดับ (40 ค)



41 ก



41 ข



41 ค

ภาพประกอบ 41 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยากของแบบทดสอบผสมเป็น 0.50 และการให้คะแนน 3 ระดับ (41 ก) 4 ระดับ (41 ข) 5 ระดับ (41 ค)

กรณีความยากแบบทดสอบผสม เป็น 0.50

จากตาราง 28 และภาพประกอบ 41 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 3 ระดับ กลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 24:8 ส่วนกลุ่ม M1 M4 และ M7 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 30:10

เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 4 ระดับ กลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 24:8 ยกเว้นกลุ่ม M1 M4 และ M8 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 30:10

เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 5 ระดับ กลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 24:8 ยกเว้นกลุ่ม M1 มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมมีความยาว 30:10

3.1.2.3 การให้คะแนนคงที่

กรณีให้คะแนน 3 ระดับ

จากตาราง 29 และภาพประกอบ 42 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 30:10 และให้คะแนน 3 ระดับ กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50 ส่วนกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.00 และกลุ่ม M6 ค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50

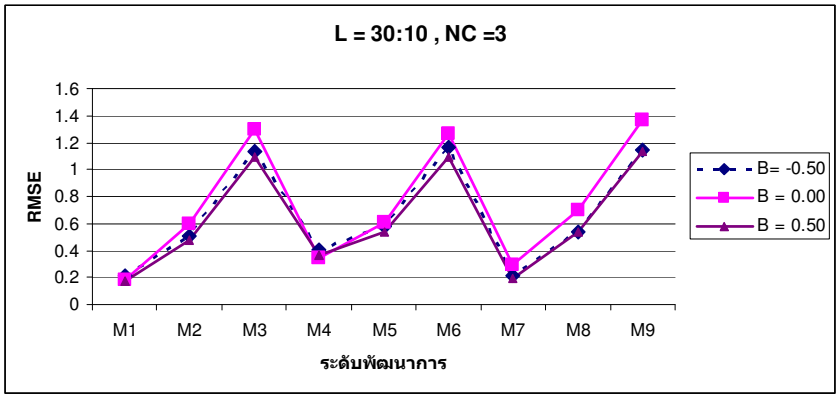
เมื่อแบบทดสอบผสมมีความยาว 24:8 และให้คะแนน 3 ระดับ กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50 ส่วนกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50 มีค่าต่ำที่สุด และกลุ่ม M8 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.00

เมื่อแบบทดสอบผสมมีความยาว15:5 และให้คะแนน 3 ระดับ กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50 ส่วนกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.00

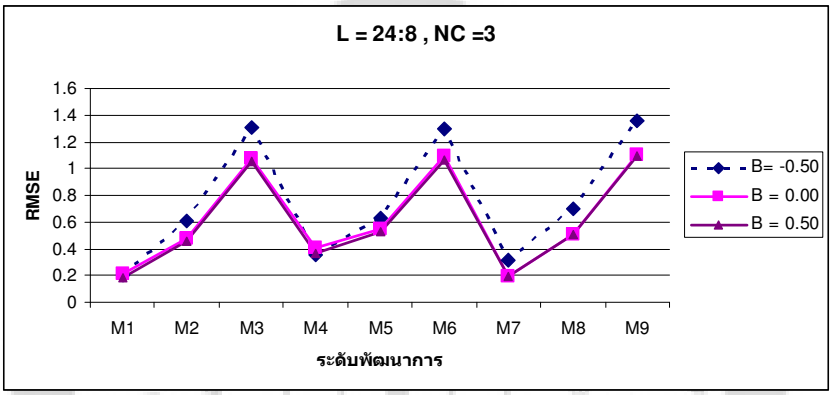
ตาราง 29 ค่าเฉลี่ยของค่า RMSE ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีการให้คะแนนและความยาวคงที่ จำแนกตามความยากและระดับพัฒนาการ

NC	L	B	M1	M2	M3	M4	M5	M6	M7	M8	M9
		-0.50	0.210	0.502	1.133	0.404	0.575	1.162	0.210	0.535	1.149
	30:10	0.00	0.181	0.598	1.294	0.347	0.606	1.262	0.293	0.698	1.366
		0.50	0.171	0.481	1.089	0.367	0.540	1.097	0.192	0.535	1.134
		-0.50	0.202	0.611	1.307	0.349	0.624	1.300	0.311	0.699	1.356
3 ระดับ	24:8	0.00	0.212	0.471	1.074	0.407	0.548	1.093	0.196	0.506	1.102
		0.50	0.181	0.460	1.052	0.370	0.526	1.065	0.193	0.508	1.094
		-0.50	0.220	0.508	1.111	0.384	0.571	1.140	0.248	0.545	1.134
	15:5	0.00	0.216	0.591	1.241	0.351	0.611	1.234	0.308	0.667	1.290
		0.50	0.191	0.489	1.086	0.367	0.546	1.095	0.224	0.543	1.126
		-0.50	0.200	0.647	1.349	0.344	0.619	1.299	0.338	0.769	1.436
	30:10	0.00	0.200	0.647	1.349	0.344	0.619	1.299	0.338	0.769	1.436
		0.50	0.168	0.544	1.157	0.344	0.573	1.148	0.222	0.535	1.212
		-0.50	0.221	0.552	1.167	0.400	0.602	1.182	0.225	0.591	1.193
4 ระดับ	24:8	0.00	0.202	0.538	1.138	0.392	0.586	1.147	0.213	0.585	1.180
		0.50	0.182	0.520	1.116	0.361	0.562	1.114	0.208	0.576	1.162
		-0.50	0.219	0.592	1.206	0.363	0.618	1.208	0.277	0.650	1.243
	15:5	0.00	0.214	0.576	1.182	0.366	0.605	1.180	0.263	0.634	1.224
		0.50	0.194	0.558	1.160	0.348	0.584	1.152	0.249	0.624	1.211
		-0.50	0.215	0.590	1.199	0.371	0.622	1.201	0.253	0.644	1.241
	30:10	0.00	0.206	0.578	1.180	0.367	0.609	1.175	0.243	0.637	1.230
		0.50	0.182	0.557	1.151	0.343	0.586	1.140	0.298	0.623	1.208
		-0.50	0.219	0.581	1.174	0.373	0.618	1.178	0.249	0.628	1.210
5 ระดับ	24:8	0.00	0.208	0.564	1.144	0.367	0.600	1.147	0.239	0.618	1.197
		0.50	0.184	0.548	1.126	0.338	0.580	1.118	0.232	0.609	1.181
		-0.50	0.233	0.609	1.199	0.359	0.629	1.191	0.294	0.670	1.244
	15:5	0.00	0.229	0.594	1.176	0.363	0.618	1.169	0.281	0.655	1.228
		0.50	0.208	0.578	1.159	0.344	0.598	1.145	0.269	0.645	1.215

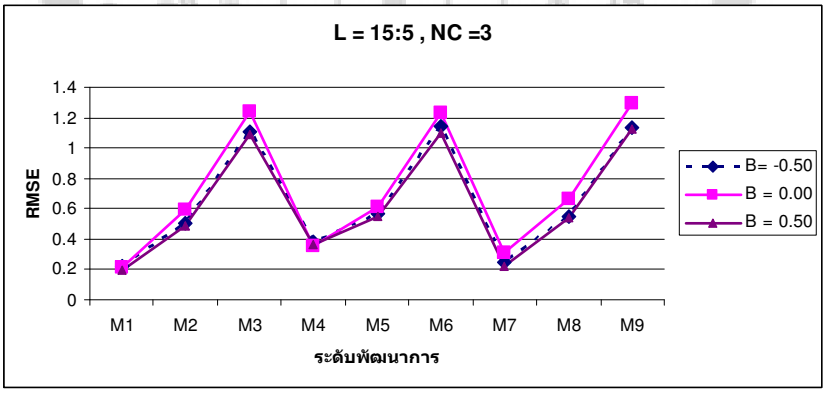
*___ หมายถึง มีค่า RMSE ต่ำที่สุด ภายใต้เงื่อนไขความยาก และ ระดับการให้คะแนนเท่ากัน



42 ก



42 ข



42 ค

ภาพประกอบ 42 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมยาว 30:10 (42 ก) 24:8 (42 ข) 15:5 (42 ค)

กรณีมีการให้คะแนน 4 ระดับ

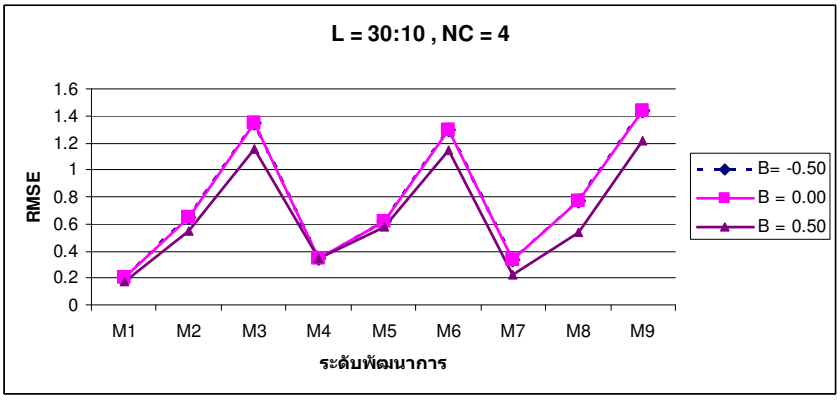
จากตาราง 29 และภาพประกอบ 43 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 30:10 และให้คะแนน 4 ระดับ กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50 ส่วนกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถค่าเท่ากันทุกระดับความยากของแบบทดสอบผสม และกลุ่ม M5 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดเท่ากันในกรณีที่แบบทดสอบผสมมีความยาก -0.50 และ 0.00

เมื่อแบบทดสอบผสมมีความยาว 24:8 และ 15:5 และให้คะแนน 4 ระดับ ทุกกลุ่มพัฒนาการ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50

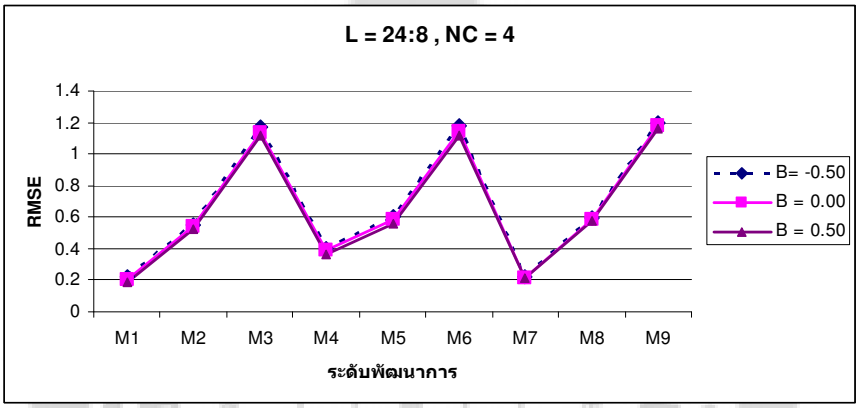
กรณีให้คะแนน 5 ระดับ

จากตาราง 29 และภาพประกอบ 44 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 30:10 และให้คะแนน 5 ระดับ กลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50 ส่วนกลุ่ม M7 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.00

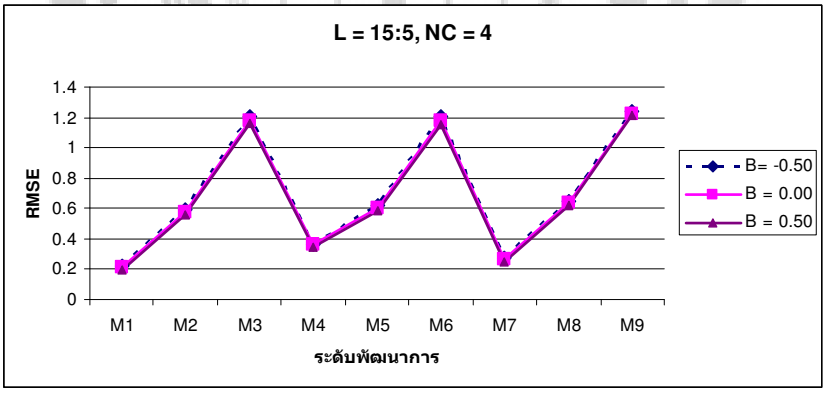
เมื่อแบบทดสอบผสมมีความยาว 24:8 และ 15:5 และให้คะแนน 5 ระดับ ทุกกลุ่มพัฒนาการมีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50



43 ก

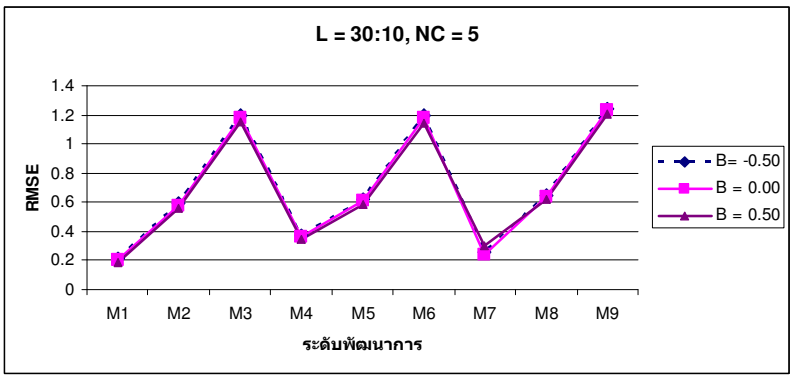


43 ข

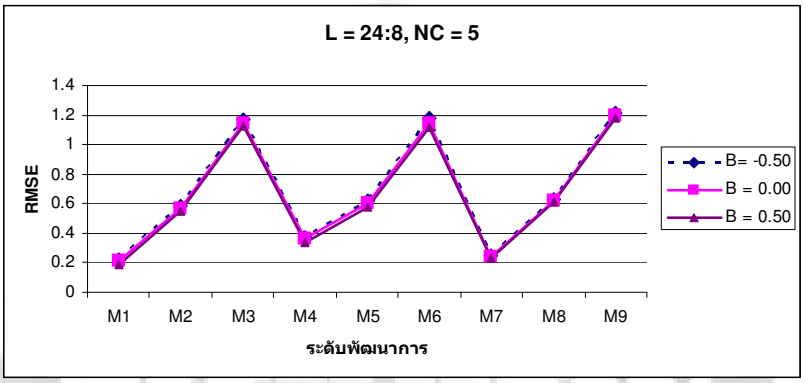


43 ค

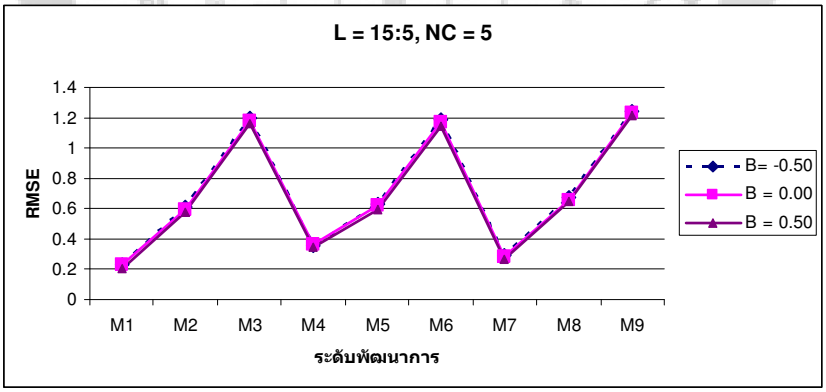
ภาพประกอบ 43 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อจำนวนรายการคำตอบเป็น 4 และความยาวเป็น 30:10 (43 ก) 24:8 (43 ข) 15:5 (43 ค)



44 ก



44 ข



44 ค

ภาพประกอบ 44 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อจำนวนรายการคำตอบเป็น 5 และความยาวเป็น 30:10 (44 ก) 24:8 (44 ข) 15:5 (44 ค)

3.2 การเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถ

ผู้วิจัยเปรียบเทียบความแตกต่างของค่าเฉลี่ยความลำเอียง (BIAS) และค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าประมาณพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันในแต่ละเงื่อนไข เพื่อทดสอบสมมติฐานการวิจัยข้อ 2 ที่ตั้งไว้ว่า *"ค่าประมาณพารามิเตอร์ความสามารถของผู้สอบที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีค่าความลำเอียง (Bias) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) เข้าใกล้ศูนย์ ทุกระดับพัฒนาการความสามารถของผู้สอบ"*

ผลการวิเคราะห์แยกนำเสนอผลการเปรียบเทียบความแม่นยำออกเป็น 2 ส่วน โดยส่วนแรกเป็นการเปรียบเทียบค่าเฉลี่ยของความลำเอียง (BIAS) ของค่าประมาณพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน และส่วนที่สองเป็นการเปรียบเทียบค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าประมาณพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันปรากฏผลดังนี้

3.2.1 การเปรียบเทียบความแตกต่างของค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ

ผู้วิจัยนำค่า BIAS ที่คำนวณจากกลุ่มระดับพัฒนาการ 9 กลุ่ม ภายใต้เงื่อนไขความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกัน มาวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (Repeated Measures) ปรากฏผลดังตาราง 30

ตาราง 30 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ย BIAS กรณีแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ

แหล่งความแปรปรวน	SS	df	MS	F	ระดับนัยสำคัญ	ขนาดของผล
ทดสอบผลตัวแปร Within						
ความยาว	0.030	2.000	0.015	8.890*	0.003	0.526
ความยาก	0.064	1.132	0.057	13.564*	0.004	0.629
การให้คะแนน	0.144	1.031	0.139	16.849*	0.003	0.678
ความยาว × ความยาก	0.084	1.556	0.054	22.652*	0.000	0.739
ความยาว × การให้คะแนน	0.019	1.814	0.01	11.727*	0.001	0.594
ความยาก × การให้คะแนน	0.029	2.004	0.014	25.641*	0.000	0.762
ความยาว × ความยาก × การให้คะแนน	0.112	1.541	0.073	31.812*	0.000	0.799

* มีนัยสำคัญทางสถิติที่ระดับ .05

ผลการวิเคราะห์ตามตาราง 30 พบว่า เมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด ความยากของแบบทดสอบผสม 3 ระดับ และการให้คะแนนของแบบทดสอบผสม 3 ขนาด ค่าเฉลี่ย BIASของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทุกระดับตัวแปร แสดงว่าเมื่อแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน ส่งผลให้ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถแตกต่างกัน

แต่อย่างไรก็ตามพบว่ามีปฏิสัมพันธ์ระหว่างความยาว กับ ความยากของแบบทดสอบ ความยาวและการให้คะแนนของแบบทดสอบ ความยาก และการให้คะแนนของแบบทดสอบ และความยาวความยากและการให้คะแนนของแบบทดสอบ ที่ระดับนัยสำคัญ .05 แสดงว่าความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถเป็นผลมาจากการส่งผลร่วมกันระหว่างตัวแปร ความยาว ความยาก และการให้คะแนนของแบบทดสอบ

เมื่อผลการทดสอบพบว่าแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน ส่งผลให้ค่า BIASของพัฒนาการความสามารถแตกต่างกัน ผู้วิจัยจึงเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ ในแต่ละคู่ระดับของ 3 เงื่อนไข ได้แก่ ความยาว ความยาก และ การให้คะแนน โดยเสนอผลการเปรียบเทียบรายคู่เป็น 3 ตอนเพื่อทดสอบสมมติฐานย่อยของสมมติฐานการวิจัยข้อ 2 จำนวน 3 ข้อดังนี้

3.2.1.1 การเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถตามสมมติฐานข้อ 2.1 ที่ตั้งไว้ว่า *"แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีค่าความลำเอียง (BIAS) และค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ทุกระดับความยาก และ ระดับการให้คะแนนของแบบทดสอบผสม"* ปรากฏผลการวิเคราะห์ดังตาราง 31

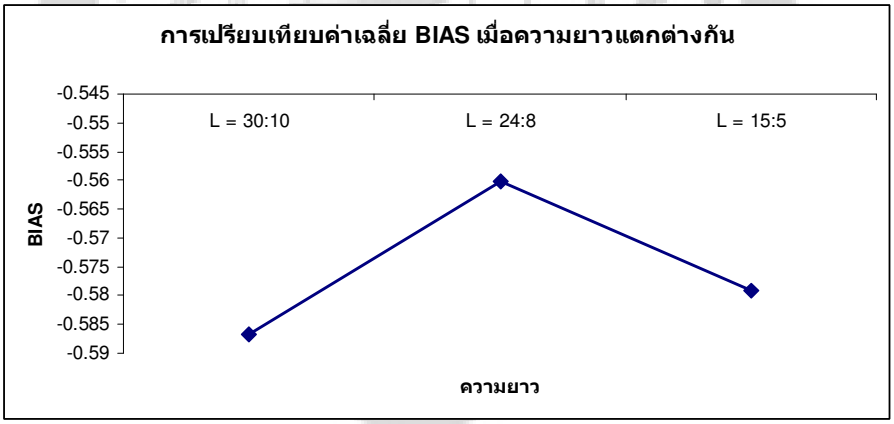
จากตาราง 31 พบว่า เมื่อความยากและการให้คะแนนคงที่ ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยาว 30:10 มีค่าสูงสุด รองลงมาคือ 15:5 และ 24:8 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ตามระดับความยาวของแบบทดสอบผสมพบว่าแบบทดสอบขนาด 30:10 มีความลำเอียงสูงกว่า ขนาด 24:8 และ ขนาด 15:5 มีความลำเอียงสูงกว่า 24:8 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบขนาด 30:10 กับ15:5 มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 31 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาว	Mean	L = 30:10	L = 24:8	L = 15:5
	L = 30:10	-0.587	-	0.033*	0.009
	L = 24:8	-0.554		-	0.024*
	L = 15:5	-0.577			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

เพื่อให้เห็นทิศทางและความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถเมื่อแบบทดสอบมีความยาวแตกต่างกัน 3 ขนาด ผู้วิจัยจึงนำเสนอเป็นแผนภาพดังภาพประกอบ 45



ภาพประกอบ 45 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน 3 ขนาด

จากภาพประกอบ 45 พบว่าทุกระดับความยาวของแบบทดสอบผสม ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเป็นลบ โดยแบบทดสอบผสมที่ยาว 24:8 มีความลำเอียงน้อยที่สุด รองลงมาคือ แบบทดสอบผสมที่ยาว 15:5 และ 30:10 ตามลำดับ

3.2.1.2 การเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

ผู้วิจัยเปรียบเทียบความลำเอียง (BIAS) ของพัฒนาการความสามารถตามสมมติฐาน ข้อ 2.2 ที่ตั้งไว้ว่า *"แบบทดสอบผสมที่มีระดับความยากของข้อสอบสอดคล้องกับระดับความสามารถผู้สอบจะมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่ระดับความยากของข้อสอบไม่สอดคล้องกับระดับความสามารถผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนนของแบบทดสอบผสม"* ปรากฏผลการวิเคราะห์ดังตาราง 32

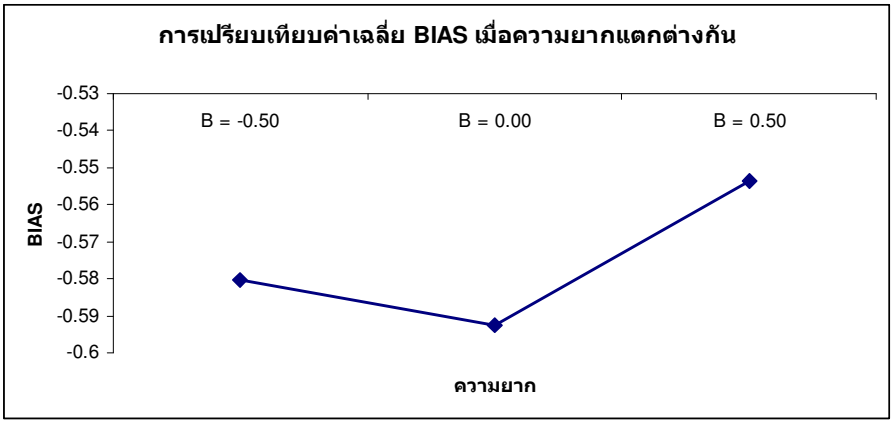
ตาราง 32 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาก	Mean	B = -0.50	B = 0.00	B = 0.50
	B = -0.50	-0.580	-	0.012	0.024*
	B = 0.00	-0.592		-	0.036*
	B = 0.50	-0.556			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 32 พบว่า เมื่อความยาวและรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่มีค่าความยาก 0.00 มีค่าเฉลี่ย BIAS สูงสุด รองลงมาคือ แบบทดสอบผสมที่มีความยาก -0.50 และ 0.50 ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ตามระดับความยากของแบบทดสอบผสมพบว่า แบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบผสมที่มีความยาก 0.50 และ แบบทดสอบผสมที่มีความยาก 0.00 มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบที่มีความยาก 0.50 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่มีความยาก 0.00 แบบทดสอบที่มีความยาก -0.50 มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ผู้วิจัยนำเสนอแผนภาพเพื่อให้เห็นทิศทางและความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน 3 ระดับ ดังภาพประกอบ 46



ภาพประกอบ 46 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน 3 ระดับ

จากภาพประกอบ 46 พบว่าทุกระดับความยากของแบบทดสอบผสม ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเป็นลบ โดยแบบทดสอบผสมที่มีความยาก 0.00 มีความลำเอียงมากที่สุด รองลงมาคือ แบบทดสอบผสมที่มีความยาก -0.50 และ 0.50 ตามลำดับ

3.2.1.3 การเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถเมื่อการให้คะแนนแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ข้อ 2.3 ที่ตั้งไว้ว่า *"แบบทดสอบผสมที่มีระดับคะแนนข้อสอบแบบให้คะแนนหลายค่ามากกว่าจะมีค่าความลำเอียง (Bias) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่มีระดับคะแนนข้อสอบแบบให้คะแนนหลายค่าน้อยกว่า ในทุกขนาดความยาวและระดับความยากของข้อสอบของแบบทดสอบผสม"* ปรากฏผลการวิเคราะห์ดังตาราง 33

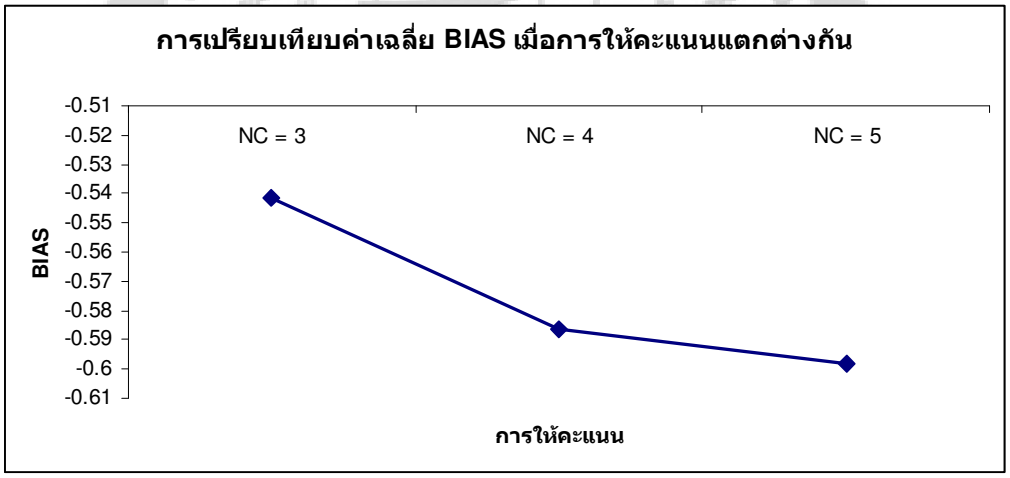
จากตาราง 33 พบว่าเมื่อแบบทดสอบผสมมีความยาวและความยากคงที่ ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่ให้คะแนน 5 ระดับมีค่าเฉลี่ย BIAS สูงสุด รองลงมาคือ แบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 3 ระดับ ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าแบบทดสอบผสมที่ให้คะแนน 5 ระดับ มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 3 ระดับ และแบบทดสอบผสมที่ให้คะแนน 4 ระดับ มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบผสมที่ให้คะแนน 3 ระดับ

ตาราง 33 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ระดับการให้คะแนน	Mean	NC=3	NC=4	NC=5
	NC=3	-0.538	-	0.044*	0.059*
	NC=4	-0.582		-	0.015*
	NC=5	-0.597			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

ผู้วิจัยนำเสนอแผนภาพเพื่อให้เห็นความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน 3 ระดับ ดังภาพประกอบ 47



ภาพประกอบ 47 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน 3 ขนาด

จากภาพประกอบ 47 พบว่าทุกระดับการให้คะแนนของแบบทดสอบผสมมีค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเป็นลบ โดยพบว่าแบบทดสอบผสมที่ให้คะแนน 5 ระดับ มีความลำเอียงมากที่สุด รองลงมาคือ แบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 3 ระดับตามลำดับ

3.2.2 การเปรียบเทียบค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถ

ผู้วิจัยเปรียบเทียบความแตกต่างของค่าเฉลี่ย RMSE ที่คำนวณจากกลุ่มพัฒนาการ 9 กลุ่ม ภายใต้เงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน โดยใช้การวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (Repeated Measures) ปรากฏผลดังตาราง 34

ตาราง 34 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ย RMSE กรณีแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ

แหล่งความแปรปรวน	SS	df	MS	F	ระดับนัยสำคัญ	ขนาดของผล
ทดสอบผลตัวแปร Within						
ความยาว	0.030	2.000	0.015	8.044*	0.004	0.501
ความยาก	0.120	2.000	0.060	35.560*	0.000	0.816
การให้คะแนน	0.028	2.000	0.014	8.314*	0.003	0.510
ความยาว × ความยาก	0.079	1.089	0.072	11.710*	0.007	0.594
ความยาว × การให้คะแนน	0.014	1.160	0.012	8.064*	0.016	0.502
ความยาก × การให้คะแนน	0.026	1.179	0.022	12.330*	0.005	0.606
ความยาว × ความยาก × การให้คะแนน	0.089	1.105	0.080	10.800*	0.009	0.574

* มีนัยสำคัญทางสถิติที่ระดับ .05

ผลการวิเคราะห์ตามตาราง 34 พบว่าเมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด ความยากของแบบทดสอบผสม 3 ระดับ และการให้คะแนนของแบบทดสอบผสม 3 ขนาด ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทุกระดับตัวแปร แสดงว่าเมื่อแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน ส่งผลให้ค่า RMSE ของพัฒนาการความสามารถแตกต่างกัน

แต่อย่างไรก็ตามพบว่ามีปฏิสัมพันธ์ระหว่างความยาว กับ ความยากของแบบทดสอบ ความยาวและการให้คะแนนของแบบทดสอบ ความยาก และการให้คะแนนของแบบทดสอบ และ ความยาวความยากและการให้คะแนนของแบบทดสอบ ที่ระดับนัยสำคัญ .05 แสดงว่าความแตกต่างของค่า RMSE ของพัฒนาการความสามารถเป็นผลมาจากการส่งผลร่วมกันระหว่างตัวแปร ความยาว ความยาก และการให้คะแนนของแบบทดสอบ

เมื่อผลการทดสอบพบว่าแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน ส่งผลให้ค่า RMSE ของพัฒนาการความสามารถแตกต่างกัน ผู้วิจัยจึงเปรียบเทียบค่าเฉลี่ยของรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถในแต่ละคู่ระดับของ 3 เงื่อนไข ได้แก่ ความยาว ความยาก และ การให้คะแนน โดยเสนอผลการเปรียบเทียบรายคู่เป็น 3 ตอนเพื่อทดสอบสมมติฐานย่อยของสมมติฐานการวิจัยข้อ 2 จำนวน 3 ข้อ ดังนี้

3.2.2.1การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนของพัฒนาการความสามารถกำลังสอง (RMSE) เพื่อทดสอบสมมติฐานข้อ 2.1 ที่ตั้งไว้ว่า *"แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ทุกระดับความยาก และ ระดับการให้คะแนนของแบบทดสอบผสม"* ปรากฏผลการวิเคราะห์ดังตาราง 35

ตาราง 35 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

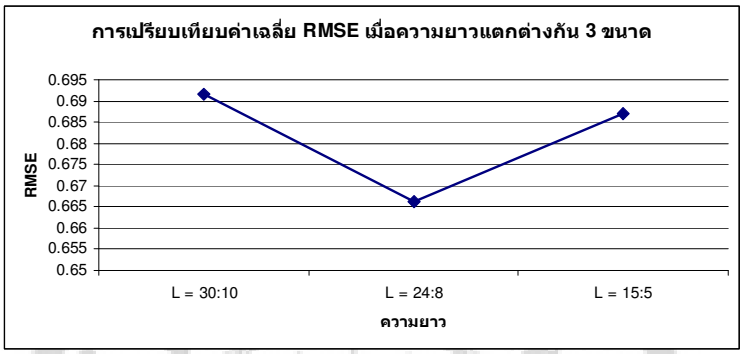
องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาว	Mean	L = 30:10	L = 24:8	L = 15:5
	L = 30:10	0.692	-	0.025*	0.005
	L = 24:8	0.666		-	0.021*
	L = 15:5	0.687			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 35 พบว่า เมื่อแบบทดสอบมีความยากและระดับการให้คะแนนคงที่ ค่าเฉลี่ยของค่า RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยาว 30:10 มีค่าเฉลี่ย RMSE สูงสุด รองลงมาคือ แบบทดสอบผสมที่มีความยาว 15:5 และ 24:8 ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ พบว่าแบบทดสอบผสมที่มีความยาว 30:10 มีค่าเฉลี่ย RMSE สูงกว่าแบบทดสอบผสมที่มีความยาว 24:8 และ แบบทดสอบผสม

ที่มีความยาว 24:8 มีค่าเฉลี่ย RMSE สูงกว่าแบบทดสอบผสมที่มีความยาว 15:5 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบผสมที่มีความยาว 30:10 กับแบบทดสอบที่มีความยาว 15:5 มีค่าเฉลี่ย RMSE แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

เพื่อให้เห็นความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถ เมื่อแบบทดสอบมีความยาวแตกต่างกัน 3 ขนาด ผู้วิจัยจึงนำเสนอเป็นแผนภาพดังภาพประกอบ 48



ภาพประกอบ 48 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน 3 ขนาด

จากภาพประกอบ 48 พบว่าค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาว 30:10 มีค่ามากที่สุด รองลงมาคือแบบทดสอบผสมที่มีความยาว 15:5 และ 24:8 ตามลำดับ

3.2.2.2 การเปรียบเทียบค่าเฉลี่ยของความคลาดเคลื่อนของพัฒนาการความสามารถกำลังสอง (RMSE) เมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมีความยากแตกต่างกันเพื่อทดสอบสมมติฐานข้อ 2.2 ที่ตั้งไว้ว่า *"แบบทดสอบผสมที่มีระดับความยากของข้อสอบสอดคล้องกับระดับความสามารถผู้สอบจะมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่ระดับความยากของข้อสอบไม่สอดคล้องกับระดับความสามารถผู้สอบในทุกขนาดความยาวและระดับการให้คะแนน"* ปรากฏผลการวิเคราะห์ดังตาราง 36

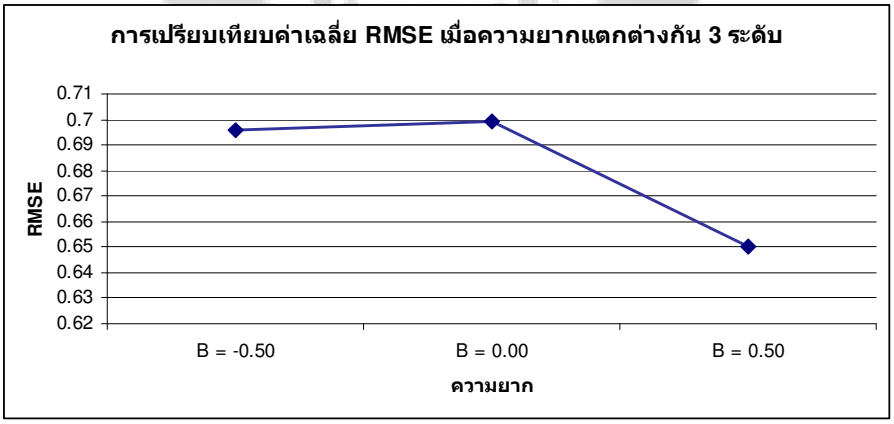
ตาราง 36 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาก	Mean	B = -0.50	B = 0.00	B = 0.50
	B = -0.50	0.696	-	0.003	0.045*
	B = 0.00	0.699		-	0.049*
	B = 0.50	0.650			-

* มีนัยสำคัญทางสถิติที่ระดับ 0.05

จากตาราง 36 พบว่าเมื่อแบบทดสอบมีความยากและระดับการให้คะแนนคงที่ ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากแบบทดสอบผสมที่มีความยาก 0.00 มีค่าเฉลี่ย RMSE สูงสุด รองลงมาคือ แบบทดสอบผสมที่มีความยาก -0.50 และ 0.50 ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ตามระดับความยากพบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ย RMSE สูงกว่าแบบทดสอบผสมที่มีความยาก 0.50 และแบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ย RMSE สูงกว่าแบบทดสอบที่มีความยาก 0.00 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ผู้วิจัยนำเสนอแผนภาพเพื่อให้เห็นทิศทางและความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน 3 ระดับดังภาพประกอบ 49



ภาพประกอบ 49 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน 3 ระดับ

จากภาพประกอบ 49 พบว่าค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาก 0.00 มีค่ามากที่สุด รองลงมาคือ แบบทดสอบผสมที่มีความยาก -0.50 และ 0.50 ตามลำดับ

3.2.2.3 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกันเพื่อทดสอบสมมติฐานข้อ 2.3 ที่ตั้งไว้ว่า *"แบบทดสอบผสมที่มีระดับคะแนนมากกว่าจะมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่ากว่าแบบทดสอบผสมที่มีระดับคะแนนน้อยกว่า ในทุกขนาดความยาวและระดับความยากของข้อสอบของแบบทดสอบผสม"* ปรากฏผลการวิเคราะห์ดังตาราง 37

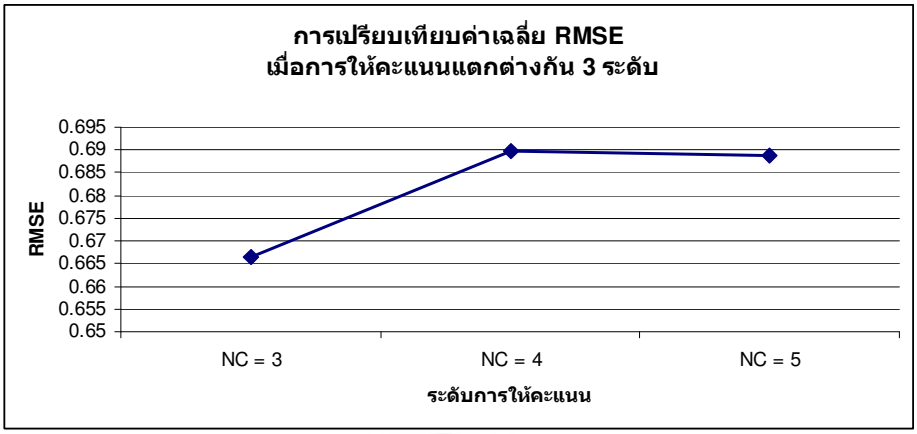
ตาราง 37 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ระดับการให้คะแนน	Mean	NC=3	NC=4	NC=5
	NC = 3	0.667	-	0.023*	0.022*
	NC = 4	0.690		-	0.001
	NC = 5	0.689			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 37 เมื่อแบบทดสอบผสมมีความยาวและความยากคงที่ แบบทดสอบผสมที่มีการให้คะแนน 4 ระดับมีค่าเฉลี่ยของค่า RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันสูงที่สุด รองลงมาคือ แบบทดสอบผสมที่ให้คะแนน 5 ระดับ และ 3 ระดับ ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ตามระดับการให้คะแนนของแบบทดสอบผสม พบว่าแบบทดสอบผสมที่ให้คะแนน 4 ระดับ มีค่าเฉลี่ย RMSE สูงกว่า แบบทดสอบผสมที่ให้คะแนน 5 ระดับ และ 3 ระดับ และแบบทดสอบผสมที่ให้คะแนน 4 ระดับ มีค่าเฉลี่ย RMSE สูงกว่าแบบทดสอบผสมที่ให้คะแนน 3 ระดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ผู้วิจัยนำเสนอแผนภาพเพื่อให้เห็นความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน 3 ระดับดังภาพประกอบ 50



ภาพประกอบ 50 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน 3 ระดับ

จากภาพประกอบ 50 พบว่าค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมให้คะแนน 4 ระดับ มีค่ามากที่สุด รองลงมาคือแบบทดสอบผสมที่ให้คะแนน 5 ระดับ และ 3 ระดับ ตามลำดับ

สรุปการตรวจสอบความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

การตรวจสอบความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ภายใต้เงื่อนไขแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน และพัฒนาการความสามารถของผู้สอบแตกต่างกัน ปรากฏผลสรุปดังนี้

1. BIAS **ของพัฒนาการความสามารถ**

เมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด ความยากของแบบทดสอบผสมแตกต่างกัน 3 ระดับ และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน 3 ระดับค่าเฉลี่ย BIAS ของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถเป็นผลมาจากการส่งผลร่วมกันระหว่างตัวแปร ความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม ผลการเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถในแต่ละคู่ระดับของ 3 เงื่อนไข ปรากฏผลดังนี้

1.1 เมื่อความยากและการให้คะแนนคงที่ ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยาว 30:10 มีค่าสูงสุด รองลงมาคือ 15:5 และ 24:8 ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าแบบทดสอบผสมที่มี

ความยาว 30:10 มีความลำเอียงสูงกว่าแบบทดสอบผสมที่มีความยาว 24:8 และ แบบทดสอบผสมที่มีความยาว 15:5 มีความลำเอียงสูงกว่าแบบทดสอบผสมที่มีความยาว 24:8 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบผสมที่มีความยาว 30:10 กับ15:5 มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

1.2 เมื่อความยาวและรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่มีความยาก 0.00 มีค่าเฉลี่ย BIAS สูงสุด รองลงมาคือ แบบทดสอบผสมที่มีความยาก -0.50 และ 0.50 ตามลำดับ และพบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบผสมที่มีความยาก 0.50 และ แบบทดสอบผสมที่มีความยาก 0.00 มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบที่มีความยาก 0.50 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่มีความยาก 0.00 แบบทดสอบที่มีความยาก -0.50 มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

1.3 เมื่อแบบทดสอบผสมมีความยาวและความยากคงที่ ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่ให้คะแนน 5 ระดับมีค่าเฉลี่ย BIAS สูงสุด รองลงมาคือ แบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 3 ระดับ ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าแบบทดสอบผสมที่ให้คะแนน 5 ระดับ มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบที่ให้คะแนน 4 ระดับ และ 3 ระดับ และแบบทดสอบที่ให้คะแนน 4 ระดับ มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบที่ให้คะแนน 3 ระดับ

2. RMSE ของพัฒนาการความสามารถ

เมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด ความยากของแบบทดสอบผสมแตกต่างกัน 3 ระดับ และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน 3 ระดับค่าเฉลี่ย RMSE ของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเป็นผลมาจากการส่งผลร่วมกันระหว่างตัวแปร ความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม ผลการเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถ ในแต่ละคู่ระดับของ 3 เงื่อนไข ปรากฏผลดังนี้

2.1 เมื่อความยากและการให้คะแนนคงที่ ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยาว 30:10 มีค่าสูงสุด รองลงมาคือ 15:5 และ 24:8 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าแบบทดสอบผสมที่มีความยาว 30:10 มี RMSE สูงกว่า แบบทดสอบผสมที่มีความยาว 24:8 และ แบบทดสอบผสมที่มีความยาว 15:5 มี RMSE สูงกว่าแบบทดสอบผสมที่มีความยาว 24:8 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบผสมที่มีความยาว 30:10 กับ 15:5 มีค่าเฉลี่ย RMSE แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

2.2 เมื่อความยาวและรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่มีค่าความยาก 0.00 มีค่าเฉลี่ย RMSE สูงสุด รองลงมาคือ แบบทดสอบผสมที่มีความยาก -0.50 และ 0.50 ตามลำดับ และพบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ย RMSE สูงกว่าแบบทดสอบผสมที่มีความยาก 0.50 และแบบทดสอบผสมที่มีความยาก 0.00 มีค่าเฉลี่ย RMSE สูงกว่าแบบทดสอบที่มีความยาก 0.50 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่มีความยาก 0.00 แบบทดสอบที่มีความยาก -0.50 มี RMSE แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

2.3 เมื่อแบบทดสอบผสมมีความยาวและความยากคงที่ ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีการให้คะแนน 4 ระดับมีค่าเฉลี่ย RMSE สูงสุดรองลงมาคือแบบทดสอบผสมที่ให้คะแนน 5 ระดับและ 3 ระดับ ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าแบบทดสอบผสมที่ให้คะแนน 4 ระดับ มีค่าเฉลี่ย RMSE สูงกว่าแบบทดสอบผสมที่ให้คะแนน 5 ระดับ และ 3 ระดับ และแบบทดสอบผสมที่ให้คะแนน 5 ระดับ มีค่าเฉลี่ย RMSE สูงกว่า แบบทดสอบผสมที่ให้คะแนน 3 ระดับ

ความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันสรุปได้ดังตาราง 38

ตาราง 38 สรุปความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ดัชนีความแม่นยำ	**ผลการเปรียบเทียบโดยรวม**	**ผลการเปรียบเทียบระดับตัวแปร**		
	ตัวแปรที่พบความแตกต่างอย่างมีนัยสำคัญที่ระดับ .05	ความยาว	ความยาก	การให้คะแนน
BIAS	ความยาว ความยาก และการให้คะแนน	**ยาว 30:10** > ยาว 15:15 > ยาว 24:8 (ในทิศทางลบ)	**ยาก 0.00** > ยาก -0.50 > ยาก 0.50 (ในทิศทางลบ)	**5 ระดับ** > 4 ระดับ > 3 ระดับ (ในทิศทางลบ)
RMSE	ความยาว ความยาก และการให้คะแนน	**ยาว 30:10** > ยาว 15:15 > ยาว 24:8	**ยาก 0.00** > ยาก -0.50 > ยาก 0.50	**4 ระดับ** > 5 ระดับ > 3 ระดับ

หมายเหตุ:

ยาว 30:10 ยาว 24:8 และ ยาว 15:5 หมายถึงแบบทดสอบผสมที่มีความยาว 30:10 24:8 และ 15:5 ตามลำดับ

ยาก -0.50 ยาก 0.00 ยาก 0.50 หมายถึงแบบทดสอบผสมที่ใช้ในการทดสอบครั้งที่ 1 มีความยากเฉลี่ยเป็น - 0.50 0.00 และ 0.50 ตามลำดับ

3 ระดับ 4 ระดับ และ 5 ระดับ หมายถึง คะแนนของข้อสอบแบบเขียนตอบในแบบทดสอบผสมมี 3 ระดับ 4 ระดับ และ 5 ระดับตามลำดับ

ตอนที่ 4 การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ผู้วิจัยคำนวณผลต่างระหว่างค่าเฉลี่ยของพัฒนาการความสามารถผู้สอบที่ได้จากการทำซ้ำ 50 รอบ กับพัฒนาการความสามารถจริงในแต่ละเงื่อนไขที่ศึกษา ได้แก่ กลุ่มผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน แบบทดสอบผสมมีความยาว ความยาก และ การให้คะแนนแตกต่างกัน การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันศึกษาและเปรียบเทียบจากค่าความลำเอียง (BIAS) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อเทียบกับพัฒนาการความสามารถจริง ผลการตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันนำเสนอเป็นสองส่วน ดังนี้

4.1 การศึกษาค่าความลำเอียง (BIAS) และค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

4.1.1 การศึกษาความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

4.1.1.1 ความยาวคงที่

4.1.1.2 ความยากคงที่

4.1.1.3 การให้คะแนนคงที่

4.1.2 การศึกษาค่ารากที่สองค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

4.1.2.1 ความยาวคงที่

4.1.2.2 ความยากคงที่

4.1.2.3 การให้คะแนนคงที่

4.2 การเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

4.2.1 การเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

4.2.2 การเปรียบเทียบค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

4.1 การศึกษาค่าความลำเอียง (BIAS) และรากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

4.1.1 การศึกษาความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ในตอนนี้เสนอผลการวิเคราะห์ตามเงื่อนไขของแบบทดสอบผสมที่ต้องการศึกษา เมื่อความยากคงที่ ความยาวคงที่ และ การให้คะแนนคงที่ปรากฏผลดังนี้

4.1.1.1 ความยาวคงที่

กรณีแบบทดสอบผสมยาว 30:10

จากตาราง 39 และภาพประกอบ 47 (ก – ค) เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก -0.50 และ 0.50 โดยส่วนใหญ่พบว่ากรณีให้คะแนน 3 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด รองลงมาคือ กรณีให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก 0.00 ค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถในกรณีให้คะแนน 3 ระดับมีค่าต่ำกว่ากรณีให้คะแนน 4 ระดับ และ 5 ระดับเฉพาะกลุ่มผู้สอบที่มีพัฒนาการเพิ่มขึ้น 0.1 (M1 M4 และ M7)เท่านั้น ส่วนกลุ่มพัฒนาการอื่นที่เหลือมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่ให้คะแนน 5 ระดับ

กรณีแบบทดสอบผสมยาว 24:8

จากตาราง 39 และภาพประกอบ 48 (ก – ค) เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 พบว่าทุกกลุ่มพัฒนาการมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่ให้คะแนน 4 ระดับ รองลงมาคือ กรณีให้คะแนน 3 ระดับ และ 5 ระดับ ตามลำดับ เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก 0.00 และ 0.50 เฉพาะกลุ่ม M1 และ M4 เท่านั้นที่พบว่ากรณีที่ให้คะแนน 4 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่ม M2 กรณีที่ให้คะแนน 5 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด และกลุ่ม M5 M6 M7 M8 และM9 กรณีที่ให้คะแนน 3 ระดับมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

กรณีแบบทดสอบผสมยาว 15:5

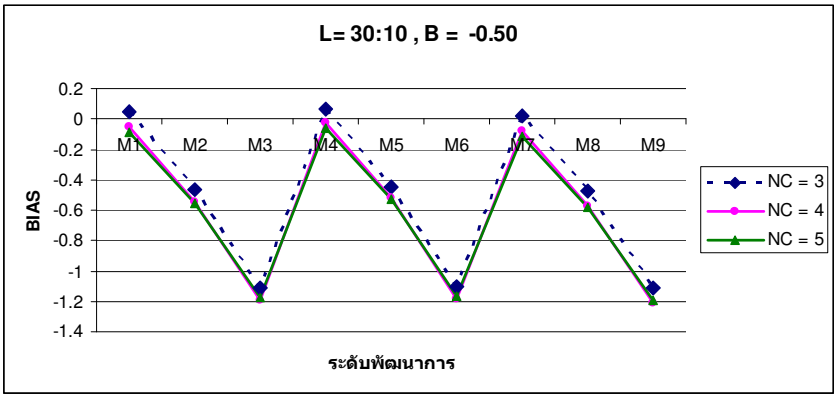
จากตาราง 39 และภาพประกอบ 49 (ก – ค) เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 และ 0.50 พบว่ากลุ่มพัฒนาการส่วนใหญ่มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่ให้คะแนน 3 ระดับ รองลงมาคือ กรณีให้คะแนน 4 ระดับ และ 5 ระดับ ยกเว้นกลุ่ม M2 และ M4 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่ให้คะแนน 4 ระดับ เมื่อแบบทดสอบผสมมีความยาว 24:8 และความยากของแบบทดสอบเป็น 0.00 พบว่ากลุ่ม M1 M4 และ M7 มีค่าความลำเอียง (BIAS) ของพัฒนาการ

ความสามารถต่ำที่สุดในกรณีที่ให้คะแนน 3 ระดับ ส่วนกลุ่ม M2 M5 M8 และM9 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่ให้คะแนน 4 ระดับ และกลุ่มที่เหลือ M3 และ M6 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่ให้คะแนน 5 ระดับ

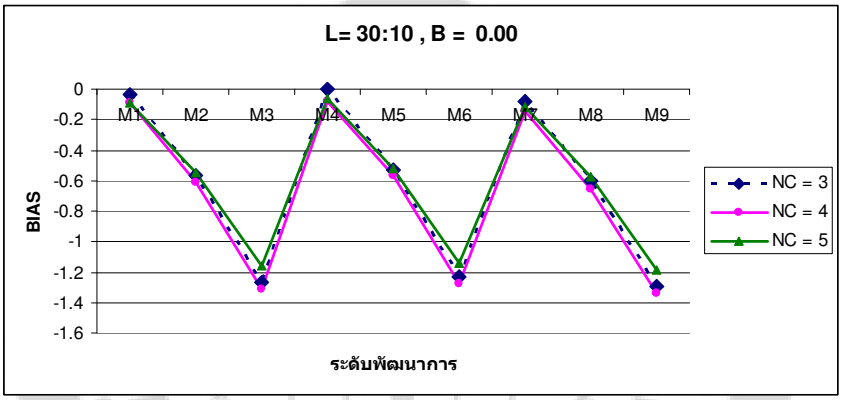
ตาราง 39 ค่าเฉลี่ยความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาว และ ความยากคงที่ จำแนกตามการให้คะแนนและระดับพัฒนาการ

L	B	NC	M1	M2	M3	M4	M5	M6	M7	M8	M9
		3	0.044	-0.461	-1.112	0.066	-0.444	-1.108	0.023	-0.476	-1.113
	-0.50	4	-0.055	-0.548	-1.196	-0.023	-0.522	-1.184	-0.082	-0.572	-1.207
		5	-0.086	-0.558	-1.178	-0.057	-0.532	-1.162	-0.114	-0.583	-1.194
		3	-0.039	-0.566	-1.268	-0.003	-0.530	-1.234	-0.078	-0.601	-1.292
30:10	0.00	4	-0.094	-0.610	-1.311	-0.085	-0.563	-1.275	-0.141	-0.657	-1.342
		5	-0.088	-0.551	-1.162	-0.059	-0.522	-1.141	-0.117	-0.579	-1.183
		3	0.020	-0.456	-1.077	0.043	-0.432	-1.060	-0.004	-0.480	-1.095
	0.50	4	-0.054	-0.523	-1.142	-0.025	-0.493	-1.120	-0.085	-0.480	-1.163
		5	-0.086	-0.537	-1.136	-0.057	-0.508	-1.113	-0.210	-0.567	-1.161
		3	-0.028	-0.570	-1.279	0.006	-0.543	-1.266	-0.062	-0.596	-1.288
	-0.50	4	-0.024	-0.510	-1.142	0.003	-0.487	-1.132	-0.051	-0.530	-1.150
		5	-0.082	-0.547	-1.153	-0.058	-0.523	-1.138	-0.107	-0.568	-1.167
		3	0.050	-0.566	-1.056	0.071	-0.411	-1.042	0.028	-0.450	-1.067
24:8	0.00	4	-0.028	-0.610	-1.119	0.000	-0.475	-1.105	-0.057	-0.528	-1.137
		5	-0.082	-0.551	-1.128	-0.056	-0.508	-1.111	-0.108	-0.562	-1.153
		3	0.037	-0.566	-1.039	0.057	-0.411	-1.024	0.015	-0.452	-1.057
	0.50	4	-0.034	-0.610	-1.100	-0.007	-0.467	-1.080	-0.061	-0.520	-1.120
		5	-0.082	-0.551	-1.112	-0.058	-0.500	-1.089	-0.107	-0.553	-1.136
		3	0.033	-0.570	-1.088	0.050	-0.446	-1.087	0.015	-0.472	-1.093
	-0.50	4	-0.072	-0.510	-1.180	-0.045	-0.528	-1.169	-0.097	-0.574	-1.189
		5	-0.114	-0.547	-1.175	-0.089	-0.546	-1.155	-0.139	-0.597	-1.190
		3	-0.034	-0.545	-1.214	-0.008	-0.521	-1.197	-0.060	-0.568	-1.225
15:5	0.00	4	-0.066	-0.540	-1.159	-0.040	-0.513	-1.142	-0.094	-0.563	-1.173
		5	-0.110	-0.560	-1.154	-0.083	-0.533	-1.135	-0.135	-0.587	-1.177
		3	0.019	-0.545	-1.068	0.040	-0.434	-1.053	-0.003	-0.475	-1.083
	0.50	4	-0.064	-0.540	-1.140	-0.036	-0.500	-1.120	-0.091	-0.555	-1.159
		5	-0.105	-0.560	-1.140	-0.079	-0.522	-1.117	-0.132	-0.579	-1.163

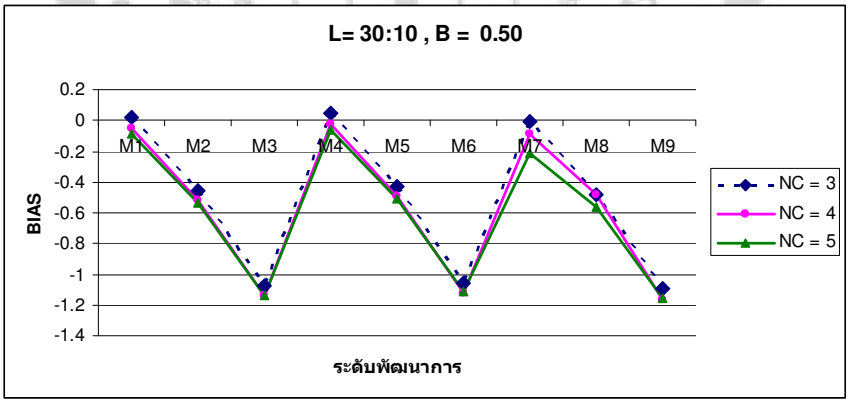
*__ หมายถึง มีค่า BIAS ต่ำสุดเมื่อแบบทดสอบผสมมีความยาว ความยากเท่ากัน



51 ก

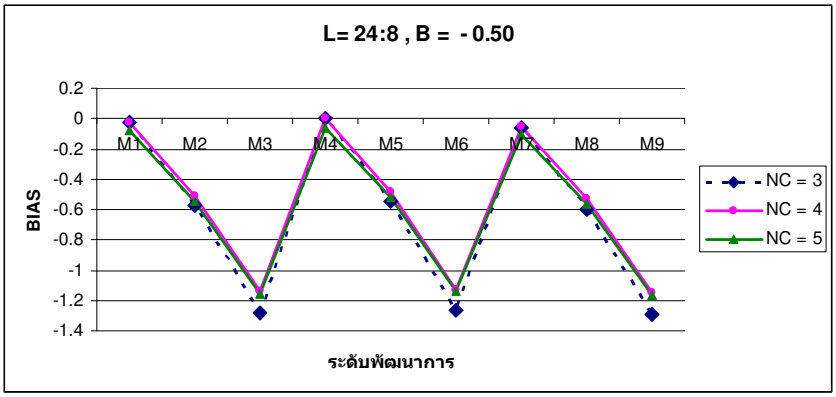


51 ข

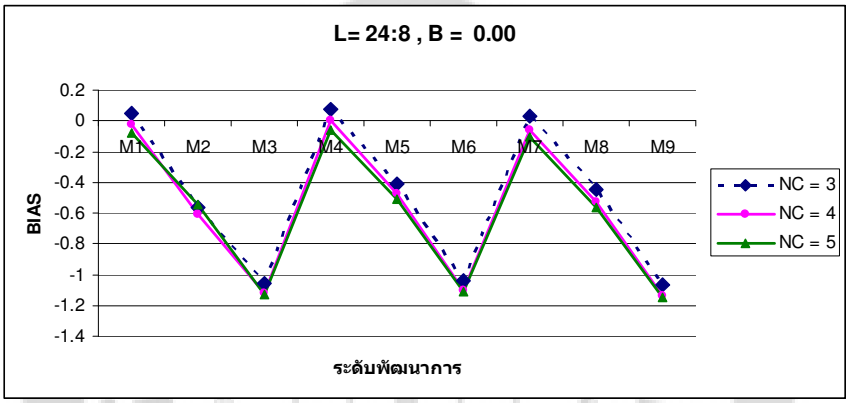


51 ค

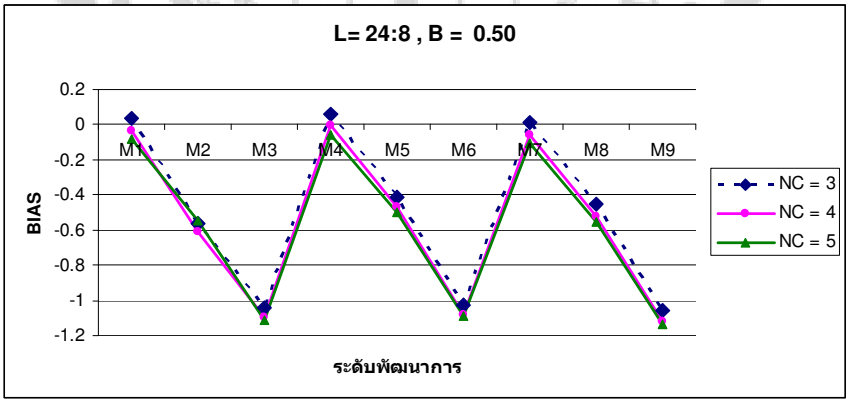
ภาพประกอบ 51 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก -0.50 (51 ก) 0.00 (51 ข) 0.50 (51 ค)



52 ก

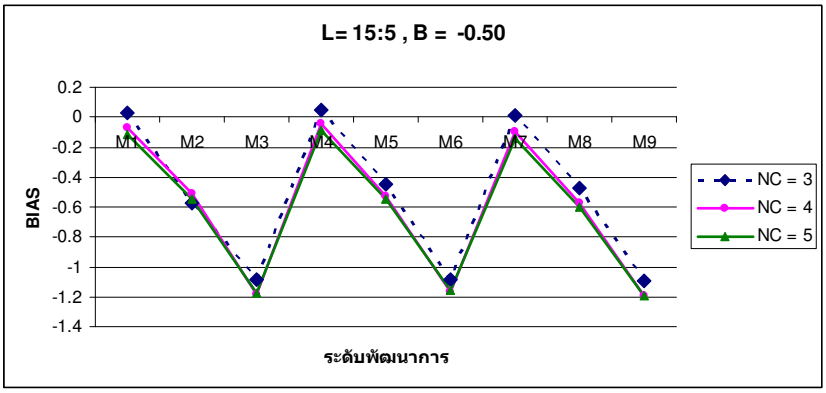


52 ข

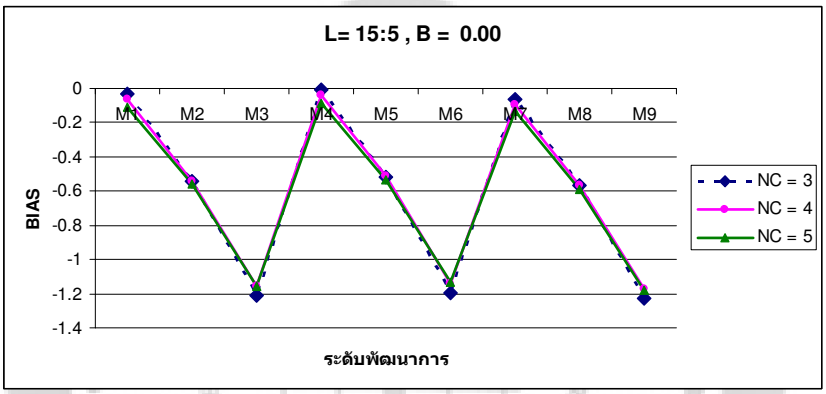


52 ค

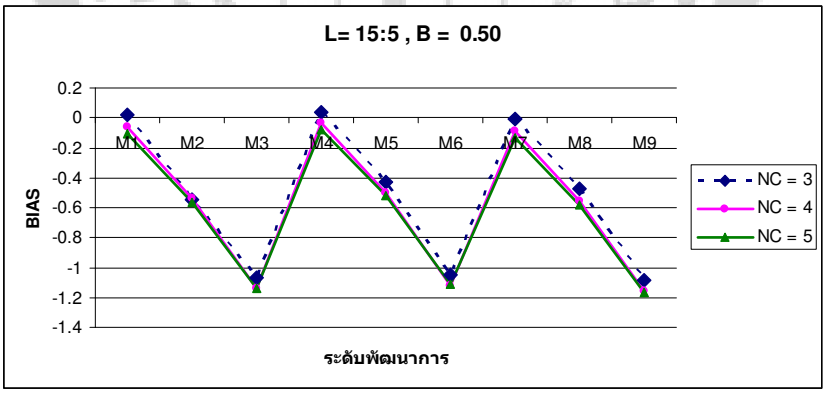
ภาพประกอบ 52 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 (52 ก) 0.00 (52 ข) 0.50 (52 ค)



53 ก



53 ข



53 ค

ภาพประกอบ 53 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก -0.50 (53 ก) 0.00 (53 ข) 0.50 (53 ค)

4.1.1.2 ความยากคงที่

กรณีแบบทดสอบผสมมีความยาก -0.50

จากตาราง 40 และภาพประกอบ 50 (ก – ค) เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 3 ระดับ พบว่ากลุ่ม M1 และ M4 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมยาว 24:8 ส่วนกลุ่ม M2 และ M5 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำสุดในกรณีที่แบบทดสอบผสมยาว 30:10 ส่วนกลุ่มที่เหลือมีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำสุดในกรณีแบบทดสอบผสมยาว 15:5

เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 4 ระดับ พบว่าในทุกกลุ่มพัฒนาการมีค่าความลำเอียง (BIAS) ต่ำที่สุดในกรณีที่แบบทดสอบผสมยาว 24:8 และพบว่า ค่าความลำเอียง (BIAS) ของพัฒนาการของกลุ่ม M2 ในกรณีแบบทดสอบผสมยาว 15:5 และ 24:8 มีค่าเท่ากัน

เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 5 ระดับ พบว่าในทุกกลุ่มพัฒนาการยกเว้นกลุ่ม M4 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมยาว 24:8 โดยพบว่ากลุ่ม M2 ค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถในกรณีแบบทดสอบผสมยาว 15:5 และ 24:8 มีค่าเท่ากัน และกลุ่ม M7 พบว่าค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถในกรณีแบบทดสอบผสมยาว 30:10 และ 24:8 มีค่าเท่ากัน ส่วนกลุ่ม M4 นั้นค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมยาว 30:10

กรณีแบบทดสอบผสมความยาก 0.00

จากตาราง 40 และภาพประกอบ 51 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยากเป็น 0.00 และให้คะแนน 3 ระดับ ในกลุ่มพัฒนาการ M1 และ M2 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำสุดในกรณีแบบทดสอบผสมยาว 15:5 ในกลุ่ม M4 แบบทดสอบผสมขนาด 30:10 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำสุด ส่วนกลุ่มที่เหลือแบบทดสอบผสมขนาด 24:8 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำสุด

เมื่อแบบทดสอบผสมมีความยากเป็น 0.00 และให้คะแนน 4 ระดับ พบว่าในทุกกลุ่มพัฒนาการ ยกเว้น M2 แบบทดสอบผสมขนาด 24:8 มีค่าความลำเอียง (BIAS) ต่ำที่สุด สำหรับกลุ่ม M2 นั้นแบบทดสอบผสมขนาด 15:5 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

เมื่อแบบทดสอบผสมมีความยากเป็น 0.00 และให้คะแนน 5 ระดับ พบว่าในทุกกลุ่มพัฒนาการ แบบทดสอบผสมขนาด 24:8 มีค่าความลำเอียง (BIAS) ต่ำที่สุด และพบว่ากลุ่ม M2 นั้นแบบทดสอบผสมขนาด 30:10 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถเท่ากับแบบทดสอบผสมขนาด 24:8

กรณีแบบทดสอบผสมความยาก 0.50

จากตาราง 40 และภาพประกอบ 52 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 3 ระดับ ในกลุ่ม M1 M4 และ M7 แบบทดสอบผสมขนาด 15:5 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด กลุ่ม M2 แบบทดสอบผสมขนาด 30:10 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่มที่เหลือ แบบทดสอบผสมขนาด 24:8 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 4 ระดับ ในกลุ่ม M2 และ M8 แบบทดสอบผสมขนาด 30:10 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่มที่เหลือ แบบทดสอบผสมขนาด 24:8 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 5 ระดับ ในกลุ่ม M2 และ M4 แบบทดสอบผสมขนาด 30:10 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด ส่วนกลุ่มที่เหลือ แบบทดสอบผสมขนาด 24:8 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุด

4.1.1.3 การให้คะแนนคงที่

กรณีให้คะแนน 3 ระดับ

จากตาราง 41 และภาพประกอบ 57 (ก – ค) พบว่าเมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมมีความยาว 30:10 ในกลุ่ม M4 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.00 กลุ่ม M8 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50 มีค่าความลำเอียง (BIAS) ส่วนกลุ่มที่เหลือส่วนใหญ่มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50

เมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมมีความยาว 24:8 ในกลุ่ม M1 และ M4 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50 กลุ่ม M2 และ M5 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.00 และ 0.50 ส่วนกลุ่มที่เหลือส่วนใหญ่มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50

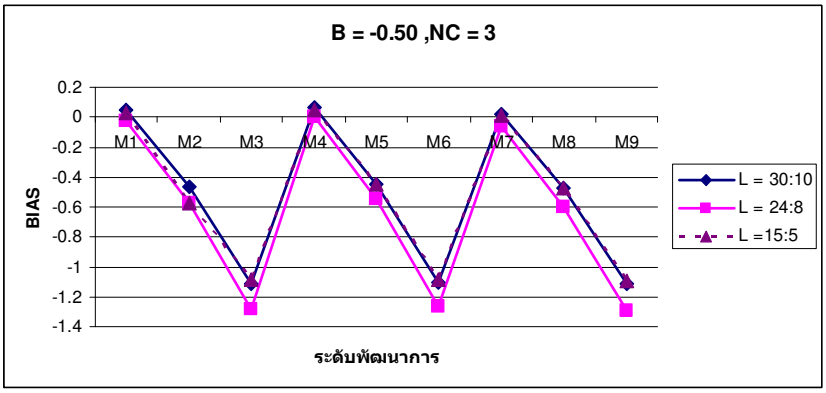
เมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมมีความยาว 15:5 ในกลุ่ม M4 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.00 กลุ่ม M8 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบ

ผสมมีความยาก -0.50 ส่วนกลุ่มที่เหลือส่วนใหญ่มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50

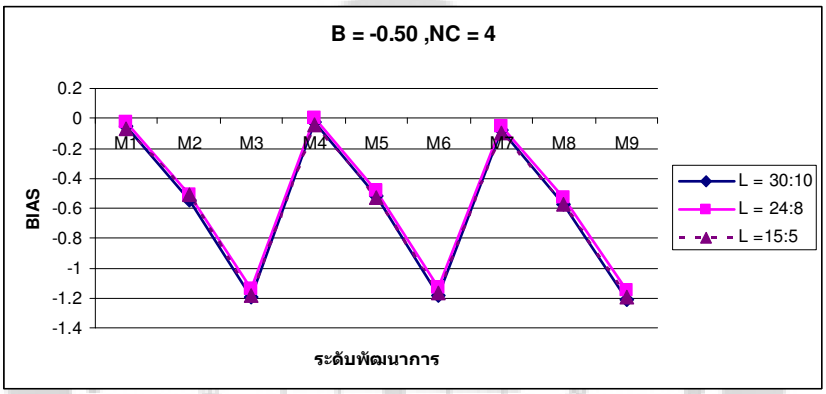
ตาราง 40 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถรายบุคคล เมื่อแบบทดสอบผสมมีความยาก และการให้คะแนนคงที่ จำแนกตามความยาวและระดับพัฒนาการ

B	NC	L	M1	M2	M3	M4	M5	M6	M7	M8	M9
		30:10	0.044	-0.461	-1.112	0.066	-0.444	-1.108	0.023	-0.476	-1.113
	3	24:8	-0.028	-0.570	-1.279	0.006	-0.543	-1.266	-0.060	-0.596	-1.288
		15:5	0.033	-0.570	-1.088	0.050	-0.446	-1.087	0.015	-0.472	-1.093
		30:10	-0.055	-0.548	-1.196	-0.023	-0.522	-1.184	-0.080	-0.572	-1.207
-0.50	4	24:8	-0.024	-0.510	-1.142	0.003	-0.487	-1.132	-0.050	-0.530	-1.150
		15:5	-0.072	-0.510	-1.180	-0.045	-0.528	-1.169	-0.100	-0.574	-1.189
		30:10	-0.086	-0.558	-1.178	-0.057	-0.532	-1.162	-0.110	-0.583	-1.194
	5	24:8	-0.082	-0.547	-1.153	-0.058	-0.523	-1.138	-0.110	-0.568	-1.167
		15:5	-0.114	-0.547	-1.175	-0.089	-0.546	-1.155	-0.140	-0.597	-1.190
		30:10	-0.039	-0.566	-1.268	-0.003	-0.530	-1.234	-0.080	-0.601	-1.292
	3	24:8	0.050	-0.566	-1.056	0.071	-0.411	-1.042	0.028	-0.450	-1.067
		15:5	-0.034	-0.545	-1.214	-0.008	-0.521	-1.197	-0.060	-0.568	-1.225
		30:10	-0.094	-0.610	-1.311	-0.085	-0.563	-1.275	-0.140	-0.657	-1.342
0.00	4	24:8	-0.028	-0.610	-1.119	0.000	-0.475	-1.105	-0.060	-0.528	-1.137
		15:5	-0.066	-0.540	-1.159	-0.040	-0.513	-1.142	-0.090	-0.563	-1.173
		30:10	-0.088	-0.551	-1.162	-0.059	-0.522	-1.141	-0.120	-0.579	-1.183
	5	24:8	-0.082	-0.551	-1.128	-0.056	-0.508	-1.111	-0.110	-0.562	-1.153
		15:5	-0.110	-0.560	-1.154	-0.083	-0.533	-1.135	-0.140	-0.587	-1.177
		30:10	0.020	-0.456	-1.077	0.043	-0.432	-1.060	0.000	-0.480	-1.095
	3	24:8	0.037	-0.566	-1.039	0.057	-0.411	-1.024	0.015	-0.452	-1.057
		15:5	0.019	-0.545	-1.068	0.040	-0.434	-1.053	0.000	-0.475	-1.083
		30:10	-0.054	-0.523	-1.142	-0.025	-0.493	-1.120	-0.080	-0.480	-1.163
0.50	4	24:8	-0.034	-0.610	-1.100	-0.007	-0.467	-1.080	-0.060	-0.520	-1.120
		15:5	-0.064	-0.540	-1.140	-0.036	-0.500	-1.120	-0.090	-0.555	-1.159
		30:10	-0.086	-0.537	-1.136	-0.057	-0.508	-1.113	-0.210	-0.567	-1.161
	5	24:8	-0.082	-0.551	-1.112	-0.058	-0.500	-1.089	-0.110	-0.553	-1.136
		15:5	-0.105	-0.560	-1.140	-0.079	-0.522	-1.117	-0.130	-0.579	-1.163

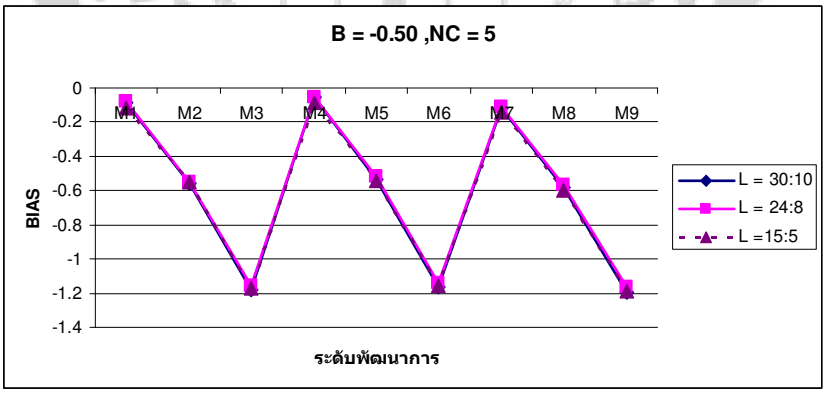
*__หมายถึง มีค่า BIAS ต่ำที่สุด ภายใต้เงื่อนไขความยาก และ การให้คะแนนเท่ากัน



54 ก

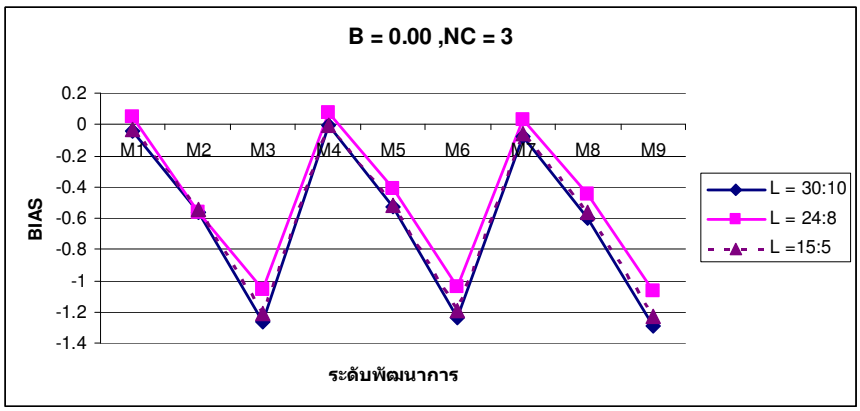


54 ข

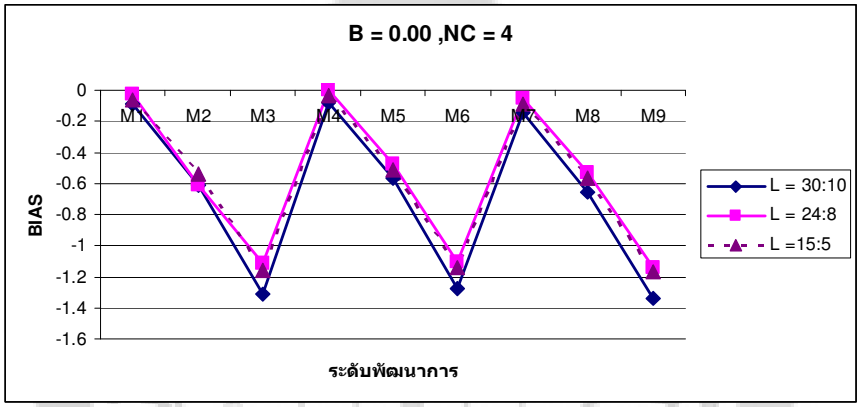


54 ค

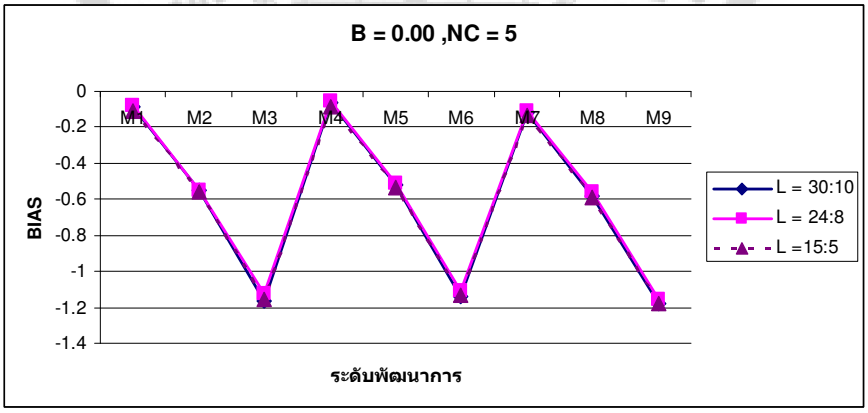
ภาพประกอบ 54 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 3 ระดับ (54 ก) 4 ระดับ (54 ข) 5 ระดับ (50 ค)



55 ก

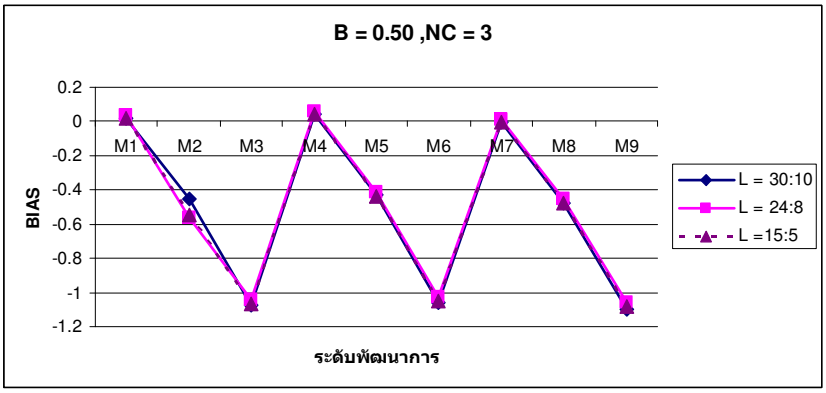


55 ข

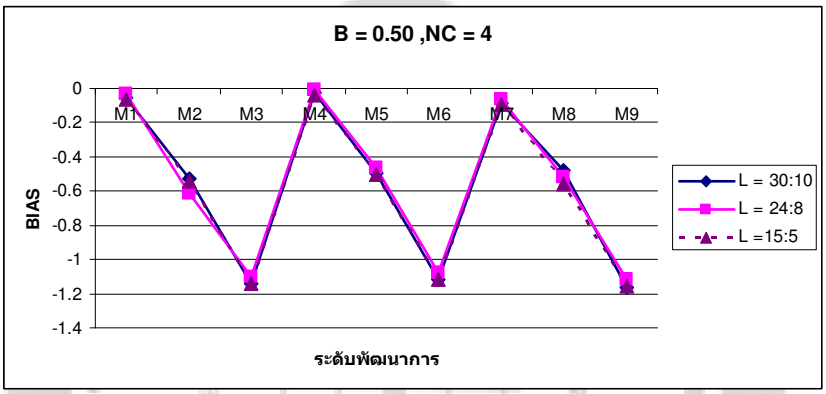


55 ค

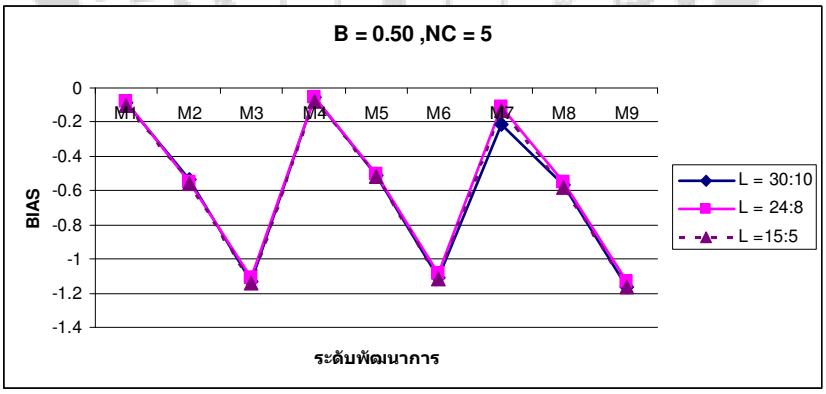
ภาพประกอบ 55 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.00 และให้คะแนน 3 ระดับ (55 ก) 4 ระดับ (55 ข) 5 ระดับ (55 ค)



56 ก



56 ข



56 ค

ภาพประกอบ 56 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 3 ระดับ (56 ก) 4 ระดับ (56 ข) 5 ระดับ (56 ค)

ตาราง 41 ค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ เมื่อแบบทดสอบผสมมีการให้คะแนน และความยาวคงที่ จำแนกตามระดับความยาก และระดับพัฒนาการ

NC	L	B	M1	M2	M3	M4	M5	M6	M7	M8	M9
		-0.50	0.044	-0.461	-1.112	0.066	-0.444	-1.108	0.023	-0.476	-1.113
	30:10	0.00	-0.039	-0.566	-1.268	-0.003	-0.530	-1.234	-0.078	-0.601	-1.292
		0.50	0.020	-0.456	-1.077	0.043	-0.432	-1.060	-0.004	-0.480	-1.095
		-0.50	-0.028	-0.570	-1.279	0.006	-0.543	-1.266	-0.062	-0.596	-1.288
3	24:8	0.00	0.050	-0.566	-1.056	0.071	-0.411	-1.042	0.028	-0.450	-1.067
		0.50	0.037	-0.566	-1.039	0.057	-0.411	-1.024	0.015	-0.452	-1.057
		-0.50	0.033	-0.570	-1.088	0.050	-0.446	-1.087	0.015	-0.472	-1.093
	15:5	0.00	-0.034	-0.545	-1.214	-0.008	-0.521	-1.197	-0.060	-0.568	-1.225
		0.50	0.019	-0.545	-1.068	0.040	-0.434	-1.053	-0.003	-0.475	-1.083
		-0.50	-0.055	-0.548	-1.196	-0.023	-0.522	-1.184	-0.082	-0.572	-1.207
	30:10	0.00	-0.094	-0.610	-1.311	-0.085	-0.563	-1.275	-0.141	-0.657	-1.342
		0.50	-0.054	-0.523	-1.142	-0.025	-0.493	-1.120	-0.085	-0.480	-1.163
		-0.50	-0.024	-0.510	-1.142	0.003	-0.487	-1.132	-0.051	-0.530	-1.150
4	24:8	0.00	-0.028	-0.610	-1.119	0.000	-0.475	-1.105	-0.057	-0.528	-1.137
		0.50	-0.034	-0.610	-1.100	-0.007	-0.467	-1.080	-0.061	-0.520	-1.120
		-0.50	-0.072	-0.510	-1.180	-0.045	-0.528	-1.169	-0.097	-0.574	-1.189
	15:5	0.00	-0.066	-0.540	-1.159	-0.040	-0.513	-1.142	-0.094	-0.563	-1.173
		0.50	-0.064	-0.540	-1.140	-0.036	-0.500	-1.120	-0.091	-0.555	-1.159
		-0.50	-0.086	-0.558	-1.178	-0.057	-0.532	-1.162	-0.114	-0.583	-1.194
	30:10	0.00	-0.088	-0.551	-1.162	-0.059	-0.522	-1.141	-0.117	-0.579	-1.183
		0.50	-0.086	-0.537	-1.136	-0.057	-0.508	-1.113	-0.210	-0.567	-1.161
		-0.50	-0.082	-0.547	-1.153	-0.058	-0.523	-1.138	-0.107	-0.568	-1.167
5	24:8	0.00	-0.082	-0.551	-1.128	-0.056	-0.508	-1.111	-0.108	-0.562	-1.153
		0.50	-0.082	-0.551	-1.112	-0.058	-0.500	-1.089	-0.107	-0.553	-1.136
		-0.50	-0.114	-0.547	-1.175	-0.089	-0.546	-1.155	-0.139	-0.597	-1.190
	15:5	0.00	-0.110	-0.560	-1.154	-0.083	-0.533	-1.135	-0.135	-0.587	-1.177
		0.50	-0.105	-0.560	-1.140	-0.079	-0.522	-1.117	-0.132	-0.579	-1.163

*__ หมายถึง มีค่า BIAS ต่ำที่สุด เมื่อภายใต้เงื่อนไขความยาว และ ระดับการให้คะแนนเท่ากัน

กรณีให้คะแนน 4 ระดับ

จากตาราง 41 และภาพประกอบ 58 (ก – ค) พบว่าเมื่อให้คะแนน 4 ระดับ และแบบทดสอบผสมมีความยาว 30:10 พบว่าในกลุ่ม M4 และ กลุ่ม M7 มีค่าความลำเอียง (BIAS) ของ

พัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50 ส่วนกลุ่มที่เหลือความลำเอียง (BIAS) ของพัฒนาการความสามารถมีค่าต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50

เมื่อให้คะแนน 4 ระดับ และแบบทดสอบผสมมีความยาว 24:8 พบว่าในกลุ่ม M1 M2 และ M7 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50 กลุ่ม M4 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.00 ส่วนกลุ่มที่เหลือพบว่ามีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50

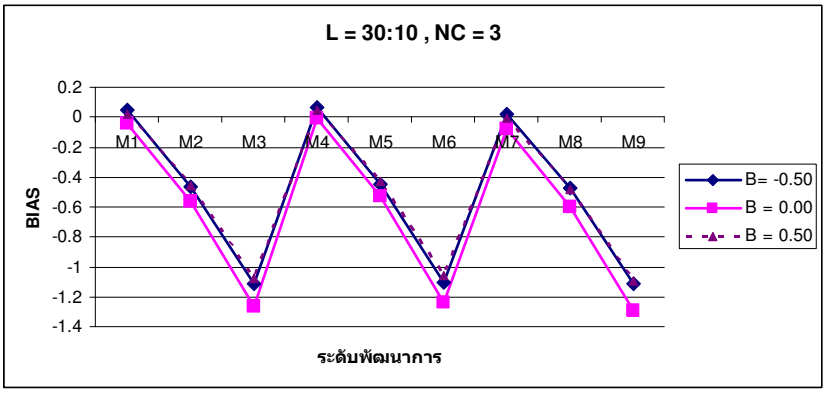
เมื่อให้คะแนน 4 ระดับ และแบบทดสอบผสมมีความยาว 15:5 พบว่าในกลุ่ม M2 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50 ส่วนกลุ่มที่เหลือส่วนใหญ่มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50

กรณีให้คะแนน 5 ระดับ

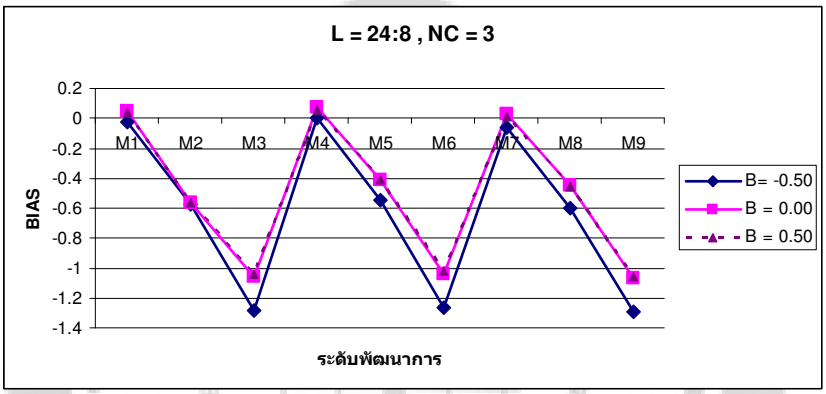
จากตาราง 41 และภาพประกอบ 59 (ก – ค) เมื่อให้คะแนน 5 ระดับ และแบบทดสอบผสมมีความยาว 30:10 พบว่าในกลุ่ม M1 M4 และ M7 มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50 ส่วนกลุ่มที่เหลือส่วนใหญ่พบว่ามีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50

เมื่อให้คะแนน 5 ระดับ และแบบทดสอบผสมมีความยาว 24:8 ในกลุ่ม M1ค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถเท่ากันทุกระดับความยาก กลุ่ม M2 และ M7 พบว่ามีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50 กลุ่ม M4 พบว่ามีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.00 ส่วนกลุ่มที่เหลือส่วนใหญ่มีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50

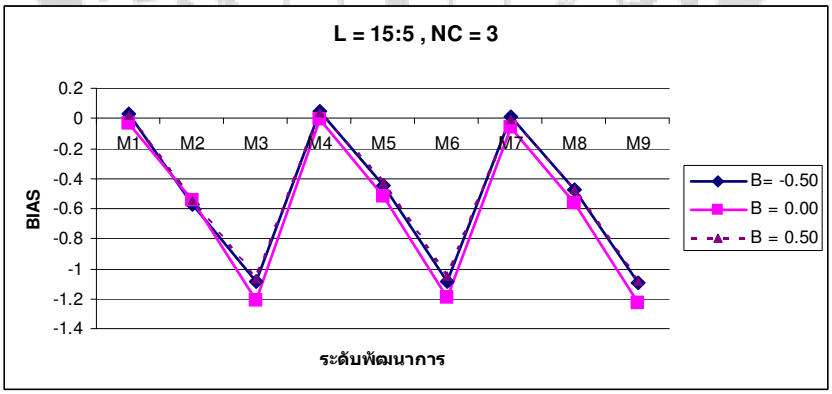
เมื่อให้คะแนน 5 ระดับ และแบบทดสอบผสมมีความยาว 15:5 ในกลุ่ม M2 พบว่ามีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก -0.50 ส่วนกลุ่มที่เหลือส่วนใหญ่พบว่ามีค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถต่ำที่สุดในกรณีที่แบบทดสอบผสมมีความยาก 0.50



57 ก

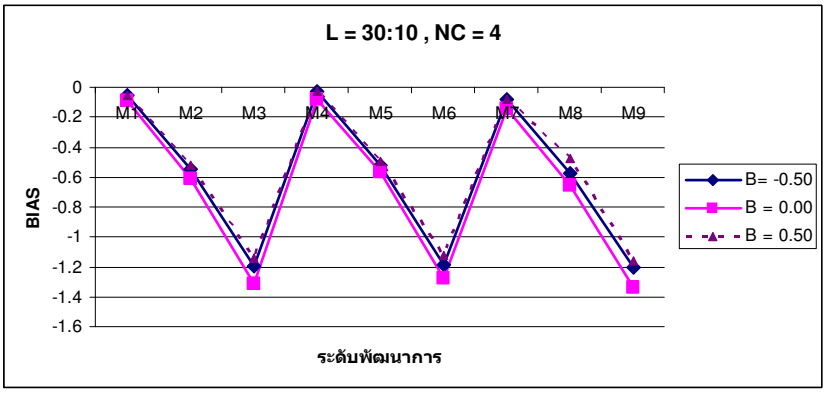


57 ข

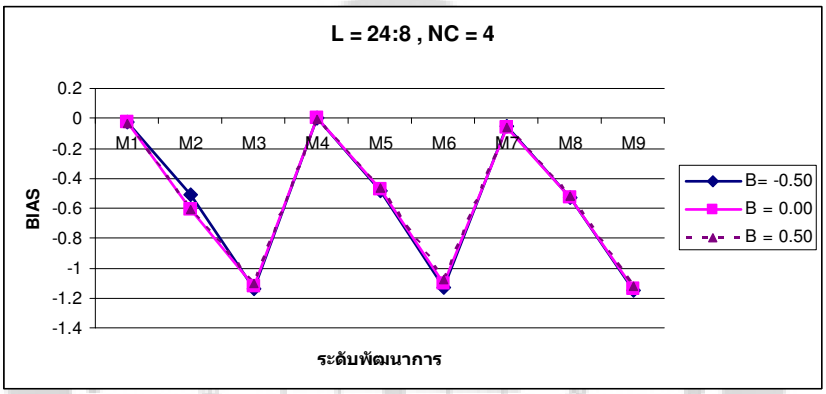


57 ค

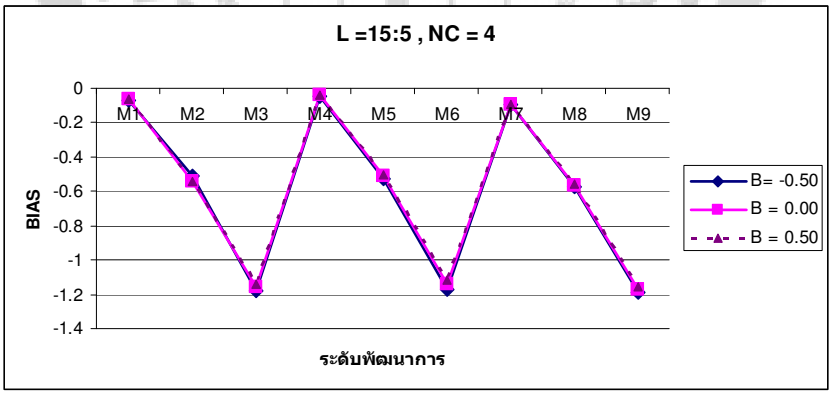
ภาพประกอบ 57 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมมีความยาว 30:10 (57 ก) 24:8 (57 ข) 15:5 (57 ค)



58 ก

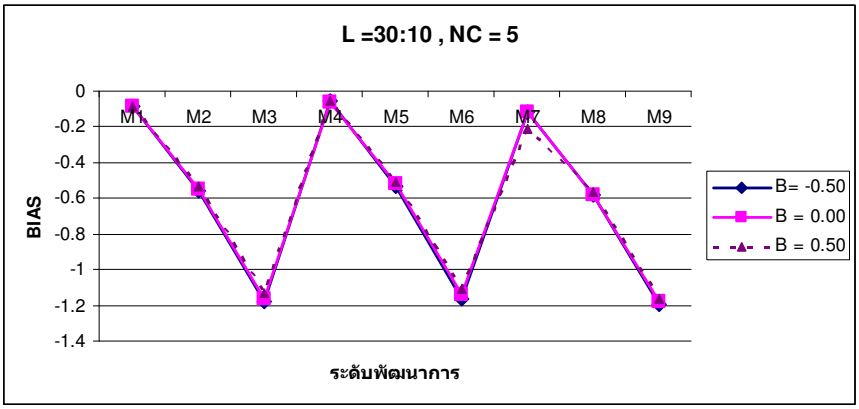


58 ข

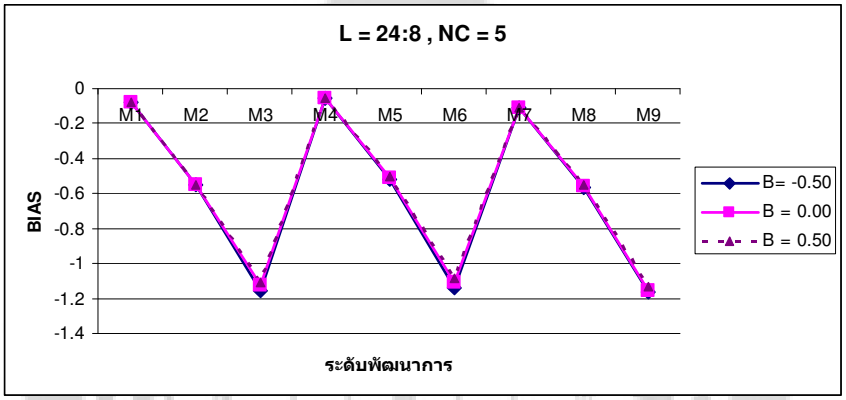


58 ค

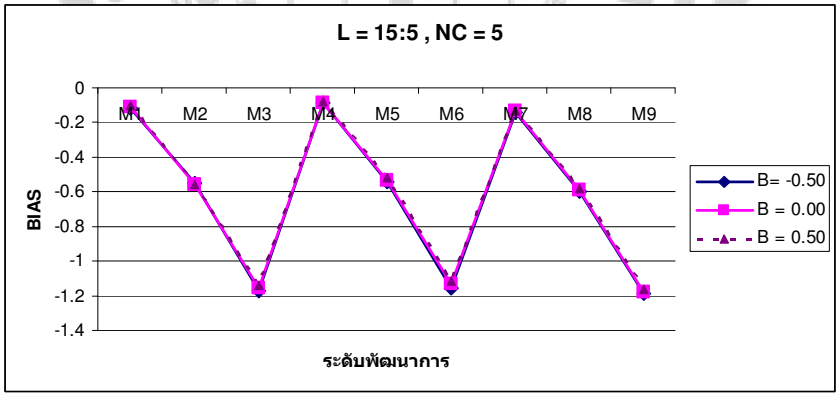
ภาพประกอบ 58 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 4 ระดับ และแบบทดสอบผสมมีความยาว 30:10 (58 ก) 24:8 (58 ข) 15:5 (58 ค)



59 ก



59 ข



59 ค

ภาพประกอบ 59 (ก – ค) แสดงค่าความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อให้คะแนน 5 ระดับ และแบบทดสอบผสมมีความยาวเป็น 30:10 (59 ก) 24:8 (59 ข) 15:5 (59 ค)

4.1.2 การศึกษาค่ารากที่สองค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

4.1.2.1 ความยาวคงที่

กรณีแบบทดสอบผสมมีความยาว 30:10

จากตาราง 42 และภาพประกอบ 60 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก -0.50 ในกลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำสุดในกรณีให้คะแนน 3 ระดับ ยกเว้นกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 4 ระดับ

เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก 0.00 กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถค่าต่ำสุดในกรณีให้คะแนน 5 ระดับ ยกเว้นกลุ่ม M1 M4 และ M7 ที่ค่า RMSE ของพัฒนาการความสามารถมีค่าต่ำที่สุดในกรณีให้คะแนน 3 ระดับ

เมื่อแบบทดสอบผสมมีความยาว 30:10 และมีความยาก 0.50 กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 3 ระดับ ยกเว้นกลุ่ม M4 ที่ค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 4 ระดับ

กรณีแบบทดสอบผสมมีความยาว 24:8

จากตาราง 42 และภาพประกอบ 61 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 24:8 และมีความยาก -0.50 ทุกกลุ่มพัฒนาการ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 4 ระดับ

เมื่อแบบทดสอบผสมที่มีความยาว 24:8 และมีความยาก 0.00 และ 0.50 กลุ่มส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 3 ระดับ ยกเว้นกลุ่ม M1 และ M4 ที่ค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 4 ระดับ

กรณีแบบทดสอบผสมมีความยาว 15:5

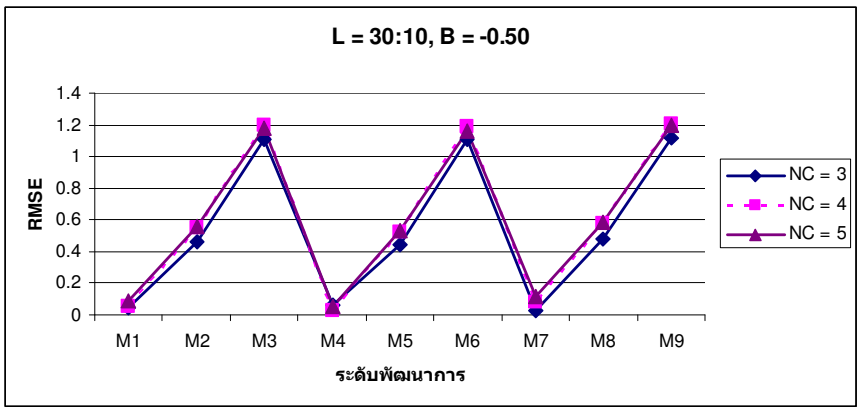
จากตาราง 42 และภาพประกอบ 61 (ก – ค) พบว่าเมื่อแบบทดสอบผสมมีความยาว 15:5 และมีความยาก -0.50 และ 0.50 ทุกกลุ่มพัฒนาการมีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 3 ระดับ ยกเว้นกลุ่ม M4 ค่า RMSE ของพัฒนาการความสามารถมีค่าต่ำสุดในกรณีให้คะแนน 4 ระดับ

เมื่อแบบทดสอบผสมมีความยาว15:5 และมีความยาก 0.00 กลุ่ม M1 M4 และ M7 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 3 ระดับ กลุ่ม M2 M5 M8 และ M9 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 4 ระดับ และกลุ่มที่เหลือมีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีให้คะแนน 5 ระดับ

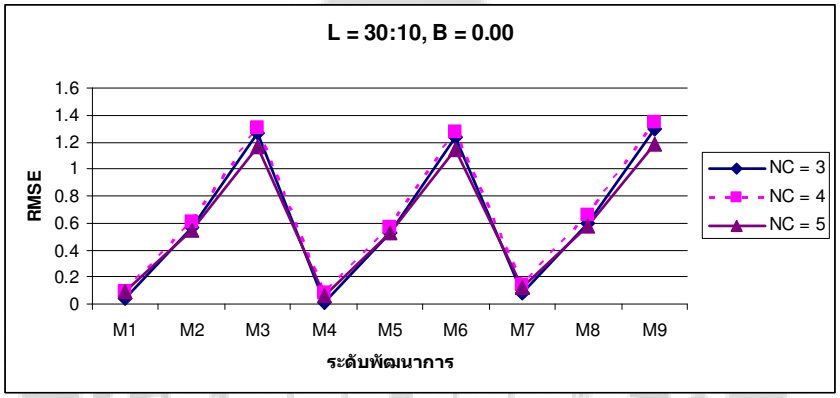
ตาราง 42 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมี ความยาว ความยาก คงที่ จำแนกตามระดับการให้คะแนนและระดับพัฒนาการ

L	B	NC	M1	M2	M3	M4	M5	M6	M7	M8	M9
		3	0.044	0.461	1.112	0.066	0.444	1.108	0.023	0.476	1.113
	-0.50	4	0.055	0.548	1.196	0.024	0.522	1.184	0.083	0.572	1.207
		5	0.086	0.558	1.178	0.058	0.532	1.162	0.114	0.583	1.194
		3	0.039	0.566	1.268	0.006	0.531	1.235	0.078	0.601	1.292
30:10	0.00	4	0.094	0.610	1.311	0.085	0.563	1.275	0.141	0.657	1.342
		5	0.088	0.551	1.162	0.059	0.522	1.141	0.117	0.579	1.183
		3	0.020	0.456	1.077	0.043	0.433	1.060	0.006	0.480	1.095
	0.50	4	0.055	0.523	1.142	0.025	0.493	1.120	0.085	0.480	1.164
		5	0.086	0.537	1.136	0.057	0.508	1.113	0.210	0.567	1.161
		3	0.029	0.570	1.279	0.008	0.543	1.266	0.062	0.596	1.288
	-0.50	4	0.025	0.510	1.142	0.006	0.487	1.132	0.052	0.530	1.150
		5	0.082	0.547	1.153	0.058	0.523	1.138	0.107	0.568	1.167
		3	0.050	0.432	1.056	0.071	0.411	1.042	0.028	0.450	1.067
24:8	0.00	4	0.029	0.502	1.119	0.005	0.475	1.105	0.057	0.528	1.137
		5	0.082	0.535	1.128	0.057	0.508	1.111	0.108	0.562	1.153
		3	0.037	0.430	1.039	0.057	0.411	1.024	0.016	0.452	1.057
	0.50	4	0.034	0.493	1.100	0.009	0.467	1.080	0.061	0.520	1.120
		5	0.082	0.526	1.112	0.058	0.500	1.089	0.107	0.553	1.136
		3	0.033	0.460	1.089	0.050	0.446	1.087	0.016	0.472	1.093
	-0.50	4	0.072	0.553	1.180	0.046	0.528	1.169	0.097	0.574	1.189
		5	0.114	0.573	1.175	0.090	0.547	1.155	0.139	0.597	1.190
		3	0.035	0.545	1.214	0.010	0.521	1.197	0.061	0.568	1.225
15:5	0.00	4	0.066	0.540	1.159	0.040	0.513	1.142	0.094	0.563	1.173
		5	0.110	0.560	1.154	0.084	0.533	1.135	0.135	0.587	1.177
		3	0.019	0.453	1.068	0.040	0.434	1.053	0.006	0.475	1.083
	0.50	4	0.064	0.528	1.140	0.036	0.500	1.120	0.091	0.555	1.159
		5	0.105	0.551	1.140	0.079	0.522	1.117	0.132	0.579	1.163

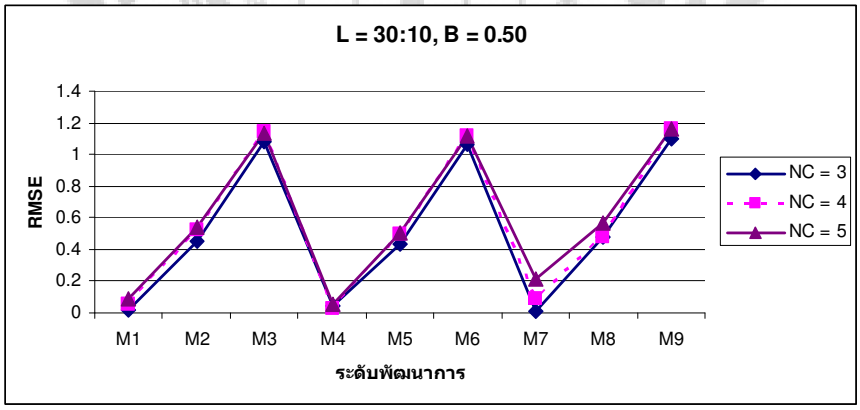
*__หมายถึง มีค่า RMSE ต่ำที่สุด ภายใต้เงื่อนไขระดับความยาว และ ความยากเท่ากัน



60 ก

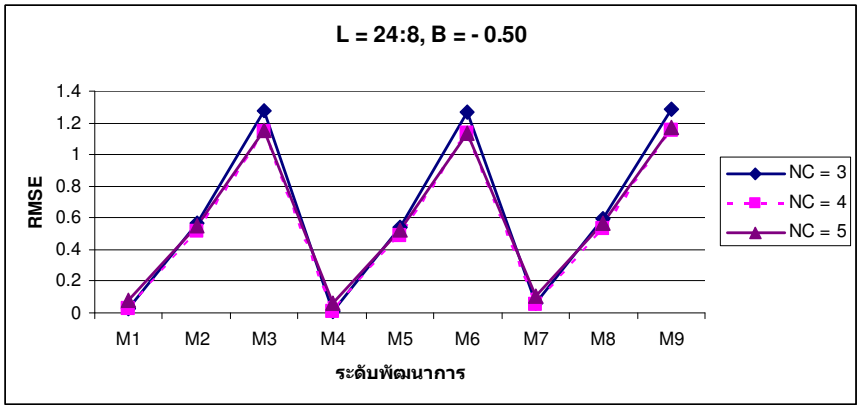


60 ข

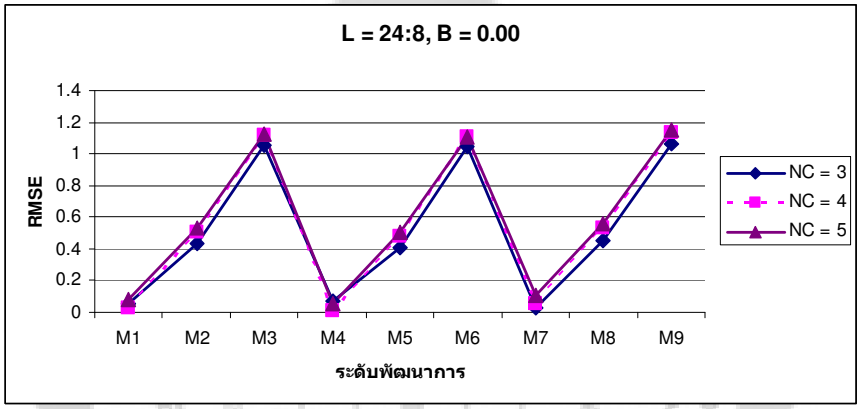


60 ค

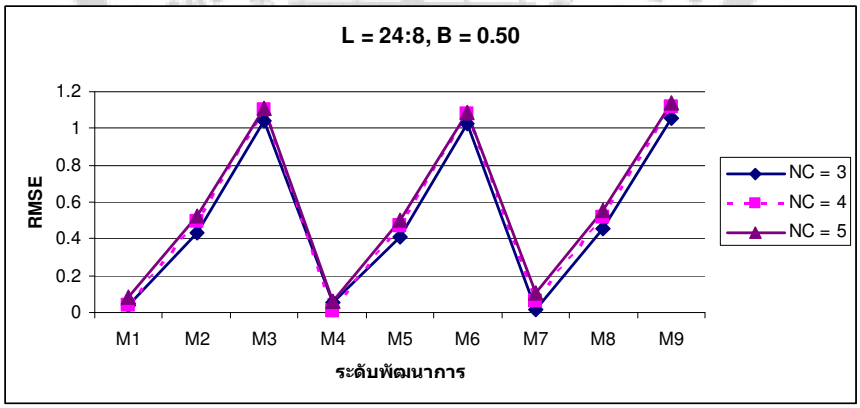
ภาพประกอบ 60 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยาวเป็น 30:10 และความยากของแบบทดสอบเป็น -0.50 (60 ก) 0.00 (60 ข) 0.50 (60 ค)



61 ก

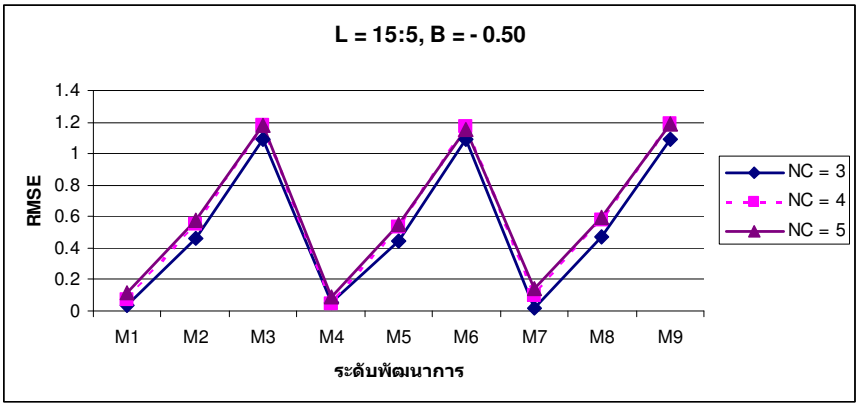


61 ข

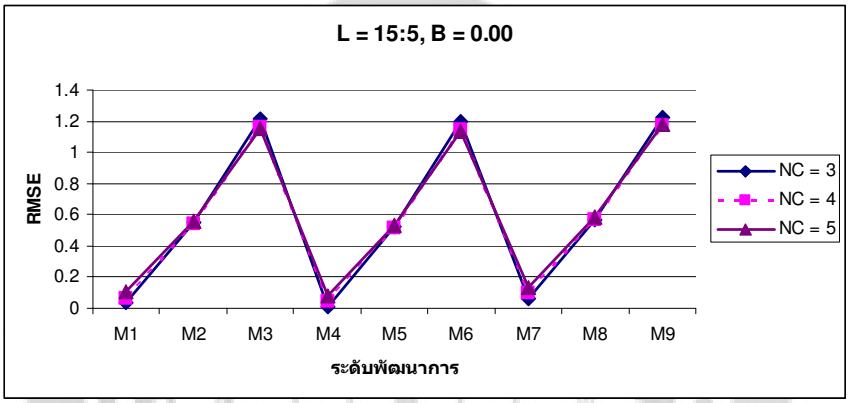


61 ค

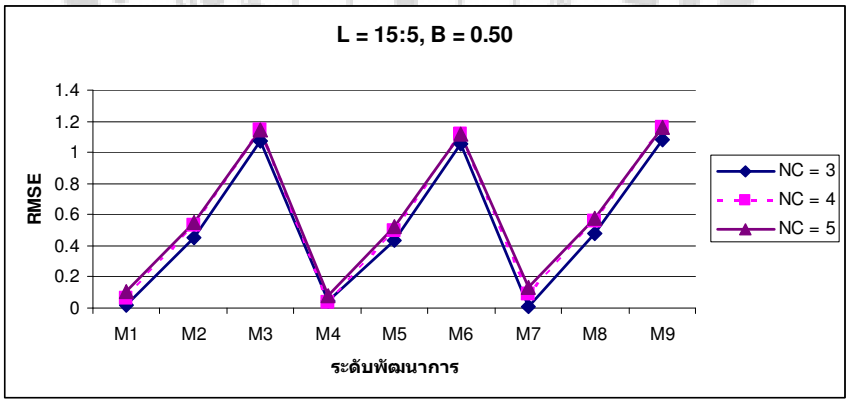
ภาพประกอบ 61 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยาวเป็น 24:8 และความยากของแบบทดสอบเป็น -0.50 (61 ก) 0.00 (61 ข) 0.50 (61 ค)



62 ก



62 ข



62 ค

ภาพประกอบ 62 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อความยาวเป็น 15:5 และความยากของแบบทดสอบเป็น -0.50 (62 ก) 0.00 (62 ข) 0.50 (62ค)

4.1.2.2 ความยากคงที่

กรณีความยากแบบทดสอบผสม เป็น -0.50

จากตาราง 43 และภาพประกอบ 63 (ก – ค) พบว่าเมื่อความยากแบบทดสอบผสมเป็น -0.50 และให้คะแนน 3 ระดับ กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 15:5 ยกเว้นกลุ่ม M1 และ M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 24:8 มีค่าต่ำที่สุด และกลุ่ม M5 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 30:10

เมื่อความยากแบบทดสอบผสม เป็น -0.50 และให้คะแนน 4 หรือ 5 ระดับ ทุกกลุ่มพัฒนาการ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 24:8

กรณีความยากแบบทดสอบผสม เป็น 0.00

จากตาราง 43 และภาพประกอบ 64 (ก – ค) พบว่าเมื่อความยากแบบทดสอบผสมเป็น 0.00 และให้คะแนน 3 ระดับ กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 24:8 ยกเว้นกลุ่ม M1 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 15:5 และกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 30:10

เมื่อความยากแบบทดสอบผสม เป็น 0.00 และให้คะแนน 4 หรือ 5 ระดับ ทุกกลุ่มพัฒนาการ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 24:8

กรณีความยากแบบทดสอบผสม เป็น 0.50

จากตาราง 43 และภาพประกอบ 65 (ก – ค) พบว่าเมื่อความยากแบบทดสอบผสมเป็น 0.50 และให้คะแนน 3 ระดับ กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 24:8 ยกเว้นกลุ่ม M1 และ M7 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 30:10 ส่วนกลุ่ม M4 และ M7 มี RMSE ของพัฒนาการความสามารถของแบบทดสอบผสมขนาด 15:5

เมื่อความยากแบบทดสอบผสมเป็น 0.50 และให้คะแนน 4 ระดับ กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 24:8 ยกเว้นกลุ่ม M8 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 30:10

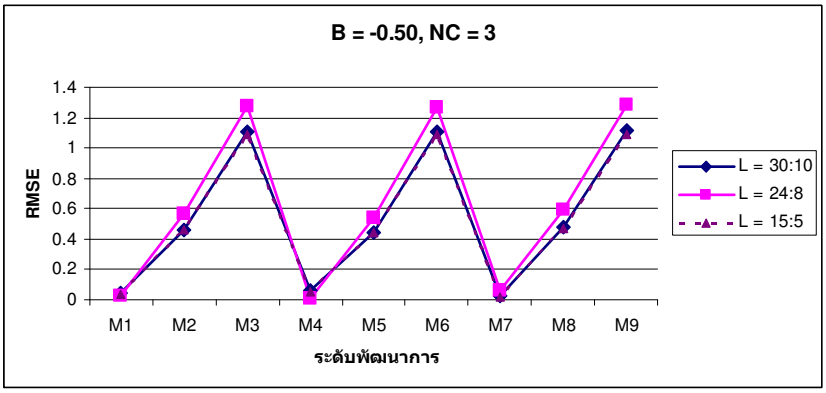
เมื่อความยากแบบทดสอบผสม เป็น 0.50 และให้คะแนน 5 ระดับ กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 24:8

มีค่าต่ำที่สุด ยกเว้นกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาว 30:10

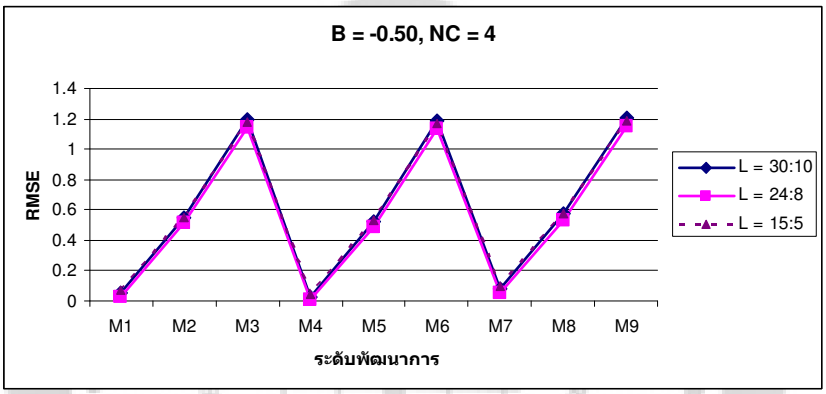
ตาราง 43 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมี ความยาก และการให้คะแนนคงที่ จำแนกตามความยาวและระดับพัฒนาการ

B	NC	L	M1	M2	M3	M4	M5	M6	M7	M8	M9
		30:10	0.044	0.461	1.112	0.066	0.444	1.108	0.023	0.476	1.113
	3	24:8	0.029	0.570	1.279	0.008	0.543	1.266	0.062	0.596	1.288
		15:5	0.033	0.460	1.089	0.050	0.446	1.087	0.016	0.472	1.093
		30:10	0.055	0.548	1.196	0.024	0.522	1.184	0.083	0.572	1.207
-0.50	4	24:8	0.025	0.510	1.142	0.006	0.487	1.132	0.052	0.530	1.150
		15:5	0.072	0.553	1.180	0.046	0.528	1.169	0.097	0.574	1.189
		30:10	0.086	0.558	1.178	0.058	0.532	1.162	0.114	0.583	1.194
	5	24:8	0.082	0.547	1.153	0.058	0.523	1.138	0.107	0.568	1.167
		15:5	0.114	0.573	1.175	0.090	0.547	1.155	0.139	0.597	1.190
		30:10	0.039	0.566	1.268	0.006	0.531	1.235	0.078	0.601	1.292
	3	24:8	0.050	0.432	1.056	0.071	0.411	1.042	0.028	0.450	1.067
		15:5	0.035	0.545	1.214	0.010	0.521	1.197	0.061	0.568	1.225
		30:10	0.094	0.610	1.311	0.085	0.563	1.275	0.141	0.657	1.342
0.00	4	24:8	0.029	0.502	1.119	0.005	0.475	1.105	0.057	0.528	1.137
		15:5	0.066	0.540	1.159	0.040	0.513	1.142	0.094	0.563	1.173
		30:10	0.088	0.551	1.162	0.059	0.522	1.141	0.117	0.579	1.183
	5	24:8	0.082	0.535	1.128	0.057	0.508	1.111	0.108	0.562	1.153
		15:5	0.110	0.560	1.154	0.084	0.533	1.135	0.135	0.587	1.177
		30:10	0.020	0.456	1.077	0.043	0.433	1.060	0.006	0.480	1.095
	3	24:8	0.037	0.430	1.039	0.057	0.411	1.024	0.016	0.452	1.057
		15:5	0.019	0.453	1.068	0.040	0.434	1.053	0.006	0.475	1.083
		30:10	0.055	0.523	1.142	0.025	0.493	1.120	0.085	0.480	1.164
0.50	4	24:8	0.034	0.493	1.100	0.009	0.467	1.080	0.061	0.520	1.120
		15:5	0.064	0.528	1.140	0.036	0.500	1.120	0.091	0.555	1.159
		30:10	0.086	0.537	1.136	0.057	0.508	1.113	0.210	0.567	1.161
	5	24:8	0.082	0.526	1.112	0.058	0.500	1.089	0.107	0.553	1.136
		15:5	0.105	0.551	1.140	0.079	0.522	1.117	0.132	0.579	1.163

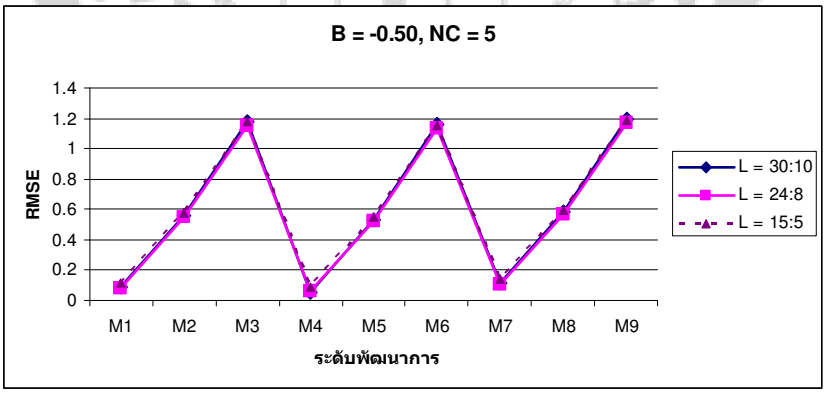
*____หมายถึง มีค่า RMSE ต่ำที่สุด ภายใต้เงื่อนไขระดับความยาก และ การให้คะแนนเท่ากัน



63 ก

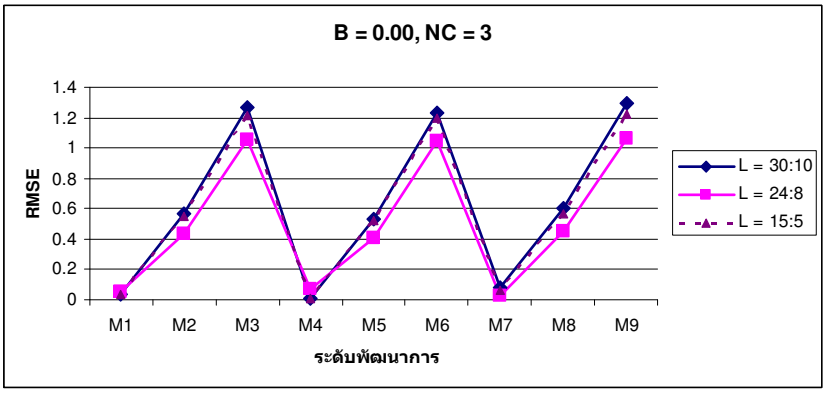


63 ข

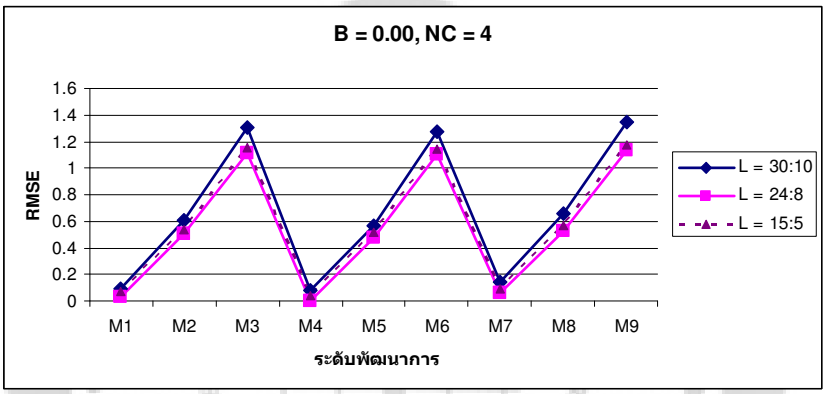


63 ค

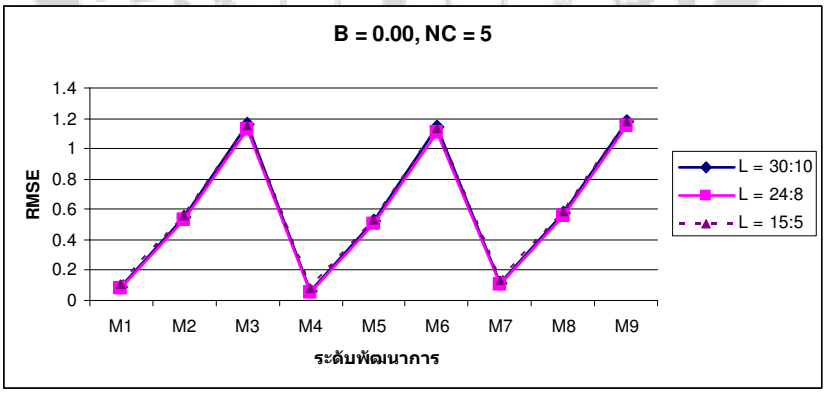
ภาพประกอบ 63 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก -0.50 และให้คะแนน 3 ระดับ (63 ก) 4 ระดับ (63 ข) 5 ระดับ (63 ค)



64 ก

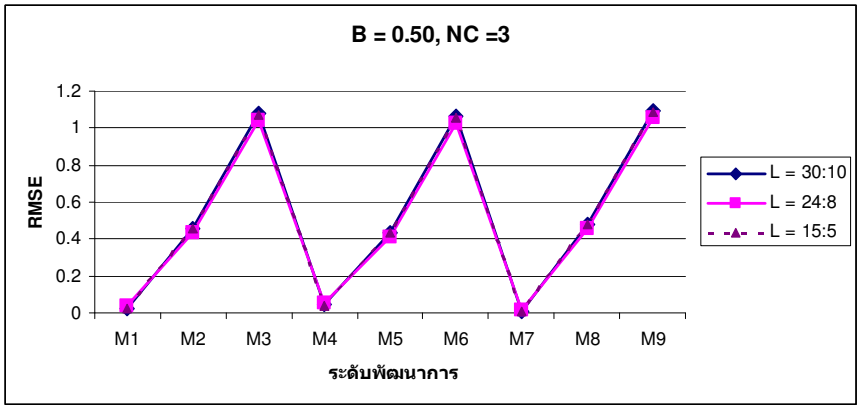


64 ข

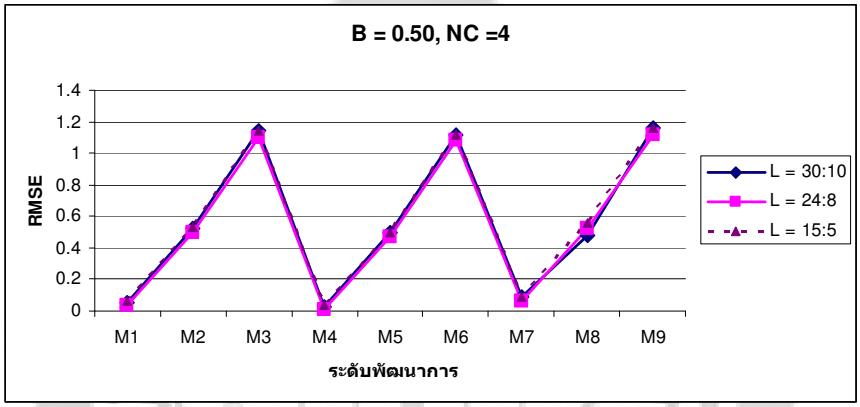


64 ค

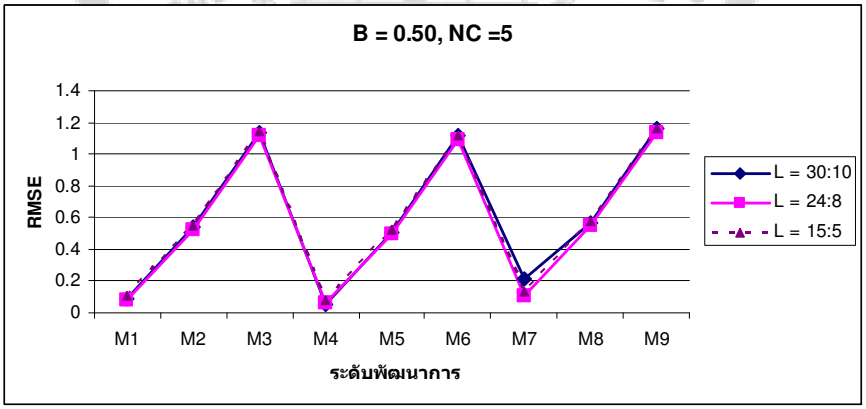
ภาพประกอบ 64 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.00 และให้คะแนน 3 ระดับ (64 ก) 4 ระดับ (64 ข) 5 ระดับ (64 ค)



65 ก



65 ข



65 ค

ภาพประกอบ 65 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อแบบทดสอบผสมมีความยาก 0.50 และให้คะแนน 3 ระดับ (65 ก) 4 ระดับ (65 ข) 5 ระดับ (65 ค)

ตาราง 44 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเมื่อแบบทดสอบผสมมี ความยาว และการให้คะแนนคงที่ จำแนกตามความยากและระดับพัฒนาการ

NC	L	B	M1	M2	M3	M4	M5	M6	M7	M8	M9
		-0.50	0.044	0.461	1.112	0.066	0.444	1.108	0.023	0.476	1.113
	30:10	0.00	0.029	0.570	1.279	0.008	0.543	1.266	0.062	0.596	1.288
		0.50	0.033	0.460	1.089	0.050	0.446	1.087	0.016	0.472	1.093
		-0.50	0.039	0.566	1.268	0.006	0.531	1.235	0.078	0.601	1.292
3 ระดับ	24:8	0.00	0.050	0.432	1.056	0.071	0.411	1.042	0.028	0.450	1.067
		0.50	0.035	0.545	1.214	0.010	0.521	1.197	0.061	0.568	1.225
		-0.50	0.020	0.456	1.077	0.043	0.433	1.060	0.006	0.480	1.095
	15:5	0.00	0.037	0.430	1.039	0.057	0.411	1.024	0.016	0.452	1.057
		0.50	0.019	0.453	1.068	0.040	0.434	1.053	0.006	0.475	1.083
		-0.50	0.055	0.548	1.196	0.024	0.522	1.184	0.083	0.572	1.207
	30:10	0.00	0.025	0.510	1.142	0.006	0.487	1.132	0.052	0.530	1.150
		0.50	0.072	0.553	1.180	0.046	0.528	1.169	0.097	0.574	1.189
		-0.50	0.094	0.610	1.311	0.085	0.563	1.275	0.141	0.657	1.342
4 ระดับ	24:8	0.00	0.029	0.502	1.119	0.005	0.475	1.105	0.057	0.528	1.137
		0.50	0.066	0.540	1.159	0.040	0.513	1.142	0.094	0.563	1.173
		-0.50	0.055	0.523	1.142	0.025	0.493	1.120	0.085	0.480	1.164
	15:5	0.00	0.034	0.493	1.100	0.009	0.467	1.080	0.061	0.520	1.120
		0.50	0.064	0.528	1.140	0.036	0.500	1.120	0.091	0.555	1.159
		-0.50	0.086	0.558	1.178	0.058	0.532	1.162	0.114	0.583	1.194
	30:10	0.00	0.082	0.547	1.153	0.058	0.523	1.138	0.107	0.568	1.167
		0.50	0.114	0.573	1.175	0.090	0.547	1.155	0.139	0.597	1.190
		-0.50	0.088	0.551	1.162	0.059	0.522	1.141	0.117	0.579	1.183
5 ระดับ	24:8	0.00	0.082	0.535	1.128	0.057	0.508	1.111	0.108	0.562	1.153
		0.50	0.110	0.560	1.154	0.084	0.533	1.135	0.135	0.587	1.177
		-0.50	0.086	0.537	1.136	0.057	0.508	1.113	0.210	0.567	1.161
	15:5	0.00	0.082	0.526	1.112	0.058	0.500	1.089	0.107	0.553	1.136
		0.50	0.105	0.551	1.140	0.079	0.522	1.117	0.132	0.579	1.163

*__หมายถึง มีค่า RMSE ต่ำที่สุด ภายใต้เงื่อนไขระดับความยาว และ ระดับการให้คะแนนเท่ากัน

4.1.2.3 ระดับการให้คะแนนคงที่

กรณีให้คะแนน 3 ระดับ

จากตาราง 44 และภาพประกอบ 66 (ก – ค) พบว่าเมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมมีความยาว 30:10 กลุ่มส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก 0.50 ส่วนกลุ่ม M1 และ M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก 0.00 และกลุ่ม M5 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก -0.50

เมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมมีขนาด 24:8 กลุ่มพัฒนาการส่วนใหญ่มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก 0.00 ส่วนกลุ่ม M1 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก 0.50 และกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก -0.50

เมื่อให้คะแนน 3 ระดับ และแบบทดสอบผสมมีขนาด 15:5 กลุ่มพัฒนาการส่วนใหญ่ค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก 0.00 ส่วนกลุ่ม M1 และ M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก 0.50 และกลุ่ม M7 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก -0.50 และ 0.50 และมีค่าเท่ากัน

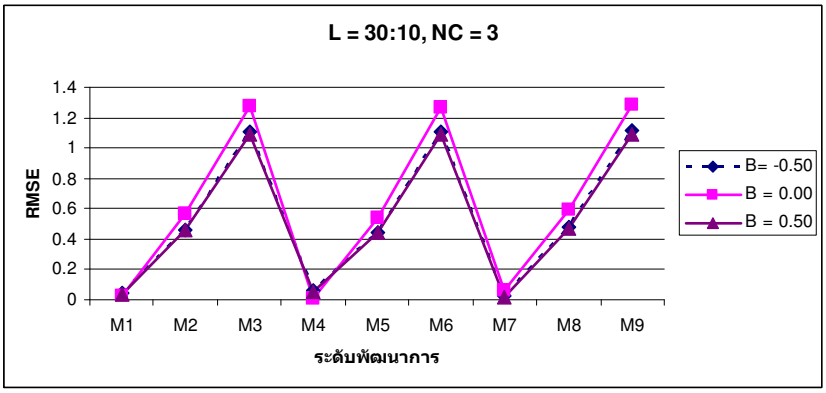
กรณีให้คะแนน 4 ระดับ

จากตาราง 44 และภาพประกอบ 67 (ก – ค) พบว่าทุกกรณีที่ให้คะแนน 4 ระดับในทุกขนาดความยาวของแบบทดสอบผสมพบว่าทุกกลุ่มมีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก 0.00

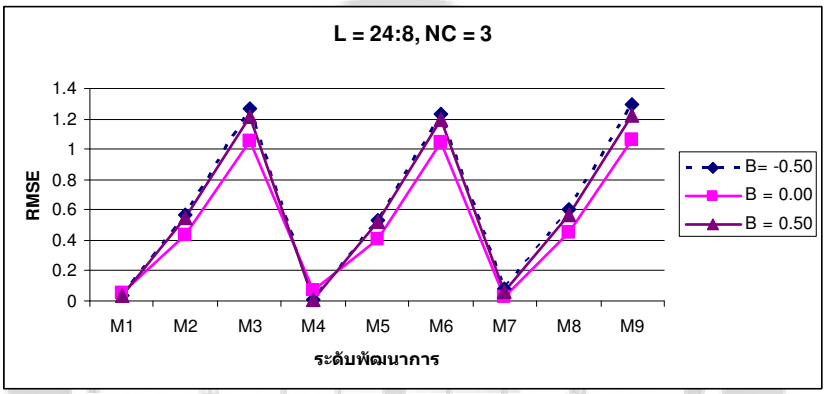
กรณีให้คะแนน 5 ระดับ

จากตาราง 44 และภาพประกอบ 64 (ก – ค) เมื่อให้คะแนน 5 ระดับ และแบบทดสอบผสมมีความยาว 30:10 และมีความยาว 24:8 พบว่าทุกกลุ่มพัฒนาการมีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก 0.00

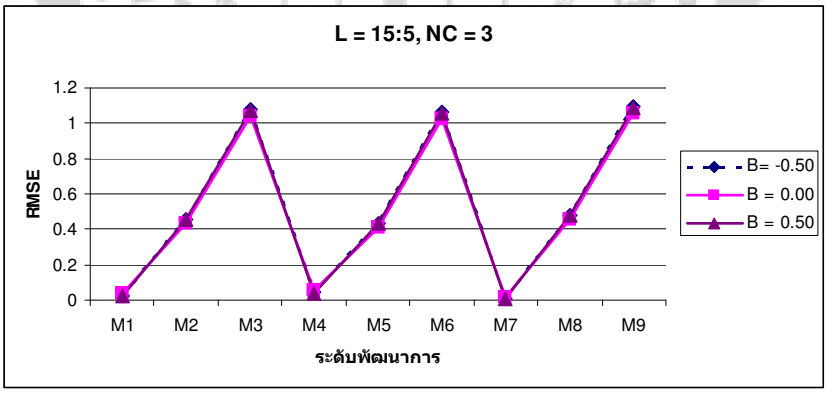
เมื่อให้คะแนน 5 ระดับ และแบบทดสอบผสมมีความยาว 15:5 กลุ่มพัฒนาการส่วนใหญ่ มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก 0.00 ยกเว้นกลุ่ม M4 มีค่า RMSE ของพัฒนาการความสามารถต่ำที่สุดในกรณีแบบทดสอบผสมมีความยาก -0.50



66 ก

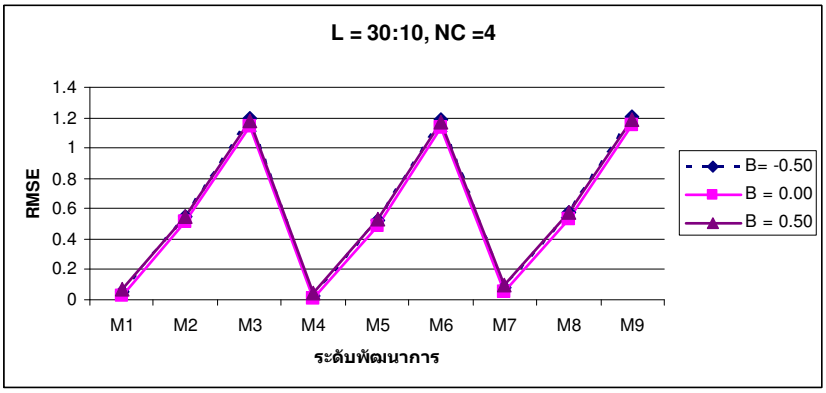


66 ข

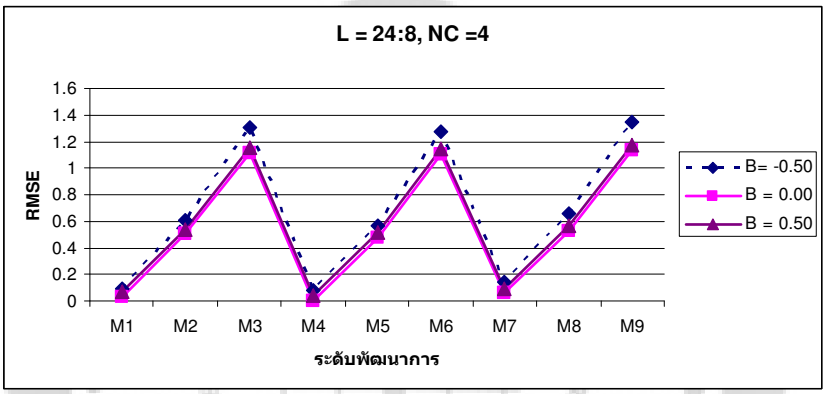


66 ค

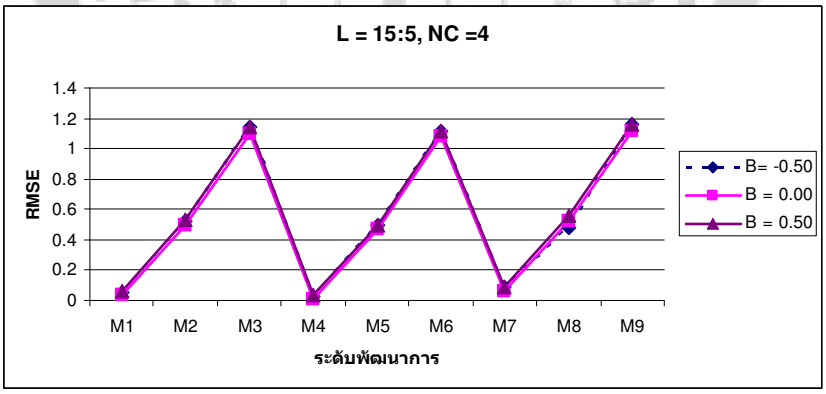
ภาพประกอบ 66 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อมีการให้คะแนน 3 ระดับ และแบบทดสอบผสมยาว 30:10 (66 ก) 24:8 (66 ข) 15:5 (66 ค)



67 ก

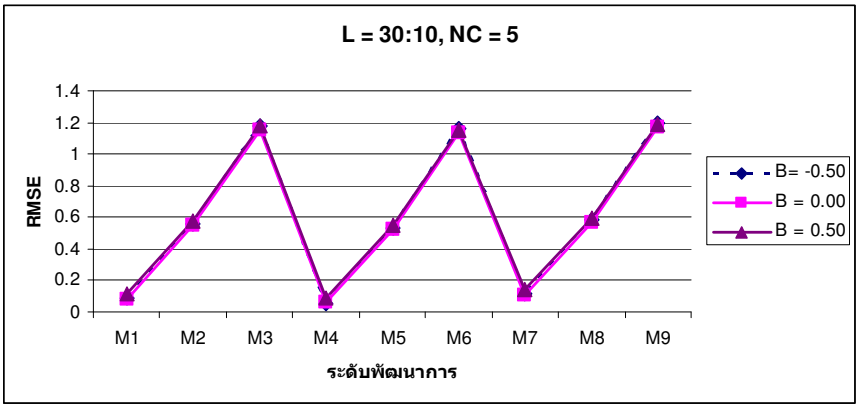


67 ข

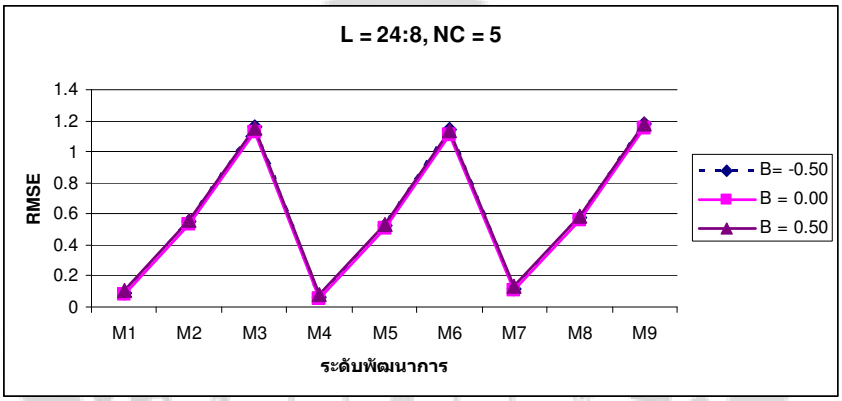


67 ค

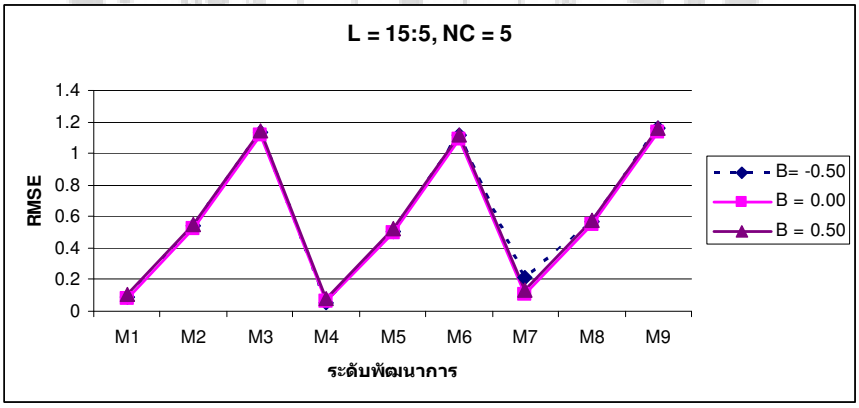
ภาพประกอบ 67 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อมีการให้คะแนน 4 ระดับ และแบบทดสอบผสมยาว 30:10 (67 ก) 24:8 (67 ข) 15:5 (67 ค)



68 ก



68 ข



68 ค

ภาพประกอบ 68 (ก – ค) แสดงค่า RMSE ของพัฒนาการความสามารถที่ประมาณค่าด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ($\hat{\theta}_2 - \hat{\theta}_1$) เมื่อเทียบกับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$) เมื่อมีการให้คะแนน 5 ระดับ และแบบทดสอบผสมยาว 30:10 (68 ก) 24:8 (68 ข) 15:5 (68 ค)

4.2 การเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ผู้วิจัยเปรียบเทียบความลำเอียง (BIAS) และค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE)ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ในแต่ละเงื่อนไข เพื่อทดสอบสมมติฐานการวิจัยข้อ 3 ที่ตั้งไว้ว่า "*เมื่อระดับพัฒนาการความสามารถของผู้สอบมีส่วนเบี่ยงเบนมาตรฐานเท่ากับหรือต่ำกว่าส่วนเบี่ยงเบนมาตรฐานของความสามารถครั้งแรก (1หรือ 0.8) ค่าประมาณความสามารถจากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตมีความแม่นยำมากกว่าเมื่อส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถของผู้สอบสูงกว่าส่วนเบี่ยงเบนมาตรฐานของความสามารถครั้งแรก*" ผลการวิเคราะห์แยกนำเสนอผลการเปรียบเทียบความแม่นยำออกเป็นสองส่วน ส่วนแรกเป็นการเปรียบเทียบความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันและส่วนที่สองเป็นการเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนของค่าประมาณความสามารถกำลังสอง (RMSE)ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันปรากฏผลตามลำดับดังนี้

4.2.1 การเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ย BIAS ที่คำนวณจากกลุ่มระดับพัฒนาการ 9 กลุ่มซึ่งแบ่งกลุ่มย่อยตามค่าส่วนเบี่ยงเบนมาตรฐานได้ 3 กลุ่ม ภายใต้เงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน โดยการวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (Repeated Measures) ปรากฏผลดังตาราง 44

ผลการวิเคราะห์ตามตาราง 44 พบว่าเมื่อค่าส่วนเบี่ยงเบนมาตรฐานของความสามารถผู้สอบในการสอบครั้งที่สองต่างกัน 3 ระดับ ค่าเฉลี่ย BIAS ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันแตกต่างอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนผลการทดสอบหลักภายในกลุ่มตัวแปรพบว่าเมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด ความยากของแบบทดสอบผสมแตกต่างกัน 3 ระดับ และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน 3 ระดับ ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทุกระดับตัวแปร แสดงว่าเมื่อแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน ส่งผลให้ค่า BIAS ของพัฒนาการความสามารถแตกต่างกัน

ตาราง 45 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกรณีแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ

แหล่งความแปรปรวน	SS	df	MS	F	ระดับนัยสำคัญ	ขนาดของผล
ทดสอบผลตัวแปรระหว่างกลุ่ม (Between)						
กลุ่ม SD	0.003	2.000	0.002	0.005	0.995	0.002
ทดสอบผลตัวแปรภายในกลุ่ม (Within)						
ความยาว	0.030	2.000	0.015	8.041*	0.006	0.573
ความยาก	0.064	1.072	0.06	10.840*	0.014	0.644
การให้คะแนน	0.144	1.032	0.139	13.461*	0.010	0.692
ความยาว × ความยาก	0.084	1.520	0.055	19.014*	0.000	0.760
ความยาว × การให้คะแนน	0.019	1.822	0.01	10.441*	0.003	0.635
ความยาก × การให้คะแนน	0.029	1.693	0.017	23.553*	0.000	0.797
ความยาว × ความยาก × การให้คะแนน	0.112	1.384	0.081	28.609*	0.000	0.827
ความยาว × กลุ่ม SD	0.005	4	0.001	0.618	0.658	0.171
ความยาก × กลุ่ม SD	0.002	4	0.000	0.197	0.935	0.062
การให้คะแนน × กลุ่ม SD	0.004	4	0.001	0.196	0.936	0.061
ความยาว × ความยาก × กลุ่ม SD	0.003	8	0.000	0.358	0.933	0.107
ความยาว × การให้คะแนน × กลุ่ม SD	0.002	8	0.000	0.561	0.799	0.158
ความยาก × การให้คะแนน × กลุ่ม SD	0.002	8	0.000	0.674	0.709	0.184
ความยาว × ความยาก × การให้คะแนน × กลุ่ม SD	0.005	16	0.000	0.597	0.870	0.166

* มีนัยสำคัญทางสถิติที่ระดับ .05

แต่อย่างไรก็ตามพบว่ามีปฏิสัมพันธ์ระหว่างความยาว กับ ความยากของแบบทดสอบผสม ความยาวและการให้คะแนนของแบบทดสอบผสม ความยาก และการให้คะแนนของแบบทดสอบผสม และ ความยาวความยากและการให้คะแนนของแบบทดสอบผสม ที่ระดับนัยสำคัญ .05 แสดงว่าความแตกต่างของค่า BIAS ของพัฒนาการความสามารถเป็นผลมาจากการส่งผลร่วมกันระหว่างตัวแปร ความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม

เมื่อผลการทดสอบพบว่าแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน ส่งผลให้ค่า BIAS ของพัฒนาการความสามารถแตกต่างกัน ผู้วิจัยจึงเปรียบเทียบค่าความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ในแต่ละคู่ระดับของ 3 เงื่อนไข ได้แก่ ความยาว ความยาก และ การให้คะแนน โดยเสนอผลการเปรียบเทียบรายคู่เป็น 3 ตอนเพื่อทดสอบของสมมติฐานการวิจัยข้อ 3 ย่อยดังนี้

4.2.1.1 การเปรียบเทียบความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์ พร้อมกัน เมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

ผู้วิจัยเปรียบเทียบความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเพื่อทดสอบของสมมติฐานการวิจัย ข้อ 3.1 ที่ตั้งไว้ว่า *"แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีแม่นยำมากกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ในทุกระดับความยาก และ ระดับการให้คะแนน"* ปรากฏผลการวิเคราะห์ดังตาราง 46

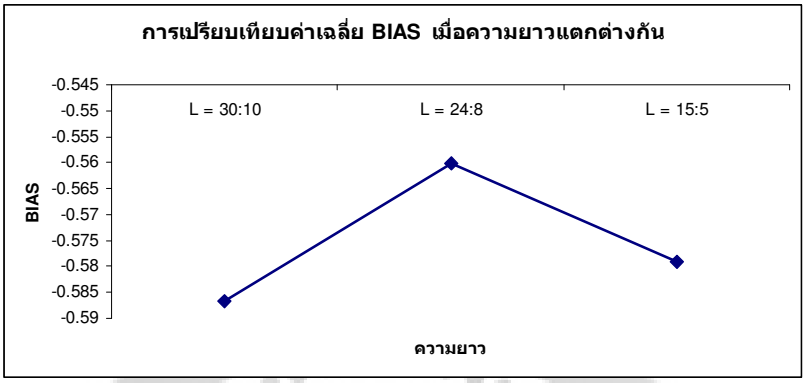
ตาราง 46 การวิเคราะห์ความแตกต่างค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาว	Mean	L = 30:10	L = 24:8	L = 15:5
	L = 30:10	-0.587	-	0.026*	0.018
	L = 24:8	-0.560		-	0.019*
	L = 15:5	-0.579			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 46 พบว่า เมื่อความยากและรูปแบบการให้คะแนนคงที่ ค่าเฉลี่ยของ BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยาว 24:8 มีค่าเฉลี่ยต่ำที่สุด รองลงมาคือ ขนาด 15:5 และ 30:10 ตามลำดับ ผลการเปรียบเทียบความแตกต่างค่าเฉลี่ย BIAS เป็นรายคู่พบว่ามี 2 คู่ที่ค่าเฉลี่ยแตกต่างกันโดยพบว่าแบบทดสอบขนาด 24:8 มี BIAS ต่ำกว่าขนาด 30:10 และแบบทดสอบขนาด 24:8 มี BIAS ต่ำกว่าขนาด 15:5 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบขนาด 30:10 มี BIAS ต่ำกว่าขนาด 15:5 อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

แสดงการเปรียบเทียบค่า BIAS ของพัฒนาการความสามารถตามขนาดความยาวของแบบทดสอบผสม ได้ดังภาพประกอบ 69



ภาพประกอบ 69 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวต่างกัน

จากภาพประกอบ 69 ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีค่าเป็นลบทุกระดับความยาว โดยที่แบบทดสอบผสมที่มีความยาว 24:8 ความลำเอียงต่ำที่สุด รองลงมาคือแบบทดสอบผสมที่มีความยาว 15:5 และ 30:10 ตามลำดับ

4.2.1.2 การเปรียบเทียบความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

ผู้วิจัยเปรียบเทียบความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เพื่อทดสอบของสมมติฐานการวิจัยข้อ 3.2 ที่ตั้งไว้ว่า *แบบทดสอบผสมที่มีระดับความยากของแบบทดสอบสอดคล้องกับระดับความสามารถเริ่มต้นของผู้สอบจะมีความแม่นยำมากกว่าแบบทดสอบผสมที่มีระดับความยากของแบบทดสอบไม่สอดคล้องกับระดับความสามารถเริ่มต้นของผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนน* ปรากฏผลการวิเคราะห์ดังตาราง 47

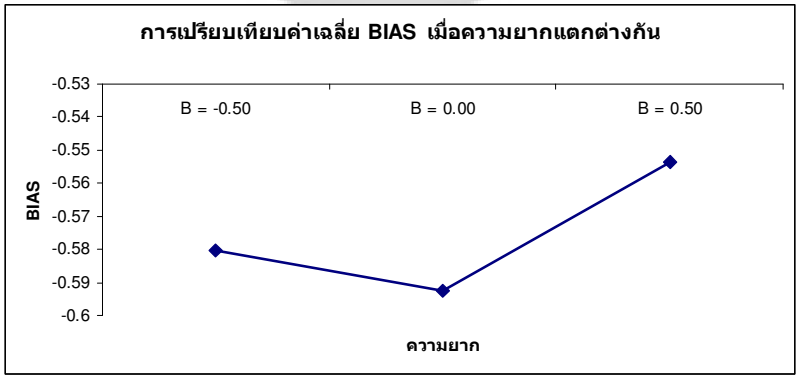
ตาราง 47 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน จำแนกตามระดับพัฒนาการความสามารถ

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาก	Mean	B = -0.50	B = 0.00	B = 0.50
	B = -0.50	-0.580	-	0.012	0.027*
	B = 0.00	-0.592		-	0.039*
	B = 0.50	-0.553			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 47 พบว่า เมื่อความยาวและรูปแบบการให้คะแนนคงที่ ค่า BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยาก 0.50 มีค่าเฉลี่ย BIAS ต่ำที่สุด รองลงมาคือแบบทดสอบที่มีความยาก -0.50 และ 0.00 ตามลำดับ ผลการเปรียบเทียบความแตกต่างค่าเฉลี่ย BIAS เป็นรายคู่พบว่ามี 2 คู่ที่ค่าเฉลี่ยแตกต่างกัน โดยแบบทดสอบที่มีความยาก 0.50 มีค่าเฉลี่ย BIAS ต่ำกว่า แบบทดสอบที่มีความยาก 0.50 และแบบทดสอบที่มีความยาก0.50 มีค่าเฉลี่ย BIAS ต่ำกว่าแบบทดสอบที่มีความยาก 0.50 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่มีความยาก -0.50 มีค่าเฉลี่ย BIAS ต่ำกว่าแบบทดสอบที่มีความยาก 0.00 อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

แสดงการเปรียบเทียบค่า BIAS ของพัฒนาการความสามารถ ตามระดับความยากของแบบทดสอบผสม ได้ดังภาพประกอบ 70



ภาพประกอบ 70 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากต่างกัน

จากภาพประกอบ 70 ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีค่าเป็นลบทุกระดับความยาก โดยที่แบบทดสอบผสมที่มีความยาก 0.50 ความลำเอียงต่ำที่สุด รองลงมาคือ แบบทดสอบผสมที่มีความยาก -0.50 และ 0.00 ตามลำดับ

4.2.1.3 การเปรียบเทียบความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีรูปแบบการให้คะแนนแตกต่างกัน

ผู้วิจัยเปรียบเทียบความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ข้อ 3.2 ที่ตั้งไว้ว่า *แบบทดสอบผสมที่มีระดับการให้คะแนนข้อสอบเขียนตอบสูงกว่า จะมีความแม่นยำมากกว่าแบบทดสอบผสมที่มีระดับการให้คะแนนข้อสอบเขียนตอบต่ำกว่า ในทุกขนาดความยาวและความยากของแบบทดสอบผสม* ปรากฏผลการวิเคราะห์ดังตาราง 48

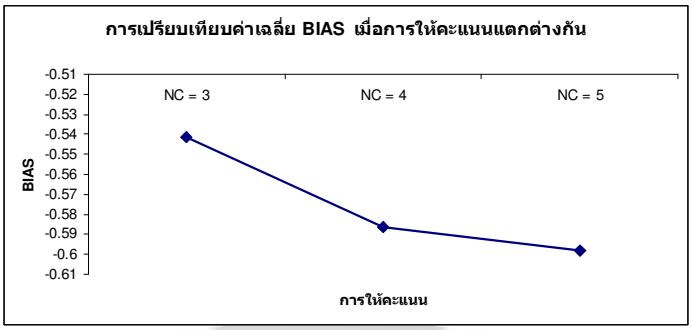
ตาราง 48 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน 3 ระดับ จำแนกตามระดับ SD ของพัฒนาการความสามารถ

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	การให้คะแนน	Mean	NC=3	NC=4	NC=5
	NC=3	-0.542	-	0.045*	0.056*
	NC=4	-0.586		-	0.012
	NC=5	-0.598			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 48 พบว่า เมื่อความยาวและความยากคงที่ ค่า BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีการให้คะแนน 3 ระดับคะแนนมีค่า BIAS ต่ำที่สุด รองลงมาคือ การให้คะแนน 4 ระดับ และ 5 ระดับตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ย BIAS เป็นรายคู่ โดยพบว่ามีจำนวน 2 คู่ที่มีค่าเฉลี่ย BIAS แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 คือ การให้คะแนน 3 ระดับ มีค่าเฉลี่ย BIAS ต่ำกว่า การให้คะแนน 4 ระดับ และ 5 ระดับ ส่วนการให้คะแนน 4 ระดับ มีค่าเฉลี่ย BIAS ต่ำกว่า การให้คะแนน 5 ระดับ อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

แสดงการเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถ ตามระดับการให้คะแนน ได้ดังภาพประกอบ 71



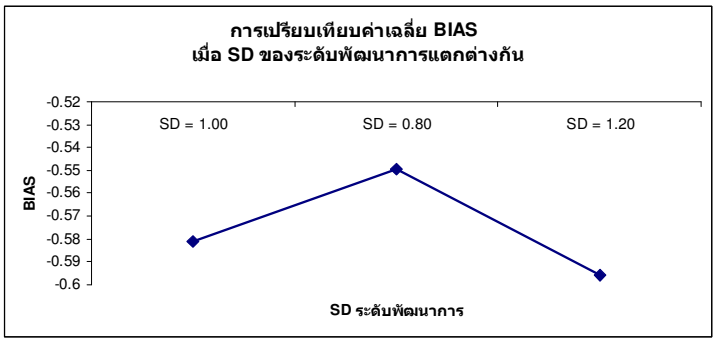
ภาพประกอบ 71 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

จากภาพประกอบ 71 ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีค่าเป็นลบทุกระดับการให้คะแนน โดยที่แบบทดสอบผสมที่ให้คะแนน 3 ระดับ มีความลำเอียงต่ำที่สุด รองลงมาคือแบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ

แม้ว่าผลการทดสอบหลักจะพบว่าค่าเฉลี่ย BIAS ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีแตกต่างตามกลุ่มส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถ 1.00 0.80 และ 1.20 อย่างไม่มีนัยสำคัญที่ระดับ .05 แต่เมื่อพิจารณาค่าเฉลี่ยของ BIAS ในแต่ละกลุ่ม SD ของระดับได้ข้อสังเกตว่า กลุ่มพัฒนาการที่มีค่าส่วนเบี่ยงเบนมาตรฐาน (SD) เป็น 0.80 และ 1.00 มีค่าเฉลี่ย BIAS ต่ำกว่ากลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 1.2

แสดงการเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถ ตามกลุ่มส่วนเบี่ยงเบนของพัฒนาการเป็น 1.00 0.80 และ 1.20 ได้ดังภาพประกอบ 72

จากภาพประกอบ 72 ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีค่าเป็นลบทุกค่าส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถ โดยที่กลุ่มพัฒนาการที่มีค่าส่วนเบี่ยงเบนมาตรฐาน (SD) เป็น 0.80 มีความลำเอียงต่ำที่สุด และ 1.00 รองลงมาคือกลุ่มพัฒนาการที่มีค่าส่วนเบี่ยงเบนมาตรฐาน (SD) เป็น 0.8 และ 1.2 ตามลำดับ



ภาพประกอบ 72 การเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อกลุ่มส่วนเบี่ยงเบนของพัฒนาการเป็น 1.00 0.80 และ 1.20

4.2.2 การเปรียบเทียบค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยของ RMSE ที่คำนวณจากกลุ่มระดับพัฒนาการ 9 กลุ่มซึ่งแบ่งย่อย 3 กลุ่มตามค่าส่วนเบี่ยงเบนมาตรฐาน ภายใต้เงื่อนไขความยาว ความยาก และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน โดยใช้การวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (Repeated Measures) ปรากฏผลดังตาราง 49

ผลการวิเคราะห์ตามตาราง 49 พบว่าผลการทดสอบหลักระหว่างกลุ่มเมื่อค่าส่วนเบี่ยงเบนมาตรฐานของความสามารถผู้สอบในการสอบครั้งที่สองต่างกัน 3 ระดับ ค่าเฉลี่ย RMSE ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันแตกต่างอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนผลการทดสอบหลักภายในกลุ่มตัวแปรพบว่าเมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด ความยากของแบบทดสอบผสม 3 ระดับ และการให้คะแนนของแบบทดสอบผสม 3 ขนาด ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทุกระดับตัวแปร แสดงว่าเมื่อแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน ส่งผลให้ค่า RMSE ของพัฒนาการความสามารถแตกต่างกัน

แต่อย่างไรก็ตามพบว่ามีปฏิสัมพันธ์ระหว่างความยาว กับ ความยากของแบบทดสอบ ความยาวและการให้คะแนนของแบบทดสอบ ความยาก และการให้คะแนนของแบบทดสอบ และความยาวความยากและการให้คะแนนของแบบทดสอบ ที่ระดับนัยสำคัญ .05 แสดงว่าความแตกต่างของค่า RMSE ของพัฒนาการความสามารถเป็นผลมาจากการส่งผลร่วมกันระหว่างตัวแปร ความยาว ความยาก และการให้คะแนนของแบบทดสอบ

ตาราง 49 การวิเคราะห์ความแปรปรวนของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกรณีแบบทดสอบผสมมีความยาวต่างกัน 3 ขนาด ความยากต่างกัน 3 ระดับ และ การให้คะแนนต่างกัน 3 ระดับ

แหล่งความแปรปรวน	SS	df	MS	F	ระดับนัยสำคัญ	ขนาดของผล
ทดสอบผลตัวแปรระหว่างกลุ่ม (Between)						
กลุ่ม SD	0.002	2.000	0.001	0.004	0.996	0.001
ทดสอบผลตัวแปรภายในกลุ่ม (Within)						
ความยาว	0.041	1.049	0.039	23.505*	0.002	0.797
ความยาก	0.068	1.004	0.068	12.417*	0.012	0.674
การให้คะแนน	0.093	2.000	0.047	15.651*	0.000	0.723
ความยาว × ความยาก	0.068	1.125	0.060	10.695*	0.013	0.641
ความยาว × การให้คะแนน	0.025	1.769	0.014	32.896*	0.000	0.846
ความยาก × การให้คะแนน	0.026	1.231	0.021	15.723*	0.004	0.724
ความยาว × ความยาก × การให้คะแนน	0.082	1.121	0.073	9.523*	0.017	0.613
ความยาว × กลุ่ม SD	0.001	4	0.001	0.202	0.932	0.063
ความยาก × กลุ่ม SD	0.002	4	0.001	0.143	0.963	0.046
การให้คะแนน × กลุ่ม SD	0.007	4	0.002	0.601	0.669	0.167
ความยาว × ความยาก × กลุ่ม SD	0.002	8	0.000	0.148	0.996	0.047
ความยาว × การให้คะแนน × กลุ่ม SD	0.000	8	0.000	0.200	0.988	0.063
ความยาก × การให้คะแนน × กลุ่ม SD	0.002	8	0.000	0.594	0.773	0.165
ความยาว × ความยาก × การให้คะแนน × กลุ่ม SD	0.005	16	0.000	0.284	0.996	0.086

* มีนัยสำคัญทางสถิติที่ระดับ .05

เมื่อผลการทดสอบพบว่าแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน ส่งผลให้ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถแตกต่างกัน ผู้วิจัยจึงเปรียบเทียบค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อน (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ในแต่ละคู่ระดับของ 3 เงื่อนไข ได้แก่ ความยาว ความยาก และ การให้คะแนน โดยเสนอผลการเปรียบเทียบรายคู่เป็น 3 ตอนเพื่อทดสอบของสมมติฐานการวิจัยย่อยข้อ 3 ดังนี้

4.2.2.1การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยของค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกันเพื่อทดสอบสมมติฐานข้อ 3.1 *ที่ว่า "แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีแม่นยำมากกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่าในทุกระดับความยาก และ ระดับการให้คะแนน"* ปรากฏผลการวิเคราะห์ดังตาราง 50

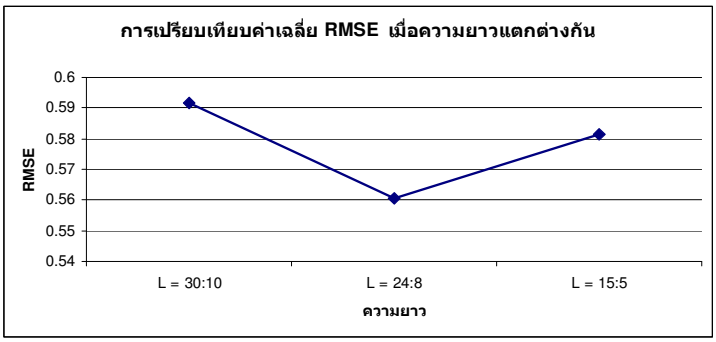
ตาราง 50 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาว	Mean	L = 30:10	L = 24:8	L = 15:5
	L = 30:10	0.592	-	0.031*	0.010
	L = 24:8	0.560		-	0.021*
	L = 15:5	0.581			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 50 พบว่า เมื่อความยากและรูปแบบการให้คะแนนคงที่ ค่า RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยาวต่างกัน มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยค่าเฉลี่ย RMSE ของแบบทดสอบขนาด 24:8 มีค่าต่ำที่สุด รองลงมาคือ แบบทดสอบขนาด 15:5 และ 30:10 ตามลำดับ เมื่อเปรียบเทียบค่าเฉลี่ย RMSE รายคู่ พบว่าแตกต่างกัน 2 คู่ โดยแบบทดสอบผสมขนาด 24:8 มีค่า RMSE ต่ำกว่าแบบทดสอบขนาด 30:10 และ แบบทดสอบผสมขนาด 24:8 มีค่า RMSE ต่ำกว่าแบบทดสอบขนาด 15:5 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบผสมขนาด 15:5 มีค่า RMSE ต่ำกว่าแบบทดสอบขนาด 30:10 อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

แสดงการเปรียบเทียบค่า RMSE ของพัฒนาการความสามารถ ตามขนาดความยาวของแบบทดสอบผสม ได้ดังภาพประกอบ 73



ภาพประกอบ 73 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาวแตกต่างกัน

จากภาพประกอบ 73 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาว 24:8 มีค่าต่ำที่สุด รองลงมาคือแบบทดสอบผสมที่มีความยาว 15:5 และ 30:10 ตามลำดับ

4.2.2.2 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนของค่าประมาณความสามารถกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเพื่อทดสอบสมมติฐาน ข้อ 3.2 ที่ตั้งไว้ว่า *แบบทดสอบผสมที่มีระดับความยากของแบบทดสอบสอดคล้องกับระดับความสามารถเริ่มต้นของผู้สอบ จะมีความแม่นยำมากกว่าแบบทดสอบผสมที่มีระดับความยากของแบบทดสอบไม่สอดคล้องกับระดับความสามารถเริ่มต้นของผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนน* ปรากฏผลการวิเคราะห์ดังตาราง 51

จากตาราง 51 พบว่า เมื่อความยาวและรูปแบบการให้คะแนนคงที่ ค่า RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยากต่างกัน มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยแบบทดสอบที่มีความยาก 0.50 มีค่าเฉลี่ย RMSE ต่ำที่สุดรองลงมาคือ แบบทดสอบที่มีความยาก -0.50 และ 0.00 ตามลำดับ เมื่อเปรียบเทียบค่าเฉลี่ย RMSE เป็นรายคู่พบว่า แตกต่างกัน 2 คู่ โดยแบบทดสอบที่มีความยาก -0.50 มีค่าเฉลี่ย RMSE ต่ำกว่า แบบทดสอบที่มีความยาก 0.00 และ แบบทดสอบที่มีความยาก 0.50 มีค่าเฉลี่ย RMSE ต่ำกว่า แบบทดสอบที่มีความยาก 0.00 อย่างมีนัยสำคัญทางสถิติที่

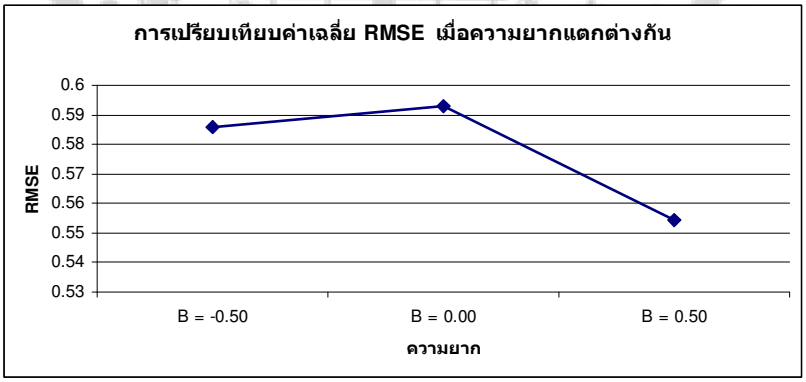
ระดับ .05 ส่วนแบบทดสอบที่มีความยาก 0.50 มีค่าเฉลี่ย RMSE ต่ำกว่าแบบทดสอบที่มีความยาก -0.50 มี อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 51 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	ความยาก	Mean	B = -0.50	B = 0.00	B = 0.50
	B = -0.50	0.586	-	0.007*	0.031
	B = 0.00	0.593		-	0.039*
	B = 0.50	0.554			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

แสดงการเปรียบเทียบค่า RMSE ของพัฒนาการความสามารถ ตามขนาดความยากของแบบทดสอบผสม ได้ดังภาพประกอบ 74



ภาพประกอบ 74 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยากแตกต่างกัน

จากภาพประกอบ 74 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีความยาก 0.50 มีค่าต่ำที่สุด รองลงมาคือแบบทดสอบผสมที่มีความยาก -0.50 และ 0.00 ตามลำดับ

4.2.2.3 การเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

ผู้วิจัยเปรียบเทียบค่าเฉลี่ยความคลาดเคลื่อนของค่าประมาณความสามารถกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เพื่อทดสอบสมมติฐานข้อ 3.2 ที่ตั้งไว้ว่า *แบบทดสอบผสมที่มีระดับการให้คะแนนข้อสอบเขียนตอบสูงกว่า จะมีความแม่นยำมากกว่าแบบทดสอบผสมที่มีระดับการให้คะแนนข้อสอบเขียนตอบต่ำกว่า ในทุกขนาดความยาวและความยากของแบบทดสอบผสม*ปรากฏผลการวิเคราะห์ดังตาราง 52

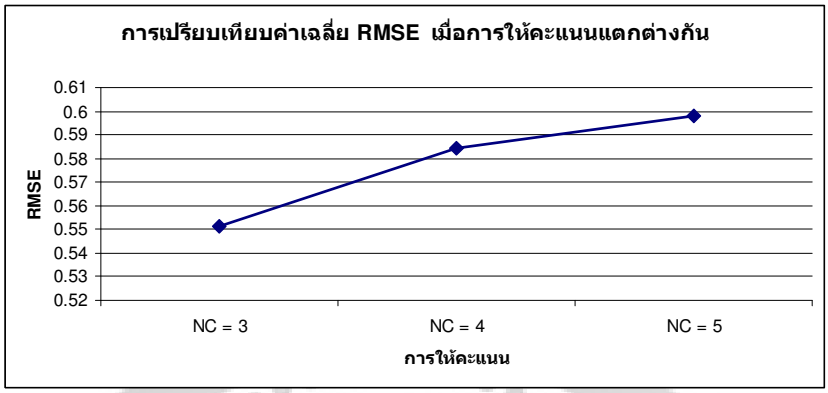
ตาราง 52 การวิเคราะห์ความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

องค์ประกอบ			การเปรียบเทียบความแตกต่าง		
ภายในกลุ่ม	การให้คะแนน	Mean	NC=3	NC=4	NC=5
	NC = 3	0.551	-	0.033*	0.047*
	NC = 4	0.584		-	0.014
	NC = 5	0.598			-

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 52 พบว่า เมื่อแบบทดสอบผสมมีความยาวและความยากคงที่ ค่า RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันในกรณีที่มีรูปแบบการให้คะแนนต่างกันมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยพบว่าการให้คะแนน 3 ระดับมีค่าเฉลี่ย RMSE ต่ำที่สุด รองลงมาคือการให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับเมื่อเปรียบเทียบค่าเฉลี่ยรายคู่พบว่า แตกต่าง 2 คู่ โดยการให้คะแนน 3 ระดับมีค่า RMSE ต่ำกว่า การให้คะแนน 4 ระดับ และ5 ระดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนให้คะแนน 4 ระดับมีค่า RMSE ต่ำกว่า การให้คะแนน 5 ระดับ อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

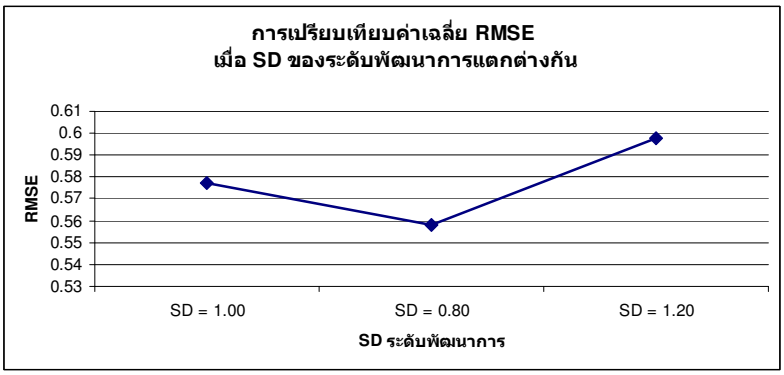
แสดงการเปรียบเทียบค่า RMSE ของพัฒนาการความสามารถ ตามขนาดความยากของแบบทดสอบผสม ได้ดังภาพประกอบ 75



ภาพประกอบ 75 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมมีการให้คะแนนแตกต่างกัน

จากภาพประกอบ 75 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบผสมให้คะแนน 3 ระดับมีค่าต่ำที่สุด รองลงมาคือแบบทดสอบผสมให้คะแนน 4 ระดับ และ 5 ระดับ ตามลำดับ

แม้ว่าผลการทดสอบจะพบว่าค่าเฉลี่ย BIAS ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันแตกต่างตามกลุ่มส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถ 1.00 0.80 และ 1.20 อย่างไม่มีนัยสำคัญที่ระดับ .05 แต่เมื่อพิจารณาค่าเฉลี่ยของ BIAS ในแต่ละกลุ่ม SD ของระดับได้ข้อสังเกตว่า เมื่อกลุ่มพัฒนาการมีค่าส่วนเบี่ยงเบนมาตรฐาน (SD) เป็น 0.80 และ 1.00 มีค่าเฉลี่ย BIAS ต่ำกว่ากลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 แสดงการเปรียบเทียบค่าเฉลี่ย BIAS ของพัฒนาการความสามารถ ตามกลุ่มส่วนเบี่ยงเบนมาตรฐานของพัฒนาการเป็น 1.00 0.80 และ 1.20 ได้ดังภาพประกอบ 76



ภาพประกอบ 76 การเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อจำแนกตามกลุ่มส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการเป็น 1.00 0.80 และ 1.20

จากภาพประกอบ 76 ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อส่วนเบี่ยงเบนมาตรฐานของกลุ่มระดับพัฒนาการเป็น 0.80 มีค่าต่ำที่สุด รองลงมาคือส่วนเบี่ยงเบนมาตรฐานของกลุ่มระดับพัฒนาการเป็น 1.00 และ 1.20 ตามลำดับ

สรุปการตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ความแม่นยำของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ภายใต้เงื่อนไขแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน และพัฒนาการความสามารถของผู้สอบแตกต่างกัน ปรากฏผลสรุปดังนี้

1. BIAS **ของพัฒนาการความสามารถ**

ค่าเฉลี่ย BIAS ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันแตกต่างตามกลุ่มส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถ 1.00 0.80 และ 1.20 อย่างไม่มีนัยสำคัญที่ระดับ .05 โดยกลุ่มพัฒนาการที่มีค่าส่วนเบี่ยงเบนมาตรฐาน (SD) เป็น 0.80 และ 1.00 มีค่าเฉลี่ย BIAS ต่ำกว่ากลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 1.2

เมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด ความยากของแบบทดสอบผสมแตกต่างกัน 3 ระดับ และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน 3 ระดับค่าเฉลี่ย BIAS ของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งความแตกต่างของค่าเฉลี่ย BIAS ของพัฒนาการความสามารถเป็นผลมาจากการส่งผลร่วมกันระหว่างตัวแปร ความยาว ความยาก และการให้คะแนนของแบบทดสอบผสม ผลการเปรียบเทียบค่าเฉลี่ยความลำเอียง (BIAS) ของพัฒนาการความสามารถ ในแต่ละคู่ระดับของ 3 เงื่อนไข ปรากฏผลดังนี้

1.1 เมื่อความยากและการให้คะแนนคงที่ ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยาว 30:10 มีค่าสูงสุด รองลงมาคือ 15:5 และ 24:8 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าแบบทดสอบผสมที่มีความยาว 30:10 มีความลำเอียงสูงกว่า แบบทดสอบผสมที่มีความยาว 24:8 และ แบบทดสอบผสมที่มีความยาว 15:5 มีความลำเอียงสูงกว่าแบบทดสอบผสมที่มีความยาว 24:8 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบผสมที่มีความยาว 30:10 กับ15:5 มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

1.2 เมื่อความยาวและรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่มีค่าความยาก 0.00 มีค่าเฉลี่ย BIAS สูงสุด รองลงมาคือ แบบทดสอบผสมที่มีความยาก -0.50 และ 0.50 ตามลำดับ และพบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบผสมที่มีความยาก 0.50 และ แบบทดสอบผสมที่มีความยาก 0.00 มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบที่มีความยาก 0.50 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่มีความยาก 0.00 แบบทดสอบที่มีความยาก -0.50 มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

1.3 เมื่อแบบทดสอบผสมมีความยาวและความยากคงที่ ค่าเฉลี่ย BIAS ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีการให้คะแนน 5 ระดับมีค่าเฉลี่ย BIAS สูงสุด รองลงมาคือ แบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 3 ระดับ ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าแบบทดสอบผสมที่ให้คะแนน 5 ระดับ มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบที่ให้คะแนน 4 ระดับ และ 3 ระดับ และแบบทดสอบที่ให้คะแนน 4 ระดับ มีค่าเฉลี่ย BIAS สูงกว่า แบบทดสอบที่ให้คะแนน 3 ระดับ

2. RMSE ของพัฒนาการความสามารถ

ค่าเฉลี่ย RMSE ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันแตกต่างตามกลุ่มส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถ 1.00 0.80 และ 1.20 อย่างไม่มีนัยสำคัญที่ระดับ .05 โดยกลุ่มพัฒนาการที่มีค่าส่วนเบี่ยงเบนมาตรฐาน (SD) เป็น 0.80 และ 1.00 มีค่าเฉลี่ย BIAS ต่ำกว่ากลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 1.2

ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 เมื่อความยาวของแบบทดสอบผสมแตกต่างกัน 3 ขนาด ความยากของแบบทดสอบผสม 3 ระดับ และการให้คะแนนของแบบทดสอบผสม 3 ขนาด ซึ่งความแตกต่างของค่าเฉลี่ย RMSE ของพัฒนาการความสามารถเป็นผลมาจากการส่งผลร่วมกันระหว่างตัวแปร ความยาว ความยาก และการ

ให้คะแนนของแบบทดสอบผสม ผลการเปรียบเทียบค่าเฉลี่ย RMSE ของพัฒนาการความสามารถ ในแต่ละคู่ระดับของ 3 เงื่อนไข ปรากฏผลดังนี้

2.1 เมื่อความยากและการให้คะแนนคงที่ ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีความยาว 30:10 มีค่าสูงสุด รองลงมาคือ 15:5 และ 24:8 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าแบบทดสอบผสมที่มีความยาว 30:10 มี RMSE สูงกว่า แบบทดสอบผสมที่มีความยาว 24:8 และ แบบทดสอบผสมที่มีความยาว 15:5 มี RMSE สูงกว่าแบบทดสอบผสมที่มีความยาว 24:8 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบผสมที่มีความยาว 30:10 กับ15:5 มี RMSE แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

2.2 เมื่อความยาวและรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่มีค่าความยาก 0.00 มีค่าเฉลี่ย RMSE สูงสุด รองลงมาคือ แบบทดสอบผสมที่มีความยาก -0.50 และ 0.50 ตามลำดับ และพบว่าแบบทดสอบผสมที่มีความยาก -0.50 มีค่าเฉลี่ย RMSE สูงกว่า แบบทดสอบผสมที่มีความยาก 0.50 และ แบบทดสอบผสมที่มีความยาก 0.00 มีค่าเฉลี่ย RMSE สูงกว่า แบบทดสอบที่มีความยาก 0.50 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่มีความยาก 0.00 แบบทดสอบที่มีความยาก -0.50 มี RMSE แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

2.3 เมื่อแบบทดสอบผสมมีความยาวและความยากคงที่ ค่าเฉลี่ย RMSE ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันจากแบบทดสอบผสมที่มีการให้คะแนน 5 ระดับมีค่าเฉลี่ย RMSE สูงสุด รองลงมาคือ แบบทดสอบผสมที่ให้คะแนน 4 ระดับ และ 3 ระดับ ตามลำดับ ผลการเปรียบเทียบค่าเฉลี่ยรายคู่พบว่าแบบทดสอบผสมที่ให้คะแนน 5 ระดับ มีค่าเฉลี่ย RMSE สูงกว่า แบบทดสอบที่ให้คะแนน 4 ระดับ และ 3 ระดับ และแบบทดสอบที่ให้คะแนน 4 ระดับ มีค่าเฉลี่ย RMSE สูงกว่า แบบทดสอบที่ให้คะแนน 3 ระดับ

ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสรุปได้ดังตาราง 53

ตาราง 53 สรุปความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

ดัชนีความแม่นยำ	การเปรียบเทียบโดยรวม		การเปรียบเทียบระดับตัวแปร			
	แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05	แตกต่างอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05	ระหว่างกลุ่ม	ภายในกลุ่ม		
			กลุ่มตามค่า SD	ความยากและการให้คะแนนคงที่	ความยาวและการให้คะแนนคงที่	ความยาวและความยากคงที่
BIAS	ความยาว ความยาก และการให้คะแนน	กลุ่มพัฒนาการตามค่า SD	**SD 1.2** > SD1.0 > SD 0.8 (ในทิศทางลบ)	**ยาว 30:10** > ยาว 15:15 > ยาว 24:8 (ในทิศทางลบ)	**ยาก 0.00** > ยาก -0.50 > ยาก 0.50 (ในทิศทางลบ)	**5 ระดับ** > 4 ระดับ > 3 ระดับ (ในทิศทางลบ)
RMSE	ความยาว ความยาก และการให้คะแนน	กลุ่มพัฒนาการตามค่า SD	**SD 1.2** > SD1.0 > SD 0.8	**ยาว 30:10** > ยาว 15:15> ยาว 24:8	**ยาก 0.00** > ยาก -0.50 > ยาก 0.50	**5 ระดับ** > 4 ระดับ > 3 ระดับ

หมายเหตุ:

ยาว 30:10 ยาว 24:8 และ ยาว 15:5 หมายถึงแบบทดสอบผสมที่มีความยาว 30:10 24:8 และ 15:5 ตามลำดับ

ยาก -0.50 ยาก 0.00 ยาก 0.50 หมายถึงแบบทดสอบผสมที่ใช้ในการทดสอบครั้งที่ 1 มีความยากเฉลี่ยเป็น - 0.50 0.00 และ 0.50 ตามลำดับ

3 ระดับ 4 ระดับ และ 5 ระดับ หมายถึง คะแนนของข้อสอบแบบเขียนตอบในแบบทดสอบผสมมี 3 ระดับ 4 ระดับ และ 5 ระดับตามลำดับ

SD 0.8 SD 1.0 และ SD 1.2 หมายถึง ส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถในกลุ่มเป็น 0.8 1.0 และ 1.2 ตามลำดับ

บทที่ 5
สรุปผล อภิปรายผลและข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการศึกษาภายใต้สถานการณ์จำลอง (Simulation Data) เพื่อตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสำหรับคำนวณพัฒนาการความสามารถของผู้สอบจำนวน 1,000 คนที่ทดสอบซ้ำสองครั้งเมื่อผู้สอบมีระดับพัฒนาการความสามารถแตกต่างกัน และแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน การตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันพิจารณาจากความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน จุดมุ่งหมายและสมมติฐานการวิจัยมีดังนี้

ความมุ่งหมายของการวิจัย

การศึกษาวิจัยครั้งนี้มีวัตถุประสงค์หลักเพื่อตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน โดยใช้แบบทดสอบผสมที่มีความยาว ความยาก และการให้คะแนนแตกต่างกันศึกษาภายใต้สถานการณ์จำลอง ซึ่งมีวัตถุประสงค์เฉพาะดังนี้

1. เพื่อศึกษาและเปรียบเทียบความสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง เมื่อผู้สอบมีพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยาก และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน
2. เพื่อศึกษาและเปรียบเทียบความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกัน
3. เพื่อศึกษาและเปรียบเทียบความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกัน

สมมติฐานของการวิจัย

ในการศึกษาภายใต้สถานการณ์จำลองครั้งนี้ผู้วิจัยใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันโดยใช้โมเดลโลจิสติกแบบสามพารามิเตอร์ร่วมกับโมเดลเจเนอรอลไรส์พาร์เชียลเครดิตประมาณ

ค่าพารามิเตอร์ความสามารถของผู้สอบในการทดสอบครั้งที่ 1 และการทดสอบครั้งที่ 2 พร้อมกันภายใต้เงื่อนไขแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนแตกต่างกัน เพื่อตรวจสอบความแม่นยำของวิธีการประมาณพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถแตกต่างกัน โดยเปรียบเทียบค่าประมาณพารามิเตอร์ความสามารถ ($\hat{\theta}$) กับความสามารถจริง (θ) ที่กำหนดตามระดับพัฒนาการความสามารถที่ศึกษา แยกตามเงื่อนไขความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสม ซึ่งผู้วิจัยได้ตั้งสมมติฐานการวิจัยไว้ดังนี้

1. ค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความสัมพันธ์ระดับสูงกับค่าความสามารถจริงที่กำหนดในทุกระดับพัฒนาการความสามารถ โดย

1.1 แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีค่าความสัมพันธ์สูงกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ในทุกระดับความยาก และ ระดับการให้คะแนน

1.2 แบบทดสอบผสมที่มีระดับความยากของข้อสอบสอดคล้องกับระดับความสามารถผู้สอบจะมีค่าความสัมพันธ์สูงกว่าแบบทดสอบผสมที่ระดับความยากของข้อสอบไม่สอดคล้องกับระดับความสามารถผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนน

1.3 แบบทดสอบผสมที่มีระดับการให้คะแนนมากกว่าจะมีค่าความสัมพันธ์สูงกว่าแบบทดสอบผสมที่มีระดับการให้คะแนนน้อยกว่า ในทุกขนาดความยาวและระดับความยาก

2. ค่าประมาณพารามิเตอร์ความสามารถของผู้สอบที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) มีค่าเข้าใกล้ศูนย์ ทุกระดับพัฒนาการความสามารถของผู้สอบ โดย

2.1 แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ทุกระดับความยาก และ ระดับการให้คะแนน

2.2 แบบทดสอบผสมที่มีระดับความยากของข้อสอบสอดคล้องกับระดับความสามารถผู้สอบจะมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่ระดับความยากของข้อสอบไม่สอดคล้องกับระดับความสามารถผู้สอบ ในทุกขนาดความยาว และ ทุกระดับการให้คะแนน

2.3 แบบทดสอบผสมที่มีระดับการให้คะแนนมากกว่าจะมีค่าความลำเอียง (BIAS) และ ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่มีระดับการให้คะแนนน้อยกว่าในทุกขนาดความยาวและระดับความยากของแบบทดสอบผสม

3. เมื่อระดับพัฒนาการความสามารถของผู้สอบมีส่วนเบี่ยงเบนมาตรฐานเท่ากับหรือต่ำกว่าส่วนเบี่ยงเบนมาตรฐานของความสามารถครั้งแรก (1.0 หรือ 0.8) ค่าประมาณความสามารถจาก

การใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความแม่นยำมากกว่าเมื่อส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถของผู้สอบสูงกว่าส่วนเบี่ยงเบนมาตรฐานของความสามารถครั้งแรก (1.2) โดย

3.1 แบบทดสอบผสมที่มีขนาดยาวกว่าจะมีแม่นยำมากกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่า ในทุกระดับความยาก และ ระดับการให้คะแนน

3.2 แบบทดสอบผสมที่มีระดับความยากของแบบทดสอบสอดคล้องกับระดับความสามารถเริ่มต้นของผู้สอบ จะมีความแม่นยำมากกว่าแบบทดสอบผสมที่มีระดับความยากของแบบทดสอบไม่สอดคล้องกับระดับความสามารถเริ่มต้นของผู้สอบ ในทุกขนาดความยาว และ ระดับการให้คะแนน

3.3 แบบทดสอบผสมที่มีระดับการให้คะแนนสูงกว่า จะมีความแม่นยำมากกว่าแบบทดสอบผสมที่มีระดับการให้คะแนนต่ำกว่า ในทุกขนาดความยาวและความยากของแบบทดสอบผสม

การกำหนดข้อมูลจำลอง

ข้อมูลที่ศึกษาเป็นข้อมูลที่ได้จากการจำลองด้วยโปรแกรม WINGEN2 (Han. 2007) โดยเป็นแบบแผนการตอบข้อสอบของผู้สอบจำนวน 1,000 คนที่ได้ทดสอบซ้ำสองครั้งด้วยแบบทดสอบผสมที่ประกอบด้วยข้อสอบแบบเลือกตอบที่ตรวจให้คะแนนสองค่าและข้อสอบแบบเขียนตอบที่ตรวจให้คะแนนหลายค่า การจำลองแบบแผนการตอบข้อสอบใช้โมเดลโลจิสติกแบบสามพารามิเตอร์ในการจำลองผลการตอบข้อสอบแบบเลือกตอบ และใช้โมเดลเจเนอรอลไรส์พาร์เชียลเครดิตจำลองผลการตอบข้อสอบแบบเขียนตอบ มีรายละเอียดการกำหนดค่าพารามิเตอร์ของผู้สอบและข้อสอบและการจำลองแบบแผนการตอบ ตามเงื่อนไขสำหรับการศึกษาครั้งนี้ ดังนี้

1) การกำหนดระดับพัฒนาการความสามารถของผู้สอบ

การศึกษาครั้งนี้กำหนดระดับความสามารถของผู้สอบ 1,000 คนในการทดสอบครั้งที่ 1 และครั้งที่ 2 ดังนี้

การทดสอบครั้งที่ 1 กำหนดให้ผู้สอบจำนวน 1,000 คน มีการแจกแจงความสามารถ (θ_1) แบบปกติ โดยที่ค่าเฉลี่ยความสามารถเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1

การทดสอบครั้งที่ 2 กำหนดให้ผู้สอบจำนวน 1,000 คน มีความสามารถ (θ_2) เฉลี่ยเพิ่มขึ้น 0.1 0.5 และ 1 ตามลำดับ และมีค่าส่วนเบี่ยงเบนมาตรฐานเป็น 1.0 0.8 และ 1.2 ตามลำดับ นำความสามารถครั้งที่สอง (θ_2) เขียนให้อยู่ในรูปเชิงเส้นของค่าความสามารถครั้งแรก (θ_1) ตามกลุ่มผู้สอบที่ใช้ศึกษาซึ่งมีลักษณะความสามารถแตกต่างกันมีทั้งสิ้น 9 กลุ่ม ดังนี้

รูปแบบที่ 1: $\theta_2 = \theta_1 + 0.1$ รูปแบบที่ 2: $\theta_2 = \theta_1 + 0.5$ รูปแบบที่ 3: $\theta_2 = \theta_1 + 1.0$

รูปแบบที่ 4: $\theta_2 = 0.8\theta_1 + 0.1$ รูปแบบที่ 5: $\theta_2 = 0.8\theta_1 + 0.5$ รูปแบบที่ 6: $\theta_2 = 0.8\theta_1 + 1.0$
รูปแบบที่ 7: $\theta_2 = 1.2\theta_1 + 0.1$ รูปแบบที่ 8: $\theta_2 = 1.2\theta_1 + 0.5$ รูปแบบที่ 9: $\theta_2 = 1.2\theta_1 + 1.0$

2) การกำหนดค่าพารามิเตอร์ของข้อสอบในแบบทดสอบผสม

ในการศึกษาครั้งนี้เป็นการศึกษาพัฒนาการความสามารถของผู้สอบจำนวน 1,000 คนที่ทำการทดสอบซ้ำสองครั้งด้วยแบบทดสอบผสมซึ่งประกอบด้วยข้อสอบแบบเลือกตอบที่ให้คะแนนสองค่าและข้อสอบแบบเขียนตอบที่ให้คะแนนหลายค่า

การวิจัยครั้งนี้ศึกษาเฉพาะพัฒนาการความสามารถที่เกิดจากการทดสอบซ้ำสองครั้งโดยการทดสอบทั้งสองครั้งใช้แบบทดสอบผสมมีความยาวและระดับการให้คะแนนเท่ากัน แต่มีความยากของแบบทดสอบที่ใช้ในการทดสอบครั้งที่ 2 สูงกว่าความยากของแบบทดสอบในการทดสอบครั้งที่ 1 ค่าพารามิเตอร์ความยากของข้อสอบทั้งสองรูปแบบในแบบทดสอบผสมทั้งสองฉบับมีการแจกแจงแบบปกติ และแบบทดสอบผสมทั้งสองฉบับมีข้อสอบร่วมที่ประกอบด้วยข้อสอบแบบเลือกตอบและข้อสอบแบบเขียนตอบประมาณ 33% ของจำนวนข้อสอบทั้งฉบับ ในการศึกษาครั้งนี้กำหนดเงื่อนไขของแบบทดสอบผสมสำหรับศึกษาความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน 3 เงื่อนไข ได้แก่ ความยาว 3 ขนาด ความยาก 3 ระดับ และการให้คะแนน 3 รูปแบบ

ความยาวของแบบทดสอบผสมกำหนดตามจำนวนข้อสอบแบบเลือกตอบต่อจำนวนข้อสอบแบบเขียนตอบเป็น 30:10 24:8 และ 15:5 ส่วนความยากของแบบทดสอบผสมเป็นความยากเฉลี่ยของแบบทดสอบทั้งฉบับโดยแบ่งเป็น 3 ระดับ ได้แก่ ค่าเฉลี่ยความยากของข้อสอบแบบทดสอบฉบับที่ 1 ต่ำกว่าระดับความสามารถเฉลี่ยของผู้สอบในการทดสอบครั้งที่ 1 (-0.5) ค่าเฉลี่ยความยากของข้อสอบแบบทดสอบฉบับที่ 1 เท่ากับระดับความสามารถเฉลี่ยของผู้สอบในการทดสอบครั้งที่ 1 (0.0) และค่าเฉลี่ยความยากของข้อสอบแบบทดสอบฉบับที่ 1 สูงกว่าระดับความสามารถเฉลี่ยของผู้สอบในการทดสอบครั้งที่ 1 (0.5) ส่วนการให้คะแนน แบ่งเป็น 3 รูปแบบตามระดับคะแนนของข้อสอบแบบเขียนตอบได้แก่ 3 ระดับ (0/1/2) 4 ระดับ (0/1/2/3) และ 5 ระดับ (0/1/2/3/4)

การวิเคราะห์ข้อมูลในการวิจัย

ผู้วิจัยตรวจสอบความถูกต้องของข้อมูลแบบแผนการตอบข้อสอบของผู้สอบ 1,000 คนที่ทำแบบทดสอบผสมฉบับที่ 1 และฉบับที่ 2 ที่จำลองได้ แล้วจึงรวมผลการตอบทั้งสองครั้งเป็นแบบแผนการตอบข้อสอบชุดเดียวกัน จากนั้นจึงประมาณค่าพารามิเตอร์ความสามารถด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกันโดยใช้โปรแกรม MULTILOG การประมาณค่ากระทำครั้งละ 1 เงื่อนไข และทำซ้ำเงื่อนไขละ 50 รอบ รวมเงื่อนไขทั้งหมดที่ศึกษาจำนวน 243 เงื่อนไข เมื่อได้ค่าประมาณพารามิเตอร์ความสามารถในการทำข้อสอบแต่ละครั้งจากวิธีการประมาณค่าพารามิเตอร์พร้อมกันแล้ว

จึงคำนวณพัฒนาการความสามารถจากผลต่างระหว่างค่าประมาณความสามารถที่ได้จากการสอบครั้งที่ 2 กับพารามิเตอร์ความสามารถจากการทดสอบครั้งที่ 1 ($\hat{\theta}_2 - \hat{\theta}_1$) ทุกเงื่อนไขศึกษา เพื่อนำไปเปรียบเทียบกับพัฒนาการความสามารถที่ได้กับพัฒนาการความสามารถจริง ($\theta_2 - \theta_1$)

การศึกษาและเปรียบเทียบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน แบ่งเป็น 4 ตอน คือ 1) การวิเคราะห์ค่าสถิติพื้นฐานของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน 2) การศึกษาความสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถกับความสามารถจริงในการทดสอบทั้งสองครั้ง 3)การศึกษาความแม่นยำของค่าประมาณพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน และ 4) การศึกษาความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน

สรุปผลการวิจัย

การวิจัยครั้งนี้สรุปผลตามวัตถุประสงค์การวิจัยได้ดังนี้

1. ทุกกลุ่มระดับพัฒนาการความสามารถ ค่าประมาณพารามิเตอร์ความสามารถจากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกัน กับความสามารถจริง มีความสัมพันธ์ทางลบในระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกเงื่อนไขของความยาว ความยาก และการให้คะแนน โดยที่

1.1 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถ ($\hat{\theta}_1$) จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง (θ_1) ทุกรูปแบบพัฒนาการ M1 ถึง M9 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถ ($\hat{\theta}_1$) จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง (θ_1) เท่ากันภายใต้เงื่อนไขแบบทดสอบผสมมีความยาว ความยาก และการให้คะแนนเท่ากัน โดยพบว่า

1.1.1 เมื่อแบบทดสอบผสมมีความยาก และการให้คะแนนคงที่ แบบทดสอบผสมที่มีขนาดยาวกว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถในทดสอบครั้งที่ 1 ที่ได้จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง ต่ำกว่าแบบทดสอบที่มีขนาดสั้นกว่า ในทิศทางลบอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

1.1.2 เมื่อแบบทดสอบผสมมีความยาว และรูปแบบการให้คะแนนคงที่ พบว่าแบบทดสอบผสมที่มีระดับความยากของแบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 ต่ำกว่าความสามารถในการทดสอบครั้งที่ 1 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถในการทดสอบครั้งที่ 1 ที่ได้จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง สูงกว่า แบบทดสอบผสมที่มีระดับความยากสูงกว่าหรือเท่ากับแบบทดสอบผสมที่มี

ระดับความยากของแบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 ในทิศทางลบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

1.1.3 เมื่อแบบทดสอบผสมมีความยาว และรูปแบบการให้คะแนนคงที่ พบว่า แบบทดสอบผสมที่ให้ระดับคะแนนน้อยกว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถในการทดสอบครั้งที่ 1 ที่ได้จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง สูงกว่าแบบทดสอบผสมที่ให้ระดับคะแนนสูงกว่า ในทิศทางลบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

1.2 ความสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถ ($\hat{\theta}_2$) ในการทดสอบครั้งที่ 2 ที่ได้จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง (θ_2) มีความสัมพันธ์ทางลบในระดับสูง ทุกเงื่อนไขที่ศึกษา โดยที่ทุกรูปแบบพัฒนาการ M1 ถึง M9 มีค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณพารามิเตอร์ความสามารถจากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริงที่ได้จากการทดสอบครั้งที่ 2 ไม่เท่ากันเมื่อมีความยาว ความยาก และการให้คะแนนเท่ากัน โดยพบว่า

1.2.1 เมื่อแบบทดสอบผสมมีความยาก และรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่มีขนาดยาวกว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณพารามิเตอร์ความสามารถ ($\hat{\theta}_2$) ในการทดสอบครั้งที่ 2 ที่ได้จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริงสูงกว่าแบบทดสอบผสมที่มีขนาดสั้นกว่าในทิศทางลบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

1.2.2 เมื่อแบบทดสอบผสมมีความยาว และรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่ใช้ในการทดสอบครั้งที่ 1 มีความยากต่ำกว่าความสามารถในการทดสอบครั้งที่ 1 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถ ($\hat{\theta}_2$) ในการทดสอบครั้งที่ 2 ที่ได้จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง (θ_2) สูงกว่าแบบทดสอบผสมที่มีความยากสูงกว่าหรือเท่ากับความสามารถในการทดสอบครั้งที่ 1 ในทิศทางลบอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

1.2.3 เมื่อแบบทดสอบผสมมีความยาว และความยากคงที่ แบบทดสอบผสมที่ให้ระดับคะแนนน้อยกว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถ ($\hat{\theta}_2$) ในการทดสอบครั้งที่ 2 ที่ได้จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง (θ_2) สูงกว่า แบบทดสอบผสมที่ให้ระดับคะแนนมากกว่า ในทิศทางลบอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

2. ความลำเอียง (BIAS) และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) ของค่าประมาณความสามารถได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกัน

2.1 พัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันในแต่ละกลุ่มระดับพัฒนาการความสามารถมีความลำเอียง (BIAS) ทางลบ โดยที่

2.1.1 เมื่อแบบทดสอบผสมมีความยาก และรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่มีความยาว 24:8 มีความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ต่ำกว่า แบบทดสอบผสมที่มีความยาว 15:5 และ 30:10 อย่างมีนัยสำคัญที่ระดับ .05

2.1.2 เมื่อแบบทดสอบผสมมีความยาว และรูปแบบการให้คะแนนคงที่ เมื่อระดับความยากของแบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 เท่ากับ 0.50 มีความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันต่ำกว่าแบบทดสอบที่มีความยาก 0.00 และ -0.50 อย่างมีนัยสำคัญที่ระดับ .05

2.1.3 เมื่อแบบทดสอบผสมมีความยาว และความยากคงที่ เมื่อแบบทดสอบผสมให้คะแนน 5 ระดับ มีความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันสูงกว่าแบบทดสอบผสมให้คะแนน 4 ระดับ และ 3 ระดับอย่างมีนัยสำคัญที่ระดับ .05

2.2 พัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันในแต่ละกลุ่มระดับพัฒนาการความสามารถมีค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) แตกต่างกัน โดยที่

2.2.1 เมื่อแบบทดสอบผสมมีความยากและรูปแบบการให้คะแนนคงที่ แบบทดสอบที่มีความยาว 30:10 มีค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) สูงกว่าแบบทดสอบที่มีความยาว 15:5 และ 24:8 อย่างมีนัยสำคัญที่ระดับ .05

2.2.2 เมื่อแบบทดสอบผสมมีความยาว และรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 มีระดับความยากเท่ากับความสามารถในการสอบครั้งที่ 1 (0.00) ค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีค่าสูงกว่าแบบทดสอบที่มีระดับความยาก -0.50 และ 0.50 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2.2.3 เมื่อแบบทดสอบผสมมีความยาว และความยากคงที่ แบบทดสอบผสมที่ให้คะแนน 3 ระดับมีค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันต่ำกว่า แบบทดสอบผสมที่ให้คะแนน 4 และ 5 ระดับ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3. ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันแตกต่างกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกัน ดังนี้

3.1 กลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนเป็น 0.80 และ 1.00 มีความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันต่ำกว่ากลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 ในทุกระดับความยาก ความยาวและรูปแบบการให้คะแนนของแบบทดสอบผสม โดยที่

3.1.1 เมื่อแบบทดสอบผสมมีความยาก และรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่มีความยาว 30:10 มีความลำเอียง (BIAS) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน สูงกว่า แบบทดสอบผสมที่มีความยาว 15:5 และ 24:8 อย่างมีนัยสำคัญที่ระดับ .05

3.1.2 เมื่อแบบทดสอบผสมมีความยาว และรูปแบบการให้คะแนนคงที่ เมื่อแบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 มีระดับความยากเท่ากับความสามารถในการสอบครั้งที่ 1 (0.00) มีความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน สูงกว่า แบบทดสอบผสมมีความยากเป็น -0.50 และ 0.50 อย่างมีนัยสำคัญที่ระดับ .05

3.1.3 เมื่อแบบทดสอบผสมมีความยาว และความยากคงที่ แบบทดสอบผสมที่ให้คะแนน 3 ระดับ มีความลำเอียง (BIAS) ของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ต่ำกว่าแบบทดสอบผสมที่ให้คะแนน 4 ระดับและ 5 ระดับอย่างมีนัยสำคัญที่ระดับ .05

3.2 กลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนเป็น 0.80 และ 1.00 มีค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ต่ำกว่า กลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนเป็น 1.2 อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 ในทุกระดับความยาว ความยาก และรูปแบบการให้คะแนนของแบบทดสอบผสม โดยที่

3.2.1 เมื่อแบบทดสอบผสมมีความยาก และการให้คะแนนคงที่ แบบทดสอบผสมที่มีความยาว 30:10 มีค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน สูงกว่า แบบทดสอบผสมที่มีความยาว 15:5 และ 24:8 อย่างมีนัยสำคัญที่ระดับ .05

3.2.2 เมื่อแบบทดสอบผสมมีความยาว และการให้คะแนนคงที่ แบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 มีความยากเท่ากับความสามารถในการสอบครั้งที่ 1 (0.00) มีค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสูงกว่าแบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 มีความยากไม่เท่ากับความสามารถในการสอบครั้งที่ 1 (-0.50 และ 0.50) อย่างมีนัยสำคัญที่ระดับ .05

3.2.3 เมื่อแบบทดสอบผสมมีความยาว และความยากคงที่ แบบทดสอบผสมที่ให้คะแนน 3 ระดับ มีค่ารากที่สองของค่าเฉลี่ยของความคลาดเคลื่อนกำลังสอง (RMSE) ต่ำกว่าแบบทดสอบผสมที่ให้คะแนน 4 ระดับและ 5 ระดับ อย่างมีนัยสำคัญที่ระดับ .05

สามารถสรุปผลโดยรวมตามสมมติฐานการวิจัยได้ดังตาราง 54

สรุปผลโดยรวม

จากตาราง 54 สรุปได้ว่าค่าประมาณพารามิเตอร์ความสามารถจากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง มีความสัมพันธ์ทางลบในระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกเงื่อนไขของความยาว ความยาก และรูปแบบการให้คะแนน และในทุกระดับพัฒนาการความสามารถทั้งหมดที่ศึกษา ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับพัฒนาการความสามารถที่มีส่วนเบี่ยงเบนเป็น 0.80 และ 1.00 สูงกว่ากลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนเป็น 1.2 ในทุกระดับความยาว ความยาก และรูปแบบการให้คะแนนของแบบทดสอบผสม อย่างไม่มีนัยสำคัญที่ระดับ .05 โดยที่

เมื่อแบบทดสอบผสมมีความยาก และรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมขนาด 24:8 มีความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันค่าสูงกว่าแบบทดสอบผสมขนาด 15:5 และ ขนาด 30:10

เมื่อแบบทดสอบผสมมีความยาว และรูปแบบการให้คะแนนคงที่ แบบทดสอบผสมที่มีระดับความยากของใช้ในการสอบครั้งที่ 1 เท่ากับความสามารถในการสอบครั้งที่ 1 (0.00) มีความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันต่ำกว่า แบบทดสอบที่มีความยากไม่เท่ากับความสามารถในการสอบครั้งที่ 1 (-0.50 และ 0.50) อย่างมีนัยสำคัญที่ระดับ .05

เมื่อแบบทดสอบผสมมีความยาว และความยากคงที่ แบบทดสอบผสมที่ให้คะแนน 3 ระดับ มีความแม่นยำของของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน สูงกว่ากรณีให้คะแนน 4 และ 5 ระดับ อย่างมีนัยสำคัญที่ระดับ .05

ตาราง 54 สรุปผลการตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีพัฒนาการความสามารถแตกต่างกัน ภายใต้เงื่อนไข ความยาว ความยาก และการให้คะแนนของแบบทดสอบผสมแตกต่างกัน โดยพิจารณาในภาพรวมและการเปรียบเทียบภายในกลุ่มและระหว่างกลุ่ม

ผล	สัมประสิทธิ์สหสัมพันธ์เพียร์สัน		ความแม่นยำของค่าประมาณพารามิเตอร์		ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน	
	ทดสอบครั้งที่ 1	ทดสอบครั้งที่ 2	BIAS	RMSE	BIAS	RMSE
โดยรวม	-มีค่าสูง ทิศทางลบ -มีค่าเท่ากันทุกกลุ่มพัฒนาการ	-มีค่าสูง ทิศทางลบ -มีค่าต่างกันตามกลุ่มพัฒนาการ	มีค่าเป็นลบเข้าใกล้ศูนย์	เข้าใกล้ศูนย์	มีค่าเป็นลบเข้าใกล้ศูนย์	เข้าใกล้ศูนย์
ภายในกลุ่ม						
กรณี 1 ความยาก และการให้คะแนนคงที่	ยาว15:15 > ยาว 24:8 > ยาว 30:10	ยาว 30:10 > ยาว 24:8 > ยาว15:15	ยาว30:10 > ยาว 15:15 > ยาว 24:8	ยาว30:10 > ยาว 15:15 > ยาว 24:8	ยาว30:10 > ยาว 15:15 > ยาว 24:8	ยาว30:10 > ยาว 15:15 > ยาว 24:8
กรณี 2 ความยาว และการให้คะแนนคงที่	ยาก -0.50 > ยาก 0.00 > ยาก 0.50	ยาก -0.50 > ยาก 0.00 > ยาก 0.50	ยาก 0.00 > ยาก -0.50 > ยาก 0.50	ยาก 0.00 > ยาก -0.50 > ยาก 0.50	ยาก 0.00 > ยาก -0.50 > ยาก 0.50	ยาก 0.00 > ยาก -0.50 > ยาก 0.50
กรณี 3 ความยาว และความยากคงที่	3 ระดับ > 4 ระดับ > 5 ระดับ	3 ระดับ > 4 ระดับ > 5 ระดับ	5 ระดับ > 4 ระดับ > 3 ระดับ	4 ระดับ > 5 ระดับ > 3 ระดับ	5 ระดับ > 4 ระดับ > 3 ระดับ	5 ระดับ > 4 ระดับ > 3 ระดับ
ระหว่างกลุ่ม กลุ่ม SD	-	-	-	-	SD 0.8 > SD1.0 > SD 1.2	SD 0.8 > SD1.0 > SD 1.2

หมายเหตุ: ____ หมายถึง มีค่าสัมบูรณ์ของสัมประสิทธิ์สหสัมพันธ์ หรือ BIAS หรือ RMSE มากที่สุด

ยาว 30:10 ยาว 24:8 และ ยาว 15:5 หมายถึงแบบทดสอบผสมที่มีความยาว 30:10 24:8 และ 15:5 ตามลำดับ

ยาก -0.50 ยาก 0.00 ยาก 0.50 หมายถึงแบบทดสอบผสมที่ใช้ในการทดสอบครั้งที่ 1 มีความยากเฉลี่ยเป็น - 0.50 0.00 และ 0.50 ตามลำดับ

3 ระดับ 4 ระดับ และ 5 ระดับ หมายถึง คะแนนของข้อสอบแบบเขียนตอบในแบบทดสอบผสมมี 3 ระดับ 4 ระดับ และ 5 ระดับ ตามลำดับ

SD 0.8 SD 1.0 SD 1.2 หมายถึง ส่วนเบี่ยงเบนมาตรฐานของระดับพัฒนาการความสามารถในกลุ่มเป็น 0.8 1.0 และ 1.2 ตามลำดับ

อภิปรายผลการวิจัย

การอภิปรายผลในการวิจัยนี้ ผู้วิจัยเสนอ 2 ประเด็นหลักคือ การอภิปรายผลตามสมมติฐานการวิจัย และการอภิปรายผลจากการจำลองข้อมูลและโปรแกรมที่ใช้ในการทำวิจัย ดังรายละเอียดต่อไปนี้

1. การอภิปรายผลตามสมมติฐานการวิจัย

เนื่องจากผลการวิจัยเกี่ยวกับความแม่นยำของวิธีการประมาณค่าพารามิเตอร์ และความแม่นยำของพัฒนาการความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์ตามสมมติฐานข้อ 2 และ 3 ได้ผลในทำนองเดียวกัน ผู้วิจัยจึงแบ่งประเด็นการอภิปรายออกเป็น 2 ประเด็นคือ การอภิปรายผลตามสมมติฐานการวิจัย ข้อ 1 และ การอภิปรายผลตามสมมติฐานการวิจัย ข้อ 2 และ ข้อ3 ดังนี้

1.1 การอภิปรายผลตามสมมติฐานการวิจัย ข้อ 1 ที่ตั้งไว้ว่า *"ค่าประมาณความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความสัมพันธ์ระดับสูงกับค่าความสามารถจริงที่กำหนดในทุกระดับพัฒนาการความสามารถ"* ผลการวิจัยพบว่าทุกกลุ่มระดับพัฒนาการความสามารถ ค่าประมาณพารามิเตอร์ความสามารถในการทำแบบทดสอบทั้งสองครั้งที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริงมีความสัมพันธ์ทางลบในระดับสูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกเงื่อนไข โดยที่ทุกรูปแบบพัฒนาการ M1 M2 M3 M4 M5 M6 M7 M8 และ M9 มีค่าสัมประสิทธิ์สหสัมพันธ์ของค่าประมาณพารามิเตอร์ความสามารถในการทดสอบครั้งที่ 1 ที่ได้จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริงเท่ากันเมื่อมีความยาว ความยาก และการให้คะแนนเท่ากัน แต่สำหรับความสัมพันธ์ระหว่างค่าประมาณพารามิเตอร์ความสามารถในการทดสอบครั้งที่ 2 ที่ได้จากการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันกับความสามารถจริง มีความแตกต่างกันตามกลุ่มรูปแบบพัฒนาการ M1 M2 M3 M4 M5 M6 M7 M8 และ M9

การที่ค่าประมาณความสามารถของผู้สอบที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีค่าเป็นลบแสดงว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันให้ค่าประมาณความสามารถที่ต่ำกว่าความเป็นจริง (Cao. 2008) แม้ว่าค่าความสัมพันธ์ระหว่างพารามิเตอร์ความสามารถที่ได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันกับค่าความสามารถจริงมีค่าเป็นลบแต่ไม่ส่งผลต่อการศึกษาครั้งนี้ เนื่องจากเป็นการศึกษาพัฒนาการความสามารถของผู้สอบโดยคำนวณจากผลต่างของค่าประมาณความสามารถในการทดสอบครั้ง 2 และครั้งที่ 1 ซึ่งพิจารณาแล้วพบว่าพัฒนาการความสามารถที่ได้มีค่าคงที่และมีลักษณะการแจกแจงที่สอดคล้องกับพัฒนาการความสามารถจริง อีกทั้งเป็นค่าประมาณความสามารถที่ถูกปรับเทียบให้อยู่บนมาตรวัดเดียวกัน อย่างไรก็ตามค่าประมาณความสามารถที่ได้

จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันอาจจะเกิดมีความบิดเบือนจากค่าความสามารถจริงได้ เนื่องจากความแปรปรวนของกลุ่มผู้สอบในการทดสอบสองครั้งไม่เท่ากัน แต่โปรแกรม MULTILOG นั้นมีข้อจำกัดเกี่ยวกับความแปรปรวนของผู้สอบที่ต้องการให้ความแปรปรวนในแต่ละกลุ่มต้องมีค่าเท่ากัน ดังนั้นการเลือกใช้โปรแกรม MULTILOG ในการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบสองกลุ่มที่มีความสามารถแตกต่างกันจึงอาจส่งผลต่อค่าประมาณความสามารถได้ การศึกษาครั้งนี้ผู้วิจัยกำหนดให้ความสามารถผู้สอบการทดสอบครั้งที่ 1 มีค่าเฉลี่ยเป็น 0 และค่าส่วนเบี่ยงเบนมาตรฐานเป็น 1 ส่วนการทดสอบครั้งที่ 2 กำหนดให้ผู้สอบมีค่าเฉลี่ยเพิ่มขึ้นแตกต่างกันและมีค่าส่วนเบี่ยงเบนมาตรฐานที่แตกต่างกัน แต่โปรแกรม MULTILOG ประมาณค่าพารามิเตอร์โดยควบคุมให้ค่าความแปรปรวนของผู้สอบในการทดสอบครั้งที่สองเป็น 1 เท่ากับค่าส่วนเบี่ยงเบนมาตรฐานในการทดสอบครั้งที่ 1 จึงอาจทำให้ค่าประมาณพารามิเตอร์ความสามารถของผู้สอบที่มีความสามารถระดับต่ำที่มีค่าน้อยกว่า 0 จึงถูกยกระดับให้ค่าพารามิเตอร์สูงขึ้นจากเดิมในขณะที่ผู้สอบกลุ่มสูงจะถูกกดระดับความสามารถให้ต่ำลงเพื่อให้ค่าเฉลี่ยความสามารถมีค่าเข้าใกล้ 0 ซึ่งเรียกว่าเป็นความบิดเบือนของมาตรวัด เปค และ ยัง (Peak; & Young, 2005) ได้เสนอวิธีการลดความบิดเบือนของมาตรวัดโดยให้ใช้วิธีการกำหนดพารามิเตอร์ข้อสอบร่วมให้มีค่าคงที่ แต่วิธีการกำหนดค่าพารามิเตอร์ข้อสอบร่วมให้คงที่ต้องใช้ร่วมกับวิธีการประมาณค่าพารามิเตอร์แยกกัน แม้ว่างานวิจัยครั้งนี้เป็นการศึกษาโดยวิธีการประมาณค่าพารามิเตอร์พร้อมกันโดยการจำลองข้อมูลซึ่งทราบค่าพารามิเตอร์ของข้อสอบแต่ผู้วิจัยไม่ได้กำหนดให้ค่าพารามิเตอร์ข้อสอบร่วมให้คงที่ เนื่องจากผู้วิจัยคำนึงถึงการประมาณค่าพารามิเตอร์ในสถานการณ์จริง ที่ไม่สามารถกำหนดพารามิเตอร์จริงได้ก่อนทำการประมาณค่าพารามิเตอร์

แม้ว่าผลการวิจัยจะสอดคล้องกับสมมติฐานที่ตั้งไว้บางส่วนแต่ข้อค้นพบที่ได้สอดคล้องกับคำแนะนำเกี่ยวกับการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกันผลการวิจัยที่ผ่านมา (Béguin; Hanson & Glas. 2000; Kim & Kolen. 2006; Kim & Lee. 2006; Li; Lissitz & Yang. 1999; Tate. 2000) ที่พบว่าเมื่อกลุ่มผู้สอบสองกลุ่มมีความสามารถไม่เท่าเทียมกันวิธีการประมาณค่าพารามิเตอร์พร้อมกันจะมีความคลาดเคลื่อนมากกว่าวิธีอื่น แต่วิธีการประมาณค่า พารามิเตอร์พร้อมกันเป็นการประมาณค่าพารามิเตอร์จากผลการทำแบบทดสอบผสมทั้งสองฉบับพร้อมกันโดยใช้การประมวลผลโดยคอมพิวเตอร์เพียง 1 ครั้ง จึงสามารถยืนยันได้ว่าค่าประมาณพารามิเตอร์ความสามารถที่ได้อยู่บนมาตรวัดเดียวกัน

นอกจากความเท่าเทียมกันของพารามิเตอร์ความสามารถในแต่ละกลุ่มผู้สอบจะมีผลต่อความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันแล้วข้อมูลจำลองที่ศึกษานั้นจะต้องสอดคล้องกับโมเดล (Kim; &Cohen. 1998, 2002; Hanson; & Beguin. 2002) จึงจะทำให้วิธีการประมาณค่าพารามิเตอร์พร้อมกันเกิดความถูกต้องแม่นยำ หากพบว่ากลุ่มผู้สอบมีระดับความสามารถ

ไม่เท่าเทียมกันและความสามารถเป็นแบบพหุมิติแต่มีความสัมพันธ์กันสูง การประมาณค่าพารามิเตอร์พร้อมกันอาจจะมีความคลาดเคลื่อนสูงกว่าการปรับเทียบโดยวิธีอื่น (Béguin; Hanson &Glas. 2000) ดังนั้นสิ่งเหล่านี้อาจเป็นสาเหตุที่ส่งผลร่วมกันต่อผลการประมาณค่าพารามิเตอร์ความสามารถในการวิจัยครั้งนี้

1.2 การอภิปรายผลตามสมมติฐานการวิจัย ข้อ 2 และ ข้อ 3 โดยสมมติฐานข้อ 2 ตั้งไว้ว่า *"ความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของค่าประมาณความสามารถกำลังสองได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว การความยาก และ การให้คะแนน ของแบบทดสอบผสมแตกต่างกัน"* และ สมมติฐานข้อ 3 ตั้งไว้ว่า *"ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อผู้สอบมีระดับพัฒนาการความสามารถเพิ่มขึ้นแตกต่างกัน ภายใต้เงื่อนไข ความยาว การความยาก และ การให้คะแนนของแบบทดสอบผสมแตกต่างกัน"*

ผลการวิจัยสอดคล้องกับสมมติฐานทั้ง 2 ข้อ โดยพบว่าเมื่อตรวจสอบความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันด้วยความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนของค่าประมาณความสามารถกำลังสองได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ได้ข้อสรุปในทำนองเดียวกันว่าเมื่อระดับพัฒนาการความสามารถที่มีส่วนเบี่ยงเบนมาตรฐานเป็น 0.80 และ 1.00 วิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความแม่นยำของสูงกว่ากลุ่มพัฒนาการที่มีส่วนเบี่ยงเบนเป็น 1.2 ในทุกระดับความยาว ความยาก และรูปแบบการให้คะแนนของแบบทดสอบผสมอย่างไม่มีนัยสำคัญที่ระดับ .05 ในขณะที่การเปรียบเทียบความแม่นยำในแต่ละเงื่อนไขพบว่าในแต่ละระดับความยาว ความยาก และการให้คะแนน กลุ่มระดับพัฒนาการที่มีความสามารถเพิ่มขึ้น 0.1 (M1 M4 และM7) มีแนวโน้มให้ค่าความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันสูงกว่ากลุ่มระดับพัฒนาการที่มีความสามารถเพิ่มขึ้น 0.5 (M2 M5 และM8) และ 1.0 (M3 M6 และM9) ซึ่งสอดคล้องกับผลการศึกษาของเปค และ ยัง (Paek; & Young. 2005) ที่พบว่าค่า MAE และRMSE มีค่าสูงขึ้นเมื่อค่าสัมบูรณ์ของพัฒนาการเฉลี่ยมีค่ามากขึ้นจนทำให้กราฟเป็นรูปพาราโบลาหงาย และมีค่าสูงสุดเมื่อส่วนเบี่ยงเบนมาตรฐานเป็น 1.2 แต่หากพัฒนาการความสามารถเพิ่มขึ้นหรือลดลง 0.1 หน่วย ค่า MAE และ RMSE จะมีค่าน้อยกว่าเมื่อผู้สอบมีพัฒนาการเฉลี่ยเพิ่มขึ้นหรือลดลงมากกว่า 0.1 หน่วย และเมื่อพัฒนาการความสามารถมีค่าส่วนเบี่ยงเบนมาตรฐานสูงกว่าส่วนเบี่ยงมาตรฐานของความสามารถเริ่มต้น (1.2) จะมีค่า MAE และ RMSE สูงกว่า เมื่อพัฒนาการความสามารถมีค่าส่วนเบี่ยงเบนมาตรฐานน้อยกว่าหรือเท่ากับส่วนเบี่ยงมาตรฐานของความสามารถเริ่มต้น (0.8 และ 1) แต่อย่างไรก็ตามเมื่อพิจารณาสมมติฐานย่อยซึ่งศึกษาในแต่ละเงื่อนไขกลับพบว่าผลการศึกษาไม่สอดคล้องตามสมมติฐานการวิจัย ดังรายละเอียดต่อไปนี้

1.2.1 เมื่อแบบทดสอบผสมมีความยาก และรูปแบบการให้คะแนนคงที่ ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบขนาด 24:8 และ 15:5 มีค่าสูงกว่า ขนาด 30:10 ซึ่งพิจารณาแล้วพบว่าไม่สอดคล้องกับจากสมมติฐานย่อยที่ตั้งไว้ว่า *แบบทดสอบขนาด 30:10 น่าจะมีความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันค่าสูงกว่า เมื่อแบบทดสอบขนาด 24:8 และ 15:5* แต่ผลที่ได้นี้กลับสอดคล้องกับผลการศึกษาของบาสตารี (Bastari. 2000) และ อัญชลี ศรีกลชาญ (2552) ที่พบว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันในกรณีกลุ่มผู้สอบมีความสามารถไม่เท่าเทียมกันจะมีความคลาดเคลื่อนลดลงเมื่อข้อสอบร่วมประกอบด้วยข้อสอบทั้ง 2 ชนิด และมีข้อสอบแบบตรวจให้คะแนนสองค่าจำนวนลดลง ซึ่งการศึกษาครั้งนี้ผู้วิจัยกำหนดจำนวนข้อสอบร่วมของแบบทดสอบ 30:10 24:8 และ 15:5 ไว้ที่ 10:4 8:3 และ 5:2 ตามลำดับ ซึ่งหากพิจารณาตามจำนวนและลักษณะข้อสอบร่วมดังกล่าวจึงอาจจะเป็นผลที่ทำให้แบบทดสอบขนาด 30:10 มีความแม่นยำต่ำกว่าแบบทดสอบขนาด 24:8 และ 15:5 เนื่องจากจำนวนข้อสอบแบบให้คะแนนสองค่าในข้อสอบร่วมของแต่ละขนาดนั่นเอง อย่างไรก็ตามไม่สามารถสรุปได้อย่างชัดเจน เนื่องจากหากพิจารณาตามสัดส่วนจำนวนข้อแล้ว จำนวนข้อสอบร่วม หรือจำนวนข้อสอบแต่ละลักษณะมีสัดส่วนที่เท่ากันหรือใกล้เคียงกันมาก

1.2.2 เมื่อแบบทดสอบผสมมีความยาว และรูปแบบการให้คะแนนคงที่ ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับความยากของแบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 เท่ากับความสามารถในการสอบครั้งที่ 1 (0.00) ต่ำกว่า กรณีแบบทดสอบผสมมีความยาก 0.50 อย่างมีนัยสำคัญที่ระดับ .05 ซึ่งไม่สอดคล้องสมมติฐานย่อยที่ตั้งไว้ว่า *“ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกันเมื่อระดับความยากของแบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 เท่ากับความสามารถในการสอบครั้งที่ 1 (0.00) น่าจะมีความแม่นยำมากกว่าเมื่อความยากของแบบทดสอบผสมที่ใช้ในการสอบครั้งที่ 1 ไม่สอดคล้องกับความสามารถในการสอบครั้งที่ 1 (0.00)”* การที่ผลการวิจัยไม่เป็นไปตามสมมติฐานอาจจะเป็นไปได้ว่าข้อสมมติฐานที่ตั้งไว้นั้นไม่ได้ครอบคลุมถึงการแจกแจงความสามารถของผู้สอบในการทดสอบในครั้งที่สองด้วย โดยผลการวิจัยของ เกา (Cao, 2008: 91) ได้ข้อสรุปเกี่ยวกับวิธีการประมาณค่าพารามิเตอร์พร้อมกันว่า การแจกแจงความสามารถของผู้สอบมีผลที่สำคัญมากต่อการปรับเทียบพารามิเตอร์โดยพบว่าหากกลุ่มผู้สอบมีความไม่เท่าเทียมกันเรื่องความสามารถแล้วความลำเอียงและค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสอง (RMSE) จะสูงกว่ากรณีผู้สอบสองกลุ่มมีความสามารถเท่าเทียมกัน

1.2.3 เมื่อแบบทดสอบผสมมีความยาว และความยากคงที่ ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อมีการให้คะแนน 3 ระดับ สูงกว่ากรณีให้คะแนน 4 และ 5 ระดับอย่างมีนัยสำคัญที่ระดับ .05 ซึ่งต่างไปจากสมมติฐานที่ตั้งไว้ว่า *“ความแม่นยำของค่าประมาณ*

ความสามารถกำลังสอง (RMSE) เมื่อมีการให้คะแนน 5 ระดับ น่าจะมีความแม่นยำสูงกว่าการให้คะแนน 3 ระดับ หรือ 4 ระดับ"

การที่แบบทดสอบผสมมีความยาว และความยากคงที่ ความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อมีการให้คะแนน 3 ระดับ สูงกว่ากรณีให้คะแนน 4 และ 5 ระดับ นั้นอาจจะเนื่องมาจากสัดส่วนจำนวนข้อสอบแต่ละลักษณะในแบบทดสอบผสม ดังผลของการศึกษาของลี ลิสิทซ์ และหยาง (Li, Lissitz ; & Yang.1999) บาสตารี (Bastari. 2000) เทท Tate (2000) และอัญชลี ศรีกลชาญ (2552) พบว่าเมื่อกลุ่มผู้สอบที่มีความสามารถไม่เท่าเทียมกัน ผลการปรับเทียบจะมีความคลาดเคลื่อนลดลงเมื่อสัดส่วนของข้อสอบทั้ง 2 ชนิดในแบบทดสอบผสมประกอบด้วยข้อสอบที่ตรวจให้คะแนนสองค่าเพิ่มมากขึ้น ในการวิจัยครั้งนี้ผู้วิจัยกำหนดเงื่อนไขให้มีระดับคะแนน 3 ระดับ 4 ระดับ และ 5 ระดับเพื่อให้น้ำหนักคะแนนที่ได้จากข้อสอบเลือกตอบและข้อสอบเขียนตอบ มีคะแนนรวมย่อยแตกต่างกัน ซึ่งเมื่อพิจารณาแล้วจะพบว่าคะแนนของข้อสอบแบบให้คะแนนหลายค่าแบบให้คะแนน 3 ระดับ (0/1/2) 4 ระดับ (0/1/2/3) และ 5 ระดับ(0/1/2/3/4) นั้นจะทำให้ข้อสอบ ขนาด 30:10 24:8 และ 15:5 มีสัดส่วนน้ำหนักคะแนนรวมของข้อสอบเลือกตอบสูงกว่าคะแนนสอบแบบให้คะแนนหลายค่า (30:20 24:16 และ 15:10) เท่ากับคะแนนสอบแบบให้คะแนนหลายค่า (30:30 24:24 และ 15:15) และ ต่ำกว่าคะแนนสอบแบบให้คะแนนหลายค่า (30:40 24:32 และ15:20) ตามลำดับ จึงเป็นได้ว่าความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อมีการให้คะแนน 3 ระดับ สูงกว่ากรณีให้คะแนน 4 และ 5 ระดับ

2. การอภิปรายผลข้อมูลที่ได้จากการจำลองข้อมูล

2.1 การกำหนดค่าพารามิเตอร์ความสามารถของผู้สอบ และข้อสอบ

ผู้วิจัยจำลองค่าพารามิเตอร์ของข้อสอบ ได้แก่ อำนาจจำแนก (a) ความยาก (b) และการเดา (c) ในแบบทดสอบผสมโดยใช้โมเดล 3PL สำหรับข้อสอบแบบเลือกตอบที่ตรวจให้คะแนนสองค่า และ โมเดล GPCM สำหรับข้อสอบแบบเขียนตอบที่ตรวจให้คะแนนหลายค่า ในการจำลองข้อมูลครั้งนี้ผู้วิจัยได้กำหนดรูปแบบการแจกแจงค่าพารามิเตอร์แต่ละค่าให้มีความสอดคล้องกับการศึกษาของเปคและยัง (Peak; & Young. 2005) โดยกำหนดให้ค่าพารามิเตอร์ความชันมีการแจกแจงแบบเอกรูปโดยมีค่าตั้งแต่ 0.8 – 1.2 กำหนดให้ค่าพารามิเตอร์การเดามีการแจกแจงแบบเอกรูปโดยมีค่าตั้งแต่ 0.01 – 0.25 และกำหนดให้ค่าพารามิเตอร์ความชันประจำรายการคำตอบข้อสอบแบบให้เขียนตอบมีการแจกแจงแบบเอกรูปมีค่าตั้งแต่ 0.8 – 1.2 ส่วนพารามิเตอร์ความยากของข้อสอบ (b) มีการแจกแจงแบบปกติ มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 เนื่องจากการวิจัยครั้งนี้ผู้วิจัยได้กำหนดเงื่อนไขเกี่ยวกับระดับความสามารถของผู้สอบในการทดสอบครั้งที่สองออกเป็น 9 กลุ่มแตกต่างตาม

ลักษณะของค่าเฉลี่ยความสามารถเพิ่มขึ้นแตกต่างกัน 3 ระดับ(0.1 0.5 และ 1.0) และส่วนเบี่ยงเบนมาตรฐานต่างกัน 3 ระดับ (1.00 0.8 และ 1.20) และกำหนดให้ค่าพารามิเตอร์ความยากของข้อสอบสำหรับการสอบในครั้งที่ 2 ที่ไม่ใช่ข้อสอบร่วมมีความยากเพิ่มขึ้นข้อละ 0.5 หน่วย สำหรับข้อสอบแบบเขียนตอบที่ให้คะแนนแบบหลายค่าให้เพิ่มความยากขึ้นขั้นละ 0.5 แต่เพื่อให้ค่าพารามิเตอร์ข้อสอบเหมาะสมกับลักษณะผู้สอบทุกกลุ่มจึงได้มีการควบคุมให้พารามิเตอร์ความยากอยู่ในช่วง -3 ถึง +3

การจำลองข้อมูลครั้งนี้ผู้วิจัยได้กำหนดค่าพารามิเตอร์ข้อสอบและค่าพารามิเตอร์ความสามารถของผู้สอบด้วยโปรแกรม WINGEN2 (Han, 2007) เพื่อให้ได้รูปแบบการตอบข้อสอบของผู้สอบที่ทำแบบทดสอบผสมแต่ละฉบับ (Test 1 และ Test 2) ลักษณะของโปรแกรม WINGEN2 มีการกำหนดค่าตั้งต้นของ Scaling factor มีค่า D = 1.0 และผู้ใช้สามารถเลือก Scaling factor เป็น ค่า D = 1.7 ได้ ในการจำลองข้อมูลครั้งนี้ ผู้วิจัยเลือก Scaling factor เป็นค่า D = 1.7 เช่นเดียวกับ อัญชลี ศรีกลชาญ (2552) ที่แนะนำว่าหากเลือก Scaling factor เป็นค่า D = 1.0 ตามที่โปรแกรมกำหนด (default) แล้วผลการจำลองข้อมูลจะได้รูปแบบการตอบข้อสอบที่มีองค์ประกอบจำนวนมากจะทำให้เกิดการละเมิดความเป็นมิติเดียวของแบบทดสอบได้

แม้ว่าการศึกษาครั้งนี้เป็นการศึกษาภายใต้สถานการณ์จำลอง แต่ผู้วิจัยได้กำหนดลักษณะพารามิเตอร์ผู้สอบและข้อสอบให้โปรแกรมจำลองข้อมูลโดยศึกษาจากเอกสารและงานวิจัยที่เกี่ยวข้องและเรียนรู้วิธีการใช้โปรแกรม เพื่อให้ข้อมูลที่ศึกษาเป็นไปตามข้อตกลงเบื้องต้นตามทฤษฎีการตอบข้อสอบ และสอดคล้องกับโมเดลที่ใช้ประมาณค่ามากที่สุด ดังนั้นจึงมั่นใจได้ว่าแบบแผนการตอบที่จำลองมานั้นสอดคล้องกับเงื่อนไขที่ต้องการศึกษามากที่สุด

2.2 การประมาณค่าพารามิเตอร์พร้อมกันด้วยโปรแกรม MULTILOG

ในการประมาณค่าพารามิเตอร์พร้อมกันผู้วิจัยใช้โปรแกรม MULTILOG ในการประมาณค่าพารามิเตอร์ความสามารถเช่นเดียวกับหลายงานวิจัย (Cao. 2008; Kinsey. 2003; Hanson; & Béguin. 2002, Kim; & Cohen. 1996; อัญชลี ศรีกลชาญ. 2552) ที่ได้ใช้โปรแกรม MULTILOG สำหรับประมาณค่าพารามิเตอร์พร้อมกัน ซึ่งเมื่อผู้วิจัยได้ผลการตอบข้อสอบของแบบทดสอบผสมสองฉบับของผู้สอบ 1,000 คนตามเงื่อนไขที่ศึกษาจากโปรแกรมจำลองข้อมูลแล้ว ผู้วิจัยได้นำผลการตอบข้อสอบทั้งสองฉบับมารวมเข้าเป็นผลการตอบของการทำข้อสอบฉบับใหม่ที่รวมเอาผลการตอบของการทำข้อสอบสองฉบับเข้าด้วยกัน ดังนั้นผลการตอบข้อสอบใหม่จึงประกอบด้วยข้อสอบเฉพาะของการทดสอบครั้งที่ 1 ข้อสอบร่วม และ ข้อสอบเฉพาะของการทดสอบครั้งที่ 2 โดยกำหนดให้ข้อสอบเฉพาะของแต่ละฉบับที่ผู้สอบไม่ได้ทำเป็น 0 ซึ่งโปรแกรม MULTILOG จะมองเป็น Missing ผู้วิจัยเขียนคำสั่ง (Syntax) เช่นเดียวกับ เกา (Cao. 2008) สำหรับ run โปรแกรมการประมาณค่าพารามิเตอร์พร้อมกันสำหรับการปรับเทียบผลการตอบข้อสอบของแบบทดสอบผสมสองครั้ง (2

กลุ่ม) โดยกำหนดให้โมเดล 3PL ประมาณค่าพารามิเตอร์ผลการตอบข้อสอบแบบให้คะแนนสองค่า และโมเดล GPCM ประมาณค่าผลการตอบข้อสอบแบบให้คะแนนหลายค่าพร้อมกัน ในการประมาณค่าพารามิเตอร์พร้อมกันใช้การ run โปรแกรมเพียงครั้งเดียว จึงเป็นสิ่งหนึ่งที่สามารถสนับสนุนได้ว่าค่าพารามิเตอร์ที่ประมาณได้อยู่บนสเกลความสามารถเดียวกันจึงสามารถนำไปเปรียบเทียบกันได้ แต่อย่างไรก็ตามสิ่งที่ต้องให้ความสำคัญเป็นพิเศษในการใช้โปรแกรม MULTILOG ในการประมาณค่าพารามิเตอร์พร้อมกันเมื่อกลุ่มผู้สอบสองกลุ่มหรือสองครั้งมีการแจกแจงความสามารถไม่เท่าเทียมกัน เนื่องจากปกติโปรแกรมจะกำหนดให้กลุ่มฐานมีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1

ในการใช้โปรแกรม MULTILOG มีข้อจำกัดเกี่ยวกับการแจกแจงความสามารถของผู้สอบจึงอาจทำให้การศึกษาพัฒนาการความสามารถของผู้สอบ ค่าพัฒนาการความสามารถของผู้สอบในกลุ่มที่มีระดับพัฒนาการความสามารถเพิ่มขึ้น 0.1 มีค่าความลำเอียง และค่ารากที่สองของค่าเฉลี่ยความคลาดเคลื่อนกำลังสองต่ำกว่า กลุ่มระดับระดับพัฒนาการความสามารถเพิ่มขึ้น 0.5 หรือ 1 เนื่องจากค่าที่เพิ่มขึ้นนี้มีค่าแตกต่างจากค่าเฉลี่ย 0 ค่อนข้างมาก

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

การศึกษาครั้งนี้ผู้วิจัยได้ศึกษาโดยการจำลองข้อมูลด้วยโปรแกรมสำเร็จรูป WINGEN2 โดยใช้โมเดล 3PL ร่วมกับ โมเดล GPCM ในการจำลองข้อมูล และประมาณค่าพารามิเตอร์ของผลการตอบที่จำลองได้ เพื่อควบคุมไม่ให้เกิดการละเมิดข้อตกลงเบื้องต้นในการใช้โมเดลการตอบข้อสอบ แต่ในสถานการณ์จริง ข้อมูลสำหรับใช้ศึกษาวิธีการประมาณค่าพารามิเตอร์พร้อมกันด้วยโมเดล 3PL ร่วมกับ GPCM อาจจะมีผลที่แตกต่างออกไป เนื่องจากความเป็นไปได้ที่จะเกิดการละเมิดข้อตกลงเบื้องต้นมีสูง โดยเฉพาะผลการตอบข้อสอบแบบผสมที่มีข้อสอบหลายลักษณะในฉบับเดียวกัน หรือการใช้โมเดลผสมสำหรับการประมาณค่าพารามิเตอร์ ตลอดจนปัจจัยอื่นที่เกี่ยวข้องกับความแม่นยำของวิธีการปรับเทียบได้เช่น จำนวนข้อสอบร่วม ลักษณะข้อสอบร่วม ลักษณะของแบบทดสอบผสม เป็นต้น จากผลการวิจัยในภาพรวม พบว่าวิธีการประมาณค่าพารามิเตอร์พร้อมกันมีความแม่นยำสูงกว่าเมื่อผู้สอบมีพัฒนาการความสามารถเพิ่มขึ้น 0.1 ส่วนกรณีที่ผู้สอบมีความสามารถเพิ่มขึ้น 0.5 หรือ 1.0 นั้นการใช้วิธีการประมาณค่าพารามิเตอร์กันอาจจะต้องพิจารณาก่อนตัดสินใจนำไปใช้ อย่างไรก็ตามลักษณะความสามารถโดยทั่วไปมักจะมีค่าเพิ่มขึ้นไม่มากกว่า 0.5 ดังนั้นวิธีการประมาณค่าพารามิเตอร์พร้อมกันจึงน่าจะมีความแม่นยำในการนำไปใช้ศึกษาพัฒนาการความสามารถของผู้สอบทางด้านการศึกษาได้ ซึ่งผู้วิจัยมีประเด็นเสนอไว้สำหรับผู้สนใจนำผลการวิจัยไปใช้ดังนี้

1.1 การศึกษาด้วยข้อมูลจำลองจะมีการควบคุมไม่ให้เกิดการละเมิดข้อตกลงเบื้องต้น หรือให้มีการละเมิดข้อลงเบื้องต้นน้อยที่สุด ดังนั้นในการศึกษาผลการวิจัยในทำนองเดียวกันกับข้อมูลในสถานการณ์จริงจึงควรได้มีการตรวจสอบข้อมูล ว่ามีลักษณะตามข้อตกลงเบื้องต้นของการใช้โมเดลการตอบข้อสอบก่อนใช้

1.2 เมื่อศึกษาพัฒนาการของผู้สอบกับข้อมูลจริงด้วยวิธีการประมาณค่าพารามิเตอร์พร้อมกันแล้วพบว่าค่าพัฒนาการนั้นเพิ่มขึ้นไม่สูงกว่า 0.1 อาจจะสามารถนำผลที่ได้ไปใช้ต่อไปได้แต่หากพบว่าค่าพัฒนาการความสามารถนั้นมีค่าสูงกว่า 0.1 จำเป็นมีการตรวจสอบหรือเปรียบเทียบผลที่ได้กับวิธีการปรับเทียบคะแนนด้วยวิธีการอื่น หรือคำนวณด้วยวิธีคำนวณพัฒนาการด้วยสูตรอื่น เพื่อสนับสนุนผลที่ได้จากการประมาณค่าพารามิเตอร์พร้อมกัน

1.3 จำนวนข้อสอบร่วม และลักษณะข้อสอบร่วมมีผลต่อคุณภาพของการปรับเทียบค่อนข้างมาก หากต้องการเปรียบเทียบปัจจัยอื่นที่ส่งผลต่อคุณภาพวิธีการประมาณค่าพารามิเตอร์พร้อมกันควรมีการควบคุมให้จำนวนข้อสอบร่วมมีความเหมาะสม และมีลักษณะเป็นตัวแทนที่ดีของข้อสอบทั้งสองฉบับ

การศึกษาวิจัยครั้งนี้เป็นการศึกษาปัจจัยที่มีผลต่อความแม่นยำของวิธีการประมาณค่าพารามิเตอร์พร้อมกัน ซึ่งเป็นเพียงจำนวนหนึ่งของปัจจัยจำนวนมากที่มีต่อการทดสอบในสถานการณ์จริง ผู้วิจัยได้ควบคุม กลุ่มตัวอย่าง สัดส่วนจำนวนข้อสอบแบบให้คะแนนสองค่าและแบบให้คะแนนหลายค่า จำนวนข้อสอบร่วม ลักษณะของข้อสอบร่วม โมเดลที่ใช้จำลองข้อมูลและประมาณค่า เพื่อยืนยันคุณภาพของวิธีการปรับเทียบ แต่อย่างไรก็ตามงานวิจัยที่ผ่านมาพบว่าความยาวของข้อสอบร่วมที่สะท้อนเนื้อหาของแบบทดสอบทั้งฉบับย่อมเป็นสิ่งดี แต่ต้องคำนึงถึงจำนวนข้อสอบร่วมว่าไม่ควรมากเกินไปเพราะอาจจะทำให้มีผลต่อคะแนนสอบได้ นอกจากนี้แม้ว่าผู้วิจัยจะมีการควบคุมให้ข้อมูลสอดคล้องกับโมเดลที่ใช้ประมาณค่าและเป็นไปตามข้อตกลงเบื้องต้นของทฤษีการตอบข้อสอบเป็นอย่างดี แต่ผลการประมาณค่าที่ได้ในหลายเงื่อนไขยังได้ผลแตกต่างไปจากสมมติฐานที่ตั้งไว้ ดังนั้นในการนำผลงานวิจัยครั้งนี้ไปใช้ต่อไป จึงควรได้มีการตรวจสอบข้อตกลงเบื้องต้นของข้อมูลตามทฤษฎีการตอบข้อสอบ แม้ว่าข้อมูลนั้นจะได้มาจากโปรแกรมการจำลองแบบแผนการตอบที่มีความถูกต้องแม่นยำ ควรมีการตรวจสอบความเป็นตัวแทนที่ดีของข้อสอบร่วมในด้านสถิติควบคู่ไปพร้อมกับความเป็นตัวแทนทางด้านเนื้อหา ตลอดจนความเป็นมิติเดียวของข้อมูล และความเท่าเทียมกันของความสามารถของผู้สอบทั้งสองครั้งซึ่งปัจจัยต่างๆ เหล่านี้ล้วนมีผลต่อความแม่นยำวิธีการประมาณค่าพารามิเตอร์พร้อมกันทั้งสิ้น

ข้อเสนอแนะในทำวิจัยครั้งต่อไป

2.1 ในการศึกษาครั้งนี้เป็นการจำลองข้อมูลเพื่อศึกษาพัฒนาการความสามารถที่ได้จากการทดสอบซ้ำโดยใช้แบบทดสอบผสมที่มีความยากต่างกัน ดังนั้นควรศึกษาผลการวัดพัฒนาการความสามารถในสถานการณ์จริงโดยใช้แบบทดสอบผสมที่มีเงื่อนไขสอดคล้องกับสถานการณ์จำลอง

2.2 ในการศึกษาครั้งนี้เป็นศึกษาภายใต้สถานการณ์จำลอง โดยกำหนดให้พัฒนาการความสามารถนั้นเกิดจากการคำนวณความแตกต่างของค่าประมาณความสามารถเพียงอย่างเดียว ยังไม่ได้มีการเปรียบเทียบว่า พัฒนาการความสามารถที่ประมาณค่าได้จากวิธีการประมาณค่าพารามิเตอร์พร้อมกันนี้ ให้ผลที่สอดคล้องหรือแตกต่างไปจากวิธีการวัดพัฒนาการที่คำนวณด้วยวิธีการวัดพัฒนาการหรือการเปลี่ยนแปลงอื่นอีกหรือไม่ ภายใต้เงื่อนไขเดียวกัน เมื่อข้อสอบแบบผสมมีความยาว ความยาก และระดับการให้คะแนนแตกต่างกัน

2.3 ในการศึกษาครั้งนี้เป็นศึกษาภายใต้สถานการณ์จำลองที่มีการสอบสองครั้ง จึงควรศึกษาผลการใช้วิธีการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อมีทดสอบมากกว่าสองครั้ง หรือใช้แบบทดสอบหลายฉบับ

2.4 ควรมีการศึกษาผลการใช้แบบทดสอบผสมสำหรับศึกษาพัฒนาการความสามารถของผู้สอบ ในสถานการณ์จริงเพื่อตรวจสอบองค์ความรู้ที่ได้จากการศึกษากับสถานการณ์การจำลอง เนื่องจากที่ผ่านมายังไม่ปรากฏว่ามีศึกษาผลการปรับเทียบคะแนนความสามารถของผู้สอบเมื่อการใช้แบบทดสอบผสมที่มีความยากต่างกันในการศึกษาพัฒนาการของผู้สอบ

2.5 เนื่องจากการศึกษาด้วยแบบทดสอบผสมโดยการใช้โมเดลการตอบข้อสอบแบบมิติเดียว ประมาณค่ามีแนวโน้มที่จะละเมิดข้อตกลงเบื้องต้นเกี่ยวกับความเป็นมิติเดียวได้ง่าย ด้วยลักษณะเฉพาะของแบบทดสอบผสมที่มีความซับซ้อน อีกทั้งมีลักษณะข้อสอบที่แตกต่างกันในการทดสอบครั้งเดียวกัน ดังนั้นควรมีการศึกษาผลการใช้โมเดลหลายมิติในการประมาณค่าผลการทดสอบด้วยแบบทดสอบผสมเพื่อดูความเหมาะสม ข้อจำกัด และความเป็นไปได้ในการนำไปใช้จริงเมื่อเทียบกับการใช้โมเดลแบบมิติเดียว

2.6 ในศึกษาครั้งนี้กำหนดให้ความแตกต่างของข้อสอบในการทดสอบครั้งที่สองมีค่าความยากเพิ่มขึ้นคงที่ข้อละ 0.5 การศึกษาครั้งต่อไปควรศึกษาผลของการเปลี่ยนแปลงค่าความยากของข้อสอบเฉพาะที่ไม่ใช่ข้อสอบร่วมว่าจะส่งผลอย่างไรต่อผลการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อขนาด ทิศทาง และการกระจายความยากของข้อสอบแตกต่างกัน

บรรณานุกรม

บรรณานุกรม

กระทรวงศึกษาธิการ. (2545). *พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 และที่แก้ไขเพิ่มเติม (ฉบับที่ 2) พ.ศ. 2545 พร้อมกฎกระทรวงที่เกี่ยวข้อง และพระราชบัญญัติการศึกษาภาคบังคับ พ.ศ. 2545.* กรุงเทพฯ.

จรรยา สิงห์ทอง. (2550). *การศึกษาผลของการวัดการเปลี่ยนแปลงห้าวิธีแบบวัดซ้ำสองครั้งภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกันและขนาดของกลุ่มตัวอย่างที่ต่างกัน.* ปริญญานิพนธ์ การศึกษาดุษฎีบัณฑิต (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

ทีมวิชาการดอทคอม. (2545). *International English Language Testing System.* สืบค้นเมื่อ 25 พฤษภาคม 2554, จาก http://www.vcharkarn.com/varticle/42

_____. (2544). *Graduate Management Admissions Test.* สืบค้นเมื่อ 25 พฤษภาคม 2554, จาก http://www.vcharkarn.com/varticle/60

_____. (2544). *Test of English as a Foreign Language.* สืบค้นเมื่อ 25 พฤษภาคม 2554, จาก http://www.vcharkarn.com/varticle/6

_____. (2544). *Graduate Record Examination.* สืบค้นเมื่อ 25 พฤษภาคม 2554, จาก http://www.vcharkarn.com/varticle/59

ธีรยุทธ ภูเขา. (2550). *การศึกษาประสิทธิภาพโมเดลสมการโครงสร้าง 3 รูปแบบ ในการศึกษาปัจจัยด้านเชาวน์ปัญญา และเชาวน์อารมณ์ที่ส่งผลต่ออัตราพัฒนาการวิชาคณิตศาสตร์.* ปริญญานิพนธ์ การศึกษาดุษฎีบัณฑิต (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.ถ่ายเอกสาร.

ประสิทธิ์ ไชยกาล. (2539). การเปรียบเทียบประสิทธิภาพระหว่างโมเดลลิสเรล 3 แบบที่ใช้ในการศึกษาตัวแปรที่สัมพันธ์กับการเปลี่ยนแปลงในระยะยาว ของผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์. วิทยานิพนธ์ครุศาสตรมหาบัณฑิต. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.ถ่ายเอกสาร.

พัชรี จันทร์เพ็ง. (2547). *การศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนโดยใช้วิธีการปรับเทียบแนวตั้ง ตามวิธีทฤษฎีตอบสนองข้อสอบร่วมกับวิธีการปรับเทียบเชิงเส้นตรง.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.ถ่ายเอกสาร.

รณชิต พฤษกรรม. (2547). *การศึกษาพัฒนาการของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาตอนต้น สังกัดกรมสามัญศึกษา เขตกรุงเทพมหานคร : การวิเคราะห์อิทธิพลสุ่มกลุ่มเวลาการอยู่รอด.*วิทยานิพนธ์ครุศาสตรมหาบัณฑิต (สถิติการศึกษา). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.ถ่ายเอกสาร.

วินิจ เทือกทอง. (2537). *การศึกษาเปรียบเทียบประสิทธิภาพของการคำนวณคะแนนเพิ่มวิธีต่างๆ ด้วยระเบียบวิธีการมอนติคาร์โล.*ปริญญานิพนธ์ การศึกษาดุษฎีบัณฑิต (วิจัยและพัฒนาหลักสูตร). กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.ถ่ายเอกสาร.

ศศิวิมล อมตชีวิน. (2546). *การเปรียบเทียบผลการวิเคราะห์เหลื่อมลำดับในโมเดลโค้งพัฒนาการที่มีช่วงเวลาการวัดแตกต่างกัน.* วิทยานิพนธ์ครุศาสตรมหาบัณฑิต.กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.ถ่ายเอกสาร.

ศิริชัย กาญจนวาสี. (2538). การวัดการเปลี่ยนแปลง. *การประชุมเชิงปฏิบัติการครั้งที่ 3 เรื่อง "หลักและวิธีวิจัยขั้นสูงเฉพาะการวิจัยและพัฒนาระบบพฤติกรรมไทยด้านต่างๆ".* สำนักงานคณะกรรมการการวิจัยแห่งชาติร่วมกับคณะกรรมการแห่งชาติเพื่อการวิจัยและพัฒนาระบบพฤติกรรมไทย. 144 – 149. กรุงเทพฯ: เดอะเบสท์ กราฟิค ดีไซด์.

ศุภลักษณ์ ใจแสวงทรัพย์. (2547). *ปัจจัยที่ส่งผลต่อคะแนนพัฒนาการวิชาคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 3.* ครุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย

สถาบันทดสอบทางการศึกษาแห่งชาติ. (2553). *คู่มือนักเรียนในการสอบ O-NET ป.6, ม.-3.* สืบค้นเมื่อ 25 พฤษภาคม 2554, จาก http://www.niets.or.th/

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2546). *คู่มือการวัดผลประเมินผลคณิตศาสตร์.* กระทรวงศึกษาธิการ กรุงเทพฯ.

สมถวิล วิจิตรวรรณา. (2543). *การเปรียบเทียบประสิทธิภาพของโมเดลโค้งพัฒนาการ ที่มีตัวแปรแฝงโมเดลพหุระดับ และโมเดลกึ่งซิมเพลกซ์ ในการวัดการเปลี่ยนแปลงระยะยาวชนิดตัวแปรเดี่ยวและตัวแปรพหุ.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต การวัดและประเมินผลการศึกษา). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.ถ่ายเอกสาร.

สุพจน์ เกิดสุวรรณ. (2545). *การพัฒนาการวัดความรู้บางส่วนของผู้ตอบแบบสอบเลือกตอบ.*วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.ถ่ายเอกสาร.

สุวิมล ว่องวานิช. (2550). *การประเมินผลการเรียนรู้แนวใหม่(พิมพ์ครั้งที่ 2).* คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.กรุงเทพฯ.

อรุณี อ่อนสวัสดิ์. (2537). *การพัฒนาวิธีการวัดการเปลี่ยนแปลงการเรียนรู้.* วิทยานิพนธ์ครุศาสตรดุษฎีบัณฑิต. (การวัดและประเมินผลการศึกษา).กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

อรุณี อ่อนสวัสดิ์. (2539, มีนาคม - เมษายน). พัฒนาการการวัดการเปลี่ยนแปลงการเรียนรู้. *วารสารการวัดผลการศึกษา.* 17(51): 1 - 18.

อวยพร เรืองตระกูล. (2544). *การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการ ตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ.* วิทยานิพนธ์ครุศาสตร์ดุษฎีบัณฑิต (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

อวยพร เรืองตระกูล. (2545, มกราคม - เมษายน). การพัฒนาและการวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ. *วารสารวิธีวิทยาการวิจัย.* 15(1): 115 - 127.

อัญชลี ศรีกลชาญ (2552).*คุณภาพของการปรับเทียบคะแนนสำหรับแบบสอบรูปแบบผสม: การประยุกต์ใช้การปรับเทียบตามทฤษฎีการตอบสนองข้อสอบด้วยวิธีโค้งคุณลักษณะและการปรับค่าพารามิเตอร์พร้อมกัน.* วิทยานิพนธ์ครุศาสตร์ดุษฎีบัณฑิต (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

Ackerman, T. (1989). Unidimensional IRT calibration of compensatory and noncompensatory items. *Applied Psychological Measurement*, 13,113-127.

Ackerman,T.,Gierl,M.J.,&Walker,C.M.(2003). *Using multidimensional item response theory to evaluate educational psychological tests*. Educational Measurement: Issues and Practice, 22, 37-51.

Adams, Raymond J.; Wilson, Mark.;& Wang, Wen-chung. (1997, March). The Multidimensional Random Coefficients Multinomial Logit Model. *Applied Psychological Measurement.* 21(1): 1-23.

Allen, N. L., Carlson, J. E., & Zelenak, C. A. (1999). *The NAEP 1996 technical report (NCES 1999-452). Washington, DC: National Center for Education Statistics*. Retrieved November 1,2008, from, http://nces.ed.gov/nationsreportcard/pdf/main1996/1999452.pdf

Archwamety, Teara.;& Tangdhanakanond, Kamonwan. (2007, January-April). Equalizing Gain Score: Relative Gain Score vs. Cumulative Relative Gain Score. *Journal of Research Methodology.* 20(1): 1-21.

Armenakis, Achilles A.; Buckley, M. Ronald.; & Bedeian, Arthur G. (1986, June). Survey Research Measurement Issues in Evaluating Change: A Laboratory Investigation. *Applied Psychological Measurement*. 10(2): 147 - 157.

Baker, F. B.,& Al-Karni, A. (1991). A comparison of two procedures for computing IRT equating coefficients. *Journal of Educational Measurement*, 28, 147–162.

Baker, F. B. (1992). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker.

Baker, F. B., & Kim, S.-H. (2004). *Item response theory: Parameter estimation techniques*. (2nd ed.). New York: Marcel Dekker.

Ban, J. –C. Hanson, B. A., Wang, T., Yi, Q., & Harris, D. J. (2001). A comparative study of on-line pretest-item calibration/scaling methods in computerized adaptive testing. *Journal of Educational Measurement*, 38, 181-212.

Bastari, B . (2000). *Linking multiple-choice and constructed-response items to a common proficiency scale*. Retrieved November 1,2008, from, http://proquest.umi.com/pqdlink?did=731794221&Fmt=7&clientId=61839&RQT=309&VName=PQD

Béguin, A. A.; Hanson, B. A.; & Glas, C. A. W. (2000). *Effect of multidimesnioanlity on separate and concurrent estimation in IRT equating.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans, LA.

Beretrvas, Susan Natasha. (2000). *To Meet or Not to Meet Standard: Proficiency Estimation Using Different Polytomous IRT Models.* University of Washington. Retrieved February 10,2008, from, http://proquest.umi.com/pqdlink?did=727733131&Fmt=2&clientId=61839&RQT=309&VName=PQD

Beretvas, S. Natasha.; & Williams, Natasha J. (2004, Winter). The Use of Hierarchical Generalized Linear Model for Item Dimensionality Assessment. *Journal of Educational Measurement.* 41(4): 379 - 395.

Bock, R. Darrell; & Mislevy, Robert J.(1982). Adaptive EAP Estimation of Ability in a Microcomputer Environment . *Applied Psychological Measurement*. 6: 431-444

Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*. 37, 29-51.

Bock, R. D. (1997). The nominal categories model. In W. J. van der Linden & R. K. Hambleton (Eds), *Handbook of modern item response theory*. (pp. 33-49). New York, NY: Springer.

Bock, R. D., & Aitken, M. (1981). Marginal maximum likelihood estmation of item parameters: Application of an EM algorithm. *Psychometrika,* 46, 443-459.

Bock, R. D., & Zimowski, M. F. (1997).Multiple group IRT. InW. J. van der Linden & R. K. Hamilton(Eds.), *Handbook of modern item response theory*. (pp. 433–448). New York: Springer-Verlag.

Bock, R. Darrell.; Thissen, David.; Zimowski, Michele F. (1997, Fall). IRT Estimate of Domain Scores. *Journal of Educational Measurement*. 34(3): 197 – 211.

Bridgeman, Brent. (1992, Fall). A Comparison of Quantitative Questions in Open-Ended and Multiple - Choice Formats. *Journal of Educational Measurement*. 29(3): 253 - 271.

Briggs, Derek C. (2003). An Introduction to Multidimensional Measurement Using Rasch Models. *Journal of Applied Measurement*. 4(1): 87 - 100. Retrieved February10,2008,from,http://epicpolicy.org/files/Briggs&Wilson_IntrotoMultidimensionalMeasurement.pdf

Briggs, Derek C.; Weeks, Jonathan P.; &Wiley, Edward. (2008).*The Impact of Vertical Scaling Decisions on Growth Projections.* Paper presented at the annual meeting of the National Council for Measurement in Education Annual Conference, March 26, 2008, New York, NY.

Brown, Nathaniel J. S. (2005). *The Multidimensional Measure of Conceptual Complexity*. Retrieved June 20, 2008, from http://bearcenter.berkeley.edu/publications/MMCC.pdf

Burton, R. (2001). Quantifying the effects of chance in multiple-choice and true/false tests: Question selection and guessing of answers. *Assessment and Evaluation in Higher Education* 26(1), 41-50.

Cao, Y. (2008). *Mixed-Format test equating: Effects of Test dimensionality and common item sets*. Doctoral Dissertation, Department of Measurement Statistics and Evaluation, Faculty of the Graduate School, University of Maryland. Retrieved April 5, 2009 from http://drum.lib.umd.edu/bitstream/1903/8843/1/umi-umd-5871.pdf

Cribbie, R. A.; & Jamieson, J. (2004). Decreases in Posttest Variance and the Measurement of Change. *Methods of Psychological Research*. Online, 9(1), 37-55. Retrieved February 8, 2008 from http://www.mpr-online.de//

Cronbach. L.J.; & Furby. L. (1970). How should we measure "change"---or should we?. Psychological Bulletin. 74: 68-80.

Cook, Karon Frances.(1996). *A Comparison of Three Polytomous Item Response Theory Models in the context of Testlet Scoring*. The University of Texas at Austin.

Davey, T.; Oshima, T. & Lee, K. (1996, December). Linking multidimensional item calibrations. *Applied Psychological Measurement*. 20(4): 405–416.

De Ayala, R.J., & Sava-Bolesta. (1999). Item parameter recovery for the nominal response model. *Applied Psychological Measurement*. 23(1): 3 -19.

de la Torre, Jimmy. (2008). Multidimensional Scoring of Abilities: The Ordered Polytomous Response Case. *Applied Psychological Measurement*.32(5): 355-370 .

DeMars, Christine E. (2004). *A Comparison of the Recovery of Parameters Using the Nominal Response and Generalized Partial Credit Models*. Poster presented at the annual meeting of American Educational Research Association, San Diego, CA. Retrieved July 8, 2009 from http://www.jmu.edu/assessment/research/faculty/demars/DemarsAERA04_recovery.PDF

Divgi, D. R. (1985). A minimum chi-square method for developing a common metric in item response theory. *Applied Psychological Measurement*, 9, 413-415.

Ding, Cody. (2003). *Exploratory Longitudinal Profile Analysis via Multidimensional Scaling*. Practical Assessment, Research & Evaluation. 8(12). Retrieved September 8, 2006 from http://PAREonline.net/getvn.asp?v=8&n=12

Dragow, F. (1989, March). An evaluation of marginal maximum lielihoo estimation for the two parameter logistic model. *Applied Psychological Measurement*. 13(1): 77 – 90.

Eid, Michael. (1996). Longitudinal Confirmatory Factor Analysis for Polytomous Item Response: Model Definition and Model Selection on the Basis of Stochastic Measurement Theory. *Method of Psychological Research Online*. 1(4): 65 - 85. Retrieved September 18, 2007 from http://www.mpr-online.de//

Embretson, Susan E. (1991, September). A Multidimensional Latent Trait Model For Measuring Learning And Change. *Psychometrika*. 56(3): 495 - 515.

Embretson, Susan E. (1992, Spring). Measuring and Validating Cognitive Modifiability as an Ability: A Study in the Spatial Domain. *Journal of Educational Measurement*. 29(1): 25 - 50.

Embretson, Susan E. (1995, Fall). A Measurement Model for Linking Individual Learning to Processes and Knowledge: Application to Mathematical Reasoning. *Journal of Educational Measurement*. 32(3). 277 - 294.

Embretson, Susan E. (2000). Multidimensional Measurement from Dynamic Tests: Abstract Reasoning Under Stress. *Multivariate Behavioral Research*. 35(4): 505 - 542.

Ercikan, K., Schwarz, R. D., Julian, M.W., Burket, G., & Link, V. (1998). Calibration and scoring of tests with multiple-choice and constructed-response item types. *Journal of Educational Measurement,* 3,137-154.

Feldt, Leonard S. (1995, Fall). Estimation of the Reliability of Differences Under Revised Reliabilities of Component Scores. *Journal of Educational Measurement.* 32(3): 295 - 301.

Fischer, Gerhard H. (2003, January). The Precision of Gain Scores Under an Item response Theory Perspective: A Comparison of Asymptotic and Exact Conditional Inference About Change. *Applied Psychological Measurement*. 27(1): 3 - 26.

Fischer, Gerhard H.; & Ponocny, Ivo. (1994, June). An Extention of the Partial Credit Model With an Application to The Measurement of Change. *Psychometrika.* 59(2): 177 - 192.

Fischer,G. H.; & Parzer, P. (1991, December). An Extension of the rating Scale Model With and Application to the Measurement of Change. *Psychometrika*. 56(4): 637 - 651.

Formann, Anton K.;& Spiel, Christiane. (1989, March). Measuring Change by Means of a Hybrid Variant of the Linear Logistics Model with Relaxed Assumptions. *Applied Psychological Measurement.* 13(1): 91 - 103.

Green, D.R., Yen, W.M. & Burket, G.R. (January 1989). Experiences in the application of item response theory in test construct. *Applied Measurement in Education*, 2(4), 297 – 312.

Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese psychological Research* 22(3): 144-149.

Hambleton, R.K., Swaminathan, H. & Rogers, H.J. (1991). *Fundamental of Item response Theory*. Newbury Park, C.A.: Sage publications.

Hamilton, L.S., Nussbaum, E.M., & Snow, R.E. (1997). Interview procedures for validating science assessments. *Applied Measurement in Education.* 10(2):181-200.

Han, K. T. (2007). *WinGen2: Windows software that generates IRT parameters and item responses [computer program].* Amherst, MA: University of Massachusetts Amherst, Center for Educational Assessment. Retrieved May 13, 2009, from http://www.umass.edu/remp/software/simcata/wingen/WinGen_Manual_Han_Hambleton_2007.pdf

_____. (2007). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement, 31:* 457-459

Hanson, B., & Béguin, 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.

Harwell, M. R., & Janosky, J. E. (1991). An empirical study of the effects of small datasets and varying prior variance on item parameter estimation in BILOG. *Applied Psychological Measurement.* 15(3): 279-291.

Heidenberg. A. J. (2000). *Reducing Pretest and Gain Score Correlation Using Item Response Theory Domain Sources*. Dissertation, Ph.D. (Research, Measurement and Statistics). Georgia: Georgia State University. Photocopied.

Hemker, B. T. (2001). *Reversibility revisited and other comparisons of three types of polytomous IRT models*. In A. Boomsma, M. A. J. van Duijn & T. A. B. Snijders (Eds.), Essays on item response theory. (pp. 277-296). New York, NY: Springer.

Hoskens, Machteld.;& Boeck, Paual De. (1995, Winter). Componential IRT Models for Polytomous Items. *Journal of Educational Measurement*. 32(4): 364 - 384.

Hu, H., Rogers, W. T., & Vukmirovic, Z. (2008). Investigation of IRT-based equating methods in the presence of outlier common items. *Applied Psychological Measurement*. 32(4): 311-333.

Hulin, Charles L.; Fritz Dragow.; & Charles K. Pasons. (1983). Item Response Theory: Application to Psychological Measurement. Homwood, Illinois: Dow Jones.

Huynh, H. (1994, March). On Equivalence Between a Partial Credit Item and a Set of Independent Rasch Binary Items. *Psychometrika*. 59(1): 111 - 119.

Jamieson, Jhon. (1995, February). Measurement of Change and The Law of Initial Values: A Computer Simulation Study. *Educational and Psychological Measurement*. 55(1): 38 - 46.

Jodoin, M. G., Keller, L. A., & Swaminathan, H. (2003). A comparison of linear, fixed common item, and concurrent parameter estimation equating procedures in capturing academic growth. *The Journal of Experimental Education*, 71(3), 229–250.

Karkee, T. & et al. (2003).*Separate versus Concurrent Calibration Methods in Vertical Scaling*. Paper presented at the Annual Meeting of the National Council on Measurement in Education (Chicago, IL, April 22-24, 2003). Retrieved January 18, 2009 from http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED478167&ERICExtSearch_SearchType_0=no&accno=ED478167

Karkee, T., Wang, Z., & Green, D. R. (2006, April). *Exploring the effects of dimensionality on three vertical scaling procedures*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Kamata, Akihito; & Tate, Richard. (2005). The Performance of a Method for the Long-term Equating of Mixed-Format Assessment. *Journal of Educational Measurement*, 42, 2: 193 - 213.

Kamata, Akihito. (2001, Spring). Item Analysis by the Hierarchical Generalized Linear Model. *Journal of Educational Measurement*. 38(1): 79 - 93.

Kelderman, Henk. (1996, June). Multidimensional Rasch Models for Partial - Credit Scoring. *Applied Psychological Measurement*. 20(2): 155 - 168.

Kelderman, Henk.; & Rijkes, Carl P.M. (1994, June). Loglinear Multidimensional IRT Models for Polytomously Scored Items. *Psychometrika*. 59(2): 149 - 176.

Kenedy, Cathleen A. (2005). *Constructing Measurement Models for MRCML Estimation: A Primer for Using the BEAR Scoring Engine*. Retrieved February 8, 2008, from http://bearcenter.berkeley.edu/publications/ConstructingMeasures.pdf

Kim, J. (2007).*A Comparison of Calibration Methods And Proficiency Estimators For Creating IRT Vertical Scales*. Theses and Dissertations. Retrieved February 8, 2009, from http://etd.lib.uiowa.edu/2007/jkim.pdf

Kim, S. (Winter 2006). A Comparative Study of IRT Fixed Parameter Calibration Methods. *Journal of Educational Measurement*, 43(4): 355–381.

Kim, S. H., & Cohen, A. S. (1992, Spring). Effects of Linking Methods on Detection of DIF. *Journal of Educational Measurement*. 29(1): 51 - 66.

_____. (1996). A Comparison of Linking and Concurrent Calibration under Item Response Theory. Retrieved January, 10, 2011, from http://www.eric.ed.gov/ERICWebPortal/contentdelivery/servlet/ERICServlet?accno=ED400274

_____. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement*, 22, 131–143.

_____. (2002). A comparison of linking and concurrent calibration under the graded response model. *Applied Psychological Measurement*, 26, 25–41.

Kim, S. (2004). *Unidimensional IRT Scale Linking Procedures for Mixed-Format Tests and Their Robustness to Multidimensionality*. Unpublished doctoral dissertation, University of Iowa, Iowa City. Retrieved February 8, 2009, from http://proquest.umicom/pqdweb?index=0&did=765922701&SrchMode=1&sid=1&Fmt=14&VInst=PROD&VType=PQD&RQT=309&VName=PQD&TS=1251015767&clientId=29945

Kim, S.H.; & 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.H. and Lee, W-C. (2004). *IRT Scale Linking Methods for Mixed-Format Tests.* ACT research report series 2004 – 2005. Retrieved June 11, 2008, from http://www.act.org/research/reports/pdf/ACT_RR2004-5.pdf

_____. (2006, March). An Extension of Four IRT Linking Methods for Mixed-Format Tests. *Journal of Educational Measurement*. 43(1): 53-76

Kim, S. H. (2006, April). *A study of IRT fixed parameter calibration methods using BILOG-MG.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Kim, S.; Walker, M.E.; & McHale, F. (2010, Spring). Comparisons among Designs for Equating Mixed-Format Tests in Large-Scale Assessments. Journal of Educational Measurement. 46(1): 36 – 53.

Kinsey, Tari L., (2003). *A Comparison of IRT and Rasch Procedures in a Mixed – Item Format Test*. Doctor of Philosophy (Educational Research), University of North Texas, USA. Retrieved June 20, 2008, from http:// digital.library.unt.edu/permalink/meta-dc-4316:1

Kolen, M. J.; & Brennan, R.L. (1995). *Test equating: Methods and practices*. New York: Springer-Verlag.

_____. (2004). *Test equating, Scaling, and Linking: Methods and Practices*. Springer Science and Business Media, Inc.

Lee, G., & Fitzpatrick, A. (2002, April). *IRT equating using anchor items having fixed c-parameters*. Paper Presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA.

Lee, Won-Chan. (2006). *Multinomial and Compound Multinomial Error Models for Tests With Complex Item Scoring*. Retrieved June 20, 2008, from http://apm.sagepub.com/cgi/content/abstract/31/4/255.

Li, Yuan H.; Tam, Hak P.; & Tompkins, Leory J. (2004). Comparison of Using the Fixed Common-Precalibrated Parameter Method and the Matched Characteristic Curve Method for Linking Multiple-Test Items. *International Journal of Testing*, 4(3), 267 – 293.

Li, Y. H., Griffith, W. D., & Tam, H. P. (1997). *Equating multiple tests via an IRT linking design: Utilizing a single set of anchor items with fixed common item parameters during the calibration process.* Paper presented at the annual meeting of the Psychometric Society, Knoxville, TN.

Li, Y.H.; Lissitz, R. W.; Yang, Y.N. (1999). *Estimating IRT Equating Coefficients for Tests with Polytomously and Dichotomously Scores Items*. Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, Canada, April 19 – 23, 1999) Retrieved Jan 20, 2009, from http://eric.ed.gov/ERICWebPortal/custom/portlets/recordDetails/detailmini.jsp?_nfpb=true&_&ERICExtSearch_SearchValue_0=ED431800&ERICExtSearch_SearchType_0=no&accno=ED431800

Li, Y. & Lissitz, R. (2000, June). An evaluation of the accuracy of multidimensional IRT linking. *Applied Psychological Measurement.* 24(2): 115–138.

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* . Dissertation submitted to the Faculty of the Graduate School of the University of Maryland, College park, in partial fulfillment Of the requirements for the degree of Doctor of Philosophy. Retrieved November 21, 2008, from www.lib.umd.edu/drum/bitstream/1903/8468/1/umi-umd-5508.pdf.

Linn, Robert L.; & Slinde, Jeffrey A. (1977, Winter). The Determination of Significance of Change Between Pre-and Posttesting Periods. *Review of Educational Research.* 47(1): 121 – 150.

Little, Todd D.; Bovaird, James A.;& Slegers, David W. (2005). *Methods for the Analysis of Change.* Retrieved January 21, 2008, from http://www.agencylab.ku.edu/~agencylab/manuscripts/(Little,%20Bovaird,%20and%20Slegers.in%20press).pdf

Lord, Frederic M. (1958). Further Problems in the Measurement of Growth. *Educational and Psychological Measurement.* 18(3): 437 – 451.

Lord, F. N., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading. MA: Addison-Wesley.

Loyd B. H., & Hoover, H.D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 17, 179 -193.

Lukhele, R., Thissen, D., & Wainer, H. (1994). On the relative value of multiple-choice, constructed response, and examinee-selected items on two achievement tests. *Journal of Educational Measurement.* 31(3): 234-250.

Mair, Patrick.;& Hatzinger, Reinhold. (2007). *Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R*. Retrieved January 8, 2008 from, http://epub.wu-wien.ac.at/dyn/virlib/wp/mediate/epub-wu-01_b9f.pdf?ID=epub-wu-01_b9f

Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14, 139-160.

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34, 207-218. Retrieved August 23, 2009 from, http://web.ebscohost.com/ehost/pdf?vid=1&hid=5&sid=ae1eb15c-0a64-4859-925e-d2f76ceab41d%40sessionmgr11

Marvelde te, Janneke M.; et al. (2006, February). Application of Multidimensional Item Response Theory Models to Longitudinal Data. *Educational and Psychological Measurement*. 66(1): 5 - 34.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika,* 47, 149-174.

May,K.;& Nicewander, W.Alan. (1998). Measuring change conventionally and adaptively. *Educational and Psychological Measurement*. 58:529 – 554.

Maydeu-Olivares, Albert. (2000). Review of MPLUS. *Multivariate Behavioral Research*. 35(4): 501 - 504.

McNemar, Quinn. (1958). On Growth Measurement. *Educational and Psychological Measurement.* 18(1): 47 – 55.

Mellenbergh. G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*. 19. 91-100.

Meiser, Thorsten. (1996, December). Loglinear Rasch Models For The Analysis of Stability and Change. *Psychometrika*. 61(4): 629 - 645.

Meiser, Thorsten.; Stern, Elsbeth.;& Langeheine, Rolf. (1998). Latent Change in Discrete Data: Unidimensional, Multidimensional, and Mixture Distribution Rasch Models for the Analysis of Repeated Observations. *Method of Psychological Research Online.* 3(2): 75 - 93. Retrieved September 18, 2007 from http://www.mpr-online.de//

Meng, H., Kolen, M. J. & Lohman, D. (2006, April). *An empirical investigation of IRT scaling methods: How different IRT models, parameter estimation procedures, proficiency estimation methods, and estimation programs affect the results of vertical scaling for the Cognitive Abilities Test.* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Meng, Xiao-Li.; Rosenthal, Robert.; & Rubin, Donald B. (1992). Comparing Correlated Correlation Coefficient. *Psychological Bulletin*. 111(1): 172 – 175.

Min, K. (2003). *The impact of scale dilation on the quality of the linking of multidimensional item response theory calibrations*. PhD thesis, Michigan State University.

Mislevy, R. J. (1986). Bayes modal estimation in item response models. *Psychometrika,* 51, 177-195.

Mislevey, R. J., & Bock, R. D. (1990). *BILOG3: Item analysis and test scoring with binary logistic models (2nd ed.).* Mooresville, IN: Scientific Software, Inc.

Mellenbergh, G. J. (1995). Conceptual notes on models for discrete polytomous item responses. *Applied Psychological Measurement*, 19, 91-100.

Muraki, Eiji. (1992, June). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement*. 16(2): 159 - 176.

Nering, Michael L. (1995, June). The Distribution of Person Fit Using True and Estimated Person Parameters. *Applied Psychological Measurement*. 19(2): 121 - 129.

Oshima, T., Davey, T., & Lee, K (2000). Multidimensional linking: four practical approaches. *Journal of Educational Measurement*, 37(4), 357-373.

Ogasawara, H. (2001). Least squares estimation of item response theory linking coefficients. *Applied Psychological Measurement*, 25, 3-24.

Paek, I., & Young, M. J. (2005). Investigation of student growth recovery in a fixed-item linking procedure with a fixed-person prior distribution for mixed-format test data. *Applied Measurement in Education*, 18, 199–215.

Park, Chang. (1997). *Accuracy of parameter estimation on polytomous IRT models*. Ed.D., University of Massachusetts Amherst. Retrieved November 1, 2008, from http://proquest.umi.com/pqdlink?did=736656651&Fmt=7&clientId=61839&RQT=309&VName=PQD

Park, Hyunjeong. (2000). *Comparison of IRT Models For Ordered Polytomous Response Data*. University of Minnesota. Retrieved November 1, 2008, from http://proquest.umi.com/pqdlink?did=728340421&Fmt=2&clientId=61839&RQT=309&VName=PQD

Pike, Gary R. (1991). *Lies, Damn Lies, and Statistics Revised: A Comparison of Three Methods of Representing Change*. Retrieved January 1, 2008, from http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/23/22/a1.pdf

Ponocny, Ivo. (2002). On the Applicability of some IRT Models for Repeated Measurement Designs: Conditions, Consequences, and Goodness-of-Fit Tests. *Method of Psychological Research Online*. 7(1):21 - 40. Retrieved September 18, 2007 from http://www.mpr-online.de//

Prieler, Jorg A.;& Raven, John. (2002). *The Measurement of Change in Groups and Individuals, with Particular Reference to the Value of Gain Scores: A New IRT-Based Methodology for the Assessment of Treatment Effects and Utilizing Gain Scores*. Retrieved October 5, 2007, from http://www.wpe.info/vault/priravf/prirav.pdf

Prowker, Adam.;& Camilli, Gregory. (2007, Spring). Looking Beyond the Overall Scores

of NAEP Assessments: Applications of Generalized Linear Mixed Modeling for Exploring Value-Added Item Difficulty Effects. *Journal of Educational Measurement.* 44(1): 69 - 87.

Reise, S.P.; & Yu, J. (1990). Parameter recovery in the grade response model using MULTILOG. *Journal of Educational Measurement*. 27(2): 133 -144.

Reynolds, T., Perkins, K., & Brutten, S. (1994). A comparative item analysis study of a language testing instrument. *Applied Psychological Measurement*, 11, 1-13.

Rosa. K., Swygert, K. A., Nelson, L. ; & Thissen. V. (2001). Item response theory applied to combinations of multiple-choice and constructed-response items—Scale scores for patterns of summed scores. In D. Thissen & M. Wainer (Eds.), *Test scoring* (pp. 253-292). Mahwah, NJ: LawrenceErlbaum Associates, Inc.

Rudinger, Georg.;& Rietz, Christian. (1998). The Neglected Time Dimension? Introducing a Longitudinal Model Testing Latent Growth Curves, Stability, and Reliability as Time Bound Processes. *Method of Psychological Research Online*. 3(2): 109 - 123. Retrieved September 18, 2007 from http://www.mpr-online.de//

Russell, M. (2000). *Summarizing Change in Test Scores: Shortcomings of Three Common Methods.* ERIC Digest Retrieved September 18, 2007, from http://www. Ericdigiests.org/2001-2/test.html.

Samejima, F. (1969). *Estimation of latent ability using a response pattern of graded scores*. Psychometric Monograph, No.17. Retrieved May 18, 2009, from http://www.psychometrika.org/journal/online/MN17.pdf

Samsa, Gregory P. (1992, Winter). *Resolution of a Regression Paradox in Pretest – Posttest Designs. Journal of Educational Measurement*. 29(4): 321 - 328.

Schmitt, Neal.; Pulakos, Elaine.;& Lieblein, Amy. (1984, Summer). Comparison of Three Techniques to Assess Group-Level Beta and Gamma Change. *Applied Psychological Measurement.* 8(3):249 -260.

Seong, T-J. (1990). Sensitivity of marginal maximum likelihood estimation of item and ability parameters to the characteristics of the prior ability distributions. *Applied Psychological Measurement*. 14(3). 299–311.

Si, Ching-Fung B. (2002). *Ability Estimation under Different Item Parameterization and Scoring Models.* Doctor of Philosophy (Educational Research) . University of North Texas. Retrieved July 13,2008, from http://digital.library.unt.edu/permalink/meta-dc-3116:1

Si, Ching - Fung.; & Schumacker, R. E. (2004). Ability Estimation under Different Item Parameterization and Scoring Models. *International Journal of Testing,* 4(2). 137-181

Simon, Mayuko Kanada .2008. *Comparison of concurrent and separate multidimensional IRT linking of item parameters*. A Doctor of Philosophy Thesis Submitted to Faculty of The Graduate School of The University of Minnesota. http://conservancy.umn.edu/bitstream/47767/1/Simon_umn_0130E_10065.pdf

Singer, Judith D. (1998, Winter). Using SAS PROC MIXED to Fit Multilevel Models, Hierarchical Models, and Individual Growth Models. *Journal of Educational and Behavioral Statistics*. 23(4): 323 - 355.

Skaggs, G., & Lissitz, R. W. (1988). Effect of examinee ability on test equating invariance. *Applied Psychological Measurement*. 14, 23–32.

Spiel, Christiane.;& Gluck, Judith. (1998). Item Response Models for Assessing Change in Dichotomous Items. *International Journal of Behavioral Development.* 22(3): 517 - 536. Retrieved January 4, 2008, from http://jbd.sagepub.com/cgi/content/abstract/22/3/517

Steyer, Rolf.; Eid, Michael.;& Schwenkmezger, Peter. (1997). Modeling True Intraindividual Change: True Change as Latent Variable. *Method of Psychological Research Online.* 2(1): 21 - 33. Retrieved September 18, 2007 from http://www.mpr-online.de//issue2/art2/steyer.pdf

Stocking, M. L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.

Stoel, Reinoud D.;& Wittenboer, Godfried van den. (2003). Time Dependence of Growth Parameters in Latent Growth Curve Models with Time Invariant Covariates. *Method of Psychological Research Online*. 8(1):21 - 41. Retrieved September 18, 2007 from http://www.mpr-online.de//

Swaminathan, H., & Gifford, J. A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics*. 7: 175-191.

Sykes & Yen. (Fall 2000). The Scaling of Mixed – Item – Format – Tests With The One – Parameter and Two – Parameter Partial Credit Models. *Journal of Educational Measurement.* 37(3): 221-244.

Tae-Je Seong. (1990). Sensitivity of Marginal Maximum Likelihood Estimation of Item and Ability Parameters to the Characteristics of the Prior Ability Distributions. *Applied Psychological Measurement,* 14(3): 299-311.

Tang, K. Linda; & Eignor, Daniel R. (1997). *Concurrent Calibration of Dichotomously and Polytomously Scored TOEFL Item Using IRT Models*. http://www.ets.org/Media/Research/pdf/RR-97-06.pdf

Tate, R. L. (2000). Performance of a proposed method for the linking of mixed format testswith constructed response and multiple choice items. *Journal of Educational Measurement.* 37: 329–346.

Tate, R. L. (2003). Equating for long-term maintenance of mixed format tests containing multiple choice and constructed response items. *Educational and Psychological Measurement*. 63. 893–914.

Tay-Lim, Brenda Siok-Hoon; Harwell, Michael. *Effects of Number of Items and Examinees on Parameter Estimation in Item Response Theory: A Research Synthesis*. Retrieved October 20, 2008, from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/16/74/4e.pdf

Thissen, D., & Steinberg, L. (1986). A taxonomy of item response models. *Psychometrika*. 51(4): 567-577.

Thissen, D. (1991). *Multilog user's guide: Multiple, categorical item analysis and test scoring using item response theory*. Chicago, IL: Scientific Software.

Tong, Ye; & Kolen, Michael J. (April 2007) .Comparisons of Methodologies and Results in Vertical Scaling for Educational Achievement Tests Vertical Scaling Methodologies and Results . *Applied Measurement in Education*. 20(2): 227– 253

Traub, R. E. (1993). *On the equivalence of traits assessed by multiple-choice and constructed- response tests.* In R. E. Bennett & W. C. Ward (Eds.), Construction versus choice in cognitive measurement (pp. 29-44). Hillsdale, NJ: Lawrence Erlbaum.

Traub, R. E. (1994). *Reliability for the Social Sciences: theory and application*. Volume 3. pp. 126 – 137. California: Sage Publication, Inc.

Tutz, G. (1990). Sequential item response models with an ordered response. *British Journal of Mathematical and Statistical Psychology*, 43, 39-55.

Tutz, G. (1997). Sequential models for ordered responses. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of modern item response theory* (pp. 139-152). New York: Springer.

van der Linden, W, J., & Hambleton, R. K. (Eds.). (1997). *Handbook of modern Hem resptmse theory*.New York: Springer.

Wollack, James A.; Jung Sung, Hyun; & Kang, Taehoon. (2005).*Longitudinal Effects of Item Parameter Drift*. Retrieved January 21, 2009, from, http://testing.wisc.edu/NCME%202005%20paper%20(Wollack,%20Sung,%20&%20Kang).pdf

Wainer, H., & Mislevey, R. J. (1990). *Item response theory, item calibration, and proficiency estimation.* In H. Wainer (Ed.), Computerized adaptive testing: A primer (pp.65-102). Hillsdale, NJ: Lawrence Erlbaum.

Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education,* 6(2): 103–118.

Wells, C. S., Subkoviak, M. J., & Serlin, R. C. (2002). The effect of item parameter drift on examinee ability estimates. *Applied Psychological Measurement*, 26(1), 77-87.

Wightman, Lawrence E. & Champlain, Andre F. (1993). *A Comparison of the Properties of IRT Parameter Esimates Using Two Different Calibration Designs*.

Willett, John B.; & Sayer, Aline G. (1994). Quantitative Methods in Psychology; Using Covariance Structure Analysis to Detect Correlates and Predictors of Individual Change Over Time. *Psychological Bulletin*. 116(2): 363 -381.

Williams, Natasha J.;& Beretvas, S. Natasha. (2006, January). DIF Identification Using HGLM for Polytomous Items. *Applied Psychological Measurement*. 30(1): 22 - 42.

Wilson, Mark.; & Masters, Geofferey N. (1993, March). The Partial Credit Model and Null Categories. *Psychometrika.* 58(1): 87 - 99.

Wu, Margaret; & Adams, Ramond. (2006). Modeling Mathematics Problem Solving Item Responses Using a Multidimensional IRT Model. *Mathematics Education Research Journal.* 18(2): 93 -113.Retrived January 21, 2008, from, http://www.merga.net.au/documents/MERJ_18_2_Wu.pdf

Yang, S. (2007). *A comparison of unidimensional and multidimensional rasch models using parameter estimates and fit indices when assumption of unidimensionality is violated* . Doctor of Philosophy, Ohio State University, Educational Policy and Leadership. Retrieved November 1, 2008, from, http://www.ohiolink.edu/etd/view.cgi?acc_num=osu1195695378

Yao, L.; & Mao, X. (2004). *Unidimensional and Multidimensional Estimation of Vertical Scaled Tests with Complex Structure*. Retrieved November 2, 2008, from http://www.ctb.com/media/articles/pdfs/ResearchArticles/Unidimensional.pdf?FOLDER%3C%3Efolder_id=2534374302134982

Yao, L.; & Boughton, K. A. (2007) . A Multidimensional Item Response Modeling Approach for Improving Subscale Proficiency Estimation and Classification. *Applied Psychological Measurement* .31(2), 83 .

Yao, Lihua.; Schwarz, Richard D. (2006, November). A Multidimensional Partial Credit Model With Associated Item and Test Statistics: An Application to Mixed-Format Tests. *Applied Psychological Measurement.* 30(6): 469 - 492.

Yen, W. M. (1993). Scaling performance assessment: Strategies for managing local item dependence. *Journal of Educational Measurement*, 30(3): 187–214.

Yen, W. M.; & Fitzpatrick, A.R. (2001). The Effects of Test Length and Sample Size on the Reliability and Equating of Tests Composed of Constructed-Response Items. *Applied Measurement in Education,* 14(1), 31–57

Yen, W. M.; & Fitzpatrick, A.R. (2006). *Item Response Theory*. Educational Measurement. In Brennan, R.L. (4th ed.pp.111 - 153). Washington, D.C. American Council on Education and Praeger Plublishers.

Yoes, Michael Ernest. (1994). *A Comparison of the Effectiveness of Item Parameter Estimation Techniques Used with The Three Parameter Logistic Item Response Theory Model.* Dissertation Abstracts International. 54(12):6497-B; June.

Yon, Haniza. (2006). Proquest Dissertations And Theses . Section 0128, Part 0288 119 pages; [Ph.D. dissertation].United States -- Michigan: Michigan State University;. Publication Number: AAT 3248633

Zhang, Jinming. (2004). *Comparison of Unidimensional and Multidimensional Approaches to IRT Parameter Estimation*. Retrieved October 21, 2008, from, http://www.ets.org/portal/site/ets/menuitem.c988ba0e5dd572bada20bc47c3921509/?vgnextoid=8548d143d9df4010VgnVCM10000022f95190RCRD&vgnextchannel=dcb3be3a864f4010VgnVCM10000022f95190RCRD

Zimowski,M. F.; Muraki, E.; Mislevy, R. J.; &Bock, R. D. (2003). *BILOG–MG (Version 7.0).* In M. duToit (Ed.), *IRT from SSI: BILOG–MG, MULTILOG, PARSCALE, TESTFACT* (chap. 2, pp. 24–256).Lincolnwood, IL: Scientific Software International.

Zwinderman, A. H.; & van der Wollenberg, A. L. (1990). Robustness of marginal maximum likelihood estimation in the Rasch model. *Applied Psychological Measurement,* 14, 73-81.

ภาคผนวก

ภาคผนวก ก

ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสม

ตาราง 55 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสม ภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก -0.50 และให้คะแนน 3 ระดับ

Item	Model	L = 30:10										L =24:8	L = 15:5
		Test A					Test B						
		a	b/b1	c/b2	b3	b4	a	b/b1	c/b2	b3	b4		
1	3PL	0.887	-2.444	0.022	-	-	1.066	-1.779	0.221	-	-	✓	✓
2	3PL	1.100	-2.262	0.026	-	-	0.825	-1.662	0.179	-	-	✓	✓
3	3PL	0.964	-1.706	0.163	-	-	1.126	-1.322	0.170	-	-	-	-
4	3PL	1.179	-1.499	0.195	-	-	1.052	-1.033	0.059	-	-	✓	✓
5	3PL	0.873	-1.316	0.239	-	-	1.057	-0.764	0.061	-	-	✓	✓
6	3PL	0.879	-1.156	0.243	-	-	0.832	-0.597	0.119	-	-	✓	-
7	3PL	1.138	-0.94	0.029	-	-	0.816	-0.505	0.05	-	-	✓	-
8	3PL	0.996	-0.792	0.134	-	-	1.166	-0.324	0.192	-	-	-	-
9	3PL	0.938	-0.654	0.100	-	-	0.954	-0.151	0.221	-	-	✓	✓
10	3PL	1.117	-0.633	0.029	-	-	0.814	-0.038	0.040	-	-	✓	✓
11	3PL	1.022	-0.375	0.022	-	-	1.024	-0.033	0.106	-	-	✓	-
12	3PL	0.937	-0.298	0.137	-	-	1.080	0.149	0.037	-	-	✓	-
13	3PL	0.852	-0.266	0.100	-	-	1.146	0.270	0.104	-	-	-	-
14	3PL	1.108	-0.183	0.084	-	-	0.809	0.270	0.119	-	-	✓	✓
15	3PL	1.110	0.094	0.168	-	-	0.905	0.410	0.102	-	-	✓	✓
16	3PL	1.143	0.337	0.151	-	-	0.971	0.467	0.057	-	-	✓	-
17	3PL	1.084	0.363	0.071	-	-	0.965	0.829	0.084	-	-	✓	-
18	3PL	1.098	0.628	0.121	-	-	0.862	1.115	0.125	-	-	-	-
19	3PL	1.084	1.112	0.023	-	-	1.097	1.283	0.033	-	-	✓	✓
20	3PL	0.988	1.788	0.087	-	-	1.031	1.428	0.055	-	-	✓	✓
21*	3PL	0.921	-1.899	0.06	-	-	0.921	-1.899	0.06	-	-	✓	✓
22*	3PL	0.880	-1.328	0.226	-	-	0.880	-1.328	0.226	-	-	✓	
23*	3PL	0.891	-1.013	0.123	-	-	0.891	-1.013	0.123	-	-	-	-
24*	3PL	0.875	-0.66	0.108	-	-	0.875	-0.66	0.108	-	-	✓	✓
25*	3PL	1.094	-0.535	0.192	-	-	1.094	-0.535	0.192	-	-	✓	-
26*	3PL	1.090	-0.269	0.014	-	-	1.090	-0.269	0.014	-	-	✓	-
27*	3PL	0.959	-0.045	0.060	-	-	0.959	-0.045	0.060	-	-	✓	✓
28*	3PL	1.162	0.349	0.049	-	-	1.162	0.349	0.049	-	-	✓	✓
29*	3PL	1.159	0.882	0.023	-	-	1.159	0.882	0.023	-	-	✓	✓
30*	3PL	1.152	1.795	0.100	-	-	1.152	1.795	0.100	-	-	-	-
1	GPCM	1.081	-2.633	-1.095	-	-	1.102	-1.198	-0.423	-	-	✓	✓
2	GPCM	1.192	-1.556	-0.271	-	-	0.819	-1.525	0.183	-	-	✓	✓
3	GPCM	0.87	-0.909	-0.075	-	-	1.224	-0.485	-0.071	-	-	-	-
4	GPCM	1.197	-0.925	0.013	-	-	1.081	-0.086	0.201	-	-	✓	-
5	GPCM	1.041	-0.055	0.011	-	-	0.919	-0.086	0.201	-	-	✓	-
6	GPCM	1.101	0.270	0.734	-	-	0.872	-0.284	0.833	-	-	✓	✓
7*	GPCM	0.851	-2.037	-0.517	-	-	0.851	0.056	1.044	-	-	✓	✓
8*	GPCM	1.091	-0.971	0.018	-	-	1.091	-0.471	0.518	-	-	-	-
9*	GPCM	1.226	-1.693	1.171	-	-	1.226	-1.693	1.171	-	-	✓	-
10*	GPCM	1.233	0.125	1.316	-	-	1.233	0.125	1.316	-	-	✓	✓

* หมายถึง ข้อสอบร่วมของแบบทดสอบสองฉบับ, ✓ หมายถึง เลือกใช้ข้อสอบข้อนั้น

ตาราง 56 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสม ภายใต้ความยาว 30:10 24:8 และ15:5 ความยาก 0.00 และให้คะแนน 3 ระดับ

Item	Model	L = 30:10										L =24:8	L = 15:5
		Test A					Test B						
		a	b/b1	c/b2	b3	b4	a	b/b1	c/b2	b3	b4		
1	3PL	0.883	-2.062	0.056	-	-	1.081	-1.049	0.246	-	-	✓	✓
2	3PL	0.969	-1.539	0.132	-	-	0.928	-1.001	0.062	-	-	✓	-
3	3PL	0.940	-1.101	0.115	-	-	1.112	-0.757	0.200	-	-	-	-
4	3PL	0.841	-1.028	0.113	-	-	1.112	-0.529	0.175	-	-	✓	✓
5	3PL	0.868	-0.834	0.015	-	-	0.857	-0.322	0.084	-	-	✓	-
6	3PL	1.114	-0.806	0.058	-	-	0.865	-0.268	0.240	-	-	✓	✓
7	3PL	0.984	-0.585	0.074	-	-	1.077	-0.212	0.192	-	-	✓	✓
8	3PL	0.839	-0.463	0.203	-	-	1.115	0.012	0.081	-	-	-	-
9	3PL	1.114	-0.444	0.124	-	-	0.992	0.276	0.114	-	-	✓	-
10	3PL	0.998	-0.203	0.236	-	-	0.831	0.339	0.222	-	-	✓	✓
11	3PL	0.876	0.037	0.150	-	-	0.959	0.426	0.199	-	-	✓	✓
12	3PL	1.118	0.156	0.061	-	-	1.093	0.603	0.083	-	-	✓	-
13	3PL	0.961	0.396	0.120	-	-	1.123	0.929	0.239	-	-	-	-
14	3PL	0.837	0.675	0.162	-	-	0.943	1.210	0.156	-	-	✓	✓
15	3PL	0.941	0.906	0.240	-	-	0.956	1.378	0.062	-	-	✓	-
16	3PL	1.192	0.945	0.056	-	-	0.826	1.410	0.235	-	-	✓	✓
17	3PL	1.137	1.029	0.086	-	-	1.175	1.640	0.035	-	-	✓	✓
18	3PL	1.040	1.279	0.057	-	-	0.812	1.818	0.189	-	-	-	-
19	3PL	0.877	1.554	0.068	-	-	0.800	1.931	0.066	-	-	✓	-
20	3PL	0.986	1.602	0.238	-	-	1.182	2.088	0.071	-	-	✓	✓
21*	3PL	1.075	-1.376	0.053	-	-	1.075	-1.376	0.053	-	-	✓	✓
22*	3PL	0.975	-0.985	0.165	-	-	0.975	-0.985	0.165	-	-	✓	-
23*	3PL	1.137	-0.749	0.156	-	-	1.137	-0.749	0.156	-	-	-	-
24*	3PL	0.905	-0.459	0.031	-	-	0.905	-0.459	0.031	-	-	✓	✓
25*	3PL	1.021	-0.160	0.043	-	-	1.021	-0.160	0.043	-	-	✓	
26*	3PL	0.849	0.385	0.116	-	-	0.849	0.385	0.116	-	-	✓	✓
27*	3PL	1.004	0.767	0.133	-	-	1.004	0.767	0.133	-	-	✓	✓
28*	3PL	0.886	0.951	0.140	-	-	0.886	0.951	0.140	-	-	-	-
29*	3PL	0.991	1.325	0.144	-	-	0.991	1.325	0.144	-	-	✓	-
30*	3PL	1.182	1.801	0.107	-	-	1.182	1.801	0.107	-	-	✓	✓
1	GPCM	1.131	-1.352	-0.56	-	-	0.824	-1.777	0.500	-	-	✓	-
2	GPCM	1.084	-1.145	-0.672	-	-	0.849	-1.761	1.193	-	-	-	-
3	GPCM	0.896	-0.896	-0.472	-	-	1.190	0.187	0.765	-	-	✓	✓
4	GPCM	1.126	-2.059	1.983	-	-	1.126	0.097	0.941	-	-	✓	✓
5	GPCM	1.144	0.238	0.625	-	-	1.143	0.739	1.425	-	-	✓	✓
6	GPCM	1.100	0.586	2.115	-	-	0.801	1.089	1.379	-	-	✓	-
7*	GPCM	0.817	-0.861	0.273	-	-	0.817	-0.861	0.273	-	-	✓	✓
8*	GPCM	0.893	-0.419	1.052	-	-	0.893	-0.419	1.052	-	-	-	-
9*	GPCM	1.183	0.661	0.726	-	-	1.183	0.661	0.726	-	-	✓	-
10*	GPCM	0.974	0.755	1.038	-	-	0.974	0.755	1.038	-	-	✓	✓

* หมายถึง ข้อสอบร่วมของแบบทดสอบสองฉบับ , ✓ หมายถึง เลือกใช้ข้อสอบข้อนั้น

ตาราง 57 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสม ภายใต้ความยาว 30:10 24:8 และ15:5 ความยากเฉลี่ย 0.50 และให้คะแนน 3 ระดับ

Item	Model	L = 30:10										L =24:8	L = 15:5
		Test A					Test B						
		a	b/b1	c/b2	b3	b4	a	b/b1	c/b2	b3	b4		
1	3PL	0.915	-2.177	0.115	-	-	1.138	-1.311	0.250	-	-	✓	-
2	3PL	0.830	-1.272	0.100	-	-	1.110	-0.664	0.014	-	-	✓	✓
3	3PL	0.930	-0.948	0.010	-	-	0.997	-0.470	0.164	-	-	-	-
4	3PL	0.858	-0.593	0.065	-	-	1.127	-0.125	0.162	-	-	✓	✓
5	3PL	0.890	-0.445	0.234	-	-	1.008	0.108	0.045	-	-	✓	-
6	3PL	0.923	-0.383	0.128	-	-	1.133	0.232	0.173	-	-	✓	✓
7	3PL	1.059	-0.170	0.051	-	-	1.193	0.304	0.021	-	-	✓	✓
8	3PL	1.163	0.155	0.054	-	-	1.123	0.576	0.035	-	-	-	-
9	3PL	1.011	0.182	0.113	-	-	0.916	0.736	0.184	-	-	✓	✓
10	3PL	0.971	0.312	0.238	-	-	1.153	0.883	0.050	-	-	✓	-
11	3PL	0.921	0.480	0.095	-	-	0.81	0.979	0.026	-	-	✓	-
12	3PL	1.113	0.541	0.011	-	-	1.141	1.035	0.090	-	-	✓	✓
13	3PL	0.982	0.781	0.026	-	-	1.042	1.346	0.088	-	-	-	-
14	3PL	1.119	1.042	0.149	-	-	0.888	1.586	0.115	-	-	✓	✓
15	3PL	1.147	1.263	0.117	-	-	1.166	1.714	0.197	-	-	✓	-
16	3PL	0.890	1.467	0.173	-	-	0.885	1.931	0.030	-	-	✓	✓
17	3PL	1.154	1.600	0.174	-	-	1.199	2.260	0.027	-	-	✓	✓
18	3PL	0.861	1.811	0.016	-	-	1.182	2.462	0.063	-	-	-	-
19	3PL	0.819	2.036	0.064	-	-	1.004	2.496	0.122	-	-	✓	✓
20	3PL	1.142	2.926	0.227	-	-	1.062	2.924	0.018	-	-	✓	-
21*	3PL	1.161	-1.097	0.099	-	-	1.161	-1.097	0.099	-	-	✓	-
22*	3PL	0.852	-0.517	0.017	-	-	0.852	-0.517	0.017	-	-	✓	✓
23*	3PL	0.828	-0.262	0.031	-	-	0.828	-0.262	0.031	-	-	-	-
24*	3PL	0.878	0.159	0.134	-	-	0.878	0.159	0.134	-	-	✓	✓
25*	3PL	0.926	0.467	0.119	-	-	0.926	0.467	0.119	-	-	✓	-
26*	3PL	0.820	0.733	0.036	-	-	0.820	0.733	0.036	-	-	✓	✓
27*	3PL	0.980	1.200	0.235	-	-	0.980	1.200	0.235	-	-	✓	✓
28*	3PL	1.154	1.550	0.135	-	-	1.154	1.550	0.135	-	-	-	-
29*	3PL	0.899	1.983	0.109	-	-	0.899	1.983	0.109	-	-	✓	✓
30*	3PL	1.022	2.927	0.153	-	-	1.022	2.927	0.153	-	-	✓	-
1	GPCM	0.944	-0.898	-0.738	-	-	1.191	-0.778	1.049	-	-	✓	✓
2	GPCM	1.064	0.023	0.226	-	-	0.946	0.375	0.929	-	-	✓	-
3	GPCM	0.804	-0.144	0.906	-	-	0.885	0.329	1.403	-	-	-	-
4	GPCM	1.057	0.505	1.064	-	-	0.864	0.201	1.575	-	-	✓	-
5	GPCM	0.814	-0.202	2.087	-	-	0.911	0.647	1.724	-	-	✓	✓
6	GPCM	0.95	-0.229	2.233	-	-	0.822	0.765	3.104	-	-	✓	✓
7*	GPCM	1.075	-0.362	1.020	-	-	1.075	-0.362	1.020	-	-	✓	✓
8*	GPCM	1.009	-0.216	1.394	-	-	1.009	-0.216	1.394	-	-	-	-
9*	GPCM	1.171	0.478	1.158	-	-	1.171	0.478	1.158	-	-	✓	-
10*	GPCM	1.197	0.209	1.712	-	-	1.197	0.209	1.712	-	-	✓	✓

* หมายถึง ข้อสอบร่วมของแบบทดสอบสองฉบับ , ✓ หมายถึง เลือกใช้ข้อสอบข้อนั้น

ตาราง 58 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสม ภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก -0.50 และให้คะแนน 4 ระดับ

Item	Model	L = 30:10										L =24:8	L = 15:5
		Test A					Test B						
		a	b/b1	c/b2	b3	b4	a	b/b1	c/b2	b3	b4		
1	3PL	0.887	-2.444	0.022	-	-	1.066	-1.779	0.221	-	-	✓	✓
2	3PL	1.100	-2.262	0.026	-	-	0.825	-1.662	0.179	-	-	✓	✓
3	3PL	0.964	-1.706	0.163	-	-	1.126	-1.322	0.170	-	-	-	-
4	3PL	1.179	-1.499	0.195	-	-	1.052	-1.033	0.059	-	-	✓	✓
5	3PL	0.873	-1.316	0.239	-	-	1.057	-0.764	0.061	-	-	✓	✓
6	3PL	0.879	-1.156	0.243	-	-	0.832	-0.597	0.119	-	-	✓	-
7	3PL	1.138	-0.94	0.029	-	-	0.816	-0.505	0.050	-	-	✓	-
8	3PL	0.996	-0.792	0.134	-	-	1.166	-0.324	0.192	-	-	-	-
9	3PL	0.938	-0.654	0.100	-	-	0.954	-0.151	0.221	-	-	✓	✓
10	3PL	1.117	-0.633	0.029	-	-	0.814	-0.038	0.04	-	-	✓	✓
11	3PL	1.022	-0.375	0.022	-	-	1.024	-0.033	0.106	-	-	✓	-
12	3PL	0.937	-0.298	0.137	-	-	1.08	0.149	0.037	-	-	✓	-
13	3PL	0.852	-0.266	0.100	-	-	1.146	0.270	0.104	-	-	-	-
14	3PL	1.108	-0.183	0.084	-	-	0.809	0.270	0.119	-	-	✓	✓
15	3PL	1.110	0.094	0.168	-	-	0.905	0.410	0.102	-	-	✓	✓
16	3PL	1.143	0.337	0.151	-	-	0.971	0.467	0.057	-	-	✓	-
17	3PL	1.084	0.363	0.071	-	-	0.965	0.829	0.084	-	-	✓	-
18	3PL	1.098	0.628	0.121	-	-	0.862	1.115	0.125	-	-	-	-
19	3PL	1.084	1.112	0.023	-	-	1.097	1.283	0.033	-	-	✓	✓
20	3PL	0.988	1.788	0.087	-	-	1.031	1.428	0.055	-	-	✓	✓
21*	3PL	0.921	-1.899	0.060	-	-	0.921	-1.899	0.060	-	-	✓	✓
22*	3PL	0.880	-1.328	0.226	-	-	0.880	-1.328	0.226	-	-	✓	
23*	3PL	0.891	-1.013	0.123	-	-	0.891	-1.013	0.123	-	-	-	-
24*	3PL	0.875	-0.660	0.108	-	-	0.875	-0.660	0.108	-	-	✓	✓
25*	3PL	1.094	-0.535	0.192	-	-	1.094	-0.535	0.192	-	-	✓	-
26*	3PL	1.090	-0.269	0.014	-	-	1.090	-0.269	0.014	-	-	✓	-
27*	3PL	0.959	-0.045	0.060	-	-	0.959	-0.045	0.060	-	-	✓	✓
28*	3PL	1.162	0.349	0.049	-	-	1.162	0.349	0.049	-	-	✓	✓
29*	3PL	1.159	0.882	0.023	-	-	1.159	0.882	0.023	-	-	✓	✓
30*	3PL	1.152	1.795	0.100	-	-	1.152	1.795	0.100	-	-	-	-
1	GPCM	1.199	-2.154	-1.088	-0.717	-	1.187	-1.055	-0.315	-0.165	-	✓	✓
2	GPCM	1.051	-3.294	-0.531	1.424	-	1.248	-1.106	-0.254	0.334	-	✓	✓
3	GPCM	0.983	-1.843	-1.213	1.357	-	1.457	-1.268	-0.230	0.650	-	-	-
4	GPCM	1.112	-0.975	-0.941	0.344	-	1.270	-0.862	-0.085	1.179	-	✓	✓
5	GPCM	1.055	-0.662	-0.371	0.270	-	0.916	-0.822	0.687	0.741	-	✓	✓
6	GPCM	1.132	-0.878	0.015	0.432	-	1.270	-0.292	0.235	0.943	-	✓	-
7*	GPCM	1.302	-1.622	-0.823	-0.092	-	1.302	-1.622	-0.823	-0.092	-	✓	✓
8*	GPCM	1.018	-1.122	-0.434	-0.086	-	1.335	-1.122	-0.434	-0.086	-	-	-
9*	GPCM	1.260	-0.778	-0.599	0.143	-	1.260	-0.778	-0.599	0.143	-	✓	-
10*	GPCM	0.908	-1.130	0.987	0.998	-	1.308	-1.130	0.987	0.998	-	✓	✓

* หมายถึง ข้อสอบร่วมของแบบทดสอบสองฉบับ , ✓ หมายถึง เลือกใช้ข้อสอบข้อนั้น

ตาราง 59 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสม ภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก 0.00 และให้คะแนน 4 ระดับ

Item	Model	L = 30:10										L =24:8	L = 15:5
		Test A					Test B						
		a	b/b1	c/b2	b3	b4	a	b/b1	c/b2	b3	b4		
1	3PL	0.883	-2.062	0.056	-	-	1.081	-1.049	0.246	-	-	✓	✓
2	3PL	0.969	-1.539	0.132	-	-	0.928	-1.001	0.062	-	-	✓	-
3	3PL	0.940	-1.101	0.115	-	-	1.112	-0.757	0.200	-	-	-	-
4	3PL	0.841	-1.028	0.113	-	-	1.112	-0.529	0.175	-	-	✓	✓
5	3PL	0.868	-0.834	0.015	-	-	0.857	-0.322	0.084	-	-	✓	-
6	3PL	1.114	-0.806	0.058	-	-	0.865	-0.268	0.240	-	-	✓	✓
7	3PL	0.984	-0.585	0.074	-	-	1.077	-0.212	0.192	-	-	✓	✓
8	3PL	0.839	-0.463	0.203	-	-	1.115	0.012	0.081	-	-	-	-
9	3PL	1.114	-0.444	0.124	-	-	0.992	0.276	0.114	-	-	✓	-
10	3PL	0.998	-0.203	0.236	-	-	0.831	0.339	0.222	-	-	✓	✓
11	3PL	0.876	0.037	0.150	-	-	0.959	0.426	0.199	-	-	✓	✓
12	3PL	1.118	0.156	0.061	-	-	1.093	0.603	0.083	-	-	✓	-
13	3PL	0.961	0.396	0.120	-	-	1.123	0.929	0.239	-	-	-	-
14	3PL	0.837	0.675	0.162	-	-	0.943	1.210	0.156	-	-	✓	✓
15	3PL	0.941	0.906	0.240	-	-	0.956	1.378	0.062	-	-	✓	-
16	3PL	1.192	0.945	0.056	-	-	0.826	1.410	0.235	-	-	✓	✓
17	3PL	1.137	1.029	0.086	-	-	1.175	1.640	0.035	-	-	✓	✓
18	3PL	1.040	1.279	0.057	-	-	0.812	1.818	0.189	-	-	-	-
19	3PL	0.877	1.554	0.068	-	-	0.800	1.931	0.066	-	-	✓	-
20	3PL	0.986	1.602	0.238	-	-	1.182	2.088	0.071	-	-	✓	✓
21*	3PL	1.075	-1.376	0.053	-	-	1.075	-1.376	0.053	-	-	✓	✓
22*	3PL	0.975	-0.985	0.165	-	-	0.975	-0.985	0.165	-	-	✓	-
23*	3PL	1.137	-0.749	0.156	-	-	1.137	-0.749	0.156	-	-	-	-
24*	3PL	0.905	-0.459	0.031	-	-	0.905	-0.459	0.031	-	-	✓	✓
25*	3PL	1.021	-0.160	0.043	-	-	1.021	-0.160	0.043	-	-	✓	
26*	3PL	0.849	0.385	0.116	-	-	0.849	0.385	0.116	-	-	✓	✓
27*	3PL	1.004	0.767	0.133	-	-	1.004	0.767	0.133	-	-	✓	✓
28*	3PL	0.886	0.951	0.140	-	-	0.886	0.951	0.140	-	-	-	-
29*	3PL	0.991	1.325	0.144	-	-	0.991	1.325	0.144	-	-	✓	-
30*	3PL	1.182	1.801	0.107	-	-	1.182	1.801	0.107	-	-	✓	✓
1	GPCM	1.131	-1.352	-0.560	1.131	-	0.824	-1.777	0.500	0.824	-	✓	-
2	GPCM	1.084	-1.145	-0.672	1.084	-	0.849	-1.761	1.193	0.849	-	✓	✓
3	GPCM	0.896	-0.896	-0.472	0.896	-	1.190	0.187	0.765	1.190	-	-	-
4	GPCM	1.126	-2.059	1.983	1.126	-	1.126	0.081	1.552	1.126	-	✓	✓
5	GPCM	1.144	0.238	0.625	1.144	-	0.801	1.089	1.379	0.801	-	✓	-
6	GPCM	1.100	0.586	2.115	1.100	-	1.143	0.739	1.425	1.143	-	✓	✓
7*	GPCM	0.817	-0.861	0.273	0.817	-	0.817	-0.861	0.273	0.817	-	✓	✓
8*	GPCM	0.893	-0.419	1.052	0.893	-	1.126	-0.403	0.441	1.126	-	-	-
9*	GPCM	1.183	0.661	0.726	1.183	-	1.183	0.661	0.726	1.183	-	✓	-
10*	GPCM	0.974	0.755	1.038	0.974	-	0.974	0.755	1.038	0.974	-	✓	✓

* หมายถึง ข้อสอบร่วมของแบบทดสอบสองฉบับ , ✓ หมายถึง เลือกใช้ข้อสอบข้อนั้น

ตาราง 60 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสม ภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก 0.50 และให้คะแนน 4 ระดับ

Item	Model	L = 30:10										L =24:8	L = 15:5
		Test A					Test B						
		a	b/b1	c/b2	b3	b4	a	b/b1	c/b2	b3	b4		
1	3PL	0.915	-2.177	0.115	-	-	1.138	-1.311	0.250	-	-	✓	✓
2	3PL	0.830	-1.272	0.100	-	-	1.110	-0.664	0.014	-	-	✓	-
3	3PL	0.930	-0.948	0.010	-	-	0.997	-0.470	0.164	-	-	-	-
4	3PL	0.858	-0.593	0.065	-	-	1.127	-0.125	0.162	-	-	✓	✓
5	3PL	0.890	-0.445	0.234	-	-	1.008	0.108	0.045	-	-	✓	-
6	3PL	0.923	-0.383	0.128	-	-	1.133	0.232	0.173	-	-	✓	✓
7	3PL	1.059	-0.170	0.051	-	-	1.193	0.304	0.021	-	-	✓	✓
8	3PL	1.163	0.155	0.054	-	-	1.123	0.576	0.035	-	-	-	-
9	3PL	1.011	0.182	0.113	-	-	0.916	0.736	0.184	-	-	✓	-
10	3PL	0.971	0.312	0.238	-	-	1.153	0.883	0.05	-	-	✓	✓
11	3PL	0.921	0.480	0.095	-	-	0.810	0.979	0.026	-	-	✓	✓
12	3PL	1.113	0.541	0.011	-	-	1.141	1.035	0.09	-	-	✓	-
13	3PL	0.982	0.781	0.026	-	-	1.042	1.346	0.088	-	-	-	-
14	3PL	1.119	1.042	0.149	-	-	0.888	1.586	0.115	-	-	✓	✓
15	3PL	1.147	1.263	0.117	-	-	1.166	1.714	0.197	-	-	✓	-
16	3PL	0.890	1.467	0.173	-	-	0.885	1.931	0.030	-	-	✓	✓
17	3PL	1.154	1.600	0.174	-	-	1.199	2.260	0.027	-	-	✓	✓
18	3PL	0.861	1.811	0.016	-	-	1.182	2.462	0.063	-	-	-	-
19	3PL	0.819	2.036	0.064	-	-	1.004	2.496	0.122	-	-	✓	-
20	3PL	1.142	2.926	0.227	-	-	1.062	2.924	0.018	-	-	✓	✓
21*	3PL	1.161	-1.097	0.099	-	-	1.161	-1.097	0.099	-	-	✓	✓
22*	3PL	0.852	-0.517	0.017	-	-	0.852	-0.517	0.017	-	-	✓	-
23*	3PL	0.828	-0.262	0.031	-	-	0.828	-0.262	0.031	-	-	-	-
24*	3PL	0.878	0.159	0.134	-	-	0.878	0.159	0.134	-	-	✓	✓
25*	3PL	0.926	0.467	0.119	-	-	0.926	0.467	0.119	-	-	✓	
26*	3PL	0.820	0.733	0.036	-	-	0.820	0.733	0.036	-	-	✓	✓
27*	3PL	0.980	1.200	0.235	-	-	0.980	1.200	0.235	-	-	✓	✓
28*	3PL	1.154	1.550	0.135	-	-	1.154	1.550	0.135	-	-	-	-
29*	3PL	0.899	1.983	0.109	-	-	0.899	1.983	0.109	-	-	✓	-
30*	3PL	1.022	2.927	0.153	-	-	1.022	2.927	0.153	-	-	✓	✓
1	GPCM	0.839	-1.430	-0.445	0.451	-	0.950	-0.368	0.301	0.641	-	✓	-
2	GPCM	1.071	-1.074	-0.606	0.935	-	0.889	-0.242	0.650	0.822	-	✓	✓
3	GPCM	0.887	-0.799	-0.386	1.089	-	1.103	0.194	0.233	1.073	-	-	-
4	GPCM	1.021	0.410	1.115	1.135	-	0.814	0.211	0.883	1.292	-	✓	✓
5	GPCM	0.931	0.885	1.161	1.293	-	0.962	0.858	1.500	2.291	-	✓	-
6	GPCM	0.863	0.757	1.065	2.401	-	0.965	0.865	2.274	2.604	-	✓	✓
7*	GPCM	0.953	-0.520	0.265	0.664	-	0.953	-0.520	0.265	0.664	-	✓	✓
8*	GPCM	1.110	-0.385	-0.283	1.670	-	1.110	-0.385	-0.283	1.670	-	-	-
9*	GPCM	0.856	-0.327	1.294	1.428	-	0.856	-0.327	1.294	1.428	-	✓	-
10*	GPCM	0.963	-0.330	1.635	2.238	-	0.963	-0.330	1.635	2.238	-	✓	✓

* หมายถึง ข้อสอบร่วมของแบบทดสอบสองฉบับ , ✓ หมายถึง เลือกใช้ข้อสอบข้อนั้น

ตาราง 61 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสมภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก -0.50 และให้คะแนน 5 ระดับ

Item	Model	L = 30:10										L =24:8	L = 15:5
		Test A					Test B						
		a	b/b1	c/b2	b3	b4	a	b/b1	c/b2	b3	b4		
1	3PL	0.887	-2.444	0.022	-	-	1.066	-1.779	0.221	-	-	✓	✓
2	3PL	1.100	-2.262	0.026	-	-	0.825	-1.662	0.179	-	-	✓	✓
3	3PL	0.964	-1.706	0.163	-	-	1.126	-1.322	0.170	-	-	-	-
4	3PL	1.179	-1.499	0.195	-	-	1.052	-1.033	0.059	-	-	✓	✓
5	3PL	0.873	-1.316	0.239	-	-	1.057	-0.764	0.061	-	-	✓	✓
6	3PL	0.879	-1.156	0.243	-	-	0.832	-0.597	0.119	-	-	✓	-
7	3PL	1.138	-0.94	0.029	-	-	0.816	-0.505	0.050	-	-	✓	-
8	3PL	0.996	-0.792	0.134	-	-	1.166	-0.324	0.192	-	-	-	-
9	3PL	0.938	-0.654	0.100	-	-	0.954	-0.151	0.221	-	-	✓	✓
10	3PL	1.117	-0.633	0.029	-	-	0.814	-0.038	0.040	-	-	✓	✓
11	3PL	1.022	-0.375	0.022	-	-	1.024	-0.033	0.106	-	-	✓	-
12	3PL	0.937	-0.298	0.137	-	-	1.08	0.149	0.037	-	-	✓	-
13	3PL	0.852	-0.266	0.100	-	-	1.146	0.270	0.104	-	-	-	-
14	3PL	1.108	-0.183	0.084	-	-	0.809	0.270	0.119	-	-	✓	✓
15	3PL	1.110	0.094	0.168	-	-	0.905	0.410	0.102	-	-	✓	✓
16	3PL	1.143	0.337	0.151	-	-	0.971	0.467	0.057	-	-	✓	-
17	3PL	1.084	0.363	0.071	-	-	0.965	0.829	0.084	-	-	✓	-
18	3PL	1.098	0.628	0.121	-	-	0.862	1.115	0.125	-	-	-	-
19	3PL	1.084	1.112	0.023	-	-	1.097	1.283	0.033	-	-	✓	✓
20	3PL	0.988	1.788	0.087	-	-	1.031	1.428	0.055	-	-	✓	✓
21*	3PL	0.921	-1.899	0.060	-	-	0.921	-1.899	0.060	-	-	✓	✓
22*	3PL	0.880	-1.328	0.226	-	-	0.88	-1.328	0.226	-	-	✓	
23*	3PL	0.891	-1.013	0.123	-	-	0.891	-1.013	0.123	-	-	-	-
24*	3PL	0.875	-0.66	0.108	-	-	0.875	-0.66	0.108	-	-	✓	✓
25*	3PL	1.094	-0.535	0.192	-	-	1.094	-0.535	0.192	-	-	✓	-
26*	3PL	1.090	-0.269	0.014	-	-	1.090	-0.269	0.014	-	-	✓	-
27*	3PL	0.959	-0.045	0.060	-	-	0.959	-0.045	0.060	-	-	✓	✓
28*	3PL	1.162	0.349	0.049	-	-	1.162	0.349	0.049	-	-	✓	✓
29*	3PL	1.159	0.882	0.023	-	-	1.159	0.882	0.023	-	-	✓	✓
30*	3PL	1.152	1.795	0.100	-	-	1.152	1.795	0.100	-	-	-	-
1	GPCM	0.824	-1.963	-1.612	-1.209	0.466	0.824	-1.463	-1.112	-0.709	0.966	✓	✓
2	GPCM	0.844	-1.730	-1.201	-0.870	0.113	0.844	-1.230	-0.701	-0.370	0.613	✓	✓
3	GPCM	0.916	-0.825	-0.822	-0.741	-0.711	0.916	-0.325	-0.322	-0.241	-0.211	-	-
4	GPCM	0.906	-1.734	-1.569	-0.382	1.209	0.906	-1.234	-1.069	0.118	1.709	✓	-
5	GPCM	1.108	-1.257	-1.129	0.130	0.193	1.108	-0.757	-0.629	0.630	0.693	✓	-
6	GPCM	1.172	-1.662	0.603	0.978	1.434	1.172	-1.162	1.103	1.478	1.934	✓	
7*	GPCM	1.400	-1.747	-1.490	-0.481	-0.060	1.400	-1.747	-1.490	-0.481	-0.060	✓	-
8*	GPCM	1.142	-1.010	-0.818	-0.329	0.087	1.142	-1.010	-0.818	-0.329	0.087	-	-
9*	GPCM	0.918	-1.384	-0.133	0.061	0.400	0.918	-1.384	-0.133	0.061	0.400	✓	✓
10*	GPCM	1.200	-0.797	0.014	0.039	0.827	1.454	-0.797	0.014	0.039	0.827	✓	✓

* หมายถึง ข้อสอบร่วมของแบบทดสอบสองฉบับ , ✓ หมายถึง เลือกใช้ข้อสอบข้อนั้น

ตาราง 62 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสม ภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก 0.00 และให้คะแนน 5 ระดับ

Item	Model	L = 30:10										L =24:8	L = 15:5
		Test A					Test B						
		a	b/b1	c/b2	b3	b4	a	b/b1	c/b2	b3	b4		
1	3PL	0.883	-2.062	0.056	-	-	1.081	-1.049	0.246	-	-	✓	✓
2	3PL	0.969	-1.539	0.132	-	-	0.928	-1.001	0.062	-	-	✓	-
3	3PL	0.940	-1.101	0.115	-	-	1.112	-0.757	0.200	-	-	-	-
4	3PL	0.841	-1.028	0.113	-	-	1.112	-0.529	0.175	-	-	✓	✓
5	3PL	0.868	-0.834	0.015	-	-	0.857	-0.322	0.084	-	-	✓	-
6	3PL	1.114	-0.806	0.058	-	-	0.865	-0.268	0.240	-	-	✓	✓
7	3PL	0.984	-0.585	0.074	-	-	1.077	-0.212	0.192	-	-	✓	✓
8	3PL	0.839	-0.463	0.203	-	-	1.115	0.012	0.081	-	-	-	-
9	3PL	1.114	-0.444	0.124	-	-	0.992	0.276	0.114	-	-	✓	-
10	3PL	0.998	-0.203	0.236	-	-	0.831	0.339	0.222	-	-	✓	✓
11	3PL	0.876	0.037	0.15	-	-	0.959	0.426	0.199	-	-	✓	✓
12	3PL	1.118	0.156	0.061	-	-	1.093	0.603	0.083	-	-	✓	-
13	3PL	0.961	0.396	0.120	-	-	1.123	0.929	0.239	-	-	-	-
14	3PL	0.837	0.675	0.162	-	-	0.943	1.210	0.156	-	-	✓	✓
15	3PL	0.941	0.906	0.240	-	-	0.956	1.378	0.062	-	-	✓	-
16	3PL	1.192	0.945	0.056	-	-	0.826	1.410	0.235	-	-	✓	✓
17	3PL	1.137	1.029	0.086	-	-	1.175	1.640	0.035	-	-	✓	✓
18	3PL	1.040	1.279	0.057	-	-	0.812	1.818	0.189	-	-	-	-
19	3PL	0.877	1.554	0.068	-	-	0.800	1.931	0.066	-	-	✓	-
20	3PL	0.986	1.602	0.238	-	-	1.182	2.088	0.071	-	-	✓	✓
21*	3PL	1.075	-1.376	0.053	-	-	1.075	-1.376	0.053	-	-	✓	✓
22*	3PL	0.975	-0.985	0.165	-	-	0.975	-0.985	0.165	-	-	✓	-
23*	3PL	1.137	-0.749	0.156	-	-	1.137	-0.749	0.156	-	-	-	-
24*	3PL	0.905	-0.459	0.031	-	-	0.905	-0.459	0.031	-	-	✓	✓
25*	3PL	1.021	-0.16	0.043	-	-	1.021	-0.16	0.043	-	-	✓	
26*	3PL	0.849	0.385	0.116	-	-	0.849	0.385	0.116	-	-	✓	✓
27*	3PL	1.004	0.767	0.133	-	-	1.004	0.767	0.133	-	-	✓	✓
28*	3PL	0.886	0.951	0.140	-	-	0.886	0.951	0.140	-	-	-	-
29*	3PL	0.991	1.325	0.144	-	-	0.991	1.325	0.144	-	-	✓	-
30*	3PL	1.182	1.801	0.107	-	-	1.182	1.801	0.107	-	-	✓	✓
1	GPCM	0.957	-1.052	-0.708	-0.539	-0.506	0.957	-0.552	-0.208	-0.039	-0.006	✓	✓
2	GPCM	0.818	-2.123	-1.090	0.408	0.663	0.818	-1.623	-0.590	0.908	1.163	✓	-
3	GPCM	1.047	-1.968	0.000	0.676	1.342	1.047	-1.468	0.500	1.176	1.842	-	-
4	GPCM	0.866	-0.803	-0.173	0.785	1.184	0.866	-0.303	0.327	1.285	1.684	✓	✓
5	GPCM	1.081	-0.589	-0.423	1.559	1.936	1.081	-0.089	0.077	2.059	2.436	✓	✓
6	GPCM	1.083	0.026	0.231	1.241	1.832	1.083	0.526	0.731	1.741	2.332	✓	-
7*	GPCM	1.160	-1.039	-0.777	-0.713	1.137	1.160	-1.039	-0.777	-0.713	1.137	✓	✓
8*	GPCM	1.079	-1.008	0.023	0.685	0.808	1.079	-1.008	0.023	0.685	0.808	-	-
9*	GPCM	1.113	-0.277	-0.008	0.385	1.175	1.113	-0.277	-0.008	0.385	1.175	✓	-
10*	GPCM	0.881	-1.003	0.929	1.370	1.580	0.881	-1.003	0.929	1.370	1.580	✓	✓

* หมายถึง ข้อสอบร่วมของแบบทดสอบสองฉบับ , ✓ หมายถึง เลือกใช้ข้อสอบข้อนั้น

ตาราง 63 ค่าพารามิเตอร์ของข้อสอบที่ใช้ในการจำลองข้อมูลผลการตอบข้อสอบของแบบทดสอบผสม ภายใต้เงื่อนไขความยาว 30:10 24:8 และ15:5 ความยาก 0.50 และให้คะแนน 5 ระดับ

Item	Model	L = 30:10										L =24:8	L = 15:5
		Test A					Test B						
		a	b/b1	c/b2	b3	b4	a	b/b1	c/b2	b3	b4		
1	3PL	0.915	-2.177	0.115	-	-	1.138	-1.311	0.250	-	-	✓	-
2	3PL	0.830	-1.272	0.100	-	-	1.110	-0.664	0.014	-	-	✓	✓
3	3PL	0.930	-0.948	0.010	-	-	0.997	-0.470	0.164	-	-	-	-
4	3PL	0.858	-0.593	0.065	-	-	1.127	-0.125	0.162	-	-	✓	✓
5	3PL	0.890	-0.445	0.234	-	-	1.008	0.108	0.045	-	-	✓	-
6	3PL	0.923	-0.383	0.128	-	-	1.133	0.232	0.173	-	-	✓	✓
7	3PL	1.059	-0.170	0.051	-	-	1.193	0.304	0.021	-	-	✓	✓
8	3PL	1.163	0.155	0.054	-	-	1.123	0.576	0.035	-	-	-	-
9	3PL	1.011	0.182	0.113	-	-	0.916	0.736	0.184	-	-	✓	✓
10	3PL	0.971	0.312	0.238	-	-	1.153	0.883	0.050	-	-	✓	-
11	3PL	0.921	0.480	0.095	-	-	0.810	0.979	0.026	-	-	✓	-
12	3PL	1.113	0.541	0.011	-	-	1.141	1.035	0.090	-	-	✓	✓
13	3PL	0.982	0.781	0.026	-	-	1.042	1.346	0.088	-	-	-	-
14	3PL	1.119	1.042	0.149	-	-	0.888	1.586	0.115	-	-	✓	✓
15	3PL	1.147	1.263	0.117	-	-	1.166	1.714	0.197	-	-	✓	-
16	3PL	0.890	1.467	0.173	-	-	0.885	1.931	0.030	-	-	✓	✓
17	3PL	1.154	1.600	0.174	-	-	1.199	2.260	0.027	-	-	✓	✓
18	3PL	0.861	1.811	0.016	-	-	1.182	2.462	0.063	-	-	-	-
19	3PL	0.819	2.036	0.064	-	-	1.004	2.496	0.122	-	-	✓	✓
20	3PL	1.142	2.926	0.227	-	-	1.062	2.924	0.018	-	-	✓	-
21*	3PL	1.161	-1.097	0.099	-	-	1.161	-1.097	0.099	-	-	✓	-
22*	3PL	0.852	-0.517	0.017	-	-	0.852	-0.517	0.017	-	-	✓	✓
23*	3PL	0.828	-0.262	0.031	-	-	0.828	-0.262	0.031	-	-	-	-
24*	3PL	0.878	0.159	0.134	-	-	0.878	0.159	0.134	-	-	✓	✓
25*	3PL	0.926	0.467	0.119	-	-	0.926	0.467	0.119	-	-	✓	-
26*	3PL	0.820	0.733	0.036	-	-	0.820	0.733	0.036	-	-	✓	✓
27*	3PL	0.980	1.200	0.235	-	-	0.980	1.200	0.235	-	-	✓	✓
28*	3PL	1.154	1.550	0.135	-	-	1.154	1.550	0.135	-	-	-	-
29*	3PL	0.899	1.983	0.109	-	-	0.899	1.983	0.109	-	-	✓	✓
30*	3PL	1.022	2.927	0.153	-	-	1.022	2.927	0.153	-	-	✓	-
1	GPCM	1.103	-0.434	-0.009	0.230	0.317	1.103	0.066	0.491	0.730	0.817	✓	✓
2	GPCM	1.121	-1.101	-0.409	0.532	1.062	1.121	-0.601	0.091	1.032	1.562	✓	-
3	GPCM	1.026	-0.519	0.001	1.002	1.120	1.026	-0.019	0.501	1.502	1.620	-	-
4	GPCM	0.863	-1.424	-0.32	1.495	2.842	0.863	-0.924	0.180	1.995	3.342	✓	✓
5	GPCM	1.145	0.231	0.577	0.724	1.874	1.145	0.731	1.077	1.224	2.374	✓	✓
6	GPCM	0.841	0.423	0.677	1.121	1.863	0.841	0.923	1.177	1.621	2.363	✓	-
7*	GPCM	1.003	-0.375	-0.079	0.334	0.957	1.003	-0.375	-0.079	0.334	0.957	✓	✓
8*	GPCM	1.118	-1.314	-0.26	0.796	2.750	1.118	-1.314	-0.26	0.796	2.750	-	-
9*	GPCM	0.884	-0.155	0.303	1.527	1.645	0.884	-0.155	0.303	1.527	1.645	✓	-
10*	GPCM	0.884	0.362	0.948	1.151	1.553	0.884	0.362	0.948	1.151	1.553	✓	✓

* หมายถึง ข้อสอบร่วมของแบบทดสอบสองฉบับ , ✓ หมายถึง เลือกใช้ข้อสอบข้อนั้น

ภาคผนวก ข

คำสั่งสำหรับโปรแกรม MULTILOG เพื่อการประมาณค่าพารามิเตอร์พร้อมกัน

คำสั่งสำหรับโปรแกรม MULTILOG เพื่อการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีขนาด 30:10

กรณีให้คะแนน 3 ระดับ

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F001\F001_01.dat',
      NITEMS = 66, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
>TEST ITEMS = (1(1)20), L3;
>TEST ITEMS = (21(1)26), NO, NC = (3(0)6),HIGH=(3(0)6);
>TEST ITEMS = (27(1)36), L3;
>TEST ITEMS = (37(1)40), NO, NC = (3(0)4),HIGH=(3(0)4);
>TEST ITEMS = (41(1)60), L3;
>TEST ITEMS = (61(1)66), NO, NC = (3(0)6),HIGH=(3(0)6);
>TMATRIX ITEMS=(21(1)26),AK,POLYNOMIAL;
>TMATRIX ITEMS=(37(1)40),AK,POLYNOMIAL;
>TMATRIX ITEMS=(61(1)66),AK,POLYNOMIAL;
>FIX ITEMS=(21(1)26),AK=2, VALUE=0.0;
>FIX ITEMS=(37(1)40),AK=2, VALUE=0.0;
>FIX ITEMS=(61(1)66),AK=2, VALUE=0.0;
>TMATRIX ITEMS=(21(1)26),CK,TRIA;
>TMATRIX ITEMS=(37(1)40),CK,TRIA;
>TMATRIX ITEMS=(61(1)66),CK,TRIA;
>ESTIMATE NCYCLES=100;
>SAVE;
>END ;
4
0129
11111111111111111111111111111111111111111111111111111111111111111111
22222222222222222222222222222222222222222222222222222222222222222222
00000000000000000000333333000000000033330000000000000000000333333
00000000000000000000000000000000000000000000000000000000000000000000
(4A1,1X,I1,1X,66A1)
```

กรณีให้คะแนน 4 ระดับ

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F002\F002_01.dat',
      NITEMS = 66, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
>TEST ITEMS = (1(1)20), L3;
>TEST ITEMS = (21(1)26), NO, NC = (4(0)6),HIGH=(4(0)6);
>TEST ITEMS = (27(1)36), L3;
>TEST ITEMS = (37(1)40), NO, NC = (4(0)4),HIGH=(4(0)4);
>TEST ITEMS = (41(1)60), L3;
>TEST ITEMS = (61(1)66), NO, NC = (4(0)6),HIGH=(4(0)6);
>TMATRIX ITEMS=(21(1)26),AK,POLYNOMIAL;
>TMATRIX ITEMS=(37(1)40),AK,POLYNOMIAL;
>TMATRIX ITEMS=(61(1)66),AK,POLYNOMIAL;
>FIX ITEMS=(21(1)26),AK=(2,3), VALUE=0.0;
>FIX ITEMS=(37(1)40),AK=(2,3), VALUE=0.0;
>FIX ITEMS=(61(1)66),AK=(2,3), VALUE=0.0;
>TMATRIX ITEMS=(21(1)26),CK,TRIA;
>TMATRIX ITEMS=(37(1)40),CK,TRIA;
>TMATRIX ITEMS=(61(1)66),CK,TRIA;
>ESTIMATE NCYCLES=100;
>SAVE;
>END ;
5
01239
11111111111111111111111111111111111111111111111111111111111111111111
22222222222222222222222222222222222222222222222222222222222222222222
00000000000000000000333333000000000033330000000000000000000000333333
00000000000000000000444444000000000044440000000000000000000000444444
00000000000000000000000000000000000000000000000000000000000000000000
(4A1,1X,I1,1X,66A1)
```

กรณีให้คะแนน 5 ระดับ

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F003\F003_01.dat',
      NITEMS = 66, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
>TEST ITEMS = (1(1)20), L3;
>TEST ITEMS = (21(1)26), NO, NC = (5(0)6),HIGH=(5(0)6);
>TEST ITEMS = (27(1)36), L3;
>TEST ITEMS = (37(1)40), NO, NC = (5(0)4),HIGH=(5(0)4);
>TEST ITEMS = (41(1)60), L3;
>TEST ITEMS = (61(1)66), NO, NC = (5(0)6),HIGH=(5(0)6);
>TMATRIX ITEMS=(21(1)26),AK,POLYNOMIAL;
>TMATRIX ITEMS=(37(1)40),AK,POLYNOMIAL;
>TMATRIX ITEMS=(61(1)66),AK,POLYNOMIAL;
>FIX ITEMS=(21(1)26),AK=(2,3,4), VALUE=0.0;
>FIX ITEMS=(37(1)40),AK=(2,3,4), VALUE=0.0;
>FIX ITEMS=(61(1)66),AK=(2,3,4), VALUE=0.0;
>TMATRIX ITEMS=(21(1)26),CK,TRIA;
>TMATRIX ITEMS=(37(1)40),CK,TRIA;
>TMATRIX ITEMS=(61(1)66),CK,TRIA;
>ESTIMATE NCYCLES=100;
>SAVE;
>END ;
6
012349
111111111111111111111111111111111111111111111111111111111111111111
222222222222222222222222222222222222222222222222222222222222222222
000000000000000000003333330000000000333300000000000000000000333333
000000000000000000004444440000000000444400000000000000000000444444
000000000000000000005555550000000000555500000000000000000000555555
000000000000000000000000000000000000000000000000000000000000000000
(4A1,1X,I1,1X,66A1)
```

คำสั่งสำหรับโปรแกรม MULTILOG เพื่อการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีขนาด 24:8

กรณีให้คะแนน 3 ระดับ

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F082\F082_01.dat',
      NITEMS = 53, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
>TEST ITEMS = (1(1)16), L3;
>TEST ITEMS = (17(1)21), NO, NC = (3(0)5),HIGH =(3(0)5);
>TEST ITEMS = (22(1)29), L3;
>TEST ITEMS = (30(1)32), NO, NC = (3(0)3),HIGH =(3(0)3);
>TEST ITEMS = (33(1)48), L3;
>TEST ITEMS = (49(1)53), NO, NC = (3(0)5),HIGH =(3(0)5);
>TMATRIX ITEMS=(17(1)21),AK,POLYNOMIAL;
>TMATRIX ITEMS=(30(1)32),AK,POLYNOMIAL;
>TMATRIX ITEMS=(49(1)53),AK,POLYNOMIAL;
>FIX ITEMS=(17(1)21),AK=2 VALUE=0.0;
>FIX ITEMS=(30(1)32),AK=2 VALUE=0.0;
>FIX ITEMS=(49(1)53),AK=2 VALUE=0.0;
>TMATRIX ITEMS=(17(1)21),CK,TRIA;
>TMATRIX ITEMS=(30(1)32),CK,TRIA;
>TMATRIX ITEMS=(49(1)53),CK,TRIA;
>ESTIMATE NCYCLES=100;
>SAVE;
>END ;
4
0129
11111111111111111111111111111111111111111111111111
22222222222222222222222222222222222222222222222222
00000000000000003333300000000333000000000000000033333
00000000000000000000000000000000000000000000000000000
(4A1,1X,I1,1X,53A1)
```

กรณีให้คะแนน 4 ระดับ

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F083\F083_01.dat',
      NITEMS = 53, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
>TEST ITEMS = (1(1)16), L3;
>TEST ITEMS = (17(1)21), NO, NC = (4(0)5),HIGH=(4(0)5);
>TEST ITEMS = (22(1)29), L3;
>TEST ITEMS = (30(1)32), NO, NC = (4(0)3),HIGH=(4(0)3);
>TEST ITEMS = (33(1)48), L3;
>TEST ITEMS = (49(1)53), NO, NC = (4(0)5),HIGH=(4(0)5);
>TMATRIX ITEMS=(17(1)21),AK,POLYNOMIAL;
>TMATRIX ITEMS=(30(1)32),AK,POLYNOMIAL;
>TMATRIX ITEMS=(49(1)53),AK,POLYNOMIAL;
>FIX ITEMS=(17(1)21),AK=(2,3), VALUE=0.0;
>FIX ITEMS=(30(1)32),AK=(2,3), VALUE=0.0;
>FIX ITEMS=(49(1)53),AK=(2,3), VALUE=0.0;
>TMATRIX ITEMS=(17(1)21),CK,TRIA;
>TMATRIX ITEMS=(30(1)32),CK,TRIA;
>TMATRIX ITEMS=(49(1)53),CK,TRIA;
>ESTIMATE NCYCLES=100;
>SAVE;
>END ;
5
01239
11111111111111111111111111111111111111111111111111111
22222222222222222222222222222222222222222222222222222
00000000000000003333300000000333000000000000000033333
00000000000000004444400000000444000000000000000044444
00000000000000000000000000000000000000000000000000000
(4A1,1X,I1,1X,53A1)
```

กรณีให้คะแนน 5 ระดับ

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F084\F084_01.dat',
      NITEMS = 53, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
>TEST ITEMS = (1(1)16), L3;
>TEST ITEMS = (17(1)21), NO, NC = (5(0)5),HIGH=(5(0)5);
>TEST ITEMS = (22(1)29), L3;
>TEST ITEMS = (30(1)32), NO, NC = (5(0)3),HIGH=(5(0)3);
>TEST ITEMS = (33(1)48), L3;
>TEST ITEMS = (49(1)53), NO, NC = (5(0)5),HIGH=(5(0)5);
>TMATRIX ITEMS=(17(1)21),AK,POLYNOMIAL;
>TMATRIX ITEMS=(30(1)32),AK,POLYNOMIAL;
>TMATRIX ITEMS=(49(1)53),AK,POLYNOMIAL;
>FIX ITEMS=(17(1)21),AK=(2,3,4), VALUE=0.0;
>FIX ITEMS=(30(1)32),AK=(2,3,4), VALUE=0.0;
>FIX ITEMS=(49(1)53),AK=(2,3,4), VALUE=0.0;
>TMATRIX ITEMS=(17(1)21),CK,TRIA;
>TMATRIX ITEMS=(30(1)32),CK,TRIA;
>TMATRIX ITEMS=(49(1)53),CK,TRIA;
>ESTIMATE NCYCLES=100;
>SAVE;
>END ;
6
012349
11111111111111111111111111111111111111111111111111111
22222222222222222222222222222222222222222222222222222
00000000000000003333300000000333000000000000000033333
00000000000000004444400000000444000000000000000044444
00000000000000005555500000000555000000000000000055555
00000000000000000000000000000000000000000000000000000
(4A1,1X,I1,1X,53A1)
```

คำสั่งสำหรับโปรแกรม MULTILOG เพื่อการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบมีขนาด 15:5

กรณีให้คะแนน 3 ระดับ

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F163\F163_01.dat',
      NITEMS = 33, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
>TEST ITEMS = (1(1)10), L3;
>TEST ITEMS = (11(1)13), NO, NC = (3(0)3),HIGH=(3(0)3);
>TEST ITEMS = (14(1)18), L3;
>TEST ITEMS = (19(1)20), NO, NC = (3(0)2),HIGH=(3(0)2);
>TEST ITEMS = (21(1)30), L3;
>TEST ITEMS = (31(1)33), NO, NC = (3(0)3),HIGH=(3(0)3);
>TMATRIX ITEMS=(11(1)13),AK,POLYNOMIAL;
>TMATRIX ITEMS=(19(1)20),AK,POLYNOMIAL;
>TMATRIX ITEMS=(31(1)33),AK,POLYNOMIAL;
>FIX ITEMS=(11(1)13),AK=2, VALUE=0.0;
>FIX ITEMS=(19(1)20),AK=2, VALUE=0.0;
>FIX ITEMS=(31(1)33),AK=2, VALUE=0.0;
>TMATRIX ITEMS=(11(1)13),CK,TRIA;
>TMATRIX ITEMS=(19(1)20),CK,TRIA;
>TMATRIX ITEMS=(31(1)33),CK,TRIA;
>SAVE;
>END ;
4
0129
11111111111111111111111111111111
22222222222222222222222222222222
00000000003330000033000000000333
00000000000000000000000000000000
(4A1,1X,I1,1X,33A1)
```

กรณีให้คะแนน 4 ระดับ

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F164\F164_01.dat',
      NITEMS = 33, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
>TEST ITEMS = (1(1)10), L3;
>TEST ITEMS = (11(1)13), NO, NC = (4(0)3),HIGH=(4(0)3);
>TEST ITEMS = (14(1)18), L3;
>TEST ITEMS = (19(1)20), NO, NC = (4(0)2),HIGH=(4(0)2);
>TEST ITEMS = (21(1)30), L3;
>TEST ITEMS = (31(1)33), NO, NC = (4(0)3),HIGH=(4(0)3);
>TMATRIX ITEMS=(11(1)13),AK,POLYNOMIAL;
>TMATRIX ITEMS=(19(1)20),AK,POLYNOMIAL;
>TMATRIX ITEMS=(31(1)33),AK,POLYNOMIAL;
>FIX ITEMS=(11(1)13),AK=(2,3), VALUE=0.0;
>FIX ITEMS=(19(1)20),AK=(2,3), VALUE=0.0;
>FIX ITEMS=(31(1)33),AK=(2,3), VALUE=0.0;
>TMATRIX ITEMS=(11(1)13),CK,TRIA;
>TMATRIX ITEMS=(19(1)20),CK,TRIA;
>TMATRIX ITEMS=(31(1)33),CK,TRIA;
>ESTIMATE NCYCLES=100;
>SAVE;
>END ;
5
01239
11111111111111111111111111111111
22222222222222222222222222222222
00000000003330000033000000000333
00000000004440000044000000000444
00000000000000000000000000000000
(4A1,1X,I1,1X,33A1)
```

กรณีให้คะแนน 5 ระดับ

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
>PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F165\F165_01.dat',
      NITEMS = 33, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
>TEST ITEMS = (1(1)10), L3;
>TEST ITEMS = (11(1)13), NO, NC = (5(0)3),HIGH=(5(0)3);
>TEST ITEMS = (14(1)18), L3;
>TEST ITEMS = (19(1)20), NO, NC = (5(0)2),HIGH=(5(0)2);
>TEST ITEMS = (21(1)30), L3;
>TEST ITEMS = (31(1)33), NO, NC = (5(0)3),HIGH=(5(0)3);
>TMATRIX ITEMS=(11(1)13),AK,POLYNOMIAL;
>TMATRIX ITEMS=(19(1)20),AK,POLYNOMIAL;
>TMATRIX ITEMS=(31(1)33),AK,POLYNOMIAL;
>FIX ITEMS=(11(1)13),AK=(2,3,4), VALUE=0.0;
>FIX ITEMS=(19(1)20),AK=(2,3,4), VALUE=0.0;
>FIX ITEMS=(31(1)33),AK=(2,3,4), VALUE=0.0;
>TMATRIX ITEMS=(11(1)13),CK,TRIA;
>TMATRIX ITEMS=(19(1)20),CK,TRIA;
>TMATRIX ITEMS=(31(1)33),CK,TRIA;
>ESTIMATE NCYCLES=100;
>SAVE;
>END ;
6
012349
111111111111111111111111111111111
222222222222222222222222222222222
000000000033300000330000000000333
000000000044400000440000000000444
000000000055500000550000000000555
000000000000000000000000000000000
(4A1,1X,I1,1X,33A1)
```

ตัวอย่าง Output จากโปรแกรม MULTILOG ในการประมาณค่าพารามิเตอร์พร้อมกัน เมื่อแบบทดสอบผสมมีขนาด 30:10

```
MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION 7.0.3
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
 >PROBLEM SCORE, INDIVIDUAL, DATA = 'C:\Program Files\MULTILOG\F001\F001_01.dat',
       NITEMS = 66, NGROUPS = 2, NEXAMINEES = 2000, NCHARS = 4;
  DATA FILE NAME IS
       C:\PROGRAM FILES\MULTILOG\F001\F001_01.DAT
 TYPE OF INPUT:
   INDIVIDUAL RESPONSE VECTORS
 >TEST ITEMS = (1(1)20), L3;
 >TEST ITEMS = (21(1)26), NO, NC = (3(0)6),HIGH=(3(0)6);
>TEST ITEMS = (27(1)36), L3;
>TEST ITEMS = (37(1)40), NO, NC = (3(0)4),HIGH=(3(0)4);
 >TEST ITEMS = (41(1)60), L3;
 >TEST ITEMS = (61(1)66), NO, NC = (3(0)6),HIGH=(3(0)6);
 >TMATRIX ITEMS=(21(1)26),AK,POLYNOMIAL;
 >TMATRIX ITEMS=(37(1)40),AK,POLYNOMIAL;
 >TMATRIX ITEMS=(61(1)66),AK,POLYNOMIAL;
 >FIX ITEMS=(21(1)26),AK=2, VALUE=0.0;
 >FIX ITEMS=(37(1)40),AK=2, VALUE=0.0;
 >FIX ITEMS=(61(1)66),AK=2, VALUE=0.0;
 >TMATRIX ITEMS=(21(1)26),CK,TRIA;
 >TMATRIX ITEMS=(37(1)40),CK,TRIA;
 >TMATRIX ITEMS=(61(1)66),CK,TRIA;
 >ESTIMATE NCYCLES=100;
 >SAVE;
 NUMBER OF CODES  4
0129
  VECTOR OF CATEGORIES FOR CODE=0
11111111111111111111111111111111111111111111111111111111111111111111
  VECTOR OF CATEGORIES FOR CODE=1
22222222222222222222222222222222222222222222222222222222222222222222
  VECTOR OF CATEGORIES FOR CODE=2
00000000000000000000333333000000000033330000000000000000000333333
  VECTOR OF CATEGORIES FOR CODE=9
00000000000000000000000000000000000000000000000000000000000000000000
(4A1,1X,I1,1X,66A1)
 MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION 7.0.3
```

```
MULTILOG for Windows 7.00.2327.2
Created on: 23 March 2011, 21:19:12
DATA PARAMETERS:
 NUMBER OF LINES IN THE DATA FILE: 2000
 NUMBER OF CATEGORICAL-RESPONSE ITEMS:  66
 NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS:   2
 TOTAL NUMBER OF "ITEMS" (INCLUDING GROUPS):  68
 NUMBER OF CHARACTERS IN ID FIELDS:  4
 MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM:  4
 THE MISSING VALUE CODE FOR CONTINUOUS DATA:  9.0000
 THE DATA WILL BE STORED IN MEMORY
ESTIMATION PARAMETERS:
  MODAL THETAS WILL BE ESTIMATED
OUTPUT-CONTROL PARAMETERS
 THE MODAL THETAS WILL BE SAVED ON FILE <jobname>.SCO
MULTILOG for Windows 7.00.2327.2
READING DATA...
KEY-
CODE CATEGORY
 0    11111111111111111111111111111111111111111111111111111111111111111
 1    22222222222222222222222222222222222222222222222222222222222222222
 2    00000000000000000000333333000000000033330000000000000000000333333
 9    00000000000000000000000000000000000000000000000000000000000000000
FORMAT FOR DATA-
 (4A1,1X,I1,1X,66A1)
FIRST OBSERVATION AS READ-
 ID    1
ITEMS 21211211111111111111111111121111121111100000000000000000000
      000000
NORML      0.000     9.000
 NORMAL PROGRAM TERMINATION
 START DATE: 03-24-2011
 START TIME: 07:03:42
 END   TIME: 07:03:43
```

Thetahat	SE	ID
-0.138	0.149	1
-0.135	0.150	2
-0.121	0.148	3
-0.116	0.148	4

-0.093	0.147	5
-0.093	0.147	6
-0.093	0.147	7
-0.119	0.147	8
-0.093	0.147	9
-0.116	0.148	10
...	...	...
-1.611	0.561	996
-1.892	0.658	997
-1.222	0.405	998
-1.892	0.658	999
-1.892	0.658	1000
-0.100	0.145	1001
-0.040	0.143	1002
-0.075	0.146	1003
-0.112	0.149	1004
-0.071	0.147	1005
-0.100	0.145	1006
-0.093	0.147	1007
-0.079	0.145	1008
-0.071	0.147	1009
-0.061	0.143	1010
-0.075	0.146	1011
-0.135	0.150	1012
-0.112	0.149	1013
...	...	...
-1.222	0.405	1994
-1.130	0.372	1995
-1.222	0.405	1996
-1.145	0.379	1997
-1.296	0.431	1998
-1.222	0.405	1999
-1.222	0.405	2000

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ ชื่อสกุล	นางสุนีย์ เงินยวง
วันเดือนปีเกิด	16 ตุลาคม 2519
สถานที่เกิด	จังหวัดตาก
สถานที่อยู่ปัจจุบัน	239/36 ตรอกหมู่บ้านสุเทพ มหาวิทยาลัยเชียงใหม่ ต.สุเทพ อ. เมือง จ.เชียงใหม่
ตำแหน่งหน้าที่การงาน	อาจารย์
สถานที่ทำงานปัจจุบัน	คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่
ประวัติการศึกษา	
พ.ศ. 2540	ศษ.บ. (คณิตศาสตร์) เกียรตินิยมอันดับ 2 จาก มหาวิทยาลัยเชียงใหม่
พ.ศ. 2546	ศษ.ม. (คณิตศาสตร์ศึกษา) จาก มหาวิทยาลัยเชียงใหม่
พ.ศ. 2554	กศ.ด. (การทดสอบและวัดผลการศึกษา) จาก มหาวิทยาลัยศรีนครินทรวิโรฒ