การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ

นางอวยพร เรืองตระกูล

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต
สาขาวิชาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยการศึกษา
คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2544
ISBN 974-17-0492-5
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

THE DEVELOPMENT AND ANALYSIS OF THE QUALITY OF GROWTH MEASUREMENT METHODS BASED ON CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY

Mrs. Auyporn Ruengtragul

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Educational Measurement and Evaluation
Department of Educational Research
Faculty of Education
Chulalongkorn University
Academic year 2001

ISBN 974-17-0492-5

หัวข้อวิทยานิพนธ์	การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ
โดย	นางอวยพร เรืองตระกูล
สาขาวิชา	การวัดและประเมินผลการศึกษา
อาจารย์ที่ปรึกษา	รองศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี
อาจารย์ที่ปรึกษาร่วม	ศาสตราจารย์กิตติคุณ ดร.นงลักษณ์ วิรัชชัย

คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้นับวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาดุษฎีบัณฑิต

..... คณบดีคณะครุศาสตร์
(รองศาสตราจารย์ ดร.ไพฑูรย์ สินลารัตน์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ศาสตราจารย์ ดร.สำเริง บุญเรืองรัตน์)

..... อาจารย์ที่ปรึกษา
(รองศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี)

..... อาจารย์ที่ปรึกษาร่วม
(ศาสตราจารย์กิตติคุณ ดร.นงลักษณ์ วิรัชชัย)

..... กรรมการ
(รองศาสตราจารย์ ดร.ดิเรก ศรีสุโข)

..... กรรมการ
(รองศาสตราจารย์ ดร.ทวีวัฒน์ ปิตยานนท์)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

อวยพร เรืองตระกูล: การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ (THE DEVELOPMENT AND ANALYSIS OF THE QUALITY OF GROWTH MEASUREMENT METHODS BASED ON CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY) อาจารย์ที่ปรึกษา: รองศาสตราจารย์ ดร. ศิริชัย กาญจนวาสี, อาจารย์ที่ปรึกษาร่วม: ศาสตราจารย์ กิตติคุณ ดร.นงลักษณ์ วิรัชชัย, 259 หน้า. ISBN 974-17-0492-5.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาและวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ 9 วิธี ประกอบด้วย วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธีได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ 2) วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน 3) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ 4) วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ และ 5) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพ วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง และ 2) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง และวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา 2 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา และ 2) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 2 กลุ่ม กลุ่มละ 698 คน และ 637 คน จากโรงเรียนสังกัดกรมสามัญศึกษาเขตกรุงเทพมหานคร 12 โรงเรียน เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับที่คู่ขนานกัน ข้อมูลที่ใช้ในการวิจัยมี 3 ชุด สองชุดแรกเป็นข้อมูลปฐมภูมิ ที่ได้จากการเก็บรวบรวมข้อมูลด้วยแบบทดสอบคณิตศาสตร์ที่ทำการวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบฉบับเดิม และแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน ส่วนข้อมูลชุดที่สาม เป็นข้อมูลทุติยภูมิที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 ที่มีการวัดซ้ำ 5 ครั้ง มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) การวิเคราะห์ข้อมูลเป็นการประมาณค่าคะแนนพัฒนาการ และการวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการในด้านความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนด้วยการวิเคราะห์โมเดลเชิงเส้นตรงระดับลดหลั่นและสถิติทดสอบ Hotelling

ผลการวิจัยสรุปได้ดังนี้

1) เมื่อเปรียบเทียบผลการประมาณค่าคะแนนพัฒนาการระหว่าง 3 กลุ่มวิธี พบว่า กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีคุณภาพไม่แตกต่างกัน

2) เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการทดสอบแบบดั้งเดิม พบว่า วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ และวิธีการวัดคะแนนพัฒนาการสัมพัทธ์มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูล 2 ชุดแรก และวิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูลชุดที่ 3

3) เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการตอบสนองข้อสอบ พบว่า วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลงมีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง

4) เมื่อเปรียบเทียบระหว่างวิธีภายในกลุ่มทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียนมีคุณภาพสูงกว่าวิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค

ภาควิชา วิจัยการศึกษา
สาขาวิชา การวัดและประเมินผลการศึกษา
ปีการศึกษา 2544

ลายมือชื่อนิสิต .....
ลายมือชื่ออาจารย์ที่ปรึกษา .....
ลายมือชื่ออาจารย์ที่ปรึกษาร่วม.....

4084934127: MAJOR EDUCATIONAL MEASUREMENT AND EVALUATION
KEY WORD: GROWTH SCORE/LATENT GROWTH MODEL/ MULTIDIMENSIONAL LATENT TRAIT MODEL FOR MEASURING LEARNING AND CHANGE/ GROWTH SCORE BASED ON CLASSICAL TEST THEORY/ GROWTH SCORE BASED ON ITEM RESPONSE THEORY.
AUYPORN RUENGTRAGUL: THE DEVELOPMENT AND ANALYSIS OF THE QUALITY OF GROWTH MEASUREMENT METHODS BASED ON CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY.
DISSERTATION ADVISOR: ASSOC. PROF. SIRICHAI KANJANAWASEE, Ph.D.,
DISSERTATION CO - ADVISOR: EMERITUS PROF. NONGLAK WIRATCHAI, Ph.D., 259 pp.
ISBN 974-17-0492-5

The purposes of this research were to develop, analyze, and compare the quality of the nine methods measuring growth. The first five growth measurement methods based on the classical test theory were 1) observed difference score method, 2) standard growth method, 3) natural logarithm of growth method, 4) relative growth method, and 5) residual growth and potential ratio method. The other two growth measurement methods based on the item response theory were 1) method of difference between true ability scores, and 2) multidimensional Rasch model for learning and change method. The last two growth measurement methods developed by the researcher were 1) relative true ability of growth and potential ratio method, and 2) relative true ability of growth and initial ability ratio method. The two samples consisted of 698 and 637 Mathayom Suksa 3 students, in 12 schools under the Department of General Education, Bangkok Metropolis. The research instruments were 3 sets of parallel mathematics tests for Mathayom Suksa 3 students. There were 3 data sets for this research. The first two sets were primary data obtaining from three - time - point repeated measures of mathematics achievement, using the same tests and using parallel tests with latin - square design . The third data set was secondary data obtaining from five - time - point measurement of mathematics achievement at Mathayom Suksa 2 level. All three data sets had been scored employing dichotomous and partial credit scoring approaches. Data were analyzed in order to estimate growth scores, and the scores quality using criterion - related validity and measurement error were compared by means of hierarchical linear model and Hotelling test.

The research results indicated that:

1) Comparing the growth score estimations across 3 theoretical approaches, there were no significance differences among the classical test theory approach, the item response theory approach and the researcher' s development approach.

2) The comparison among methods within the classical test theory approach revealed that the methods of observed difference score and the relative growth method had better quality than other methods for the first 2 data sets; and that the residual growth and potential ratio method had better quality than other methods for the third data set.

3) The comparison between methods within the item response theory approach revealed that the method of multidimensional Rasch model for learning and change had better quality than the method of difference between true ability scores.

4) The comparison between methods within the researcher's development approach revealed that the relative true ability of growth and initial ability ratio method had better quality than the relative true ability of growth and potential ratio method for the dichotomous data sets.

Department	Educational research	Student's signature .....
Field of study	Educational Measurement and Evaluation	Advisor's signature .....
Academic year	2001	Co - Advisor's signature .....

กิตติกรรมประกาศ

วิทยานิพนธ์เล่มนี้สำเร็จได้ก็ด้วยความกรุณาอย่างสูงยิ่งจากอาจารย์ที่ปรึกษาวิทยานิพนธ์ รองศาสตราจารย์ ดร.ศิริชัย กาญจนวาสี และอาจารย์ที่ปรึกษาวิทยานิพนธ์ร่วม ศาสตราจารย์กิตติคุณ ดร.นงลักษณ์ วิรัชชัย ที่ได้ให้ทั้งความรู้ในหลักวิชามากมายอันทรงคุณค่ายิ่ง คำแนะนำที่สูงค่าที่เป็นประโยชน์ตั้งแต่ในวิชาสัมมนาซึ่งทำให้ผู้วิจัยได้หัวข้อในการทำวิทยานิพนธ์ ทั้งยังได้ดูแลเอาใจใส่คอยติดตามความก้าวหน้าของผู้วิจัยตลอดมา รวมทั้งเป็นต้นแบบที่ล้ำค่ายิ่งของการทำหน้าที่อาจารย์ที่ปรึกษาวิทยานิพนธ์ ผู้วิจัยรู้สึกซาบซึ้งและขอกราบขอบพระคุณเป็นอย่างยิ่งด้วยความเคารพมา ณ ที่นี้ด้วย

ขอกราบขอบพระคุณคณะกรรมการสอบวิทยานิพนธ์ทุกท่าน ที่ได้กรุณาสละเวลาอันมีค่ามาร่วมสอบวิทยานิพนธ์ ตลอดจนให้ข้อเสนอแนะที่เป็นประโยชน์ต่อการทำวิทยานิพนธ์ฉบับนี้

ขอกราบขอบพระคุณคณาจารย์ภาควิชาวิจัยทุกท่าน ที่เป็นผู้ประสิทธิ์ประสาทวิชาความรู้ทั้งปวงให้แก่ผู้วิจัย และยังให้การสนับสนุนให้ผู้วิจัยเรียนต่อในระดับปริญญาเอก ตลอดจนคอยติดตามให้กำลังใจตลอดระยะเวลาของการศึกษา โดยเฉพาะอย่างยิ่งศาสตราจารย์กิตติคุณ ดร.นงลักษณ์ วิรัชชัย ศาสตราจารย์ดร. สมหวัง พิธิยานุวัฒน์ และศาสตราจารย์ดร. อุทุมพร จามรมาน ที่เป็นผู้กระตุ้นให้ผู้วิจัยเรียนต่อ ซึ่งผู้วิจัยรู้สึกซาบซึ้งในพระคุณเป็นอย่างยิ่ง

ขอขอบคุณรองศาสตราจารย์ ดร.สุวิมล ว่องวาณิช ที่ห่วงใย ให้กำลังใจ และช่วยเหลือตลอดมา

ขอขอบคุณบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย และทบวงมหาวิทยาลัย ที่ให้ทุนอุดหนุนการวิจัย

ขอขอบคุณ คุณสมถวิล วิจิตรวรรณา ที่ให้ความอนุเคราะห์ให้ใช้ฐานข้อมูลจากงานวิจัยมาเป็นข้อมูลในการวิจัยครั้งนี้ และขอขอบคุณนักเรียนชั้นมัธยมศึกษาปีที่ 3 และนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 โรงเรียนในสังกัดกรมสามัญศึกษา ในเขตกรุงเทพมหานคร จำนวน 18 โรงเรียน ที่เป็นกลุ่มตัวอย่างในการวิจัยครั้งนี้ ตลอดจนอาจารย์ในโรงเรียนทั้ง 18 โรงที่ให้ความร่วมมือในการเก็บรวบรวมข้อมูล

อวยพร เรืองตระกูล

สารบัญ

หน้า

บทคัดย่อภาษาไทย ..... ง

บทคัดย่อภาษาอังกฤษ ..... จ

กิตติกรรมประกาศ ..... ฉ

สารบัญ ..... ช

สารบัญตาราง ..... ฌ

สารบัญแผนภาพ ..... ด

บทที่

1 บทนำ 1

ความเป็นมาและความสำคัญของปัญหา 1

วัตถุประสงค์ของการวิจัย 12

สมมติฐานในการวิจัย 13

ข้อตกลงเบื้องต้น 15

ขอบเขตการวิจัย 15

นิยามศัพท์เฉพาะ 17

ประโยชน์ที่คาดว่าจะได้รับ 19

2 เอกสารและงานวิจัยที่เกี่ยวข้อง 21

ตอนที่ 1 การวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม 21

ตอนที่ 2 การวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ 37

ตอนที่ 3 การเปรียบเทียบวิธีการวัดคะแนนพัฒนาการระหว่างวิธีการวัดตามทฤษฎีการทดสอบแบบดั้งเดิม กับทฤษฎีการตอบสนองข้อสอบ 55

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง 69

ตอนที่ 5 การพัฒนาวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ 78

3 วิธีดำเนินการศึกษา 82

ตอนที่ 1 การพัฒนาวิธีการวัดคะแนนพัฒนาการ 82

ตอนที่ 2 การสร้างและพัฒนาเครื่องมือ 85

ตอนที่ 3 การตรวจสอบคุณสมบัติของแบบทดสอบ 95

สารบัญ(ต่อ)

หน้า

ตอนที่ 4 ประชากรและกลุ่มตัวอย่าง 99

ตอนที่ 5 การเก็บรวบรวมข้อมูล 102

ตอนที่ 6 การวิเคราะห์ข้อมูล 103

4 ผลการวิเคราะห์ข้อมูล 112

ตอนที่ 1 ผลการวิเคราะห์คะแนนที่ได้จากการวัดเบื้องต้น. 115

1.1 ผลการวิเคราะห์คะแนนที่ได้จากการวัดด้วยค่าสถิติบรรยาย 115

1.2 ผลการทดสอบเงื่อนไข sphericity ของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดระหว่างเวลา 124

1.3 ผลการทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดระหว่างเวลา 126

ตอนที่ 2 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลพัฒนาการที่มีตัวแปรแฝงกับข้อมูลเชิงประจักษ์ และประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์สำหรับใช้เป็นคะแนนเกณฑ์ 132

2.1 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ 132

2.2 ผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียน.....137

ตอนที่ 3 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลของแต่ละกลุ่มวิธีและวิธีที่ใช้เป็นเกณฑ์ 144

3.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) 144

3.2 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) 147

3.3 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลที่ได้จากการวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ...150

3.4 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลที่ได้จากการวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน(equivalence 2) 153

สารบัญ(ต่อ)

หน้า

ตอนที่ 4 ผลการวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการ 155

4.1 ผลการวิเคราะห์ค่าความตรงตามเกณฑ์สัมพันธ์ 155

4.2 ผลการวิเคราะห์ค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการ 181

ตอนที่ 5 ผลการเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ 184

5.1 ผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ของวิธีการวัดคะแนนพัฒนาการ 184

5.2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการ 185

5 สรุป อภิปรายผล และข้อเสนอแนะ 192

สรุปผลการวิจัย 194

อภิปรายผลการวิจัย. 199

ข้อเสนอแนะ 205

รายการอ้างอิง 209

ภาคผนวก 214

ภาคผนวก ก รายชื่อผู้เชี่ยวชาญ 215

ภาคผนวก ข รายชื่อโรงเรียนกลุ่มตัวอย่าง 217

ภาคผนวก ค ตัวอย่างคำสั่งและผลการวิเคราะห์ข้อมูลโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงด้วยโปรแกรม LISREL 8.30 for student.....219

ภาคผนวก ง ตัวอย่างคำสั่งและผลการวิเคราะห์ข้อมูลเพื่อประมาณค่าความสามารถที่แท้จริงของข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาคด้วยโปรแกรม BILOG 3 238

ภาคผนวก จ ตัวอย่างคำสั่งและผลการวิเคราะห์ข้อมูลเพื่อประมาณค่าความสามารถที่แท้จริงของข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วนด้วยโปรแกรม PARSCALE 2 252

ภาคผนวก ฉ ตัวอย่างผลการวิเคราะห์ข้อมูลเพื่อประมาณค่าความสามารถที่แท้จริงของข้อมูลด้วยโมเดล MRMLC ด้วยโปรแกรม LPCM - WIN 1.0 255

ประวัติผู้วิจัย 259

สารบัญตาราง

ตารางที่ **หน้า**

2.1 สรุปคุณสมบัติของโมเดลที่วัดคะแนนพัฒนาการ 41

2.2 โครงสร้างข้อมูลที่ใช้เก็บข้อมูล 3 ครั้ง..... 46

2.3 สรุปผลการเปรียบเทียบวิธีการวัดคะแนนพัฒนาการระหว่างวิธีการวัดในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ 55

2.4 งานวิจัยที่เกี่ยวกับการวัดคะแนนพัฒนาการ 77

2.5 ค่า Vθ และ Sθ ที่มีค่า θ_2, θ_1 ต่างกัน 81

3.1 ลักษณะข้อมูล 3 ชุด และการใช้ข้อมูลแต่ละชุดในการประมาณค่าคะแนนพัฒนาการ...85

3.2 ค่าความสัมพันธ์ของคะแนนความรู้บางส่วนระหว่างวิธีการให้คะแนนแบบอัตนัยแบบอาร์โนลด์ (Arnold) และแบบอัตนัยประยุกต์ 89

3.3 จำนวนกลุ่มตัวอย่างที่ใช้ในการทดลองแบบทดสอบแต่ละฉบับจำแนกตามโรงเรียน.... 94

3.4 ค่าสถิติพื้นฐานของผลการวิเคราะห์รายข้อของแบบทดสอบคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 3 จำนวน 4 ฉบับ 94

3.5 คุณภาพของแบบทดสอบคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับ..... 95

3.6 ค่า eigen ร้อยละของความแปรปรวน ความแปรปรวนสะสมของแบบทดสอบจำแนกตามวิธีการให้คะแนน 96

3.7 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันของแบบทดสอบทั้ง 3 ฉบับจำแนกตามวิธีการตรวจให้คะแนน 96

3.8 ค่า eigen ร้อยละของความแปรปรวน ความแปรปรวนสะสมของแบบทดสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค และแบบความรู้บางส่วน จำแนกตามแบบทดสอบ ..97

3.9 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันของแบบทดสอบทั้ง 5 ฉบับ จำแนกตามวิธีการตรวจให้คะแนน 98

3.10 การทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนจากแบบทดสอบทั้ง 3 ฉบับ.....99

3.11 การทดสอบความแตกต่างของความแปรปรวนของแบบสอบทั้ง 3 ฉบับ.....99

3.12 จำนวนของนักเรียนที่เป็นกลุ่มตัวอย่างที่ใช้ในการวิจัยจำแนกตามข้อมูล 3 ชุด.....102

3.13 จำนวนครั้งในการวัดที่นำมาประมาณค่าอัตราพัฒนาการของผู้สอบแต่ละคนเพื่อใช้เป็น เกณฑ์ โดยจำแนกตามชุดของข้อมูลและแบบการให้คะแนน 106

สารบัญตาราง (ต่อ)

ตารางที่ **หน้า**

4.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม (repeat) ที่ได้จากการทดสอบ 3 ครั้งจำแนกตามวิธีการตรวจให้คะแนน 117

4.2 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน(counterbalance) ที่ได้จากการทดสอบ 3 ครั้ง จำแนกตามวิธีการตรวจให้คะแนน 120

4.3 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence) ที่ได้จากการทดสอบ จำแนกตามวิธีการตรวจให้คะแนน 123

4.4 ผลการทดสอบ sphericity ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์จำแนกตามจำนวนครั้งของการทดสอบ และแบบทดสอบที่ใช้ 125

4.5 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ระหว่างเวลาของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม (repeat) 126

4.6 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม (repeat) 126

4.7 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ระหว่างเวลาของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) 127

4.8 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) 128

4.9 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) 129

4.10 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) 129

4.11 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) 130

4.12 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) 130

สารบัญตาราง (ต่อ)

ตารางที่ **หน้า**

4.13 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลพัฒนาการที่มีตัวแปรแฝงของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม(repeat).....133

4.14 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลพัฒนาการที่มีตัวแปรแฝงของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) 134

4.15 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลพัฒนาการที่มีตัวแปรแฝงของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) 135

4.16 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลพัฒนาการตัวแปรแฝงของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2).....136

4.17 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม (repeat) ที่มีการให้คะแนนแบบทวิวิภาค(dichotomous).....145

4.18 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม (repeat) ที่มีการให้คะแนนความรู้บางส่วน (partial credit).....147

4.19 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการให้คะแนนแบบทวิวิภาค (dichotomous) 148

4. 20 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการให้คะแนนความรู้บางส่วน (partial credit) 150

4. 21 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลชุดวัด3 ครั้ง ซ้ำด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการให้คะแนนแบบทวิวิภาค (dichotomous) 151

4.22 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการให้คะแนนความรู้บางส่วน (partial credit) 153

สารบัญตาราง (ต่อ)

ตารางที่ **หน้า**

4. 23 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการให้คะแนนแบบทวิวิภาค (dichotomous) 154

4.24 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการที่ได้จากวิธีต่างๆกับคะแนพัฒนาการที่เป็นเกณฑ์ จำแนกตามกลุ่มวิธีการวัด ชุดของข้อมูลและวิธีการตรวจให้คะแนน.....156

4.25 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของวิธีการวัดต่างๆกับคะแนนพัฒนาการที่เป็นเกณฑ์ จำแนกตามลักษณะของข้อมูล วิธีการตรวจให้คะแนน 160

4.26 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของวิธีการวัดต่างๆกับคะแนนพัฒนาการที่เป็นเกณฑ์ ของข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการให้คะแนนความรู้บางส่วน (partial credit) จำแนกตามกลุ่มวิธีการวัด 162

4.27 ค่าสถิติ t ที่ใช้ในการเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์รายคู่ของคะแนนพัฒนาการของข้อมูลชุดวัดซ้ำ 3 ครั้ง (equivalence 1) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) 163

4.28 ผลการเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์รายคู่ของคะแนนพัฒนาการจำแนกตามวิธีการวัด ชุดของข้อมูลและวิธีการตรวจให้คะแนน 165

4.29 ค่าสหสัมพันธ์ระหว่างอันดับของคะแนนพัฒนาการที่ได้จากวิธีต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ จำแนกตามกลุ่มวิธีการวัด ชุดของข้อมูลและวิธีการตรวจให้คะแนน 167

4.30 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสหสัมพันธ์อันดับของอันดับคะแนนพัฒนาการของวิธีการวัดต่างๆกับอันดับคะแนนพัฒนาการที่เป็นเกณฑ์ จำแนกตามลักษณะของข้อมูล วิธีการตรวจให้คะแนน 170

4.31 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการให้คะแนนความรู้บางส่วน (partial credit) จำแนกตามกลุ่มวิธีการวัด 173

สารบัญตาราง (ต่อ)

ตารางที่ **หน้า**

4.32 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ ของข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการให้คะแนนความรู้บางส่วน (partial credit) จำแนกตามกลุ่มวิธีการวัด 175

4.33 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ ของข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการให้คะแนนแบบทวิวิภาค (dichotomous) จำแนกตามกลุ่มวิธีการวัด .....

177

4.34 ค่าสถิติ t ในการทดสอบความแตกต่างของค่าสหสัมพันธ์อันดับเป็นรายคู่ของคะแนนพัฒนาการ จำแนกตามชุดของข้อมูลวัดซ้ำ.....178

4.35 ผลการเปรียบเทียบค่าสหสัมพันธ์อันดับเป็นรายคู่ของคะแนนพัฒนาการ จำแนกตามวิธีการวัด ชุดของข้อมูลและวิธีการตรวจให้คะแนน 180

4.36 ค่าความลำเอียงและค่ารากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน จำแนกตามชุดของข้อมูลและวิธีการให้คะแนน 183

5.1 สรุปผลการเปรียบเทียบคุณภาพของคะแนนพัฒนาการระหว่างกลุ่มวิธีการวัดและภายในกลุ่มวิธีการวัด จำแนกตามชุดของข้อมูล และวิธีการตรวจให้คะแนน 199

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญแผนภาพ

แผนภาพที่ **หน้า**

2.1 โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงในการวิเคราะห์ข้อมูลระยะยาว 34

2.2 ลักษณะของโมเดลพัฒนาการตัวแปรแฝง 36

2.3 โมเดล MRMLC 43

2.4 ความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบถูกกับความสามารถที่เพิ่มขึ้น 79

4.1 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม จากการวัด 3 ครั้ง ที่มีการตรวจให้คะแนนแบบทวิวิภาค 118

4.2 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม ที่มีการตรวจให้คะแนนความรู้บางส่วน 118

4.3 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน จากการวัด 3 ครั้ง ที่มีการตรวจให้คะแนนแบบทวิวิภาค 121

4.4 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน จากการวัด 3 ครั้ง ที่มีการตรวจให้คะแนนความรู้บางส่วน 121

4.5 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน จากการวัด 5 ครั้ง ที่มีการตรวจให้คะแนนแบบทวิวิภาค 124

4.6 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน จากการวัด 5 ครั้ง ที่มีการตรวจให้คะแนนความรู้บางส่วน 124

4.7 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์เฉลี่ยของนักเรียนในการวัด 3 ครั้ง ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม ที่มีการตรวจให้คะแนนแบบทวิวิภาค 127

4.8 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์เฉลี่ยของนักเรียนในการวัด 3 ครั้ง ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม ที่มีการตรวจให้คะแนนความรู้บางส่วน 127

4.9 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนในการวัด 3 ครั้ง ของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนานที่มีการตรวจให้คะแนนแบบทวิวิภาค 128

4.10 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนในการวัด 3 ครั้ง ของชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน ที่มีการตรวจให้คะแนนความรู้บางส่วน 128

สารบัญแผนภาพ(ต่อ)

แผนภาพที่ **หน้า**

4.11 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนที่ได้จากข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน ที่มีการตรวจคะแนนแบบทวิวิภาค 129

4.12 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนที่ได้จากข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน ที่มีการตรวจให้คะแนนความรู้บางส่วน ..... .129

4.13 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนที่ได้จากข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) 131

4.14 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนที่ได้จากการวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) 131

4.15 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำด้วยแบบทดสอบเดิม (repeat)และมีการให้คะแนนแบบทวิวิภาค (dichotomous) 137

4.16 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำด้วยแบบทดสอบเดิม (repeat) และมีการให้คะแนนความรู้บางส่วน (partial credit) 138

4.17 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) และมีการให้คะแนนแบบทวิวิภาค (dichotomous) 139

4.18 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) และมีการให้คะแนนความรู้บางส่วน (partial credit) 140

4.19 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) และมีการให้คะแนนแบบทวิวิภาค (dichotomous) 141

4.20 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) และมีการให้คะแนนความรู้บางส่วน (partial credit) 142

สารบัญแผนภาพ(ต่อ)

แผนภาพที่ **หน้า**

4.21 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) 143

4.22 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) 186

4.23 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) 187

4.24 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) 187

4.25 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ..188

4.26 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) 188

4.27 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit).....189

4.28 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน (counterbalance 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) 189

5.1 ความสัมพันธ์ระหว่างตำแหน่งของค่าสุดโต่งกับสมการที่ใช้ในการทำนาย 202

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

การวัดคะแนนพัฒนาการเป็นวิธีวิทยาการวัดที่ได้รับความสนใจจากศาสตร์เกือบทุกสาขา จุดมุ่งหมายในการวัดคะแนนพัฒนาการที่สำคัญสรุปได้ 3 ประการคือ ประการแรกเพื่อพัฒนาวิธีการวัดที่ถูกต้อง ประการที่สองเพื่อประมาณค่าคะแนนพัฒนาการ ประการที่สามเพื่อศึกษารูปแบบของพัฒนาการ (Burr and Nesselroade, 1990: 3-4) สำหรับในสาขาการศึกษานั้นการวัดคะแนนพัฒนาการเป็นที่สนใจของนักวัดผลมาตั้งแต่ยุคเริ่มต้นของการวัดตั้งแต่ปีค.ศ. 1924 (Embretson, 1991: 415) เพราะการวัดคะแนนพัฒนาการไม่เพียงแต่ให้ภาพพัฒนาการของการเรียนรู้ของผู้เรียนว่ามีมากน้อยเพียงใด แต่คะแนนพัฒนาการยังเป็นตัวบ่งชี้ที่ใช้วัดประสิทธิภาพหรือคุณภาพของตัวแปรจัดกระทำที่ให้แก่ผู้เรียน และสามารถนำไปใช้ในการประเมินระบบการจัดการศึกษา ว่าบรรลุวัตถุประสงค์หรือไม่อย่างไร (Pike, 1991: 501; Willett, 1994: 671)

การวัดคะแนนพัฒนาการในยุคต้นๆนั้นเป็นการวัดตามทฤษฎีการทดสอบแบบดั้งเดิม โดยมีหลักการว่าคะแนนที่วัดได้โดยตรงเป็นผลรวมของ คะแนนความสามารถที่แท้จริงและความคลาดเคลื่อนจากการวัด โดยที่ค่าเฉลี่ยของความคลาดเคลื่อนในการวัดมีค่าเท่ากับ 0 และความคลาดเคลื่อนในการวัดไม่สัมพันธ์กับคะแนนจริง และความคลาดเคลื่อนในการวัดแต่ละครั้งไม่สัมพันธ์กัน (Lord, 1980: 3-9) และคะแนนพัฒนาการได้จากคะแนนความแตกต่างที่วัดได้ (observed difference score) ซึ่งประมาณค่าได้จากผลต่างระหว่างคะแนนสอบหลังเรียนกับคะแนนสอบก่อนเรียน วิธีการนี้เป็นวิธีที่ง่ายและเป็นวิธีที่นิยมใช้เพราะคะแนนพัฒนาการที่ได้เป็นตัวประมาณค่าที่ไม่ลำเอียง (Raykov, 1993: 54) วิธีการวัดวิธีนี้มีจุดอ่อนตรงข้อตกลงเบื้องต้นที่ว่าคะแนนความคลาดเคลื่อนในการสอบก่อนเรียนและหลังเรียนไม่สัมพันธ์กัน นักวัดผลการศึกษาหลายท่านพบว่าตามข้อตกลงเบื้องต้นนี้ทำให้คะแนนพัฒนาการที่วัดได้มีค่าความเที่ยงต่ำและคะแนนพัฒนาการที่ได้มีความสัมพันธ์ทางลบกับคะแนนสอบก่อนเรียน (Baird, 1988; Bereiter, 1963; Cronbach and Furby, 1970 Hanson, 1988; Linn and Slinde, 1977; Terenzini, 1989; Thorndike, 1966 cited in Pike, 1991: 501) ในประเด็นของค่าความเที่ยงต่ำนั้น K. Tatsuoka (1975 cited in Boonruangrutana, 1978) ได้พิสูจน์ให้เห็นว่าค่าความเที่ยงของคะแนนพัฒนาการมีค่าสูงขึ้นมากเมื่อยอมให้ค่าความคลาดเคลื่อนของการวัดทั้ง 2 ครั้งสัมพันธ์กัน (จาก

เดิมมีค่า 0.16 เพิ่มเป็น 0.5691) นักวัดผลการศึกษาได้พยายามคิดวิธีการวัดคะแนนพัฒนาการวิธีใหม่ ๆ ที่มีประสิทธิภาพยิ่งขึ้น ซึ่งสามารถแยกได้เป็น 2 กลุ่ม (the International Encyclopedia of Education, 1994) กลุ่มแรกเป็นวิธีการวัดคะแนนพัฒนาการแบบดั้งเดิม (traditional methods of measuring changes) ที่มีการวัด 2 ครั้ง (two-wave) และกลุ่มที่สองเป็นการวัดคะแนนพัฒนาการแนวใหม่ (modern methods of measuring changes) ที่มีการวัดตั้งแต่ 2 ครั้งขึ้นไป

ผลการพัฒนาวิธีการวัดคะแนนพัฒนาการแบบดั้งเดิมตามทฤษฎีการทดสอบแบบดั้งเดิมมี 7 วิธีดังต่อไปนี้

1) วิธีการวัดรูปแบบคะแนนจริงของพัฒนาการ (true - score models of change) Lord ได้นำเสนอวิธีการวัดคะแนนพัฒนาการครั้งแรกในปี 1956 (Pike, 1991: 502) โดยนิยามว่าคะแนนพัฒนาการประมาณค่าได้จากผลการทำนายคะแนนเพิ่มแท้จริงด้วยคะแนนสอบก่อนเรียนและหลังเรียน จุดเด่นของวิธีนี้คือ ผู้เรียนสองคนที่มีคะแนนความแตกต่างระหว่างคะแนนสอบก่อนเรียนและหลังเรียนเท่ากัน ผู้ที่ได้คะแนนสอบก่อนเรียนสูงกว่าจะได้คะแนนพัฒนาการสูงกว่า ผู้เรียนที่เก่งได้คะแนนพัฒนาการมากกว่าผู้เรียนที่อ่อน ผลการวัดคะแนนพัฒนาการของลอร์ด จะให้ผลคงเส้นคงวา แต่วิธีนี้มีจุดอ่อนอยู่ที่ คะแนนพัฒนาการขึ้นอยู่กับขนาดกลุ่มตัวอย่าง ต้องใช้กลุ่มตัวอย่างขนาดใหญ่มาประมาณค่าเพื่อให้ค่าสัมประสิทธิ์การถดถอยมีความคลาดเคลื่อนลดลง (อรุณี อ่อนสวัสดิ์, 2537: 4; วินิจ เทือกทอง, 2537: 4) จุดอ่อนอีกประการหนึ่งคือ พิสัยของคะแนนพัฒนาการของลอร์ดแคบกว่าพิสัยของคะแนนความแตกต่างระหว่างคะแนนสอบก่อนเรียนและหลังเรียน (อรุณี อ่อนสวัสดิ์, 2537: 4)

2) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือ (residual change score) Du Bois เสนอวิธีการวัดคะแนนพัฒนาการส่วนที่เหลือ ในปี 1957 (Pike, 1991: 502) โดยนิยามว่าคะแนนพัฒนาการประมาณค่าได้จากความแตกต่างระหว่างคะแนนสอบหลังเรียนกับคะแนนผลการทำนายคะแนนสอบหลังเรียนด้วยคะแนนสอบก่อนเรียน วิธีนี้มีจุดเด่นในด้านที่คะแนนพัฒนาการเป็นอิสระจากคะแนนสอบก่อนเรียน (Raykov, 1993: 54; อรุณี อ่อนสวัสดิ์, 2537: 3) และสามารถกำจัดความสัมพันธ์ลวงระหว่างคะแนนพัฒนาการกับคะแนนสอบก่อนเรียนได้ วิธีนี้มีจุดอ่อน 3 ประการคือ ประการแรกถ้าคะแนนสอบก่อนเรียนและหลังเรียนมีความสัมพันธ์กันสูง ความเที่ยงของคะแนนพัฒนาการจะมีค่าต่ำ ประการที่สองคะแนนพัฒนาการส่วนที่เหลือไม่สามารถแสดงถึงปริมาณการเปลี่ยนแปลงรายบุคคลได้โดยตรง ประการที่สามคะแนนพัฒนาการส่วนที่เหลือขึ้นอยู่กับขนาดกลุ่มตัวอย่าง (อรุณี อ่อนสวัสดิ์, 2537: 3; วินิจ เทือกทอง, 2537: 3)

3) วิธีการวัดคะแนนพัฒนาการที่เป็นอิสระจากคะแนนสอบก่อนเรียน (base - free measure of change) Tucker, Damarin และ Messic นำเสนอวิธีการวัดคะแนนพัฒนาการที่เป็นอิสระจากคะแนนสอบก่อนเรียน ในปี 1966 (Pile, 1991: 502; Raykov, 1993: 54) โดย

นิยามว่าคะแนนพัฒนาการประมาณค่าได้จากความแตกต่างระหว่างคะแนนสอบหลังเรียนกับคะแนนทำนายหลังเรียนด้วยคะแนนจริงก่อนเรียน จุดเด่นวิธีนี้คือการกำจัดความสัมพันธ์ลวงระหว่างคะแนนพัฒนาการกับคะแนนสอบก่อนเรียน (Zimmerman & Williams, 1982: 142) จุดอ่อนของวิธีนี้ คือคะแนนพัฒนาการมีค่าทั้งบวกและลบ ทำให้ยากในการตีความเปรียบเทียบว่าใครมีการเปลี่ยนแปลงการเรียนรู้มากหรือน้อยกว่าใคร (อรุณี อ่อนสวัสดิ์, 2537: 3)

4) วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน (standard score method) Labouvie เสนอวิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐานในปี 1982 (Burr and Nesselroade, 1990: 17) โดยนิยามว่าคะแนนพัฒนาการประมาณค่าได้จากความแตกต่างระหว่างคะแนนมาตรฐานหลังเรียนกับคะแนนมาตรฐานก่อนเรียน จุดเด่นวิธีนี้ คือการแก้ปัญหาลักษณะการแจกแจงที่ต่างกันของคะแนนสอบก่อนเรียนและคะแนนสอบหลังเรียน เมื่อแปลงคะแนนให้เป็นคะแนนมาตรฐานจะได้คะแนนอยู่ในมาตร(scale)เดียวกัน ซึ่งทำให้คะแนนพัฒนาการที่ได้สามารถนำมาเปรียบเทียบกันได้ระหว่างผู้สอบ หรือระหว่างวิชาที่ต่างกัน

5) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ (logarithm of observed score method) Tornqvist, Vartia, and Vartia เสนอแนวคิดในการใช้ลอการิทึมธรรมชาติในการวัดคะแนนพัฒนาการในปี 1985 (Burr and Nesselroade, 1990: 13) โดยนิยามว่าคะแนนพัฒนาการประมาณค่าได้จากความแตกต่างระหว่างลอการิทึมธรรมชาติของคะแนนสอบหลังเรียนกับลอการิทึมธรรมชาติของคะแนนสอบก่อนเรียน วิธีนี้มีจุดเด่นในการแก้ปัญหาในกรณีที่คะแนนสอบก่อนเรียนและคะแนนสอบหลังเรียนไม่เป็นฟังก์ชั่นแบบบวก เมื่อใช้ค่าลอการิทึมธรรมชาติทำให้คะแนนทั้ง 2 ตัวมีการแจกแจงแบบสมมาตร มีฟังก์ชั่นแบบบวก และมีคุณสมบัติเป็นการแจกแจงปกติ (symmetric, additive and normalized properties) จุดอ่อนของวิธีนี้ คะแนนมาตรฐานมิใช่ฟังก์ชั่นเชิงเส้นของคะแนนพัฒนาการ แต่เป็นผลรวมของคะแนนพัฒนาการกับคะแนนเบี่ยงเบนจากค่าเฉลี่ย ถ้าความแปรปรวนของคะแนนการวัดทั้ง 2 ครั้งไม่เท่ากัน ค่าคะแนนพัฒนาการลอการิทึมธรรมชาติจะมีความลำเอียงโดยมีค่าสูงเกินกว่าที่ควรจะเป็น (นงลักษณ์ วิรัชชัย, 2542; 263)

6) วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ ศิริชัย กาญจนวาสี (อรุณี อ่อนสวัสดิ์, 2537) เสนอแนวคิดในการวัดพัฒนาการโดยนิยามว่าคะแนนพัฒนาการสัมพัทธ์ประมาณค่าได้จากอัตราส่วนระหว่างผลต่างของคะแนนสอบก่อนเรียนกับหลังเรียนกับผลต่างของคะแนนเต็มกับคะแนนสอบก่อนเรียน วิธีนี้มีจุดเด่น คือ คะแนนพัฒนาการสัมพัทธ์ จะแก้ปัญหาในเรื่องของผู้เรียนที่ได้คะแนนพัฒนาการเท่ากัน แต่มีศักยภาพในการเปลี่ยนแปลงไม่เท่ากัน เพราะคะแนนพัฒนาการสัมพัทธ์จะบอกถึงอัตราการเปลี่ยนแปลงเมื่อเทียบกับศักยภาพที่ผู้เรียนมีอยู่ วิธีนี้จะช่วยแก้ปัญหาเรื่องอิทธิพลเพดาน (ceilling effect) ได้อีกทางหนึ่ง

ต่อมาวินิจ เทือกทอง (2537) ได้นำแนวคิดวิธีประมาณค่าคะแนนพัฒนาการสัมพัทธ์ของศิริชัย กาญจนวาสี ไปศึกษาเปรียบเทียบและประยุกต์ใช้กับคะแนนพัฒนาการ (gain score) ที่ประมาณค่าจาก 3 วิธี คือวิธีประมาณค่าจากความแตกต่างระหว่างคะแนนสอบก่อนเรียนและหลังเรียน วิธีการวัดคะแนนพัฒนาการแบบวิธีเศษเหลือ และวิธีการวัดคะแนนพัฒนาการของลอร์ด โดยประยุกต์เป็น 2 แบบ แบบที่ 1 เป็นการนำคะแนนพัฒนาการทั้ง 3 วิธี มาเทียบกับศักยภาพของผู้สอบแต่ละคน ศักยภาพของผู้สอบประมาณค่าจากผลต่างระหว่างคะแนนเต็มกับคะแนนสอบก่อนเรียน แบบที่ 2 เป็นการนำคะแนนพัฒนาการทั้ง 3 วิธี มาเทียบส่วนร้อยกับฐานคะแนนของผู้สอบแต่ละคน คะแนนที่เป็นฐานได้แก่คะแนนสอบก่อนเรียน จุดเด่นของวิธีการเทียบคะแนนพัฒนาการกับศักยภาพของผู้สอบนี้ คือการคำนึงถึงศักยภาพของผู้สอบแต่ละคน ดังนั้นคะแนนพัฒนาการของผู้สอบแต่ละคนจะแตกต่างกันตามศักยภาพของผู้สอบ และมีการขจัดอิทธิพลเพดานของผู้สอบแต่ละคนด้วย จุดเด่นของวิธีการเทียบคะแนนพัฒนาการกับส่วนร้อยของฐานคะแนนผู้สอบ คือมีการคำนึงถึงพื้นความรู้เดิมของผู้สอบ คะแนนพัฒนาการที่ได้จึงเป็นการเทียบกับความรู้เดิมที่มีอยู่ ดังนั้นผู้สอบ 2 คนที่ได้คะแนนพัฒนาการเท่ากัน แต่มีความรู้เดิมไม่เท่ากัน จะได้คะแนนพัฒนาการที่ต่างกัน ผู้สอบที่มีความรู้เดิมสูง จะมีคะแนนพัฒนาการต่ำกว่าผู้สอบที่มีความรู้เดิมต่ำ และจากการวิจัยพบว่าวิธีการวัดคะแนนพัฒนาการแบบวิธีเศษเหลือเทียบกับศักยภาพของผู้สอบมีความคลาดเคลื่อนต่ำกว่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแบบประมาณค่าจากความแตกต่างระหว่างคะแนนสอบก่อนเรียนและหลังเรียนเทียบกับศักยภาพของผู้สอบ และของวิธีการวัดคะแนนพัฒนาการของลอร์ดเทียบกับศักยภาพของผู้สอบ และพบว่าวิธีการเทียบคะแนนพัฒนาการกับส่วนร้อยของฐานคะแนนผู้สอบทั้ง 3 วิธี มีความคลาดเคลื่อนของคะแนนพัฒนาการสูงกว่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแบบวิธีเศษเหลือเทียบกับศักยภาพของผู้สอบ

7) วิธีการวัดคะแนนพัฒนาการโดยขจัดอิทธิพลเพดาน อรุณี อ่อนสวัสดิ์ได้พัฒนาวิธีการวัดคะแนนพัฒนาการโดยขจัดอิทธิพลเพดานในปี 2537 โดยอาศัยแนวคิดตามทฤษฎีการเรียนรู้ของบลูม และการขจัดอิทธิพลเพดาน โดยนิยามว่าคะแนนพัฒนาการการเรียนรู้เป็นฟังก์ชั่นของพื้นความรู้เดิม การเรียนการสอนและอิทธิพลเพดาน จุดเด่นวิธีนี้คือ สามารถกำจัดอิทธิพลเพดานได้และมีแนวโน้มว่าคะแนนพัฒนาการมีค่าความเที่ยงสูงกว่าคะแนนพัฒนาการที่ได้จากการประมาณค่าความแตกต่างระหว่างคะแนนสอบก่อนเรียนกับหลังเรียน และคะแนนพัฒนาการของลอร์ด วิธีการวัดคะแนนพัฒนาการโดยขจัดอิทธิพลเพดานมีจุดอ่อนคือผลการวัดคะแนนพัฒนาการจะมีประสิทธิภาพดีในสถานการณ์ที่ความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนและหลังเรียนมีค่าสูง กลุ่มตัวอย่างมีขนาดปานกลาง และมีข้อคำถามน้อยข้อเท่านั้น

นอกจากนี้วิธีนี้ค่าความเที่ยงของคะแนนพัฒนาการไม่ต่างจากวิธีอื่นๆ และวิธีการประมาณค่าค่อนข้างยุ่งยากซับซ้อน

ผลการพัฒนาวิธีการวัดคะแนนพัฒนาการแนวใหม่ตามทฤษฎีการทดสอบแบบดั้งเดิมที่มีการวัดตั้งแต่ 2 ครั้งขึ้นไป เป็นการประยุกต์ใช้โมเดลสมการโครงสร้าง (Structural Equation Model=SEM) ในการวัดคะแนนพัฒนาการ (Tucker, 1958; Meredith, 1960; McArdle and Hamagami, 1991) โมเดลที่ได้รับความนิยมและใช้กันมากในการวัดคะแนนพัฒนาการระยะยาวที่มีการวัดตั้งแต่ 3 ครั้งขึ้นไปนั้น ได้แก่ โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (Latent Growth Curve Model) ที่มีแนวคิดว่าคะแนนการวัดแต่ละครั้งเป็นคะแนนประกอบ (composite score) ที่ประกอบด้วยองค์ประกอบ 3 ส่วนคือ ส่วนแรกเป็นองค์ประกอบในสถานะเริ่มต้น (initial factor) ส่วนที่ 2 เป็นองค์ประกอบของคะแนนพัฒนาการทั้งหมด (overall change factor) และส่วนที่ 3 เป็นองค์ประกอบเฉพาะ ซึ่งเป็นความคลาดเคลื่อนในการวัด (Raykov, 1994) โมเดลนี้ถือเป็นโมเดลที่ดีที่สุดในการวัดคะแนนพัฒนาการและเป็นวิธีที่ใช้วิธีการวิเคราะห์องค์ประกอบในโมเดลสมการโครงสร้าง (Stoolmiler, 1995: 105) ในต่างประเทศมีนักวัดผลการศึกษานำโมเดลนี้ไปใช้อย่างกว้างขวาง เช่น McArdle และ Epstein (1987), McArdle (1988), Ducan และ Ducan (1994) และ Stoolmiler (1994) เป็นต้น สำหรับในประเทศไทยนั้นผู้ที่นำโมเดลนี้ไปใช้ได้แก่ อิทธิพงษ์ ตั้งสกุลเรืองไล (2541) มนต์ทิวา ไชยแก้ว (2542) และสมถวิล วิจิตรวรรณา (2543)

สำหรับวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบแยกออกได้เป็น 2 แนวคิด ดังนี้

แนวคิดแรก เป็นการประยุกต์ราส์ชโมเดล ในการวัดคะแนนสอบที่มีการวัดซ้ำโดย Fischer สร้างพารามิเตอร์ของข้อสอบให้อยู่ในลักษณะฟังก์ชั่นเชิงเส้น และเรียกชื่อโมเดลใหม่ว่า โมเดลการทดสอบโลจิสติกเชิงเส้น (Linear Logistic Test Model = LLTM) หรือโมเดลราส์ชโครงสร้าง (Structural Rasch Model) ในการวัดคะแนนพัฒนาการตามวิธีนี้ Fischer (1997, 187-190) อธิบายว่าเมื่อมีการวัดคะแนนสอบก่อนเรียนและหลังเรียน คะแนนพัฒนาการจะถูกสร้างให้อยู่ในรูปโมเดลของคะแนนพัฒนาการของพารามิเตอร์ของข้อสอบ (ในรูปของฟังก์ชั่นของอิทธิพลจากตัวแปรจัดกระทำ) และมีข้อตกลงเบื้องต้นว่า พารามิเตอร์ของผู้สอบ (θ_v) มีค่าคงที่ในการวัดแต่ละครั้ง วิธีการนี้ใช้ได้กับแบบวัดที่เป็นเอกมิติ (unidimension) และพหุมิติ (multidimension) ข้อสอบที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และพหุวิภาค (Polytomous) จุดเด่นของวิธีการวัดคะแนนพัฒนาการวิธีนี้คือที่การนำค่าความยากของข้อสอบมาใช้ในการประมาณค่าความสามารถที่แท้จริงที่เป็นคะแนนพัฒนาการ วิธีการนี้สามารถใช้ข้อสอบคนละชุดกันในการสอบแต่ละครั้ง ซึ่งจะป้องกันปัญหาที่เกิดจากการจำข้อสอบได้ และลดปัญหา

จากอิทธิพลความคงที่ของการตอบและอิทธิพลจากการฝึกทำข้อสอบ ซึ่งเป็นการช่วยป้องกันไม่ให้เกิดความลำเอียง (Embretson, 1991: 499) โมเดลที่ใช้ในการวัดคะแนนพัฒนาการวิธีนี้มีหลายแบบได้แก่ 1) Linear Logistic Model with Relaxed Assumption (LLRA) พัฒนาโดย Fischer ในปี 1974 เป็นการวัดคะแนนพัฒนาการที่มีการวัด 2 ครั้งที่ใช้แบบทดสอบชุดเดิม การตรวจข้อสอบเป็นคะแนนแบบทวิวิภาค (dichotomous) ไม่จำกัดจำนวนมิติ (dimension) 2) Hybrid - LLRA พัฒนาโดย Fischer ในปี 1977 สำหรับการวัดคะแนนพัฒนาการที่ไม่ต้องใช้ข้อสอบชุดเดิมในการวัดซ้ำ 3) Linear Rating Scale Model (LRSM) พัฒนาโดย Fischer และ Parzer (1991) ในปี 1991 ใช้สำหรับการวัดคะแนนพัฒนาการที่เครื่องมือวัดเป็นมาตรประมาณค่า (rating scale) ที่มีการวัดมากกว่า 2 ครั้ง โดยใช้ข้อคำถามชุดเดิมในการวัดแต่ละครั้ง 4) Linear Partial Credit Model (LPCM) พัฒนาโดย Fischer และ Ponocny (1994) ในปี 1994 ใช้สำหรับการวัดคะแนนพัฒนาการในการสอบที่มีการวัดมากกว่า 2 ครั้งและมีการตรวจให้คะแนนความรู้บางส่วน (partial credit model) ข้อจำกัดของโมเดลเหล่านี้คือ การประมาณค่าคะแนนต้องใช้วิธีการ conditional maximum likelihood (CML) และการทดสอบสมมติฐานเกี่ยวกับค่าพารามิเตอร์ต้องใช้ conditional likelihood ratio(CLR) ทดสอบความมีนัยสำคัญ ซึ่งต้องมีโปรแกรมคอมพิวเตอร์เฉพาะสำหรับวิเคราะห์ข้อมูล

แนวคิดที่สอง เป็นการประมาณค่าคะแนนพัฒนาการจากผลต่างของพารามิเตอร์ของผู้สอบ (θ) ที่ได้จากการสอบก่อนเรียนและหลังเรียน วิธีการวัดคะแนนพัฒนาการตามแนวคิดนี้มีหลายโมเดล เช่น โมเดลการวิเคราะห์โครงสร้างตัวแปรแฝงระยะยาว (longitudinal latent structure analysis) ของ Anderson (1985) โมเดลเส้นโค้งของทฤษฎีการตอบสนองข้อสอบที่กลมกลืน (fit) กับเส้นโค้งพัฒนาการเป็นรายบุคคล ของ Bock (1976) และโมเดลการประมาณค่าพารามิเตอร์ด้วยเส้นโค้งพัฒนาการแบบ exponential และ hyperbolic ของ Keats (1983) เป็นต้น ในช่วงเวลาต่อมามีการพัฒนาโมเดลเหล่านี้ให้สามารถประมาณค่าคะแนนพัฒนาการที่จริงได้ ได้แก่ 1) โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (Multidimensional Rasch Model for Learning and Change: MRMLC) Embretson (1991a; 1991b) ได้เสนอโมเดลราส์ชพหุมิติสำหรับการวัดคะแนนพัฒนาการในปี 1991 โดยมีแนวคิดว่าคะแนนพัฒนาการเป็นผลรวมของความสามารถเริ่มต้นกับความสามารถที่พัฒนาเพิ่ม (modifiabilities) โมเดลนี้ใช้สำหรับการวัดตั้งแต่ 2 ครั้งขึ้นไป โดยใช้กับข้อสอบต่างกันแต่มีข้อสอบร่วม 1 ฉบับ และมีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) วิธีการนี้มีการพัฒนาให้ดีขึ้นโดยการจัดสอบแบบจตุรัสละติน (latin square) เพื่อให้เกิดการถ่วงสมดุลย์ (counterbalanced) 2) การวัดคะแนนพัฒนาการจากผลต่างของความสามารถของผู้สอบก่อนเรียนและหลังเรียน May และ Nicewander (1998) ได้เสนอสูตรการวัดคะแนนพัฒนาการจากความแตกต่างของ

ความสามารถที่แท้จริง $(\theta_2-\theta_1)$ ที่ได้จากการสอบหลังเรียน (posttest) กับการสอบก่อนเรียน (pretest) โดยใช้วิธีการประมาณค่าความสามารถที่แท้จริง (θ) ในแต่ละครั้งแบบ EAP (expected a posteriori) จุดเด่นของวิธีนี้คือเป็นวิธีการที่ง่ายและสามารถแก้ไขปัญหาในเรื่องของความคลาดเคลื่อนของค่าคะแนนพัฒนาการที่เกิดขึ้นจากปฏิสัมพันธ์ระหว่างค่าความยากของข้อสอบกับความสามารถเริ่มต้นที่แท้จริงของผู้สอบ ที่เกิดจากการหาคะแนนพัฒนาการด้วยผลต่างของคะแนนดิบที่ได้จากการสอบ 2 ครั้ง จุดอ่อนของวิธีนี้คือ คะแนนพัฒนาการของความสามารถที่แท้จริงที่เท่ากัน อาจมาจากความสามารถที่แท้จริงเริ่มต้นที่แตกต่างกัน และคะแนนพัฒนาการของความสามารถที่แท้จริงยังมีอิทธิพลเพดานติดอยู่

เมื่อพิจารณากลุ่มแนวคิดที่วัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ จะเห็นได้ว่าวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริงตามวิธีการของ May และ Nicewander (1998) โดยการหาผลต่างระหว่างความสามารถที่แท้จริง $(\theta_2-\theta_1)$ จากการสอบหลังเรียน (posttest) กับการสอบก่อนเรียน (pretest) เป็นวิธีที่ง่าย วิธีนี้ถึงแม้จะมีจุดเด่นด้านการแก้ปัญหาในเรื่องการมีปฏิสัมพันธ์ระหว่างความสามารถเริ่มต้นของผู้สอบกับค่าความยากของข้อสอบ แต่ยังมีจุดอ่อนอยู่ 2 ประการ คือ

ประการที่หนึ่ง คะแนนพัฒนาการที่ประมาณค่าจากความแตกต่างของความสามารถที่แท้จริงยังมีอิทธิพลเพดานติดอยู่ กล่าวคือผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงจะมีพิสัยของศักยภาพการพัฒนาแคบกว่าพิสัยของผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ คะแนนพัฒนาการของผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงจึงต่ำกว่าคะแนนของผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ เนื่องจากติดอิทธิพลเพดาน

เพื่อปรับแก้จุดอ่อนข้อนี้ของวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง ผู้วิจัยจึงได้ประยุกต์แนวคิดของ ศิริชัย กาญจนวาสี (2532, อ้างถึงในอรุณี อ่อนสวัสดิ์, 2537: 23) มาใช้แก้ปัญหาเรื่องอิทธิพลเพดานโดยพัฒนาเป็นวิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์โดยประมาณค่าจากอัตราส่วนระหว่างผลต่างของความสามารถที่แท้จริงก่อนและหลังเรียนกับศักยภาพการพัฒนาได้ โดยนิยามศักยภาพการพัฒนาว่าเป็นผลต่างของความสามารถสูงสุดกับความสามารถที่แท้จริงก่อนเรียน โดยใช้สูตรดังนี้

$$S\theta = \frac{\theta_2 - \theta_1}{\mathrm{Max} - \theta_1}$$

$S\theta$ = วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา

θ_1 = ความสามารถที่แท้จริงก่อนเรียน

θ_2 = ความสามารถที่แท้จริงหลังเรียน

Max = ความสามารถสูงสุด

แต่เนื่องจากค่า θ เป็นคะแนนมาตรฐานที่มีค่าตั้งแต่ $-\alpha$ ถึง $+\alpha$ ซึ่งสามารถกำหนดช่วงได้ตั้งแต่ -4 ถึง +4 เพราะสามารถครอบคลุมการแจกแจงของข้อมูลได้ถึงร้อยละ 99.994 (Kirk, 1995: 798) ดังนั้น Max ที่เป็นความสามารถที่แท้จริงสูงสุดสำหรับการวิจัยครั้งนี้จึงกำหนดให้มีค่าเท่ากับ 4

โดยมีข้อตกลงเบื้องต้นว่า การสอบทั้ง 2 ครั้ง ใช้วิธีการประมาณค่าแบบ EAP (expected a posteriori)

วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์วิธีนี้ผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงจะมีคะแนนพัฒนาการของความสามารถที่แท้จริงสูงกว่าผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ เมื่อมีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากัน ตัวอย่างเช่น นาย ก และนาย ข มีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากันคือ 1 นาย ก มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ 2 นาย ข มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ 1 ความสามารถสูงสุดเท่ากับ 4 นาย ก จะมีคะแนนพัฒนาการของความสามารถที่แท้จริงเท่ากับ 0.5 ซึ่งสูงกว่าคะแนนพัฒนาการของความสามารถที่แท้จริงของนาย ข ที่มีค่าเท่ากับ 0.33

ประการที่สอง การตีความของคะแนนพัฒนาการที่ได้จากวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริงที่เท่ากันกรณีที่ผู้สอบมีความสามารถที่แท้จริงก่อนเรียนที่ไม่เท่ากันยังไม่เป็นธรรม

ผู้วิจัยมีความคิดว่าผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงมีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากันกับผู้ที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ จะมีโอกาสที่จะพัฒนาความสามารถที่แท้จริงได้มากกว่าผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ เพราะในกรณีที่คะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากันนั้นผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงน่าจะใช้ความสามารถที่แท้จริงเพิ่มขึ้นน้อยกว่าผู้ที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ ที่จะทำให้ได้คะแนนพัฒนาการเท่ากัน ดังนั้นความสามารถที่แท้จริงที่พัฒนาขึ้นจึงน่าจะได้จากวิธีการเทียบคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงกับความสามารถที่แท้จริงก่อนเรียนด้วย ตามแนวคิดของ วินิจ เทือกทอง (2537) ดังนั้นผู้วิจัยจึงประยุกต์วิธีการของวินิจ เทือกทอง พัฒนาวิธีการวัดคะแนนพัฒนาการของความสามารถที่แท้จริงสัมพัทธ์ โดยประมาณค่าจากอัตราส่วนระหว่างผลต่างระหว่างความสามารถที่แท้จริงกับความสามารถที่แท้จริงก่อนเรียน โดยใช้สูตรดังนี้

$$V\theta = \frac{\theta_2 - \theta_1}{\theta_1}$$

$V\theta$ = วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน

θ_1 = ความสามารถที่แท้จริงก่อนเรียน

θ_2 = ความสามารถที่แท้จริงหลังเรียน

เนื่องจากค่า θ_1 เป็นคะแนนที่ได้จากการประมาณค่าความสามารถที่แท้จริงด้วยฟังก์ชั่นโลจิสติกโมเดลตามทฤษฎีการตอบสนองข้อสอบซึ่งมีค่าได้ตั้งแต่ $-\alpha$ ถึง $+\alpha$ ดังนั้น θ_1 จึงมีโอกาสที่จะมีค่าเป็นศูนย์ได้ ซึ่งจะทำให้ไม่สามารถหาค่าของคะแนนพัฒนาการได้ และถ้า θ_1 มีค่าเป็นลบ และความแตกต่างของความสามารถที่แท้จริงมีค่าเป็นบวก แต่คะแนนพัฒนาการที่ได้จะมีค่าเป็นลบซึ่งไม่สอดคล้องกับความเป็นจริง เช่น คนที่มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ -1 และมีพัฒนาการเพิ่มขึ้นมาอย่างสูงโดยมีความสามารถที่แท้จริงหลังเรียนเท่ากับ 2 จะได้คะแนนพัฒนาการเป็น -3 เป็นต้น

ดังนั้นผู้วิจัยจึงได้ปรับค่า θ ให้มีค่าเป็นบวกเสมอโดยการปรับด้วยการนำค่า Max ไปบวกค่า θ ทุกค่าในสมการ ผลการปรับได้สูตรใหม่ดังนี้

$$V\theta = \frac{(\text{Max} + \theta_2) - (\text{Max} + \theta_1)}{(\text{Max} + \theta_1)}$$

$$V\theta = \frac{\text{Max} + \theta_2 - \text{Max} - \theta_1}{(\text{Max} + \theta_1)}$$

$$\therefore V\theta = \frac{\theta_2 - \theta_1}{(\text{Max} + \theta_1)}$$

จากสูตรที่ปรับได้ใหม่นี้จะเห็นได้ว่าคล้ายกับสูตรแรกที่พัฒนาขึ้น ต่างกันที่สูตรแรกจะหารผลต่างของความสามารถที่แท้จริงด้วยผลต่างของ Max กับ θ_1 แต่สูตรที่สองจะหารผลต่างของความสามารถที่แท้จริงด้วยผลบวกของ Max กับ θ_1 ซึ่งทั้งสองสูตรมีข้อตกลงเบื้องต้นเหมือนกันคือการวัดทั้ง 2 ครั้ง ใช้วิธีการประมาณค่าแบบ EAP (expected a posteriori)

วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียนวิธีนี้ผู้ที่มีความสามารถที่แท้จริงก่อนเรียนสูงจะมีคะแนนพัฒนาการต่ำกว่าผู้ที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ เมื่อมีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากัน ตัวอย่างเช่น นาย ก และนาย ข มีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากันคือ 1 นาย ก มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ 1 นาย ข มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ -2 นาย ก จะมีคะแนนพัฒนาการเท่ากับ 0.2 ซึ่งต่ำกว่าคะแนนพัฒนาการของนาย ข ที่มีค่าเท่ากับ 0.5

โดยสรุปในการวิจัยครั้งนี้ผู้วิจัยได้พัฒนาวิธีการวัดคะแนนพัฒนาการตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่มีการวัด 2 ครั้งขึ้นมาเพื่อแก้จุดอ่อนของวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง ได้แก่ วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์ ซึ่งมี 2 วิธีคือ

1) เมื่อเทียบกับศักยภาพการพัฒนา
2) เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน

การจัดการศึกษาตามแนวพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 ในปัจจุบันเน้นผู้เรียนสำคัญที่สุด กระบวนการจัดการศึกษาต้องส่งเสริมให้ผู้เรียนสามารถพัฒนาตามธรรมชาติและเต็มศักยภาพ (มาตราที่ 22 พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542) ระบบการวัดและประเมินผลการศึกษาต้องเน้นการวัดพัฒนาการของผู้เรียน โดยมุ่งวัดคะแนนพัฒนาการทั้งด้านความรู้ ทักษะและคุณธรรม (ศิริชัย กาญจนวาสี, 2543) จึงทำให้การวัดคะแนนพัฒนาการเข้ามามีบทบาทในการวัดและประเมินผลการศึกษามากขึ้น

จากการศึกษาค้นคว้าดังเสนอในตอนต้นผู้วิจัยพบว่าวิธีการวัดคะแนนพัฒนาการนั้นมีหลายกลุ่ม ทั้งกลุ่มทฤษฎีการทดสอบแบบดั้งเดิมและกลุ่มทฤษฎีการตอบสนองข้อสอบแต่ละกลุ่มมีหลากหลายวิธี ตั้งแต่วิธีที่ง่ายที่สุด จนถึงวิธีที่ซับซ้อน วิธีการวัดคะแนนพัฒนาการแต่ละวิธียังไม่มีข้อสรุปที่ชัดเจนว่าวิธีไหนดีที่สุดสำหรับสถานการณ์ใด จากการศึกษาค้นคว้าถึงแม้จะพบว่าโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (Latent Growth Curve Model: LGCM) เป็นโมเดลที่ได้รับความนิยมในการใช้วัดการเปลี่ยนแปลงหรือวัดคะแนนพัฒนาการระยะยาว โดยได้รับการยอมรับว่าโมเดลนี้เป็นโมเดลที่ดีที่สุดในการวัดคะแนนพัฒนาการจากนักวัดผลการศึกษาหลายท่าน ได้แก่ Rogosa, Brandt and Zimowski, 1982; Rogosa and Willett, 1985 (cited in Stoolmiller, 1995) จุดเด่นของโมเดลนี้คือมีการวัดหลายครั้งจึงมีข้อสารสนเทศมากเพียงพอที่จะบอกได้ว่าแบบแผนของพัฒนาการเป็นรูปแบบใด เป็นเส้นตรง หรือเส้นโค้ง คะแนนพัฒนาการรายบุคคลต่างกันหรือไม่ต่างกัน นอกจากนั้นยังมีการนำค่าความคลาดเคลื่อนเข้าไปร่วมในการประมาณค่าคะแนนพัฒนาการอีกด้วย กล่าวได้ว่ายิ่งมีการวัดหลายครั้งจะทำให้การประมาณค่าคะแนนพัฒนาการของผู้เรียนแต่ละคนได้ถูกต้องยิ่งขึ้น และมีความเที่ยงของคะแนนพัฒนาการเพิ่มขึ้น (Willett and Sayer, 1994: 377) ข้อจำกัดของโมเดลนี้อยู่ที่ความสิ้นเปลืองในการวัดหลายครั้ง นอกจากนั้นวิธีการตรวจสอบโมเดลว่ามีความสอดคล้องกับข้อมูลเชิงประจักษ์หรือไม่นั้นจำเป็นต้องอาศัยความรู้ในสถิติขั้นสูง จึงไม่ค่อยเหมาะสมในทางปฏิบัติ

ผู้วิจัยมีความเห็นว่าการนำโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (LGCM) ไปใช้ในโรงเรียนเพื่อวิเคราะห์ประมาณค่าคะแนนพัฒนาการของนักเรียนในโรงเรียนคงเป็นไปได้ยาก เนื่องจากมีข้อจำกัดหลายประการ ประการแรกครูมีภาระงานมาก ระยะเวลาของการสอนมี

จำกัดและเนื้อหาที่สอนมีมากไม่สามารถที่จะทำการวัดซ้ำ ๆ กันหลายครั้งได้ ประการที่สองการที่จะสร้างโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง ต้องอาศัยความรู้ทางด้านการวัดผลและความรู้ทางด้านสถิติขั้นสูง ซึ่งครูส่วนมากยังไม่สะดวกที่จะนำไปใช้ ผู้วิจัยจึงสนใจที่จะศึกษาวิธีการวัดคะแนนพัฒนาการที่เป็นวิธีที่ง่าย ไม่ซับซ้อน ไม่เสียเวลาในการวัด และเป็นวิธีที่สะดวกสำหรับครูในโรงเรียนนำไปใช้ปฏิบัติได้

จากกลุ่มวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมรวม 7 วิธี ผู้วิจัยใช้เกณฑ์ในการเลือกวิธีการวัดคะแนนพัฒนาการรวม 2 ประการคือ ต้องเป็นวิธีการที่ให้ผลการวัดมีค่าความเที่ยงและมีค่าความคลาดเคลื่อนน้อย จากการวิจัยของอรุณี อ่อนสวัสดิ์ (2537) พบว่าวิธีการวัดคะแนนพัฒนาการโดยขจัดอิทธิพลเพดานทุกวิธีที่อรุณี อ่อนสวัสดิ์ นำเสนอมีค่าความเที่ยงและค่าความคลาดเคลื่อนไม่แตกต่างจากวิธีการหาคะแนนความแตกต่างระหว่างคะแนนดิบและวิธีการหาคะแนนพัฒนาการของลอร์ดแต่มีวิธีการที่ยุ่งยากกว่า และจากงานวิจัยของวินิจ เทือกทอง (2537) พบว่าคะแนนความแตกต่างระหว่างคะแนนดิบ คะแนนพัฒนาการของลอร์ดมีค่าความคลาดเคลื่อนต่ำกว่าคะแนนพัฒนาการส่วนที่เหลือ สำหรับวิธีหาคะแนนความแตกต่างระหว่างคะแนนดิบนั้นเป็นวิธีที่ง่ายที่สุด และเป็นตัวประมาณค่าที่ไม่ลำเอียงในการวัดพัฒนาการ (Raykov, 1993: 54) ดังนั้นผู้วิจัยจึงเลือกศึกษาวิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ และวิธีวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ ซึ่งเป็นวิธีที่มีค่าความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบกับวิธีการวัดคะแนนพัฒนาการที่มีการขจัดอิทธิพลเพดานแบบวิธีเทียบส่วน ตามผลการวิจัยของวินิจ เทือกทอง (2537) ดังนั้นในกลุ่มของทฤษฎีการทดสอบแบบดั้งเดิม ผู้วิจัยใช้วิธีการวัดคะแนนพัฒนาการทั้งหมด 5 วิธี คือ

1. วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ (observed difference score method) ในที่นี้ใช้สัญลักษณ์ DS

2. วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน (standard score method) ในที่นี้ใช้สัญลักษณ์ SC

3. วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ (logarithm of observed score method) ในที่นี้ใช้สัญลักษณ์ NL

4. วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ (relative growth method) ในที่นี้ใช้สัญลักษณ์ SR

5. วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ(residual growth and potential ratio method) ในที่นี้ใช้สัญลักษณ์ SRR

สำหรับวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบนั้น วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง (Lθ) และวิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (MRMLC) เป็นวิธีการประมาณค่าคะแนนพัฒนาการที่สามารถหาเป็นรายบุคคลได้ และมีวิธีการที่ค่อนข้างสะดวก มีโปรแกรมสำเร็จรูปที่ใช้ในการวิเคราะห์ ดังนั้นผู้วิจัยจึงใช้วิธีการวัดคะแนนพัฒนาการในกลุ่มนี้ 2 วิธี คือ

1. วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง (method of difference between true ability scores) ในที่นี้ใช้สัญลักษณ์ Lθ

2. วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (Multidimensional Rasch Model for Learning and Change: MRMLC)

เนื่องจากวิธีการวัดคะแนนพัฒนาการตามทฤษฎีตอนสนองข้อสอบวิธีแรกยังมีจุดอ่อนด้านอิทธิพลเพดานและความไม่ยุติธรรมสำหรับคนที่มีความสามารถที่แท้จริงเริ่มต้นต่ำ ผู้วิจัยจึงได้พัฒนาเพื่อปรับแก้จุดอ่อนให้เป็นวิธีการวัดคะแนนพัฒนาการตามทฤษฎีตอบสนองข้อสอบเพิ่มขึ้นอีก 2 วิธี คือ

1.วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา (relative true ability of growth and potential ratio method) ในที่นี้ใช้สัญลักษณ์ Sθ

2.วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน (relative true ability of growth and initial ability ratio method) ในที่นี้ใช้สัญลักษณ์ Vθ

ในการวิจัยครั้งนี้ผู้วิจัยจะใช้คะแนนพัฒนาการ (GS) ซึ่งประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (LGCM) ที่ได้รับการยอมรับว่าเป็นโมเดลที่ดีที่สุดสำหรับการวัดคะแนนพัฒนาการระยะยาวเป็นเกณฑ์ในการพิจารณาเปรียบเทียบคุณภาพของคะแนนพัฒนาการที่ได้จากการวัดในแต่ละกลุ่มวิธีในทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีในทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีที่ผู้วิจัยพัฒนา

วัตถุประสงค์ในการวิจัย

ในการวิจัยครั้งนี้ผู้วิจัยมีวัตถุประสงค์เพื่อพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ โดยมีวัตถุประสงค์เฉพาะดังนี้

1. เพื่อวิเคราะห์การประมาณค่าคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ (DS) 2) วิธีการ

วัดคะแนนพัฒนาการจากคะแนนมาตรฐาน (SC) 3) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ (NL) 4) วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ (SR) และ 5) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพ (SRR) ตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง ($L\theta$) และ 2) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (MRMLC) และวิธีที่ผู้วิจัยพัฒนา 2 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา ($S\theta$) และ 2) วิธีการวัดความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน ($V\theta$)

2. เพื่อเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธี ตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี และวิธีที่ผู้วิจัยพัฒนาขึ้น 2 วิธี ในด้านความตรงตามเกณฑ์สัมพันธ์และความคลาดเคลื่อนของวิธีการวัด

สมมติฐานในการวิจัย

1. สมมติฐานเกี่ยวกับความตรงตามเกณฑ์สัมพันธ์

ในการวิจัยครั้งนี้ความตรงตามเกณฑ์สัมพันธ์ จะพิจารณาจากค่าความสัมพันธ์ 2 แบบคือ 1) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการที่ได้จากวิธีวัดแต่ละวิธีกับคะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (GS) ที่ใช้เป็นเกณฑ์ 2) ค่าสหสัมพันธ์อันดับระหว่างอันดับของคะแนนพัฒนาการที่ได้จากวิธีการวัดคะแนนพัฒนาการแต่ละวิธีกับอันดับของคะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (GS) ที่ใช้เป็นเกณฑ์ โดยผู้วิจัยได้ตั้งสมมติฐานด้านความตรงตามเกณฑ์สัมพันธ์ดังนี้

1.1 คะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีความตรงตามเกณฑ์สัมพันธ์อยู่ในระดับสูง

1.2 คะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม

1.3 เมื่อเปรียบเทียบคะแนนพัฒนาการภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม คะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ (DS) และคะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการสัมพัทธ์ (SR) มีความตรงตาม

เกณฑ์สัมพันธ์สูงกว่าวิธีอื่นๆ เพราะ DS เป็นตัวประมาณค่าที่ไม่ลำเอียงในการวัดคะแนนพัฒนาการที่แท้จริง (Raykov, 1993: Willet, 1994) และเพราะ DS กับ SR เป็นวิธีวัดคะแนนพัฒนาการที่ไม่ขึ้นอยู่กับความสัมพันธ์ของคะแนนสอบก่อนเรียนและหลังเรียน รวมทั้งไม่ขึ้นกับคะแนนของกลุ่ม

1.4 เมื่อเปรียบเทียบคะแนนพัฒนาการภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ คะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง ($D\theta$) มีความตรงตามเกณฑ์สัมพันธ์สูงที่สุด เพราะ $D\theta$ ได้จากการประมาณค่า θ ด้วยโมเดลโลจิสติก 3 พารามิเตอร์ ส่วน MRMLC ได้จากการประมาณค่า θ ด้วยโมเดลโลจิสติก 1 พารามิเตอร์ ดังนั้นค่า θ ที่ประมาณได้ใน $D\theta$ จึงใกล้เคียงกับความสามารถที่แท้จริงมากกว่า MRMLC

1.5 เมื่อเปรียบเทียบคะแนนพัฒนาการภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา คะแนนพัฒนาการที่ได้จากวิธีการวัดคะแนนพัฒนาการทั้ง 2 วิธี มีความตรงตามเกณฑ์สัมพันธ์ไม่ต่างกัน เนื่องจากเป็นวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบเหมือนกัน

2. สมมติฐานด้านความคลาดเคลื่อน

ความคลาดเคลื่อนจะพิจารณาจากค่าสถิติ 2 ค่า คือ 1) ค่าเฉลี่ยของความแตกต่างระหว่างคะแนนพัฒนาการที่ได้จากแต่ละวิธีกับคะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (GS) ที่ใช้เป็นเกณฑ์ 2) ค่ารากที่สองของค่าเฉลี่ยกำลังที่สองของความแตกต่างของคะแนนพัฒนาการที่ได้จากวิธีวัดแต่ละวิธีกับคะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (GS) ที่ใช้เป็นเกณฑ์ โดยผู้วิจัยได้ตั้งสมมติฐานด้านความคลาดเคลื่อนไว้ดังนี้

2.1 คะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีความคลาดเคลื่อนต่ำกว่าคะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม

2.2 เมื่อเปรียบเทียบคะแนนพัฒนาการภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม คะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ (DS) และคะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการสัมพัทธ์ (SR) มีความคลาดเคลื่อนต่ำกว่าวิธีอื่นๆ

2.3 เมื่อเปรียบเทียบคะแนนพัฒนาการภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ คะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง ($D\theta$) มีความคลาดเคลื่อนต่ำกว่า MRMLC

2.4 เมื่อเปรียบเทียบคะแนนพัฒนาการภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา คะแนนพัฒนาการที่ได้จากวิธีการวัดคะแนนพัฒนาการทั้ง 2 วิธี มีความคลาดเคลื่อนไม่ต่างกัน

ข้อตกลงเบื้องต้น

1. คะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง ซึ่งเป็นโมเดลที่ได้รับการยอมรับในการวัดคะแนนพัฒนาการระยะยาวว่าเป็นโมเดลที่ดีที่สุด (Stoolmiler, 1995) การวิจัยนี้จึงถือว่าเป็นคะแนนเกณฑ์ของพัฒนาการที่แท้จริง

2. แบบทดสอบที่ใช้เป็นเครื่องมือในการวัดความรู้บางส่วนในการวิจัยครั้งนี้เป็นแบบทดสอบวิชาคณิตศาตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ที่ผู้วิจัยสร้างขึ้นโดยประยุกต์วิธีการให้คะแนนความรู้บางส่วนของแบบทดสอบอัตนัยของ Wright และ Masters (1982) โดยผู้วิจัยได้ประยุกต์ข้อสอบอัตนัยให้เป็นข้อสอบปรนัยแบบเลือกตอบ โดยแต่ละตัวเลือกสะท้อนถึงวิธีการคิดที่แตกต่างกัน เมื่อผู้ตอบเลือกตัวเลือกใดจึงถือว่าเป็นคะแนนความรู้บางส่วนตามตัวเลือกนั้น

ขอบเขตการวิจัย

1. การวิจัยครั้งนี้เป็นการพัฒนาและวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม ทฤษฎีการตอบสนองข้อสอบ และวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา ผู้วิจัยกำหนดวิธีการวัดคะแนนพัฒนาการ 3 กลุ่ม จำนวน 9 วิธี ดังต่อไปนี้

1.1 กลุ่มวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมมีจำนวน 5 วิธี คือ

1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ (DS)

2) วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน (SC)

3) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ (NL)

4) วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ (SR)

5) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพ (SRR)

1.2 กลุ่มวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบมี 2 วิธีคือ

1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง ($L\theta$)

2) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (MRMLC)

1.3 กลุ่มวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามี 2 วิธีคือ

1) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา

2) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน

ทั้งนี้ผู้วิจัยกำหนดให้วิธีการวัดคะแนนพัฒนาการตามโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (LGCM) เป็นวิธีเกณฑ์ในการวิเคราะห์เปรียบเทียบ เนื่องจาก Rogosa, Brandt and Zimowski, 1982; Rogosa and Willett, 1985 (cited in Stoolmiller, 1995) กล่าวว่าวิธีการดังกล่าวเป็นวิธีการที่ดีที่สุดในการวัดคะแนนพัฒนาการแต่มีความยุ่งยากและซับซ้อนในการประมาณค่าคะแนนพัฒนาการ

2. เนื่องจากการวิเคราะห์เปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการในการวิจัยครั้งนี้รวม 9 วิธี มีวิธีการวัดคะแนนพัฒนาการตามโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (LGCM) เป็นเกณฑ์ ผู้วิจัยจึงใช้ค่าสถิติในการตรวจสอบคุณภาพของวิธีการวัดคะแนนพัฒนาการดังนี้

2.1 ความตรงตามเกณฑ์สัมพันธ์ระหว่างคะแนนพัฒนาการที่ได้จากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (LGCM) ที่ใช้เป็นเกณฑ์กับคะแนนพัฒนาการที่ได้จากแต่ละวิธีทั้ง 9 วิธี เนื่องจากคะแนนที่ได้จากการตรวจข้อสอบมีทั้งแบบทวิวิภาคและพหุวิภาค การประมาณค่าความตรงตามเกณฑ์สัมพันธ์จึงใช้ทั้งแบบสหสัมพันธ์ของเพียร์สันและสหสัมพันธ์อันดับของสเปียร์แมน

2.2 ค่าความคลาดเคลื่อนของคะแนนพัฒนาการของผู้เรียนที่ได้จากแต่ละวิธีทั้ง 9 วิธี

3. ข้อมูลที่ใช้ในการวิจัยครั้งนี้มี 3 ชุด ดังนี้

3.1 ชุดที่ 1 เป็นฐานข้อมูลวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 เก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในภาคเรียนที่ 1 ปีการศึกษา 2543 จำนวน 698 คน ในโรงเรียนสังกัดกรมสามัญศึกษาเขตกรุงเทพมหานคร ผู้วิจัยเก็บรวบรวมข้อมูลด้วยตนเองจากการสอบซ้ำ (repeat) 3 ครั้งโดยใช้แบบทดสอบฉบับเดียวกัน โดยมีการเก็บข้อมูล 3 ครั้ง ระยะเวลาในการเก็บข้อมูลห่างกันครั้งละ 4 สัปดาห์

3.2 ชุดที่ 2 เป็นฐานข้อมูลวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 เก็บรวบรวมข้อมูลจากกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 3 ในภาคเรียนที่ 1 ปีการศึกษา 2543 จำนวน

637 คน ในโรงเรียนสังกัดกรมสามัญศึกษาเขตกรุงเทพมหานคร ผู้วิจัยเก็บรวบรวมข้อมูลด้วยตนเองจากการสอบซ้ำแบบจตุรัสละติน (Latin square) เพื่อให้เกิดการถ่วงสมดุลย์ (counterbalance) โดยผู้วิจัยแบ่งกลุ่มตัวอย่างออกเป็น 3 กลุ่ม สอบกลุ่มละ 3 ครั้ง ใช้แบบทดสอบ 3 ฉบับ โดยแต่ละฉบับใช้สอบในครั้งเดียวกันฉบับละกลุ่ม ระยะเวลาในการเก็บข้อมูลห่างกันครั้งละ 4 สัปดาห์

3.3 ชุดที่ 3 เป็นฐานข้อมูลวิชาคณิตศาสตร์ที่วัดความสามารถในการแก้โจทย์ปัญหาระดับชั้นมัธยมศึกษาปีที่ 2 เก็บรวบรวมจากกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 2 ในภาคเรียนที่ 1 ปีการศึกษา 2542 จำนวน 433 คน ในโรงเรียนขยายโอกาสทางการศึกษาขั้นพื้นฐาน สังกัดสำนักงานการประถมศึกษาจังหวัดสมุทรสงคราม ซึ่งเป็นฐานข้อมูลในงานวิจัยของสมถวิล วิจิตรวรรณา ที่มีการเก็บข้อมูล 5 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence) ระยะเวลาในการเก็บข้อมูลห่างกันครั้งละ 3 สัปดาห์

4. ข้อมูลทั้ง 3 ชุด แต่ละชุดมีการตรวจให้คะแนน 2 แบบดังนี้

4.1 ตรวจให้คะแนนแบบทวิวิภาค (dichotomous)

4.2 ตรวจให้คะแนนความรู้บางส่วน (partial credit)

นิยามศัพท์เฉพาะ

คะแนนพัฒนาการ หมายถึงคะแนนการเปลี่ยนแปลงที่แสดงถึงพัฒนาการทางการเรียนรู้ของผู้เรียนหลังจากที่ได้รับการเรียนการสอน ซึ่งในการวิจัยครั้งนี้เป็นคะแนนการเปลี่ยนแปลงของนักเรียนในระดับมัธยมศึกษาปีที่ 2 ในโรงเรียนขยายโอกาสทางการศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดสมุทรสงคราม และนักเรียนในระดับมัธยมศึกษาปีที่ 3 สังกัดกรมสามัญศึกษาในเขตกรุงเทพมหานคร ในวิชาคณิตศาสตร์ ซึ่งวัดได้จากวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา

คะแนนพัฒนาการที่แท้จริง หมายถึงคะแนนพัฒนาการที่ใช้เป็นเกณฑ์ในการวิจัยครั้งนี้ได้จากการประมาณค่าด้วยโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่มีการวัด 3 ครั้ง สำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคณิตศาสตร์ฉบับเดิม (repeat) ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคณิตศาสตร์ที่คู่ขนานกัน (counterbalance) และมีการวัด 3-5 ครั้ง สำหรับข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคณิตศาสตร์ที่คู่ขนานกัน (equivalence)

วิธีการวัดคะแนนพัฒนาการ หมายถึงกระบวนการที่นำคะแนนที่ได้จากผลการสอบตั้งแต่ 2 ครั้งขึ้นไปมาจัดกระทำทางคณิตศาสตร์ ให้ได้คะแนนพัฒนาการ ในการวิจัยครั้งนี้มี 3 กลุ่มวิธี ได้แก่ กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธีคือ วิธีการวัดคะแนนพัฒนาการจาก

ความแตกต่างระหว่างคะแนนดิบ วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ และวิธีการวัดคะแนนพัฒนาการส่วนเหลือเทียบกับศักยภาพของผู้สอบ กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ 2 วิธีคือ วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง และวิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการวัดการเปลี่ยนแปลง (MRMLC) กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา 2 วิธีคือ วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา และวิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน

คะแนนพัฒนาการที่ได้จากวิธีการวัดทั้ง 9 วิธี มีนิยามดังต่อไปนี้

1. **วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ** (observed difference score method) คะแนนพัฒนาการหมายถึงผลต่างระหว่างคะแนนสอบหลังเรียนกับคะแนนสอบก่อนเรียน

2. **วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน** (standard score method) คะแนนพัฒนาการหมายถึงผลต่างระหว่างคะแนนมาตรฐานของคะแนนสอบหลังเรียนกับคะแนนมาตรฐานของคะแนนสอบก่อนเรียน

3. **วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ** (logarithm of observed score method) คะแนนพัฒนาการหมายถึงผลต่างระหว่างลอการิทึมธรรมชาติของคะแนนสอบหลังเรียนกับลอการิทึมธรรมชาติของคะแนนสอบก่อนเรียน

4. **วิธีการวัดคะแนนพัฒนาการสัมพัทธ์** (relative growth method) คะแนนพัฒนาการหมายถึงอัตราส่วนร้อยละระหว่างผลต่างระหว่างคะแนนสอบหลังเรียนกับคะแนนสอบก่อนเรียนกับผลต่างระหว่างคะแนนเต็มกับคะแนนสอบก่อนเรียน

5. **วิธีการวัดคะแนนพัฒนาการส่วนเหลือเทียบกับศักยภาพของผู้สอบ** (residual growth and potential ratio method) คะแนนพัฒนาการหมายถึงอัตราส่วนร้อยละระหว่างผลต่างระหว่างคะแนนสอบหลังเรียนกับค่าทำนายคะแนนสอบหลังเรียนด้วยคะแนนสอบก่อนเรียน กับผลต่างระหว่างคะแนนเต็มกับคะแนนสอบก่อนเรียน

6. **วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง** (method of difference between true ability scores) คะแนนพัฒนาการหมายถึงผลต่างระหว่างความสามารถที่แท้จริงหลังเรียนกับความสามารถที่แท้จริงก่อนเรียน

7. **วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง** (Multidimensional Rasch Model for Learning and Change) คะแนน

พัฒนาการหมายถึงความสามารถที่แท้จริงที่เพิ่มขึ้นเนื่องจากอิทธิพลของการเรียนรู้ที่วิเคราะห์ด้วยโปรแกรม LPCM

8. **วิธีการวัดคะแนนพัฒนาการความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา** (relative true ability of growth and potential ratio method) คะแนนพัฒนาการหมายถึงอัตราส่วนระหว่างผลต่างระหว่างความสามารถที่แท้จริงหลังเรียนกับความสามารถที่แท้จริงก่อนเรียน กับผลต่างระหว่างความสามารถสูงสุด กับความสามารถที่แท้จริงก่อนเรียน

9. **วิธีการวัดคะแนนพัฒนาการความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน** (relative true ability of growth and initial ability ratio method) คะแนนพัฒนาการหมายถึงอัตราส่วนระหว่างผลต่างระหว่างความสามารถที่แท้จริงหลังเรียนกับความสามารถที่แท้จริงก่อนเรียน กับความสามารถที่แท้จริงก่อนเรียน

คุณภาพของวิธีการวัดคะแนนพัฒนาการ หมายถึงคุณสมบัติของวิธีการวัดคะแนนพัฒนาการที่สามารถให้คะแนนพัฒนาการได้ถูกต้อง โดยตรวจสอบได้จากความตรงตามเกณฑ์สัมพันธ์ ความคลาดเคลื่อนในการวัด โดยมีคะแนนพัฒนาการที่ได้จากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (LGCM) เป็นเกณฑ์

ประโยชน์ที่คาดว่าจะได้รับ

ประโยชน์ทางด้านวิชาการ

1. ได้ขยายองค์ความรู้ในด้านการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมและตามทฤษฎีการตอบสนองข้อสอบ

2. แนวทางในการศึกษาการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบที่มีการวัด 2 ครั้ง โดยใช้เครื่องมือต่างชุดกัน และมีการตรวจให้คะแนนต่างกัน

3. ได้วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์ตามทฤษฎีการตอบสนองข้อสอบ

4. ได้รูปแบบการสร้างแบบทดสอบเลือกตอบที่ใช้ในการวัดความรู้บางส่วน

ประโยชน์ทางด้านการนำไปใช้

1. ได้ข้อสารสนเทศเกี่ยวกับวิธีการวัดคะแนนพัฒนาการแบบต่างๆที่แสดงถึงผลการวัดที่สอดคล้องกับพัฒนาการที่แท้จริงของแต่ละบุคคล ซึ่งครูสามารถที่จะใช้เป็นสารสนเทศในการที่จะเลือกวิธีที่เหมาะสมกับสภาพการเรียนการสอนในโรงเรียน

2. ได้วิธีการวัดคะแนนพัฒนาการที่เหมาะสมกับสภาพของโรงเรียนที่ครูสามารถนำไปวัดคะแนนพัฒนาการรายบุคคลของนักเรียนเพื่อใช้ในการประเมินผลการเรียนรายบุคคล

หรือประยุกต์ใช้ในการประเมินการเรียนการสอนของครู ตลอดจนประเมินผลสัมฤทธิ์ของโรงเรียนและหลักสูตร รวมทั้งเป็นข้อมูลสำหรับรองรับการประกันคุณภาพการศึกษา

3. ได้วิธีการวัดคะแนนพัฒนาการที่เหมาะสมที่สามารถประยุกต์ใช้กับงานวิจัยเชิงทดลองและโครงการฝึกอบรมต่าง ๆ ที่ต้องการวัดพัฒนาการรายบุคคล

4. ได้สารสนเทศเกี่ยวกับพัฒนาการทางด้านคณิตศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 ในโรงเรียนขยายโอกาสทางการศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดสมุทรสงคราม และของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ในโรงเรียนที่สังกัดกรมสามัญศึกษา เขตกรุงเทพมหานคร ซึ่งจะเป็นประโยชน์ต่อครูผู้สอนในการพัฒนาการเรียนการสอนในห้องเรียน และเป็นประโยชน์ต่อผู้บริหารทุกระดับที่จะใช้สารสนเทศในการวางแผนพัฒนาการศึกษาของเยาวชนไทยในด้านความสามารถทางคณิตศาสตร์

5. ได้แบบทดสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ในระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับ ที่พัฒนาจนมีคุณภาพ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง โดยขอนำเสนอเป็น 5 ตอน ดังนี้ 1) การวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม 2) การวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ 3) การเปรียบเทียบวิธีการวัดคะแนนพัฒนาการระหว่างวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม กับทฤษฎีการตอบสนองข้อสอบ 4) งานวิจัยที่เกี่ยวข้อง 5) การพัฒนาวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ โดยมีรายละเอียดดังต่อไปนี้

ตอนที่ 1 การวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม

1.1 แนวคิดในการวัดคะแนนพัฒนาการ

การวัดคะแนนพัฒนาการเป็นวิธีวิทยาการวัดที่ได้รับความสนใจอย่างมากทั้งทางสังคมศาสตร์และพฤติกรรมศาสตร์ โดยเฉพาะอย่างยิ่งในด้านพัฒนาการของมนุษย์ การวัดคะแนนพัฒนาการในยุคแรก ๆ นั้นเป็นข้อมูลที่ได้จากการวัดซ้ำ 2 ครั้ง ที่วัดคุณลักษณะเดียวกัน โดยมีแนวคิดว่าพัฒนาการที่เกิดขึ้นนั้นเกิดจากการจัดกระทำ (treatment) ที่ได้รับ การวัดคะแนนพัฒนาการจึงเป็นการวัดก่อนและหลังจากได้รับการจัดกระทำ โดยมีข้อตกลงว่าความคลาดเคลื่อนของการวัดคะแนนพัฒนาการก่อนและหลังจากที่ได้รับการจัดกระทำมีค่าเท่ากันและหักล้างหมดไป จึงทำให้การวัดคะแนนพัฒนาการในระยะเริ่มแรกไม่สนใจในเรื่องของความคลาดเคลื่อนในการวัดที่เกิดขึ้นที่วัดในคุณลักษณะเดียวกัน (Baird, 1988; Nuttall, 1986 cited in Pike, 1991)

ความพยายามในการวัดคะแนนพัฒนาการในสมัยเริ่มแรกมีการพัฒนาค่อนข้างช้า เนื่องจากประสบปัญหาในเรื่องคุณภาพด้านความเที่ยงและความตรงของคะแนนพัฒนาการที่ได้จากวิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ (difference scores) เพราะมีผู้โต้แย้งกันมากมายในเรื่องของการขาดคุณสมบัติในด้านความเที่ยงและความตรง Bereiter (citecd in Embretson, 1991) ได้ระบุว่าคะแนนพัฒนาการที่ได้จากวิธีการวัดวิธีนี้มีปัญหา 3 ประการด้วยกันคือ ประการแรก ความเที่ยงของคะแนนพัฒนาการมีความสัมพันธ์ผกผันกับค่าความสัมพันธ์ระหว่างคะแนนสอบก่อนและหลังการสอน กล่าวคือ ถ้าคะแนนที่ได้จากการสอบทั้ง 2 ครั้งนั้นมีความสัมพันธ์กันสูง คะแนนพัฒนาการก็จะมีค่าความเที่ยงต่ำ และถ้าคะแนนที่ได้จากการสอบทั้ง 2 ครั้ง มีความสัมพันธ์กันต่ำ คะแนนพัฒนาการจะมีค่าความเที่ยงสูง ซึ่ง

Bereiter ได้ชี้ว่า ถ้าคะแนนที่ได้จากการสอบครั้งแรก (pretest) และการสอบครั้งหลัง (posttest) มีความสัมพันธ์กันต่ำ จะทำให้ยากในการตีความคะแนนพัฒนาการ เพราะความสัมพันธ์ที่ต่ำนี้แสดงว่าการสอบทั้ง 2 ครั้งนี้ไม่ได้วัดในมิติเดียวกัน ประการที่สอง ผู้เรียนที่มีความสามารถเริ่มต้นที่แตกต่างกัน คะแนนพัฒนาการที่วัดได้ไม่อยู่บนมาตรเดียวกัน (same scale) กล่าวคือ ผู้เรียนที่มีความสามารถเริ่มต้นสูง คะแนนพัฒนาการที่วัดได้มีความหมายแตกต่างจากคะแนนพัฒนาการที่เท่ากันที่มาจากผู้เรียนที่มีความสามารถเริ่มต้นต่ำกว่า ประการที่สามคะแนนพัฒนาการมีความสัมพันธ์ต่ำกับความสามารถเริ่มต้น กล่าวคือ คะแนนพัฒนาการมีความสัมพันธ์ลวงทางลบกับความสามารถเริ่มต้น ซึ่ง Lord (cited in Embretson, 1991) ได้อธิบายว่าความคลาดเคลื่อนในการวัดครั้งแรกยังคงมีอยู่ในคะแนนพัฒนาการ แต่มีเครื่องหมายเปลี่ยนไป จึงทำให้เกิดความสัมพันธ์ทางลบระหว่างคะแนนพัฒนาการและความสามารถเริ่มต้น

1.2 วิธีการวัดคะแนนพัฒนาการแบบดั้งเดิม

นงลักษณ์ วิรัชชัย (2542) ได้แบ่งวิธีการวัดคะแนนพัฒนาการออกเป็น 2 กลุ่ม คือ กลุ่มแรกเป็นการวัดคะแนนพัฒนาการที่มีการวัด 2 ครั้ง ใช้สถิติที่ไม่ยุ่งยาก ซึ่งเรียกว่าเป็นวิธีการวัดคะแนนพัฒนาการแบบดั้งเดิม กลุ่มที่ 2 เป็นการวัดคะแนนพัฒนาการที่มีการวัดมากกว่า 2 ครั้งขึ้นไป และใช้โมเดลต่าง ๆ เช่น โมเดลลิสเรล โมเดลเชิงเส้นระดับลดหลั่นมาประมาณค่าคะแนนพัฒนาการ และเรียกว่าเป็นวิธีการวัดคะแนนพัฒนาการแนวใหม่ สำหรับวิธีการวัดคะแนนพัฒนาการแบบดั้งเดิม มีรายละเอียดดังนี้

1.2.1 วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ (observed difference score)

วิธีนี้คะแนนพัฒนาการหมายถึง ผลต่างระหว่างคะแนนสอบหลังเรียนและคะแนนสอบก่อนเรียน โดยมีข้อตกลงว่าการสอบทั้ง 2 ครั้ง ได้วัดในคุณลักษณะเดียวกัน โดยใช้แบบสอบชุดเดิมหรือแบบสอบคู่ขนาน (Pike, 1991) ซึ่งมีสมการดังนี้

$$DS = X_2 - X_1$$

เมื่อ DS คือ วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ

X_2 คือ คะแนนสอบหลังเรียน

X_1 คือ คะแนนสอบก่อนเรียน

จุดเด่นของวิธีนี้คือเป็นวิธีการที่ง่ายที่สุดของการวัดคะแนนพัฒนาการที่ใช้การวัด 2 ครั้งเป็นวิธีเริ่มแรกของการวัดคะแนนพัฒนาการที่ได้รับความนิยมสูง เพราะเป็นตัวประมาณค่าที่ไม่ลำเอียงในการหาคะแนนพัฒนาการที่แท้จริง (Raykov, 1993; Willett, 1994)

ในช่วงทศวรรษที่ 1960 ถึง 1970 ความนิยมในวิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบได้ลดลง เนื่องจากมีผู้ชี้ประเด็นปัญหาในด้านทฤษฎีการวัดแบบดั้งเดิมที่ว่า คะแนนที่สังเกตได้เท่ากับผลบวกระหว่างคะแนนจริงกับความคลาดเคลื่อนในการวัด (measure-ment error) โดยที่มีข้อตกลงเบื้องต้นว่าความคลาดเคลื่อนในการวัดเป็นอิสระต่อกัน ดังนั้น คะแนนพัฒนาการที่หาด้วยวิธีนี้ไม่สนใจในเรื่องความคลาดเคลื่อนในการวัดที่เกิดขึ้นโดยให้ความคลาดเคลื่อนในการวัดของคะแนนสอบหลังเรียนและก่อนเรียนมีค่าเท่ากัน และหักล้างกันหมดไปจึงเป็นการฝ่าฝืนข้อตกลงเบื้องต้น (Pike, 1991; Willett, 1994) ดังสมการ

$$
\begin{aligned}
Y_{POST} &= T_{POST} + E_{POST} \\
Y_{PRE} &= T_{PRE} + E_{PRE} \quad , \\
G &= Y_{POST} - Y_{PRE} \\
&= (T_{POST} - T_{PRE}) + (E_{POST} - E_{PRE}) \\
&= T_{POST} - T_{PRE}
\end{aligned}
$$

ภายใต้ข้อตกลงเบื้องต้นว่า $E_{POST} - E_{PRE} = 0$

จุดอ่อนของวิธีการวัดวิธีนี้คือมีค่าความเที่ยงต่ำ คะแนนพัฒนาการไม่เป็นอิสระจากคะแนนเริ่มต้น (Raykov, 1993; Willett, 1994) คะแนนพัฒนาการมีความสัมพันธ์ทางลบกับคะแนนเริ่มต้น ซึ่งเป็นความสัมพันธ์เทียม (Spurious correlation) (Pike, 1991; Rogosa & Willett, 1985; Willett, 1994) และข้อมูลไม่เป็นไปตามข้อตกลงเบื้องต้น

ในประเด็นของค่าความเที่ยงของคะแนนพัฒนาการมีค่าต่ำนั้น K. Tatsuoka (1975, cited in Boonruangrutana, 1978) ได้ชี้ให้เห็นว่าเกิดจากข้อตกลงเบื้องต้นที่ว่าคะแนนความคลาดเคลื่อนที่ได้จากการวัดทั้ง 2 ครั้งไม่สัมพันธ์ จึงทำให้ค่าความเที่ยงของคะแนนพัฒนาการมีค่าต่ำ ซึ่ง Goulet, Linn และ Tatsuoka (1975 cited in Boonruangrutana, 1978) ได้เสนอวิธีการประมาณค่าพัฒนาการที่แท้จริง โดยผ่อนคลายข้อตกลงเบื้องต้นที่ว่าคะแนนความคลาดเคลื่อนที่ได้จากการวัดทั้ง 2 ครั้งไม่สัมพันธ์ และพบว่าค่าความเที่ยงของคะแนนพัฒนาการมีค่าสูงขึ้นกว่าเดิมมาก (เดิม 0.16 เพิ่มเป็น 0.5691)

1.2.2 **วิธีการวัดคะแนนพัฒนาการที่แท้จริงของลอร์ด** (estimated true gain score) วิธีนี้ Lord เสนอขึ้นในปี 1956 เป็นครั้งแรก (Pike, 1991) โดย Lord ได้นิยามคะแนนพัฒนาการว่าเป็นผลการทำนายคะแนนเพิ่มแท้จริงด้วยหลักการถดถอยพหุคูณ โดยมีคะแนนสอบก่อนเรียนและหลังเรียนเป็นตัวพยากรณ์ เขียนเป็นสมการได้ดังนี้

$$L_i = W_x X_i + W_y Y_i + K$$

หรือ $$V_i = Y + b_{VX,Y}(X_i - \overline{X}) + b_{vy,x}(Y_i - \overline{Y})$$

โดยที่ L_i คือ วิธีการวัดคะแนนพัฒนาการของลอร์ดสำหรับคนที่ i

W_X คือ สัมประสิทธิ์ถดถอยพหุคะแนนความแตกต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียนลงบนคะแนนสอบก่อนเรียน

W_Y คือ สัมประสิทธิ์ถดถอยพหุคะแนนความแตกต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียนลงบนคะแนนสอบหลังเรียน

K คือ ค่าคงที่สำหรับสมการถดถอยพหุ

V_i คือ คะแนนความแตกต่างระหว่างคะแนนจริงก่อนเรียนและคะแนนจริงหลังเรียน

$b_{VX,Y}$ คือ สัมประสิทธิ์ถดถอยความแตกต่างระหว่างคะแนนจริงก่อนเรียนและคะแนนจริงหลังเรียนบนคะแนนสอบก่อนเรียนเมื่อควบคุมคะแนนสอบหลังเรียน

$b_{VY,X}$ คือ สัมประสิทธิ์ถดถอยความแตกต่างระหว่างคะแนนจริงก่อนเรียนและคะแนนจริงหลังเรียนลงบนคะแนนสอบหลังเรียนเมื่อควบคุมคะแนนสอบก่อนเรียน

จุดเด่นของวิธีนี้คือคะแนนความแตกต่างระหว่างคะแนนดิบก่อนเรียนและหลังเรียนเท่ากัน ผู้ที่ได้คะแนนก่อนเรียนสูงกว่าจะได้คะแนนพัฒนาการสูง เด็กเก่งได้คะแนนพัฒนาการมากกว่าเด็กอ่อน และผลการหาคะแนนพัฒนาการของลอร์ดจะให้ผลคงเส้นคงวา

จุดอ่อนของวิธีนี้ คือคะแนนขึ้นอยู่กับกลุ่มตัวอย่าง ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ เพื่อให้ค่าสัมประสิทธิ์การถดถอยมีความหมายที่ควรจะเป็น และพิสัยของคะแนนพัฒนาการของลอร์ดแคบกว่าพิสัยของคะแนนความแตกต่างระหว่างคะแนนดิบก่อนเรียนและหลังเรียน (อรุณี อ่อนสวัสดิ, 2537)

1.2.3 **วิธีการหาคะแนนพัฒนาการส่วนที่เหลือ** (residual change score)

วิธีนี้เป็นวิธีที่ DuBois เสนอขึ้นในปี 1957 (Pike, 1991) คะแนนพัฒนาการในวิธีนี้เป็นการประมาณค่าจากผลต่างระหว่างคะแนนสอบหลังเรียนและคะแนนผลการทำนายหลังเรียนด้วยคะแนนสอบก่อนเรียน เขียนเป็นสมการได้ดังนี้

$$R_i = Y_i - Y'_i$$

$$R_i = Y_i - [\bar{Y} + B_{YX}(X_i - \bar{X})]$$

หรือ $$R_i = Y_i - [\bar{Y} + r_{XY} / S^2_X(X_i - \bar{X})]$$

โดยที่ R_i คือ วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือของคนที่ i

Y_i คือ คะแนนสอบหลังเรียนของคนที่ i

Y'_i คือ คะแนนทำนายผลการสอบหลังเรียนของคนที่ i

$\bar{Y}$ คือ ค่าเฉลี่ยคะแนนสอบหลังเรียนของกลุ่ม

$\bar{X}$ คือ ค่าเฉลี่ยคะแนนสอบก่อนเรียนของกลุ่ม

B_{XY} คือ สัมประสิทธิ์ถดถอยคะแนนสอบหลังเรียนลงบนคะแนนสอบก่อนเรียน

X_i คือ คะแนนสอบก่อนเรียนของคนที่ i

r_{XY} คือ สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนและหลังเรียน

S^2_X คือ ความแปรปรวนของคะแนนสอบก่อนเรียน

วิธีการวัดคะแนนพัฒนาการ วิธีนี้มีจุดเด่นในด้านคะแนนพัฒนาการเป็นอิสระจากคะแนนสอบก่อนเรียน (Raykov, 1993; อรุณี อ่อนสวัสดิ์, 2537) และสามารถกำจัดความสัมพันธ์ลวงระหว่างคะแนนพัฒนาการกับคะแนนสอบก่อนเรียนได้ (อรุณี อ่อนสวัสดิ์, 2537)

วิธีนี้มีจุดอ่อนคือ ถ้าคะแนนสอบก่อนและหลังเรียนมีความสัมพันธ์กันสูง ความเที่ยงของคะแนนพัฒนาการก็จะมีค่าต่ำ คะแนนพัฒนาการส่วนที่เหลือไม่สามารถแสดงถึงปริมาณการเปลี่ยนแปลงรายบุคคลได้โดยตรง และคะแนนพัฒนาการขึ้นอยู่กับกลุ่มตัวอย่าง (อรุณี อ่อนสวัสดิ์, 2537)

1.2.4 วิธีหาคะแนนพัฒนาการที่เป็นอิสระจากคะแนนสอบก่อนเรียน (base-free measure of change)

วิธีนี้เป็นวิธีที่ Tucker, Damarin และ Messick ได้เสนอขึ้นในปี 1966 (Pike, 1991) โดยคะแนนพัฒนาการประมาณค่าจากผลต่างระหว่างคะแนนสอบหลังเรียนกับคะแนนทำนายหลังเรียนด้วยคะแนนจริงก่อนเรียน เขียนเป็นสมการได้ดังนี้

$$B_i = T_i - Y_i^*$$

หรือ $$B_i = Y_i - [\bar{Y} + B_{YX} / R_{XX'}(X_i - \bar{X})]$$

หรือ $$B_i = Y_i - [\bar{Y} + S_{TX,TY} / S^2_{TX}(X_i - \bar{X})]$$

โดยที่ B_i คือ วิธีการวัดคะแนนพัฒนาการที่เป็นอิสระจากคะแนนสอบก่อนเรียน

Y_i^* คือ คะแนนทำนายผลการวัดหลังเรียนด้วยคะแนนจริงก่อนเรียน

$R_{XX'}$ คือ ค่าความเที่ยงของเครื่องมือวัดก่อนเรียน

$S_{TX,TY}$ คือ ความแปรปรวนร่วมระหว่างคะแนนจริงก่อนเรียนและหลังเรียน

S^2_{TX} คือ ความแปรปรวนคะแนนจริงก่อนเรียน

วิธีนี้มีจุดเด่นในด้านคะแนนพัฒนาการเป็นอิสระจากคะแนนจริงก่อนเรียน และสามารถกำจัดความสัมพันธ์ลวงระหว่างคะแนนพัฒนาการกับคะแนนสถานภาพเริ่มต้นได้ (อรุณี อ่อนสวัสดิ์, 2537)

จุดอ่อนของวิธีนี้มีลักษณะอย่างเดียวกับคะแนนพัฒนาการส่วนที่เหลือ เพราะคะแนนพัฒนาการส่วนที่เหลือเป็นกรณีเฉพาะของคะแนนพัฒนาการที่เป็นอิสระจากคะแนนจริงก่อนเรียน เมื่อความเที่ยงในการวัดก่อนเรียนเป็น 1 นอกจากนี้ความคลาดเคลื่อนของคะแนนพัฒนาการวิธีนี้ยังมีแนวโน้มที่จะสูงกว่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการ 2 วิธีที่กล่าวมาก่อนข้างต้น (อรุณี อ่อนสวัสดิ์, 2537)

1.2.5 วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน (standard score method)

วิธีนี้ Labouvie เสนอขึ้นในปี 1982 โดยคะแนนพัฒนาการประมาณค่าจากผลต่างระหว่างคะแนนมาตรฐานหลังเรียนกับคะแนนมาตรฐานก่อนเรียนเขียนเป็นสมการได้ดังนี้

$$SC = Z_2 - Z_1$$

SC = วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน

Z_1 = คะแนนมาตรฐานก่อนเรียน

Z_2 = คะแนนมาตรฐานหลังเรียน

จุดเด่นวิธีนี้คือแก้ปัญหาในด้านของคะแนนสอบก่อนเรียนและคะแนนสอบหลังเรียนที่มีการแจกแจงแตกต่างกัน เมื่อแปลงคะแนนให้เป็นคะแนนมาตรฐานจะได้คะแนนอยู่ในมาตรเดียวกัน (scale) ซึ่งทำให้คะแนนพัฒนาการสามารถนำมาเปรียบเทียบกันได้ระหว่างคนหรือระหว่างวิชาที่ต่างกัน

1.2.6 วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ (logarithm of observed score method)

วิธีนี้ Tornqvist, Vartia, and Vartia ได้เสนอขึ้นในปี 1985 (Burr and Nesselroade, 1990) โดยคะแนนพัฒนาการประมาณค่าจากผลต่างระหว่างลอการิทึมธรรมชาติของคะแนนสอบหลังเรียนกับลอการิทึมธรรมชาติของคะแนนสอบก่อนเรียน เขียนเป็นสมการดังนี้

$$NL = \ln X_2 - \ln X_1$$

NL = วิธีการวัดคะแนนพัฒนาการจากวิธีลอการิทึมของคะแนนดิบ

$\ln X_1$ = ลอการิทึมธรรมชาติของคะแนนสอบก่อนเรียน

$\ln X_2$ = ลอการิทึมธรรมชาติของคะแนนสอบหลังเรียน

วิธีนี้มีจุดเด่นในการแก้ปัญหาในกรณีที่คะแนนสอบก่อนเรียนและคะแนนสอบหลังเรียนไม่เป็นฟังก์ชั่นแบบบวก จึงแปลงค่าคะแนนทั้ง 2 ตัวให้เป็นค่าลอการิทึมธรรมชาติซึ่งทำให้ คะแนนลอการิทึมธรรมชาติทั้ง 2 ตัว มีการแจกแจงแบบสมมาตร เป็นฟังก์ชั่นแบบบวก และมีคุณสมบัติเป็นการแจกแจงปกติ (symmetric, additive and normalized properties)

1.2.7 วิธีการวัดคะแนนพัฒนาการสัมพัทธ์

วิธีนี้เป็นวิธีที่ ศิริชัย กาญจนวาสี (2532) เสนอขึ้นมา โดยคะแนนพัฒนาการสัมพัทธ์ประมาณค่าอัตราส่วนร้อยละระหว่างผลต่างของคะแนนสอบครั้งหลังกับคะแนนสอบครั้งแรกกับผลต่างของคะแนนเต็มกับคะแนนสอบครั้งแรก เขียนเป็นสมการได้ดังนี้

$$SR = 100(Y - X) / (F - X)$$

โดยที่ F คือ คะแนนเต็มในการวัด

Y คือ คะแนนสอบครั้งหลัง

X คือ คะแนนสอบครั้งแรก

วิธีนี้ ศิริชัย กาญจนวาสี เสนอว่าเพื่อลดปัญหาการถดถอยเข้าสู่ส่วนกลาง รวมทั้งคำนึงถึงอัตราความงอกงามเนื่องจากได้นำทั้งพัฒนาการสัมบูรณ์ และพัฒนาการที่น่าจะพัฒนาได้ของแต่ละคนมาคิด ซึ่งเป็นการแก้ปัญหาเรื่องอิทธิพลเพดานได้อีกทางหนึ่ง (อรุณี อ่อนสวัสดิ์, 2537)

1.2.8 วิธีการวัดคะแนนพัฒนาการโดยขจัดอิทธิพลเพดาน (ceilling effect)

วิธีนี้พัฒนาโดยอรุณี อ่อนสวัสดิ์ ในปี 2537 ซึ่งอาศัยแนวคิดตามทฤษฎีการเรียนรู้ของบลูมและอิทธิพลเพดาน โดยมีแนวคิดว่าพัฒนาการการเรียนรู้เป็นฟังก์ชั่นของพื้นความรู้เดิม การเรียนการสอนและอิทธิพลเพดาน ดังนั้น อรุณีจึงได้นำเสนอตัวแปรที่เกี่ยวข้องดังนี้

พัฒนาการการเรียนรู้ ประมาณค่าจากผลต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียน ($T_y - T_x$)

พื้นความรู้เดิมใช้การถดถอยผลต่างระหว่างคะแนนจริงลงบนคะแนนจริงก่อนเรียน (W_1T_X)

การเรียนการสอนเป็นตัวแปรไม่ทราบค่า (V)

อิทธิพลเพดาน เป็นตัวแปรที่ไม่สามารถหาข้อยุติได้ว่าควรใช้ตัวแปรลักษณะใดจึงเหมาะสม อรุณีจึงได้ตั้งข้อสันนิษฐานเป็น 4 แนวคิด แต่ละแนวคิดมีทางปฏิบัติ 2 แนวทาง คือแนวทางแรกเป็นการวัดด้วยการประมาณค่าจริง ส่วนแนวทางที่สองเป็นการวัดด้วยคะแนนดิบ แนวคิดเกี่ยวกับ อิทธิพลเพดานที่ต่างกัน ทั้ง 4 แนวคิดได้แก่

แนวคิดที่ 1 อิทธิพลเพดานคือ การถดถอยความแตกต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียน ลงบนคะแนนจริงความแตกต่างระหว่างคะแนนเต็มและคะแนนสอบหลังเรียน ($W_2T_{(F-Y)}$)

แนวคิดที่ 2 อิทธิพลเพดาน คือ การถดถอยความแตกต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียน ลงบนคะแนนจริงความแตกต่างระหว่างคะแนนเต็มและคะแนนสอบหลังเรียนร่วมกับการถดถอยผลต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียน ลงบนคะแนนจริงผลต่างระหว่างคะแนนเต็มและคะแนนสอบก่อนเรียน($W_2T_{(F-Y)}$และ $W_2T_{(F-X)}$)

แนวคิดที่ 3 อิทธิพลเพดาน คือ การถดถอยผลต่างระหว่างคะแนนจริงลงบนคะแนนจริงผลต่างระหว่างคะแนนเต็มและคะแนนสอบหลังเรียน เมื่อควบคุมคะแนนจริงก่อนเรียน ($W_{2\text{-}1}T_{(F-Y)}$)

แนวคิดที่ 4 อิทธิพลเพดานคือสัดส่วนระหว่างการถดถอยผลต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียนลงบนคะแนนจริงผลต่างระหว่างคะแนนเต็มและคะแนนสอบหลังเรียนกับคะแนนจริงก่อนเรียน ($W_2T_{(F-Y)}$ / T_X)

จาก 4 แนวคิดของอิทธิพลเพดาน และ 2 แนวทางของคะแนนที่ใช้ โดยให้แนวทางที่ 1 (A) วัดด้วยค่าประมาณคะแนนจริง และแนวทางที่ 2 (B) วัดด้วยคะแนนดิบ วิธีการวัดการเปลี่ยนแปลงการเรียนรู้ของอรุณี จึงแบ่งได้เป็น 8 วิธีดังนี้

1) $C1A_i = (1 - W_2)\, T_{Yi} - T_{Xi}$

2) $C1B_i = (1 - W_2)\, Y_i - X_i$

3) $C2A_i = (1 - W_2)\, T_{Yi} - (1 - W_3)\, T_{Xi}$

4) $C2B_i = (1 - W_2)\, Y_i - (1 - W_3)\, X_i$

5) $C3A_i = (1 - W_{2.1})\, T_{Yi} - T_{Xi}$

6) $C3B_i = (1 - W_{2.1})\, Y_i - X_i$

7) $C4A_i = (1 - W_2 / T_{Xi})\, T_{Yi} - T_{Xi} + W_2 F / T_{Xi}$

8) $C4B_i = (1 - W_2 / X_i)\, Y_i - X_i + W_2 F / X_i$

โดยที่ W_2 คือ สัมประสิทธิ์ถดถอย $T_Y - T_X$ ลงบน $T_{(F-Y)}$

$W_2 = -r_{DY} S_D / R_{YY'} S_Y$

W_3 คือ สัมประสิทธิ์ถดถอย $T_Y - T_X$ ลงบน $T_{(F-X)}$

$W_3 = -r_{DX} S_D / R_{XX'} S_X$

$W_{2.1}$ คือ สัมประสิทธิ์ถดถอย $T_Y - T_X$ ลงบน $T_{(F-X)}$ เมื่อควบคุม T_X

จุดเด่นของวิธีการวัดนี้คือสามารถกำจัดอิทธิพลเพดานได้ และมีค่าความเที่ยงสูงกว่า วิธีการวัดคะแนนพัฒนาการโดยวิธีการประมาณค่าจากผลต่างระหว่างคะแนนสอบก่อนเรียนและหลังเรียน และวิธีวัดพัฒนาการของลอร์ด

จุดอ่อนคือผลการวัดจะมีประสิทธิภาพดีในสถานการณ์ที่ความสัมพันธ์ระหว่างความสามารถก่อนเรียนและหลังเรียนมีค่าสูง กลุ่มตัวอย่างขนาดปานกลาง และมีข้อคำถามน้อยเท่านั้น

1.2.9 วิธีการวัดคะแนนพัฒนาการโดยการเทียบส่วนร้อยของคะแนนพัฒนาการ

วิธีนี้วินิจ เทือกทอง พัฒนาขึ้นในปี 2537 เพื่อแก้จุดอ่อนของการวัดคะแนนพัฒนาการที่มีอิทธิพลของคะแนนเพดาน โดยการใช้ฐานในการเปรียบเทียบใหม่ให้เทียบเป็นร้อยละ ซึ่งวินิจได้เสนอฐานที่ใช้ในการเทียบ คือคะแนนสอบครั้งแรก (พื้นความรู้เดิม) โดยการนำค่าคะแนนพัฒนาการ (gain score) ที่ได้คำนวณจาก 3 วิธี คือ คะแนนพัฒนาการวิธีดั้งเดิม คะแนนพัฒนาการวิธีเศษเหลือ และคะแนนพัฒนาการวิธีของลอร์ด มาเทียบส่วนร้อยกับฐานที่เป็นคะแนนสอบครั้งแรก จุดเด่นของวิธีนี้คือจะทำให้อัตราพัฒนาการของผู้สอบแต่ละคนต่างกันตามพื้นความรู้เดิมของแต่ละคนที่มีอยู่ถึงแม้ว่าจะได้คะแนนพัฒนาการเท่ากัน คนที่มีพื้นความรู้เดิมสูง จะมีอัตราคะแนนพัฒนาการที่ต่ำกว่าคนที่มี ความรู้เดิมต่ำ จุดอ่อนวิธีนี้คือมีค่าความคลาดเคลื่อนสูงกว่าคะแนนพัฒนาการที่หาแบบวิธีเศษเหลือเทียบกับศักยภาพของผู้สอบ วิธีการวัดคะแนนพัฒนาการโดยการเทียบส่วนร้อยของคะแนนพัฒนาการมีดังนี้

1) วิธีดั้งเดิมเทียบกับคะแนนครั้งแรก มีสมการดังนี้

$$VPD = \frac{100D}{X}$$

VPD คือ วิธีการวัดคะแนนพัฒนาการแบบวิธีดั้งเดิมเทียบกับคะแนนครั้งแรก

D คือ คะแนนพัฒนาการวิธีดั้งเดิม

X คือ คะแนนสอบครั้งแรก

2) วิธีส่วนที่เหลือหรือวิธีเศษเหลือเทียบกับคะแนนครั้งแรก มีสมการดังนี้

$$VPR = \frac{100R}{X}$$

VPR คือ วิธีการวัดคะแนนพัฒนาการแบบวิธีเศษเหลือเทียบกับคะแนนครั้งแรก

R คือ คะแนนพัฒนาการวิธีเศษเหลือ

X คือ คะแนนสอบครั้งแรก

3) วิธีของลอร์ดเทียบกับคะแนนครั้งแรก มีสมการดังนี้

$$VPL = \frac{100L}{X}$$

VPL คือ วิธีการวัดคะแนนพัฒนาการแบบวิธีของลอร์ดเทียบกับคะแนนครั้งแรก

L คือ คะแนนพัฒนาการวิธีของลอร์ด

X คือ คะแนนสอบครั้งแรก

1.2.10 วิธีการวัดคะแนนพัฒนาการโดยการเทียบคะแนนพัฒนาการกับศักยภาพของผู้สอบ

วิธีนี้วินิจ เทือกทอง พัฒนาขึ้นในปี 2537 เพื่อแก้จุดอ่อนของการวัดคะแนนพัฒนาการที่มีอิทธิพลของคะแนนเพดาน โดยการใช้ฐานในการปรับเทียบใหม่ คือคะแนนศักยภาพของผู้สอบที่ประมาณค่าจากผลต่างของคะแนนเต็มกับคะแนนสอบครั้งหลังซึ่งใช้แนวคิดของศิริชัย กาญจนวาสี โดยการนำค่าคะแนนพัฒนาการ (gain score) ที่ได้จาก 3 วิธี คือ คะแนนพัฒนาการวิธีดั้งเดิม คะแนนพัฒนาการวิธีเศษเหลือ และคะแนนพัฒนาการวิธีของลอร์ด มาเทียบส่วนร้อยกับคะแนนศักยภาพ จุดเด่นของวิธีนี้จะทำให้อัตราพัฒนาการที่ได้ของผู้สอบแต่ละคนแตกต่างกันตามศักยภาพของผู้สอบถึงแม้ว่าจะมีคะแนนพัฒนาการที่ต่างกันก็ตาม ผู้สอบที่มีศักยภาพสูงจะมีอัตราคะแนนพัฒนาการที่สูงกว่าผู้สอบที่มีศักยภาพต่ำ จุดอ่อนของวิธีนี้คือมีค่าความคลาดเคลื่อนสูงกว่าวิธีที่หาคะแนนพัฒนาการโดยไม่ขจัดอิทธิพลเพดาน ได้แก่วิธีดั้งเดิม วิธีเศษเหลือและวิธีของลอร์ด วิธีการวัดคะแนนพัฒนาโดยการเทียบคะแนนพัฒนาการกับศักยภาพของผู้สอบมีวิธี ต่าง ๆ ดังนี้

1) วิธีดั้งเดิมเทียบกับศักยภาพของผู้สอบ มีสมการดังนี้

$$SRD = \frac{100D}{F - X}$$

SRD คือ วิธีการวัดคะแนนพัฒนาการแบบวิธีดั้งเดิมเทียบกับศักยภาพของผู้สอบ

D คือ คะแนนพัฒนาการวิธีดั้งเดิม

F คือ คะแนนเต็ม

X คือ คะแนนสอบครั้งแรก

2) วิธีเศษเหลือเทียบกับศักยภาพของผู้สอบ มีสมการดังนี้

$$SRR = \frac{100R}{F - X}$$

SRR คือ วิธีการวัดคะแนนพัฒนาการแบบวิธีเศษเหลือเทียบกับศักยภาพของผู้สอบ

R คือ คะแนนพัฒนาการวิธีเศษเหลือ

F คือ คะแนนเต็ม

X คือ คะแนนสอบครั้งแรก

3) วิธีของลอร์ดเทียบกับศักยภาพของผู้สอบ มีสมการดังนี้

$$SRL = \frac{100L}{F - X}$$

SRL คือ วิธีการวัดคะแนนพัฒนาการแบบวิธีของลอร์ดเทียบกับศักยภาพของผู้สอบ

D คือ คะแนนพัฒนาการวิธีของลอร์ด

F คือ คะแนนเต็ม

X คือ คะแนนสอบครั้งแรก

1.3 วิธีการวัดคะแนนพัฒนาการแนวใหม่

สำหรับวิธีการวัดคะแนนพัฒนาการในแนวใหม่ที่มีการวัดมากกว่า 2 ครั้งที่ยึดตามทฤษฎีการทดสอบแบบดั้งเดิมนั้นได้อาศัยพัฒนาการความก้าวหน้าทางเทคโนโลยีของเครื่องคอมพิวเตอร์ ทำให้สามารถนำวิธีการวิเคราะห์ทางสถิติขั้นสูงที่ซับซ้อนมาประยุกต์ใช้กับการวัดการเปลี่ยนแปลง และในปัจจุบันมีใช้กันมาก ได้แก่ การประยุกต์ใช้โมเดลสมการโครงสร้าง (Structure Equation Model) (Tucker, 1958; Meredith, 1960; McArdle และ Hamagami 1991) การวัดคะแนนพัฒนาการที่ประยุกต์ใช้โมเดลสมการโครงสร้างนั้น ได้มีการพัฒนาเป็นโมเดลต่าง ๆ ขึ้นมากมาย แต่โมเดลที่ได้รับความนิยมและใช้กันมากในการวัดการเปลี่ยนแปลงระยะยาวนั้น ได้แก่ โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (Latent Growth Curve Model)

โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (Latent Growth Curve Model: LGCM)

โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง ได้พัฒนามาจากโมเดลพัฒนาการเชิงเส้นโค้ง (growth curve) ซึ่งเป็นโมเดลที่ Wishart (1934) ได้ใช้ศึกษาน้ำหนักที่เพิ่มขึ้น (weight gain) ในหมูเบคอน ซึ่งต่อมา Rao (1958) และ Tucker (1958) ได้เสนอให้ใช้วิธีการวิเคราะห์องค์ประกอบมาใช้ในการพัฒนา โมเดลพัฒนาการเชิงเส้นโค้ง (growth curve) ใช้การวิเคราะห์องค์ประกอบแบบการวิเคราะห์ส่วนประกอบมุขสำคัญ (principal component analysis) มาศึกษาแบบแผนพัฒนาการรายบุคคล แบบแผนพัฒนาการแปลผลจากคะแนนประกอบ (composite score) ที่น้ำหนักแต่ละส่วนได้มาจากการหมุนแกน วิธีนี้มีจุดอ่อนที่ไม่สามารถประมาณค่าความแปรปรวนส่วนที่เหลือหรือความคลาดเคลื่อนได้ ต่อมามีผู้ปรับโมเดลนี้หลายครั้ง

เช่น Scher, Young และ Meredith ภายหลัง Meredith และ Tisak (1988, 1990) ได้นำเสนอการวิเคราะห์โมเดลพัฒนาการเชิงเส้นโค้งรายบุคคลโดยการใช้โมเดลสมการโครงสร้าง(SEM) วิเคราะห์ตัวแปรแฝงและความคลาดเคลื่อนของส่วนที่เหลือ และตรวจสอบโมเดลด้วยสมการโครงสร้างเชิงยืนยัน และเรียกวิธีการวิเคราะห์นี้ว่า โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (Latent Growth Curve Model: LGCM) โมเดลนี้ถือเป็นโมเดลที่ดีที่สุดในการศึกษาพัฒนาการที่ใช้วิธีการวิเคราะห์องค์ประกอบในโมเดลสมการโครงสร้าง (Stoolmiler, 1995: 105) และมีนักวัดผลการศึกษานำไปใช้อย่างกว้างขวาง เช่น McArdle และ Epstein (1987), McArdle และ Epstein (1987), McArdle (1988), McArdle และ Hamagami (1991, 1995), Ducan และ Ducan (1994) และ Stoolmiler (1994) เป็นต้น สำหรับในประเทศไทยนั้นผู้ที่นำไปใช้ได้แก่ อิทธิพงษ์ ตั้งสกุลเรืองไล (2541) มนต์ทิวา ไชยแก้ว (2542) และสมถวิล วิจิตรวรรณา (2542)

แนวคิดและระเบียบวิธีของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงในการวัดคะแนนพัฒนาการ

ในการวัดคะแนนพัฒนาการคุณลักษณะแฝงของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง มีแนวคิดว่าพัฒนาการเชิงเส้นตรงรายบุคคล ในการวัด 1 คน ในเวลา 1 ครั้ง (ซึ่งมีการวัดหลายๆครั้ง) คะแนนการวัดในแต่ละครั้ง เป็นคะแนนประกอบ(composite score) ที่ประกอบด้วยองค์ประกอบ 3 ส่วนคือ ส่วนแรกเป็นองค์ประกอบในสถานะเริ่มต้น (initial factor) ส่วนที่ 2 เป็นองค์ประกอบของการเปลี่ยนแปลงทั้งหมด (overall change factor) และส่วนที่ 3 เป็นองค์ประกอบเฉพาะซึ่งเป็นความคลาดเคลื่อนในการวัด (Raykov, 1994) ซึ่งมีสมการดังนี้

$$Y(t,n) = L(n) + B(t)\, S(n) + E(t,n)$$

Y(t,n) คือ ตัวแปรหรือค่าคะแนนที่สังเกตได้ในวัดครั้งที่ t

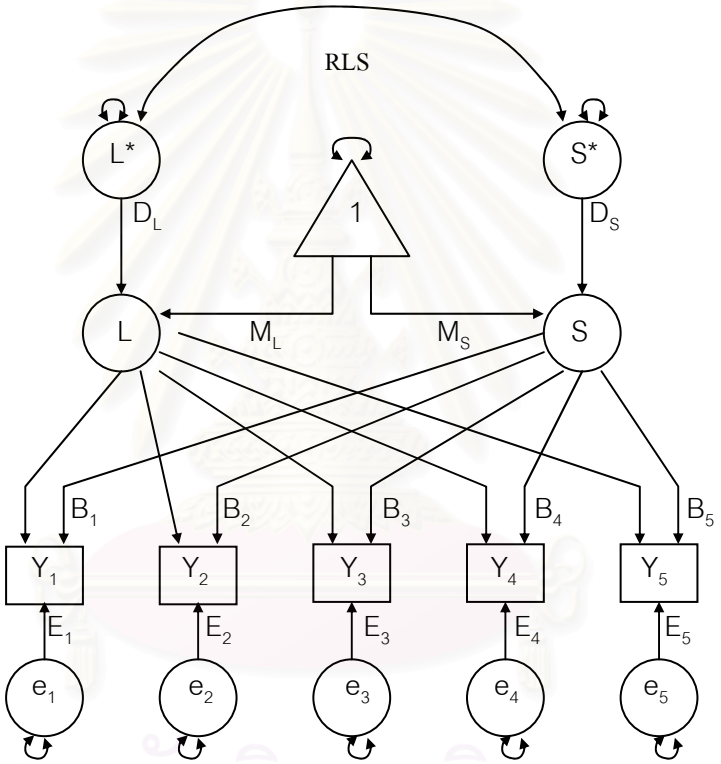
L(n) คือ ตัวแปรแฝงแสดงความแตกต่างระหว่างบุคคลในผลการปฏิบัติงานหรือความสามารถที่เป็นค่าเริ่มต้น

S(n) คือ ตัวแปรแฝงความชันที่แสดงถึงความแตกต่างระหว่างบุคคลในอัตราการเปลี่ยนแปลงของผลการปฏิบัติงานหรือความสามารถตลอดช่วงเวลามีเครื่องหมาย + หรือ – แทนทิศทางของการเปลี่ยนแปลง ค่าตัวแปรแฝงความชันจะมีค่าคงที่สำหรับทุกคน แต่ตัวแปรแฝงความชันจะมีส่วนอธิบายตัวแปรที่สังเกตได้แตกต่างกัน ตามค่าสัมประสิทธิ์ B(t) หรือน้ำหนักองค์ประกอบ ซึ่งมีค่าแตกต่างกันในแต่ละบุคคล

E(n) คือ ตัวแปรแฝงความคลาดเคลื่อนสุ่ม หรือคะแนนเศษเหลือที่เป็นตัวแปรสุ่ม ซึ่งมีค่าเฉลี่ยเป็น 0 และค่าสหสัมพันธ์กับตัวแปรอื่น ๆ เป็น 0 ทุกช่วงเวลา

B(t) คือ สัมประสิทธิ์ของเวลา t

โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงนี้ ภายหลัง McArdle และ Hamagami (1991, 1995) ได้พัฒนาเพิ่มขึ้นโดยใช้หลักการความแปรปรวนแบบวัดซ้ำ โมเดลออโตรีเกรสซีพและการวิเคราะห์องค์ประกอบระยะยาว ทำให้โมเดลมีเทอมคงที่ (constant) ในโมเดล และโมเดลไม่ต้องวิเคราะห์ด้วยการวิเคราะห์องค์ประกอบ 2 ขั้นตอน และยังใช้ตรวจสอบโค้งพัฒนาการแบบต่าง ๆ ได้อีกด้วย



แผนภาพที่ 2.1 โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงในการวิเคราะห์ข้อมูลระยะยาว (McArdle and Hamagami, 1995; อิทธิพงศ์ ตั้งสกุลเรืองไล, 2541)

แผนภาพที่ 2.1 มีการกำหนดชื่อย่อและสัญลักษณ์แทนค่าพารามิเตอร์ต่าง ๆ ในโมเดลดังต่อไปนี้

รูปสี่เหลี่ยม คือ ตัวแปรสังเกตได้

รูปวงรี คือ ตัวแปรแฝง

รูปสามเหลี่ยม คือ ค่าคงที่

ลูกศรทางเดียว คือ สัมประสิทธิ์การถดถอยของตัวแปรแฝงบนตัวแปรสังเกตได้

ลูกศรสองทาง คือ ค่าความสัมพันธ์แบบสมมาตร ได้แก่ ความแปรปรวน-ความแปรปรวนร่วม

ชื่อตัวแปรต่าง ๆ ได้แก่

$y_{(t)}$ คือ ตัวแปรสังเกตได้ในการวัดครั้งที่ t

L คือ ตัวแปรแฝงที่เป็นผลการวัดครั้งแรก

S คือ ตัวแปรแฝงความชันหรืออัตราการเปลี่ยนแปลง

L* คือ ความคลาดเคลื่อนของตัวแปรแฝงผลการวัดครั้งแรก

S* คือ ความคลาดเคลื่อนของตัวแปรแฝงความชัน

$E_{(t)}$ คือ ความคลาดเคลื่อนในการวัดครั้งที่ t

คือ ค่าคงที่ (constant) กำหนดให้เป็นตัวแปรสังเกตได้ และมีค่าเท่ากับ 1

$B_{(t)}$ คือ ค่าน้ำหนักองค์ประกอบพื้นฐานหรือพารามิเตอร์ที่เป็นอัตราการเปลี่ยนแปลงในการวัดครั้งที่ t ค่า $B_{(t)}$ ของทุกตัวแปรสังเกตได้อาจมีค่าเท่ากันหรืออาจมีค่าต่างกันตามที่นักวิจัยกำหนดหรืออาจเป็นพารามิเตอร์อิสระ ที่บอกลักษณะของพัฒนาการ

M_L คือ ค่าเฉลี่ยของตัวแปรแฝง L

M_S คือ ค่าเฉลี่ยของตัวแปรแฝงความชันหรือตัวแปรแฝง S

D_L คือ ส่วนเบี่ยงเบนมาตรฐานของตัวแปรแฝง L

D_S คือ ส่วนเบี่ยงเบนมาตรฐานของตัวแปรแฝงความชันหรือตัวแปรแฝง S

D_e คือ ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนในการวัดครั้งที่ t และ

RLS คือ ความแปรปรวนร่วมระหว่างความคลาดเคลื่อนของตัวแปรแฝง L กับความคลาดเคลื่อนของตัวแปรแฝงความชันหรืออัตราการเปลี่ยนแปลง S หรือค่าสัมประสิทธิ์สหสัมพันธ์ (correlation coefficient) ระหว่างองค์ประกอบของตัวแปรที่สัมพันธ์กับการเปลี่ยนแปลง

ลักษณะของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงดังกล่าวมีลักษณะเด่นคือ ความยืดหยุ่น (flexible) ของโมเดลที่อาจดัดแปลงโมเดลให้เหมาะสมกับโค้งพัฒนาการรูปแบบต่าง ๆ ได้ โดยการกำหนดค่าพารามิเตอร์ B(t) ในงานวิจัยของ McArdle and Hamagami (1991) มีการกำหนดค่าพารามิเตอร์แตกต่างกันเป็น 4 แบบ ซึ่งจะได้โมเดล 4 รูปแบบดังนี้ (อิทธิพงษ์ ตั้งสกุลเรืองไล, 2541)

1) โมเดลพัฒนาการเชิงเส้นโค้ง (nonlinear growth model) โมเดลนี้จะมีโค้งพัฒนาการที่ไม่เป็นเส้นตรง โดยการกำหนดพารามิเตอร์ B(t) ซึ่งอยู่ในรูปเวคเตอร์ให้เป็น [0, 2, -2, 1, -1] ซึ่งนักวิจัยสามารถกำหนดค่า B(t) ด้วยเวคเตอร์ที่มีค่าต่าง ๆ กัน เพื่อให้ได้โค้งพัฒนาการที่เหมาะกับข้อมูลได้

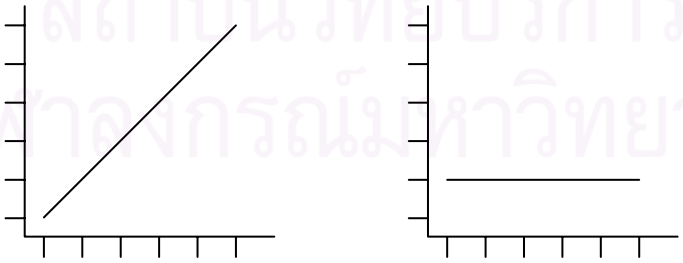
2) โมเดลพัฒนาการที่กำหนดพารามิเตอร์คงที่ (fixed arrows growth model) ในกรณีที่นักวิจัยทราบรูปแบบโค้งพัฒนาการชัดเจน หรือทราบค่าพารามิเตอร์ B(t) จากงานวิจัยที่มีผู้ทำไว้แล้ว นักวิจัยอาจกำหนดค่าเวคเตอร์ B(t) ตามลักษณะโค้งที่ต้องการได้

3) โมเดลพัฒนาการเชิงเส้นตรง (linear growth model) ในกรณีนี้ นักวิจัยเชื่อว่าข้อมูลที่เป็นการศึกษาระยะยาว มีลักษณะการเปลี่ยนแปลง หรือโค้งพัฒนาการเป็นแบบเส้นตรง นักวิจัยกำหนดค่าพารามิเตอร์เป็นเวคเตอร์ [0, 1, 2, 3, 4]

4) โมเดลพัฒนาการพัฒนาที่ไม่มีความชัน (no slope baseline growth model) มีการกำหนดค่าความชันพื้นฐานเป็น 0 โมเดลนี้นักวิจัยกำหนดค่าเวคเตอร์ในพารามิเตอร์ B(t) มีค่าเท่ากับ [0, 0, 0, 0, 0] ซึ่งหมายความว่า ลักษณะของโมเดลควรจะสอดคล้องกับข้อมูลน้อยที่สุด และใช้เป็นพื้นฐานในการเปรียบเทียบประสิทธิภาพของโมเดลอื่น ๆ ได้

ลักษณะของโมเดลพัฒนาการทั้ง 4 รูปแบบ แสดงได้ดังแผนภาพที่ 2.2

โมเดลพัฒนาการเชิงเส้นโค้ง โมเดลพัฒนาการที่มีการกำหนดค่าพารามิเตอร์คงที่



โมเดลพัฒนาการเชิงเส้นตรง โมเดลพัฒนาการที่ไม่มีความชัน

แผนภาพที่ 2.2 ลักษณะของโมเดลพัฒนาการตัวแปรแฝง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 2 การวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ

2.1 แนวคิดในการวัดคะแนนพัฒนาการ

จากการที่การวัดตามแนวทฤษฎีการตอบสนองข้อสอบได้มีการนำค่าพารามิเตอร์ของข้อสอบ และความสามารถที่แท้จริงของผู้สอบมาร่วมในการอธิบายความน่าจะเป็นที่จะตอบข้อสอบถูกในการสอบแต่ละครั้ง โดยอยู่ในรูปของฟังก์ชั่นเพิ่มทางเดียว (monotonically increasing function) ซึ่งจะแสดงให้เห็นถึงคนที่มีความสามารถสูงจะมีความน่าจะเป็นที่จะตอบถูกมากกว่าคนที่มีความสามารถต่ำบนคุณลักษณะเดียวกัน (Bock, 1997; Hambleton and Swaminathan, 1985) ดังนั้น การวัดพัฒนาการตามแนวทฤษฎีการตอบสนองข้อสอบจึงมีหลายโมเดล ซึ่งสามารถแยกโมเดลต่าง ๆ ออกเป็น 2 แนวคิด คือ แนวคิดแรก คะแนนพัฒนาการที่ได้ในระหว่างเวลา เกิดจากการเปลี่ยนแปลงของค่าพารามิเตอร์ของข้อคำถามในระหว่างเวลามากกว่าที่จะเกิดจากการเปลี่ยนแปลงของค่าพารามิเตอร์ของคนที่ถือว่าคงที่ (Fischer, 1997) แนวคิดที่สอง คะแนนพัฒนาการที่ได้ในระหว่างเวลา เกิดจากการเปลี่ยนแปลงของค่าพารามิเตอร์ของคนในระหว่างเวลา โดยมีค่าพารามิเตอร์ของข้อคำถามคงที่ (Embretson, 1991)

2.2 วิธีการวัดคะแนนพัฒนาการที่คะแนนพัฒนาการมาจากค่าพารามิเตอร์ของข้อคำถาม

Fischer ได้เสนอแนวคิดนี้ในปี 1973 โดยเสนอว่าผู้สอบหรือคนไข้ที่ตอบข้อสอบหรือคำถามในแบบสอบถามที่ให้ตอบหลาย ๆ ครั้ง หรือหลังจากที่ได้รับการจัดกระทำ ในแต่ละช่วงเวลานั้น การเปลี่ยนแปลงของคำตอบที่ได้ควรจะมาจากการเปลี่ยนแปลงของค่าพารามิเตอร์ของข้อสอบมากกว่าการเปลี่ยนแปลงของค่าพารามิเตอร์ของคน ซึ่งมีค่าคงที่ โดย Fischer ได้แนวคิดมาจากบทความของ Scheiblechner (1972) ที่นำค่าความยากในโมเดลราส์ชมาเป็นตัวแปรตามที่ทำนายด้วยกระบวนการทางพุทธิพิสัย (cognitive operations) Fischer ได้ขยายแนวคิดนี้ออกไปเป็นโมเดลใหม่ที่นำค่าความยากของข้อสอบในโมเดลราสชมาจัดให้อยู่ในรูปฟังก์ชั่นของพารามิเตอร์พื้นฐาน (basic parameter: η) ซึ่งพารามิเตอร์พื้นฐานได้แก่สิ่งที่ได้จากประสบการณ์ การฝึกอบรม หรือการรักษา เป็นต้น (Fischer, 1997) และเรียกโมเดลนี้ว่าโมเดลแบบทดสอบโลจิสติกเชิงเส้นตรง (Linear Logistic Test Model: LLTM) โดยมีสูตรดังนี้

$$P(+ \mid S_v I_i) = \frac{\exp(\theta_v - \beta_i)}{1 + \exp(\theta_v - \beta_i)}$$

$$= \frac{\exp(\theta_v - \sum_j q_{ij}\eta_j - c)}{1 + \exp(\theta_v - \sum_j q_{ij}\eta_j - c)}$$

+ แทน คำตอบที่ถูก/ตอบว่าใช่

θ_v แทน พารามิเตอร์ของคนที่ v

β_i แทน พารามิเตอร์ของข้อสอบ (ความยาก) ของข้อที่ i

η_i แทน พารามิเตอร์พื้นฐาน

q_{ij} แทน น้ำหนัก

c แทน ค่าคงที่

และมี metrix θ แสดงถึงพารามิเตอร์ของค่าความยากในช่วงเวลาต่าง ๆ

$$\theta = \begin{pmatrix} \eta_2 & \eta_3 & \eta_4 & \ldots.. & \eta_k \\ 0 & 0 & 0 & \ldots.. & 0 \\ 1 & 0 & 0 & \ldots.. & 0 \\ 0 & 1 & 0 & \ldots.. & 0 \\ 0 & 0 & 1 & \ldots.. & 0 \\ . & . & . & & . \\ . & . & . & & . \\ 0 & 0 & 0 & & 1 \end{pmatrix}$$

โดยที่ $\beta_i = \eta_i$, สำหรับ i = 2, …k

ค่า 0 ในแถวแรกหมายถึง $\beta_i = \eta_i = 0$ สำหรับทำเป็นค่าปกติ (normalization) LLTM ใช้กับการวัด 2 ครั้ง ข้อมูลแบบทวิวิภาค (dichotomous) ที่เป็นเอกมิติ (unidimension) โดยมีวิธีการวิเคราะห์แบบเดียวกับราส์ชโมเดล (Rasch model: RM) โดยใช้การวิเคราะห์ด้วยวิธีไลค์ลิฮูดสูงสุดแบบมีเงื่อนไข (conditional maximum likelihood: CML) และวิธี marginal maximum likelihood (MML)

ต่อมาได้มีการขยาย LLTM เป็นโมเดลต่าง ๆ ดังนี้

1) โมเดลโลจิสติกเชิงเส้นตรงที่มีการผ่อนคลายข้อตกลงเบื้องต้น (Linear Logistic Model with Relaxed Assumption: LLRA) ในปี 1974 Fischer (1983; 1989) ได้พัฒนาโมเดล LLTM ให้สามารถใช้ได้กว้างขวางขึ้น โดยใช้ได้กับข้อมูลที่ไม่จำกัดจำนวนมิติ (dimension) แต่ยังคงใช้กับการวัด 2 ครั้ง ข้อมูลที่เป็นแบบทวิวิภาค (dichotomous) โดยมีสูตรดังนี้

$$P(+ \mid I_i, S_v, T_1) = \frac{\exp(\theta_{vi})}{1 + \exp(\theta_{vi})}$$

$$P(+ \mid I_i, S_v, T_2) = \frac{\exp(\theta_{vi} + \delta_v)}{1 + \exp(\theta_{vi} + \delta_v)}$$

$$\delta_v = \sum_j q_{vj} \eta_j + \sum_{j<1} q_{vi} q_{vl} \rho_{jl} + \tau$$

T_1, T_2 แทน การวัดครั้งที่ 1, 2

θ_{vi} แทน ความสามารถของคนที่ v วัดด้วยข้อสอบข้อที่ i

q_{vj}, q_{vl} แทน ชุดของการจัดกระทำ (treatment) M_j, M_l ที่ให้แก่คนที่ v ระหว่างการวัดครั้งที่ 1, 2

η_j แทน อิทธิพลของการจัดกระทำ (treatment) M_j

τ แทน อิทธิพลแนวโน้มที่เกิดขึ้นระหว่างช่วงเวลา T_1 และ T_2

ρ_{jl} แทน ปฏิสัมพันธ์ระหว่างการจัดกระทำ (treatment) M_j, M_l

โดยที่ η_j และ τ ประมาณค่าด้วยวิธี CML และทดสอบสมมติฐานด้วยอัตราส่วนไลค์ลิฮูดแบบมีเงื่อนไข (conditional likelihood ratio: CLR)

$$\chi^2 \sim -2 \{\ln L_0 - \ln L_1\}; \; df = U_1 = U_0$$

U_0, U_1 แทนจำนวนพารามิเตอร์ในการทดสอบ H_0 และ H_1

H_0 : All $\eta_j - \rho_i = \tau = 0$ (ไม่มีการเปลี่ยนแปลง)

H_1 : $\eta_j = \rho_{jl} = 0$ สำหรับทุกการจัดกระทำ (treatment)

2) Hybrid – LLRA ในปี 1977 Fischer (1989) ได้ขยายโมเดล LLRA ออกไป โดย Hybrid – LLRA ใช้กับการวัดซ้ำด้วยข้อสอบที่ต่างชุดกัน โดยมีสูตรดังนี้

$$P(+ \mid S_v, D_i, T_t) = \frac{\exp(\theta_{vi} + \sigma_{it} + \delta_{gt})}{1 + \exp(\theta_{vi} + \sigma_{it} + \delta_{gt})}$$

โดยที่ $$\delta_{gt} = \sum_j \bar{q}_{gjt} + \eta_j + \sum_{j<l} \bar{q}_{gj} + \bar{q}_{gl} + \rho_{jl} + (T_t - T_l)\tau$$

S_v แทน คนที่ v ในกลุ่ม g

θ_{vi} แทน ความสามารถของคนที่ v ในคุณลักษณะที่ i

σ_{it} แทน ค่าความยากของข้อสอบที่วัดคุณลักษณะที่ i ในเวลา t

$\bar{q}_{gjt}$ แทน ชุดของ treatment M_j ที่ให้กับคนที่ v ในกลุ่ม g ในเวลา t (จำนวนหน่วยของการรักษาหรือการสอน)

η_j แทน อิทธิพลจากการจัดกระทำ (treatment) M_j

δ_{gt} แทน ขนาดของการเปลี่ยนแปลงในคนที่ v ในกลุ่ม g เวลา t

3) โมเดลมาตรประมาณค่าเชิงเส้นตรง (Linear Rating Scale Model: LRSM) Fischer และ Parzer (1991) ได้เสนอโมเดลนี้ในปี 1991 โดยได้ขยายมาตรประมาณค่า (Rating Scale) ของ Andrich (1978) ออกไปให้อยู่ในรูปของ LLTM โดยกระจายค่าพารามิเตอร์ของแต่ละขั้น (category) ออกไปในรูปของผลรวมของพารามิเตอร์อิทธิพล โดยมีสูตรดังนี้

$$P(K_{vi} = k/\theta_v;\beta_j;\omega_0,...\omega_m) = \frac{\exp[k(\theta_v+\beta_i)+\omega_k]}{\sum_{h=0}^{m}\exp[h(\theta_v+\beta_i)+\omega_h]}$$

K_{vi} แทน ค่าในขั้น (category) ของคนที่ v และข้อที่ I (ซึ่งมีค่า = 0, 1, 2, …, h, …m)

θ_v แทน ความสามารถของคนที่ v

β_i แทน ค่าความยากของข้อที่ i

$\omega_0, \ldots\omega_m$ แทน ค่าพารามิเตอร์ของ category

และ $\beta_i = \sum_{j=1}^{\eta} q_{ij}\alpha_j + c \;;\; i = 1, 2, ..., t$

โดยที่ q_{ij} เป็นน้ำหนักของพารามิเตอร์พื้นฐานที่ j ของข้อที่ i

α_j เป็นพารามิเตอร์พื้นฐานที่ j

LRSM ใช้กับการวัด 2 ครั้ง ที่วัดซ้ำด้วยข้อคำถามชุดเดิมที่เป็นมาตรประมาณค่า (rating scale) ที่วัดเพียงเอกมิติ (unidimension) โดยใช้ CML ในการประมาณค่าพารามิเตอร์ด้วยวิธี Newton-Raphson Algorithm และทดสอบสมมติฐานด้วย CLR

4) โมเดลการให้คะแนนความรู้บางส่วนเชิงเส้นตรง (Linear Partial Credit Model: LPCM) Fischer และ Ponocny (1994) ได้เสนอโมเดลนี้ในปี 1994 โดยขยายโมเดลการให้คะแนนความรู้บางส่วน (partial credit) ของ Andersen (1983) และ Masters (1982) ด้วยการหาค่าความน่าจะเป็นของการตอบข้อคำถามที่ i คนที่ v ในระดับความคิดเห็นที่ h โดยมีสูตรดังนี้

$$P(A_{vih} = 1/S_v, I_i) = \frac{\exp(h\theta_v + \beta_{ih})}{\sum_{t=0}^{m_i} \exp(t\theta_v + \beta_{it})}$$

$$\beta_{ih} = \sum_{j=1}^{p} w_{ihj}\alpha_j + hc$$

โดยที่ A_{vih} แทนการตอบของคนที่ v บนข้อที่ i ใน category ที่ h

$m_i + 1$ แทนจำนวนขั้น (category) ของข้อที่ i คือ c_O, …, c_m

β_{ih} แทนระดับความยากของขั้น (category) ที่ h ข้อที่ i

θ_v แทนความสามารถของคนที่ v

α_j แทนพารามิเตอร์พื้นฐานของการจัดกระทำ (treatment) ที่ j

w_{ihj} แทนน้ำหนักของพารามิเตอร์พื้นฐานที่ j ของข้อที่ i ขั้น (category) ที่ h

c แทนค่าคงที่ปกติ (normalization constant)

LPCM ใช้กับการวัด 2 ครั้ง ด้วยข้อคำถามชุดเดิมที่เป็นการให้คะแนนความรู้บางส่วน (partial credit) ที่วัดเอกมิติ (unidimension) โดยใช้ CML ในการประมาณค่าพารามิเตอร์โดยใช้วิธี A Quasi-Newton ที่เรียกว่า Broyden-Fletcher-Goldfarb-Shanno (BFGS) ในการหาอนุพันธ์แบ่งส่วนลำดับที่สอง (second-order partial อนุพันธ์ (derivative)s) ซึ่งให้ผลไม่ต่างจากวิธี Newton-Raphson แต่ BFGS ให้การคำนวณที่เร็วกว่า และทดสอบสมมติฐานด้วย CLR

ตารางที่ 2.1 สรุปคุณสมบัติของโมเดลที่วัดคะแนนพัฒนาการ

จำนวนครั้งที่วัด	ลักษณะแบบสอบ	ลักษณะของข้อมูล		
		ทวิวิภาค (dichotomous)	พหุภาค (polytomous)	
			ขั้น(category)	ลำดับ(ordered)
2	ชุดเดิม	LLRA	LLRA	LRSM, LPCM
2	ต่างชุดกัน	Hybrid – LLRA	-	LRSM, LPCM
มากกว่า 2	ชุดเดิม/ต่างชุดกัน	LLTM	Fischer, 1974 เสนอแนวคิด ยังไม่ได้ทดสอบแนวคิด	-

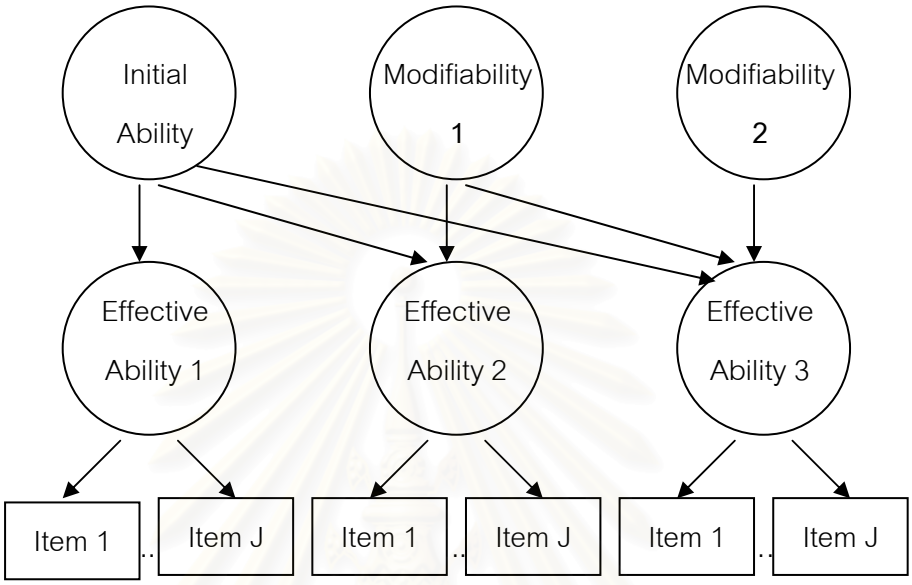
ในแนวคิดของ LLTM นี้ จะเห็นได้ว่ามีจุดเด่นในด้านของการนำค่าความยากของข้อสอบมามีส่วนร่วมในการประมาณค่าความสามารถ และสามารถที่จะใช้ข้อสอบต่างชุดกันเพื่อเลี่ยงปัญหาการจำข้อสอบได้ แต่ในโมเดลชุดนี้มีจุดอ่อนอยู่ที่พัฒนาการที่หาได้อยู่ในรูปของทั้งกลุ่ม ไม่ได้หาความสามารถที่พัฒนาได้รายบุคคล ดังนั้นการหาพัฒนาการด้วยโมเดลเหล่านี้จึงเป็นเพียงการทดสอบอิทธิพลของการจัดกระทำ (treatment) ระหว่างเวลาว่ามีความแตกต่างระหว่างกลุ่มที่ได้รับ/ไม่ได้รับหรือไม่ ซึ่งเหมือนกับการทดสอบด้วย ANOVA และ MANOVA แต่วิธีนี้จะดีกว่าในจุดที่เป็นการทดสอบจากความสามารถที่แท้จริง ไม่ใช่ทดสอบจากค่าคะแนนที่สังเกตได้ (observed score) ที่มีค่าความคลาดเคลื่อนอยู่ด้วย

2.3 วิธีการวัดคะแนนพัฒนาการที่คะแนนพัฒนาการมาจากค่าพารามิเตอร์ของผู้สอบ

คะแนนพัฒนาการวิธีนี้ประมาณค่าจากผลต่างของค่าพารามิเตอร์ของผู้สอบนั่นคือ คะแนนพัฒนาการได้จากการเปลี่ยนแปลงของค่าความสามารถของผู้สอบ (θ) ที่มีค่าพารามิเตอร์ของข้อคำถาม/ข้อสอบคงที่นั้น ผู้พัฒนาโมเดลตามแนวคิดนี้หลายคน เช่น Bock (1976) ได้ประมาณค่าคะแนนพัฒนาการโดยการหาเส้นโค้งของโมเดลทฤษฎีการตอบสนองข้อสอบที่กลมกลืน (fit) กับเส้นโค้งพัฒนาการ แต่ไม่สำเร็จ โดย Bock ได้เสนอว่าเป็นการยากที่โมเดลของพัฒนาการจะกลมกลืน (fit) กับเส้นโค้งพัฒนาการทางสมองได้ เนื่องมาจากค่าความเที่ยงของการวัดแต่ละครั้งมีค่าต่ำเกินไปที่จะใช้ทดสอบสมมติฐานเกี่ยวกับแนวโน้มได้ Keats (1983) ได้เสนอการประมาณค่าพารามิเตอร์ด้วยโมเดลโค้งพัฒนาการแบบยกกำลัง (exponential) และโค้งไฮเปอร์โบลา (hyperbolic) ในปี 1985 Anderson ได้เสนอโมเดลราส์ชพหุมิติสำหรับการวัดพัฒนาการของคนที่มีการวัดหลาย ๆ ครั้ง ซึ่งโมเดลของ Anderson ให้ค่าความยากของข้อสอบมีค่าคงที่ และความสามารถของผู้สอบเปลี่ยนไปตามช่วงเวลา แต่โมเดลของ Anderson ไม่สามารถให้ค่าคะแนนพัฒนาการเป็นรายบุคคลได้ (Embretson, 1991a; 1991b) ต่อมาได้มีโมเดลที่พัฒนาและสามารถประมาณค่าคะแนนพัฒนาการเป็นรายบุคคลได้ดังนี้

1) โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (Multidimensional Rasch Model for Learning and Change: MRMLC) ในปี 1991 Embretson เสนอแนวคิดในการใช้โมเดลราส์ชพหุระดับวัดคะแนนพัฒนาการที่ได้จากการวัดตั้งแต่ 2 ครั้งขึ้นไปใช้กับข้อมูลแบบทวิวิภาค (dichotomous) โดยนิยามว่าความสามารถที่ถูกวัดมี 2 ประเภท คือ ความสามารถที่เป็นผล (effective ability) และความสามารถในการเรียนรู้ หรือความสามารถที่เพิ่มขึ้น (learning ability or modifiability) ความสามารถที่เป็นผล เป็นศักยภาพแฝงของผู้สอบที่อยู่

ภายใต้เงื่อนไขของเวลาในการวัด ความสามารถที่เป็นผลในการวัดครั้งแรกก็คือ ความสามารถเริ่มต้น (initial ability) ความสามารถในการเรียนรู้เป็นความสามารถเพิ่มขึ้นในระหว่างการวัด 2 ครั้งใด ๆ ที่อยู่ติดกัน ความสามารถที่เป็นผลในการวัดครั้งหลัง ๆ จะประกอบด้วยความสามารถเริ่มต้น และความสามารถในการเรียนรู้ 1 หรือมากกว่า 1 ตัว



แผนภาพที่ 2.3 โมเดล MRMLC

MRMLC มีข้อตกลงเบื้องต้นว่าความสามารถที่เพิ่มขึ้นของผู้สอบระหว่างเวลามีโครงสร้างเป็นแบบ Wiener Simplex กล่าวคือ ความสามารถของผู้สอบจะมีแต่เพิ่มขึ้น ไม่มีลดลงจากเดิม

จากโมเดลราส์ชพื้นฐาน

$$P(x_{ij} = 1/\theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)} \quad \text{.....} \quad (1)$$

θ_j คือความสามารถของคนที่ j ที่จะตอบข้อที่ i ถูก

b_i คือค่าความยากของข้อสอบข้อที่ i

ในการขยายโมเดลราส์ชเพื่อใช้ในการวัดพัฒนาการสำหรับความสามารถ M ตัว ในการตอบข้อสอบที่มีการวัด k ครั้ง สำหรับการวัดครั้งแรก (K = 1), ความสามารถจะขึ้นอยู่กับ

ความสามารถเริ่มต้น และในการวัดครั้งหลัง (K > 1) ความสามารถจะขึ้นอยู่กับ k – 1 ตัว ของความสามารถที่เพิ่มขึ้น (Modifiability) และความสามารถเริ่มต้น

MRMLC กำหนดให้ค่าอำนาจจำแนกของข้อสอบภายในการวัด K ครั้ง สำหรับวัดความสามารถ M ตัว อยู่ในรูปเมทริกซ์ K × M ดังนี้

$$A = \begin{bmatrix} 1 & 0 & ..... & 0 \\ 1 & 1 & ..... & 0 \\ . & & & \\ . & & & \\ 1 & 1 & ..... & 1 \end{bmatrix}$$

เมื่อ K = M; โครงสร้างของค่าอำนาจจำแนกระหว่างเวลาของการวัดนี้ จะเหมือนกับโมเดล Wiener Simplex ของน้ำหนักองค์ประกอบของ Joreskog

ถ้าให้อำนาจจำแนกของข้อสอบทุกข้อในการวัดครั้งเดียวกันมีค่าเท่ากัน จะได้โมเดลดังนี้

$$P(x_{i(k)j} = 1/\theta_j, b_i, a_k) = \frac{\exp(\sum_m a_{km}\theta_{jm} - b_i)}{1 + \exp(\sum_m a_{km}\theta_{jm} - b_i)} \quad \text{.....} \quad (2)$$

โดยที่ θ_j คือ เวคเตอร์ของความสามารถ ที่ซึ่ง θ_{j1} เป็นความสามารถที่ k = 1, θ_{j2} ... θ_{jm} เป็นความสามารถที่เพิ่มขึ้น เมื่อ k > 1

b_i คือ ค่าความยาก

a_k คือ เวคเตอร์ของค่าอำนาจจำแนก a_{km} เป็นค่าอำนาจจำแนกของข้อสอบในการวัดครั้งที่ k บนมิติของความสามารถ m a_{k1}, ... a_{km} เป็นสมาชิกในเมทริกซ์ K × M ของเมทริกซ์ A a_k เป็นเวคเตอร์แถว

จากสมการที่ (2) ค่าอำนาจจำแนกในโมเดลราส์ชโดยทั่วไป จะกำหนดให้มีค่าเป็น 1 ดังนั้น ค่าอำนาจจำแนกของข้อสอบทุกข้อมีค่าเป็น 1 ใส่ในสมการที่ (2) และเพราะว่าการสอบ k ครั้ง เกี่ยวข้องกับความสามารถที่เพิ่มขึ้น k ตัว เท่านั้น, a_{km} จึงสามารถตัดออกจากโมเดลได้ โดยเขียนในรูปผลรวมถึง k คือ $\sum_{m=1}^{k}\theta_{jm}$ ดังนั้น โมเดลราส์ชพหุมิติสำหรับการวัดการเปลี่ยนแปลงจึงสามารถเขียนในรูป

$$P(x_{i(k)j} = 1/\theta_j, b_i) = \frac{\exp(\sum_{m=1}^{k}\theta_{jm} - b_i)}{1 + \exp(\sum_{m=1}^{k}\theta_{jm} - b_i)} \quad \text{.....} \quad (3)$$

ดังนั้น ข้อสอบ 1 ข้อ ที่วัดในครั้งที่ k เกี่ยวข้องกับความสามารถที่ k ความสามารถเริ่มต้น และความสามารถที่เพิ่มขึ้น k-1 ตัว ความสามารถที่เปลี่ยนแปลงไปจากการวัดครั้งที่ k-1 กับครั้งที่ k ขึ้นอยู่กับความสามารถที่เพิ่มขึ้นตัวที่ m ซึ่งเกี่ยวกับการวัดครั้งที่ k

สมการที่ (3) แสดงให้เห็นถึงโมเดล MRMLC ที่เป็นพหุมิติ อย่างไรก็ตามใน MRMLC ภายในการวัดแต่ละครั้ง ความน่าจะเป็นของการตอบข้อสอบถูก สามารถหาได้จากโมเดลเอกมิติ (unidimension)

สำหรับทุกข้อภายในการวัดครั้งที่ k และทุก k ความน่าจะเป็นของการตอบข้อสอบถูกขึ้นอยู่กับความสามารถประกอบ θ_{jk}^{c} (composite ability) ตัวเดียวกัน ซึ่งเป็นผลรวมของความสามารถเริ่มต้นกับความสามารถที่เพิ่มขึ้น k-1 ตัว

ดังนั้น ความน่าจะเป็นที่คนที่ j จะตอบข้อที่ i ถูกในการวัดครั้งที่ k เป็นไปตามสมการ

$$P(x_{i(k)j} = 1/\theta_{jk}^{c}, b_i) = \frac{\exp(\theta_{jk}^{c} - b_i)}{1 + \exp(\theta_{jk}^{c} - b_i)} \qquad \text{.....} \quad (4)$$

ที่ซึ่ง θ_{jk}^{c} คือความสามารถของคนที่ j ในการวัดครั้งที่ k

b_i คือค่าความยากของข้อที่ i

สมการที่ (4) นี้เป็นโมเดลราสชเอกมิติภายในการวัดครั้งที่ k คะแนนรวมหาได้จากการประมาณค่า θ_{jk}^{c} ถึงแม้ว่า θ_{jk}^{c} จะเป็นพหุมิติ เพราะได้มาจากความสามารถเริ่มต้นและความสามารถที่เพิ่มขึ้นหลายตัวใน MRMLC

การประมาณค่าพารามิเตอร์ ได้มีการพัฒนาตัวประมาณค่าแบบไลค์ลิฮูดสูงสุดแบบมีเงื่อนไข (conditional maximum likelihook: CML) ใช้ในการประมาณค่าพารามิเตอร์ของโมเดลนี้

วิธีการดำเนินการ (a) ออกแบบโครงสร้างของข้อมูล ซึ่งควรจะใส่รวมอยู่ใน likelihood ของข้อมูล (b) หาค่าสถิติสำหรับคนภายในการวัดแต่ละครั้ง คือ θ_{jk}^{c}

ในการวัดหลายครั้ง จำเป็นที่ข้อสอบแต่ละข้อต้องถูกวัดในแต่ละครั้ง แต่เพื่อให้สอดคล้องกับข้อตกลงเบื้องต้นในเรื่องของความเป็นอิสระของข้อ (local independent) ดังนั้นข้อสอบแต่ละข้อจึงควรที่จะถูกวัดเพียงครั้งเดียว ซึ่งจะช่วยป้องกันอิทธิพลต่าง ๆ ได้ เช่น การจำได้ อิทธิพลจากความคงที่ในการตอบ และอิทธิพลจากการฝึกทำ เป็นต้น ดังนั้น ในการออกแบบการเก็บข้อมูลจึงควรที่จะแบ่งเป็นกลุ่มของข้อสอบที่สอบกับกลุ่มคนไม่ให้ซ้ำกัน เช่น ในการออกแบบการเก็บข้อมูล 3 ครั้ง จะแบ่งคนออกเป็น 3 กลุ่ม และแบ่งข้อสอบออกเป็นชุดดังในตารางที่ 2.2

ตารางที่ 2.2 โครงสร้างข้อมูลที่ใช้เก็บข้อมูล 3 ครั้ง

กลุ่มคนสอบ	ข้อสอบ			
	ข้อสอบร่วม	ข้อสอบชุดที่ 1	ข้อสอบชุดที่ 2	ข้อสอบชุดที่ 3
1	pretest	pretest	posttest 1	posttest 2
2	pretest	posttest 2	pretest	posttest 1
3	pretest	posttest 1	posttest 2	pretest

ในการวัดครั้งแรก ข้อสอบแต่ละชุดได้ถูกกำหนดให้วัดตามกลุ่มที่ได้ออกแบบไว้

ในการวัดครั้งที่ 2 θ_{jk}^{c} ในการวัดแต่ละครั้งให้ค่าประมาณเหมือนกับค่า θ_j ในเวคเตอร์ M มิติ ตามที่กำหนดในสมการที่ (2) ภายในการวัดครั้งที่ k ความน่าจะเป็นที่จะตอบข้อที่ i ถูกต้องของคนที่ j ต้องมีความสามารถ θ_j และมีน้ำหนัก a_k ซึ่งมีค่าเหมือนกับความสามารถประกอบตัวเดียว θ_{jk}^{c}

$$P(x_{i(k)j} = 1/a_k, b_i, \theta_j) = P(x_{i(k)j} = 1/b_i, \theta_{jk}^{c}) \quad \text{.....} \quad (5)$$

ในการประมาณค่า b_i ใน (2) หรือ (3) ถูกพัฒนาให้อยู่ในรูป θ_{jk}^{c}

แม้ว่า MRMLC จะสามารถใช้กับการวัดหลายครั้ง เพื่อให้โมเดลนี้ชัดเจน จึงขอเสนอในรูปการวัด 2 ครั้ง

เริ่มต้นที่สมการ (5) ความน่าจะเป็นที่การตอบข้อที่ $x_{i(k)j}$ เท่ากับ x (x มีค่าเป็น 0 หรือ 1) สามารถที่จะกำหนดภายในการวัดแต่ละครั้ง ความน่าจะเป็นของการตอบ คือ $P_{i(kj} = x/\theta_{jk}^{c}$, ε_i โดยที่ $\varepsilon_i = -b_i$ เป็นค่าความง่าย ดังนั้น ความน่าจะเป็นที่จะตอบข้อที่ i ในแต่ละครั้งของการสอบ มีดังนี้

$$P(\theta_1^c) = \frac{\exp(\theta_{j1}^c + \varepsilon_i)^{x_{i(k)j}}}{1 + \exp(\theta_{i1}^c + \varepsilon_i)} \quad \text{.....} \quad (6)$$

$$P(\theta_2^c) = \frac{\exp(\theta_{j2}^c + \varepsilon_i)^{x_{i(k)j}}}{1 + \exp(\theta_{j2}^c + \varepsilon_i)}$$

โดยที่ $P(\theta_1^c)$ เป็นความน่าจะเป็นที่คนที่ j จะตอบข้อที่ i ถูกในการวัดครั้งที่ 1

$P(\theta_2^c)$ เป็นความน่าจะเป็นที่คนที่ j จะตอบข้อที่ i ถูกในการวัดครั้งที่ 2

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

k_{ig} เป็นตัวแปรทวิวิภาคที่จะระบุถึงครั้งที่ของการวัด ที่ข้อสอบข้อที่ i ถูกสอบในกลุ่ม g

$$k_{ig} = \begin{cases} 1 & \text{ถ้าข้อสอบข้อที่ i ถูกวัดในครั้งที่ 1} \\ 0 & \text{ถ้าข้อสอบข้อที่ i ถูกวัดในครั้งที่ 2} \end{cases}$$

P_g คือความน่าจะเป็นของแบบแผนการตอบในกลุ่ม g สำหรับข้อสอบ I ข้อ บนเวคเตอร์ของความสามารถส่วนประกอบ θ_j^c สำหรับ θ_{j1}^c, ..., θ_{jk}^c และเวคเตอร์ของความยากของข้อสอบ $\boldsymbol{\varepsilon}$ สำหรับ $\boldsymbol{\varepsilon}_i$, ..., $\boldsymbol{\varepsilon}_I$ และเวคเตอร์ครั้งที่ของการสอบของข้อ k_g สำหรับ k_{ig}, ..., k_{Ig}

$$P_g(x_{1(k)j},...x_{I(k)j}/\theta_j^c,\boldsymbol{\varepsilon},k_g) = \pi_i P(\theta_1^c)^{k_{ig}} \pi_i P(\theta_2^c)^{1-k_{ig}} \quad \text{.....} \quad (7)$$

แทนค่า $P(\theta_1^c)$ และ $P(\theta_2^c)$ จากสมการ (6) ใน (7)

$$\begin{aligned} \ln P_g(x_{1(k)j},...x_{I(k)j}/\theta_j^c,\boldsymbol{\varepsilon},k_g) &= \Sigma_i k_{ig} x_{i(k)j}(\theta_{j1}^c+\varepsilon_i) \\ &-\ln(1-\exp(\theta_{j1}^c+\varepsilon_i))+\Sigma_i(1-k_{ig})(x_{i(k)j}(\theta_{j2}^c+\varepsilon_i)) \\ &-\ln(1+\exp(\theta_{j2}^c+\varepsilon_i)) \quad \text{.....} \quad (8) \end{aligned}$$

r_{jkg} คือคะแนนรวมสำหรับการวัดครั้งที่ k ภายในกลุ่ม g สำหรับคนที่ j มีสมการดังนี้

$$\begin{aligned} r_{j1g} &= \Sigma_i k_{ig} x_{i(k)j} \\ r_{j2g} &= \Sigma_i (1-k_{ig}) x_{i(k)j} \quad \text{.....} \quad (9) \end{aligned}$$

เมื่อแยกผลบวกในสมการ (8) และแทนค่าคะแนนรวมจากสมการ (9)

$$\begin{aligned} \ln P_g(x_{1(k)j},...x_{I(k)j}/\theta_j^c,\boldsymbol{\varepsilon},k_g) &= r_{j1g}\theta_{j1}^c + r_{j2g}\theta_{j2}^c \\ &+\Sigma_i k_{ig} x_{i(k)j}\varepsilon_i + \Sigma_i(1-k_{ig})x_{i(k)j}\varepsilon_i \\ &-\Sigma_i k_{ig}\ln(1+\exp(\theta_{j1}^c+\varepsilon_i)) \\ &-\Sigma_i(1-k_{ig})\ln(1+\exp(\theta_{j2}^c+\varepsilon_i)) \quad \text{.....} \quad (10) \end{aligned}$$

เพื่อพัฒนาตัวประมาณค่า CML สำหรับค่าพารามิเตอร์ ε_i จึงต้องใช้เวคเตอร์ของค่าคาดหวังของการตอบในรูปคะแนนดิบ r_{jkg} log ของความน่าจะเป็นที่คนที่ j ได้คะแนนดิบ r_{j1g}

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

และ r_{j2g} ในการวัด 2 ครั้ง เงื่อนไขของ θ_{j1}^{c} และ θ_{j2}^{c} ถูกกำหนดโดยการเพิ่ม k_{ig} เข้าไปในผลรวมภายในการวัดครั้งที่ k ให้ความน่าจะเป็นของคะแนน r_{j1g} และ r_{j2g} ดังสมการต่อไปนี้

$$\ln P(r_{j1g}, r_{j2g} / \theta_{j1}^{c}, \theta_{j2}^{c}, k_g) = r_{j1g}\theta_{j1}^{c} + r_{j2g}\theta_{j2}^{c} + \ln\gamma_{rg1}$$
$$+ \ln\gamma_{rg2} - \sum_{i} k_{ig} \ln(1+\exp(\theta_{j1}^{c} + \varepsilon_i))$$
$$- \sum_{i} (1-k_{ig}) \ln(1+\exp(\theta_{j2}^{c} + \varepsilon_j)) \quad \text{.....} \quad (11)$$

โดยที่ γ_{rg1} และ γ_{rg2} เป็นสมาชิกของฟังก์ชั่นสมมาตรของคะแนนดิบของคนที่ j ในกลุ่ม g ภายใต้การวัดครั้งที่ 1 และ 2 ตามลำดับ

ความน่าจะเป็นของเวคเตอร์การตอบสนองข้อสอบ, $x_{1(k)j}$, ..., $x_{l(k)j}$ สำหรับคนที่ j ในกลุ่ม g (x_j) บนคะแนนดิบ ได้มาจากการลบความน่าจะเป็นของเวคเตอร์การตอบสนองด้วยความน่าจะเป็นของคะแนนดิบ ได้สมการดังนี้

$$\ln P_g(x_j / r_{j1g}, r_{j2g}) = \sum_{i} x_{i(k)j}\varepsilon_i - \ln r_{rg1} - \ln\gamma_{rg2} \quad \text{.....} \quad (12)$$

ผลรวมระหว่างคนและกลุ่มให้ค่า log likelihood ของข้อมูลเป็น ln L ดังสมการ

$$\ln L = \sum_{g}\sum_{i} m_{ig}\varepsilon_i - \sum_{g}\sum_{k}\sum_{r=1}^{l_{kg}-1} n_{rgk} \ln\gamma_{rgk} \quad \text{.....} \quad (13)$$

โดยที่ m_{ig} คือความถี่ marginal สำหรับผ่านข้อที่ i ในกลุ่ม g

n_{rgk} คือความถี่ของคะแนน r ในกลุ่ม g ในการวัดครั้งที่ k

l_{kg} คือจำนวนข้อที่ใช้ในการวัดครั้งที่ k ในกลุ่ม g

ในสมการที่ (13) ความสามารถถูกแยกตามการวัดแต่ละครั้งภายในกลุ่มแต่ละกลุ่ม ดังนั้น สมาชิกของฟังก์ชั่นสมมาตร γ_{rgk} ถูกคำนวณระหว่างเวลาภายในเงื่อนไขเดียวกันภายในกลุ่ม

สมการสำหรับประมาณค่าแบบไลค์ลิฮูดสูงสุด (maximum likelihood) มีสมการดังนี้

$$0 = \sum_{g} m_{ig} - \sum_{g}\sum_{k}\sum_{r}^{l_{kg}-1} n_{rgk} \exp(\varepsilon_i)\left(\frac{\gamma_{rgk}^{(i)} - 1}{\gamma_{rgk}}\right) \quad \text{.....} \quad (14)$$

ที่ซึ่ง $\gamma^{(i)}_{rgk}$ - 1 เป็น partial deviative ของ γ_{rgk} ที่สัมพันธ์กับ ε_i สมการของ ε_i สามารถถูกหาค่าด้วยการทวนซ้ำด้วยวิธีของ gradient หรือ Newton-Raphson ค่าความง่าย ε_i ถูกให้แปลงค่าเป็นค่าความยาก b_i

การประมาณค่า CML สามารถหาได้จากโปรแกรมคอมพิวเตอร์ที่ยอมให้กำหนดค่าพารามิเตอร์ และแยก marginalization ระหว่างการวัดและระหว่างกลุ่ม ตัวอย่างเช่น โปรแกรม LLTM ของ Fischer และ Forman (1972) และ LINLOG ของ Whitely และ Nieh (1981) ความไม่แน่นอนของมาตร สามารถหาค่าใหม่ได้ด้วยโปรแกรมเหล่านี้ โดย a) ให้ค่าเฉลี่ยของความยากเป็น 0 b) เชื่อมโยงการประมาณค่าระหว่างกลุ่มโดยการใช้ข้อสอบที่อยู่ในการวัดครั้งเดียวกันของแต่ละกลุ่ม c) กำหนดให้ค่าอำนาจจำแนกมีค่าเป็น 1 ภายในการวัดครั้งเดียวกัน

การประมาณค่าความสามารถ (Ability estimates) ความสามารถใน MRMLC สามารถประมาณค่าโดยใช้ไลค์ลิฮูดสูงสุด (maximum likelihood) โดยใช้วิธี Newton-Raphson พหุระดับ ซึ่งวิธีนี้ต้องใช้ฟังก์ชั่น log likelihood สำหรับรูปแบบการตอบที่แตกต่างกัน (differentiable) 2 ครั้ง กับความสามารถ θ_j

เริ่มต้นโดยการที่คนที่ j ที่อยู่ในกลุ่ม g ได้ทำข้อสอบ I ข้อ ตามเงื่อนไขของข้อสอบในเวคเตอร์ k_g ดังในสมการที่ (8) เพราะว่าข้อที่ i ควรถูกใช้สอบในบางการวัด ค่าอำนาจจำแนกบนมิติภายใต้การวัดครั้งที่ k ($a_{i(k)m}$) คล้ายกับสมการ (7), MRMLC เขียนอยู่ในรูปน้ำหนัก $a_{i(k)m}$ ที่ซึ่งถูกกำหนดให้มีค่าเท่ากันทุกข้อ ภายในการวัดครั้งที่ k บนมิติที่ m ดังสมการ

$$P(x_{i(k)j} = 1/b_i, a_k, \theta_j) = P(\theta_j) = \frac{\exp(\sum_m a_{i(k)m}\theta_{jm} - b_i)}{1 + \exp(\sum_m a_{i(k)m}\theta_{jm} - b_i)} \quad \text{.....} \quad (15)$$

log ความน่าจะเป็นของเวคเตอร์การตอบของคนที่ j, x_j ประกอบด้วยความน่าจะเป็นทั้งสำหรับข้อที่ทำได้ $P(\theta_j)$ และข้อที่ทำไม่ได้ $1\text{-}P(\theta_j)$ จาก (3) เขียนเป็นสมการได้ดังนี้

$$\ln P(x_j / b, A_g, \theta_j) = \sum_i [x_{i(k)j} \ln P(\theta_j) + (1 - x_{i(k)j}) \ln(1 - P(\theta_j)] \quad \text{.....} \quad (16)$$

ที่ซึ่ง A_g คือเมทริกซ์ I×M ของน้ำหนักของข้อสอบ I ข้อ บนมิติ m มิติ ที่ใช้สอบคนที่ j ภายในกลุ่ม g สมาชิกของ A_g คือ $a_{i(k)m}$

partial อนุพันธ์ (derivative) อันดับแรกของสมการ (16) ที่สัมพันธ์กับ θ_{jm} เขียนเป็นสมการดังนี้

$$\frac{\partial \ln P}{\partial \theta_{jm}} = \sum_i a_{i(k)m}[x_{i(k)j} - P(\theta_j)] \quad \text{.....} \quad (17)$$

partial อนุพันธ์ (derivative) อันดับ 2 ของฟังก์ชั่น log likelihood เขียนเป็นสมการของ Variance ได้ดังนี้

$$\frac{\partial^2}{\partial \theta_{jm}^2} = -\sum_i a_{i(k)m}^2 P(\theta_j)(1-P(\theta_j)) \quad \text{.....} \quad (18)$$

สำหรับ covariance เขียนได้เป็นสมการดังนี้

$$\frac{\partial^2}{\partial \theta_{j1}\theta_{j2}} = -\sum_i a_{i(k)1}a_{i(k)2} P(\theta_j)(1-P(\theta_j)) \quad \text{.....} \quad (19)$$

วิธี Newton-Raphson ทำการทวนซ้ำเพื่อแก้สมการการประมาณค่าที่ใช้กันอยู่ทั่วไปได้ถูกปรับปรุงโดยการลบอัตราส่วนของเวคเตอร์ของอนุพันธ์ (derivative) ครั้งแรกออกจากเมทริกซ์ของอนุพันธ์ (derivative) อันดับ 2 จนกระทั่งได้ค่าลู่เข้าสู่เกณฑ์ที่กำหนด เมทริกซ์สารสนเทศ (H_θ) สำหรับความสามารถให้คูณด้วย –1 เท่าของค่าคาดหวังของอนุพันธ์ (derivative) อันดับ 2 ซึ่งมีสมาชิกของความสามารถเริ่มต้นและความสามารถเพิ่มขึ้น 1 ตัว ($a_{i(k)m}$) และค่าอำนาจจำแนกที่เป็นค่าทวิวิภาคสำหรับข้อที่ i บนมิติ m เมื่อการวัดอยู่ภายใต้เงื่อนไข k :

$$H_\theta = \begin{bmatrix} \sum_i a_{i(k)1}^2 P(\theta_j)(1-P(\theta_j)) & \sum_i a_{i(k)1}a_{i(k)2} P(\theta_j)(1-P(\theta_j)) \\ \sum_i a_{i(k)1}a_{i(k)2} P(\theta_j)(1-P(\theta_j)) & \sum_i a_{i(k)2}^2 P(\theta_j)(1-P(\theta_j)) \end{bmatrix} \quad \text{.....} \quad (20)$$

วิธีไลค์ลิฮูดสูงสุด (maximum likelihood: ML) สำหรับประมาณค่าความสามารถ สามารถใช้ได้กับโมเดลที่การแจกแจงไม่ได้กำหนดไว้แต่ต้น แต่ไม่สามารถที่จะประมาณค่าคะแนนที่เป็น 0 กับคะแนนเต็มได้ วิธีการประมาณค่าด้วย Bayesian สามารถประมาณค่าคะแนนที่ได้ 0 หรือเต็มได้ แต่การแจกแจงต้องมีลักษณะพิเศษ

คุณสมบัติของโมเดล

จำนวนมิติ (Dimensionality) MRMLC ได้กำหนดไว้อย่างชัดเจนแล้วว่าการเปลี่ยนแปลงเป็นแบบพหุมิติระหว่างการวัด ถ้าข้อสอบชุดเดิมใช้สอบหลายครั้ง การตอบไม่ได้มีค่า

ความคาดหวังค่าเดียวกัน เวคเตอร์ของค่าอำนาจจำแนกของข้อสอบที่มี 2 ค่าสำหรับมิติที่ M สำหรับการวัดที่ k, a_k และเวคเตอร์ความสามารถบนมิติที่ M, θ_j

$$P(x_{i(k)j} = 1/b_i, a_k, \theta_j) \neq P(x_{i(k')j} = 1/b_i, a_k, \theta_j) \quad \text{.....} \quad (21)$$

สำหรับทุกคู่ของ k, k' ที่ซึ่ง $k \neq k'$ มีเพียงกรณีเดียวเท่านั้นที่ผิดปกติ คือกรณีที่ความสามารถเป็น 0 สำหรับค่าอำนาจจำแนกไม่ตรงกันระหว่างครั้งที่ k กับ k' เพราะว่าค่าอำนาจจำแนกสำหรับความสามารถ a_k ไม่ได้แปรค่าระหว่างข้อภายในการวัดครั้งที่ k

MRMLC ถือว่าเป็นเอกมิติภายในการวัดแต่ละครั้ง ความสามารถส่วนประกอบเดี่ยวให้ความน่าจะเป็นของการตอบสนองข้อสอบค่าเดียวกันในรูปของเวคเตอร์ความสามารถและความสามารถที่เพิ่มขึ้น

ความคลาดเคลื่อนมาตรฐานของความสามารถ ในการประมาณค่าด้วย ML ความแปรปรวนของความคลาดเคลื่อนของความสามารถ เมื่อพารามิเตอร์ของข้อถูกกำหนดให้อยู่บนเส้นทแยงของ inverse ของเมทริกซ์สารสนเทศ กำหนดให้คนที่ j ได้รับการทดสอบด้วยข้อสอบ 1 ชุด ในการสอบครั้งที่ 1 ($1 \leq i \leq S_1$-1) และข้อสอบอีกชุดใช้สอบในครั้งที่ 2 ($S_1 \leq i \leq I$) ภายใต้ MRMLC ทุกข้อมีค่าอำนาจจำแนกเท่ากันบนความสามารถเริ่มต้น ดังนั้น $a_{i(k)1} = 1$ สำหรับทุก i และ k ค่าอำนาจจำแนกของความสามารถเพิ่มขึ้น ขึ้นอยู่กับครั้งที่ของการวัด ดังนั้นสำหรับข้อสอบที่สอบครั้งแรก ค่าอำนาจจำแนกเท่ากับ 0 ($a_{i(2)2} = 0$ สำหรับทุก i = 1, ..., S_1-1) และข้อสอบชุดที่สอบในครั้งที่ 2 ค่าอำนาจจำแนกเท่ากับ 1 ($a_{i(2)2} = 1$ สำหรับ i = S_1, ..., I) จากน้ำหนักที่มีรูปแบบนี้ และจากสมการที่ (18), (19) เมทริกซ์สารสนเทศมีสมการดังนี้

$$H_\theta = \begin{bmatrix} \sum_{i=1}^{I} P(\theta_j)(1-P(\theta_j)) & \sum_{i=S_1}^{I} P(\theta_j)(1-P(\theta_j)) \\ \sum_{i=S_1}^{I} P(\theta_j)(1-P(\theta_j)) & \sum_{i=S_1}^{I} P(\theta_j)(1-P(\theta_j)) \end{bmatrix} \quad \text{.....} \quad (22)$$

ผลรวมระหว่างข้อถูกแบ่งได้ดังสมการต่อไปนี้

$$\sum_{i=1}^{I} P(\theta_j)(1-P(\theta_j)) = \sum_{i=1}^{S_1-1} P(\theta_j)(1-P(\theta_j)) + \sum_{i=S_1}^{I} P(\theta_j)(1-P(\theta_j)) \quad \text{.....} \quad (23)$$

inverse ของ H_θ ให้เมทริกซ์ covariance อย่างง่ายสัมพัทธ์ของ θ_j มีสมการ คือ

$$S_\theta = \begin{bmatrix} \dfrac{1}{\sum_{i=1}^{S_1-1} P(\theta_j)(1-P(\theta_j))} & \dfrac{-1}{\sum_{i=1}^{S_1-1} P(\theta_j)(1-P(\theta_j))} \\ \dfrac{-1}{\sum_{i=1}^{S_1-1} P(\theta_j)(1-P(\theta_j))} & \dfrac{\left(\sum_{i=1}^{I} P(\theta_j)(1-P(\theta_j))\right)}{k} \end{bmatrix} \quad \text{.....} \quad (24)$$

โดยที่ $k = \left(\sum_{i=1}^{S_1-1} P(\theta_j)(1-P(\theta_j))\right)\left(\sum_{i=S_1}^{I} P(\theta_j)(1-P(\theta_j))\right)$

เส้นทแยงของ S_θ คือความแปรปรวนของความคลาดเคลื่อน (error variances) และนอกเส้นทแยงมุมคือค่า covariance ของความคลาดเคลื่อน

วัตถุประสงค์หลัก (Specific Objectivity) การประมาณค่าพารามิเตอร์ของข้อสอบเป็นวัตถุประสงค์หลัก ถ้าพารามิเตอร์ของคนถูกตัดออกไปก่อนที่จะเข้าสู่กระบวนการประมาณค่า ให้ใช้สมการที่ (14) ที่ซึ่งใช้การประมาณค่าพารามิเตอร์ด้วยตัวประมาณค่า CML

ค่าของมาตร (Scale values) ในสมการ (3) ความไม่แน่นอนของมาตรของ MRMLC ได้แก้ใหม่ด้วยการกำหนดค่าเฉลี่ยของความยากให้เป็น 0 ($b_i = 0$) และค่าอำนาจจำแนก (a_{km}) ให้เป็น 1 การที่ต้องการน้ำหนักเป็น 1 เพื่อให้ความสามารถมีผลกระทบกับความน่าจะเป็นในการตอบถูกข้อที่ i เท่ากัน ถ้าไม่กำหนดข้อจำกัด สูตรที่ต่างกันของโมเดลแสดงให้เห็นว่าความสามารถของคนมีผลกระทบที่แปรค่าได้ต่อความน่าจะเป็นในการตอบข้อสอบข้อที่ i ถูก ซึ่งสามารถทำให้เหมาะสมยิ่งขึ้น เห็นอย่างชัดเจนว่า indeterminacies ของมาตร สามารถแก้ไขใหม่ภายใต้การแจกแจงของความสามารถ ด้วยการกำหนดให้ค่าเฉลี่ยของความสามารถเริ่มต้นเป็น 0 และความแปรปรวนเป็น 1 ค่าเฉลี่ยของความสามารถที่เพิ่มขึ้นต้องไม่กำหนด ยอมให้มีค่าเพิ่มขึ้นหรือลดลงโดยรวม สูตรของ Thissen และ Steinberg (1986) สำหรับโมเดลราส์ชได้ขยายออกไปโดยให้ค่าอำนาจจำแนกใน MRMLC มีความชันร่วม (d_m) บนมิติ m มากกว่าให้น้ำหนักเป็น 1 ในสมการที่ (3) ให้เป็นโมเดลที่ไม่เป็นมาตรฐาน (unstandardized) เป็นโมเดลที่เกิดจากการรวมความสามารถระหว่างมิติ m ถึง k โมเดลมาตรฐาน (standardized) ของ MRMLC มีสมการดังนี้

$$P(x_{i(k)j} = 1) = \frac{\exp\left(\sum_{m}^{k} d_m \theta_{jm} - b_i\right)}{1 + \exp\left(\sum_{m}^{k} d_m \theta_{jm} - b_i\right)} \quad \text{.....} \quad (25)$$

สมการที่ (25) เป็นสมการที่ความสามารถมีผลกระทบไม่เท่ากันต่อความน่าจะเป็นในการตอบถูก d_m จะแปรค่าระหว่างมิติในสมการ (3) ความสามารถที่มีผลกระทบไม่เท่ากันถูกซ่อนไว้ในความแปรปรวนของความสามารถที่ไม่ได้กำหนดค่า ดังนั้น ความสามารถที่มี

ผลกระทบเพียงเล็กน้อยในโมเดลที่ไม่เป็นมาตรฐาน (unstandardized) จะมีความแปรปรวนต่ำ สามารถแสดงให้เห็นชัดได้ด้วยการกำหนดให้ความสามารถมีการแจกแจงเป็นปกติพหุตัวแปร และความสามารถไม่สัมพันธ์กัน เพราะความแปรปรวนของความสามารถส่วนประกอบ $s^2_{\theta_k^c}$ ที่การวัดครั้งที่ k มีสมการเป็น

$$s^2_{\theta_k^c} = \sum_m^k d_m^2 s^2_{\theta_m} \qquad \text{.....} \quad (26)$$

โมเดลมาตรฐาน (standardized) กับ โมเดลไม่เป็นมาตรฐาน (unstandardized) มีประโยชน์แตกต่างกัน โมเดลไม่เป็นมาตรฐาน (unstandardized) เป็นโมเดลที่เหมาะกับความสามารถที่เพิ่มขึ้นที่มีผลโดยตรงต่อการเพิ่มขึ้นในการตอบถูกที่เกี่ยวกับเงื่อนไขของการจัดกระทำในมิติที่ m ดังนั้น ความสามารถที่เพิ่มขึ้นของโมเดลไม่เป็นมาตรฐาน (unstandardized) ควรจะมีผลกระทบแฝงบนความสามารถโดยตรง ความสามารถเริ่มต้นและความสามารถที่เพิ่มขึ้นในโมเดลมาตรฐาน (standardized) ควรจะใช้ (ถ้าค่าเฉลี่ยถูกลบออก) ในรูปของการอิงกลุ่ม (norm-referenced) ในขณะที่ความแปรปรวนถูกกำหนดให้เป็น 1

ต่อมา Embretson (1995) ได้พัฒนาโมเดล MRMLC ให้เป็น MRMLC+ โดยโมเดลนี้การวัดพัฒนาการจะขึ้นอยู่กับความสามารถของคนและค่าพารามิเตอร์ของข้อสอบที่เปลี่ยนไปในระหว่างการวัด โดยมีสูตรดังนี้

$$P(X_{i(k)j} = 1/\theta_j, \phi, S_i) = \frac{\exp\left(\sum_m^k \theta_{jm} - \sum_n S_{in}\phi_n + f\right)}{1 + \exp\left(\sum_m^k \theta_{jm} - \sum_n S_{in}\phi_n + f\right)}$$

S_{in} แทนค่าของข้อสอบที่ i ที่อยู่บนองค์ประกอบ n

ϕ_n แทนน้ำหนักองค์ประกอบที่ n

f แทนค่าคงที่ปกติ (normalization constant)

θ_{jm} แทนความสามารถที่เพิ่มขึ้นในตัวที่ m ของคนที่ j

MRMLC+ ประมาณค่าพารามิเตอร์ด้วย CML ด้วยวิธี Newton-Raphson

2) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง (method of difference between true ability scores) เสนอโดย May และ Nicewander ในปี 1998 การวัดคะแนนพัฒนาการด้วยวิธีนี้เป็นวิธีที่ง่ายสุด โดยใช้แนวคิดเดียวกับการวัดคะแนนพัฒนาการด้วยวิธีการประมาณค่าจากความแตกต่างระหว่างคะแนนดิบ แต่วิธีที่เสนอนี้มีโมเดลในการประมาณค่าความสามารถที่แท้จริงแบบ logistic 3 พารามิเตอร์ที่ใช้วิธีประมาณ

ค่าพารามิเตอร์ด้วยวิธี EAP (expected a posteriori) ซึ่ง May และ Nicewander เสนอว่าเป็นวิธีที่สามารถแก้ปัญหาความคลาดเคลื่อนในการวัดคะแนนพัฒนาการในกรณีที่มีปฏิสัมพันธ์ระหว่างผู้สอบที่มีระดับความสามารถเริ่มต้นไม่เท่ากันกับข้อสอบที่มีความยากง่ายต่างกัน โดยมีสูตรดังนี้

$L\theta = \theta_2 - \theta_1$

$L\theta$ แทนคะแนนพัฒนาการจากวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง

θ_1 แทนความสามารถที่แท้จริงก่อนเรียน

θ_2 แทนความสามารถที่แท้จริงหลังเรียน

การประมาณค่าพารามิเตอร์ด้วยวิธีการของเบย์ส (Bayes)

การประมาณค่าพารามิเตอร์ด้วยวิธีการของเบย์ส (Bayes) เป็นค่าที่ได้มาจากค่าเฉลี่ยของการแจกแจงภายหลังของค่า θ (posterior distribution of θ) ซึ่งให้รูปแบบการตอบ X_i ซึ่งสามารถประมาณค่าได้ถูกต้องด้วยวิธี gaussion quadrature (Bock and Mislevy, 1982) ฟังก์ชั่นของตัวประมาณค่า EAP (expected a posteriori) มีดังนี้

$$\bar{\theta}_i = \frac{\sum_{k=1}^{q} X_k P(X_i / X_k) A(X_k)}{\sum_{k=1}^{q} P(X_i / X_k) A(X_k)}$$

$A(X_k)$ แทนน้ำหนักที่เป็นบวกของฟังก์ชันการกระจาย X

โดยมี posterior standard deviation (PSD) ดังนี้

$$PSD(\bar{\theta}_i) \cong \frac{\sum_{k=1}^{q} (X_k - \bar{\theta}_i)^2 P(X_i / X_k) A(X_k)}{\sum_{k=1}^{q} P(X_i / X_k) A(X_k)}$$

EAP จะประมาณค่าได้ค่าความคลาดเคลื่อนเฉลี่ยน้อยกว่าตัวประมาณค่าตัวอื่น ๆ เช่น ML sample mean ของ EAP จะเป็นค่าประมาณของค่าเฉลี่ยของประชากรของคุณลักษณะแฝงที่ไม่ลำเอียง sample standard deviation จะมีค่าน้อยกว่า latent population แต่ไม่มีผลกับการประมาณค่าเพราะเมื่อทำคะแนนที่ได้ให้เป็นค่ามาตรฐานแล้ว จะเป็นการปรับค่า sample standard deviation ไปด้วย

สูตรของ gaussian quadrature มีดังนี้

$$\bar{P}(X) \approx \sum_{k=1}^{q} P(X / X_k) A(X_k)$$

ตอนที่ 3 การเปรียบเทียบวิธีการวัดคะแนนพัฒนาการระหว่างวิธีการวัดตามทฤษฎีการทดสอบแบบดั้งเดิมกับทฤษฎีการตอบสนองข้อสอบ

ตารางที่ 2.3 สรุปผลการเปรียบเทียบวิธีการวัดคะแนนพัฒนาการระหว่างวิธีการวัดในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ

	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
กลุ่มวิธีใน CTT 1. ใช้การวัด 2 ครั้ง 1.1 วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ	-	คะแนนที่ได้จากการวัดทั้ง 2 ครั้ง อยู่ในมาตรเดียวกัน มีการแจกแจงเหมือนกัน สามารถที่จะหาความแตกต่างโดยตรงได้ **สูตร** $DS = X_2 - X_1$	1. เป็นตัวประมาณค่าที่ไม่ลำเอียงในการวัดพัฒนาการ 2. เป็นวิธีที่ง่าย (Raykov, 1993)	1. คนที่มีความสามารถเริ่มต้นต่างกัน ถึงแม้จะมีคะแนนพัฒนาการที่เท่ากัน แต่ก็มีความหมายต่างกัน 2. คะแนนพัฒนาการมีความสัมพันธ์ทางลบกับความสามารถเริ่มต้น	1. คะแนนพัฒนาการที่วัดด้วยวิธีนี้กับวิธีของ ลอร์ด มีค่าความคลาดเคลื่อนต่ำกว่าวิธีของคะแนนการเปลี่ยนแปลงส่วนที่เหลือ (วินิจ เทือกทอง, 2537) 2. คะแนนพัฒนาการที่วัดด้วยวิธีนี้กับวิธีของลอร์ด และวิธีวัดการเปลี่ยนแปลงการเรียนรู้โดยขจัดอิทธิพลเพดานทุกวิธีที่เสนอโดยอรุณี มีค่าความเที่ยงและค่าความคลาดเคลื่อนไม่แตกต่างกัน (อรุณี อ่อนสวัสดิ์, 2537)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
1.2 วิธีการวัดคะแนนพัฒนาการของลอร์ด	Lord (1956)	คะแนนพัฒนาการที่แท้จริงได้มาจากการถดถอยพหุคูณของคะแนนดิบก่อนเรียนและหลังเรียน **สูตร** $v_i = y + b_{vx.y}(x_i - \bar{x}) + b_{vy.x}(y_i - \bar{y})$	คะแนนพัฒนาการจะต่างกันแม้ว่าจะมีคะแนนความแตกต่างระหว่างคะแนนดิบก่อนและหลังเรียนเท่ากัน แต่มีคะแนนดิบก่อนเรียนต่างกัน	1. ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ในการประมาณค่าเพื่อให้สัมประสิทธิ์การถดถอยมีความคลาดเคลื่อนต่ำ 2. พิสัยของคะแนนพัฒนาการวิธีนี้แคบกว่าพิสัยของคะแนนความแตกต่างระหว่างคะแนนดิบก่อนเรียนและหลังเรียน 3 เป็นวิธีที่วัดยากกว่าวิธีวัดคะแนนความแตกต่างระหว่างคะแนนดิบ	1. คะแนนพัฒนาการที่วัดด้วยวิธีนี้มีค่าความคลาดเคลื่อนไม่แตกต่างจากวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ (วินิจ เทือกทอง, 2537) 2. คะแนนพัฒนาการที่วัดด้วยวิธีนี้และวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบและวิธีวัดการเปลี่ยนแปลงการเรียนรู้โดยขจัดอิทธิพลเพดานทุกวิธีที่เสนอโดย อรุณี มีค่าความเที่ยงและค่าความคลาดเคลื่อนไม่แตกต่างกัน (อรุณี อ่อนสวัสดิ์, 2537)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
1.3 วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือ	Du Bois (1957)	คะแนนพัฒนาการควรที่จะเป็นอิสระจากคะแนนก่อนเรียนโดยคะแนนพัฒนาการประมาณค่าจากความแตกต่างระหว่างคะแนนการวัดหลังเรียนกับคะแนนผลการทำนายหลังเรียนด้วยคะแนนก่อนเรียน **สูตร** $R_i = Y_i - [\ \bar{Y} + B_{YX} (X_i - \bar{X})\ ]$	1. คะแนนพัฒนาการเป็นอิสระจากคะแนนก่อนเรียน 2. สามารถกำจัดความสัมพันธ์ลวงระหว่างคะแนนพัฒนาการกับคะแนนก่อนเรียน	1. ค่าความคลาดเคลื่อนของสัมประสิทธิ์การถดถอยขึ้นอยู่กับขนาดของกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ 2. มีอิทธิพลเพดานเกิดขึ้น 3. วิธีการวัดยากกว่าวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ	คะแนนพัฒนาการที่วัดด้วยวิธีนี้มีความคลาดเคลื่อนสูงกว่าวิธีของลอร์ด และวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ (วินิจ เทือกทอง, 2537)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
1.4 วิธีการวัดคะแนนพัฒนาการที่เป็นอิสระจากคะแนนก่อนเรียน	Tucker, Damarin and Messick (1966)	คะแนนพัฒนาการเป็นอิสระจากคะแนนก่อนเรียน โดยคะแนนพัฒนาการประมาณค่าจากความแตกต่างระหว่างคะแนนหลังเรียนกับคะแนนผลการทำนายหลังเรียนด้วยคะแนนจริงก่อนเรียน **สูตร** $B_i = Y_i - [\ \bar{Y} + B_{YX} / R_{XX'} (X_i - \bar{X})\ ]$	สามารถกำจัดความสัมพันธ์ลวงระหว่างคะแนนพัฒนาการกับคะแนนก่อนเรียน	1. คะแนนพัฒนาการขึ้นอยู่กับค่าความเที่ยงของเครื่องมือในการวัดก่อนเรียน ถ้าความเที่ยงเป็น1 จะได้คะแนนพัฒนาการเหมือนวิธีการวัดคะแนนการเปลี่ยนแปลงส่วนที่เหลือ 2. ค่าความคลาดเคลื่อนของสัมประสิทธิ์การถดถอยขึ้นอยู่กับขนาดของกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ 3 วิธีการวัดยากกว่าวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ	-

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
1.5 วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน	Labouvie (1982)	คะแนนที่วัดได้ในครั้งแรกและครั้งหลังไม่อยู่ในมาตรการเดียวกัน การแจกแจงไม่เหมือนกัน จึงแปลงเป็นคะแนนมาตรฐาน **สูตร** $SC = Z_2 - Z_1$	1. คะแนนพัฒนาการสามารถเปรียบเทียบระหว่างวิชากันได้ 2. คะแนนพัฒนาการสามารถที่จะบวกลบกันได้		-
1.6 วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ	Tornqvist, Vartia and Vartia (1985)	คะแนนที่วัดได้ในครั้งแรกและครั้งหลังไม่อยู่ในมาตรเดียวกัน การแจกแจงไม่เหมือนกัน และไม่เป็นโค้งปกติ และมีแนวโน้มว่าจะเป็นสัดส่วนกัน จึงแปลงเป็นค่าลอการิทึมธรรมชาติ **สูตร** $NL = \ln X_1 - \ln X_2$	คะแนนพัฒนาการที่ได้จะมีการแจกแจงปกติ มีคุณสมบัติของการบวก		-

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
1.7 วิธีการวัดคะแนนพัฒนาการสัมพัทธ์	ศิริชัย กาญจนวาสี (2532)	อัตราความงอกงามขึ้นอยู่กับการเปลี่ยนแปลงสัมบูรณ์และศักยภาพที่จะสามารถพัฒนาได้ของผู้สอบ ซึ่งจะทำให้แต่ละคนมีอัตราความงอกงามแตกต่างกันตามศักยภาพของตนที่มีอยู่ และสามารถช่วยให้มีการขจัดอิทธิพลเพดาน **สูตร** $SR = \left(\frac{y - x}{F - x}\right)100$	1. มีการขจัดอิทธิพลเพดาน 2. สามารถที่จะเปรียบเทียบอัตราความงอกงามของคนที่มีการเปลี่ยนแปลงสัมบูรณ์เท่ากัน (y – x)	มีค่าความคลาดเคลื่อนสูงกว่าวิธีการวัดพัฒนาการที่หาจากวิธีเศษเหลือเทียบกับศักยภาพของผู้สอบ (วินิจ เทือกทอง, 2537)	วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบมีค่าความคลาดเคลื่อนต่ำกว่าวิธีประมาณพัฒนาการสัมพัทธ์และคะแนนเพิ่มวิธีของลอร์ดที่เทียบกับศักยภาพของผู้สอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
1.8 วิธีการวัดคะแนนพัฒนาการโดยขจัดอิทธิพลเพดาน	อรุณี อ่อนสวัสดิ์ (2537)	คะแนนพัฒนาการเป็นฟังก์ชั่นของพื้นความรู้เดิมการเรียนการสอนและอิทธิพลเพดาน	สามารถขจัดอิทธิพลเพดาน	1. คะแนนพัฒนาการทุกวิธีที่มีการขจัดอิทธิพลเพดานมีค่าความเที่ยงความคลาดเคลื่อนในการวัดคะแนนพัฒนาการไม่แตกต่างจากวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบและวิธีการวัดคะแนนการเปลี่ยนแปลงของลอร์ด 2. วิธีวัดยากกว่าวิธีอื่น ๆ ที่กล่าวมาแล้ว	คะแนนพัฒนาการทุกวิธีที่มีการขจัดอิทธิพลเพดานมีค่าความเที่ยงความคลาดเคลื่อนในการวัดคะแนนพัฒนาการไม่แตกต่างจากวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบและวิธีการวัดคะแนนพัฒนาการของลอร์ด (อรุณี อ่อนสวัสดิ์, 2537)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
1.9 วิธีการวัดคะแนนพัฒนาการโดยเทียบส่วนร้อยของคะแนนพัฒนาการ คะแนนพัฒนาการที่ใช้มี 3 วิธี 1) วิธีดั้งเดิม 2) วิธีของลอร์ด 3) วิธีเศษเหลือ	วินิจ เทือกทอง (2537)	คะแนนพัฒนาการที่มีการเทียบเป็นร้อยละของพื้นความรู้เดิมของผู้สอบ (คะแนนสอบครั้งแรกของผู้สอบ) จะทำให้คะแนนพัฒนาการของแต่ละคนแตกต่างกันไปตามพื้นความรู้เดิมของผู้สอบ	1. มีการขจัดอิทธิพลเพดาน 2. สามารถที่จะเปรียบเทียบอัตราความงอกงามของผู้ที่มีคะแนนเพิ่มเท่ากัน	มีค่าความคลาดเคลื่อนสูงกว่าวิธีหาคะแนนพัฒนาการที่ไม่มีการขจัดอิทธิพลเพดานได้แก่วิธีดั้งเดิม วิธีของลอร์ดและวิธีเศษเหลือ	มีค่าความคลาดเคลื่อนสูงกว่าวิธีหาคะแนนพัฒนาการที่ไม่มีการขจัดอิทธิพลเพดานได้แก่วิธีดั้งเดิม วิธีของลอร์ดและวิธีเศษเหลือ และเมื่อเทียบกับวิธีที่มีการขจัดอิทธิพลเพดานด้วยกันจะมีค่าความคลาดเคลื่อนสูงกว่าวิธีหาคะแนนพัฒนาการเทียบกับศักยภาพของผู้สอบเมื่อคะแนนพัฒนาการได้มาจากวิธีเศษเหลือ
1.10 วิธีการวัดคะแนนพัฒนาการโดยเทียบคะแนนพัฒนาการกับศักยภาพของผู้สอบ คะแนนพัฒนาการที่ใช้มี 3 วิธี เหมือนข้อ 1.9	วินิจ เทือกทอง (2537)	คะแนนพัฒนาการที่มีการเทียบกับศักยภาพของผู้สอบจะทำให้คะแนนพัฒนาการของแต่ละคนแตกต่างกันไปตามศักยภาพ	1. มีการขจัดอิทธิพลเพดาน 2. สามารถที่จะเปรียบเทียบอัตราความงอกงามของผู้ที่มีคะแนนเพิ่มเท่ากัน	มีค่าความคลาดเคลื่อนสูงกว่าวิธีหาคะแนนพัฒนาการที่ไม่มีการขจัดอิทธิพลเพดานได้แก่วิธีดั้งเดิม วิธีของลอร์ดและวิธีเศษเหลือ	มีค่าความคลาดเคลื่อนสูงกว่าวิธีหาคะแนนพัฒนาการที่ไม่มีการขจัดอิทธิพลเพดานได้แก่วิธีดั้งเดิม วิธีของลอร์ดและวิธีเศษเหลือ และเมื่อเทียบกับวิธีที่มีการขจัดอิทธิพลเพดานด้วยกันมีค่าความคลาดเคลื่อนต่ำกว่าวิธีอื่นๆ เมื่อคะแนนพัฒนาการได้มาจากการวัดด้วยวิธีเศษเหลือ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
2. ใช้การวัดมากกว่า 2 ครั้ง 2.1 โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง	McArdle and Hamagami (1991)	อัตราพัฒนาการเป็นตัวแปรแฝงที่มีค่าต่างกันระหว่างบุคคลหาจากผลรวมของผลคูณระหว่างคะแนนน้ำหนักองค์ประกอบของการวัด กับคะแนนผลการสอบแต่ละครั้ง	1. สามารถหาแบบแผนของพัฒนาการได้ 2. มีการนำค่าความคลาดเคลื่อนนำมาคิด	1. ต้องมีการวัดหลายครั้ง 2. ต้องมีความรู้ในเรื่องโมเดลสมการโครงสร้าง	เป็นโมเดลที่ได้รับการยอมรับว่าเป็นโมเดลที่ดีที่สุดในการศึกษาพัฒนาการที่ใช้วิธีวิเคราะห์องค์ประกอบในโมเดลสมการโครงสร้าง (Stoolmiler, 1994)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
กลุ่มวิธี IRT 1. คะแนนพัฒนาการเกิดจากการเปลี่ยนแปลงของค่าพารามิเตอร์ของข้อสอบระหว่างเวลา 1.1 วิธีการวัดพัฒนาการความสามารถที่แท้จริง โดยใช้โมเดล LLTM	Fischer (1973)	ผู้สอบถูกทดสอบด้วยแบบสอบหรือแบบสอบถามหลาย ๆ ครั้ง หรือหลาย ๆ สถานการณ์ ในช่วงเวลาระหว่างการวัดผู้สอบได้รับ treatments ดังนั้นพัฒนาการที่เกิดขึ้นระหว่างการวัดสามารถที่จะทำให้อยู่ในรูปของการเปลี่ยนแปลงของค่าพารามิเตอร์ของข้อสอบ	1. ใช้กับการวัดตั้งแต่ 2 ครั้งขึ้นไป / ข้อสอบชุดเดิมหรือต่างชุดกัน 2. ใช้ได้กับแบบวัดที่เป็น dichotomous 3. ใช้การวิเคราะห์ข้อมูลที่ได้จากการวัดหลายครั้งพร้อมกัน	ไม่สามารถหาคะแนนพัฒนาการเป็นรายบุคคลได้	ยังไม่เคยมีการวิจัยเปรียบเทียบถึงประสิทธิภาพของผลการวัดเมื่อเทียบกับวิธีการวัดวิธีอื่น ๆ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
1.2 LLRA	Fischer (1974)	1. อิทธิพลของการจัดกระทำที่เกิดขึ้นระหว่างเวลาเกิดในรูปของค่าพารามิเตอร์ของข้อสอบ ซึ่งส่งผลต่อคะแนนพัฒนาการ 2. ใช้กับการวัด 2 ครั้ง ข้อมูลเป็น dichotomous	สามารถยืดหยุ่นในข้อตกลงเบื้องต้นเกี่ยวกับข้อสอบที่ต้องเป็นแบบเอกพันธ์ของโมเดล LLTM	1. ไม่สามารถวัดความแตกต่างระหว่างบุคคลในการเรียนรู้เพราะอิทธิพลแนวโน้มและตัวแปรจัดกระทำถูกกำหนดให้มีค่าเท่ากันทุกคน 2. ใช้ข้อสอบชุดเดิมในการสอบทั้ง 2 ครั้ง	-
1.3 LRSM	Fischer and Parzer (1991)	1. อิทธิพลของการจัดกระทำที่เกิดขึ้นระหว่างเวลาเกิดในรูปของค่าพารามิเตอร์ของข้อสอบ ซึ่งส่งผลต่อคะแนนพัฒนาการ 2. ใช้กับการวัดตั้งแต่ 2 ครั้งขึ้นไป 3. ใช้กับข้อมูลที่เป็น rating scale เป็นโมเดลเอกมิติ	1. ใช้ข้อสอบคนละชุดในการสอบแต่ละครั้งเพื่อป้องกันการจำได้ โดยใช้ข้อสอบร่วมจำนวนหนึ่ง 2. สามารถทดสอบพัฒนาการที่เพิ่มขึ้นระหว่างกลุ่ม	ไม่สามารถวัดความแตกต่างระหว่างบุคคลในการเรียนรู้เพราะอิทธิพลแนวโน้มและตัวแปรจัดกระทำถูกกำหนดให้มีค่าเท่ากันทุกคน	-

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
1.4 LPCM	Fischer and Ponocny (1994)	1. อิทธิพลของการจัดกระทำที่เกิดขึ้นระหว่างเวลาเกิดในรูปของค่าพารามิเตอร์ของข้อสอบ ซึ่งส่งผลต่อคะแนนพัฒนาการ 2. ใช้กับการวัดตั้งแต่ 2 ครั้งขึ้นไป 3. ใช้กับข้อมูลที่เป็น partial credit เป็นโมเดลเอกมิติ	1. ใช้ข้อสอบคนละชุดในการสอบแต่ละครั้งเพื่อป้องกันการจำได้ โดยใช้ข้อสอบร่วมจำนวนหนึ่ง 2. สามารถทดสอบพัฒนาการที่เพิ่มขึ้นระหว่างกลุ่ม	ไม่สามารถวัดความแตกต่างระหว่างบุคคลในการเรียนรู้ เพราะอิทธิพลแนวโน้มและตัวแปรจัดกระทำถูกกำหนดให้มีค่าเท่ากันทุกคน	-

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
2. คะแนนพัฒนาการเกิดจากการเปลี่ยนแปลงของค่าพารามิเตอร์ของคน ระหว่างเวลา 2.1 MRMLC	Embretson (1991, 1995)	1. อิทธิพลของการจัดกระทำที่เกิดขึ้นระหว่างเวลาเกิดในรูปของค่าพารามิเตอร์ของข้อสอบ ซึ่งส่งผลต่อคะแนนพัฒนาการ 2. ใช้กับการวัดตั้งแต่ 2 ครั้งขึ้นไป 3. ใช้กับข้อมูลที่เป็น dichotomousโมเดลพหุมิติ	1. ใช้ข้อสอบคนละชุดในการสอบแต่ละครั้งเพื่อป้องกันการจำได้ โดยใช้ข้อสอบร่วมจำนวนหนึ่ง 2. สามารถหาคะแนนพัฒนาการระหว่างบุคคล	-	-

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

วิธีวัด	ผู้พัฒนา	แนวคิด	จุดเด่น	จุดอ่อน	ผลการวิจัยที่ทำไปแล้ว
2.2 วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง	May และ Nicewander (1998)	พัฒนาการที่เกิดขึ้นเป็นพัฒนาการของตัวแปรแฝงหรือความสามารถที่แท้จริง ซึ่งสามารถที่จะวัดได้ด้วยวิธีของ IRT โดยใช้การวัดเพียง 2 ครั้ง **สูตร** $L\theta = \theta_2 - \theta_1$	1. เป็นวิธีที่ง่าย 2. มีการนำค่าพารามิเตอร์ของข้อสอบมาร่วมในการประมาณค่าความสามารถที่แท้จริง 3. สามารถที่จะเปรียบเทียบพัฒนาการของความสามารถที่แท้จริงระหว่างคน และระหว่างวิชาที่ต่างกันได้ 4. ขจัดอิทธิพลของปฏิสัมพันธ์ระหว่างความยากของข้อสอบกับความสามารถเริ่มต้นของผู้สอบที่มีต่อคะแนนการเปลี่ยนแปลง	1. ต้องใช้กลุ่มตัวอย่างในการวิเคราะห์เป็นจำนวนมาก 2. ไม่สามารถขจัดอิทธิพลเพดาน	การวัดความแตกต่างระหว่างความสามารถที่แท้จริง สามารถขจัดอิทธิพลของปฏิสัมพันธ์ระหว่างความยากของข้อสอบกับความสามารถเริ่มต้นของผู้สอบที่มีต่อคะแนนพัฒนาการเมื่อใช้การวัดพัฒนาการด้วยการหาความแตกต่างของคะแนนดิบ (May and Nicewander (1998)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

4.1 งานวิจัยในต่างประเทศ

1) Embretson (1991b) ได้เสนอโมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (multidimensional Rasch model for learning and change: MRMLC) เพื่อใช้วัดคะแนนพัฒนาการ โดยพัฒนาจากโมเดลราส์ชที่เป็นพหุมิติด้วยการเพิ่ม Wiener simplex structure เพื่อหาคะแนนพัฒนาการรายบุคคลในรูปของความสามารถของคนที่เพิ่มขึ้นระหว่างเวลา โดยที่ค่าพารามิเตอร์ของข้อสอบมีค่าคงที่ MRMLC ไม่ต้องใช้แบบสอบฉบับเดิมเพื่อต้องการหลีกเลี่ยงอิทธิพลที่เกิดจากการจำได้ ซึ่งจะทำให้เกิดการตอบอย่างลำเอียง ซึ่งส่งผลให้เกิดการละเมิดข้อตกลงเบื้องต้นในเรื่องของการเป็นอิสระกันระหว่างข้อ (local independence) Embretson ได้ทดลองใช้กับข้อมูลจริงที่ใช้แบบสอบวัดความสามารถในการเรียนรู้มิติสัมพันธ์ (SLAT) โดยใช้ข้อสอบ 76 ข้อ Pretest 24 ข้อ Posttest 2 ครั้ง ๆ ละ 24 ข้อ และ 4 ข้อที่เพิ่มใน Pretest เพื่อใช้เชื่อมโยงในการประมาณค่าระหว่างกลุ่ม Embretson ได้หาสหสัมพันธ์ระหว่าง MRMLC กับคะแนนมาตรฐานที่ได้จากการวัดการเปลี่ยนแปลงแบบดั้งเดิมในการสอบ Pretest คะแนนที่เพิ่มขึ้นครั้งที่ 1, 2 พบว่ามีความสัมพันธ์กันสูงทุกค่า (0.994, 0.953 และ 0.949) ในขณะที่ถ้าใช้โมเดลของ Anderson หาค่าสหสัมพันธ์กับคะแนนมาตรฐาน จะได้ค่าสหสัมพันธ์ของทั้ง 2 วิธีในการสอบ Pretest คะแนนที่เพิ่มขึ้นครั้งที่ 1, 2 มีค่าต่ำ (0.994, 0.598, 0.389)

2) Fischer และ Parzer (1991) ได้เสนอวิธีการใช้โมเดลการให้คะแนนเป็นระดับคะแนนที่มีช่วงคะแนนเท่ากันทุกช่วงเชิงเส้นตรง (linear rating scale model: LRSM) ในการศึกษาการเปลี่ยนแปลง Fischer ใช้ conditional maximum likelihood (CML) ในการประมาณค่าพารามิเตอร์ และใช้ likelihood-ratio ทดสอบ H_O: ไม่มีการเปลี่ยนแปลง Fischer และ Parzer ได้ทดสอบแนวคิดนี้ด้วยการใช้ข้อมูลจริงจากคนไข้ที่มีความผิดปกติทางจิต ซึ่งเป็นคนไข้ภายนอกของคลินิกของมหาวิทยาลัยเวียนนา โดยแบ่งคนไข้ออกเป็น 2 กลุ่มกลุ่มทดลองเป็นกลุ่มที่ให้ยาในการรักษาและมีการเข้าฝึกอบรมการผ่อนคลายด้วยการพบปะสนทนาในกลุ่มเป็นเวลา 7 วัน จำนวน 37 คน กลุ่มควบคุมเป็นกลุ่มให้ยาอย่างเดียว จำนวน 50 คน โดยการให้คนไข้ประเมินตนเองในแบบวัดอาการในด้านต่าง ๆ ของคนไข้ ซึ่งเป็นแบบวัดมาตรประมาณค่า 4 ช่วง ในการประเมินครั้งแรกเมื่อคนไข้มารับการรักษา และครั้งที่ 2 เมื่อเวลาผ่านไป 1 ปี ผลการวิจัยพบว่ากลุ่มทดลองไม่มีการเปลี่ยนแปลงเกิดขึ้นจากการทดสอบทั้ง 2 ครั้ง และมีปฏิสัมพันธ์ระหว่างลักษณะของคนไข้กับการจัดกระทำ (treatment) ที่ให้

จากผลการทดลองใช้กับข้อมูลจริง Fischer และ Parzer ได้เสนอว่าข้อมูลจริงที่ใช้นี้ไม่เป็นไปตามข้อตกลงที่ว่าจะต้องเป็นแบบวัดเอกมิติเท่านั้น แต่ข้อมูลที่ใช้จริงเป็นแบบวัดพหุมิติประเด็นที่ 2

วิธีนี้ใช้กับข้อมูลที่มีเพียง 1 ข้อ หรือ 1 ลักษณะอาการที่ตอบในการวัดหลายครั้ง แต่ข้อมูลที่ใช้จริงมีข้อคำถามหลายลักษณะอาการ ดังนั้นจึงยากที่จะทำการทดสอบสมมติฐานได้

3) Fischer และ Ponocny (1994) ได้เสนอวิธีการใช้โมเดลการให้คะแนนความรู้บางส่วนเชิงเส้นตรง (linear partial credit model: LPCM) ที่อยู่ในตระกูลของการให้คะแนนพหุวิภาคของโมเดลราส์ชเอกมิติมาวัดคะแนนพัฒนาการแทนการใช้โมเดลมาตรประเมินค่าเชิงเส้น (linear rating scale model: LRSM) ที่พบว่าเมื่อใช้กับข้อมูลจริงในการวัดคะแนนการเปลี่ยนแปลงของอาการคนไข้ทางจิตมีข้อจำกัดในเรื่องค่าพารามิเตอร์ของข้อกระทง (β_i) พารามิเตอร์ของช่วงคะแนน (ω_h) และพารามิเตอร์ของคน (θ_u) ที่จะทำให้ข้อมูลที่ได้เหมาะกับโมเดลในการวัดพัฒนาการ (Fischer & Parzer, 1991) PCM มีความยืดหยุ่นกว่า LRSM และ LRSM เป็นเพียงกรณีหนึ่งของ LPCM แต่เนื่องจากการวิเคราะห์ LPCM มีค่าพารามิเตอร์ในการประมาณค่ามากกว่า LRSM ดังนั้น Fischer และ Ponocny จึงได้พัฒนา CML ที่ใช้ใน LRSM ให้มีประสิทธิภาพมากขึ้น โดยใช้ Broyden-Fletcher Goldfarb Shanno (quasi-Newton) ในการประมาณค่าพารามิเตอร์ แทนการประมาณค่าด้วย Newton Raphson algorithm ที่ใช้ใน LRSM

ในการทดลองโมเดลกับข้อมูลจริง Fischer และ Ponocny ได้ใช้กลุ่มตัวอย่างคนไข้ที่ผิดปกติทางจิตชุดเดิมกับที่วิเคราะห์ด้วย LRSM ในการทดลองของ LPCM นี้ ได้ทดลองกับคนไข้กลุ่มทดลอง 37 คน ที่ให้ยาและเข้ารับการฝึกอบรมการผ่อนคลายความเครียดด้วยการมีส่วนร่วมในกลุ่มและกลุ่มควบคุม 50 คน ที่รักษาด้วยยาอย่างเดียวทั้ง 2 กลุ่ม ให้ประเมินอาการของตนเองด้วยข้อคำถาม 24 ข้อ ที่เป็นมาตรประมาณค่า 4 ช่วงในช่วงแรกที่เข้ารับการรักษา และหลังจากนั้น 1 ปี ทั้ง 2 กลุ่มประเมินตนเองใหม่ด้วยข้อคำถาม 24 ข้อเดิมที่ปรับใหม่โดยให้คะแนนเป็นส่วนๆ ไม่เท่ากันตามความหมายของ LPCM ผลการวิจัยพบว่า กลุ่มทดลองไม่มีการเปลี่ยนแปลงในอาการป่วย แสดงว่าการเข้าฝึกอบรมไม่ได้มีผลให้อาการป่วยทางจิตบรรเทาลง ซึ่งสอดคล้องกับผลการวิจัยจาก LRSM และยังให้ผลการวิจัยอีกว่าการใช้ยามีผลทำให้เกิดการเปลี่ยนแปลง

จากงานวิจัยชิ้นนี้ผู้วิจัยได้นำเสนอโปรแกรมสำเร็จรูปที่ใช้ชื่อว่า LPCM ซึ่งใช้สำหรับการวิจัยที่วัดซ้ำที่มีประสิทธิภาพ มีความถูกต้องในการประมาณค่าอิทธิพลของตัวแปรมากกว่าการประมาณค่าพารามิเตอร์ของข้อ มีประโยชน์และใช้ง่ายสำหรับการวิเคราะห์ข้อมูลที่เป็นการตอบแบบพหุวิภาค (polytomous respones) และสามารถทดสอบสมมติฐานการเปลี่ยนแปลง (testing hypotheses on change)

4) Embretson (1995) ได้พัฒนาโมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (MRMLC) โดยการนำกระบวนการและโครงสร้างของความรู้ใส่เข้าไปในโมเดลราส์ช โดยตั้งชื่อโมเดลใหม่นี้ว่า MRMLC+ ซึ่งโมเดลนี้สามารถหาพัฒนาการรายบุคคล โดยพิจารณาจากทั้งความสามารถที่เพิ่มขึ้น

ระหว่างเวลา และจากพัฒนาการในกระบวนการเรียนและความรู้ที่เพิ่มขึ้น ซึ่งอยู่ในรูปของพารามิเตอร์ของข้อสอบ โดย Embretson นำโมเดลนี้ไปใช้กับพนักงานใหม่ของ Air Force จำนวน 578 คน ที่ได้รับการอบรมขั้นพื้นฐาน 3 สัปดาห์ โดยใช้แบบทดสอบการแก้ปัญหาทางคณิตศาสตร์ ทำการทดสอบ 3 ครั้ง ด้วยแบบสอบที่ไม่ซ้ำกัน คือสอบก่อนสอน สอบระหว่างสอน และสอบหลังสอน โดยแบบทดสอบก่อนสอนและแบบทดสอบหลังสอน แต่ละฉบับมี 10 ข้อ ประกอบด้วยปัญหาเรื่องอัตราส่วนจำนวน 7 ข้อ และ 3 ข้อในเรื่องของอัตราส่วนที่ขยายตัวแปรและจำนวนสมการมากขึ้น และแบบทดสอบระหว่างการสอนจะมีเรื่องอัตราส่วน 7 ข้อ และมี 3 ข้อที่เป็นข้อสอบร่วม จัดการสอบแบบจตุรัสละติน (latin square) ผลการวิจัยพบว่าจำนวนคำในโจทย์ปัญหา ระดับความสามารถในการอ่าน จำนวนตัวแปร มีค่าสหสัมพันธ์สูงกับค่าความยากที่ได้จากโมเดลราส์ช พารามิเตอร์ของ MRMLC+ ประมาณค่าด้วย CML โดยใช้โปรแกรม LINLOG ความสามารถเริ่มต้นและความสามารถที่เพิ่มขึ้น (Initial competency level and modifiability) ของแต่ละคนถูกประมาณค่าด้วยวิธี Newton-Raphson การแปลความหมายของระดับความสามารถเริ่มต้นและความสามารถที่เพิ่มขึ้นได้จากการนำระดับความสามารถของแต่ละคนมาใส่ในมาตรของค่าความยากของข้อสอบที่แยกให้อยู่ในรูปของชนิดของจำนวนคำในโจทย์ปัญหา ระดับความสามารถในการอ่านจำนวนตัวแปร ซึ่งผลจากการวิจัยพบว่า MRMLC+ มีประโยชน์ในทำความเข้าใจลักษณะและระดับของความสามารถ พารามิเตอร์ของคนสามารถประยุกต์ใช้ได้กับข้อสอบที่มีความซับซ้อนในด้านของการวัดความรู้ที่มีอยู่

5) May and Nicewander (1998) ได้ศึกษาวิธีการวัดคะแนนความแตกต่าง (difference score) ที่จะช่วยป้องกันหรือลดขนาดของอิทธิพลที่เกิดจากความยากของแบบสอบในการสอบครั้งแรกที่มีความลำเอียงต่อคะแนนพัฒนาการของผู้สอบที่มีความสามารถแตกต่างกัน May and Nicewander ได้ทำการศึกษาด้วยวิธีการจำลองข้อมูล (Simulation) แบ่งคนออกเป็น 3 กลุ่ม ตามระดับความสามารถจริง (θ) คือกลุ่มเก่ง ปานกลาง อ่อน และสุ่มข้อสอบมา 4 ประเภทคือ ข้อสอบที่มีความยากใกล้เคียงกัน (Uniform) ข้อสอบที่มีระดับความยากกระจายเป็นโค้งปกติ (Symmetric) ข้อสอบที่ง่ายและข้อสอบที่ยาก สุ่มข้อสอบประเภทละ 50 ข้อ และใช้โมเดลของเปลี่ยนแปลงแฝง (Latent change) 3 โมเดล คือโมเดลที่ 1 พัฒนาการคงที่ ($\theta_2 = \theta_1 + 1$) โมเดลที่ 2 พัฒนาการที่เพิ่มขึ้น ($\theta_2 = 1.2\ \theta_1 + 1$) และโมเดลที่ 3 พัฒนาการที่ลดลง ($\theta_2 = 0.8\ \theta_1 + 1$) และใช้วิธีวัดหาคะแนนพัฒนาการ 2 แบบ คือ คะแนนพัฒนาการที่ได้จากความแตกต่างของคะแนนดิบในการวัดครั้งที่ 2 และคะแนนดิบในการวัดครั้งแรก และคะแนนพัฒนาการที่ได้จากผลต่างของการประมาณค่าด้วย IRT 3 พารามิเตอร์ ในการวัดครั้งที่ 2 และครั้งที่ 1 ($\theta_2 - \theta_1$) โดยใช้วิธีประมาณค่าแบบ EAP(θ) (expected a posteriori estimates) ผลการวิจัยพบว่าคะแนนพัฒนาการที่มาจากผลต่างของคะแนนดิบตอบถูกในการวัดครั้งที่ 1 และ 2 นั้นในทุกประเภทของแบบสอบ และทุก

กลุ่มความสามารถของผู้สอบ และคะแนนพัฒนาการทุกคะแนนบิดเบือนจากคะแนนพัฒนาการที่แท้จริง โดยเฉพาะอย่างยิ่งในกลุ่มคนเก่ง จะได้คะแนนพัฒนาการที่ต่ำกว่าความเป็นจริงมากกว่ากลุ่มอ่อนและกลุ่มปานกลาง สำหรับกลุ่มอ่อนนั้นจะได้คะแนนพัฒนาการต่ำกว่าความเป็นจริงมากในกรณีข้อสอบยาก สำหรับคะแนนพัฒนาการที่ได้จากคะแนนความแตกต่างที่วิเคราะห์ด้วย IRT แบบ EAP (θ) พบว่า คะแนนพัฒนาการที่ได้มีค่าใกล้เคียงกับคะแนนพัฒนาการที่แท้จริงเกือบทุกกรณี ยกเว้นในกรณีคนเก่งที่ใช้แบบสอบง่าย และคนอ่อนที่ใช้แบบสอบยาก ที่คะแนนความแตกต่างที่ได้จะต่ำกว่าความเป็นจริงเล็กน้อย ซึ่งมีค่าน้อยเมื่อเทียบกับคะแนนพัฒนาการที่ได้จากผลต่างของคะแนนดิบ สำหรับแบบสอบที่ใช้แบบสอบปรับเหมาะ คะแนนพัฒนาการที่ได้จากวิธี EAP(θ) จะมีค่าใกล้เคียงค่าคะแนนพัฒนาการที่แท้จริงทุกกรณี

4.2 งานวิจัยในประเทศไทย

1) อรุณี อ่อนสวัสดิ์ (2537) ได้พัฒนาวิธีการวัดคะแนนพัฒนาการในสถานการณ์ที่มีการสอบก่อนเรียนและหลังเรียน และเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการที่พัฒนากับวิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนจริง วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างความสามารถ วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ และวิธีวัดคะแนนพัฒนาการการเรียนรู้ของลอร์ด ในการพัฒนาได้อาศัยทฤษฎีการเรียนรู้ของบลูมและคำนึงถึงอิทธิพลเพดาน 4 แนวคิด แต่ละแนวคิดมีทางปฏิบัติเป็น 2 แนวทาง แนวทางหนึ่ง วัดด้วยค่าประมาณคะแนนจริง ส่วนอีกแนวทางหนึ่งวัดด้วยคะแนนดิบผู้วิจัยจึงเสนอวิธีวัดพัฒนาการการเรียนรู้ได้ 8 วิธี ใช้เทคนิคมอนติคาร์โลในการจำลองปัญหาการวิจัยด้วยเครื่องคอมพิวเตอร์ ในสภาพการเรียนรู้ 3 ลักษณะ คือ แบบรอบรู้ แบบทั่วไป และแบบกึ่งรอบรู้ ความสัมพันธ์ระหว่างความสามารถก่อนเรียนและหลังเรียน 3 ระดับ คือ 0.4, 0.6 และ 0.8 กลุ่มตัวอย่าง 3 ขนาด คือ 30, 50 และ 100 คน และแบบสอบมีความยาก 3 ระดับ คือ 30, 60 และ 120 ข้อ

ผลการวิจัยพบว่าทุกวิธีที่อรุณีพัฒนาขึ้นมีความเที่ยงสูงกว่าวิธีการวัดคะแนนพัฒนาการโดยวิธีการประมาณค่าจากความแตกต่างระหว่างคะแนนดิบก่อนและหลังเรียน และวิธีการวัดการเปลี่ยนแปลงของลอร์ด และมีข้อจำกัดในด้านผลการวัดมีประสิทธิภาพดีในบางสถานการณ์ คือ สถานการณ์ที่มีความสัมพันธ์ระหว่างความสามารถก่อนเรียนและหลังเรียนสูง กลุ่มตัวอย่างมีขนาดปานกลาง และแบบสอบมีน้อยข้อ

2) วินิจ เทือกทอง (2537) ได้ศึกษาเปรียบเทียบประสิทธิภาพของวิธีการประมาณค่าคะแนนพัฒนาการ 5 วิธี คือวิธีดั้งเดิม วิธีเศษเหลือ วิธีของลอร์ด วิธีของศิริชัย ซึ่งในการวิจัยครั้งนี้หมายถึงคะแนนพัฒนาการที่ได้จากวิธีดั้งเดิม วิธีเศษเหลือ และวิธีของลอร์ด หารด้วยคะแนนความแตกต่างระหว่าง

คะแนนเต็มและคะแนนสอบครั้งแรกคูณด้วย 100 และวิธีเทียบส่วนร้อย ซึ่งในการวิจัยครั้งนี้หมายถึง คะแนนพัฒนาการที่ได้จากวิธีดั้งเดิม วิธีเศษเหลือ และวิธีของลอร์ด หารด้วยคะแนนสอบครั้งแรกคูณด้วย 100 โดยมีรูปแบบการแจกแจงที่แตกต่างกันตามสถานการณ์ดังต่อไปนี้คือ แยกตามลักษณะการแจกแจงของคะแนนการวัดครั้งแรกและคะแนนการวัดครั้งหลังที่มีลักษณะการกระจาย 4 แบบคือ เบ้ขวา-เบ้ซ้าย เบ้ขวา-ปกติ ปกติ-เบ้ซ้าย และปกติ-ปกติ ค่าคะแนนพัฒนาการมีค่าเฉลี่ยเพิ่มจากคะแนนการวัดครั้งแรกเป็น 3 แบบ คือ 0.8σ, 0.5σ และ 0.2σ ส่วนค่าสหสัมพันธ์ระหว่างคะแนนการวัดครั้งแรกกับคะแนนพัฒนาการมีค่า 3 ขนาดคือ 0.4, 0.6 และ 0.8 ขนาดตัวอย่างมี 3 ขนาดคือ 10, 30 และ 50 ความเบ้ของข้อมูลมี 2 แบบคือ 0.4 และ 0.7 และความโด่งของข้อมูลมี 2 แบบคือ 2 และ 3 การพิจารณาค่าประสิทธิภาพนั้นจะพิจารณาจากค่าความคลาดเคลื่อนที่น้อยที่สุด

การดำเนินการวิจัยครั้งนี้ ใช้การจำลองสถานการณ์ด้วยเทคนิคมอนติคาร์โล โดยการคำนวณหาค่าคะแนนเพิ่มของแต่ละวิธีการ ตามแต่ละสถานการณ์ จำนวน 1,000 ครั้ง

ผลการวิจัยพบว่า เมื่อไม่ได้คำนึงถึงเพดานของคะแนนการสอบวัดแล้ว วิธีการคำนวณหาค่าคะแนนพัฒนาการด้วย วิธีดั้งเดิมหรือวิธีของลอร์ด มีค่าความคลาดเคลื่อนน้อยที่สุด แต่เมื่อมีการปรับแก้เพดานของคะแนนการสอบวัด วิธีการคำนวณหาค่าคะแนนพัฒนาการด้วย วิธีเศษเหลือเทียบกับศักยภาพของผู้สอบ มีค่าความคลาดเคลื่อนน้อยที่สุด

3) ประสิทธิ์ ชัยกาล (2539) ได้เปรียบเทียบประสิทธิภาพระหว่างโมเดลลิสเรล 3 แบบ ที่ใช้ในการศึกษาตัวแปรที่สัมพันธ์กับพัฒนาการของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ได้แก่ โมเดลลิสเรลที่มีการวัดพัฒนาการในรูปโมเดลพื้นฐานการวิเคราะห์องค์ประกอบระยะยาว โมเดลวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้ตัวเดียว และโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้หลายตัว กลุ่มตัวอย่างคือนักเรียนชั้นประถมศึกษาปีที่ 6 โรงเรียนสังกัดกรุงเทพมหานคร จำนวน 606 คน เครื่องมือที่ใช้ในการวิจัยได้แก่แบบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ โดยเก็บข้อมูลในการวัดตัวแปรด้านผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ และเจตคติต่อพฤติกรรมการเรียนคณิตศาสตร์ 3 ครั้ง และวัดตัวแปรด้านความถนัดทางคณิตศาสตร์ 1 ครั้ง

ผลการวิจัยพบว่า โมเดลลิสเรลที่มีการวัดพัฒนาการในรูปโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้หลายตัว มีประสิทธิภาพสูงที่สุด เพราะมีความคลาดเคลื่อนของโมเดลต่ำ และบ่งชี้ได้ว่าความสัมพันธ์ของตัวแปรที่สัมพันธ์กับพัฒนาการมีความไม่แปรเปลี่ยน รองลงมาคือโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้หลายตัวมีประสิทธิภาพสูงสุด เพราะโมเดลสามารถประมาณค่าพารามิเตอร์ที่บ่งชี้พัฒนาการที่เกิดขึ้นตลอดช่วงเวลาได้ และโมเดลมีความคลาดเคลื่อนต่ำสุด

4) เอื้อมพร หลินเจริญ (2539) ได้พัฒนาโมเดลการศึกษาพัฒนาการคุณลักษณะทางวิทยาศาสตร์ของนักเรียน ซึ่งประกอบด้วยตัวแปรสังเกตได้ 14 ตัวแปร และตัวแปรแฝง 4 ตัวแปร ได้แก่ ลักษณะของนักเรียน สภาพแวดล้อมทางครอบครัว คุณลักษณะทางวิทยาศาสตร์ก่อนเรียนและหลังเรียน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่อยู่ในโรงเรียนสังกัดกรมสามัญศึกษาในเขตกรุงเทพมหานคร และในจังหวัดกำแพงเพชรที่อยู่ในอำเภอเมืองและนอกเขตอำเภอเมืองจำนวน 450 คน ผลการวิจัยพบว่าโมเดลการศึกษาพัฒนาการคุณลักษณะทางวิทยาศาสตร์ที่ผู้วิจัยพัฒนาขึ้น มีความสอดคล้องกับข้อมูลเชิงประจักษ์โมเดลไม่แปรเปลี่ยนระหว่างเวลา แสดงให้เห็นว่าการวัดพัฒนาการทั้ง 2 ครั้ง เป็นการวัดตัวแปรแฝงตัวเดียวกัน ส่วนผลการทดสอบความไม่แปรเปลี่ยนระหว่างกลุ่ม พบว่าโมเดลยังไม่สอดคล้องกับข้อมูลเชิงประจักษ์ ค่ามัชฌิมเลขคณิตของคะแนนพัฒนาการคุณลักษณะทางวิทยาศาสตร์ของนักเรียนที่ศึกษาในกรุงเทพมหานคร สูงกว่านักเรียนที่ศึกษาในจังหวัดกำแพงเพชรที่อยู่นอกเขตอำเภอเมืองอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

5) อิทธิพงษ์ ตั้งสกุลเรืองไล (2541) ได้ทำการเปรียบเทียบประสิทธิภาพของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง 4 รูปแบบ คือ โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงและกำหนดค่าพารามิเตอร์อิสระ โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงและกำหนดค่าพารามิเตอร์คงที่ โมเดลพัฒนาการเชิงเส้นตรง และโมเดลพัฒนาการพื้นฐานที่ไม่มีค่าความชัน โดยศึกษาการเปลี่ยนแปลงระยะยาวของพัฒนาการด้านผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ และพัฒนาการทางกายภาพด้านน้ำหนักและส่วนสูงของนักเรียนประถมศึกษา ซึ่งใช้ดัชนีในการเปรียบเทียบ ได้แก่ ค่าสถิติไค-สแควร์ ดัชนีวัดระดับความกลมกลืน ดัชนีรากกำลังที่สองเฉลี่ยของเศษเหลือ และค่าความคลาดเคลื่อนในรูปคะแนนมาตรฐานสูงสุด กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นประถมศึกษาโรงเรียนในสังกัดกรุงเทพมหานครและกรมสามัญศึกษา จำนวน 406 คน ในสังกัดสำนักงานการประถมศึกษา จังหวัดพิษณุโลก จำนวน 592 คน เครื่องมือที่ใช้ในการวิจัยได้แก่ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ และแบบบันทึกการตรวจสุขภาพของนักเรียน โดยเก็บข้อมูลระยะยาว 5 ครั้ง วิเคราะห์ข้อมูลโดยใช้โมเดลลิสเรลแบบมีตัวแปรแฝง

ผลการวิจัยพบว่า โค้งพัฒนาการผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนมีลักษณะเป็นเส้นโค้งพาราโบลาคว่ำ ในขณะที่น้ำหนักและส่วนสูงของนักเรียนมีลักษณะเป็นเส้นโค้งพาราโบลาหงาย และโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงและกำหนดค่าพารามิเตอร์อิสระชนิดความแปร-ปรวนของความคลาดเคลื่อนไม่เท่ากัน มีประสิทธิภาพในการศึกษาพัฒนาการเกี่ยวกับการวัดพัฒนาการในระยะยาวของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ และพัฒนาการทางกายภาพด้านน้ำหนักและส่วนสูงได้ดีที่สุด

6) มนต์ทิวา ไชยแก้ว (2542) ได้เปรียบเทียบผลการวิเคราะห์การเปลี่ยนแปลงระยะยาวที่ใช้โมเดลประยุกต์โค้งพัฒนาการที่มีตัวแปรแฝงระหว่างโมเดลที่มีข้อมูลครบสมบูรณ์จากการวัดระยะยาว 5 ครั้ง กับโมเดลที่มีการวัดไม่ครบสมบูรณ์และโมเดลที่มีข้อมูลขาดหาย โมเดลการวัดไม่ครบสมบูรณ์ประกอบด้วยโมเดลที่มีการวัด 3 ครั้งรวม 6 โมเดล และโมเดลที่มีการวัด 4 ครั้ง รวม 4 โมเดล ซึ่งมีรูปแบบการวัดไม่ครบสมบูรณ์แตกต่างกัน โมเดลที่มีข้อมูลขาดหายประกอบด้วยโมเดลที่มีอัตราการขาดหายแตกต่างกันเป็น 5% 10% 15% และ 20% รวม 4 โมเดล ตามลำดับ ข้อมูลสำหรับการวิจัยเป็นข้อมูลทุติยภูมิรวม 4 ชุด ชุดแรกเป็นฐานข้อมูลการวัดระยะยาว 5 ครั้ง ของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ จากนักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 406 คน ในโรงเรียนสังกัดกรุงเทพมหานคร ฐานข้อมูลอีก 3 ชุด ที่เหลือเป็นข้อมูลอัตราการเข้าเรียนในระดับประถมศึกษา มัธยมศึกษาและอุดมศึกษาจากประเทศต่าง ๆ จำนวน 125 , 118 และ 112 ประเทศ เก็บข้อมูลในปี ค.ศ. 1975, 1980, 1985, 1990 และ 1995 ตามลำดับ การวิเคราะห์ความกลมกลืนของโมเดลประยุกต์โค้งพัฒนาการที่มีตัวแปรแฝงกับข้อมูลประจักษ์ทั้ง 4 ชุด ใช้โปรแกรมลิสเรล เวอร์ชั่น 8.10 ด้วยรูปแบบคำสั่งสัญลักษณ์แรม และโปรแกรมอีคิวเอส สำหรับวินโดว์ เวอร์ชั่น 5.7 บี

ผลการวิจัยสรุปได้ว่า 1) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการเปลี่ยนแปลงระยะยาวของโมเดลที่มีการวัดไม่ครบสมบูรณ์กับโมเดลที่มีการวัดครบสมบูรณ์ส่วนใหญ่มีค่าสูงกว่า.70 โดยโมเดลที่วัดได้ 4 ครั้งมีค่าสัมประสิทธิ์สหสัมพันธ์สูงกว่าโมเดลที่วัดได้ 3 ครั้ง และโมเดลที่มีช่วงระยะเวลาที่วัดนานมีค่าสัมประสิทธิ์สหสัมพันธ์สูงกว่าโมเดลที่ช่วงระยะเวลาที่วัดสั้น 2) ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสองระหว่างการเปลี่ยนแปลงระยะยาวจากโมเดลที่มีการวัดครบสมบูรณ์กับโมเดลที่วัดได้ 3 ครั้งมีค่าสูงกว่าโมเดลที่วัดได้ 4 ครั้ง 3) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการเปลี่ยนแปลงระยะยาวของโมเดลที่มีการวัดครบสมบูรณ์กับโมเดลที่มีอัตราการขาดหายของข้อมูลต่ำที่มีค่าสูงกว่าโมเดลที่มีอัตราการขาดหายสูง 4) ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ยกำลังสองระหว่างการเปลี่ยนแปลงระยะยาวจากโมเดลที่มีการวัดครบสมบูรณ์กับโมเดลมีอัตราการขาดหายต่ำมีค่าสูงกว่าโมเดลที่มีอัตราการขาดหายสูง

7) สมถวิล วิจิตรวรรณา (2543) ได้ทำการเปรียบเทียบประสิทธิภาพของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง โมเดลพหุระดับ และโมเดลกึ่งซิมเพลกซ์ ในการวัดพัฒนาการระยะยาวทั้งตัวแปรเดี่ยวและตัวแปรพหุ โดยใช้ดัชนีความสอดคล้องของโมเดลกับข้อมูล ได้แก่ สถิติไค-สแควร์ ดัชนี GFI และ CFI ดัชนีรากกำลังสองเฉลี่ยของเศษเหลือ (RMR) และดัชนีความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ (RMSEA) ซึ่งการวิจัยครั้งนี้ได้นำเสนอโมเดลการวัดพัฒนาการโมเดลใหม่ คือโมเดลกึ่งซิมเพลกซ์ที่มีตัวแปรแฝงพัฒนาการที่สามารถอธิบายคะแนนเริ่มต้นและอัตราพัฒนาการที่แท้จริง ข้อมูลในงานวิจัยนี้ได้มาจากประชากรนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 469 คน จากโรงเรียนขยายโอกาสทาง

การศึกษา สังกัดสำนักงานการประถมศึกษาจังหวัดสมุทรสงคราม ภาคเรียนที่ 1 ปีการศึกษา 2542 ทำการทดสอบ 5 ครั้ง ห่างกันครั้งละ 3 สัปดาห์ เครื่องมือในการวิจัยครั้งนี้เป็นแบบทดสอบคู่ขนานวิชาคณิตศาสตร์ 5 ฉบับ ที่วัด 2 คุณลักษณะ คือการคิดคำนวณและการแก้โจทย์ปัญหา ซึ่งแบบทดสอบฉบับที่ 4 ไม่เป็นคู่ขนานกับแบบทดสอบฉบับอื่น ๆ แต่ในการวัดคุณลักษณะการแก้โจทย์ปัญหาแบบสอบทั้ง 5 ฉบับ คู่ขนานกัน การวิเคราะห์ข้อมูลโมเดลพหุระดับใช้โปรแกรม HLM ส่วนโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง และโมเดลกึ่งซิมเพลกซ์ที่มีตัวแปรแฝงพัฒนาการ ใช้โปรแกรมการวิเคราะห์ความแปรปรวนร่วม EQS

ผลการวิจัยพบว่า โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงเป็นโมเดลที่ใช้อธิบายการวัดพัฒนาการระยะยาวที่มีประสิทธิภาพสูงสุด ทั้งชนิดตัวแปรเดี่ยว และตัวแปรพหุ รองลงมาคือโมเดลกึ่งซิม เพลกซ์ที่มีตัวแปรแฝงพัฒนาการ และโมเดลพหุระดับ

โดยสรุปจากการศึกษาค้นคว้างานวิจัยที่เกี่ยวกับการวัดพัฒนาการ พบว่า วิธีการวัดพัฒนาการรายบุคคลเพื่อให้สอดคล้องกับพระราชบัญญัติการศึกษาแห่งชาติ 2542 นั้น ส่วนใหญ่จะเป็นวิธีการวัดที่ต้องใช้การวัดหลายครั้งและเป็นวิธีการวัดที่ต้องใช้ความสามารถในการวิเคราะห์สูง (รายละเอียดดังตารางที่ 2.4) ซึ่งไม่เหมาะที่จะนำไปใช้ในระดับสถานศึกษา ดังนั้นผู้วิจัยจึงต้องการศึกษาหาวิธีการวัดพัฒนาการวิธีที่ง่าย มีจำนวนครั้งในการวัดน้อยครั้งและเหมาะสมกับสถานภาพของสถานศึกษา โดยใช้โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (Latent Growth Curve Model: LGCM) เป็นวิธีเกณฑ์ ซึ่งเป็นโมเดลที่ได้รับการยอมรับว่าเป็นโมเดลที่ดีที่สุดในการศึกษาพัฒนาการ และเป็นโมเดลที่ได้รับความนิยมในการนำไปใช้วัดพัฒนาการจากนักวัดผลการศึกษาหลายท่าน ได้แก่ Rogosa, Brandt and Zimowski, 1982; Rogosa and Willett, 1985 (cited in Stoolmiller, 1995) ผู้วิจัยสนใจที่จะเปรียบเทียบคุณภาพของคะแนนพัฒนาการของวิธีการวัดคะแนนพัฒนาการที่มีการวัด 2 ครั้ง ในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีตามทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาโดยเลือกเฉพาะวิธีการวัดที่ไม่ยากในการประมาณค่าคะแนนพัฒนาการ

วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบจะดูเหมือนยากกว่าวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม แต่วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบมีจุดเด่นในด้านสามารถที่จะแก้ปัญหาที่เกิดขึ้นในการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม ในเรื่องของค่าความเที่ยงของคะแนนพัฒนาการต่ำ และคะแนนพัฒนาการที่เท่ากันมีความหมายต่างกันเมื่อความสามารถเริ่มต้นไม่เท่ากัน (Embretson, 1991b: 511; 1995: 279) และยังแก้ปัญหาในเรื่องของปฏิสัมพันธ์ระหว่างความสามารถเริ่มต้นกับความยากของข้อสอบที่มีผลต่อคะแนนพัฒนาการ (May and Nicewander, 1998) และด้วยประสิทธิภาพของเครื่องคอมพิวเตอร์ซึ่งจะช่วยให้ครูสามารถหาคะแนนพัฒนาการได้ง่ายจากโปรแกรมคอมพิวเตอร์สำเร็จรูป ดังนั้นทั้งสองกลุ่มวิธีดังกล่าวจึงน่าจะเป็นวิธีที่เหมาะสมและสะดวกสำหรับสภาพการเรียนการสอนที่ครูสามารถนำไปใช้ได้จริง และจะเป็น

ประโยชน์ในการเลือกใช้วิธีการวัดคะแนนพัฒนาการที่เหมาะสมที่จะให้ผลการวัดที่ถูกต้องเหมาะสมกับสภาพการณ์ในสถานศึกษาเพื่อให้ครู/อาจารย์สามารถนำไปใช้ได้

ตารางที่ 2.4 งานวิจัยที่เกี่ยวกับการวัดพัฒนาการ

งานวิจัย	โมเดล/สูตรที่ใช้	วัดพัฒนาการด้าน	พัฒนาการที่วัดได้	จำนวนครั้งที่วัด	การวิเคราะห์
กลุ่มวิธี IRT 1. Embretson: 1991, 1995	MRMLC	คณิตศาสตร์	รายบุคคล	3	ง่าย, มีโปรแกรมสำเร็จรูป
2. Fischer และ Parzer: 1991	LRSM	คนไข้	เป็นกลุ่ม	2	ง่าย, มีโปรแกรมสำเร็จรูป
3. Fischer และ Ponocny: 1994	LPCM	คนไข้	เป็นกลุ่ม	2	ง่าย, มีโปรแกรมสำเร็จรูป
4. May และNicewander: 1998	$L\theta = \theta_2 - \theta_1$	จำลองข้อมูล	รายบุคคล	2	ง่าย, มีโปรแกรมสำเร็จรูป
กลุ่มวิธี CTT 5. อรุณี อ่อนสวัสดิ์: 2537	ฟังก์ชั่นของพื้นความรู้เดิมการเรียนการสอนและอิทธิพลเพดาน	จำลองข้อมูล	รายบุคคล	2	ยาก
6. วินิจ เทือกทอง: 2537	$SRR = \frac{R}{F-X} \times 100$	จำลองข้อมูล	รายบุคคล	2	ปานกลาง
7. ประสิทธิ์ ไชยกาล: 2539	โมเดลวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้ตัวเดียว,หลายตัว	คณิตศาสตร์	เปรียบเทียบโมเดล	3	ยาก
8. เอื้อมพร หลินเจิญ: 2539	โมเดลวิเคราะห์องค์ประกอบระยะยาว	วิทยาศาสตร์	ทดสอบโมเดล	2	ยาก
9. อิทธิพงษ์ ตั้งสกุลเรืองไล: 2541	โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง	คณิตศาสตร์ น้ำหนักส่วนสูง	เปรียบเทียบโมเดล	5	ยาก
10. มนต์ทิวา ไชยแก้ว: 2542	โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง	คณิตศาสตร์ อัตราการเข้าเรียนในระดับประถมศึกษา มัธยมศึกษา อุดมศึกษา	เปรียบเทียบโมเดลเมื่อมีข้อมูลขาดหาย	5	ยาก
11. สมถวิล วิจิตรวรรณา: 2543	โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง โมเดลพหุระดับ และโมเดลกึ่งซิมเพลกซ์	คณิตศาสตร์	เปรียบเทียบโมเดล	5	ยาก

ตอนที่ 5 การพัฒนาวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ

จากการวิเคราะห์วิธีการวัดคะแนนพัฒนาการตามวิธีการของ May และ Nicewander นั้น พบว่าคะแนนพัฒนาการที่ได้ ($\theta_2 - \theta_1$) ยังมีจุดอ่อนอยู่ 2 ประการ คือ

ประการที่หนึ่งพัฒนาการของความสามารถที่แท้จริงยังมีอิทธิพลเพดานติดอยู่ กล่าวคือ ผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงจะมีพิสัยของศักยภาพที่จะพัฒนาได้แคบกว่าของผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนที่ต่ำกว่า ซึ่งจะทำให้ผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงมีคะแนนพัฒนาการของความสามารถที่แท้จริงต่ำกว่าผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ เนื่องจากติดอิทธิพลเพดาน เพื่อปรับแก้จุดอ่อนข้อนี้ของวิธีการวัดความแตกต่างของความสามารถที่แท้จริง ผู้วิจัยจึงได้ประยุกต์แนวคิดของ ศิริชัย กาญจนวาสี (อ้างถึงในอรุณี อ่อนสวัสดิ์, 2537: 23) ซึ่งเป็นการวัดในแนวคิดของทฤษฎีการทดสอบแบบดั้งเดิมมาใช้ในทฤษฎีของการตอบสนองข้อสอบ โดยวิธีการของศิริชัย กาญจนวาสี ต้องการลดปัญหาการถดถอยเข้าสู่ส่วนกลาง และคำนึงถึงอัตราความงอกงาม โดยนำทั้งการเปลี่ยนแปลงสัมบูรณ์ และการเปลี่ยนแปลงที่น่าจะพัฒนาได้ของแต่ละคนมาคิด ซึ่งเป็นการแก้ปัญหาเรื่องอิทธิพลเพดานได้ทางหนึ่ง ดังนั้นผู้วิจัยจึงได้ประยุกต์เป็นวิธีการวัดคะแนนพัฒนาการของความสามารถที่แท้จริงสัมพัทธ์ โดยนำเอาผลต่างระหว่างความสามารถที่แท้จริงมาเทียบกับศักยภาพที่จะพัฒนาได้ ซึ่งศักยภาพที่จะพัฒนาได้ในที่นี้หมายถึงผลต่างระหว่างความสามารถสูงสุดกับความสามารถที่แท้จริงก่อนเรียน โดยใช้สูตรดังนี้

$$S\theta = \frac{\theta_2 - \theta_1}{Max - \theta_1}$$

$S\theta$ = วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา

θ_1 = ความสามารถที่แท้จริงก่อนเรียน

θ_2 = ความสามารถที่แท้จริงหลังเรียน

Max = ความสามารถสูงสุด

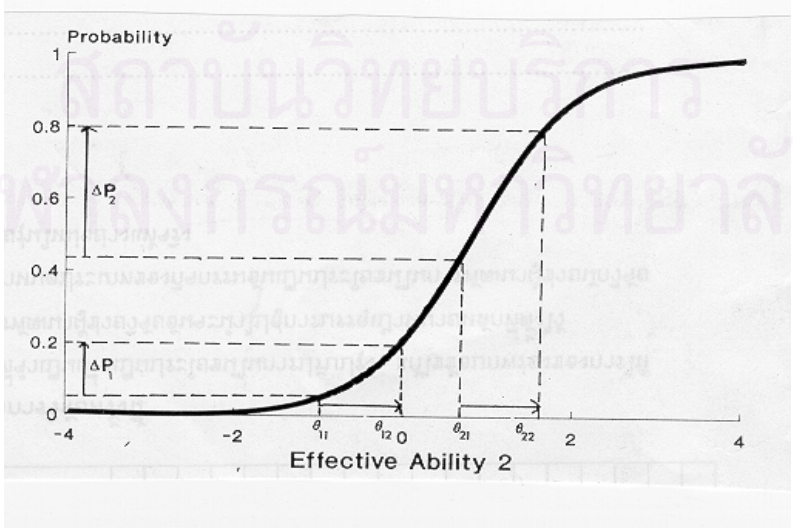
แต่เนื่องจากค่า θ เป็นคะแนนมาตรฐานที่มีค่าจาก $-\alpha$ ถึง $+\alpha$ ซึ่งสามารถกำหนดช่วงได้ตั้งแต่ 4 ถึง -4 เพราะสามารถครอบคลุมการแจกแจงของข้อมูลได้ถึงร้อยละ 99.994 (Kirk, 1995: 798) ดังนั้น Max ที่เป็นความสามารถที่แท้จริงสูงสุดสำหรับการวิจัยครั้งนี้จึงกำหนดให้มีค่าเท่ากับ 4

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

โดยมีข้อตกลงเบื้องต้นว่า การสอบทั้ง 2 ครั้ง ใช้วิธีการประมาณค่าแบบ EAP (expected a posteriori)

วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนาวิธีนี้ผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงจะมีคะแนนพัฒนาการของความสามารถที่แท้จริงสูงกว่าผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ เมื่อมีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากัน ตัวอย่างเช่น นาย ก และนาย ข มีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากันคือ 1 นาย ก มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ 2 นาย ข มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ 1 ความสามารถสูงสุดเท่ากับ 4 นาย ก จะมีคะแนนพัฒนาการของความสามารถที่แท้จริงเท่ากับ 0.5 ซึ่งสูงกว่าคะแนนพัฒนาการของความสามารถที่แท้จริงของนาย ข ที่มีค่าเท่ากับ 0.33

ประการที่สอง การตีความของคะแนนพัฒนาการของความสามารถที่แท้จริงที่เท่ากันที่มาจากผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนที่ไม่เท่ากันยังไม่เป็นธรรม Embretson (1991b) ชี้ให้เห็นว่าในทฤษฎีของการตอบสนองข้อสอบนั้น ความสามารถที่แท้จริงกับความน่าจะเป็นในการตอบข้อสอบถูกมีความสัมพันธ์ที่ไม่เป็นเชิงเส้น ผู้วิจัยมีความคิดว่าผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงมีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากันกับผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ น่าจะมีโอกาสที่จะพัฒนาความสามารถที่แท้จริงสูงกว่าผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำกว่า เพราะคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากันนั้นผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงน่าจะใช้ความสามารถที่แท้จริงเพิ่มขึ้นน้อยกว่าผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ หรือมีโอกาสที่จะพัฒนามาก/ง่ายกว่าที่จะทำให้ได้ขนาดของพัฒนาการเท่ากันดังแผนภาพที่ 2.4



แผนภาพที่ 2.4 ความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบถูก กับความสามารถที่เพิ่มขึ้น

ดังนั้นโอกาสที่จะพัฒนาความสามารถที่แท้จริงจึงน่าจะได้จากวิธีการของการเทียบผลต่างระหว่างความสามารถที่แท้จริงกับความสามารถที่แท้จริงก่อนเรียน ซึ่งเป็นแนวคิดของวินิจ เทือกทอง (2537) ที่ใช้ในทฤษฎีของการทดสอบแบบดั้งเดิม ดังนั้นผู้วิจัยจึงประยุกต์วิธีการของวินิจ เทือกทอง ได้เป็นสูตรดังนี้

$$v\theta = \frac{\theta_2 - \theta_1}{\theta_1}$$

$V\theta$ = วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน

θ_1 = ความสามารถที่แท้จริงก่อนเรียน

θ_2 = ความสามารถที่แท้จริงหลังเรียน

เนื่องจากค่า θ_1 เป็นคะแนนที่ได้จากการประมาณค่าความสามารถที่แท้จริงด้วยฟังก์ชั่นโลจิสติกโมเดลตามทฤษฎีการตอบสนองข้อสอบซึ่งมีค่าได้ตั้งแต่ $-\alpha$ ถึง $+\alpha$ ดังนั้น θ_1 จึงมีโอกาสที่จะมีค่าเป็นศูนย์ได้ ซึ่งจะทำให้ไม่สามารถหาค่าของคะแนนพัฒนาการได้ และถ้า θ_1 มีค่าเป็นลบและความแตกต่างของความสามารถที่แท้จริงมีค่าเป็นบวก แต่คะแนนพัฒนาการที่ได้จะมีค่าเป็นลบซึ่งไม่สอดคล้องกับความเป็นจริง เช่น คนที่มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ -1 และมีพัฒนาการเพิ่มขึ้นมาอย่างสูงโดยมีความสามารถที่แท้จริงหลังเรียนเท่ากับ 2 จะได้คะแนนพัฒนาการเป็น -3 เป็นต้น

ดังนั้นผู้วิจัยจึงได้ปรับค่า θ ให้มีค่าเป็นบวกเสมอโดยการปรับสูตรด้วยการนำค่า Max ไปบวกกับค่า θ ทุกค่าได้สูตรปรับใหม่ดังนี้

$$V\theta = \frac{(Max + \theta_2) - (Max + \theta_1)}{(Max + \theta_1)}$$

$$V\theta = \frac{Max + \theta_2 - Max - \theta_1}{(Max + \theta_1)}$$

$$\therefore V\theta = \frac{\theta_2 - \theta_1}{(Max + \theta_1)}$$

โดยมีข้อตกลงเบื้องต้นว่า การวัดทั้ง 2 ครั้ง ใช้วิธีการประมาณค่าแบบ EAP (expected a posteriori)

วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียนวิธีนี้ผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนสูงจะมีคะแนนพัฒนาการของความสามารถที่แท้จริงต่ำกว่าผู้สอบที่มีความสามารถที่แท้จริงก่อนเรียนต่ำ เมื่อมีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากัน ตัวอย่างเช่น นาย ก และนาย ข มีคะแนนความแตกต่างระหว่างความสามารถที่แท้จริงเท่ากันคือ 1 นาย ก มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ 1 นาย ข มีความสามารถที่แท้จริงก่อนเรียนเท่ากับ -2 นาย ก จะมีคะแนนพัฒนาการของความสามารถที่แท้จริงเท่ากับ 0.2 ซึ่งต่ำกว่าคะแนนพัฒนาการของความสามารถที่แท้จริงของนาย ข ที่มีค่าเท่ากับ 0.5

ตารางที่ 2.5 ค่า Vθ และ Sθ ที่มีค่า θ_2, θ_1 ต่างกัน

θ_2	θ_1	$\theta_2 - \theta_1$	max	Vθ	Sθ
4	3	1	4	0.1429	1.0000
2	1	1	4	0.2000	0.3333
1	0	1	4	0.2500	0.2500
0	-1	1	4	0.3333	0.2000
-1	-2	1	4	0.5000	0.1667
-2	-3	1	4	1.0000	0.1429

ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงได้พัฒนาวิธีการวัดคะแนนพัฒนาการตามแนวคิดทฤษฎีการตอบสนองข้อสอบที่มีการวัด 2 ครั้งขึ้นมาเพื่อแก้จุดอ่อนของวิธีการวัดความแตกต่างของความสามารถที่แท้จริง ได้แก่ วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์ ซึ่งมี 2 วิธีคือ

1) เมื่อเทียบกับศักยภาพการพัฒนา โดยมีสูตรดังนี้

$$S\theta = \frac{\theta_2 - \theta_1}{Max - \theta_1}$$

2) เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน โดยมีสูตรดังนี้

$$V\theta = \frac{\theta_2 - \theta_1}{Max + \theta_1}$$

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาและวิเคราะห์ตรวจสอบคุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม และทฤษฎีการตอบสนองข้อสอบ วิธีการวัดคะแนนพัฒนาการในการวิจัยครั้งนี้ประกอบด้วยวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธี ทฤษฎีการตอบสนองข้อสอบ 2 วิธี และวิธีที่ผู้วิจัยพัฒนา 2 วิธี รวมทั้งหมด 9 วิธี การเสนอวิธีการดำเนินการวิจัยแบ่งเป็น 6 ตอนได้แก่ 1) การพัฒนาวิธีการวัดคะแนนพัฒนาการ 2) การสร้างและพัฒนาเครื่องมือ 3) การตรวจสอบคุณสมบัติของแบบทดสอบ 4) ประชากรและกลุ่มตัวอย่าง 5) การเก็บรวบรวมข้อมูล และ 6) การวิเคราะห์ข้อมูล ซึ่งมีรายละเอียดในการดำเนินการวิจัยดังนี้

ตอนที่ 1 การพัฒนาวิธีการวัดคะแนนพัฒนาการ

1.1 การศึกษาค้นคว้าแนวคิด วิธีการวัดคะแนนพัฒนาการจากเอกสาร งานวิจัยตลอดจนตำราที่เกี่ยวข้องและการรวบรวมวิธีการวัดคะแนนพัฒนาการมาวิเคราะห์ดังรายละเอียดต่อไปนี้

1.1.1 จัดกลุ่มวิธีการวัดคะแนนพัฒนาการแยกออกเป็น 2 กลุ่มวิธีคือ กลุ่มวิธีทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีทฤษฎีการตอบสนองข้อสอบ

1.1.2 วิเคราะห์จุดอ่อนและจุดแข็งของวิธีการวัดคะแนนพัฒนาการกลุ่มวิธีทฤษฎีการทดสอบแบบดั้งเดิม ที่มีการวัด 2 ครั้ง (two waves) 10 วิธีได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ 2) วิธีการวัดคะแนนพัฒนาการของลอร์ด 3) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือ 4) วิธีการวัดคะแนนพัฒนาการที่เป็นอิสระจากคะแนนสอบก่อนเรียน 5) วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน 6) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ 7) วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ 8) วิธีการวัดคะแนนพัฒนาการโดยขจัดอิทธิพลเพดาน 9) วิธีการวัดคะแนนพัฒนาการโดยการเทียบส่วนร้อยของคะแนนพัฒนาการ และ 10) วิธีการวัดคะแนนพัฒนาการโดยการเทียบคะแนนพัฒนาการกับศักยภาพของผู้สอบ

1.1.3 วิเคราะห์จุดอ่อนและจุดแข็งของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม ที่มีการวัดมากกว่า 2 ครั้ง (multi-wave) ได้แก่วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง (Latent Growth Curve Model)

1.1.4 วิเคราะห์จุดอ่อนและจุดแข็งของวิธีการวัดคะแนนพัฒนาการกลุ่มวิธีทฤษฎีการตอบสนองข้อสอบ 5 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง 2) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดล LLRA 3) วิธีการวัดคะแนนพัฒนาการโดยโมเดล

LRSM 4) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดล LPCM และ 5) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดล MRMLC

1.2 การกำหนดวิธีการวัดคะแนนพัฒนาการ จุดอ่อนและจุดแข็งของวิธีการวัดคะแนนพัฒนาการ 10 วิธีในกลุ่มทฤษฎีการทดสอบแบบดั้งเดิม และวิธีการวัดคะแนนพัฒนาการ 5 วิธีในกลุ่มทฤษฎีการตอบสนองข้อสอบ ในข้อ1.1 ผู้วิจัยได้กำหนดวิธีการวัดคะแนนพัฒนาการในแต่ละกลุ่มทฤษฎีที่จะนำมาศึกษา โดยเลือกวิธีการวัดคะแนนพัฒนาการที่เป็นวิธีที่ง่าย ไม่ซับซ้อน ไม่เสียเวลามากในการวัด มีค่าความเที่ยงสูงและค่าความคลาดเคลื่อนในการวัดต่ำ ผลการกำหนดวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีทฤษฎีการตอบสนองข้อสอบมีดังนี้

1.2.1 วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม มี 5 วิธี ดังนี้

1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ

$$DS = X_2 - X_1$$

เมื่อ DS คือวิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ

X_2 คือคะแนนสอบหลังเรียน

X_1 คือคะแนนสอบก่อนเรียน

2) วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน

$$SC = Z_2 - Z_1$$

เมื่อ SC คือ วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน

Z_1 คือ คะแนนมาตรฐานของคะแนนสอบก่อนเรียน

Z_2 คือ คะแนนมาตรฐานของคะแนนสอบหลังเรียน

3) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ

$$NL = \ln X_2 - \ln X_1$$

เมื่อ NL คือ วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ

$\ln X_1$ คือ ลอการิทึมธรรมชาติของคะแนนสอบก่อนเรียน

$\ln X_2$ คือ ลอการิทึมธรรมชาติของคะแนนสอบหลังเรียน

4) วิธีการวัดคะแนนพัฒนาการสัมพัทธ์.

$$SR = \frac{100\,(y - x)}{F - X}$$

เมื่อ SR คือวิธีการวัดคะแนนพัฒนาการสัมพัทธ์

F คือ คะแนนเต็มในการวัด

Y คือ คะแนนสอบครั้งหลัง

X คือ คะแนนสอบครั้งแรก

5) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ

$$SRR = \frac{100R}{F - X}$$

เมื่อ SRR คือ วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ

R คือ คะแนนพัฒนาการวิธีส่วนที่เหลือ

F คือ คะแนนเต็ม

X คือ คะแนนสอบครั้งแรก

1.2.2 วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ มี 2 วิธีดังนี้

1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง

$$L\theta = \theta_2 - \theta_1$$

เมื่อ $L\theta$ วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง

θ_1 คือความสามารถที่แท้จริงก่อนเรียน

θ_2 คือความสามารถที่แท้จริงหลังเรียน

2) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (MRMLC)

1.3 การพัฒนาวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ ผู้วิจัยได้ประยุกต์แนวคิดในการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมของ ศิริชัย กาญจนวาสี (2532) (อ้างถึงในอรุณี อ่อนสวัสดิ์, 2537: 23) และแนวคิดของวิทิจ เทือกทอง (2537) มาใช้ในการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ กรณีการวัด 2 ครั้ง เพื่อให้ได้วิธีการที่ง่าย ไม่ซับซ้อน ได้เป็นวิธีการวัด 2 วิธี ดังนี้

1) เมื่อเทียบกับศักยภาพการพัฒนา

$$S\theta = \frac{\theta_2 - \theta_1}{Max - \theta_1}$$

เมื่อ $S\theta$ คือ วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา

θ_1 คือ ความสามารถที่แท้จริงก่อนเรียน

θ_2 คือ ความสามารถที่แท้จริงหลังเรียน

Max คือ ความสามารถสูงสุด

2) เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน

$$V\theta = \frac{\theta_2 - \theta_1}{Max + \theta_1}$$

เมื่อ Vθ คือ วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริง

โดยมีข้อตกลงเบื้องต้นว่า การวัดทั้ง 2 ครั้ง ใช้วิธีการประมาณค่าแบบ EAP (expected a posteriori)

1.4 การกำหนดวิธีการวัดคะแนนพัฒนาการที่จะใช้เป็นเกณฑ์ ผู้วิจัยได้ใช้โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงซึ่งเป็นโมเดลที่มีผู้นิยมใช้กันมากในการวัดคะแนนพัฒนาการมาเป็นเกณฑ์ โดยใช้โมเดลที่มีการวัด 3 ครั้งสำหรับข้อมูลทุกชุด และใช้โมเดลที่มีการวัด 5 ครั้งสำหรับข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน ซึ่งเป็นข้อมูลจากฐานข้อมูลวิทยานิพนธ์ของสมถวิล วิจิตรวรรณา รายละเอียดดังตาราง 3.1

ตารางที่ 3.1 ลักษณะข้อมูล 3 ชุด และการใช้ข้อมูลแต่ละชุดในการประมาณค่าคะแนนพัฒนาการ

	ข้อมูลชุดที่ 1	ข้อมูลชุดที่ 2	ข้อมูลชุดที่ 3
แบบทดสอบ	คณิตศาสตร์ ม. 3	คณิตศาสตร์ ม. 3	คณิตศาสตร์ ม. 2
จำนวนครั้งที่วัด	3	3	5
การตรวจให้คะแนน	dichotomous, partial credit	dichotomous, partial credit	dichotomous
วิธีการวัด	ใช้คะแนนครั้งที่	ใช้คะแนนครั้งที่	ใช้คะแนนครั้งที่
กลุ่มวิธี CTT			
1. DS	1, 3	1, 3	1, 3 และ 1, 5
2. SC	1, 3	1, 3	1, 3 และ 1, 5
3. NL	1, 3	1, 3	1, 3 และ 1, 5
4. SR	1, 3	1, 3	1, 3 และ 1, 5
5. SRR	1, 3	1, 3	1, 3 และ 1, 5
กลุ่มวิธี IRT			
1. Lθ	1, 3	1, 3	1, 3 และ 1, 5
2. MRMLC	1, 3	1, 3	1, 3 และ 1, 5
กลุ่มวิธีที่พัฒนา			
1. Sθ	1, 3	1, 3	1, 3 และ 1, 5
2. Vθ	1, 3	1, 3	1, 3 และ 1, 5
เกณฑ์สำหรับการเปรียบเทียบ			
GS	1-3	1-3	1-5

ตอนที่ 2 การสร้างและพัฒนาเครื่องมือ

เครื่องมือในการวิจัยครั้งนี้เป็นแบบทดสอบผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 ในภาคเรียนที่ 1 ซึ่งผู้วิจัยใช้เป็นเครื่องมือในการเก็บข้อมูล โดยมีขั้นตอนในการสร้างและพัฒนาเครื่องมือดังนี้

2.1 การพัฒนารูปแบบของแบบทดสอบที่ใช้วัดความรู้บางส่วน

2.1.1 การศึกษารูปแบบของแบบทดสอบที่ใช้วัดความรู้บางส่วน

ผู้วิจัยได้ศึกษาเอกสาร งานวิจัยที่เกี่ยวกับการให้คะแนนความรู้บางส่วน พบว่า แบบทดสอบอัตนัยเป็นเครื่องมือที่ดีที่สามารถวัดความรู้บางส่วนและให้คะแนนตามความรู้บางส่วนที่มีอยู่ของผู้เรียน แต่จากประสบการณ์ในการสอนวิชาคณิตศาสตร์และสถิติของผู้วิจัยรวมทั้งจากการสัมภาษณ์อาจารย์ผู้สอนวิชาคณิตศาสตร์ในระดับมัธยมศึกษา 3 ท่าน และจากการสัมภาษณ์นักเรียนในระดับมัธยมศึกษา 2 คน พบว่าถ้าออกข้อสอบคณิตศาสตร์เป็นแบบอัตนัยให้แสดงวิธีทำ ผู้เรียนส่วนใหญ่จะไม่ทำแบบทดสอบอัตนัย ส่งการะดาษเปล่า หรือเขียนแต่คำตอบ ไม่แสดงวิธีทำ โดยเฉพาะอย่างยิ่งถ้าเป็นผู้เรียนที่เรียนอ่อนจะไม่ลองคิดและไม่ทำเลย เพราะผู้เรียนไม่รู้จะตั้งต้นเขียนแสดงวิธีทำในแต่ละบรรทัดอย่างไรแต่จะคิดคำนวณเป็นคำตอบออกมา ดังนั้นเมื่อเป็นแบบทดสอบอัตนัยให้แสดงวิธีทำผู้เรียนจึงไม่มั่นใจในการเขียนแสดงวิธีทำจึงไม่ทำ แต่เมื่อเป็นแบบทดสอบแบบเลือกตอบผู้เรียนสามารถหาคำตอบได้ ดังนั้นผู้วิจัยจึงได้ใช้กรอบแนวคิดของไรท์และมาสเตอร์ส (Wright and Masters,1982: 41) ในการให้คะแนนความรู้บางส่วนในแบบทดสอบคณิตศาสตร์ที่เป็นข้อสอบอัตนัยแสดงวิธีทำมาปรับเป็นแบบทดสอบเลือกตอบที่วัดความรู้บางส่วน โดยมีวิธีการดังนี้

ตัวอย่างคำถาม $\sqrt{(9.0 / 0.3) - 5}$ มีค่าเท่ากับเท่าไร

ไรท์และมาสเตอร์ส ได้ให้วิธีการตรวจให้คะแนนดังนี้

ผู้เรียนไม่ทำขั้นตอนใดเลย			ได้ 0 คะแนน
9.0 / 0.3	=	30	ได้ 1 คะแนน
30 - 5	=	25	ได้ 2 คะแนน
$\sqrt{25}$	=	5	ได้ 3 คะแนน

ผู้วิจัยจึงได้ประยุกต์ตัวอย่างคำถามเป็นข้อสอบเลือกตอบย่อย 4 ตัวเลือกจำนวน 3 ข้อ ดังนี้

ข้อที่ 1 9.0 / 0.3 มีค่าเท่ากับเท่าไร

ก. 0.03 ข. 0.3 ค. 3.0 ง. 30.0

ข้อที่ 2 30 - 5 มีค่าเท่ากับเท่าไร

ก. 29.5 ข. 25.0 ค. 15.0 ง. 6.0

ข้อที่ 3 $\sqrt{(9.0 / 0.3) - 5}$ มีค่าเท่ากับเท่าไร

ก. 5 ข. 12.5 ค. 15.0 ง. 25.0

การให้คะแนน ผู้วิจัยได้ให้คะแนนข้อสอบเลือกตอบข้อละ 1 คะแนน ถ้าผู้เรียนทำถูกทั้ง 3 ข้อจะได้คะแนนเท่ากับทำข้อสอบอัตนัยถูกได้คะแนนเต็ม 3 คะแนน แต่ถ้าผู้เรียนทำข้อ 1 ถูกเพียงข้อเดียวจะ

ได้คะแนนความรู้ในส่วนนี้ 1 คะแนน วิธีนี้มีข้อดีที่ว่าถึงแม้ผู้เรียนจะทำข้อ 1 ผิด คือไม่มีความรู้ในส่วนของการหารเลขที่เป็นทศนิยม แต่ถ้าทำข้อที่ 2 ถูกจะได้คะแนน 1 คะแนนในส่วนที่เป็นความรู้ในเรื่องของการลบ ผู้วิจัยจึงได้ใช้แนวคิดนี้ประยุกต์ข้อสอบอัตนัย 1 ข้อคะแนนเต็ม 3 คะแนนมาเป็นข้อสอบปรนัยแบบเลือกตอบย่อย 3 ข้อๆละ 1 คะแนน โดยรวมคะแนนรายข้อรวม 3 ข้อเป็นคะแนนเต็มเท่ากับข้อสอบอัตนัย 1 ข้อ

นอกจากนี้ผู้วิจัยยังได้ศึกษารูปแบบของการให้ความรู้บางส่วนซึ่งมีผู้พยามยามคิดวิธีการให้คะแนนความรู้บางส่วนสำหรับแบบทดสอบที่เป็นแบบเลือกตอบหลายวิธี เช่น วิธีการให้คะแนนแบบประเพณีนิยม วิธีการแก้การเดา วิธีการให้คะแนนแบบคูมบ์ (the Coombs scoring) และวิธีการให้คะแนนของอาร์โนลด์ (the Arnold scoring, อ้างถึงใน พรทิพย์ ไชยโส, 2534) เป็นต้น แต่วิธีการให้คะแนนของอาร์โนลด์ (the Arnold scoring) นอกจากจะให้คะแนนตามความรู้บางส่วนแล้วยังมีการแก้ปัญหาการเดาซึ่งพัฒนาขึ้นมาจากโมเดลทางคณิตศาสตร์ที่ให้คะแนนที่เพิ่มขึ้นจากการเดาสุ่มเป็นศูนย์ และสามารถปรับการลงโทษผู้รู้ผิดได้ และยังทำให้คะแนนที่ได้จากแบบทดสอบมีความเที่ยงสูงขึ้น เมื่อเทียบกับการให้คะแนนแบบประเพณีนิยม (พรทิพย์ ไชยโส, 2534: 34)

ดังนั้นผู้วิจัยจึงได้ใช้วิธีการให้คะแนนของอาร์โนลด์ (the Arnold scoring) แบบทดสอบเลือกตอบแบบประยุกต์ที่ผู้วิจัยประยุกต์จากแนวคิดการให้คะแนนความรู้บางส่วนของไรท์และมาสเตอร์ส (Wright and Masters) และแบบทดสอบอัตนัย มาใช้ในการทำการทดลองรูปแบบการให้คะแนนความรู้บางส่วนว่าวิธีไหนจะให้คะแนนความรู้บางส่วนได้สัมพันธ์กับการให้คะแนนความรู้บางส่วนในแบบทดสอบอัตนัยมากที่สุด

ผู้วิจัยได้ออกข้อสอบโดยใช้เนื้อหาในวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 ในภาคเรียนที่ 1 โดยออกแบบเป็นแบบทดสอบ 3 แบบ จำนวน 2 ฉบับ โดยแต่ละฉบับแบ่งเป็นตอนๆ จำนวน 3 ตอนคือ

ตอนที่ 1 เป็นแบบทดสอบอัตนัยแสดงวิธีทำจำนวน 6 ข้อ

ตอนที่ 2 เป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก จำนวน 6 ข้อ ซึ่งเป็นโจทย์แบบเดียวกับข้อสอบอัตนัย โดยให้นักเรียนกาเครื่องหมาย X ลงในตัวเลือกที่นักเรียนคิดว่าผิดแน่นอน และกาเครื่องหมาย √ ลงในตัวเลือกที่นักเรียนคิดว่าถูกแน่นอน ตัวเลือกที่ไม่ทราบว่าถูกหรือผิดไม่ต้องกาเครื่องหมาย ซึ่งผู้วิจัยจะนำมาตรวจให้คะแนนตามวิธีของอาร์โนลด์ (Arnold) โดยมีการให้คะแนนดังนี้ (พรทิพย์ ไชยโส, 2534: 71)

-1/ 3 คะแนนเมื่อตัดตัวเลือกที่ถูกโดยเข้าใจว่าเป็นตัวลวง ถึงแม้จะตัดตัวลวงตัวอื่นๆถูกบ้าง

0 คะแนนเมื่อไม่สามารถตัดตัวลวงตัวใดๆออกได้เลย

1/ 9 คะแนนเมื่อตัดตัวลวงออกได้ถูกต้อง 1 ตัว

1/ 3 คะแนนเมื่อตัดตัวลวงออกได้ถูกต้อง 2 ตัว

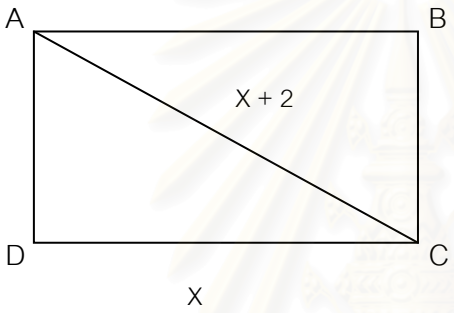
1 คะแนนเมื่อตัดตัวลวงออกได้ถูกต้องทั้งหมด 3 ตัว

ตอนที่ 3 เป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก จำนวน 18 ข้อ ซึ่งเป็นโจทย์แบบเดียวกับข้อสอบอัตนัยแต่มีการแตกข้อสอบออกเป็น 3 ข้อต่อข้อสอบอัตนัย 1 ข้อ ที่ผู้วิจัยประยุกต์ขึ้นตามแนวคิดการให้คะแนนของ ไรท์และมาสเตอร์ส

ตัวอย่างข้อสอบ

ตอนที่ 1 ให้นักเรียนแสดงวิธีทำลงในตัวข้อสอบ

1. กำหนดให้ ABCD เป็นสี่เหลี่ยมผืนผ้า AC ยาว X + 2 หน่วย BC ยาว 8 หน่วย และ CD ยาว X หน่วย ดังรูป จงหาพื้นที่ของสามเหลี่ยม ABC



……..

……

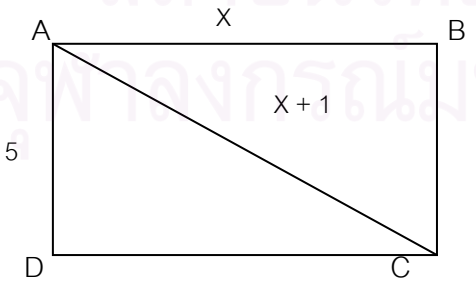
……

……

……

ตอนที่ 2 ให้นักเรียนพิจารณาตัวเลือกแต่ละตัวว่าตัวเลือกตัวใดผิด แล้วกาเครื่องหมาย X ลงบนกระดาษคำตอบให้ตรงกับตัวเลือกที่นักเรียนทราบว่าผิดแน่นอน และให้นักเรียนกาเครื่องหมาย √ ลงบนกระดาษคำตอบให้ตรงกับตัวเลือกที่นักเรียนทราบว่าถูกแน่นอน

1. กำหนดให้ ABCD เป็นสี่เหลี่ยมผืนผ้า AC ยาว X + 1 หน่วย AB ยาว X หน่วย และ AD ยาว 5 หน่วย ดังรูป สามเหลี่ยม ABC มีพื้นที่กี่ตารางหน่วย

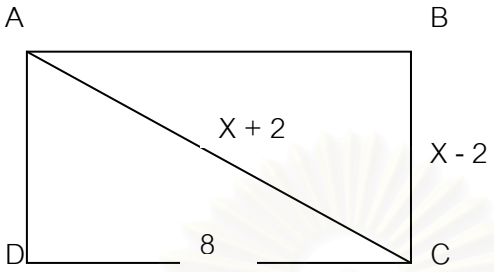


ก. 24 ข. 30 ค. 48 ง. 60

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 3 ให้นักเรียนเลือกคำตอบที่ถูกต้องที่สุดเพียงคำตอบเดียว แล้วกาเครื่องหมาย X ลงบนกระดาษคำตอบให้ตรงกับข้อที่เลือก

คำถาม กำหนดให้ ABCD เป็นสี่เหลี่ยมผืนผ้า AC ยาว X + 2 หน่วย BC ยาว X - 2 หน่วย และ CD ยาว 8 หน่วย ดังรูป



1. AB ยาวเท่าไร

ก. $2\sqrt{2X}$ ข. X ค. $\sqrt{2X^2+8}$ ง. $2X^2+8$

2. X มีค่าเท่ากับข้อใด

ก. 6 ข. 8 ค. 10 ง. 12

3. สามเหลี่ยม ABC มีพื้นที่กี่ตารางหน่วย

ก. 24 ข. 30 ค. 48 ง. 60

2.1.2 การทดลองรูปแบบของเครื่องมือวัดความรู้บางส่วน

ผู้วิจัยได้นำแบบทดสอบทั้ง 2 ฉบับไปทดสอบกับนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 4 ที่เรียนเนื้อหาคณิตศาสตร์ในระดับชั้นมัธยมศึกษาปีที่ 3 มาแล้วในวันที่ 30 เมษายน 2543 ที่โรงเรียนศึกษานารี จำนวน 44 คน และวันที่ 6 พฤษภาคม 2543 ที่โรงเรียนมัธยมวัดสิงห์จำนวน 63 คน โดยแจกแบบทดสอบทั้งฉบับ A และฉบับ B อย่างเป็นระบบให้แก่นักเรียนทั้ง 2 โรงเรียน ได้ผลของความสัมพันธ์ระหว่างการให้คะแนนความรู้บางส่วนในวิธีต่างๆกับคะแนนจากแบบทดสอบอัตนัยดังในตารางที่ 3.2

ตารางที่ 3.2 ค่าความสัมพันธ์ของคะแนนความรู้บางส่วนระหว่างวิธีการให้คะแนนแบบอัตนัย แบบอาร์โนลด์ (Arnold) และแบบอัตนัยประยุกต์

แบบทดสอบ	แบบอัตนัย		แบบอาร์โนลด์		แบบอัตนัยประยุกต์	
	ฉบับ A (49 คน)	ฉบับ B (58 คน)	ฉบับ A (49 คน)	ฉบับ B (58 คน)	ฉบับ A (49 คน)	ฉบับ B (58 คน)
แบบอัตนัย	1.00	1.00	.514**	.434**	.659**	.619**
แบบอาร์โนลด์	.514**	.434**	1.00	1.00	.432**	.430**
แบบอัตนัยประยุกต์	.659**	.619**	.432**	.430**	1.00	1.00

** มีความสัมพันธ์ที่ระดับความมีนัยสำคัญ 0.01

จากผลการทดลองแบบทดสอบทั้ง 2 ฉบับให้ผลสอดคล้องกันว่าคะแนนจากแบบสอบอัตนัยมีความสัมพันธ์กับคะแนนจากแบบทดสอบแบบเลือกตอบ(แบบอัตนัยประยุกต์)ที่ผู้วิจัยประยุกต์ขึ้นสูงกว่าคะแนนจากแบบทดสอบที่ให้คะแนนแบบอาร์โนลด์ (the Arnold scoring)

นอกจากนั้นผู้วิจัยได้สัมภาษณ์นักเรียนทั้ง 2 โรงเรียน ๆ ละ 3 คนเกี่ยวกับการทำแบบทดสอบอัตนัย แบบทดสอบแบบเลือกตอบที่ผู้วิจัยประยุกต์และแบบทดสอบที่ให้คะแนนแบบอาร์โนลด์ (the Arnold scoring) พบว่านักเรียนทั้ง 2 โรงเรียนให้คำตอบที่ตรงกันว่าอยากทำแบบทดสอบแบบเลือกตอบที่ผู้วิจัยประยุกต์มากที่สุดเพราะคุ้นเคยและง่ายในการทำข้อสอบ และถ้าให้เลือกทำแบบทดสอบระหว่างแบบทดสอบอัตนัยกับแบบทดสอบที่ให้คะแนนแบบอาร์โนลด์ (Arnold) อยากทำแบบทดสอบ อัตนัยมากกว่าแบบทดสอบที่ให้คะแนนแบบอาร์โนลด์ (the Arnold scoring) เพราะในการทำแบบทดสอบอัตนัยสามารถรู้ตัวได้เลยว่าทำได้หรือไม่ได้ ซึ่งผิดกับแบบทดสอบที่ให้คะแนนแบบอาร์โนลด์ (the Arnold scoring) ที่เมื่อทำแล้วไม่มั่นใจว่าจะทำถูกหรือทำผิด มีเพียง 1 คนที่ตอบว่าชอบแบบทดสอบที่ให้คะแนนแบบอาร์โนลด์ (the Arnold scoring) มากกว่าแบบทดสอบอัตนัยเพราะกลัวในเรื่องของการตรวจให้คะแนนที่ไม่เป็นปรนัยทำให้อาจเกิดความลำเอียงหรือไม่ยุติธรรมในการตรวจให้คะแนนได้

รูปแบบของแบบทดสอบที่ใช้ในการวัดความรู้บางส่วน

จากผลการทดลองเครื่องมือที่ใช้วัดความรู้บางส่วนข้างต้น ในการวิจัยครั้งนี้ผู้วิจัยจึงใช้แบบทดสอบแบบเลือกตอบ 4 ตัวเลือกที่ผู้วิจัยประยุกต์ขึ้นตามแนวคิดการให้คะแนนข้อสอบอัตนัยของไรท์และมาสเตอร์ (Wright and Masters, 1982) เป็นเครื่องมือในการเก็บข้อมูล โดยให้ข้อสอบอัตนัย 1 ข้อ ประยุกต์เป็นชุดของข้อสอบแบบเลือกตอบที่มี 4 ตัวเลือก ชุดละ 3 ข้อ โดยเป็นข้อสอบแบบเลือกตอบย่อย 2 ข้อ และข้อสอบเลือกตอบหลัก 1 ข้อ

2.1.3 วิธีการตรวจให้คะแนนความรู้บางส่วนของข้อมูลปฐมภูมิ

แบบทดสอบที่ผู้วิจัยพัฒนาเป็นชุดของข้อสอบเลือกตอบที่มี 4 ตัวเลือก มีวิธีการตรวจให้คะแนนความรู้บางส่วนโดยการรวมคะแนนจากชุดของข้อสอบเลือกตอบทั้ง 3 ข้อให้เป็นคะแนนความรู้บางส่วน 1 ข้อใหญ่ ถ้าตอบถูกข้อย่อย 1 ข้อจะได้คะแนน 1 คะแนน ดังนั้นในชุดของข้อสอบเลือกตอบแต่ละชุดจะมีช่วงคะแนนความรู้บางส่วนระหว่าง 0 ถึง 3 คะแนน

ส่วนวิธีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผู้วิจัยตรวจให้คะแนนเฉพาะข้อสอบเลือกตอบข้อหลัก ถ้าตอบถูก 1 ข้อจะได้คะแนน 1 คะแนน ดังนั้นในชุดของข้อสอบเลือกตอบแต่ละชุดจะมีช่วงคะแนนแบบทวิวิภาคระหว่าง 0 ถึง 1 คะแนน

2.1.4 วิธีการตรวจให้คะแนนความรู้บางส่วนของข้อมูลทุติยภูมิ

ผู้วิจัยได้ใช้ข้อมูลจากฐานข้อมูลงานวิจัยของสมถวิล วิจิตรวรรณา (2543) ที่มีเครื่องมือในการเก็บรวบรวมข้อมูลเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 2 เป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก ที่คู่ขนานกันจำนวน 5 ฉบับๆ ละ 25 ข้อ ซึ่งวัดความสามารถในการแก้โจทย์ปัญหาเกี่ยวกับการ บวก / ลบ จำนวนลบ ตรวจให้คะแนนแบบทวิภาค (dichotomous) ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน การตรวจให้คะแนนความรู้บางส่วนในข้อมูลชุดนี้ไม่สามารถใช้วิธีเดียวกับข้อ 2.1.3 ได้ ผู้วิจัยจึงดำเนินการเพื่อตรวจให้คะแนนความรู้บางส่วนของข้อมูลชุดนี้โดยมีวิธีการดังนี้

1) ศึกษารูปแบบคำตอบของการบวก/ลบจำนวนลบ ผู้วิจัยพบว่า Tatsuoka (1983 cited in Wang, Winson and Adams, 1997: 142) ได้กล่าวถึงรูปแบบคำตอบของการลบกันของจำนวนลบในวิชาคณิตศาสตร์ เช่น -7 -(-15) ว่า มากกว่าร้อยละ 99 ของคำตอบสามารถที่แยกออกได้เป็น 2 องค์ประกอบ (component) คือ เครื่องหมาย และจำนวนที่เป็นบวก (sign and absolute value) โดยมี 4 รูปแบบของคำตอบ ได้แก่ (1,1), (1,0), (0,1), และ (0,0) เมื่อตัวเลขแรกในวงเล็บแทน เครื่องหมาย ตัวเลขที่ 2 ในวงเล็บแทน จำนวนที่เป็นบวก และ 1 หมายถึงการตอบถูก ส่วน 0 หมายถึงการตอบผิด

2) ศึกษาลักษณะเฉพาะของข้อสอบ (item specification) และตัวเลือกในแบบทดสอบที่วัดความสามารถในการแก้โจทย์ปัญหาแต่ละข้อของแบบทดสอบทั้ง 5 ฉบับพบว่าแบบทดสอบในส่วนที่วัดความสามารถในการแก้โจทย์ปัญหานั้นมีรูปแบบของคำตอบในตัวเลือกดังนี้

1. มโนทัศน์ถูก การคำนวณถูก
2. มโนทัศน์ถูก การคำนวณผิด
3. มโนทัศน์ผิด การคำนวณถูก
4. มโนทัศน์ผิด การคำนวณผิด

3) ตรวจสอบรูปแบบคำตอบของตัวเลือกแต่ละข้อ ผู้วิจัยได้ดำเนินการระบุรูปแบบคำตอบของตัวเลือกในข้อสอบแต่ละข้อ และตรวจสอบว่าตัวเลือกในข้อสอบแต่ละข้อมีรูปแบบตรงกับที่ผู้วิจัยระบุไว้หรือไม่โดยดำเนินการดังนี้

1. สุ่มแบบทดสอบมา 1 ฉบับ (ฉบับที่ 3) แปลงข้อสอบที่วัดความสามารถในการแก้โจทย์ปัญหาทั้ง 25 ข้อที่เป็นแบบเลือกตอบให้เป็นข้อสอบอัตนัยแสดงวิธีทำ

2. นำแบบทดสอบไปทดสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 3 โรงเรียนวัดประดู่ในทรงธรรมจำนวน 2 ห้องเรียน 64 คน ในเดือน สิงหาคม 2544 โดยแบ่งนักเรียนออกเป็น 2 กลุ่มนักเรียนกลุ่มที่ 1 ทำข้อสอบอัตนัยข้อคี่ นักเรียนกลุ่มที่ 2 ทำข้อสอบอัตนัยข้อคู่ ผลการตรวจข้อสอบ อัตนัยที่ให้นักเรียนแสดงวิธีทำพบว่าคำตอบของนักเรียนที่ตรงกับตัวเลือกแต่ละตัวใน

ข้อสอบแต่ละข้อมีรูปแบบการคำนวณตามที่ผู้วิจัยวิเคราะห์ไว้ถึงร้อยละ 70 ดังนั้นผู้วิจัยจึงแก้ไขรูปแบบการคำนวณของตัวเลือกนั้นให้ถูกต้อง ซึ่งรูปแบบการคำนวณที่ไม่ตรงกับที่ผู้วิจัยระบุไว้ในแต่ละตัวเลือกโดยมากมีลักษณะดังนี้คือ

รูปแบบการคำนวณที่กำหนด	รูปแบบการคำนวณที่นักเรียนทำ
มโนทัศน์ถูก การคำนวณผิด	มโนทัศน์ผิด การคำนวณถูก
มโนทัศน์ผิด การคำนวณถูก	มโนทัศน์ถูก การคำนวณผิด

ผู้วิจัยได้ตรวจให้คะแนนความรู้บางส่วนของข้อมูลทุติยภูมิชุดนี้ดังนี้

1. ตัวเลือกที่มีมโนทัศน์ถูก การคำนวณถูก ให้ (1,1) คะแนน = 2 คะแนน
2. ตัวเลือกที่มีมโนทัศน์ถูก การคำนวณผิด ให้ (1,0) คะแนน =1 คะแนน
3. ตัวเลือกที่มีมโนทัศน์ผิด การคำนวณถูก ให้ (0,1) คะแนน =1 คะแนน
4. ตัวเลือกที่มีมโนทัศน์ผิด การคำนวณผิด ให้ (0,0) คะแนน =0 คะแนน

ข้อมูลชุดนี้จึงมีคะแนนความรู้บางส่วนในแต่ละข้อ ระหว่าง 0 ถึง 2 คะแนน

ตัวอย่างการให้คะแนน

ถ้าเลขจำนวนหนึ่งรวมกับ 8 ได้เท่ากับ -2 จงหาเลขจำนวนนั้น

ก. -10 ข. -6 ค. 6 ง. 10

ข้อ ก จะได้คะแนนความรู้บางส่วน 2 คะแนน เพราะ มโนทัศน์ถูก การคำนวณถูก คือ - 8 - 2 = - 10

ข้อ ข จะได้คะแนนความรู้บางส่วน 1 คะแนน เพราะมโนทัศน์ถูก การคำนวณผิด คือ - 8 - 2 = - 6

ข้อ ค จะได้คะแนนความรู้บางส่วน 0 คะแนน เพราะมโนทัศน์ผิด การคำนวณผิด คือ - 8 + 2 = 6

ข้อ ง จะได้คะแนนความรู้บางส่วน 1 คะแนน เพราะมโนทัศน์ผิด การคำนวณถูก คือ 8 + 2 = 10

2.2 การสร้างและการพัฒนาแบบทดสอบเลือกตอบประยุกต์

หลังจากที่ผู้วิจัยได้รูปแบบของแบบทดสอบวัดความรู้บางส่วนสำหรับการเก็บข้อมูล ผู้วิจัยจึงได้ดำเนินการสร้างและพัฒนาคุณภาพของแบบทดสอบเลือกตอบประยุกต์ในวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 เพื่อให้มีคุณภาพด้านความตรง ความเที่ยง และคุณภาพรายข้อ โดยมีรายละเอียดดังนี้

2.2.1 การสร้างและตรวจสอบความตรงของแบบทดสอบเลือกตอบประยุกต์

ผู้วิจัยได้ดำเนินการสร้างและตรวจสอบความตรงของแบบทดสอบเลือกตอบประยุกต์ดังนี้

1) นำเนื้อหาในแต่ละบทและวัตถุประสงค์รายวิชาของวิชาคณิตศาสตร์ ค 011 ระดับชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533) ที่กำหนดไว้ในคู่มือครูของสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี ไปให้ผู้เชี่ยวชาญจำนวน 7 คน ที่มีประสบการณ์ในการสอนคณิตศาสตร์ในระดับชั้นมัธยมศึกษาปีที่ 3 และสำเร็จการศึกษาในระดับมหาบัณฑิต ให้น้ำหนักความสำคัญของแต่ละเนื้อหาและแต่ละจุดประสงค์

2) สร้างตารางวิเคราะห์เนื้อหาและพฤติกรรม ให้มีน้ำหนักความสำคัญของเนื้อหาและวัตถุประสงค์แต่ละข้อตามน้ำหนักความสำคัญที่ได้จากผู้เชี่ยวชาญในข้อ 1 เพื่อประมาณค่าจำนวนข้อสอบในแต่ละเนื้อหาและวัตถุประสงค์ ซึ่งได้ข้อสอบทั้งหมด 20 ข้อ

3) สร้างแบบทดสอบเลือกตอบ 4 ตัวเลือก จำนวน 4 ฉบับ ๆ ละ 20 ข้อให้มีลักษณะคู่ขนานกันในด้านเนื้อหา นำแบบทดสอบเลือกตอบทั้ง 4 ฉบับ (A, B, C และ D) ให้ผู้เชี่ยวชาญทางด้านวัดผลและด้านการสอนคณิตศาสตร์จำนวน 7 คน ที่มีประสบการณ์ด้านการสอนคณิตศาสตร์และสำเร็จการศึกษาในระดับมหาบัณฑิตทางด้านวัดและประเมินผลการศึกษาตรวจสอบว่าข้อสอบแต่ละข้อสอดคล้องกับเนื้อหาและจุดประสงค์ที่ต้องการวัดหรือไม่ตามตารางวิเคราะห์เนื้อหาและพฤติกรรม (ตรวจสอบความตรงตามเนื้อหา) และนำข้อเสนอแนะของผู้เชี่ยวชาญมาปรับปรุงข้อสอบ

4) นำแบบทดสอบเลือกตอบ ทั้ง 4 ฉบับ ที่ปรับปรุงแล้ว มาปรับเป็นแบบทดสอบเลือกตอบประยุกต์ 4 ตัวเลือก โดยขยายข้อสอบแต่ละข้อให้เป็นข้อสอบย่อยแบบเลือกตอบ 3 ข้อย่อย ดังนั้นแบบทดสอบเลือกตอบแต่ละฉบับที่มีจำนวน 20 ข้อ จึงขยายออกเป็นแบบทดสอบเลือกตอบประยุกต์ 4 ตัวเลือก จำนวน 60 ข้อ นำแบบทดสอบเลือกตอบประยุกต์ทั้ง 4 ฉบับไปให้ผู้เชี่ยวชาญทางด้านวัดผลและด้านการสอนคณิตศาสตร์จำนวน 7 คน ตรวจสอบว่าข้อสอบย่อยในแต่ละข้อเป็นการวัดความรู้บางส่วนในข้อสอบข้อหลักหรือไม่ และนำข้อเสนอแนะของผู้เชี่ยวชาญมาปรับปรุงข้อสอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

2.2.2 การวิเคราะห์คุณภาพรายข้อ

เมื่อผู้วิจัยสร้างและพัฒนาแบบทดสอบเลือกตอบประยุกต์ จนมีคุณภาพด้านความตรง จึงทำการวิเคราะห์คุณภาพรายข้อของแบบทดสอบดังนี้

1) นำแบบทดสอบเลือกตอบประยุกต์ทั้ง 4 ฉบับที่ปรับปรุงเรียบร้อยแล้ว ไปทดสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 4 จำนวน 20 คน เพื่อทดสอบความเหมาะสมของภาษา แล้วนำมาปรับปรุงแก้ไข และเพื่อใช้ในการประมาณเวลาในการสอบซึ่งพบว่าต้องใช้เวลาในการสอบ 2 คาบ คือ100 นาที

2) นำแบบทดสอบเลือกตอบประยุกต์ทั้ง 4 ฉบับที่ปรับปรุงแล้ว ไปทดสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 4 จำนวน 6 โรงเรียน จำนวนนักเรียนทั้งหมด 734 คน เพื่อหาคุณภาพของข้อสอบรายข้อ รายละเอียดดังตารางที่ 3. 3

ตารางที่ 3.3 จำนวนกลุ่มตัวอย่างที่ใช้ในการทดลองแบบทดสอบเลือกตอบประยุกต์แต่ละฉบับ จำแนกตามโรงเรียน

โรงเรียน	ฉบับ A	ฉบับ B	ฉบับ C	ฉบับ D
พระโขนงวิทยาลัย	38	42	35	34
เตรียมอุดมศึกษาพัฒนาการ	39	37	36	36
พุทธจักรวิทยา	30	27	27	28
มัธยมวัดสิงห์	30	30	28	29
สุวรรณาราม	19	21	21	20
ศึกษานารี	40	36	27	24
รวม	196	193	174	171

3) นำผลการสอบในข้อ 2 ไปวิเคราะห์คุณภาพรายข้อตามทฤษฎีการทดสอบแบบดั้งเดิมด้วยโปรแกรม CTIA การที่ผู้วิจัยไม่ได้วิเคราะห์ข้อสอบรายข้อตามทฤษฎีการตอบสนองข้อสอบนั้นเป็นเพราะว่าแบบทดสอบเลือกตอบประยุกต์ที่ผู้วิจัยสร้างขึ้นในแต่ละชุดข้อสอบมีข้อสอบย่อย 3 ข้อซึ่งมีความสัมพันธ์กันระหว่างข้อสอบย่อย ดังนั้นจึงไม่เป็นไปตามข้อตกลงเบื้องต้นของการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบในด้าน local independence ผลการวิเคราะห์รายข้อของแบบทดสอบเลือกตอบประยุกต์ ดังตารางที่ 3.4

ตารางที่ 3.4 ค่าสถิติพื้นฐานของผลการวิเคราะห์รายข้อของแบบทดสอบเลือกตอบประยุกต์วิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 4 ฉบับ

ค่าสถิติ	ฉบับ A	ฉบับ B	ฉบับ C	ฉบับ D
ค่าเฉลี่ยความยาก	0.467	0.451	0.434	0.465
ค่าเฉลี่ยอำนาจจำแนก	0.488	0.480	0.466	0.527
ความเที่ยง (KR 20)	0.918	0.921	0.914	0.932
ความคลาดเคลื่อนมาตรฐานของการวัด (SEM)	3.428	3.328	3.379	3.369

4) เพื่อให้สอดคล้องกับวิธีการวัดคะแนนพัฒนาการตามทฤษฎีของการตอบสนองข้อสอบของโมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (MRMLC) ที่ Embretson เสนอให้ใช้วิธีการสอบแบบจตุรัสละติน (latin square) โดยมีข้อสอบที่มีค่าความยากปาน

กลางเป็นข้อสอบร่วม เพื่อแก้ปัญหาของข้อสอบที่จะไม่มีคุณสมบัติของความเป็นอิสระของข้อสอบ (local independence) ซึ่งผู้วิจัยได้วางแผนในการเก็บข้อมูลพัฒนาการ 3 ครั้ง ดังนั้นผู้วิจัยจึงนำแบบทดสอบเลือกตอบประยุกต์ทั้ง 4 ฉบับที่คู่ขนานกันมาจัดชุดแบบทดสอบใหม่ให้เหลือเพียง 3 ฉบับ (ก, ข และ ค) โดยนำข้อสอบย่อยที่เป็นข้อหลักแต่ละข้อจากทั้ง 4 ฉบับมาคัดเลือกให้เหลือเพียง 3 ข้อ โดยเลือกข้อสอบข้อที่มีค่าความยากใกล้เคียงกัน แล้วทำการสุ่มข้อสอบแต่ละข้อออกเป็นแบบทดสอบ 3 ฉบับ ดังนั้นแบบทดสอบเลือกตอบประยุกต์ฉบับใหม่ทั้ง 3 ฉบับจะมีข้อสอบหลัก 22 ข้อ (66 ข้อย่อย) โดย มี 18 ข้อที่เป็นข้อสอบต่างกัน และมี 4 ข้อที่เป็นข้อสอบร่วม มีค่าความยากเฉลี่ยทั้ง 3 ฉบับใกล้เคียงกัน ดังรายละเอียดในตารางที่ 3. 5

ตารางที่ 3. 5 คุณภาพของแบบทดสอบเลือกตอบประยุกต์วิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับ

ค่าสถิติ	ฉบับที่ ก	ฉบับที่ ข	ฉบับที่ ค
ค่าเฉลี่ยความยาก	0.448	0.442	0.499
ค่าเฉลี่ยอำนาจจำแนก	0.509	0.549	0.576
ความเที่ยง (KR 20)	0.913	0.918	0.928

สำหรับข้อมูลทุติยภูมิ ซึ่งได้จากฐานข้อมูลงานวิจัยของสมถวิล วิจิตรวรรณา ที่เก็บข้อมูล 5 ครั้งด้วยแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 ที่เป็นแบบทดสอบเลือกตอบ 4 ตัวเลือกที่คู่ขนานกันจำนวน 5 ฉบับๆ ละ 55 ข้อ โดยมีค่าความยากเฉลี่ย 0.49, 0.44, 0.49, 0.51 และ 0.52 ตามลำดับ และมีค่าอำนาจจำแนกเฉลี่ย 0.31, 0.30, 0.30, 0.41 และ 0.33 ตามลำดับ

ตอนที่ 3 การตรวจสอบคุณสมบัติของแบบทดสอบเลือกตอบประยุกต์

3.1 การตรวจสอบความเป็นเอกมิติ (unidimensional) ของแบบทดสอบเลือกตอบประยุกต์

สำหรับแบบทดสอบเลือกตอบประยุกต์ที่ผู้วิจัยสร้างขึ้นทั้ง 3 ฉบับ เพื่อให้เป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ผู้วิจัยจึงได้ตรวจสอบจำนวนองค์ประกอบในแบบทดสอบด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (exploratory factor analysis) แบบ alpha โดยใช้เกณฑ์ค่า eigenvalues ตั้งแต่ 1 ขึ้นไปนับเป็น 1 องค์ประกอบ จากการวิเคราะห์องค์ประกอบเชิงสำรวจ พบว่าแบบทดสอบทั้ง 3 ฉบับที่ตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) ทุกฉบับมีจำนวนองค์ประกอบเพียงองค์ประกอบเดียว โดยมีค่า eigenvalues เท่ากับ 3.609, 3.715 และ 3.875 ตามลำดับ สำหรับแบบทดสอบที่ตรวจให้คะแนน

แบบทวิวิภาค (dichotomous) และแบบทดสอบที่ตรวจให้คะแนนความรู้บางส่วน (partial credit) มีค่า eigenvalues เท่ากับ 5.831, 6.007 และ 6.739 ตามลำดับ (รายละเอียดอยู่ในตารางที่ 3.6) และผู้วิจัยได้ทำการวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis) เพื่อยืนยันว่าแบบทดสอบแต่ละฉบับมี 1 องค์ประกอบจริง ผลการวิเคราะห์พบว่าแบบทดสอบทุกฉบับทุกแบบการให้คะแนนมี 1 องค์ประกอบจริง โดยแบบทดสอบที่ตรวจให้คะแนนแบบทวิวิภาค (dichotomous) มีค่าไค-สแควร์เท่ากับ 207.30, 203.39 และ 206.77 ค่า p เท่ากับ 0.48, 0.48 และ 0.43 โดยมีชั้นแห่งความเป็นอิสระเท่ากับ 207, 203 และ 204 ตามลำดับ ส่วนแบบทดสอบที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีค่าไค-สแควร์เท่ากับ 196.76, 186.60 และ 195.88 ค่า p เท่ากับ 0.45, 0.45 และ 0.41 โดยมีชั้นแห่งความเป็นอิสระเท่ากับ 195, 186 และ 192 ตามลำดับ (รายละเอียดดังตารางที่ 3.7)

ตารางที่ 3.6 ค่า eigen ร้อยละของความแปรปรวน ความแปรปรวนสะสมของแบบทดสอบจำแนกตามแบบทดสอบ และวิธีการตรวจให้คะแนน

องค์ประกอบ	ฉบับ ก			ฉบับ ข			ฉบับ ค		
	ค่า eigen	% of Variance	Cumulative %	ค่า eigen	% of Variance	Cumulative %	ค่า eigen	% of Variance	Cumulative %
ให้คะแนนแบบทวิวิภาค									
1	3.609	16.406	16.406	3.715	16.889	16.889	3.875	17.614	17.614
ให้คะแนนความรู้บางส่วน									
1	5.831	26.506	26.506	6.007	27.303	27.303	6.739	30.633	30.633

ตารางที่ 3.7 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันของแบบทดสอบ
ทั้ง 3 ฉบับจำแนกตามวิธีการตรวจให้คะแนน

ค่าสถิติ	ฉบับ ก		ฉบับ ข		ฉบับ ค	
	dichotomous	partial credit	dichotomous	partial credit	dichotomous	partial credit
χ^2	207.30	196.76	203.39	186.60	206.77	195.88
p	0.48	0.45	0.48	0.45	0.43	0.41
df	207	195	203	186	204	192
RMR	0.058	0.052	0.056	0.050	0.057	0.049
GFI	0.90	0.91	0.91	0.91	0.90	0.91
LSR	3.06	2.59	2.12	2.27	2.58	2.36

สำหรับข้อมูลที่เป็นทุติยภูมินั้นเพื่อให้เป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ผู้วิจัยจึงได้ตรวจสอบจำนวนองค์ประกอบในแบบทดสอบด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (exploratory factor analysis) แบบ alpha โดยใช้เกณฑ์ค่า eigenvalues ตั้งแต่ 1 ขึ้นไปนับเป็น 1 องค์ประกอบ ผลการวิเคราะห์พบว่าแบบทดสอบฉบับที่ 1, 2 และ 4 ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีจำนวนองค์ประกอบทั้งหมด 2 องค์ประกอบ ส่วนแบบทดสอบฉบับที่ 3 และ 5 มีเพียงองค์ประกอบเดียว โดยมีค่า eigenvalues อยู่ในช่วง 1.001 ถึง 4.272 (รายละเอียดอยู่ในตารางที่ 3.8) และผู้วิจัยได้ทำการวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis) เพื่อยืนยันว่าแบบทดสอบแต่ละฉบับมี 1 องค์ประกอบจริง ผลการวิเคราะห์พบว่าแบบทดสอบทุกฉบับทุกแบบการให้คะแนนมี 1 องค์ประกอบจริง โดยแบบทดสอบที่ข้อมูลมีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) มีค่าไค-สแควร์เท่ากับ 244.71, 222.47, 242.40, 213.76 และ 228.71 ค่า p เท่ากับ 0.48, 0.46, 0.48, 0.47 และ 0.44 โดยมีชั้นแห่งความเป็นอิสระเท่ากับ 244, 221, 242, 213 และ 226 ตามลำดับ ส่วนแบบทดสอบที่ข้อมูลมีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีค่าไค-สแควร์เท่ากับ 239.01, 239.99, 247.28, 223.77 และ 231.43 ค่า p เท่ากับ 0.49, 0.49, 0.50, 0.44 และ 0.46 โดยมีชั้นแห่งความเป็นอิสระเท่ากับ 239, 240, 248, 221 และ 230 ตามลำดับ (รายละเอียดดังตารางที่ 3.9)

ตารางที่ 3.8 ค่า eigen ร้อยละของความแปรปรวน ความแปรปรวนสะสมของแบบทดสอบจำแนกตามแบบทดสอบ และวิธีการตรวจให้คะแนน

	dichotomous			partial credit		
องค์ประกอบ	ค่า eigen	% of Variance	Cumulative %	ค่าeigen	% of Variance	Cumulative %
แบบทดสอบฉบับที่ 1						
1	1.233	4.931	4.931	1.753	7.011	7.011
2	1.059	4.237	9.167	1.105	4.421	11.432
แบบทดสอบฉบับที่ 2						
1	1.740	6.959	6.959	1.412	5.649	5.649
2	1.333	5.332	12.291	1.001	4.002	9.651
แบบทดสอบฉบับที่ 3						
1	1.926	7.703	7.703	1.881	7.523	7.523
แบบทดสอบฉบับที่ 4						
1	3.052	12.206	12.206	2.950	11.800	11.800
2	1.529	6.118	18.324	1.190	4.762	16.562
แบบทดสอบฉบับที่ 5						
1	4.272	17.087	17.087	3.225	12.901	12.901

ตารางที่ 3.9 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันของแบบทดสอบทั้ง 5 ฉบับ จำแนกตามวิธีการตรวจให้คะแนน

ค่าสถิติ	ฉบับที่ 1	ฉบับที่ 2	ฉบับที่ 3	ฉบับที่ 4	ฉบับที่ 5
			ให้คะแนนแบบทวิวิภาค		
χ^2	244.71	222.47	242.40	213.76	228.71
p	0.48	0.46	0.48	0.47	0.44
df	244	221	242	213	226
RMR	0.043	0.043	0.040	0.042	0.039
GFI	0.96	0.96	0.96	0.96	0.96
LSR	2.87	3.37	2.53	2.77	2.33
			ให้คะแนนความรู้บางส่วน		
χ^2	239.01	239.99	247.28	223.77	231.43
p	0.49	0.49	0.50	0.44	0.46
df	239	240	248	221	230
RMR	0.042	0.041	0.042	0.039	0.039
GFI	0.96	0.96	0.95	0.96	0.96
LSR	2.98	2.42	2.38	2.63	3.04

3.2 การตรวจสอบความเป็นคู่ขนานของแบบทดสอบเลือกตอบประยุกต์

แบบทดสอบที่คู่ขนานกัน (parallel tests) นอกจากจะต้องมีโครงสร้างเนื้อหาที่คู่ขนานกัน ซึ่งตรวจสอบโดยผู้เชี่ยวชาญแล้ว ยังต้องมีการตรวจสอบทางด้านค่าสถิติ โดยการทดสอบความเท่ากันของค่าเฉลี่ยและความแปรปรวนของคะแนนจากแบบทดสอบ (ศิริชัย กาญจนวาสี, 2544: 17) ผู้วิจัยจึงได้ทำการทดสอบค่าเฉลี่ยของคะแนนจากแบบทดสอบที่สร้างขึ้นทั้ง 3 ฉบับเพื่อดูว่ามีค่าแตกต่างกันหรือไม่ โดยใช้สถิติทดสอบ One - Way ANOVA และทดสอบความแปรปรวนของคะแนนจากแบบทดสอบเลือกตอบประยุกต์ทั้ง 3 ฉบับมีค่าต่างกันหรือไม่ (test of homogeneity of variances) โดยใช้ Levene Statistic ผลการทดสอบพบว่าค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ ระดับมัธยมศึกษาปีที่ 3 จากแบบทดสอบเลือกตอบประยุกต์ทั้ง 3 ฉบับมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทั้งข้อมูลชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ($F = 0.14$, $p = 0.87$) และให้คะแนนความรู้บางส่วน (partial credit) ($F = 2.18$, $p = 0.11$) (รายละเอียดดังตารางที่ 3.10) และความแปร ปรวนของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ ระดับมัธยมศึกษาปีที่ 3 จากแบบทดสอบเลือกตอบประยุกต์ทั้ง 3 ฉบับ มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญ ทั้งข้อมูลชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) (Levene Statistic = 0.121, $p = 0.886$) และให้คะแนนความรู้บางส่วน (partial credit) (Levene Statistic = 0.021, $p = 0.979$) (รายละเอียดดังตารางที่ 3.11) ผลการทดสอบ

แสดงให้เห็นว่าแบบทดสอบเลือกตอบประยุกต์ทั้ง 3 ฉบับมีคุณสมบัติของแบบทดสอบคู่ขนาน เพราะมีค่าเฉลี่ยและความแปรปรวนของคะแนนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ตารางที่ 3.10 การทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนจากแบบทดสอบเลือกตอบประยุกต์ทั้ง 3 ฉบับ

แหล่งความแปรปรวน	dichotomous					partial credit				
	SS	df	MS	F	Sig.	SS	df	MS	F	Sig.
Between Groups	4.94	2	2.47	0.14	0.87	578.92	2	289.46	2.18	0.11
Within Groups	9139.25	510	17.92			67822.26	510	132.98		
Total	9144.19	512				68401.17	512			

ตารางที่ 3.11 การทดสอบความแตกต่างของความแปรปรวนของแบบสอบทั้ง 3 ฉบับ

dichotomous				partial credit			
Levene Statistic	df1	df2	Sig.	Levene Statistic	df1	df2	Sig.
0.121	2	510	0.886	0.021	2	510	0.979

ตอนที่ 4 ประชากรและกลุ่มตัวอย่าง

ในการวิจัยครั้งนี้ผู้วิจัยมีข้อมูลทั้งหมด 3 ชุด ซึ่งมีลักษณะดังนี้

ข้อมูลชุดที่ 1 เป็นข้อมูลที่ได้จากการวัดซ้ำ 3 ครั้งด้วยแบบทดสอบเดิม (repeated measure)

ข้อมูลชุดที่ 2 เป็นข้อมูลที่ได้จากการวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนานที่มีการจัดแบบทดสอบให้กับกลุ่มตัวอย่างแบบละติน (latin square) เพื่อให้เกิดความสมดุลย์กัน (counterbalance)ระหว่างผู้สอบกับแบบทดสอบที่ใช้และครั้งที่สอบ

ข้อมูลชุดที่ 3 เป็นข้อมูลที่ได้จากการวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence) ข้อมูลชุดที่ 1 และชุดที่ 2 เป็นข้อมูลที่ผู้วิจัยสร้างเครื่องมือและเก็บข้อมูลเอง ส่วนข้อมูลชุดที่ 3 เป็นข้อมูลทุติยภูมิที่ได้จากฐานข้อมูลงานวิจัยของ สมถวิล วิจิตรวรรณา (2543)

ดังนั้นประชากรและกลุ่มตัวอย่างจึงมีรายละเอียดดังนี้

ประชากร

ประชากรมี 2 กลุ่ม กลุ่มแรกคือนักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัดกรมสามัญศึกษา ในเขตกรุงเทพมหานคร ในปีการศึกษา 2543 มีจำนวนทั้งสิ้น 53,774 คน (ศูนย์สารนิเทศ,

กระทรวงศึกษาธิการ, 2543) กลุ่มที่สองเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 สังกัดสำนักงานการประถมศึกษา จังหวัดสมุทรสงคราม ในปีการศึกษา 2542

ประชากรกลุ่มแรกเป็นประชากรของข้อมูลปฐมภูมิที่ผู้วิจัยเก็บรวบรวมข้อมูลเอง ได้เป็นข้อมูลชุดที่ 1 และชุดที่ 2

ประชากรกลุ่มที่สองเป็นประชากรของข้อมูลทุติยภูมิที่ได้จากฐานข้อมูลงานวิจัยของสมถวิล วิจิตรวรรณา (2543)

กลุ่มตัวอย่าง

การกำหนดขนาดกลุ่มตัวอย่างและวิธีการสุ่มกลุ่มตัวอย่างสำหรับข้อมูลชุดที่ 1 และชุดที่ 2 ผู้วิจัยทำการคำนวณขนาดตัวอย่างที่ใช้ในการวิจัยครั้งนี้โดยใช้สูตรจาก ศิริชัย กาญจนวาสี ทวีวัฒน์ ปิตยานนท์ และดิเรก ศรีสุโข (2537: 26) ดังนี้

$$n_{\mu} = \frac{NZ^2\sigma^2}{NE^2 + Z^2\sigma^2}$$

โดย n_{μ} คือขนาดตัวอย่างที่พอเหมาะ

N คือจำนวนประชากร

Z คือค่าความเชื่อมั่นที่ผู้วิจัยกำหนดไว้ที่ 95%

σ^2 คือค่าความแปรปรวนของคะแนนสอบวิชาคณิตศาสตร์ของประชากร

E คือค่าความคลาดเคลื่อนสูงสุดที่จะยอมรับได้ ซึ่งผู้วิจัยตั้งไว้ที่ 0.10 ของ σ

จากการคำนวณขนาดตัวอย่างที่ผู้วิจัยยอมให้มีความคลาดเคลื่อนเกิดขึ้นสูงสุดได้ไม่เกินร้อยละ 10 ของความแปรปรวนของประชากร จะได้ขนาดตัวอย่าง 382 คน แต่เพื่อให้การวิจัยครั้งนี้มีความแกร่ง (robustness) ดังนั้นผู้วิจัยจึงเพิ่มขนาดตัวอย่างเป็น 698 คน และ 637 คน สำหรับข้อมูลชุดที่ 1 และข้อมูลชุดที่ 2 ตามลำดับ โดยใช้วิธีการสุ่มแบบหลายขั้นตอน (multistage) ดังนี้

ขั้นที่ 1 สุ่มสหวิทยาเขต โดยสุ่มสหวิทยาเขตในเขตพระนครกับเขตธนบุรี สุ่มมาเขตละ 2 สหวิทยาเขต จากทั้งหมด 24 สหวิทยาเขต (กองแผนงานกรมสามัญศึกษา: 2543) สุ่มอย่างง่าย (simple sampling) สหวิทยาเขตในเขตพระนครและเขตธนบุรี อย่างละ 1 สหวิทยาเขต เพื่อเป็นกลุ่มตัวอย่างสำหรับเก็บข้อมูลชุดที่ 1 และข้อมูลชุดที่ 2 ตามลำดับ

ขั้นที่ 2 สุ่มโรงเรียน โดยสุ่มโรงเรียนภายในสหวิทยาเขตที่ได้รับการสุ่มมาแล้วในขั้นที่ 1 ด้วยการสุ่มอย่างง่ายสหวิทยาเขตละ 3 โรงเรียน

ขั้นที่ 3 สุ่มห้องเรียน โดยสุ่มห้องเรียนภายในโรงเรียนที่ได้รับการสุ่มมาแล้วในขั้นที่ 2 ด้วยการสุ่มอย่างง่ายโรงเรียนละ 3 ห้องเรียน

สำหรับข้อมูลชุดที่ 3 ที่เป็นข้อมูลทุติยภูมิ ซึ่งเก็บข้อมูลจากนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 ในโรงเรียนขยายโอกาสทางการศึกษาทุกโรงเรียนในสังกัดสำนักงานการประถมศึกษา จังหวัดสมุทรสงคราม จำนวน 15 โรงเรียน จำนวน 469 คน แต่ผู้วิจัยได้ข้อมูลมาเพียง 13 โรงเรียน จำนวน 433 คน

เนื่องจากในการหาคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ โมเดลการวิเคราะห์จะต้องใช้คำตอบทุกข้อมาวิเคราะห์ ดังนั้นผู้วิจัยจึงตัดนักเรียนที่สอบไม่ครบทั้ง 3 ครั้งออก สำหรับข้อมูลชุดที่ 1 ที่เป็นข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม และข้อมูลชุดที่ 2 ที่เป็นข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน และตัดนักเรียนที่สอบไม่ครบทั้ง 5 ครั้งออกในข้อมูลชุดที่ 3 ที่เป็นข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน

ข้อมูลชุดที่ 1 ชุดวัดซ้ำด้วยแบบทดสอบเดิม เก็บข้อมูลมาทั้งหมด 698 คน ได้มีการตัดข้อมูลนักเรียนที่สอบไม่ครบทั้ง 3 ครั้งออก เหลือนักเรียนจำนวน 635 คน (ร้อยละ 90.97) โดยนักเรียนที่ถูกตัดออกอยู่ในกลุ่มสูง (สูงกว่า 1 SD) ร้อยละ 4.84 อยู่ในกลุ่มปานกลาง (- 1 SD ถึง 1 SD) ร้อยละ 74.19 และอยู่ในกลุ่มต่ำ (ต่ำกว่า -1 SD) ร้อยละ 20.97 โดยในข้อมูลชุดนี้มีค่าเฉลี่ยของผลการสอบ 3 ครั้งเท่ากับ 27.56 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 11.16 จากคะแนนเต็ม 66 คะแนน

ข้อมูลชุดที่ 2 ชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน เก็บข้อมูลมาทั้งหมด 637 คน ได้มีการตัดข้อมูลนักเรียนที่สอบไม่ครบทั้ง 3 ครั้งออก เหลือนักเรียนจำนวน 581 คน (ร้อยละ 91.21) โดยนักเรียนที่ถูกตัดออกอยู่ในกลุ่มสูง (สูงกว่า 1 SD) ร้อยละ 14.81 อยู่ในกลุ่มปานกลาง (- 1 SD ถึง 1 SD) ร้อยละ 85.19 ส่วนในกลุ่มต่ำ (ต่ำกว่า -1 SD) นั้นไม่มีการตัดออกเลย ซึ่งนักเรียนที่ถูกตัดออกในกลุ่มสูงและปานกลางมีจำนวนมากเนื่องจากว่ามีนักเรียนในโรงเรียนหนึ่งได้ขาดสอบในครั้งที่ 2 เป็นจำนวนมากซึ่งเป็นนักเรียนที่มีความรับผิดชอบสูงทางโรงเรียนจึงให้ไปทำกิจกรรมของโรงเรียนซึ่งตรงกับเวลาที่สอบและนักเรียนในกลุ่มนี้เป็นนักเรียนที่มีความสามารถทางการเรียนอยู่ในขั้นดีจึงทำให้จำนวนนักเรียนที่ถูกคัดออกเป็นนักเรียนในกลุ่มสูงและปานกลางมีจำนวนมาก ข้อมูลชุดนี้มีค่าเฉลี่ยของผลการสอบ 3 ครั้งเท่ากับ 29.57 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 13.84 จากคะแนนเต็ม 66 คะแนน

ข้อมูลชุดที่3 ชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน เก็บข้อมูลมาทั้งหมด 433 คน ได้มีการตัดข้อมูลนักเรียนที่สอบไม่ครบทั้ง 5 ครั้งออก เหลือนักเรียนจำนวน 386 คน (ร้อยละ 89.15) โดยนักเรียนที่ถูกตัดออกอยู่ในกลุ่มสูง (สูงกว่า 1 SD) ร้อยละ 17.02 อยู่ในกลุ่มปานกลาง (- 1 SD ถึง 1 SD) ร้อยละ 72.34 และอยู่ในกลุ่มต่ำ (ต่ำกว่า -1 SD) ร้อยละ 10.64 ข้อมูลชุดนี้มีค่าเฉลี่ยของผลการสอบ 3 ครั้งเท่ากับ 8.73 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.41 จากคะแนนเต็ม 25 คะแนน (รายละเอียดดังตารางที่ 3.12)

ตารางที่ 3.12 จำนวนนักเรียนที่เป็นกลุ่มตัวอย่างจำแนกตามชุดของข้อมูล

รายชื่อโรงเรียน	จำนวนนักเรียนที่สอบ	จำนวนนักเรียนที่สอบครบทั้ง 3 ครั้ง	รายชื่อโรงเรียน	จำนวนนักเรียนที่สอบ	จำนวนนักเรียนที่สอบครบทั้ง 5 ครั้ง
ข้อมูลชุดที่ 1			**ข้อมูลชุดที่** 3		
สหวิทยาเขตราชนครินทร์			1. วัดช่องลม	22	16
1. วัดสุทธิวราราม	118	114	2. วัดลาดแป้ง	53	52
2. ยานนาเวศวิทยาคม	130	119	3. วัดดาวโด่ง	30	25
3. นนทรีวิทยา	127	119	4. วัดบางน้อย	38	35
สหวิทยาเขตเพชรเกษม			5. วัดปรกสุธรรมาราม	51	48
4. วัดรางบัว	87	81	6. วัดสาธุชนาราม	32	26
5. ราชวินิตบางแคปานขำ	140	116	7. วัดโคกเกตุ	39	34
6. ปัญญาวรคุณ	96	86	8. บ้านคลองบางกก	28	24
รวม	698	635(90.97%)	9. บ้านคลองสมบูรณ์	16	15
ข้อมูลชุดที่ 2			10. ไทยรัฐวิทยา(70)	19	17
สหวิทยาเขตกรุงเกษม			11. วัดปากสมุทร	38	34
1. เทพศิรินทร์	92	83	12. ศรัทธาธรรม	34	28
2. วัดสระเกศ	77	75	13. บ้านลาดใหญ่	33	30
3.ไตรมิตรวิทยาลัย	130	120			
สหวิทยาเขตปากน้ำภาษีเจริญ					
4. ประดู่ในทรงธรรม	119	109			
5. วัดนวลนรดิศ	126	123			
6. จันทร์ประดิษฐารามวิทยาคม	93	71			
รวม	637	581(91.21%)	**รวม**	433	386(89.15%)

ตอนที่ 5 **การเก็บรวบรวมข้อมูล**

ข้อมูลชุดที่ 1 ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) เป็นข้อมูลที่ได้จากการทดสอบซ้ำ 3 ครั้ง แต่ละครั้งห่างกัน 4 สัปดาห์ โดยใช้แบบทดสอบฉบับเดิม

ข้อมูลชุดที่ 2 ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) เป็นข้อมูลที่ได้จากการทดสอบซ้ำ 3 ครั้ง แต่ละครั้งห่างกัน 4 สัปดาห์ โดยใช้แบบทดสอบต่างฉบับกันจำนวน 3 ฉบับที่คู่ขนานกัน ซึ่งแบบทดสอบแต่ละฉบับจะใช้ทดสอบทั้ง 3 ครั้งในกลุ่มตัวอย่างที่ต่างกลุ่มกัน และกลุ่มตัวอย่างแต่ละกลุ่มจะต้องถูกทดสอบด้วยแบบทดสอบทั้ง 3 ฉบับโดยทดสอบครั้งละฉบับในลักษณะของการสอบที่มีความสมดุลย์กันของแบบทดสอบ (counterbalance) โดยการจัดรูปการทดสอบแบบจตุรัสละติน (latin square) เพื่อให้เป็นไปตามข้อตกลงเบื้องต้นในเรื่องของความเป็นอิสระต่อกันในข้อสอบแต่ละข้อ (local independence) (Embretson, 1995: 283) ตามแนวคิดในการวัดคะแนนพัฒนาการของทฤษฎีการตอบสนองข้อสอบ และเพื่อไม่ให้มีอิทธิต่างๆมาทำให้เกิดความคลาดเคลื่อนจากการวัดซ้ำด้วยแบบทดสอบฉบับเดิมได้ เช่นอิทธิพลจากการจำได้ อิทธิพลจาก

การคงที่ของการตอบ และอิทธิพลจากการมีทักษะที่ได้จากการทำข้อสอบซ้ำ (Anastasi, 1988 cited in Embretson, 1991: 499)

ข้อมูลชุดที่ 3 ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน (equivalence) เป็นข้อมูลที่ได้จากการทดสอบซ้ำ 5 ครั้ง แต่ละครั้งห่างกัน 3 สัปดาห์ โดยใช้แบบทดสอบคู่ขนานกันจำนวน 5 ฉบับ ทดสอบครั้งละ 1 ฉบับ เพื่อป้องกันการจำข้อสอบได้ ตามแนวคิดในการวัดคะแนนพัฒนาการแบบดั้งเดิม

การเก็บข้อมูลชุดที่ 1 และ ข้อมูลชุดที่ 2 กำหนดเก็บข้อมูล 3 ครั้ง ห่างกันครั้งละ 1 เดือน คือ

ครั้งที่ 1 สัปดาห์ที่ 3 - 4 ของเดือน มิถุนายน 2543

ครั้งที่ 2 สัปดาห์ที่ 3 - 4 ของเดือน กรกฎาคม 2543

ครั้งที่ 3 สัปดาห์ที่ 3 - 4 ของเดือน สิงหาคม 2543

การเก็บข้อมูลชุดที่ 3 กำหนดเก็บข้อมูล 5 ครั้ง ห่างกันครั้งละ 3 สัปดาห์ คือ

ครั้งที่ 1 สัปดาห์ที่ 3 ของเดือนมิถุนายน 2542

ครั้งที่ 2 สัปดาห์ที่ 2 ของเดือนกรกฎาคม 2542

ครั้งที่ 3 สัปดาห์ที่ 1 ของเดือนสิงหาคม 2542

ครั้งที่ 4 สัปดาห์ที่ 4 ของเดือนสิงหาคม 2542

ครั้งที่ 5 สัปดาห์ที่ 3 ของเดือนกันยายน 2542

ตอนที่ 6 การวิเคราะห์ข้อมูล

ผู้วิจัยนำข้อมูลชุดวัดซ้ำด้วยแบบทดสอบเดิม ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน และข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน มาตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) แล้วดำเนินการวิเคราะห์ข้อมูลทุกชุด ดังนี้

ตอนที่ 1 การวิเคราะห์ค่าสถิติเบื้องต้น

1.1 การวิเคราะห์ค่าสถิติพื้นฐานของข้อมูล เพื่อตรวจสอบลักษณะของข้อมูลทั่วไป ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด ค่าต่ำสุด ค่าความโด่งและค่าความเบ้ของคะแนนที่ได้จากการวัดทุกครั้ง โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS / PC+

1.2 การวิเคราะห์ความแตกต่างของค่าเฉลี่ยของคะแนนที่ได้จากการวัดระหว่างช่วงเวลาของข้อมูลแต่ละชุด ด้วยวิธีการวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (repeated measures anlysis of variance) (Daniel, 1995: 316) เพื่อตรวจสอบความแตกต่างของค่าเฉลี่ยของคะแนนที่ได้จากการวัดระหว่างช่วงเวลาของการวัดข้อมูลแต่ละชุด ซึ่งในการวิเคราะห์ความแปรปรวนแบบวัดซ้ำนี้

ต้องมีการตรวจสอบเงื่อนไข sphericity ของคะแนนที่ได้จากการวัดระหว่างช่วงเวลาของข้อมูลแต่ละชุด โดยใช้การทดสอบของ Mauchly

การทดสอบเงื่อนไข sphericity เป็นการตรวจสอบว่าเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของคะแนนแต่ละชุดเมื่อแปลงเป็นเมทริกซ์ของตัวแปรใหม่ที่เป็นตัวแปรของการวัดซ้ำจะมีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมเป็นเมทริกซ์เอกลักษณ์หรือไม่ ถ้าไม่เป็นเมทริกซ์เอกลักษณ์จะต้องมีการปรับค่า degrees of freedom ซึ่งมีผลกระทบต่อการทดสอบ F test ในการทดสอบความแปรปรวน ซึ่งมีรายละเอียดดังนี้

ทดสอบสมมติฐานความแปรปรวนของการวัดซ้ำโดยทั่วไปจะต้องมีการแปลงการวัดซ้ำ k ครั้งนั้นให้เป็นตัวแปรใหม่ (k - 1) ตัว โดยมีเงื่อนไขของการเป็น sphericity กล่าวคือ เมทริกซ์ของความแปรปรวนร่วมของตัวแปรใหม่ต้องมีลักษณะเป็น diagonal matrix ที่มีความแปรปรวนเท่ากัน และมีความแปรปรวนร่วมเป็น 0 โดยให้เมทริกซ์เดิมเป็น $\sum$ และเมทริกซ์ใหม่เป็น C

$$C'\sum C = \sigma^2 I = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ . \\ . \\ k-1 \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & 2 & 3 & . \; . & k-1 \end{array} \\ \begin{pmatrix} \sigma^2 & 0 & 0 & . \; . & 0 \\ 0 & \sigma^2 & 0 & . \; . & 0 \\ 0 & 0 & \sigma^2 & . \; . & 0 \\ . & . & . & . \; . & . \\ . & . & . & . \; . & . \\ 0 & 0 & 0 & . \; . & \sigma^2 \end{pmatrix} \end{array}$$

ในการทดสอบเงื่อนไข sphericity นี้ใช้สถิติทดสอบของ Mauchly ถ้าผลการทดสอบด้วยค่าไคร์-สแควร์พบว่าเงื่อนไขของ sphericity ยังเป็นจริงดังนั้น ค่า $\hat{\varepsilon}$ มีค่าเท่ากับ 1 แต่ถ้าเงื่อนไขของ sphericity ไม่เป็นจริง $\hat{\varepsilon}$ จะมีค่าต่างออกไป (Stevens: 1996) ดังนี้

Huynh and Feldt (1976) เสนอดังนี้ $\tilde{\varepsilon} = \dfrac{n(k-1)\hat{\varepsilon} - 2}{(k-1)[(n-1) - (k-1)\hat{\varepsilon}]}$

ซึ่งค่า $\hat{\varepsilon}$ มีผลต่อค่า degrees of freedom ของการทดสอบ F test คือ

$$\nu_1 = (k-1)\hat{\varepsilon}$$

$$\nu_2 = (n-1)(k-1)\hat{\varepsilon}$$

ส่วน Greenhouse and Geisser (1959) เสนอค่า degrees of freedom ดังนี้

$(k-1)$ ให้เป็น $[1/(k-1)](k-1) = 1$

$(n-1)(k-1)$ $[1/(K-1)](n-1)(k-1) = n-1$

ตอนที่ 2 การวิเคราะห์เพื่อประมาณค่าคะแนนพัฒนาการของผลสัมฤทธิ์ทางคณิตศาสตร์

วิเคราะห์เพื่อสร้างและตรวจสอบโมเดลโค้งพัฒนาการตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์โดยตรวจสอบประสิทธิภาพของโมเดล หรือตรวจสอบความสอดคล้องกลมกลืนระหว่างโมเดลกับข้อมูล โดยใช้โปรแกรม LISREL 8.3 ซึ่งจะทำการทดสอบ 5 รูปแบบคือ (นงลักษณ์ วิรัชชัย, 2542: 280-281; อิทธิ์พงษ์ ตั้งสกุลเรืองไล, 2541)

1) โมเดลพัฒนาการที่ไม่มีความชัน (no slope baseline growth model: NSB)

2) โมเดลพัฒนาการเชิงเส้น (linear growth model: LIN)

3) โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์คงที่ (latent growth curve model with fixed parameter: FIX)

4) โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์อิสระ (latent growth curve model with free parameter: FRE)

5) โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์อิสระและมีความแปรปรวนของความคลาดเคลื่อนไม่เท่ากัน (latent growth curve model with free parameter and unequal disturbance variance: UDV)

ในการตรวจสอบประสิทธิภาพของโมเดลโค้งพัฒนาการตัวแปรแฝงในด้านความสอดคล้องกลมกลืนของข้อมูลกับโมเดลนั้นจะใช้ดัชนีในการตรวจสอบดังนี้

1) ค่าไค-สแควร์ (chi - squares statistics) ถ้าค่าไค-สแควร์มีค่ายิ่งใกล้ศูนย์มากเท่าไร แสดงว่าโมเดลสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ Saris และ Stronkhorst (1984 อ้างถึงใน นงลักษณ์ วิรัชชัย, 2542) เสนอว่าค่าไค-สแควร์ควรมีค่าเท่ากับองศาอิสระเมื่อโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

2) ดัชนีวัดความกลมกลืน (goodness of fit index: GFI) มีค่าเข้าใกล้ 1.00 แสดงว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

3) ดัชนีรากที่สองของค่าเฉลี่ยกำลังสองของเศษ (root mean squared residual: RMR) มีค่ายิ่งใกล้ศูนย์ แสดงว่าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

4) ค่าความคลาดเคลื่อนที่เป็นคะแนนมาตรฐานสูงสุด (largest standardized residual: LSR) ถ้าโมเดลมีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ LSR มีค่าไม่ควรเกิน 2.00

2.2 ประมาณค่าอัตราพัฒนาการของผู้สอบแต่ละคนเพื่อใช้เป็นเกณฑ์ซึ่งมีจำนวนครั้งของการวัดแตกต่างกัน โดยแยกตามชุดของข้อมูลและแบบการให้คะแนน ดังนี้

ตารางที่ 3.13 จำนวนครั้งในการวัดที่นำมาประมาณค่าอัตราพัฒนาการของผู้สอบแต่ละคนเพื่อใช้เป็นเกณฑ์ โดยจำแนกตามชุดของข้อมูลและแบบการให้คะแนน

วิธีการให้คะแนน	ชุดของข้อมูล		
	ชุดที่ 1	ชุดที่ 2	ชุดที่ 3
dichotomous	วัด 3 ครั้ง	วัด 3 ครั้ง	วัด 3 และ 5 ครั้ง
partial credit	วัด 3 ครั้ง	วัด 3 ครั้ง	วัด 3 และ 5 ครั้ง

ตอนที่ 3 การวิเคราะห์ประมาณค่าคะแนนพัฒนาการรายบุคคล

การวิเคราะห์ประมาณค่าคะแนนพัฒนาการรายบุคคลที่ได้จากวิธีการวัดคะแนนพัฒนาการแต่ละวิธีใน 3 กลุ่มวิธี ได้แก่กลุ่มวิธีตามทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีตามทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีตามทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา รวมทั้งคะแนนพัฒนาการจากวิธีเกณฑ์ มีรายละเอียดดังนี้

3.1 กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม

วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม มีทั้งหมด 5 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ 2) วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน 3) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ 4) วิธีการวัดคะแนนพัฒนาการสัมพันธ์ 5) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้สอบ ซึ่งทั้ง 5 วิธีผู้วิจัยใช้การประมาณค่าคะแนนพัฒนาการด้วยโปรแกรม SPSS for WIN 10.07

3.2 กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ

วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบ มีทั้งหมด 2 วิธี ได้แก่

1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริงสำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ประมาณค่าพารามิเตอร์ของความสามารถที่แท้จริงแบบ 3 พารามิเตอร์ ใช้โปรแกรม BILOG 3 ของ Mislevy และ Bock (1990) และข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ประมาณค่าพารามิเตอร์ของความสามารถที่แท้จริง โดยใช้โปรแกรม PARSCALE 2 ของ Muraki และ Bock (1993)

2) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (Multidimensional Rasch Model for Learning and Change: MRMLC) ประมาณค่าคะแนนพัฒนาการด้วยโปรแกรม LPCM – WIN 1.0 ของ Fischer และ Ponocny – Seliger (1998)

3.3 กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา

วิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีทั้งหมด 2 วิธีได้แก่

1) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ประมาณค่าพารามิเตอร์ของความสามารถที่แท้จริงแบบ 3 พารามิเตอร์ โดยใช้โปรแกรม BILOG 3 และข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ประมาณค่าพารามิเตอร์ของความสามารถที่แท้จริง โดยใช้โปรแกรม PARSCALE 2

2) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ประมาณค่าพารามิเตอร์ของความสามารถที่แท้จริงแบบ 3 พารามิเตอร์ โดยใช้โปรแกรม BILOG 3 และข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ประมาณค่าพารามิเตอร์ของความสามารถที่แท้จริง โดยใช้โปรแกรม PARSCALE 2

ตอนที่ 4 การวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการ

คุณภาพของวิธีการวัดคะแนนพัฒนาการจะพิจารณาจากค่าความตรงตามเกณฑ์สัมพันธ์ และค่าความคลาดเคลื่อนของคะแนนพัฒนาการ โดยมีรายละเอียดการวิเคราะห์ดังนี้

4.1 การวิเคราะห์ค่าความตรงตามเกณฑ์สัมพันธ์

ในการวิจัยครั้งนี้วิเคราะห์ค่าความตรงตามเกณฑ์สัมพันธ์จากค่าความสัมพันธ์ 2 แบบ คือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการที่ได้จากวิธีการวัดแต่ละวิธีกับคะแนนพัฒนาการที่เป็นเกณฑ์ และค่าสหสัมพันธ์อันดับระหว่างอันดับของคะแนนพัฒนาการที่ได้จากวิธีการวัดแต่ละวิธีกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ โดยมีรายละเอียดดังนี้

4.1.1 การวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์มีขั้นตอนดังนี้

1) หาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการของแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนากับคะแนนพัฒนาการที่ประมาณค่าจากโมเดลโค้งพัฒนาการที่มีตัวแปรแฝงเกณฑ์ โดยใช้สูตร Pearson's Product Moment

2) วิเคราะห์ความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการระหว่างกลุ่มวิธี 3 กลุ่มวิธี คือกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยผู้วิจัยอาศัยแนวคิดในด้านของการวิเคราะห์ของโมเดลเชิงเส้นตรงระดับลดหลั่น (Hierarchical Linear Models: HLM) 2 ระดับ ที่ใช้โปรแกรมสำเร็จรูป HLM for win 5.04 ของ Bryk, Raudenbush และ Congdon

เพื่อตรวจสอบดูว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการมีความผันแปรระหว่างกลุ่มวิธีวัดหรือไม่ ซึ่งในการวิเคราะห์ด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น 2 ระดับ ครั้งนี้ถึงแม้จะมีขนาดกลุ่มตัวอย่างขนาดเล็ก แต่จากตัวอย่างการประยุกต์ใช้ HLM ในการวิเคราะห์อภิมาน (meta - analysis) ของ Bryk และ Raudenbush (1992) ซึ่งใช้ขนาดตัวอย่างเพียง 19 ตัวอย่างก็สามารถวิเคราะห์ข้อมูลได้ ดังนั้นผู้วิจัยจึงประยุกต์ใช้ HLM 2 ระดับในการตรวจสอบความผันแปรของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการระหว่างกลุ่มวิธีวัดโดยมีขั้นตอนการวิเคราะห์ดังนี้ (Raudenbush, 1992; ศิริชัย กาญจนวาสี, 2540 และ อิทธิฤทธิ์ พงษ์ปิยะรัตน์, 2542)

การวิเคราะห์โมเดลศูนย์ (null model)

เป็นการวิเคราะห์เพื่อให้เห็นภาพรวมของค่าสัมประสิทธิ์สหสัมพันธ์ โดยไม่มีการนำตัวแปรอิสระอื่นๆเข้ามาร่วมในการวิเคราะห์ เพื่อประมาณค่าพารามิเตอร์ของค่าสัมประสิทธิ์สหสัมพันธ์ และตรวจสอบว่ามีความแปรปรวนระหว่างหน่วย(กลุ่มวิธีการวัด)หรือไม่ ซึ่งมีการวิเคราะห์ดังนี้

1. การวิเคราะห์ภายในหน่วย

$$Y_{ij} = b_{01} + e_{ij}$$

2. การวิเคราะห์ระหว่างหน่วย

$$b_{01} = r_{00} + U_{0j}$$

โดย Y_{ij} คือค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการ

b_{01} คือค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการในแต่ละกลุ่มวิธี

r_{00} คือค่าเฉลี่ยรวมของสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการทุกกลุ่มวิธี

e_{ij} คือค่าความแปรปรวนส่วนที่เหลือในการวิเคราะห์ระดับภายในแต่ละวิธีการวัด

U_{0j} คือค่าความแปรปรวนส่วนที่เหลือในการวิเคราะห์ระดับระหว่างกลุ่มวิธีการวัด

การวิเคราะห์โมเดล HLM จะแบ่งค่าพารามิเตอร์ออกเป็นอิทธิพลคงที่ (fixed effect) และอิทธิพลสุ่ม (random effect) ซึ่งจะมีการทดสอบอิทธิพลคงที่ ($H_0 : r_{00} = 0$) โดยใช้สถิติทดสอบที (t-test) ถ้าผลการทดสอบ พบว่า มีนัยสำคัญทางสถิติแสดงว่าค่าเฉลี่ยรวมของสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการมีค่าไม่เท่ากับศูนย์ ส่วนการทดสอบอิทธิพลสุ่ม ($H_0 : Var(b_{01}) = 0$; $H_0 : Var(U_{0j}) = 0$) โดยใช้สถิติทดสอบไค-สแควร์ ถ้าผลการทดสอบพบว่ามีนัยสำคัญทางสถิติแสดงว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการในแต่ละกลุ่มวิธีมีความแตกต่าง (มีความแปรปรวน) ถ้าผลการทดสอบไม่มีนัยสำคัญทางสถิติแสดงค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการในแต่ละกลุ่มวิธีไม่แตกต่างกัน

ถ้าผลการทดสอบพบว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการมีความผันแปรระหว่างกลุ่มวิธี 3 กลุ่มวิธีในการวัดคะแนนพัฒนาการ จะทำการทดสอบรายคู่ของกลุ่มวิธีเพื่อดูว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการระหว่างกลุ่มวิธี 2 กลุ่มใดที่มีความผันแปรระหว่างกลุ่มวิธีในการวัดพัฒนาการ โดยใช้การวิเคราะห์ของโมเดลเชิงเส้นตรงระดับลดหลั่น (Hierarchical Linear Models: HLM) 2 ระดับ ที่ใช้โปรแกรมสำเร็จรูป HLM for win 5.04 ของ Bryk, Raudenbush และ Congdon

3) วิเคราะห์ความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการภายในแต่ละกลุ่มวิธีทั้ง 3 กลุ่มวิธี คือในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยทำการตรวจสอบค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของวิธีวัดพัฒนาการเป็นรายคู่ภายในแต่ละกลุ่มวิธีเพื่อดูว่าค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการคู่ใดบ้างที่มีความแตกต่างกันโดยใช้สูตรการทดสอบของ Hotelling (1940 cited in Lindeman, Merenda and Gold, 1980) ดังนี้

$$t = (r_{XZ} - r_{YZ})\sqrt{\frac{(n-3)(1+r_{XY})}{2(1-r_{XY}^2-r_{XZ}^2-r_{YZ}^2+2r_{XY}r_{XZ}r_{YZ})}}$$

โดยที่ n คือขนาดของกลุ่มตัวอย่าง

r_{XY} คือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y

r_{XZ} คือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร X และตัวแปร Z

r_{YZ} คือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร Y และตัวแปร Z

และมี df = n - 3

4.1.2 การวิเคราะห์ค่าสหสัมพันธ์อันดับ มีขั้นตอนดังนี้

1) หาค่าสหสัมพันธ์อันดับระหว่างอันดับของคะแนนพัฒนาการของแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา กับอันดับของคะแนนพัฒนาการที่ได้จากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง โดยใช้สูตร Spearman's Rho

2) วิเคราะห์ความแตกต่างของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการระหว่างกลุ่มวิธี 3 กลุ่มวิธี คือกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีที่ผู้วิจัยพัฒนา โดยผู้วิจัยอาศัยแนวคิดในด้านของการวิเคราะห์ของโมเดลเชิงเส้นตรงระดับลดหลั่น (Hierarchical Linear Models: HLM) 2 ระดับ ที่ใช้โปรแกรมสำเร็จรูป HLM for win 5.04 ของ Bryk, Raudenbush และ Congdon เพื่อตรวจสอบดูว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการมีความผันแปรระหว่างกลุ่มวิธีวัดหรือไม่

ถ้าผลการทดสอบพบว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการมีความผันแปรระหว่างกลุ่มวิธี 3 กลุ่มวิธีในการวัดพัฒนาการ จะทำการทดสอบรายคู่ของกลุ่มวิธีเพื่อดูว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการระหว่างกลุ่มวิธี 2 กลุ่มใดที่มีความผันแปรระหว่างกลุ่มวิธีในการวัดพัฒนาการ โดยใช้การวิเคราะห์ของโมเดลเชิงเส้นตรงระดับลดหลั่น (Hierarchical Linear Models: HLM) 2 ระดับ ที่ใช้โปรแกรมสำเร็จรูป HLM for win 5.04 ของ Bryk, Raudenbush และ Congdon

3) วิเคราะห์ความแตกต่างของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการภายในแต่ละกลุ่มวิธีทั้ง 3 กลุ่มวิธี คือในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยทำการตรวจสอบค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีวัดคะแนนพัฒนาการเป็นรายคู่ภายในแต่ละกลุ่มวิธีเพื่อดูว่าค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการคู่ใดบ้างที่มีความแตกต่างกันโดยใช้สูตรการทดสอบของ Hotelling (1940 cited in Lindeman, Merenda and Gold, 1980)

4.2 การวิเคราะห์ค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการ

ในการวิจัยครั้งนี้วิเคราะห์ค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการของแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยมีขั้นตอนดังนี้

4.2.1 แปลงคะแนนพัฒนาการที่ได้จากวิธีการวัดคะแนนพัฒนาการของแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา และคะแนนพัฒนาการที่ได้จากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่เป็นวิธีเกณฑ์ให้เป็นคะแนนมาตรฐาน (Z- score)

4.2.2 วิเคราะห์ค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการของแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และในกลุ่มวิธีที่ผู้วิจัยที่พัฒนา โดยวิเคราะห์จากค่าสถิติ 2 ค่าคือ

4.2.2.1 ค่าความลำเอียง (bias) หาจากค่าเฉลี่ยของความแตกต่าง (average difference) ระหว่างคะแนนพัฒนาการที่ได้จากวิธีการวัดแต่ละวิธีกับคะแนนพัฒนาการที่ได้จากวิธีที่เป็นเกณฑ์ โดยมีสูตรดังนี้

$$\text{bias} = \frac{\sum_{k}^{n}(\hat{\theta}_k - \theta_k)}{n}$$

4.2.2.2 ค่ารากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (RMSE: root mean squared error) โดยมีสูตรดังนี้

$$RMSE = \left[\frac{\sum_{k}^{n} (\hat{\theta}_k - \theta_k)^2}{n} \right]^{\frac{1}{2}}$$

$\hat{\theta}_k$ = คะแนนพัฒนาการ/คะแนนพัฒนาการความสามารถที่แท้จริงที่แปลงเป็นคะแนนมาตรฐานแล้วของคนที่ k

θ_k = อัตราพัฒนาการที่หาได้จากโมเดลโค้งพัฒนาการตัวแปรแฝงที่แปลงเป็นคะแนนมาตรฐานแล้วของคนที่ k

ตอนที่ 5 การเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ

ในการเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ จะทำการเปรียบเทียบระหว่างกลุ่มวิธีการวัด และภายในกลุ่มวิธีทั้ง 3 กลุ่มวิธีการวัด ด้วยการวิเคราะห์แผนภาพการกระจายของค่าความสัมประสิทธิ์สหสัมพันธ์และค่ารากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน (RMSE: root mean squared error)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ผู้วิจัยขอนำเสนอผลการวิเคราะห์ข้อมูลออกเป็น 5 ตอน ตอนแรก ผลการวิเคราะห์ค่าสถิติเบื้องต้น ได้แก่ ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลทั้ง 3 ชุด คือ 1) ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคณิตศาสตร์ฉบับเดิม (repeat) 2) ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) และ 3) ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบที่คู่ขนานกัน (equivalence) การวิเคราะห์ความแตกต่างของค่าเฉลี่ยของคะแนนที่ได้จากการวัดระหว่างช่วงเวลา ตอนที่ 2 ผลการตรวจสอบความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง 5 รูปแบบ เพื่อนำโมเดลที่สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์มากที่สุดมาประมาณค่าคะแนนพัฒนาการเพื่อใช้เป็นคะแนนเกณฑ์ ตอนที่ 3 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลของแต่ละกลุ่มวิธี ได้แก่ กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาขึ้น ตอนที่ 4 ผลการวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการ ด้านความตรงตามเกณฑ์สัมพันธ์ และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธี ตอนที่ 5 ผลการเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ

ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อภาษาอังกฤษที่ใช้สื่อความหมายแทนชุดข้อมูล แทนโมเดล แทนตัวแปรและค่าสถิติดังนี้

สัญลักษณ์แทนชุดข้อมูล

repeat	หมายถึง คะแนนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 ที่ได้จากข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม โดยทำการวัดซ้ำ 3 ครั้ง
counterbalance	หมายถึง คะแนนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 ที่ได้จากข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน โดยทำการวัดซ้ำ 3 ครั้ง
equivalence 1	หมายถึง คะแนนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 2 ที่ได้จากข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน
equivalence 2	หมายถึง คะแนนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 2 ที่ได้จากข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน

สัญลักษณ์แทนวิธีการตรวจให้คะแนน

dichotomous	หมายถึง คะแนนที่ได้จากวิธีการตรวจให้คะแนนแบบทวิวิภาค โดยมีคะแนนเป็น 1 คะแนน เมื่อตอบถูก และเป็น 0 คะแนน เมื่อตอบผิด
partial credit	หมายถึง คะแนนที่ได้จากวิธีการตรวจให้คะแนนความรู้บางส่วน โดยมีคะแนนตั้งแต่ 0 ถึง 3 คะแนน สำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) และข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน(counterbalance) และมีคะแนนตั้งแต่ 0 ถึง 2 คะแนนสำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน (equivalence)

สัญลักษณ์แทนโมเดล

NSB	หมายถึง โมเดลพัฒนาการที่ไม่มีความชัน (no slope baseline growth model)
LIN	หมายถึง โมเดลพัฒนาการเชิงเส้น (linear growth model)
FIX	หมายถึง โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์คงที่ (latent growth curve model with fixed parameter)
FRE	หมายถึง โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดค่าพารามิเตอร์อิสระ (latent growth curve model with free parameter)
UDV	หมายถึง โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์อิสระ และมีความแปรปรวนของความคลาดเคลื่อนไม่เท่ากัน (latent growth curve model with free parameter and unequal disturbance variance)

สัญลักษณ์แทนค่าสถิติ

GFI	หมายถึง ดัชนีวัดความกลมกลืนที่ใช้วัดความสอดคล้องระหว่างโมเดลกับข้อมูล
RMR	หมายถึง ดัชนีรากที่สองของค่าเฉลี่ยกำลังสองของเศษที่ใช้วัดความสอดคล้องระหว่างโมเดลกับข้อมูล
LSR	หมายถึง ค่าความคลาดเคลื่อนที่เป็นคะแนนมาตรฐานสูงสุดที่ใช้วัดความสอดคล้องระหว่างโมเดลกับข้อมูล

สัญลักษณ์แทนตัวแปร

dscore 1 - dscore 5	หมายถึงคะแนนที่ได้จากการสอบครั้งที่ 1-5 ที่ตรวจให้คะแนนแบบทวิวิภาค (dichotomous)
pscore 1- pscore 5	หมายถึง คะแนนที่ได้จากการสอบครั้งที่ 1-5 ที่ตรวจให้คะแนนความรู้บางส่วน (partial credit)
GROWTH	หมายถึง อัตราพัฒนาการรายบุคคลที่ได้จากการประมาณค่าด้วยโมเดลพัฒนาการที่มีตัวแปรแฝงที่มีความสอดคล้องกับข้อมูลเชิงประจักษ์
DS	หมายถึงวิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ
SC	หมายถึงวิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน
NL	หมายถึงวิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ
SR	หมายถึงวิธีการวัดคะแนนพัฒนาการสัมพัทธ์
SRR	หมายถึงวิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพของผู้ตอบ
Lθ	หมายถึงวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง
MRMLC	หมายถึงวิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลตามแนวคิดของ Embretson
Sθ	หมายถึงวิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา
Vθ	หมายถึงวิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน
GS	หมายถึงวิธีการวัดคะแนนพัฒนาการตามโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 1 ผลการวิเคราะห์คะแนนที่ได้จากการวัดเบื้องต้น

การวิเคราะห์คะแนนที่ได้จากการวัดเบื้องต้น ผู้วิจัยได้แบ่งเป็น 2 ตอน คือ ตอนแรกเป็นการวิเคราะห์คะแนนที่ได้จากการวัดด้วยค่าสถิติบรรยาย ตอนที่สองเป็นการทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนที่ได้จากการวัดระหว่างเวลา ซึ่งมีรายละเอียดดังนี้

1.1 ผลการวิเคราะห์คะแนนที่ได้จากการวัดด้วยค่าสถิติบรรยาย

การวิเคราะห์ข้อมูลในเบื้องต้นนี้เป็นการวิเคราะห์ด้วยค่าสถิติพื้นฐานเพื่อบรรยายลักษณะของคะแนนที่ได้จากการวัด ได้แก่ ค่าต่ำสุด (minimum) ค่าสูงสุด (maximum) ค่าเฉลี่ย (mean) ค่ามัธยฐาน (median) ค่าฐานนิยม (mode) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ค่าสัมประสิทธิ์ความแปรผัน (coefficient of variation) ความแปรปรวน (variance) ค่าความเบ้ (skewness) ค่าความโค้ง (kurtosis) โดยผู้วิจัยได้แยกออกเป็นข้อมูล 3 ชุด ซึ่งมีผลการวิเคราะห์ดังนี้

1.1.1. **ข้อมูลชุดที่ 1 ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (**repeat) ได้แก่ คะแนนที่ได้จากการวัดด้วยแบบทดสอบฉบับเดิม 3 ครั้ง จากกลุ่มตัวอย่าง 635 คน ซึ่งมีการตรวจให้คะแนน 2 แบบ คือแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) มีรายละเอียดของคะแนนดังนี้

1.1.1.1 ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ซึ่งมีคะแนนเต็ม 22 คะแนน คะแนนรวมต่ำที่สุดที่ได้จากการวัดครั้งที่ 1 และ 2 มีค่าเท่ากัน คือ 0 คะแนน ครั้งที่ 3 มีค่าต่ำสุดคือ 1 คะแนน ส่วนคะแนนรวมสูงสุดจากการวัดครั้งที่ 1 คือ 21 คะแนน คะแนนรวมสูงสุดจากการวัดครั้งที่ 2 และ 3 มีค่าเท่ากันคือ 22 คะแนน คะแนนเฉลี่ยจากการวัดครั้งที่ 3 มีค่าสูงสุด คือ 10.115 คะแนน ในขณะที่คะแนนเฉลี่ยจากการวัดครั้งที่ 1 และ 2 มีค่าใกล้เคียงกัน คือ 6.992 และ 7.980 คะแนน ตามลำดับ โดยมีร้อยละ 50 ของกลุ่มตัวอย่างซึ่งมีความสามารถอยู่ในระดับปานกลาง มีพิสัยของคะแนนแคบ และมีค่าใกล้เคียงกันในการวัดครั้งที่ 1 และ 2 (PR_{75} – PR_{25} อยู่ในช่วง 5-9 และ 5-10 คะแนน) สำหรับการวัดครั้งที่ 3 นั้น มีพิสัยของคะแนนกว้างกว่า และมีค่าสูงสุด สูงกว่าการวัดในครั้งที่ 1 และ 2 (PR_{75} – PR_{25} อยู่ในช่วง 6-13 คะแนน) คะแนนจากการวัดครั้งที่ 1 และ 2 มีค่ามัธยฐานเท่ากันคือ 7 คะแนน ในขณะที่การวัดครั้งที่ 3 มีค่าเท่ากับ 9 คะแนน สำหรับความแปรปรวนของคะแนนจากการวัดทั้ง 3 ครั้ง พบว่า การวัดครั้งที่ 3 มีความแปรปรวนสูงสุด (CV (%) = 52.912) รองลงมาได้แก่การวัดครั้งที่ 2 และการวัดครั้งที่ 1 ตามลำดับ (CV (%) = 52.005, 46.825)

เมื่อพิจารณารูปแบบโค้งการแจกแจงจากค่าความเบ้และความโด่งพบว่า คะแนนจากการวัดทั้ง 3 ครั้ง มีลักษณะการแจกแจงที่มีลักษณะคล้ายกันคือเบ้ขวา โดยคะแนนจากการวัดครั้งที่ 2 มีลักษณะการเบ้มากที่สุด รองลงมาคือคะแนนจากการวัดครั้งที่ 1 และครั้งที่ 3 ตามลำดับ (Sk = 0.995, 0.889 และ 0.697 ตามลำดับ) สำหรับลักษณะความโด่งของลักษณะการแจกแจงของคะแนนจากการวัดทั้ง 3 ครั้งนั้น ความโด่งของคะแนนจากการวัดครั้งที่ 1 และครั้งที่ 2 มีลักษณะโด่งกว่าโค้งปกติ โดยคะแนนจากการวัดครั้งที่ 1 โด่งมากกว่าการวัดครั้งที่ 2 (Ku = 1.184, 0.615) ส่วนคะแนนจากการวัดครั้งที่ 3 นั้น มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ (Ku = -0.412) แต่อย่างไรก็ตามคะแนนจากการวัดในแต่ละครั้งยังมีลักษณะของลักษณะการแจกแจงเป็นโค้งปกติ เมื่อพิจารณาจากค่าความคลาดเคลื่อนของความเบ้และความโด่งที่มีค่าเท่ากันทั้ง 3 ครั้ง เท่ากับ 0.097 และ 0.194 ตามลำดับ ซึ่งมีค่าไม่เกิน 2.00 หรือไม่น้อยกว่า – 2.00 ยังถือได้ว่ามีลักษณะการแจกแจงเป็นโค้งปกติ (SPSS Base 8.0, 1998) (รายละเอียดดังตารางที่ 4.1 และแผนภาพที่ 4.1)

1.1.1.2. ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีคะแนนเต็ม 66 คะแนน คะแนนรวมต่ำสุดจากการวัดครั้งที่ 1 คือ 3 คะแนน การวัดครั้งที่ 2 และครั้งที่ 3 มีคะแนนต่ำสุดเท่ากันคือ 8 คะแนน ส่วนคะแนนรวมสูงสุดจากการวัดครั้งที่ 1 มีค่าต่ำกว่าจากการวัดครั้งที่ 2 และครั้งที่ 3 คือได้ 65 คะแนน การวัดครั้งที่ 2 และครั้งที่ 3 มีคะแนนสูงสุดเท่ากันคือ 66 คะแนน คะแนนเฉลี่ยจากการวัดครั้งที่ 3 มีค่าสูงสุดคือ 32.857 คะแนน รองลงมาคือคะแนนจากการวัดครั้งที่ 2 และครั้งที่ 1 โดยมีคะแนนเท่ากับ 26.402 และ 23.027 คะแนน ตามลำดับ โดยที่ร้อยละ 50 ของกลุ่มตัวอย่างซึ่งมีความสามารถอยู่ในระดับปานกลางของคะแนนจากการวัดครั้งที่ 3 มีพิสัยกว้างที่สุด รองลงมาคือคะแนนจากการวัดครั้งที่ 2 และครั้งที่ 1 ตามลำดับ ($PR_{75} – PR_{25}$ อยู่ในช่วงคะแนน 20-44, 18-32 และ 17-28 ตามลำดับ) ซึ่งคะแนนจากการวัดครั้งที่ 3 มีค่ามัธยฐานสูงที่สุด คือ 29 คะแนน ในขณะที่คะแนนจากการวัดครั้งที่ 2 และครั้งที่ 1 มีค่ามัธยฐานใกล้เคียงกัน คือ 23, 21 คะแนน ตามลำดับ สำหรับความแปรปรวนของคะแนนจากการวัดครั้งที่ 3 มีค่าสูงที่สุด (CV (%) = 46.876) รองลงมาได้แก่คะแนนจากการวัดครั้งที่ 2 และครั้งที่ 1 ตามลำดับ (CV (%) = 45.762, 38.893)

เมื่อพิจารณารูปแบบโค้งการแจกแจงจากค่าความเบ้และความโด่งพบว่า คะแนนจากการวัดทั้ง 3 ครั้ง มีลักษณะการแจกแจงคล้ายคลึงกันคือมีลักษณะเบ้ขวา โดยคะแนนจากการวัดครั้งที่ 2 มีลักษณะเบ้มากที่สุด รองลงมาคือคะแนนจากการวัดครั้งที่ 1 และครั้งที่ 3 ตามลำดับ (Sk = 1.112, 1.075 และ 0.705) สำหรับลักษณะความโด่งของลักษณะการแจกแจงของคะแนนจากการวัดทั้ง 3 ครั้งนั้น ความโด่งของคะแนนจากการวัดครั้งที่ 1 และครั้งที่ 2 มีลักษณะโด่งกว่าโค้งปกติ โดยคะแนนจากการวัดครั้งที่ 1 โด่งมากกว่าการวัดครั้งที่ 2 (Ku = 1.810,

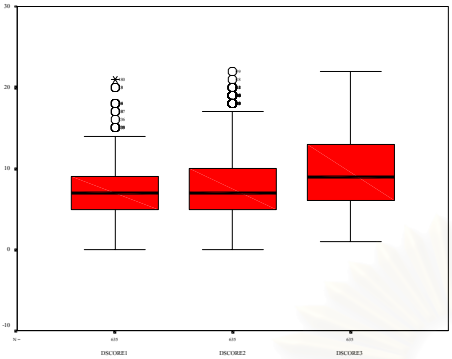
0.632) ส่วนคะแนนจากการวัดครั้งที่ 3 มีลักษณะการแจกแจงของคะแนนที่แบนกว่าโค้งปกติ (Ku = -0.579) คะแนนจากการวัดทั้ง 3 ครั้ง มีลักษณะของลักษณะการแจกแจงเป็นโค้งปกติ เพราะค่าความคลาดเคลื่อนของค่าความเบ้และค่าความโด่ง ซึ่งมีค่าเท่ากันทั้ง 3 ครั้งคือ 0.097 และ 0.194 ตามลำดับ ซึ่งมีค่าไม่เกิน 2.00 จึงยังถือว่าคะแนนจากการวัดในแต่ละครั้งมีลักษณะการแจกแจงเป็นโค้งปกติ (รายละเอียดดังตารางที่ 4.1 และแผนภาพที่ 4.2)

ตารางที่ 4.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่ได้จากการทดสอบ 3 ครั้งจำแนกตามวิธีการตรวจให้คะแนน

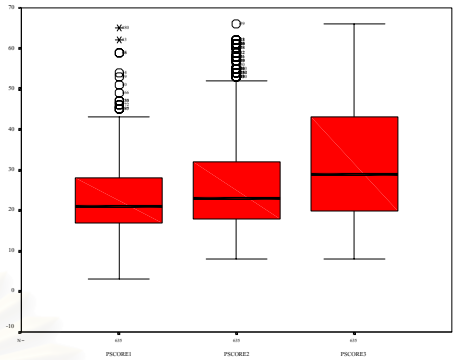
	dichotomous			partial credit		
	DSCORE1	DSCORE2	DSCORE3	PSCORE1	PSCORE2	PSCORE3
N	635	635	635	635	635	635
Minimum	0	0	1	3	8	8
Maximum	21	22	22	65	66	66
Mean	6.992	7.980	10.115	23.027	26.402	32.857
Median	7.000	7.000	9.000	21.000	23.000	29.000
Mode	5	7	6	20	20	19
SD	3.274	4.150	5.352	8.956	12.082	15.402
CV (%)	46.825	52.005	52.912	38.893	45.762	46.876
Variance	10.721	17.225	28.641	80.212	145.973	237.221
Skewness	0.889	0.995	0.697	1.075	1.112	0.705
SE(Sk)	0.097	0.097	0.097	0.097	0.097	0.097
Kurtosis	1.184	0.615	-0.412	1.810	0.632	-0.579
SE(Ku)	0.194	0.194	0.194	0.194	0.194	0.194

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

แผนภาพที่ 4.1 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำ



ด้วยแบบทดสอบฉบับเดิม จากการวัด 3 ครั้ง ที่มีการตรวจให้คะแนนแบบทวิ



ด้วยแบบทดสอบฉบับเดิม ที่มีการตรวจให้คะแนนความรู้บางส่วน

โดยภาพรวมสรุปได้ว่าลักษณะลักษณะการแจกแจงของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) มีลักษณะลักษณะการแจกแจงที่คล้ายคลึงกันมาก

1.1.2. **ข้อมูลชุดที่ 2 ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน(counterbalance)** ได้แก่คะแนนที่ได้จากการวัดด้วยแบบทดสอบคู่ขนาน 3 ฉบับ ในการทดสอบ 3 ครั้ง ซึ่งมีการจัดกลุ่มการวัดแบบจตุรัสละติน (latin square) กล่าวคือในการวัดแต่ละครั้งจะต้องใช้แบบทดสอบทั้ง 3 ฉบับในกลุ่มตัวอย่างที่ต่างกลุ่มกัน ซึ่งกลุ่มตัวอย่างที่ใช้มีทั้งหมด 581 คน มีการตรวจให้คะแนน 2 แบบ คือแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) โดยมีรายละเอียดของคะแนนแต่ละครั้งดังนี้

1.1.2.1. ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) มีคะแนนเต็ม 22 คะแนน คะแนนรวมที่ได้จากการวัดครั้งที่ 2 มีค่าต่ำสุดคือ 0 คะแนน ในขณะที่คะแนนรวมที่ได้จากการวัดครั้งที่ 1 และครั้งที่ 3 มีค่าต่ำสุดเท่ากันคือ 1 สำหรับคะแนนรวมสูงสุดจากการวัดครั้งที่ 1 มีค่าเท่ากับ 21 คะแนน ส่วนคะแนนรวมสูงสุดจากการวัดครั้งที่ 2 และครั้งที่ 3 มีค่าเท่ากันคือ 22 คะแนน คะแนนเฉลี่ยจากการวัดทั้ง 3 ครั้งมีค่าใกล้เคียงกัน โดยมีลำดับสูงที่สุดคือคะแนนจากการวัดครั้งที่ 3 รองลงมาคือคะแนนจากการวัดครั้งที่ 2 และครั้งที่ 1 ซึ่งมีคะแนนดังนี้ 10.071, 9.429 และ 8.544 คะแนนตามลำดับ โดยมีร้อยละ 50 ของกลุ่มตัวอย่างซึ่งมีความสามารถอยู่ในระดับปานกลาง มีพิสัยของคะแนนกว้างที่สุดคือคะแนนจากการวัดครั้งที่ 3 (PR_{75}

PR_{25} อยู่ในช่วงคะแนน 6-14) รองลงมาคือคะ 119 การวัดครั้งที่ 1 และครั้งที่ 2 ซึ่งมีค่าใกล้เคียงกัน (PR_{75} – PR_{25} อยู่ในช่วงคะแนน 5-11 และ 6-11 ตามลำดับ) ซึ่งมีค่ามัธยฐานของคะแนนใกล้เคียงกัน เรียงตามลำดับคือ คะแนนจากการวัดครั้งที่ 3 ครั้งที่ 2 และครั้งที่ 1 โดยมีคะแนนเป็น 9, 8 และ 7 คะแนนตามลำดับ สำหรับความแปรปรวนของคะแนนจากการวัดทั้ง 3 ครั้ง พบว่าการวัดครั้งที่ 3 มีค่าสูงที่สุดและมีค่าใกล้เคียงกับการวัดครั้งที่ 2 (CV (%) = 55.049, 54.481) และการวัดครั้งที่ 1 มีค่าความแปรปรวนต่ำสุด (CV (%) = 51.592)

เมื่อพิจารณารูปแบบโค้งการแจกแจงจากค่าความเบ้และความโด่งพบว่า คะแนนจากการวัดทั้ง 3 ครั้ง มีลักษณะลักษณะการแจกแจงที่คล้ายกันคือ เบ้ขวา โดยคะแนนจากการวัดครั้งที่ 2 มีลักษณะเบ้มากที่สุด รองลงมาคือการวัดครั้งที่ 1 และครั้งที่ 3 ตามลำดับ (Sk = 0.935, 0.868 และ 0.748) ส่วนลักษณะความโด่งของลักษณะการแจกแจงของคะแนนจากการวัดทั้ง 3 ครั้ง มีลักษณะความโด่งแบบเดียวกันคือแบนกว่าโค้งปกติโดยคะแนนจากการวัดครั้งที่ 3 มีลักษณะลักษณะการแจกแจงที่แบนที่สุด รองลงมาคือครั้งที่ 2 และครั้งที่ 1 ตามลำดับ (Ku = -0.544, -0.143 และ –0.055) แต่เมื่อพิจารณาจากค่าความคลาดเคื่อนของความเบ้และความโด่งพบว่ามีค่า 0.101 และ 0.202 ซึ่งมีค่าเท่ากันทั้ง 3 ครั้ง ซึ่งมีค่าไม่เกิน 2.00 ดังนั้นจึงสรุปได้ว่าคะแนนจากการวัดในแต่ละครั้งยังคงมีลักษณะลักษณะการแจกแจงเป็นโค้งปกติ (รายละเอียดดังตารางที่ 4.2 และแผนภาพที่ 4.3)

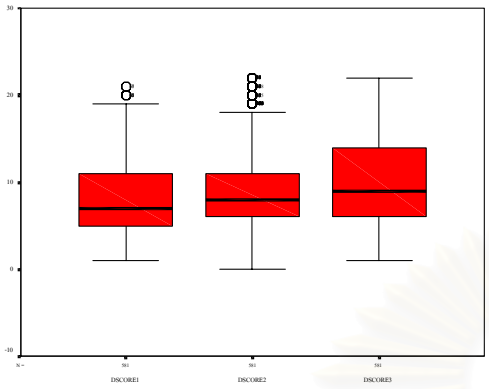
1.1.2.2. ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีคะแนนเต็ม 66 คะแนน คะแนนรวมต่ำสุดจากการวัดครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 คือ 8, 7 และ 9 คะแนน ตามลำดับ ส่วนคะแนนรวมสูงสุดจากการวัดครั้งที่ 1 มีค่าต่ำสุดคือ 63 คะแนน คะแนนรวมสูงสุดจากการวัดครั้งที่ 2 และครั้งที่ 3 มีค่าเท่ากันคือ 66 คะแนน คะแนนเฉลี่ยจากการวัดครั้งที่ 3 มีค่าสูงสุดคือ 32.394 คะแนน รองลงมาคือคะแนนจากการวัดครั้งที่ 2 และครั้งที่ 1 ซึ่งมีคะแนนเป็น 30.349 และ 27.329 คะแนน ตามลำดับ โดยมีร้อยละ 50 ของกลุ่มตัวอย่างซึ่งมีความสามารถอยู่ในระดับปานกลาง มีพิสัยที่กว้างที่สุด ได้แก่คะแนนจากการวัดครั้งที่ 3 (PR_{75} – PR_{25} อยู่ในช่วงคะแนน 20-45) รองลงมาคือคะแนนจากการวัดครั้งที่ 2 (PR_{75} – PR_{25} อยู่ในช่วงคะแนน 19-37) และคะแนนจากการวัดครั้งที่ 1 ตามลำดับ (PR_{75} – PR_{25} อยู่ในช่วงคะแนน 18-32) ตามลำดับ ซึ่งคะแนนจากการวัดครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 มีค่ามัธยฐานเท่ากับ 23, 25 และ 27 คะแนน ตามลำดับ ส่วนความแปรปรวนของคะแนนจากการวัดทั้ง 3 ครั้งนั้น พบว่ามีค่าใกล้เคียงกันทั้ง 3 ครั้ง เรียงตามลำดับคือ ครั้งที่ 1 ครั้งที่ 2 และครั้งที่ 3 (CV (%) = 47.104, 48.888 และ 49.241)

เมื่อพิจารณารูปแบบโค้งการแจกแจงจากค่าความเบ้และความโด่งพบว่า คะแนนจากการวัดทั้ง 3 ครั้ง มีลักษณะการแจกแจงที่คล้ายกันคือ มีลักษณะเบ้ขวา โดยมีคะแนน

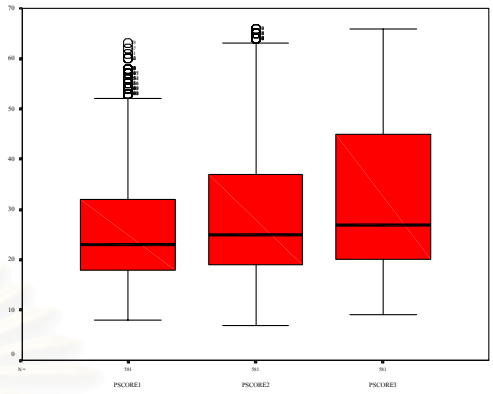
จากการวัดครั้งที่ 1 มีลักษณะเบ้มากที่สุด รองล 120 ะแนนจากการวัดครั้งที่ 2 และครั้งที่ 3 ตามลำดับ (Sk = 1.013, 0.957 และ 0.732) ส่วนลกษณะความโด่งของลักษณะการแจกแจงของคะแนนนั้นพบว่าคะแนนจากการวัดครั้งที่ 1 มีความโด่งใกล้เคียงกับโค้งปกติ (Ku = 0.020) ส่วนคะแนนจากการวัดครั้งที่ 2 และครั้งที่ 3 มีลักษณะเหมือนกันคือ ลักษณะการแจกแจงมีลักษณะแบนกว่าโค้งปกติ โดยคะแนนจากการวัดครั้งที่ 3 มีลักษณะแบนมากกว่าคะแนนจากการวัดครั้งที่ 2 (Ku = - 0.687 และ – 0.251) แต่เมื่อพิจารณาจากค่าความคลาดเคลื่อนของความเบ้และความโด่งของลักษณะการแจกแจงของคะแนนจากการวัดทั้ง 3 ครั้ง พบว่ามีค่า 0.101 และ 0.202 ซึ่งมีค่าเท่ากันทั้ง 3 ครั้ง และมีค่าไม่เกิน 2.00 จึงสรุปได้ว่าคะแนนจากการวัดในแต่ละครั้งยังคงมีลักษณะลักษณะการแจกแจงเป็นโค้งปกติ (รายละเอียดดังตารางที่ 4.2 และแผนภาพที่ 4.4)

ตารางที่ 4.2 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน(counterbalance) ที่ได้จากการทดสอบ 3 ครั้ง จำแนกตามวิธีการตรวจให้คะแนน

	dichotomous			partial credit		
	DSCORE1	DSCORE2	DSCORE3	PSCORE1	PSCORE2	PSCORE3
N	581	581	581	581	581	581
Minimum	1	0	1	8	7	9
Maximum	21	22	22	63	66	66
Mean	8.544	9.429	10.071	27.329	30.349	32.394
Median	7.000	8.000	9.000	23.000	25.000	27.000
Mode	6	6	5	22	22	19
SD	4.408	5.137	5.544	12.873	14.837	15.951
CV (%)	51.592	54.481	55.049	47.104	48.888	49.241
Variance	19.428	26.387	30.738	165.714	220.131	254.422
Skewness	0.868	0.935	0.748	1.013	0.957	0.732
SE(Sk)	0.101	0.101	0.101	0.101	0.101	0.101
Kurtosis	-0.055	-0.143	-0.544	0.020	-0.251	-0.687
SE(Ku)	0.202	0.202	0.202	0.202	0.202	0.202



แผนภาพที่ 4.3 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนานจากการวัด 3 ครั้ง ที่มีการตรวจให้คะแนนแบบทวิวิภาค



แผนภาพที่ 4.4 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนานจากการวัด 3 ครั้ง ที่มีการตรวจให้คะแนนความรู้บางส่วน

โดยภาพรวมสรุปได้ว่าลักษณะลักษณะการแจกแจงของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน(counterbalance) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) มีลักษณะลักษณะการแจกแจงที่คล้ายคลึงกันมาก

1.1.3. **ข้อมูลชุดที่ 3 ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน** (equivalence) ได้แก่คะแนนที่ได้จากการวัดด้วยแบบทดสอบคู่ขนานกัน 5 ฉบับ ที่ใช้สอบครั้งละฉบับ จากกลุ่มตัวอย่าง 386 คน ซึ่งมีการตรวจให้คะแนน 2 แบบ คือแบบทวิวิภาค (dichotomous) และแบบให้คะแนนความรู้บางส่วน (partial credit) ซึ่งมีรายละเอียดของคะแนนดังนี้

1.1.3.1.ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) มีคะแนนเต็ม 25 คะแนน คะแนนรวมที่ได้จากการวัดในครั้งที่ 1 ครั้งที่ 3 และครั้งที่ 5 มีคะแนนต่ำสุดเท่ากันคือ 1 คะแนน ส่วนครั้งที่ 2 และครั้งที่ 4 มีคะแนนต่ำสุดเท่ากันคือ 2 คะแนน ส่วนคะแนนรวมสูงสุดจากการวัดครั้งที่ 1 ถึงครั้งที่ 5 เรียงตามลำดับดังนี้ 16, 18, 19, 20 และ 22 คะแนน คะแนนเฉลี่ยจากการวัดทั้ง 5 ครั้ง มีคะแนนจากการวัดครั้งที่ 5 มีค่าเฉลี่ยสูงสุด คือ 9.969 คะแนน รองลงมาคือครั้งที่ 4 ครั้งที่ 3 ครั้งที่ 2 และครั้งที่ 1 ตามลำดับ โดยมีคะแนนเฉลี่ยเท่ากับ 9.916 8.552, 7.643 และ 7.412 คะแนน โดยมีร้อยละ 50 ของกลุ่มตัวอย่างซึ่งมีความสามารถอยู่ในระดับปานกลาง มีพิสัยของคะแนนกว้างที่สุดได้แก่คะแนนจากการวัดครั้งที่ 4 และครั้งที่ 5 ที่มีค่าใกล้เคียงกัน ($PR_{75} - PR_{25}$ อยู่ในช่วง 7-13 คะแนน และช่วง 6-13 คะแนน) รองลงมาได้แก่คะแนนจากการวัดครั้งที่ 3, ครั้งที่ 2 และครั้งที่ 1 ตามลำดับ ($PR_{75} - PR_{25}$ อยู่ในช่วง 6-11, 6-10 และ 6-9

คะแนน) ซึ่งมีค่ามัธยฐานของคะแนนจากการวัดครั้งที่ 1 ถึงครั้งที่ 5 ใกล้เคียงกันคือ 7, 7, 8, 9 และ 9 คะแนน ตามลำดับ สำหรับความแปรปรวนของคะแนนจากการวัดทั้ง 5 ครั้ง พบว่า คะแนนจากการวัดครั้งที่ 5 มีค่าสูงที่สุด (CV (%) = 43.214) รองลงมาคือคะแนนจากการวัดครั้งที่ 4 (CV (%) = 40.742) ส่วนคะแนนจากการวัดในครั้งที่ 1 ถึงครั้งที่ 3 มีค่าใกล้เคียงกัน (CV (%) = 36.697, 38.192 และ 37.149 ตามลำดับ)

เมื่อพิจารณารูปแบบโค้งการแจกแจงจากค่าความเบ้และความโด่งพบว่า คะแนนจากการวัดครั้งที่ 1 ถึงครั้งที่ 5 มีลักษณะลักษณะการแจกแจงที่คล้ายกันคือมีลักษณะเบ้ขวา โดยมีคะแนนจากการวัดครั้งที่ 2 มีลักษณะเบ้มากที่สุด (Sk = 0.710) รองลงมาได้แก่คะแนนจากการวัดครั้งที่ 5, ครั้งที่ 1, ครั้งที่ 3 และครั้งที่ 4 ตามลำดับ (Sk = 0.564, 0.464, 0.395 และ 0.329) ส่วนลักษณะความโด่งของลักษณะการแจกแจงพบว่าคะแนนจากการวัดครั้งที่ 1 ถึงครั้งที่ 2 มีลักษณะโด่งสูงกว่าโค้งปกติ โดยคะแนนจากการวัดครั้งที่ 2 มีลักษณะโด่งที่สุด (Ku = 0.582) รองลงมาคือคะแนนจากการวัดครั้งที่ 1 (Ku = 0.408) ส่วนครั้งที่ 3 มีลักษณะใกล้เคียงกับโค้งปกติ (Ku = 0.014) สำหรับคะแนนจากการวัดครั้งที่ 4 และครั้งที่ 5 นั้น มีลักษณะลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีคะแนนจากการวัดครั้งที่ 4 แบนกว่าการวัดครั้งที่ 5 (Ku = -0.673, -0.464) แต่เมื่อพิจารณาค่าความคลาดเคลื่อนของความเบ้ พบว่ามีค่าตั้งแต่ 0.124 ถึง 0.127 และค่าความคลาดเคลื่อนของความโด่งมีค่าตั้งแต่ 0.248 ถึง 0.253 ในการวัดทั้ง 5 ครั้ง ซึ่งมีค่าไม่ถึง 2.00 ดังนั้น จึงสรุปได้ว่าคะแนนจากการวัดในแ 123 ยังคงมีลักษณะลักษณะการแจกแจงเป็นโค้งปกติ (รายละเอียดอยู่ในตารางที่ 4.3 และแผนภาพที่ 4.5)

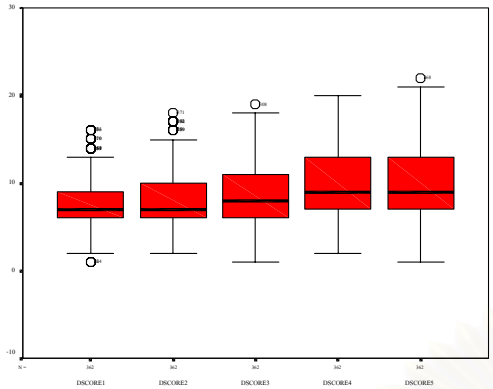
1.1.3.2. ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีคะแนนเต็ม 50 คะแนน คะแนนรวมที่ได้จากการวัดครั้งที่ 1 มีคะแนนต่ำสุด ต่ำกว่าคะแนนจากการวัดในครั้งอื่น ๆ คือ 5 คะแนน ส่วนคะแนนรวมต่ำสุดจากการวัดในครั้งที่ 2 ถึงครั้งที่ 4 มีค่าใกล้เคียงกันคือ 14, 14, 13 และ 14 คะแนน ตามลำดับ ส่วนคะแนนรวมสูงสุดจากการวัดครั้งที่ 1 และครั้งที่ 2 มีค่าเท่ากัน และมีค่าต่ำกว่าคะแนนจากครั้งอื่น ๆ คือ 39 คะแนน ส่วนคะแนนครั้งที่ 3 ถึงครั้งที่ 5 มีคะแนนรวมสูงสุดคือ 41, 43 และ 46 คะแนน ตามลำดับ ส่วนคะแนนเฉลี่ยของคะแนนที่ได้จากการวัดทั้ง 5 ครั้ง คะแนนเฉลี่ยมีค่าสูงสุดจากการวัดครั้งที่ 5 คือ 29.003 คะแนน โดยมีคะแนนเฉลี่ยจากการวัดครั้งที่ 4 ถึงครั้งที่ 1 เรียงตามลำดับดังนี้ 28.770, 27.337, 24.759 และ 24.736 คะแนน ซึ่งมีร้อยละ 50 ของกลุ่มตัวอย่างซึ่งมีความสามารถอยู่ในระดับปานกลาง มีพิสัยของคะแนนกว้างที่สุดได้แก่คะแนนจากการวัดในครั้งที่ 4 และครั้งที่ 5 ที่มีค่าใกล้เคียงกัน (PR_{75} – PR_{25} อยู่ในช่วง 24-33 คะแนน และ 24-34 คะแนน) รองลงมาได้แก่คะแนนจากการวัดในครั้งที่ 3, ครั้งที่ 2 และครั้งที่ 1 ซึ่งมีค่าใกล้เคียงกัน (PR_{75} – PR_{25} อยู่ในช่วง 24-31, 21-28 และ 22-28 คะแนน) โดยมีค่ามัธยฐานของคะแนนจากการวัดครั้งที่ 4 มีค่าสูงสุดและมีค่าใกล้เคียงกับค่ามัธยฐานองคะแนนจากการวัดครั้งที่ 5 และครั้งที่ 3 คือ 28.5, 28 และ 27.5 คะแนน ตามลำดับ และค่ามัธยฐานจากการวัดในครั้งที่ 1 และครั้งที่ 2 มีค่าใกล้เคียงกันคือ 25 และ 24 คะแนน สำหรับความแปรปรวนของคะแนนครั้งที่ 2 มีค่าสูงสุด (CV (%) = 22.860) รองลงมาคือคะแนนจากการวัด

ครั้งที่ 4, ครั้งที่ 1, ครั้งที่ 2 และครั้งที่ 3 ตามลำดับ (CV (%) = 21.126, 19.490, 19.031 และ 17.913)

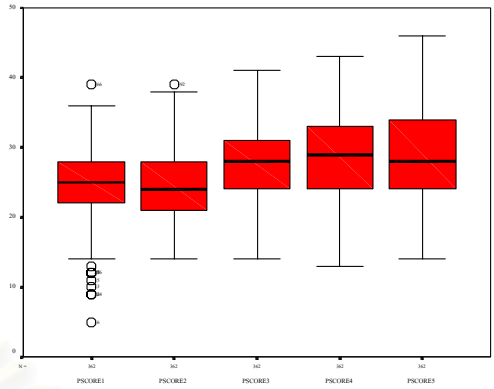
เมื่อพิจารณารูปแบบโค้งการแจกแจงจากค่าความเบ้และความโด่งพบว่า คะแนนจากการวัดครั้งที่ 1 มีลักษณะเบ้ซ้าย (Sk = -0.493) ส่วนคะแนนจากการวัดครั้งที่ 2 ถึงครั้งที่ 5 มีลักษณะคล้ายคลึงกัน คือมีลักษณะเบ้ขวา โดยมีการวัดในครั้งที่ 5 มีคะแนนเบ้ขวามากที่สุด (Sk = 0.331) รองลงมาคือคะแนนจากการวัดครั้งที่ 2 และ ครั้งที่ 4 ส่วนครั้งที่ 3 มีลักษณะใกล้เคียงกับโค้งปกติ (Sk = 0.273, 0.132 และ 0.039 ตามลำดับ) สำหรับความโด่งของลักษณะการแจกแจงของคะแนนจากการวัดทั้ง 5 ครั้ง พบว่าคะแนนจากการวัดครั้งที่ 1 มีลักษณะโด่งสูงกว่าโค้งปกติ (Ku = 1.542) ส่วนคะแนนจากการวัดในครั้งที่ 2 ถึงครั้งที่ 2 มีลักษณะลักษณะการแจกแจงที่แบนกว่าลักษณะการแจกแจงโค้งปกติ โดยมีคะแนนจากการวัดในครั้งที่ 4 แบนที่สุด (Ku = -0.639) รองลงมาคือคะแนนจากการวัดครั้งที่ 5, ครั้งที่ 3 และครั้งที่ 2 ตามลำดับ (Ku = -0.516, -0.268 และ – 0.128) แต่เมื่อพิจารณาค่าความคลาดเคลื่อนของความเบ้ พบว่ามีค่าตั้งแต่ 0.124 ถึง 0.127 และค่าความคลาดเคลื่อนของความโด่งมีค่าตั้งแต่ 0.248 ถึง 0.253 จากการวัดทั้ง 5 ครั้ง ซึ่งมีค่าไม่ถึง 2.00 จึงสรุปได้ว่าคะแนนจากการวัดในแต่ละครั้งยังคงมีลักษณะลักษณะการแจกแจงเป็นโค้งปกติ (รายละเอียดดังตารางที่ 4.3 และแผนภาพที่ 4.6)

ตารางที่ 4.3 ผลการวิเคราะห์ค่าสถิติพื้นฐานของข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence) ที่ได้จากการทดสอบ จำแนกตามวิธีการตรวจให้คะแนน

	dichotomous					partial credit				
	DSCORE1	DSCORE2	DSCORE3	DSCORE4	DSCORE5	PSCORE1	PSCORE2	PSCORE3	PSCORE4	PSCORE5
N	386	378	386	370	386	386	378	386	370	386
Minimum	1	2	1	2	1	5	14	14	13	14
Maximum	16	18	19	20	22	39	39	41	43	46
Mean	7.412	7.643	8.552	9.916	9.969	24.736	24.759	27.337	28.770	29.003
Median	7.000	7.000	8.000	9.000	9.000	25.000	24.000	27.500	28.500	28.000
Mode	7	6	7	8	6	26	24	30	26	24
SD	2.720	2.919	3.177	4.040	4.308	4.821	4.712	4.897	6.078	6.630
CV (%)	36.697	38.192	37.149	40.742	43.214	19.490	19.031	17.913	21.126	22.860
Variance	7.396	8.522	10.092	16.321	18.555	23.244	22.199	23.985	36.947	43.951
Skewness	0.464	0.710	0.395	0.329	0.564	-0.493	0.273	0.039	0.132	0.331
SE(SK)	0.124	0.125	0.124	0.127	0.124	0.124	0.125	0.124	0.127	0.124
Kurtosis	0.408	0.582	0.014	-0.673	-0.464	1.542	-0.128	-0.268	-0.639	-0.516
SE(KU)	0.248	0.250	0.248	0.253	0.248	0.248	0.250	0.248	0.253	0.248



แผนภาพที่ 4.5 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน จากการวัด 5 ครั้ง ที่มีการตรวจให้คะแนนแบบทวิวิภาค



แผนภาพที่ 4.6 แผนภูมิ boxplot ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน จากการวัด 5 ครั้ง ที่มีการตรวจให้คะแนนความรู้บางส่วน

โดยภาพรวมสรุปได้ว่า ลักษณะการแจกแจงของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนานนี้มีลักษณะไม่ค่อยสอดคล้องกันของข้อมูลที่ตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) ในการวัดครั้งที่ 1 ถึงการวัดครั้งที่ 3 โดยการวัดครั้งที่ 1 คะแนนมีลักษณะลักษณะการแจกแจงที่เบ้ตรงกันข้ามกัน ส่วนคะแนนจากการวัดครั้งที่ 2 และครั้งที่ 3 คะแนนมีลักษณะโด่งที่ตรงกันข้ามกัน

1.2 ผลการทดสอบเงื่อนไข sphericity ของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดระหว่างเวลา

ก่อนที่ผู้วิจัยจะทำการทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำนั้น ผู้วิจัยได้ทำการตรวจสอบความเป็น sphericity ของเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัด 3 ครั้ง สำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนานที่มีการวัด 3 ครั้ง (equivalence 1) และข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนานที่มีการวัด 5 ครั้ง (equivalence 2) เพื่อดูว่าเมื่อมีการแปลงเป็นตัวแปรการวัดซ้ำแล้ว จะมีเมทริกซ์ของความแปรปรวน-ความแปรปรวนร่วมเป็นเมทริกซ์เอกลักษณ์หรือไม่ เพื่อที่จะได้เลือกสูตรการหาค่า degrees of freedom ได้ถูกต้อง เมื่อทำการทดสอบความแปรปรวนแบบวัดซ้ำ (repeated measure)

จากการทดสอบ sphericity ของข้อมูลทุกชุด โดยใช้การทดสอบของ Mauchly พบว่าข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วย

แบบทดสอบคู่ขนาน (counterbalance) และข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนานที่มีการวัด 5 ครั้ง (equivalence 2) เมื่อตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) เมื่อแปลงเป็นตัวแปรการวัดซ้ำ (time) จะมีเมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมเป็นเมทริกซ์เอกลักษณ์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ดังนั้นในการทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลชุดต่างๆ เหล่านี้ผู้วิจัยจะใช้การประมาณค่าแบบ Greenhouse-Geisser แทนการประมาณค่าแบบ sphericity assumed สำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนานที่มีการวัด 3 ครั้ง (equivalence 1) ที่มีการให้คะแนนแบบทวิวิภาค (dichotomous) และแบบให้คะแนนความรู้บางส่วน (partial credit) เมื่อแปลงเป็นตัวแปรวัดซ้ำ (time) จะมีเมทริกซ์ความแปรปรวน-ความแปร ปรวนร่วม ไม่เป็นเมทริกซ์เอกลักษณ์ จึงยังคงใช้การประมาณค่าแบบ sphericity assumed (รายละเอียดดังตารางที่ 4.4)

ตารางที่ 4.4 ผลการทดสอบ sphericity ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์จำแนกตามจำนวนครั้งของการทดสอบ และแบบทดสอบที่ใช้

แบบทดสอบ	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Repeat							
- dichotomous	0.882	79.215**	2.000	0.000	0.895	0.897	0.500
- partial credit	0.786	152.750**	2.000	0.000	0.823	0.825	0.500
Counterbalance							
- dichotomous	0.966	19.741**	2.000	0.000	0.968	0.971	0.500
- partial credit	0.927	44.135**	2.000	0.000	0.932	0.935	0.500
equivalence 1							
- dichotomous	0.994	2.127	2.000	0.342	0.994	1.000	0.500
- partial credit	0.990	3.699	2.000	0.155	0.990	0.996	0.500
equivalence 2							
- dichotomous	0.794	82.856**	9.000	0.000	0.886	0.896	0.250
- partial credit	0.848	59.313**	9.000	0.000	0.914	0.925	0.250

** $p < 0.01$

1.3 ผลการทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดระหว่างเวลา

ในการทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ โดยทดสอบด้วย One-Way ANOVA repeated measure ผู้วิจัยขอนำเสนอรายละเอียดดังนี้

1.3.1 ข้อมูลชุดที่ 1 ข้อมูลวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ พบว่าค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยค่าเฉลี่ยจากการวัดครั้งที่ 1 มีค่าต่ำกว่าค่าเฉลี่ยจากการวัดครั้งที่ 2 และครั้งที่ 3 และค่าเฉลี่ยจากการวัดครั้งที่ 2 มีค่าต่ำกว่าค่าเฉลี่ยจากการวัดครั้งที่ 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทั้งคะแนนที่เป็นแบบทวิวิภาค (dichotomous) และให้คะแนนความรู้บางส่วน (partial credit) (รายละเอียดดังตารางที่ 4.5-4.6 และแผนภาพที่ 4.7-4.8)

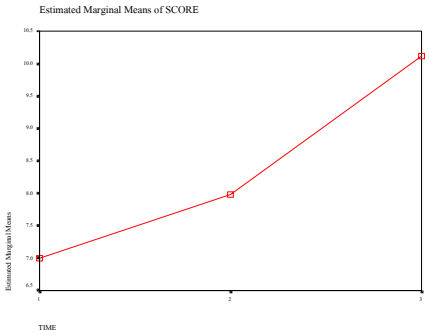
ตารางที่ 4.5 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ระหว่างเวลาของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat)

Source	dichotomous					partial credit				
	Type III SS	df	MS	F	Sig.	Type III SS	df	MS	F	Sig.
TIME	3235.776	1.790	1808.198	193.812**	0.000	31683.367	1.647	19238.184	312.532**	0.000
Error(TIME)	10584.890	1134.545	9.330			64272.633	1044.135	61.556		

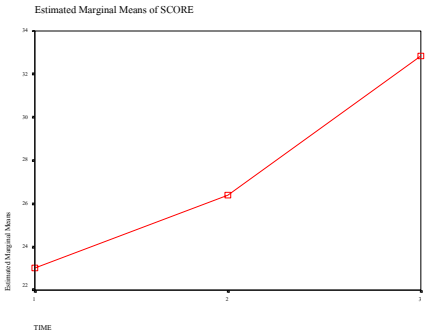
** $p < 0.01$

ตารางที่ 4.6 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat)

TIME	dichotomous			partial credit		
	Mean Difference	SE	Sig.	Mean Difference	SE	Sig.
ครั้งที่ 1 กับ ครั้งที่ 2	-0.987	0.140	0.000	-3.375	0.333	0.000
ครั้งที่ 1 กับ ครั้งที่ 3	-3.123	0.187	0.000	-9.830	0.482	0.000
ครั้งที่ 2 กับ ครั้งที่ 3	-2.135	0.156	0.000	-6.455	0.369	0.000



แผนภาพที่ 4.7 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์เฉลี่ยของนักเรียนในการวัด 3 ครั้ง ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม ที่มีการตรวจให้คะแนนแบบทวิวิภาค



แผนภาพที่ 4.8 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์เฉลี่ยของนักเรียนในการวัด 3 ครั้ง ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม ที่มีการตรวจให้คะแนนความรู้บางส่วน

1.3.2 **ข้อมูลชุดที่ 2 ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance)** ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ พบว่าค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยค่าเฉลี่ยจากการวัดครั้งที่ 1 มีค่าต่ำกว่าค่าเฉลี่ยจากการวัดครั้งที่ 2 และ 3 และค่าเฉลี่ยจากการวัดครั้งที่ 2 มีค่าต่ำกว่าค่าเฉลี่ยจากการวัดครั้งที่ 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทั้งคะแนนที่เป็นทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) (รายละเอียดดังตารางที่ 4.7-4.8 และแผนภาพที่ 4.9-4.10)

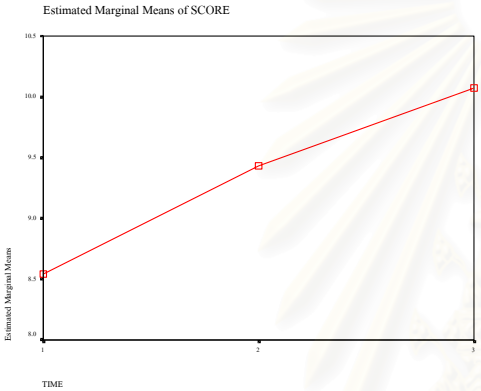
ตารางที่ 4.7 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ระหว่างเวลาของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance)

Source	dichotomous					partial credit				
	Type III SS	df	MS	F	Sig.	Type III SS	df	MS	F	Sig.
TIME	682.785	1.935	352.836	53.281**	0.000	7545.966	1.863	4049.897	105.163**	0.000
Error(TIME)	7432.548	1122.377	6.622			41618.034	1080.684	38.511		

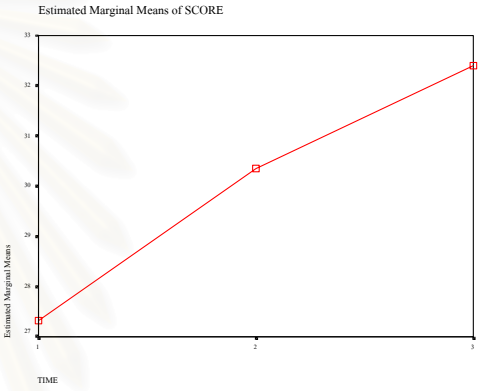
** $p < 0.01$

ตารางที่ 4.8 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance)

TIME	dichotomous			partial credit		
	Mean Difference	SE	Sig.	Mean Difference	SE	Sig.
ครั้งที่ 1 กับ ครั้งที่ 2	-0.885	0.137	0.000	-3.021	0.306	0.000
ครั้งที่ 1 กับ ครั้งที่ 3	-1.527	0.160	0.000	-5.065	0.389	0.000
ครั้งที่ 2 กับ ครั้งที่ 3	-0.642	0.147	0.000	-2.045	0.354	0.000



แผนภาพที่ 4.9 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนในการวัด 3 ครั้ง ของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนานที่มีการตรวจให้คะแนนแบบทวิวิภาค



แผนภาพที่ 4.10 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนในการวัด 3 ครั้ง ของชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน ที่มีการตรวจให้คะแนนความรู้บางส่วน

1.3.3 **ข้อมูลชุดที่ 3 ข้อมูลวัดซ้ำด้วยแบบทดสอบคู่ขนาน** (equivalence) ซึ่งผู้วิจัยได้แยกผลการวิเคราะห์ออกเป็นการวัด 3 ครั้ง และการวัด 5 ครั้ง ดังนี้

1.3.3.1 ข้อมูลวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ พบว่าค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์มีความแตแต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยค่าเฉลี่ยจากการวัดครั้งที่ 1 กับครั้งที่ 2 ไม่แตกต่างกัน ส่วนค่าเฉลี่ยจากการวัดในครั้งที่ 1 และครั้งที่ 2 มีค่าต่ำกว่าค่าเฉลี่ยจากการวัดครั้งที่ 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทั้งคะแนนที่เป็นทวิวิภาค (dichotomous) และ ตรวจให้คะแนนความรู้บางส่วน (partial credit) (รายละเอียดดังตารางที่ 4.9-4.10 และแผนภาพที่ 4.11-4.12)

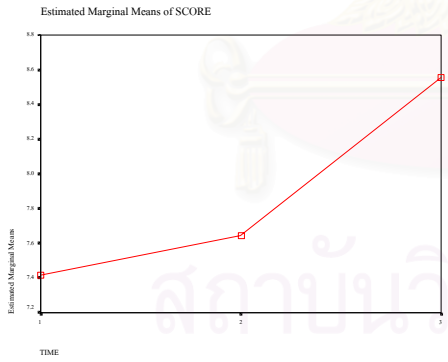
ตารางที่ 4.9 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1)

Source	dichotomous					partial credit				
	Type III SS	df	MS	F	Sig.	Type III SS	df	MS	F	Sig.
TIME	273.926	2.000	136.963	22.942**	0.000	1671.466	2.000	835.733	47.801**	0.000
Error(TIME)	4501.407	754.000	5.970			13182.534	754.000	17.483		

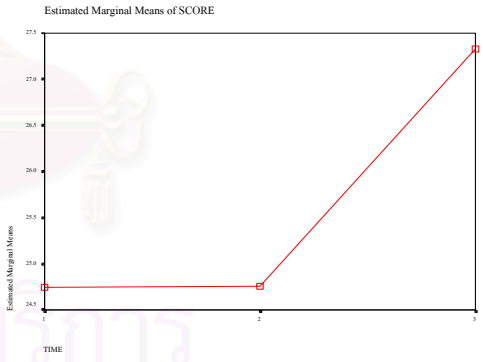
** p < 0.01

ตารางที่ 4.10 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1)

TIME	dichotomous			partial credit		
	Mean Difference	SE	Sig.	Mean Difference	SE	Sig.
ครั้งที่ 1 กับ ครั้งที่ 2	-0.228	0.172	0.186	-0.013	0.315	0.966
ครั้งที่ 1 กับ ครั้งที่ 3	-1.138	0.183	0.000	-2.582	0.308	0.000
ครั้งที่ 2 กับ ครั้งที่ 3	-0.910	0.178	0.000	-2.569	0.289	0.000



แผนภาพที่ 4.11 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนที่ได้จากข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน ที่มีการตรวจคะแนนแบบทวิวิภาค



แผนภาพที่ 4.12 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนที่ได้จากข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน ที่มีการตรวจให้คะแนนความรู้บางส่วน

1.3.3.2 ข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ พบว่าค่าเฉลี่ยของคะแนนมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยค่าเฉลี่ยจากการวัดครั้งที่ 1 กับครั้งที่ 2 และค่าเฉลี่ยจากการวัดครั้งที่ 4 กับครั้งที่ 5 ไม่แตกต่างกัน ส่วนค่าเฉลี่ยจากการวัดครั้งที่ 1 และครั้งที่ 2 มีค่าต่ำกว่าค่าเฉลี่ยจากการวัดครั้งที่ 3, ครั้งที่ 4 และครั้งที่ 5 ค่าเฉลี่ยจากการวัดครั้งที่ 3 มีค่าต่ำกว่าค่าเฉลี่ยจากการวัดครั้งที่ 4 และครั้งที่ 5 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทั้งคะแนนที่เป็น dichotomous และ partial credit (รายละเอียดดังตารางที่ 4.11-4.12 และแผนภาพที่ 4.13-4.14)

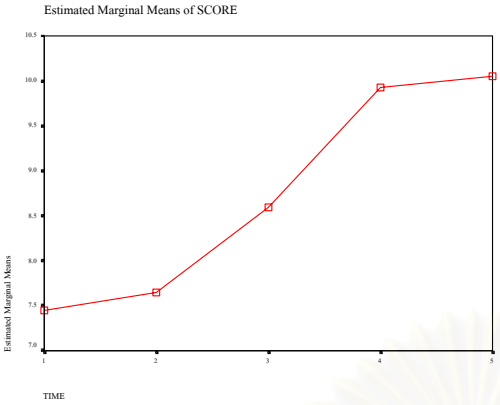
ตารางที่ 4.11 ผลการวิเคราะห์ความแปรปรวนของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2)

Source	dichotomous					partial credit				
	Type III SS	df	MS	F	Sig.	Type III SS	df	MS	F	Sig.
TIME	2178.196	3.546	614.309	66.112**	0.000	6333.110	3.656	1732.211	72.437**	0.000
Error(TIME)	11893.804	1280.022	9.292			31562.090	1319.846	23.913		

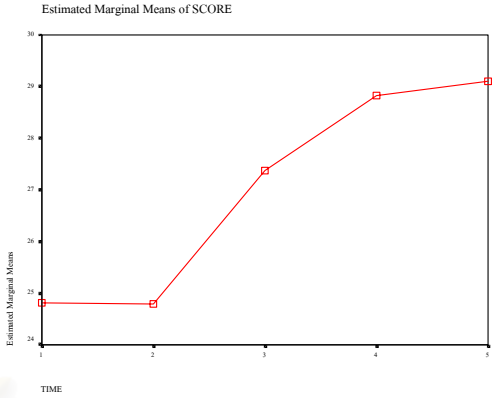
** $p < 0.01$

ตารางที่ 4.12 ผลการเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากข้อมูลวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2)

TIME	dichotomous			partial credit		
	Mean Difference	SE	Sig.	Mean Difference	SE	Sig.
ครั้งที่ 1 กับ ครั้งที่ 2	-0.207	0.176	0.239	0.030	0.319	0.924
ครั้งที่ 1 กับ ครั้งที่ 3	-1.149	0.189	0.000	-2.552	0.316	0.000
ครั้งที่ 1 กับ ครั้งที่ 4	-2.483	0.227	0.000	-3.997	0.382	0.000
ครั้งที่ 1 กับ ครั้งที่ 5	-2.608	0.239	0.000	-4.282	0.391	0.000
ครั้งที่ 2 กับ ครั้งที่ 3	-0.942	0.184	0.000	-2.583	0.298	0.000
ครั้งที่ 2 กับ ครั้งที่ 4	-2.276	0.228	0.000	-4.028	0.346	0.000
ครั้งที่ 2 กับ ครั้งที่ 5	-2.401	0.228	0.000	-4.312	0.360	0.000
ครั้งที่ 3 กับ ครั้งที่ 4	-1.334	0.218	0.000	-1.445	0.348	0.000
ครั้งที่ 3 กับ ครั้งที่ 5	-1.459	0.227	0.000	-1.729	0.369	0.000
ครั้งที่ 4 กับ ครั้งที่ 5	-0.124	0.209	0.552	-0.285	0.332	0.393



แผนภาพที่ 4.13 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนที่ได้จากข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)



แผนภาพที่ 4.14 พัฒนาการของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนที่ได้จากการวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 2 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลพัฒนาการที่มีตัวแปรแฝงกับข้อมูลเชิงประจักษ์ และประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์สำหรับใช้เป็นคะแนนเกณฑ์

การวิเคราะห์ในตอนที่ 2 นี้ ผู้วิจัยได้แบ่งออกเป็น 2 ขั้นตอน ขั้นแรกเป็นการตรวจสอบความสอดคล้องกลมกลืนของโมเดลพัฒนาการที่มีตัวแปรแฝง 5 รูปแบบกับข้อมูลเชิงประจักษ์ เพื่อให้ได้โมเดลที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลทุกชุด ขั้นที่ 2 เป็นการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนแต่ละคนของข้อมูลทุกชุด ซึ่งมีรายละเอียดดังนี้

2.1 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์

ผู้วิจัยได้สร้างและทำการตรวจสอบโมเดลทั้งหมด 5 โมเดล คือ โมเดลพัฒนาการที่ไม่มีความชัน (NSB) โมเดลพัฒนาการเชิงเส้น (LIN) โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์คงที่ (FIX) โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดค่าพารามิเตอร์อิสระ (FRE) และโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์อิสระและมีความแปรปรวนของค่าความคลาดเคลื่อนไม่เท่ากัน (UDV) ซึ่งมีรายละเอียดดังนี้

2.1.1 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat)

2.1.1.1 จากการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) พบว่าโมเดล NSB, LIN, FIX, FRE และ UDV มีค่า χ^2 เท่ากับ 595.7359, 36.6842, 75.9146, 14.1507 และ 0.3580 ตามลำดับ มี df เท่ากับ 7, 3, 5, 2 และ 1 ตามลำดับ โดยมีนัยสำคัญทางสถิติที่ 0.0000 สำหรับโมเดล NSB, LIN และ FIX ส่วนโมเดล FRE มีนัยสำคัญทางสถิติที่ระดับ 0.0008 และโมเดล UDV มีนัยสำคัญทางสถิติที่ระดับ 0.5496 เมื่อพิจารณาจากดัชนีค่า ไค-สแควร์ (χ^2) ที่มีค่าใกล้กับองศาอิสระ ดัชนีวัดความกลมกลืน (GFI) ที่มีค่าใกล้ 1.00 ดัชนีรากที่สองของค่าเฉลี่ยกำลังสอง (RMR) มีค่าใกล้ศูนย์ และค่าความคลาดเคลื่อนที่เป็นคะแนนมาตรฐานสูงสุด (LSR) มีค่าใกล้ศูนย์ สรุปได้ว่าข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) โมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุดคือโมเดล UDV (รายละเอียดดังตารางที่ 4.13)

2.1.1.2 ผลการตรวจสอบโมเดลพัฒนาการตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) พบว่าโมเดล NSB, LIN, FIX, FRE และ UDV มีค่า χ^2 เท่ากับ 925.3652, 58.2175, 262.3240, 32.2845 และ 4.3220 ตามลำดับ มี df เท่ากับ 7, 3, 5, 2 และ 2 ตามลำดับ โดยมีนัยสำคัญทางสถิติที่ 0.0000 สำหรับโมเดล NSB, LIN, FIX และ FRE ส่วนโมเดล UDV มีนัยสำคัญทางสถิติที่ระดับ 0.1152 เมื่อพิจารณาจากดัชนีค่า χ^2, GFI ถึงแม้ว่า RMR ของโมเดล FRE จะมีค่าต่ำกว่าก็ตาม แต่เมื่อเทียบจากค่าดัชนีอื่น ๆ แล้ว โมเดล UDV สอดคล้องกลมกลืนกับข้อมูลมากกว่าโมเดล FRE ดังนั้นสรุปได้ว่าโมเดลที่มีความสอดคล้องกลมกลืนกับข้อมูลชุดนี้มากที่สุดคือโมเดล UDV (รายละเอียดดังตารางที่ 4.13)

ตารางที่ 4.13 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลพัฒนาการที่มีตัวแปรแฝงของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม

ค่าสถิติ	NSB	LIN	FIX	FRE	UDV
dichotomous					
χ^2	595.7359	36.6842	75.9146	14.1507	0.3580
P	0.0000	0.0000	0.0000	0.0008	0.5496
df	7	3	5	2	1
RMR	17.8966	2.8089	14.9782	0.1173	0.1792
GFI	0.6761	0.9727	0.9464	0.9891	0.9997
LSR	9.1954	4.9078	0.0000	2.0697	0.5980
partial credit					
χ^2	925.3652	58.2175	262.3240	32.2845	4.3220
P	0.0000	0.0000	0.0000	0.0000	0.1152
df	7	3	5	2	2
RMR	177.1295	21.9577	328.8671	0.9070	4.4712
GFI	0.5908	0.9580	0.8546	0.9758	0.9966
LSR	8.6520	5.8607	0.0000	3.1622	1.9971

2.1.2 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) (รายละเอียดดังตารางที่ 4.14)

2.1.2.1 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์คณิตศาสตร์ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) พบว่าโมเดล NSB, LIN, FIX, FRE และ UDV มีค่า χ^2 เท่ากับ 684.6046, 29.4135, 19.8082, 11.8846 และ 0.2429 ตามลำดับ มี df เท่ากับ 7, 3, 5, 2 และ 1 ตามลำดับ โดยมีนัยสำคัญทาง

งถิติที่ 0.0000 สำหรับโมเดล NSB และ LIN และมีนัยสำคัญทางสถิติที่ระดับ 0.0014, 0.0026 และ 0.6221 สำหรับโมเดล FIX. FRE และ UDV ตามลำดับ เมื่อพิจารณาจากค่า χ^2, GFI, RMR และ LSR สรุปได้ว่าโดมล UDV มีความสอดคล้องกลมกลืนกับข้อมูลชุดนี้มากที่สุด

2.1.2.2 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) พบว่าโมเดล NSB, LIN, FIX, FRE และ UDV มีค่า χ^2 เท่ากับ 314.9616, 33.0227, 43.7118, 32.5184 และ 0.0590 ตามลำดับ มี df เท่ากับ 7, 3, 5, 2 และ 1 ตามลำดับ โดยมีนัยสำคัญทางสถิติที่ 0.0000 สำหรับโมเดล NSB, LIN, FIX และ FRE ส่วนโมเดล UDV มีนัยสำคัญทางสถิติที่ระดับ 0.8081 และเมื่อพิจารณาจากค่า χ^2, GFI, RMR และ LSR สรุปได้ว่าโมเดลที่มีความสอดคล้องกลมกลืนกับข้อมูลชุดนี้มากที่สุดคือโมเดล UDV

ตารางที่ 4.14 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลพัฒนาการที่มีตัวแปรแฝงของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance)

ค่าสถิติ	NSB	LIN	FIX	FRE	UDV
Dichotomous					
χ^2	648.6046	29.4135	19.8082	11.8846	0.2429
P	0.0000	0.0000	0.0014	0.0026	0.6221
df	7	3	5	2	1
RMR	93.3561	1.3673	4.9140	0.6428	0.0871
GFI	0.7201	0.9769	0.9835	0.9900	0.9998
LSR	0.0000	3.1269	0.0000	2.7251	0.4934
partial credit					
χ^2	314.9616	33.0227	43.7118	32.5184	0.0590
P	0.0000	0.0000	0.0000	0.0000	0.8081
df	7	3	5	2	1
RMR	97.9944	13.7699	0.0718	10.1034	0.1929
GFI	0.7664	0.9731	0.9647	0.9735	0.9999
LSR	3.3615	3.9885	0.0000	5.1800	0.2431

2.1.3 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 1) (รายละเอียดดังตารางที่ 4.15)

2.1.3.1 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)

พบว่าโมเดล NSB, LIN, FIX, FRE และ UDV มีค่า χ^2 เท่ากับ 57.5404, 6.1253, 53.9702, 1.4442 และ 0.8037 ตามลำดับ มี df เท่ากับ 7, 3, 5, 2 และ 1 ตามลำดับ โดยมีนัยสำคัญทางสถิติที่ 0.0000 สำหรับโมเดล NSB และ FIX ส่วนโมเดล LIN, FRE และ UDV มีนัยสำคัญทางสถิติที่ระดับ 0.1057, 0.4857 และ 0.3700 ตามลำดับ เมื่อพิจารณาจากค่า χ^2, RMR, GFI และ LSR สรุปได้ว่าโมเดล FRE มีความสอดคล้องกลมกลืนกับข้อมูลมากที่สุด

2.1.3.2 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) พบว่าโมเดล NSB, LIN, FIX, FRE และ UDV มีค่า χ^2 เท่ากับ 99.4240, 28.0775, 536.8730, 3.7515 และ 0.2901 ตามลำดับ มี df เท่ากับ 7, 3, 5, 2 และ 1 ตามลำดับ โดยมีนัยสำคัญทางสถิติที่ 0.0000 สำหรับโมเดล NSB, LIN และ FIX ส่วนโมเดล FRE และ UDV มีนัยสำคัญทางสถิติที่ระดับ 0.1532 และ 0.5901 ตามลำดับ เมื่อพิจารณาจากค่า χ^2 และ GFI โมเดล UDV มีความสอดคล้องกลมกลืนกับข้อมูลมากกว่าโมเดล FRE ถึงแม้ว่าจะมีค่า RMR สูงกว่าโมเดล FRE และจากค่า LSR พบว่าโมเดล FRE ไม่สามารถคำนวณได้ เมื่อกำหนดให้หาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างความคลาดเคลื่อนของตัวแปรแฝงความสามารถเริ่มต้นกับความคลาดเคลื่อนของตัวแปรแฝงอัตราพัฒนาการ ดังนั้นจึงสรุปได้ว่าโมเดล UDV มีความสอดคล้องกับข้อมูลชุดนี้มากที่สุด

ตารางที่ 4.15 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลพัฒนาการที่มีตัวแปรแฝงของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1)

ค่าสถิติ	NSB	LIN	FIX	FRE	UDV
dichotomous					
χ^2	57.5404	6.1253	53.9702	1.4442	0.8037
P	0.0000	0.1057	0.0000	0.4857	0.3700
df	7	3	5	2	1
RMR	5.3896	1.5119	11.2394	0.0561	0.1617
GFI	0.9275	0.9922	0.9386	0.9981	0.9990
LSR	2.5920	2.2866	0.0000	0.7938	0.8963
partial credit					
χ^2	99.4240	28.0775	536.8730	3.7515	0.2901
P	0.0000	0.0000	0.0000	0.1532	0.5901
df	7	3	5	2	1
RMR	39.1350	19.1659	64.9407	0.6287	1.5597
GFI	0.8930	0.9658	0.6300	0.9952	0.9996
LSR	1.8818	-	0.0000	-	0.5413

2.1.4 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 2) (รายละเอียดดังตารางที่ 4.16)

2.1.4.1 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) พบว่าโมเดล NSB, LIN, FIX, FRE และ UDV มีค่า χ^2 เท่ากับ 437.0513, 63.0241, 230.0805, 292.8251 และ 13.8025 ตามลำดับ มี df เท่ากับ 18, 14, 16, 11 และ 9 ตามลำดับ โดยมีนัยสำคัญทางสถิติที่ 0.0000 สำหรับโมเดล และ NSB, LIN, FIX และ FRE ตามลำดับ ส่วนโมเดล UDV มีนัยสำคัญทางสถิติที่ระดับ 0.1295 เมื่อพิจารณาจากค่า χ^2, RMR, GFI และ LSR พบว่าโมเดล UDV มีความสอดคล้องกลมกลืนกับข้อมูลชุดนี้มากที่สุด

2.1.4.2 ผลการตรวจสอบโมเดลพัฒนาการที่มีตัวแปรแฝงที่สอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) พบว่าโมเดล NSB, LIN, FIX, FRE และ UDV มีค่า χ^2 เท่ากับ 2610.3476, 61.4955, 574.1280, 32.3422 และ 30.1783 ตามลำดับ มี df เท่ากับ 18, 14, 16, 11 และ 9 ตามลำดับ โดยมีนัยสำคัญทางสถิติที่ 0.0000 สำหรับโมเดล NSB, LIN และ FIX ส่วนโมเดล FRE และ UDV มีนัยสำคัญทางสถิติที่ระดับ 0.0001 และ 0.0004 ตามลำดับ เมื่อพิจารณาจากค่า χ^2, RMR, GFI และ LSR พบว่าโมเดล UDV มีความสอดคล้องกลมกลืนกับข้อมูลชุดนี้มากที่สุดเมื่อเทียบกับโมเดลทั้ง 4 นี้ แต่จากค่า χ^2, RMR, GFI และ LSR บ่งชี้ว่ามีโมเดลอื่นนอกเหนือจากโมเดลเหล่านี้ที่มีความสอดคล้องกลมกลืนกับข้อมูลชุดนี้มากกว่า

ตารางที่ 4.16 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลพัฒนาการตัวแปรแฝงของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ของข้อมูลวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2)

ค่าสถิติ	NSB	LIN	FIX	FRE	UDV
dichotomous					
χ^2	437.0513	63.0241	230.0805	292.8251	13.8025
P	0.0000	0.0000	0.0000	0.0000	0.1295
df	18	14	16	11	9
RMR	14.6863	3.8489	34.8659	11.1722	0.9097
GFI	0.6669	0.9464	0.8624	0.7747	0.9715
LSR	5.1087	5.0921	0.0000	-	2.6901
partial credit					
χ^2	2610.3476	61.4955	574.1280	37.3422	30.1783
P	0.0000	0.0000	0.0000	0.0001	0.0004
df	18	14	16	11	9
RMR	229.7677	19.9125	300.2399	8.6920	7.0382
GFI	0.3730	0.9523	0.7615	0.9702	0.9750
LSR	0.0000	4.0928	-	5.1364	4.9838

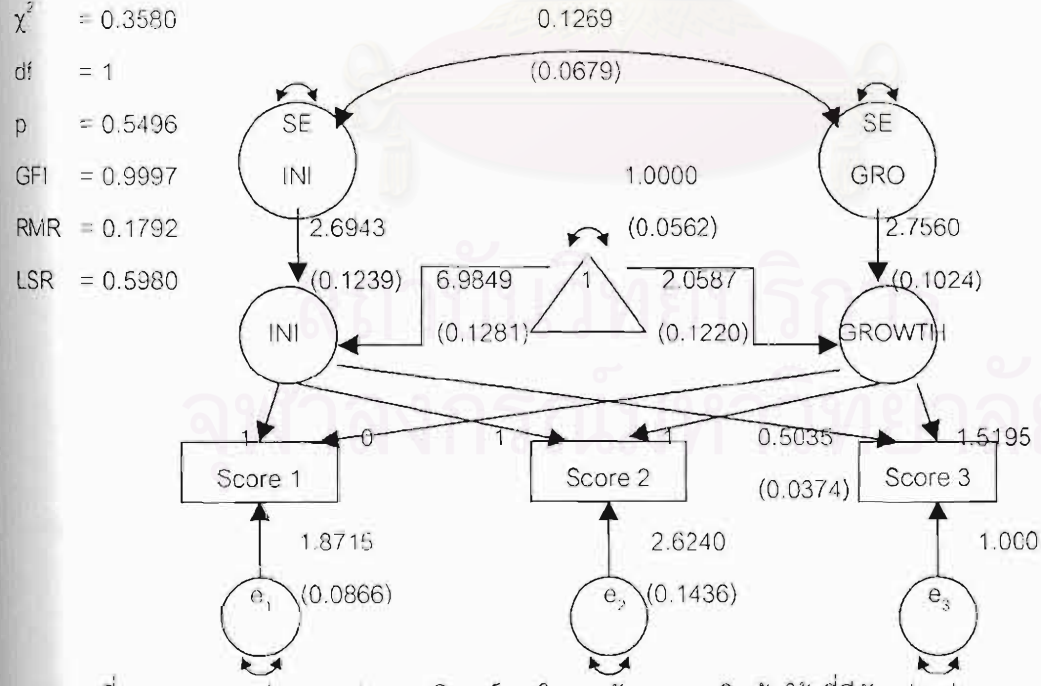
2.2 ผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียน

หลังจากการตรวจสอบโมเดลพัฒนาการตัวแปรแฝงแล้ว พบว่าโมเดลที่สอดคล้องกับข้อมูลมากที่สุดคือโมเดลใดแล้ว ผู้วิจัยจะนำผลการประมาณค่าพารามิเตอร์ของโมเดลนั้นมาประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียน โดยมีรายละเอียดสำหรับข้อมูลแต่ละชุดดังนี้

2.2.1 ผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) แยกตามวิธีการให้คะแนนดังนี้

2.2.1.1 ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) จากการตรวจสอบความสอดคล้องกลมกลืนของโมเดลทั้ง 5 โมเดลกับข้อมูล พบว่าโมเดล UDV เป็นโมเดลเดียวที่มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ โดยมีค่า χ^2= 0.3580, df = 1 และ p = 0.5496 มีผลการประมาณค่าพารามิเตอร์ที่เป็นค่าเฉลี่ยของคะแนนการวัดครั้งแรกเท่ากับ 6.9849 (SE = 0.1281, t = 54.5307) ค่าเฉลี่ยของอัตราพัฒนาการมีค่าเท่ากับ 2.0587 (SE = 0.1220, t = 16.8795) มีอัตราพัฒนาการในครั้งที่ 1 และ 3 เป็นค่าคงที่เท่ากับ 0 และ 1.5195 ตามลำดับ และอัตราพัฒนาการในการวัดครั้งที่ 2 ประมาณค่าได้เท่ากับ 0.5035 (SE = 0.0374, t = 13.4760) (รายละเอียดดังแผนภาพที่ 4.15) และมีผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนเป็นรายบุคคล ดังสมการ

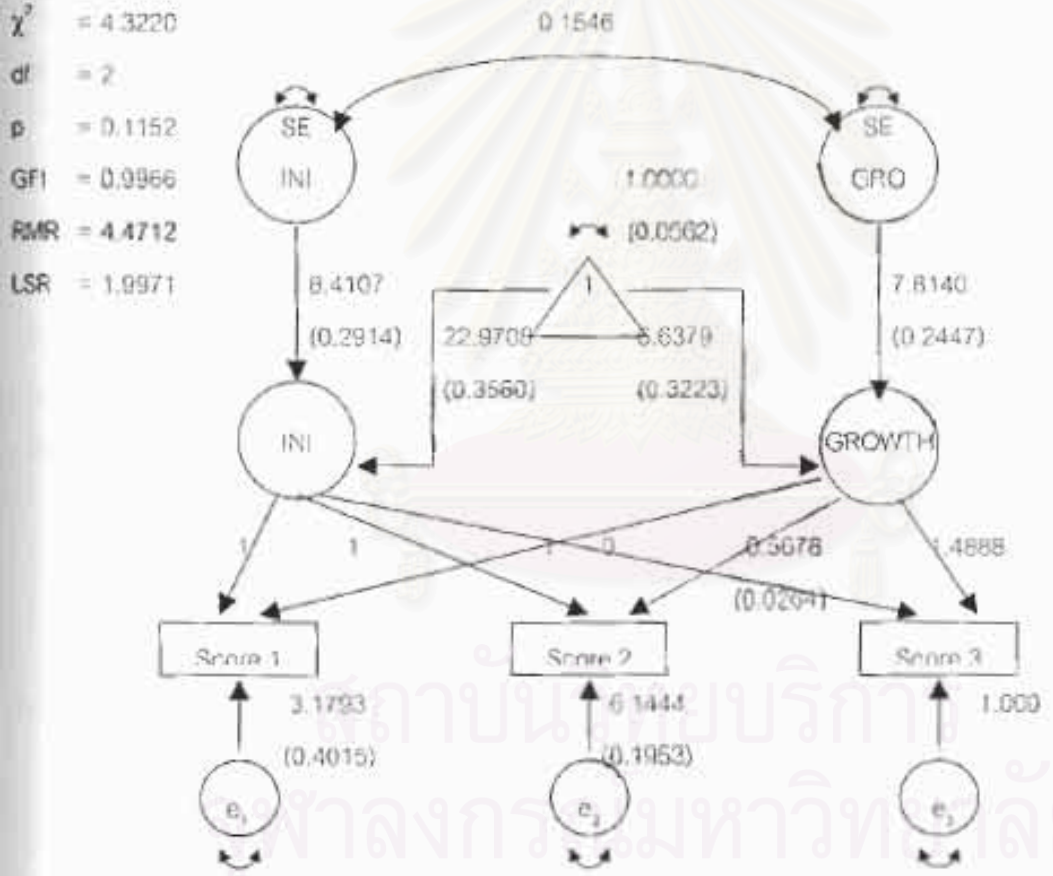
$$GROWTH = -0.3182\ SCORE1 - 0.0806\ SCORE2 + 0.5742\ SCORE3$$



แผนภาพที่ 4.15 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) และมีการให้คะแนนแบบทวิวิภาค (dichotomous)

2.2.1.2 **ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน** (partial credit) **ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลทั้ง 5 โมเดลกับข้อมูล** พบว่าโมเดล UDV เป็นโมเดลเดียว**ที่มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์** โดยมีค่า χ^2 = 4.3220, df = 2 **และ** p = 0.1152 **มีผลการประมาณค่าพารามิเตอร์ที่เป็นค่าเฉลี่ยของคะแนนการวัดครั้งแรกเท่ากับ** 22.9708 (SE = 0.3560, t = 64.5257) **ค่าเฉลี่ยของอัตราพัฒนาการมีค่าเท่ากับ** 6.6379 (SE = 0.3223, t = 20.5935) **มีอัตราพัฒนาการในการวัดครั้งที่ 1 และ 3 เป็นค่าคงที่เท่ากับ** 0 **และ** 1.488 **ตามลำดับ และอัตราพัฒนาการในการวัดครั้งที่** 2 ประมาณค่าได้เท่ากับ 0.5678 (SE = 0.0264, t = 21.4984) **(รายละเอียดดังแผนภาพที่ 4.16) และมีผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนเป็นรายบุคคล ดังสมการ**

GROWTH = -0.4932 SCORE 1 – 0.0751 SCORE 2 + 0.6496 SCORE 3

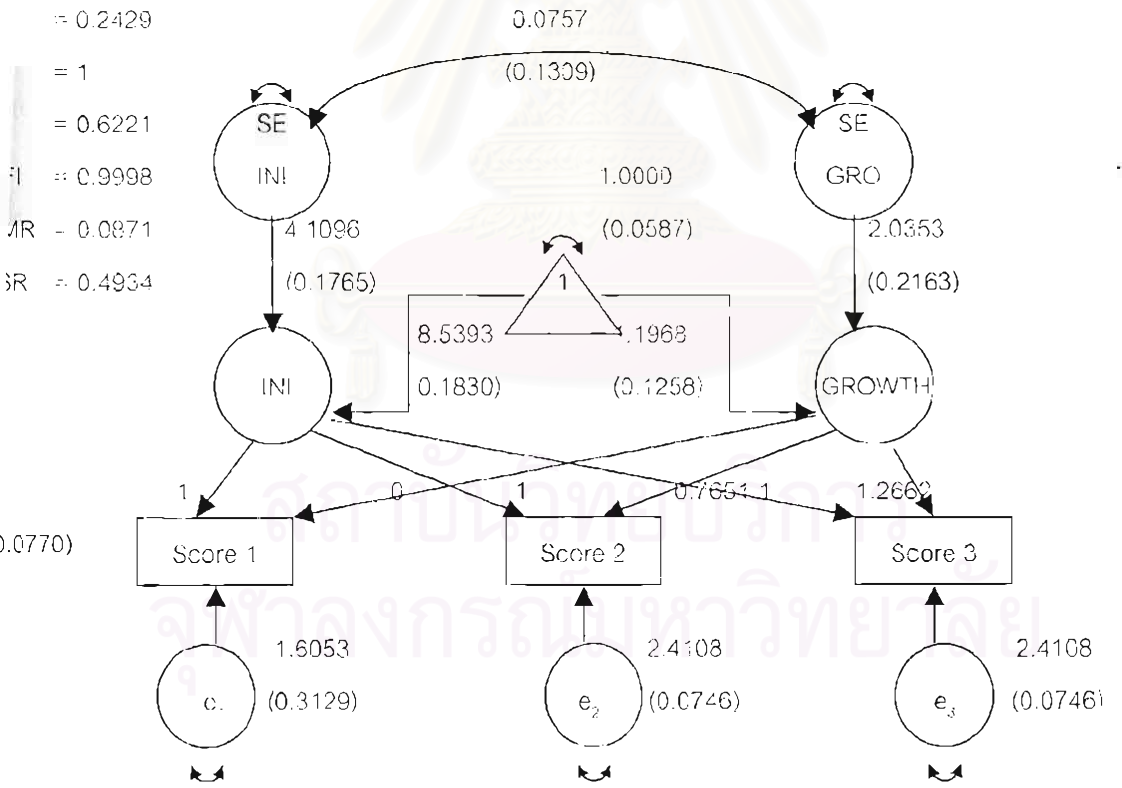


แผนภาพที่ 4.16 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงรองผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) และมีการให้คะแนนความรู้บางส่วน (partial credit)

2.2.2 การประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนที่ได้การวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) แยกตามวิธีการให้คะแนนดังนี้

2.2.2.1 ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) จากการทดสอบความสอดคล้องกลมกลืนของโมเดลทั้ง 5 โมเดลกับข้อมูล พบว่าโมเดล UDV เป็นโมเดลที่มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ โดยมีค่า χ^2= 0.2429, df = 1 และ 0.6221 มีผลการประมาณค่าพารามิเตอร์ที่เป็นค่าเฉลี่ยของคะแนนการวัดครั้งแรกเท่ากับ 693 (SE = 0.1830, t = 46.6731) ค่าเฉลี่ยของอัตราพัฒนาการมีค่าเท่ากับ 1.1968 = 0.1258, t = 9.5107) มีอัตราพัฒนาการในครั้งที่ 1 และ 3 เป็นค่าคงที่เท่ากับ 0 และ 362 ตามลำดับ และอัตราพัฒนาการในการวัดครั้งที่ 2 ประมาณค่าได้เท่ากับ 0.7651 = 0.0770, t = 9.9387) (รายละเอียดดังแผนภาพที่ 4.17) และมีผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนเป็นรายบุคคล ดังสมการ

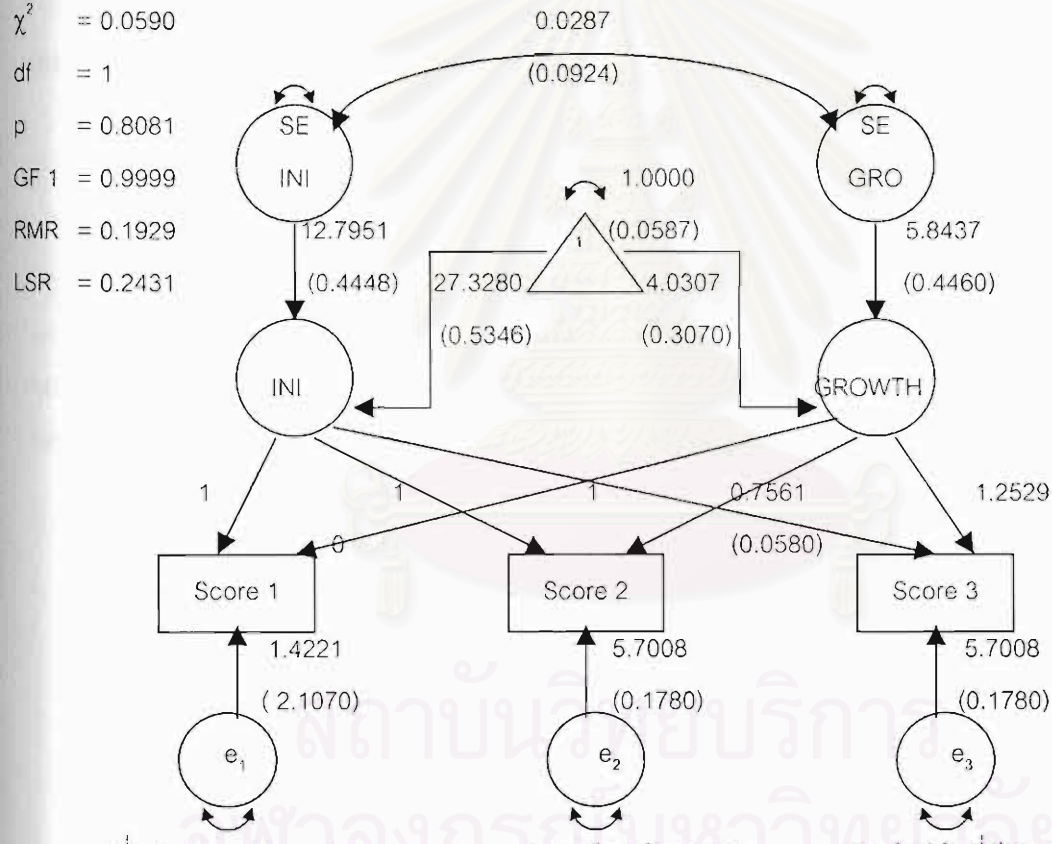
GROWTH = -0.3527 SCORE 1 + 0.1219 SCORE 2 + 0.3042 SCORE 3



ผนภาพที่ 4.17 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) และมีการให้คะแนนแบบทวิวิภาค (dichotomous)

2.2.2.2 ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลทั้ง 5 โมเดลกับข้อมูล พบว่าโมเดล UDV เป็นโมเดลเดียวที่มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ โดยมีค่า χ^2= 0.0590, df = 1 และ p = 0.8081 มีผลการประมาณค่าพารามิเตอร์ที่เป็นค่าเฉลี่ยของคะแนนการวัดครั้งแรกเท่ากับ 27.3280 (SE = 0.5346, t = 51.1233) ค่าเฉลี่ยของอัตราพัฒนาการมีค่าเท่ากับ 4.0307 (SE = 0.3070, t = 13.1277) มีอัตราพัฒนาการในการวัดครั้งที่ 1 และ 3 เป็นค่าคงที่เท่ากับ 0 และ1.2529 ตามลำดับ และอัตราพัฒนาการในการวัดครั้งที่ 2 ประมาณค่าได้เท่ากับ 0.7561 (SE = 0.0580, t = 13.0323) (รายละเอียดดังแผนภาพที่4.18) และมีผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนเป็นรายบุคคล ดังสมการ

GROWTH = -0.6107 SCORE 1 + 0.2249 SCORE 2 + 0.3977 SCORE 3

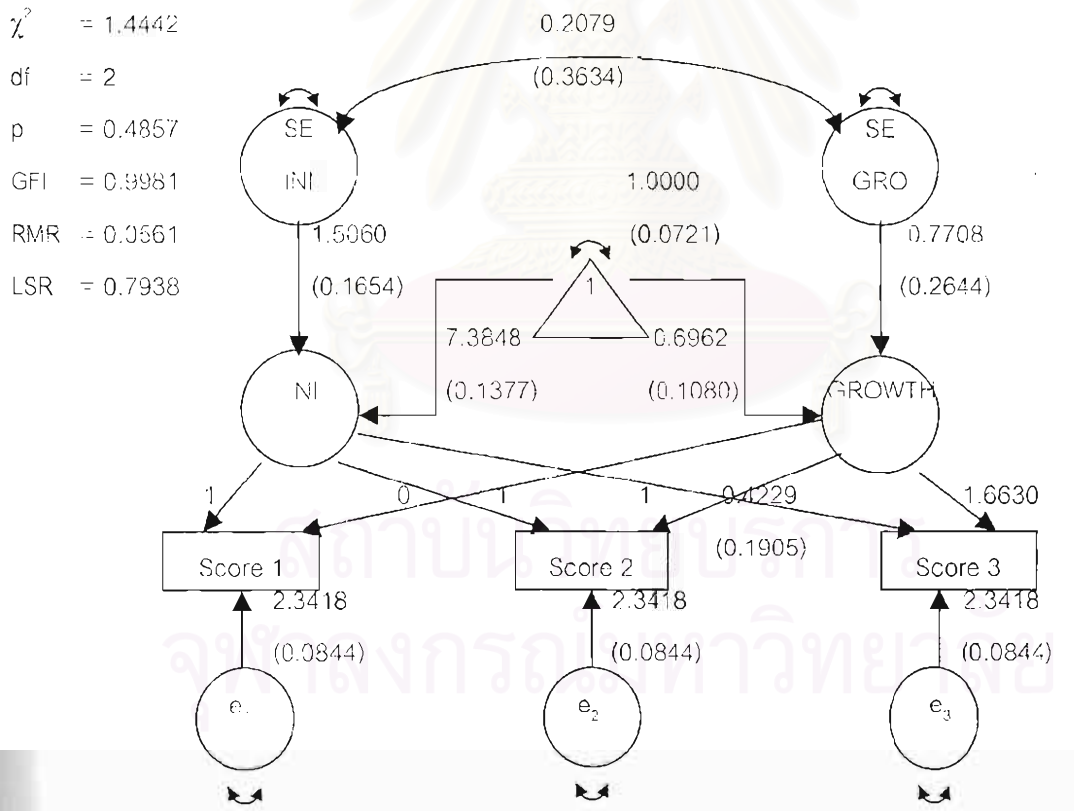


แผนภาพที่ 4.18 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) และมีการให้คะแนนความรู้บางส่วน (partial credit)

2.2.3 ผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนที่ได้จากการวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) แยกตามวิธีการให้คะแนนดังนี้

2.2.3.1 ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) จากการตรวจสอบความสอดคล้องกลมกลืนของโมเดลทั้ง 5 โมเดลกับข้อมูล พบว่าโมเดล FRE มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์มากกว่าโมเดลอื่นๆ เมื่อพิจารณาจากค่า p, RMR, GFI และ LSR โมเดล FRE มีค่า χ^2= 1.4442, df = 2 และ p = 0.4857 มีผลการประมาณค่าพารามิเตอร์ที่เป็นค่าเฉลี่ยของคะแนนการวัดครั้งแรกเท่ากับ 7.3848 (SE = 0.1377, t = 53.6154) ค่าเฉลี่ยของอัตราพัฒนาการมีค่าเท่ากับ 0.6962 (SE = 0.1080, t = 6.4468) โดยมีอัตราพัฒนาการในครั้งที่ 1 และ 3 เป็นค่าคงที่เท่ากับ 0 และ 1.6630 ตามลำดับ และอัตราพัฒนาการในการวัดครั้งที่ 2 ประมาณค่าได้เท่ากับ 0.4229 (SE = 0.1905, t = 2.2196) (รายละเอียดดังแผนภาพที่ 4.19) และมีผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนเป็นรายบุคคล ดังสมการ

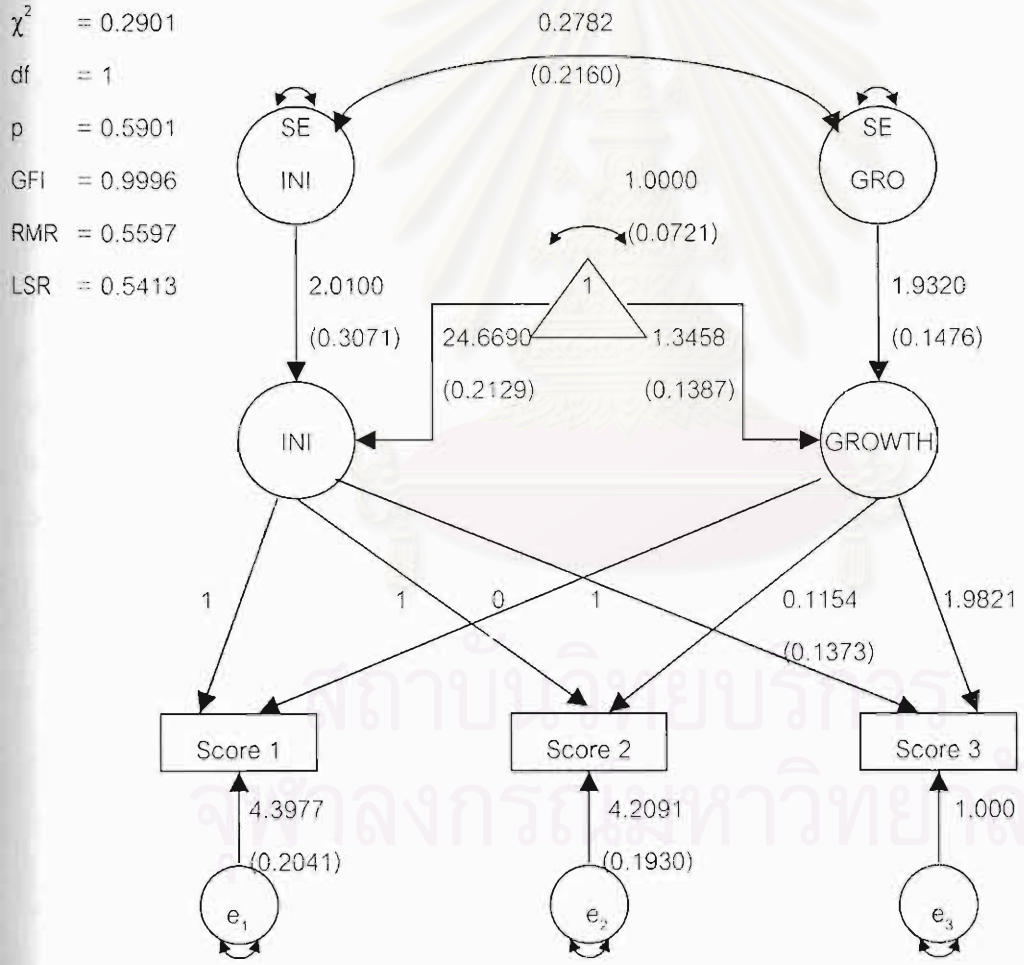
GROWTH = -0.0155 SCORE 1 + 0.0187 SCORE 2 + 0.1188 SCORE 3



แผนภาพที่ 4.19 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) และมีการให้คะแนนแบบทวิวิภาค (dichotomous)

2.2.3.2 ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลทั้ง 5 โมเดลกับข้อมูล พบว่าโมเดล UDV มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์มากกว่าโมเดลอื่นๆ เมื่อพิจารณาจากค่า p และ LSR โมเดล UDV มีค่า χ^2= 0.2901, df = 1 และ p = 0.5901 มีผลการประมาณค่าพารามิเตอร์ที่เป็นค่าเฉลี่ยของคะแนนการวัดครั้งแรกเท่ากับ 24.6690 (SE = 0.2129, t = 115.8715) ค่าเฉลี่ยของอัตราพัฒนาการมีค่าเท่ากับ 1.3458 (SE = 0.1387, t = 9.7010) มีอัตราพัฒนาการในการวัดครั้งที่ 1 และ 3 เป็นค่าคงที่เท่ากับ 0 และ 1.9821 ตามลำดับ และอัตราพัฒนาการในการวัดครั้งที่ 2 ประมาณค่าได้เท่ากับ 0.1154 (SE = 0.1373, t = 0.8409) (รายละเอียดดังแผนภาพที่ 4.20) และมีผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนเป็นรายบุคคล ดังสมการ

GROWTH = -0.0458 SCORE 1 – 0.0459 SCORE 2 + 0.3790 SCORE 3

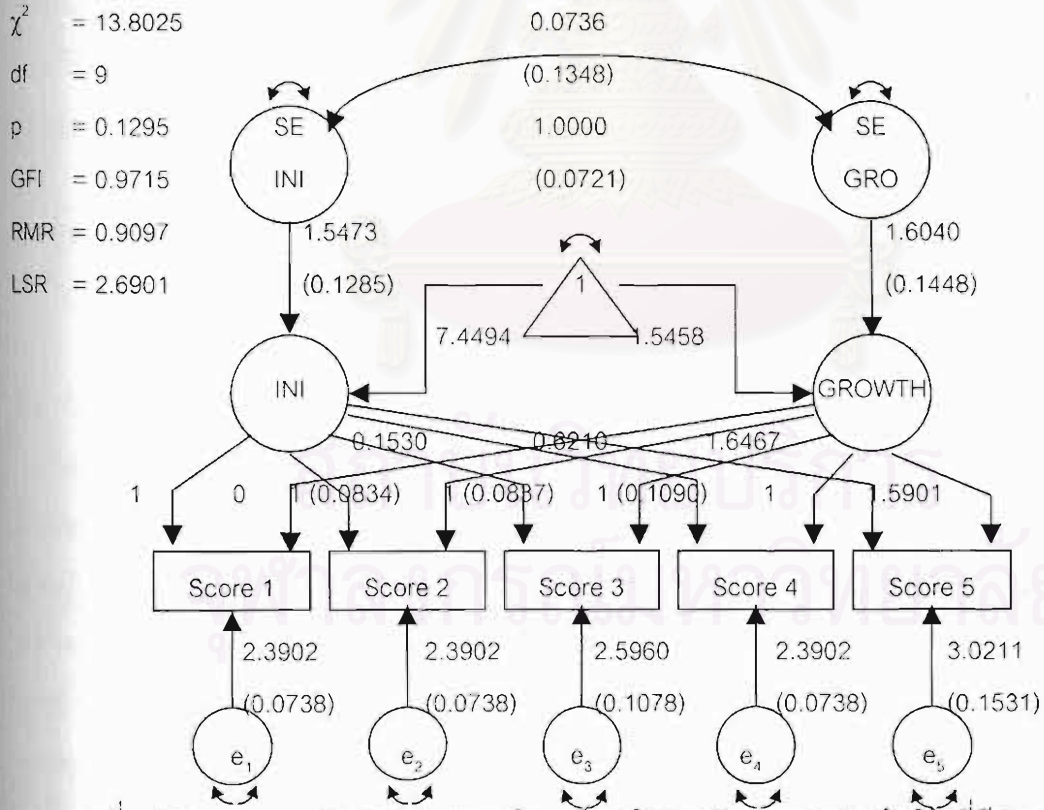


แผนภาพที่ 4.20 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) และมีการให้คะแนนความรู้บางส่วน (partial credit)

2.2.4 ผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนที่ได้จากการวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) แยกตามวิธีการให้คะแนนดังนี้

2.2.4.1 ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) จากการตรวจสอบความสอดคล้องกลมกลืนของโมเดลทั้ง 5 โมเดลกับข้อมูล พบว่าโมเดล UDV เป็นโมเดลเดียวที่มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ โดยมีค่า χ^2= 13.8025, df = 9 และ p = 0.1295 มีผลการประมาณค่าพารามิเตอร์ที่เป็นค่าเฉลี่ยของคะแนนการวัดครั้งแรกเท่ากับ 7.4494 (SE = 0.1343, t = 55.4607) ค่าเฉลี่ยของอัตราพัฒนาการมีค่าเท่ากับ 1.5458 (SE = 0.1385, t = 11.1642) มีอัตราพัฒนาการในครั้งที่ 1 และ 5 มีค่าคงที่คือ 0 และ 1.5901 ตามลำดับ ส่วนอัตราพัฒนาการในการวัดครั้งที่ 2, 3 และ 4 ประมาณค่าได้เท่ากับ 0.1530 (SE = 0.0834, t = 1.8340), 0.6210 (SE = 0.0837, t = 7.4183) และ 1.6467 (SE = 0.1090, t = 15.1133) ตามลำดับ (รายละเอียดดังแผนภาพที่ 4.21) และมีผลการประมาณค่าคะแนนพัฒนาการทางคณิตศาสตร์ของนักเรียนเป็นรายบุคคล ดังสมการ

GROWTH = -0.0844 SCORE 1 - 0.0562 SCORE 2 + 0.0254 SCORE 3
+ 0.2187 SCORE 4 + 0.1304 SCORE 5



แผนภาพที่ 4.21 ผลการประมาณค่าพารามิเตอร์ของโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงของผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)

2.2.4.2 ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลทั้ง 5 โมเดลกับข้อมูล พบว่าโมเดล UDV มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์มากกว่าโมเดลอื่น ๆ แต่โมเดล UDV ไม่สอดคล้องกับข้อมูลเชิงประจักษ์อย่างมีนัยสำคัญ เมื่อพิจารณาจากค่า p มีค่าเท่ากับ 0.004 และ RMR พบว่ามีค่าเกิน 1.00 (7.0382) และค่า LSR มีค่าเกิน 2.00 (4.9839) ซึ่งแสดงว่ายังสามารถที่จะปรับหาโมเดลอื่นที่มีความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์มากกว่าโมเดล UDV ดังนั้นผู้วิจัยจึงไม่ได้หาคะแนนพัฒนาการรายบุคคลของข้อมูลชุดนี้

ตอนที่ 3 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลของแต่ละกลุ่มวิธีและวิธีที่ใช้เป็นเกณฑ์

ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคล จะนำเสนอแยกตามกลุ่มวิธีที่ใช้หาคะแนนพัฒนาการ ได้แก่กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยแยกตามข้อมูลที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) วัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) และ วัดซ้ำด้วยแบบทดสอบคู่ขนาน (equivalence) และแยกตามวิธีการให้คะแนนแบบทวิวิภาค (dichotomous) และให้คะแนนความรู้บางส่วน (partial credit) ซึ่งมีรายละเอียดดังนี้

3.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ซึ่งแยกตามวิธีการให้คะแนนดังนี้

3.1.1 ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)

ผลการวิเคราะห์ค่าสถิติพื้นฐาน พบว่าคะแนนพัฒนาการที่ได้จากวิธีเกณฑ์ คือ GS มีค่าเฉลี่ยเท่ากับ 2.9400 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.5469 มีค่าต่ำสุดเท่ากับ - 2.5836 ค่าสูงสุดเท่ากับ 10.1506 คะแนน GS มีลักษณะการแจกแจงแบบเบ้ขวา (Sk = 0.5122) มีลักษณะแบนกว่าโค้งปกติเล็กน้อย (Ku = -0.1906) ส่วนคะแนนพัฒนาการที่ได้จากกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่ามีค่าเฉลี่ยอยู่ระหว่าง 0 ถึง 21.6724 ส่วนเบี่ยงเบนมาตรฐานมีค่าอยู่ระหว่าง 0.6131 ถึง 70.3620 ค่าต่ำสุดอยู่ระหว่าง - 87.5000 ถึง -1.9459 มีค่าสูงสุดอยู่ระหว่าง 2.8332 ถึง 1124.2540 เมื่อพิจารณารูปแบบโค้งการแจกแจงพบว่า คะแนนพัฒนาการจากทุกวิธีมีลักษณะการแจกแจงแบบเบ้ขวา โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีลักษณะการแจกแจงเบ้ขวามากที่สุด (Sk = 8.1221) รองลงมาคือวิธีการวัด DS, SR, SC และ NL ตามลำดับ ส่วนลักษณะของความโด่งของลักษณะการแจกแจงมีความโด่งกว่าโค้งปกติ โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีลักษณะโด่งที่สุด (Ku = 109.5332) รองลงมาได้แก่วิธีการวัด NL, SC, DS และ SR ตามลำดับ เมื่อเทียบกับวิธีเกณฑ์แล้วพบว่าคะแนนพัฒนาการจากวิธีการวัด DS

มีลักษณะการแจกแจงแบบเบ้ขวาใกล้เคียงกับ GS มากที่สุด (Sk = 0.5388) และมีลักษณะความโด่งสูงกว่าโด่งปกติเล็กน้อย (Ku = 0.2604) ซึ่งถือได้ว่ามีลักษณะการแจกแจงคล้ายกับ GS มากกว่าคะแนนพัฒนาการจากวิธีการวัดวิธีอื่น ๆ สำหรับคะแนนพัฒนาการจากกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่าทั้งคะแนนพัฒนาการจากวิธีการวัด Lθ และ MRMLC มีค่าเฉลี่ยเท่ากับ 0 และ 0.6811 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ0.6862 และ 1.1256 ค่าต่ำสุดเป็น - 2.1357 และ - 2.2851 และมีค่าสูงสุดเป็น 2.5454 และ 4.8703 ตามลำดับ และมีลักษณะการแจกแจงแบบเบ้ขวาใกล้เคียงกับ GS (Sk = 0.4884 และ 0.6714 ตามลำดับ) โดยคะแนนพัฒนาการจากทั้ง 2 วิธี มีลักษณะของความโด่งสูงกว่าโด่งปกติ (Ku = 0.7881 และ 0.7947 ตามลำดับ) ซึ่งถือได้ว่ามีลักษณะการแจกแจงค่อนข้างต่างจาก GS ส่วนในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา คะแนนพัฒนาการจากวิธีการวัด Sθ และ Vθ มีค่าเฉลี่ยเท่ากับ - 0.0125 และ 0.0063 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ0.2024 และ 0.1710 มีค่าต่ำสุดเป็น -1.1302 และ - 0.4074 และมีค่าสูงสุดเป็น 0.5659 และ 0.7559 ตามลำดับ คะแนนพัฒนาการจากวิธีการวัด Sθ มีลักษณะการแจกแจงแบบเบ้ซ้าย และมีลักษณะของความโด่งสูงกว่าโด่งปกติค่อนข้างมาก ส่วนคะแนนพัฒนาการจากวิธีการวัด Vθ มีลักษณะการแจกแจงแบบเบ้ขวา และมีลักษณะของความโด่งสูงกว่ากว่าโด่งปกติเล็กน้อย

โดยสรุป คะแนนพัฒนาการจากวิธีการวัด DS มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการจากวิธีการวัด GS มากที่สุด รองลงมาคือ คะแนนพัฒนาการจากวิธีการวัด SR (รายละเอียดดังตารางที่ 4.17)

ตารางที่ 4.17 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการให้คะแนนแบบทวิวิภาค(dichotomous)

คะแนนพัฒนาการ	N	Mean	Median	Mode	SD	Variance	Skewness	Kurtosis	Minimum	Maximum
วิธีที่เป็นเกณฑ์										
GS	635	2.9400	2.7556	0.8770	2.5469	6.4869	0.5182	-0.1906	-2.5836	10.1506
วิธี CTT										
DS	635	3.1228	3.0000	3.0000	4.7043	22.1300	0.5388	0.2604	-8.0000	18.0000
SC	635	0.0000	-0.1064	-0.3473	1.0083	1.0166	0.3518	0.3342	-2.7223	3.2532
NL	634	0.3329	0.3365	0.0000	0.6131	0.3759	0.1290	0.9415	-1.9459	2.8332
SR	635	21.6724	16.7500	0.0000	34.5272	1192.1276	0.3970	0.1313	-87.5000	100.0000
SRR	635	8.4482	-6.6844	-29.5508	70.3620	4950.8084	8.1221	109.5332	-66.3529	1124.2540
วิธี IRT										
Lθ	635	0.0000	-0.0940	-2.1357	0.6862	0.4709	0.4887	0.7881	-2.1357	2.5454
MRMLC	615	0.6811	0.5766	0.0000	1.1256	1.2670	0.6714	0.7947	-2.2851	4.8703
วิธีที่พัฒนาขึ้น										
Sθ	635	-0.0125	-0.0217	-1.1302	0.2024	0.0410	-0.7142	3.8632	-1.1302	0.5659
Vθ	635	0.0063	-0.0247	-0.4074	0.1710	0.0292	0.8417	1.1136	-0.4074	0.7559

3.1.2 ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit)

ผลการวิเคราะห์ค่าสถิติพื้นฐาน พบว่าคะแนนพัฒนาการที่ได้จากวิธีเกณฑ์คือ GS มีค่าเฉลี่ยเท่ากับ 8.0041 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 7.5665 มีค่าต่ำสุดเท่ากับ 6.3101 ค่าสูงสุดเท่ากับ 31.8971 คะแนน GS มีลักษณะการแจกแจงแบบเบ้ขวา (Sk = 0.7101) มีความโด่งใกล้เคียงกับโค้งปกติ (Ku = 0.0242) ส่วนคะแนนพัฒนาการจากกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่ามีค่าเฉลี่ยอยู่ระหว่าง 0 ถึง 25.1715 ส่วนเบี่ยงเบนมาตรฐานมีค่าอยู่ระหว่าง 0.4194 ถึง 143.2436 มีค่าต่ำสุดอยู่ระหว่าง - 52.0576 ถึง - 0.7802 มีค่าสูงสุดอยู่ระหว่าง 2.1203 ถึง 3282.0804 และมีรูปแบบโค้งของลักษณะการแจกแจงแบบเบ้ขวา โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีลักษณะการแจกแจงเบ้ขวามากที่สุด (Sk = 19.1905) รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด DS, SR, SC และ NL ตามลำดับ ส่วนลักษณะของความโด่ง คะแนนพัฒนาการจากทุกวิธีมีความโด่งกว่าโค้งปกติ ยกเว้นคะแนนพัฒนาการจากวิธีการวัด SR ที่มีความโด่งใกล้เคียงกับโค้งปกติ (Ku = - 0.0424) โดย คะแนนพัฒนาการจากวิธีการวัด SSR มีความโด่งสูงกว่าโค้งปกติมากที่สุด (Ku = 430.4999) รองลงมาได้แก่คะแนนพัฒนาการจากวิธีการวัด NL, SC และ DS ตามลำดับ เมื่อเทียบกับวิธีเกณฑ์คะแนนพัฒนาการจากวิธีการวัด SR มีลักษณะการแจกแจงใกล้เคียงกับ GS มากที่สุด (Sk = 0.7726, Ku = 0.0427) สำหรับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบพบว่าคะแนนพัฒนาการจากวิธีการวัด Lθ มีค่าเฉลี่ยเท่ากับ -0.1060 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 1.1327 มีค่าต่ำสุดเป็น -2.6526 และค่าสูงสุดเป็น 2.9407 มีลักษณะการแจกแจงเบ้ขวา (Sk = 0.2335) และมีความโด่งแบนกว่าโค้งปกติ (Ku = -0.5312) ซึ่งมีลักษณะการแจกแจงต่างจาก GS ส่วนกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า คะแนนพัฒนาการจากวิธีการวัด Sθ และ Vθ มีค่าเฉลี่ยเท่ากับ -0.0312 และ -0.0206 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.3709 และ 0.2388 มีค่าต่ำสุดเป็น -1.0972 และ -0.5319 มีค่าสูงสุดเป็น 0.8116 และ 0.7092 ตามลำดับ คะแนนพัฒนาการจากวิธีการวัด Sθ มีลักษณะการแจกแจงค่อนข้างปกติ มีความโด่งแบนกว่าโค้งปกติ คะแนนพัฒนาการจากวิธีการวัด Vθ มีลักษณะการแจกแจงแบบเบ้ขวา มีความโด่งแบนกว่าโค้งปกติ

โดยสรุป คะแนนพัฒนาการจากวิธีการวัด SR มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการจากวิธีการวัด GS มากที่สุด รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด DS (รายละเอียดดังตารางที่ 4.18)

ตารางที่ 4.18 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการให้คะแนนความรู้บางส่วน (partial credit)

คะแนนพัฒนาการ	N	Mean	Median	Mode	SD	Variance	Skewness	Kurtosis	Minimum	Maximum
วิธีที่เป็นเกณฑ์										
GS	635	8.0041	6.6029	2.3507	7.5665	57.2519	0.7101	0.0242	-6.3101	31.8971
วิธี CTT										
DS	635	9.8299	8.0000	3.0000	12.1408	147.4000	0.7749	0.3001	-15.0000	50.0000
SC	635	0.0000	-0.1200	-0.2086	0.8760	0.7674	0.5762	0.6244	-2.5347	2.9857
NL	635	0.3192	0.2877	0.0000	0.4194	0.1759	0.4316	0.8347	-0.7802	2.1203
SR	635	25.1715	17.3913	0.0000	31.8453	1014.1240	0.7726	-0.0424	-50.0000	100.0000
SRR	635	13.3923	-8.3336	-24.6210	143.2436	20518.7248	19.1965	430.4999	-52.0576	3282.0804
วิธี IRT										
L0	635	-0.1060	-0.1343	-0.7107	1.1327	1.2830	0.2335	-0.5312	-2.6526	2.9407
วิธีที่พัฒนา										
S0	635	-0.0312	-0.0380	-1.0972	0.3709	0.1376	0.0990	-0.3365	-1.0972	0.8116
V0	635	-0.0206	-0.0305	-0.5319	0.2388	0.0570	0.3545	-0.4400	-0.5319	0.7092

3.2 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ซึ่งแยกวิธีการให้คะแนนดังนี้

3.2.1 ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)

ผลการวิเคราะห์ค่าสถิติพื้นฐาน พบว่าคะแนนพัฒนาการที่ได้จากวิธีเกณฑ์ คือ GS มีค่าเฉลี่ยเท่ากับ 1.1994 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.4241 มีค่าต่ำสุดเท่ากับ 2.7588 ค่าสูงสุดเท่ากับ 5.4425 คะแนน GS มีลักษณะการแจกแจงแบบเบ้ขวา (Sk = 0.2386) มีลักษณะแบนกว่าโค้งปกติ (Ku = -0.3520) ส่วนคะแนนพัฒนาการจากกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมพบว่ามีค่าเฉลี่ยอยู่ระหว่าง 0 ถึง 24.8943 ส่วนเบี่ยงเบนมาตรฐานมีค่าอยู่ระหว่าง 0.5244 ถึง 109.4771 มีค่าต่ำสุดอยู่ระหว่าง -200.0000 ถึง -1.7918 มีค่าสูงสุดอยู่ระหว่าง 1.7047 ถึง 1143.1193 คะแนนพัฒนาการจากวิธีการวัด DS และ SRR มีลักษณะการแจกแจงแบบเบ้ขวา โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีลักษณะการแจกแจงเบ้ขวามากที่สุด (Sk = 4.5348) รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด DS (Sk = 0.1186) ส่วนคะแนนพัฒนาการจากวิธีการวัด SC มีลักษณะการแจกแจงใกล้เคียงกับโค้งปกติ (Sk = 0.0351) ส่วนคะแนนพัฒนาการจากวิธีการวัด NL และ SR มีลักษณะการแจกแจงแบบเบ้ซ้าย โดยคะแนนพัฒนาการจากวิธีการวัด SR เบ้ซ้ายมากกว่า NL ส่วนลักษณะของความโด่ง พบว่าคะแนนพัฒนาการจากทุกวิธีมีลักษณะโด่งสูงกว่าโค้งปกติ โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีลักษณะโด่งที่สุด (Ku = 31.1927) รองลงมาได้แก่คะแนนพัฒนาการจากวิธีการวัด SR, NL และ

SC ตามลำดับ ส่วนคะแนนพัฒนาการจากวิธีการวัด DS มีความโด่งใกล้เคียงกับโค้งปกติ สำหรับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบพบว่าคะแนนพัฒนาการจากวิธีการวัด Lθ และ MRMLC มีค่าเฉลี่ยเท่ากับ 0 และ 0.3107 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.6862 และ 0.9017 มีค่าต่ำสุดเป็น-2.5742 และ -2.5068 และมีค่าสูงสุดเป็น 2.3773 และ 2.8707 ตามลำดับ โดยคะแนนพัฒนาการจากวิธีการวัด Lθ มีลักษณะการแจกแจงค่อนข้างเป็นโค้งปกติ (Sk = -0.0969, Ku = 0.0269) ส่วนคะแนนพัฒนาการจากวิธีการวัด MRMLC มีลักษณะการแจกแจงแบบเบ้ขวาแบนกว่าโค้งปกติเล็กน้อย สำหรับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา คะแนนพัฒนาการจากวิธีการวัด Sθ และ Vθ มีค่าเฉลี่ยเท่ากับ - 0.0131 และ 0.0085 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.2145 และ 0.1696 มีค่าต่ำสุดเป็น - 1.0529 และ - 0.4634 มีค่าสูงสุดเป็น 0.5912 และ 0.5975 ตามลำดับ คะแนนพัฒนาการจากวิธีการวัด Sθ มีลักษณะการแจกแจงแบบเบ้ซ้าย และมีลักษณะโด่งกว่าโค้งปกติ ส่วนคะแนนพัฒนาการจากวิธีการวัด Vθ มีลักษณะการแจกแจงแบบเบ้ขวา และมีลักษณะโด่งแบนกว่าโค้งปกติ

โดยสรุป คะแนนพัฒนาการจากวิธีการวัด Vθ มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการจากวิธีการวัด GS มากที่สุด รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด MRMLC (รายละเอียดดังตารางที่ 4.19)

ตารางที่ 4.19 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการให้คะแนนแบบทวิวิภาค (dichotomous)

คะแนนพัฒนาการ	N	Mean	Median	Mode	SD	Variance	Skewness	Kurtosis	Minimum	Maximum
วิธีที่เป็นเกณฑ์										
GS	581	1.1994	1.0488	0.7446	1.4241	2.0281	0.2386	-0.3520	-2.7588	5.4426
วิธี CTT										
DS	581	1.5267	1.0000	0.0000	3.8678	14.9601	0.1186	0.0594	-12.0000	15.0000
SC	581	0.0000	0.0063	-0.7912	0.7479	0.5593	0.0351	0.2491	-2.7866	2.5020
NL	581	0.1422	0.1542	0.0000	0.5244	0.2749	-0.1175	0.7694	-1.7918	1.7047
SR	581	13.3271	10.5263	0.0000	39.3763	1550.4898	-0.1868	2.2790	200.0000	100.0000
SRR	581	24.8943	-10.0091	-36.3359	109.4771	11985.2330	4.5348	31.1927	-106.3918	1143.1193
วิธี IRT										
Lθ	581	0.0000	-0.0415	-2.5742	0.6862	0.4709	-0.0969	0.0269	-2.5742	2.3773
MRMLC	550	0.3107	0.2456	0.0000	0.9017	0.8130	0.1179	-0.1541	-2.5068	2.8707
วิธีที่พัฒนา										
Sθ	581	-0.0131	-0.0116	-1.0529	0.2145	0.0460	-1.0050	2.9151	-1.0529	0.5912
Vθ	581	0.0085	-0.0100	-0.4634	0.1696	0.0288	0.3074	-0.3897	-0.4634	0.5975

3.2.2 **ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน** (partial credit)

ผลการวิเคราะห์ค่าสถิติพื้นฐาน พบว่า คะแนนพัฒนาการที่ได้จากวิธีเกณฑ์คือ GS มีค่าเฉลี่ยเท่ากับ 3.0191 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4.7847 มีค่าต่ำสุดเท่ากับ -13.1232 ค่าสูงสุดเท่ากับ 18.5019 คะแนน GS มีลักษณะการแจกแจงแบบเบ้ขวา (Sk= 0.2730) มีความโด่งสูงกว่าโค้งปกติ (Ku = 0.5119) ส่วนคะแนนพัฒนาการจากกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่ามีค่าเฉลี่ยอยู่ระหว่าง 0 ถึง 28.5681 ส่วนเบี่ยงเบนมาตรฐานมีค่าอยู่ระหว่าง 0.3269 ถึง 110.6482 มีค่าต่ำสุดระหว่าง - 208.3333 ถึง - 0.9933 มีค่าสูงสุดอยู่ระหว่าง 1.1206 ถึง 981.0600 โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีลักษณะการแจกแจงแบบเบ้ขวามากที่สุด (Sk = 3.6976) รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด DS (Sk = 0.2398) ส่วนคะแนนพัฒนาการจากวิธีการวัด SR มีลักษณะการแจกแจงแบบเบ้ซ้ายมากที่สุด (Sk = - 0.5741) รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด NL (Sk = -0.2000) ส่วนคะแนนพัฒนาการจากวิธีการวัด SC มีลักษณะการแจกแจงค่อนข้างเป็นโค้งปกติ (Sk = -0.0322) โดยคะแนนพัฒนาการจากทุกวิธีมีลักษณะการแจกแจงโด่งกว่าโค้งปกติ โดยมีคะแนนพัฒนาการจากวิธีการวัด SRR มีความโด่งมากที่สุด (Ku = 19.0804) รองลงมาได้แก่คะแนนพัฒนาการจากวิธีการวัด SR, SC, DS และ NL ตามลำดับ สำหรับวิธีของทฤษฎีการตอบสนองข้อสอบ คะแนนพัฒนาการจากวิธีการวัด LΘ มีค่าเฉลี่ยเท่ากับ 0.3681 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.6974 มีค่าต่ำสุดเป็น -1.9618 และค่าสูงสุดเป็น 2.3631 มีลักษณะการแจกแจงแบบเบ้ขวา (Sk = 0.1423) และมีความโด่งสูงกว่าโค้งปกติเล็กน้อย (Ku = 0.1859) ส่วนกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา คะแนนพัฒนาการจากวิธีการวัด SΘ และ VΘ มีค่าเท่ากับ 0.0990 และ 0.1004 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.2062 และ 0.1886 มีค่าต่ำสุดเป็น - 0.8349 และ- 0.4549 มีค่าสูงสุดเป็น 0.6795 และ 0.9525 ตามลำดับ คะแนนพัฒนาการจากวิธีการวัดSΘ มีลักษณะการแจกแจงใกล้เคียงกับโค้งปกติ และสูงโด่งกว่าโค้งปกติ ส่วนคะแนนพัฒนาการจากวิธีการวัด VΘ มีลักษณะการแจกแจงแบบเบ้ขวา มีความโด่งสูงกว่าโค้งปกติ

โดยสรุป คะแนนพัฒนาการจากวิธีการวัด DS มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการจากวิธีการวัด GS มากที่สุด รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด LΘ (รายละเอียดดังตารางที่ 4.20)

ตารางที่ 4. 20 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการให้คะแนนความรู้บางส่วน (partial credit)

คะแนนพัฒนาการ	N	Mean	Median	Mode	SD	Variance	Skewness	Kurtosis	Minimum	Maximum
วิธีที่เป็นเกณฑ์										
GS	581	3.0191	2.5749	-2.8155	4.7847	22.8932	0.2730	0.5119	-13.1232	18.5019
วิธี CTT										
DS	581	5.0654	5.0000	-3.0000	9.3788	87.9612	0.2398	0.7243	-30.0000	37.0000
SC	581	0.0000	-0.0169	-0.2757	0.6183	0.3823	-0.0322	1.2386	-2.5232	2.0520
NL	581	0.1525	0.1542	0.0000	0.3269	0.1069	-0.2000	0.5825	-0.9933	1.1206
SR	581	16.3123	10.8696	0.0000	37.5870	1412.7818	-0.5741	5.2697	-208.3333	100.0000
SRR	581	28.5681	-12.7023	418.2025	110.6482	12243.0200	3.6976	19.0804	-79.4110	981.0600
วิธี IRT										
LΘ	581	0.3681	0.3722	-0.3364	0.6974	0.4863	0.1423	0.1859	-1.9618	2.3631
วิธีที่พัฒนา										
SΘ	581	0.0990	0.0821	-0.8349	0.2062	0.0425	0.0574	1.4595	-0.8349	0.6795
VΘ	581	0.1004	0.1009	-0.4549	0.1886	0.0356	0.6605	1.5090	-0.4549	0.9525

3.3 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลที่ได้จากการวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ซึ่งแยกตามวิธีการให้คะแนนดังนี้

3.3.1 ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)

ผลการวิเคราะห์ค่าสถิติพื้นฐาน พบว่าคะแนนพัฒนาการที่ได้จากวิธีเกณฑ์ คือ GS มีค่าเฉลี่ยเท่ากับ 1.0441 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.3909 มีค่าต่ำสุดเท่ากับ 0.1353 ค่าสูงสุดเท่ากับ 2.3795 คะแนน GS มีลักษณะการแจกแจงแบบเบ้ขวา (Sk = 0.4371) มีลักษณะโด่งกว่าโค้งปกติเล็กน้อย (Ku = 0.1483) ส่วนคะแนนพัฒนาการจากกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่ามีค่าเฉลี่ยอยู่ระหว่าง - 0.0169 ถึง 4.8451 ส่วนเบี่ยงเบนมาตรฐานมีค่าอยู่ระหว่าง 0.5242 ถึง 21.0636 มีค่าต่ำสุดอยู่ระหว่าง - 71.4286 ถึง - 2.3979 มีค่าสูงสุดอยู่ระหว่าง 2.4849 ถึง 68.6418 โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีลักษณะการแจกแจงเบ้ขวา คะแนนพัฒนาการจากวิธีการวัด SR และ NL มีลักษณะการแจกแจงแบบเบ้ซ้าย คะแนนพัฒนาการจากวิธีการวัด SRR, SR และ NL มีลักษณะการแจกแจงโด่งกว่าโค้งปกติ โดยคะแนนพัฒนาการจากวิธีการวัด NL มีลักษณะโด่งมากที่สุด (Ku = 2.5623) ส่วนคะแนนพัฒนาการจากวิธีการวัด DS และ SC มีลักษณะการแจกแจงใกล้เคียงกับลักษณะการแจกแจงปกติ (Sk = 0.0683, 0.0204; Ku = 0.0432, - 0.0086 ตามลำดับ) สำหรับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบพบว่าคะแนนพัฒนาการจากวิธีการวัด LΘ และ MRMLC มีค่าเฉลี่ยเท่ากับ 0.0016 และ 0.2005 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.8553 และ 0.7641 มีค่าต่ำสุดเป็น - 2.5213 และ - 3.0215 และมีค่าสูงสุดเป็น 2.4378 และ 3.1908 ตามลำดับ โดยคะแนนพัฒนาการจากวิธี

การวัด Lθ มีลักษณะการแจกแจงใกล้เคียงกับลักษณะการแจกแจงปกติ (Sk = - 0.0424, Ku = 0.0968) ส่วนคะแนนพัฒนาการจากวิธีการวัด MRMLC มีลักษณะการแจกแจงใกล้เคียงกับโค้งปกติ แต่มีความโค้งสูงกว่าโค้งปกติ สำหรับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่าคะแนนพัฒนาการจากวิธีการวัด Sθ และ Vθ มีค่าเฉลี่ยเท่ากับ - 0.0306 และ 0.0185 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.2579 และ 0.2109 มีค่าต่ำสุดเป็น - 1.0808 และ - 0.4701 ค่าสูงสุดเป็น 0.5855 และ 0.6354 ตามลำดับ โดยคะแนนพัฒนาการจากวิธีการวัด Sθ มีลักษณะการแจกแจงแบบเบ้ซ้าย และมีลักษณะโด่งกว่าโค้งปกติ ส่วนคะแนนพัฒนาการจากวิธีการวัด Vθ มีลักษณะการแจกแจงแบบเบ้ขวา และมีลักษณะโด่งใกล้เคียงกับโค้งปกติ

โดยสรุป คะแนนพัฒนาการจากวิธีการวัด Vθ มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการจากวิธีการวัด GS มากที่สุด รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด DS (รายละเอียดดังตารางที่ 4.21)

ตารางที่ 4. 21 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลชุดวัด 3 ครั้ง ซ้ำด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการให้คะแนนแบบทวิวิภาค (dichotomous)

คะแนนพัฒนาการ	N	Mean	Median	Mode	SD	Variance	Skewness	Kurtosis	Minimum	Maximum
วิธีที่เป็นเกณฑ์										
GS	378	1.0441	1.0134	0.7165	0.3909	0.1528	0.4371	0.1483	0.1353	2.3795
วิธี CTT										
DS	386	1.1399	1.0000	-1.0000	3.5458	12.5726	0.0683	0.0432	-10.0000	11.0000
SC	386	-0.0169	-0.0424	-0.7082	1.1683	1.3648	0.0204	-0.0086	-3.6387	3.3304
NL	386	0.1414	0.1335	0.0000	0.5242	0.2748	-0.1115	2.5623	-2.3979	2.4849
SR	386	4.8451	5.2632	0.0000	21.0636	443.6770	-0.3864	0.7404	-71.4286	62.5000
SRR	386	1.2458	-2.3071	-8.9125	19.4574	378.5839	0.6397	0.7420	-54.5021	68.6418
วิธี IRT										
Lθ	386	-0.0024	0.0016	-1.0234	0.8553	0.7316	-0.0424	0.0968	-2.5213	2.4378
MRMLC	386	0.2270	0.2005	0.0000	0.7641	0.5839	-0.0625	1.0370	-3.0215	3.1908
วิธีที่พัฒนา										
Sθ	386	-0.0306	0.0004	-0.2562	0.2579	0.0665	-1.0690	2.3187	-1.0808	0.5855
Vθ	386	0.0185	0.0004	-0.2555	0.2109	0.0445	0.4193	0.0351	-0.4701	0.6354

3.3.2 ข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit)

ผลการวิเคราะห์ค่าสถิติพื้นฐาน พบว่า คะแนนพัฒนาการที่ได้จากวิธีเกณฑ์คือ GS มีค่าเฉลี่ยเท่ากับ 8.0875 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.7722 มีค่าต่ำสุดเท่ากับ 3.4965 ค่าสูงสุดเท่ากับ 13.1090 คะแนนพัฒนาการจากวิธีการวัด GS มีลักษณะการแจกแจงใกล้เคียงกับโค้งปกติ โดยมีลักษณะแบนกว่าปกติเล็กน้อย (Sk = - 0.0010, Ku = - 0.1807) ส่วนคะแนนพัฒนาการจากกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่ามีค่าเฉลี่ยอยู่ระหว่าง 0 ถึง 8.0189 ส่วนเบี่ยงเบนมาตรฐานมีค่าอยู่ระหว่าง 0.2627 ถึง 22.6409 มีค่าต่ำสุดอยู่ระหว่าง - 73.6842 ถึง - 0.6008 มีค่าสูงสุดอยู่ระหว่าง 1.4351 ถึง 77.3572 โดย NL มีลักษณะการแจกแจงเบ้ขวามากที่สุด (Sk = 1.0105) และมีความโด่งสูงกว่าโค้งปกติมากที่สุด (Ku = 3.0864) คะแนนพัฒนาการจากวิธีการวัด DS, SC และ SRR มีลักษณะการแจกแจงแบบเบ้ขวา (Sk = 0.3555, 0.3633 และ 0.3587 ตามลำดับ) คะแนนพัฒนาการจากวิธีการวัด DS และ SC มีความโด่งใกล้เคียงกับโค้งปกติ (Ku = 0.0600, 0.0695 ตามลำดับ) ส่วนคะแนนพัฒนาการจากวิธีการวัด SRR มีความโด่งสูงกว่าโค้งปกติ ส่วนคะแนนพัฒนาการจากวิธีการวัด SR มีลักษณะการแจกแจงแบบเบ้ซ้าย โดยมีความโด่งสูงกว่าโค้งปกติ สำหรับทฤษฎีการตอบสนองข้อสอบ พบว่าคะแนนพัฒนาการจากวิธีการวัด Lθ มีค่าเฉลี่ยเท่ากับ 0.0761 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.8270 มีค่าต่ำสุดเป็น - 2.4321 และค่าสูงสุดเป็น 2.4811 มีลักษณะการแจกแจงเบ้ซ้ายและมีความโด่งแบนกว่าโค้งปกติเล็กน้อย ส่วนกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาพบว่าคะแนนพัฒนาการจากวิธีการวัด Sθ และ Vθ มีค่าเฉลี่ยเท่ากับ - 0.0075 และ 0.0426 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.2312 และ 0.2128 มีค่าต่ำสุดเป็น - 0.9562 และ - 0.4665 และมีค่าสูงสุดเป็น 0.5141 และ 0.7817 ตามลำดับ คะแนนพัฒนาการจากวิธีการวัด Sθ มีลักษณะการแจกแจงแบบเบ้ซ้าย มีความโด่งสูงกว่าโค้งปกติ คะแนนพัฒนาการจากวิธีการวัด Vθ มีลักษณะการแจกแจงแบบเบ้ขวา มีความโด่งใกล้เคียงกับโค้งปกติ

โดยสรุป คะแนนพัฒนาการจากวิธีการวัด Lθ มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการจากวิธีการวัด GS มากที่สุด รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด DS (รายละเอียดดังตารางที่ 4.22)

ตารางที่ 4.22 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการให้คะแนนความรู้บางส่วน (partial credit)

คะแนนพัฒนาการ	N	Mean	Median	Mode	SD	Variance	Skewness	Kurtosis	Minimum	Maximum
วิธีที่เป็นเกณฑ์										
GS	378	8.0875	8.0782	5.4250	1.7722	3.1406	-0.0010	-0.1807	3.4965	13.1090
วิธี CTT										
DS	386	2.6010	2.0000	3.0000	5.9484	35.3833	0.3555	0.0600	-14.0000	21.0000
SC	386	0.0000	-0.1220	-1.1477	1.2241	1.4983	0.3633	0.0695	-3.4100	3.8076
NL	386	0.1063	0.0785	0.0000	0.2627	0.0690	1.0105	3.0864	-0.6008	1.4351
SR	386	8.0189	9.0909	0.0000	22.6409	512.6108	-0.2585	0.2255	-73.6842	60.8696
SRR	386	1.1287	0.7654	-18.1268	20.8581	435.0605	0.3587	0.4177	-54.7599	77.3572
วิธี IRT										
LΘ	386	0.0761	0.0444	-2.4321	0.8270	0.6839	-0.1140	-0.1080	-2.4321	2.4811
วิธีที่พัฒนา										
SΘ	386	-0.0075	0.0113	-0.9562	0.2312	0.0534	-0.8261	1.1743	-0.9562	0.5141
VΘ	386	0.0426	0.0102	-0.4665	0.2128	0.0453	0.4558	0.0669	-0.4665	0.7817

3.4 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพัฒนาการรายบุคคลที่ได้จากการวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)

ผลการวิเคราะห์ค่าสถิติพื้นฐาน พบว่าคะแนนพัฒนาการที่ได้จากวิธีเกณฑ์คือ GS มีค่าเฉลี่ยเท่ากับ 2.6420 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.2415 มีค่าต่ำสุดเท่ากับ – 0.0742 ค่าสูงสุดเท่ากับ 5.8222 คะแนนพัฒนาการจากวิธีการวัด GS มีลักษณะการแจกแจงแบบเบ้ขวา โดยมีลักษณะแบนกว่าโค้งปกติ ส่วนคะแนนพัฒนาการจากกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่ามีค่าเฉลี่ยอยู่ระหว่าง -0.0274 ถึง 13.3496 ส่วนเบี่ยงเบนมาตรฐานมีค่าอยู่ระหว่าง 0.5680 ถึง 27.0393 มีค่าต่ำสุดอยู่ระหว่าง - 54.7331 ถึง - 1.6094 มีค่าสูงสุดอยู่ระหว่าง 2.5649 ถึง 107.8898 โดยคะแนนพัฒนาการจากวิธีการวัด SC และ SR มีลักษณะการแจกแจงใกล้เคียงกับโค้งปกติ ส่วนคะแนนพัฒนาการจากวิธีอื่นๆมีลักษณะการแจกแจงแบบเบ้ขวา ส่วนลักษณะความโด่งพบว่าคะแนนพัฒนาการจากวิธีการวัด NL และ SRR มีความโด่งสูงกว่าโค้งปกติ คะแนนพัฒนาการจากวิธีการวัด DS, SC และ SR มีลักษณะของโค้งแบนกว่าโค้งปกติ สำหรับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบพบว่า คะแนนพัฒนาการจากวิธีการวัด LΘ และ MRMLC มีค่าเฉลี่ยเท่ากับ - 0.0115 และ 0.4944 ส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 0.9807 และ 0.9132 มีค่าต่ำสุดเป็น - 2.6160 และ - 1.9954 ค่าสูงสุดเป็น 2.6215 และ 3.3583 ตามลำดับ โดยคะแนนพัฒนาการจากวิธีการวัด LΘ มีลักษณะการแจกแจงใกล้เคียงกับโค้งปกติ (Sk = 0.0912;

Ku = -0.0696) คะแนนพัฒนาการจากวิธีการวัด MRMLC มีลักษณะการแจกแจงแบบเบ้ขวา มีความโด่งแบนกว่าโค้งปกติ สำหรับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาพบว่าคะแนนพัฒนาการจากวิธีการวัด Sθ และ Vθ มีค่าต่ำสุดเป็น - 1.3095 และ - 0.4664 และมีค่าสูงสุดเป็น 0.5919 และ 0.7854 โดยคะแนนพัฒนาการจากวิธีการวัด Sθ มีลักษณะการแจกแจงแบบเบ้ซ้าย และมีความโด่งสูงกว่าโค้งปกติ ส่วนคะแนนพัฒนาการจากวิธีการวัด Vθ มีลักษณะการแจกแจงแบบเบ้ขวา และมีความโด่งใกล้เคียงกับโค้งปกติ

โดยสรุป คะแนนพัฒนาการจากวิธีการวัด DS มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการจากวิธีการวัด GS มากที่สุด รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด MRMLC (รายละเอียดดังตารางที่ 4.23)

ตารางที่ 4. 23 ค่าสถิติพื้นฐานของคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆของข้อมูลวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการให้คะแนนแบบทวิวิภาค (dichotomous)

คะแนนพัฒนาการ	N	Mean	Median	Mode	SD	Variance	Skewness	Kurtosis	Minimum	Maximum
วิธีที่เป็นเกณฑ์										
GS	362	2.6420	2.5624	1.5314	1.2415	1.5414	0.3883	-0.5465	-0.0742	5.8222
วิธี CTT										
DS	386	2.5570	2.0000	0.0000	4.5187	20.4168	0.2491	-0.5196	-9.0000	14.0000
SC	386	-0.0274	-0.1053	-0.7991	1.2122	1.4695	0.0986	-0.4281	-3.2826	2.9485
NL	386	0.2743	0.2513	0.0000	0.5680	0.3226	0.2358	0.8097	-1.6094	2.5649
SR	386	13.3496	11.7647	0.0000	26.3551	694.5890	-0.0239	-0.1120	-72.7273	82.3529
SRR	386	1.1372	-5.0861	-21.9764	27.0393	731.1211	0.9470	0.8350	-54.7331	107.8898
วิธี IRT										
Lθ	386	-0.0115	-0.0467	1.6427	0.9807	0.9618	0.0912	-0.0696	-2.6160	2.6215
MRMLC	386	0.4944	0.4184	0.0000	0.9132	0.8340	0.2265	-0.2237	-1.9954	3.3583
วิธีที่พัฒนา										
Sθ	386	-0.0347	-0.0106	0.3604	0.2909	0.0846	-1.0585	2.8690	-1.3095	0.5919
Vθ	386	0.0187	-0.0120	0.4772	0.2447	0.0599	0.5640	-0.0386	-0.4664	0.7854

ตอนที่ 4 ผลการวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการ

คุณภาพของวิธีการวัดคะแนนพัฒนาการในการวิจัยครั้งนี้พิจารณาจากคุณภาพ 2 ด้าน ได้แก่ 1) ความตรงตามเกณฑ์สัมพันธ์ 2) ค่าความคลาดเคลื่อนของคะแนนพัฒนาการ ซึ่งมีรายละเอียดของการวิเคราะห์ดังนี้

4.1 ผลการวิเคราะห์ค่าความตรงตามเกณฑ์สัมพันธ์

ค่าความตรงตามเกณฑ์สัมพันธ์ พิจารณาจากค่าความสัมพันธ์ 2 แบบ ได้แก่ 1) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการที่เป็นเกณฑ์กับคะแนนพัฒนาการที่ได้จากวิธีการวัดวิธีแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา และ 2) ค่าสหสัมพันธ์อันดับระหว่างอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์กับอันดับของคะแนนพัฒนาการที่ได้จากวิธีการวัดวิธีแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยรายละเอียดดังนี้

4.1.1 ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการ

ผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการ จะนำเสนอเป็น 3 ขั้นตอนคือ 1) ผลการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการ 2) ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการระหว่างกลุ่มวิธี 3 กลุ่มวิธี 3) ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการภายในแต่ละกลุ่มวิธี โดยมีรายละเอียดดังนี้ดังนี้

1) ผลการประมาณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการของแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา กับคะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่ใช้เป็นเกณฑ์ โดยใช้สูตร Pearson's Product Moment พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการที่ได้จากวิธีการวัดแต่ละวิธีกับคะแนนพัฒนาการที่เป็นเกณฑ์ มีความสัมพันธ์ทางบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทุกวิธี และความสัมพันธ์ส่วนใหญ่มีค่าค่อนข้างสูง (ความสัมพันธ์มีค่าอยู่ระหว่าง 0.2329 ถึง 0.9929) เมื่อพิจารณาแยกตามกลุ่มวิธีพบว่าในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ข้อมูลที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับฉบับเดิม (repeat) และข้อมูลที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) คะแนนพัฒนาการจากวิธีการวัด DS มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงที่สุด ส่วนข้อมูลที่ได้จากการวัดซ้ำ 3 ครั้ง และ 5 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 1 และ equivalence

2) คะแนนพัฒนาการจากวิธีการวัด SRR มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงที่สุด ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบพบว่า ข้อมูลทุกชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) คะแนนพัฒนาการจากวิธีการวัด MRMLC มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัดLθ ส่วนข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีคะแนนพัฒนาการที่ได้จากวิธีการวัดวิธีเดียวคือวิธีการวัด Lθ ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาพบว่าคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ ในกรณีที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และคะแนนพัฒนาการจากวิธีการวัด Sθ มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Vθ ในกรณีที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit)

ในภาพรวม พบว่าคะแนนพัฒนาการ DS มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงที่สุด สำหรับข้อมูลที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) และข้อมูลที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) และคะแนนพัฒนาการ SRR มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงที่สุด สำหรับข้อมูลที่ได้จากการวัดซ้ำ 3 ครั้ง และ 5 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 1 และ equivalence 2) (รายละเอียดดังตารางที่ 4.24)

ตารางที่ 4.24 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการที่ได้จากวิธีต่างๆกับคะแนนพัฒนาการที่เป็นเกณฑ์ จำแนกตามกลุ่มวิธีการวัด ชุดของข้อมูลและวิธีการตรวจให้คะแนน

	repeat		counterbalance		equivalence 1		equivalence 2
	dichotomous	partial credit	dichotomous	partial credit	dichotomous	partial credit	dichotomous
วิธีCTT							
DS	0.9658**	0.9929**	0.9492**	0.9535**	0.7227**	0.7155**	0.7061**
SC	0.8000**	0.8242**	0.8579**	0.8960**	0.6602**	0.7085**	0.5879**
NL	0.7999**	0.8755**	0.8082**	0.8604**	0.6136**	0.6043**	0.5630**
SR	0.9583**	0.9459**	0.8738**	0.8071**	0.7108**	0.7528**	0.7228**
SRR	0.5478**	0.2329**	0.4304**	0.2864**	0.9733**	0.9532**	0.7642**
วิธี IRT							
Lθ	0.7797**	0.8067**	0.6967**	0.7439**	0.5593**	0.4358**	0.5420**
MRMLC	0.9480**	-	0.9418**	-	0.6761**	-	0.6610**
วิธีที่พัฒนา							
Sθ	0.6757**	0.7925**	0.6201**	0.7314**	0.4411**	0.4244**	0.4599**
Vθ	0.7879**	0.8072**	0.6959**	0.6638**	0.5850**	0.4065**	0.5536**

** P < 0.01

2) ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการระหว่างกลุ่มวิธี 3 กลุ่มวิธี คือ กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยการตรวจสอบว่ามีความแปรผันอันเนื่องมาจากกลุ่มวิธีการวัดหรือไม่ ในการวิจัยครั้งนี้ค่าสัมประสิทธิ์สหสัมพันธ์ที่นำมาวิเคราะห์มีจำนวนน้อย จึงอาจจะไม่เหมาะสมที่จะตรวจสอบค่าความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการโดยใช้การวิเคราะห์ด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น (Hiearchical Linear Model: HLM) แต่เนื่องจากยังไม่มีวิธีการวิเคราะห์วิธีอื่นที่เหมาะสมกว่าวิธีนี้ที่จะทำการตรวจสอบได้ว่ามีความแปรผันอันเนื่องมาจากกลุ่มวิธีการวัดหรือไม่ ดังนั้นผู้วิจัยจึงใช้การวิเคราะห์ด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น 2 ระดับ ซึ่งผลการวิเคราะห์มีรายละเอียดดังนี้ (รายละเอียดดังตารางที่ 4.25)

(1) ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) เมื่อตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 9 ค่า มีค่าเท่ากับ 0.807 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($t = 17.464$, $df = 2$, $p = 0.000$) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาขึ้นมีค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.814, 0.864 และ 0.732 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์ สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ ($\chi^2 = 0.940$, $df = 2$, $p > 0.500$) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน ส่วนข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 8 ค่ามีค่าเท่ากับ 0.785 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($t = 9.467$, $df = 2$, $p = 0.000$) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.774, 0.807 และ 0.800 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่าแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ ($\chi^2 = 0.027$, $df = 2$, $p > 0.500$) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน

(2) ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 9 วิธี มีค่าเท่ากับ 0.764 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($t = 13.430$, $df = 2$, $p = 0.000$) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.784, 0.819 และ 0.658 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ ($\chi^2 = 1.063$, $df = 2$, $p>0.500$) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน ส่วนข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 8 วิธี มีค่าเท่ากับ 0.743 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($t = 10.097$, $df = 2$, $p = 0.000$) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.761, 0.744 และ 0.698 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ ($\chi^2 = 0.132$, $df = 2$, $p >0.500$) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน

(3) ข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 9 ค่า มีค่าเท่ากับ 0.636 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($t = 9.343$, $df = 2$, $p = 0.000$) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.736, 0.618 และ 0.513 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของแต่ละกลุ่มวิธีทั้ง 3 กลุ่ม พบว่าแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ ($\chi^2 = 5.222$, $df = 2$, $p = 0.072$) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน ส่วนข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการวิเคราะห์

พบว่าค่าเฉลี่ยสัมประสิทธิ์ สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 8 ค่า มีค่าเท่ากับ 0.545 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 4.779, df = 2, p = 0.05) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.747, 0.436 และ 0.415 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (χ^2 = 19.465, df = 2, p = 0.000) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าแตกต่างกัน

(4) ข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 9 ค่า มีค่าเท่ากับ 0.601 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 12.342, df = 2, p = 0.000) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.669, 0.602 และ 0.507 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2 = 5.840, df = 2, p = 0.052) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.25 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของวิธีการวัดต่างๆกับคะแนนพัฒนาการที่เป็นเกณฑ์ จำแนกตามลักษณะของข้อมูล วิธีการตรวจให้คะแนน

ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
repeat	dichotomous	intercept, Goo	0.807	0.046	2	17.464	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.002	0	2	0.940	>0.500
		level R	0.139	0.019			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
repeat	partial credit	intercept, Goo	0.785	0.083	2	9.467	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.002	0	2	0.027	>0.500
		level R	0.234	0.055			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
counterbalance	dichotomous	intercept, Goo	0.764	0.057	2	13.430	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.010	0.000	2	1.063	>0.500
		level R	0.170	0.029			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
counterbalance	partial credit	intercept, Goo	0.743	0.074	2	10.097	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.009	0.000	2	0.132	>0.500
		level R	0.207	0.043			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
equivalence 1	dichotomous	intercept, Goo	0.636	0.068	2	9.343	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.090	0.008	2	5.222	0.072
		level R	0.125	0.016			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
equivalence 1	partial credit	intercept, Goo	0.545	0.114	2	4.779	0.050
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.179	0.032	2	19.465	0.000
		level R	0.113	0.013			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
equivalence 2	dichotomous	intercept, Goo	0.601	0.049	2	12.342	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.067	0.004	2	5.840	0.052
		level R	0.084	0.007			

โดยสรุป พบว่าข้อมูลทุกชุดค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการระหว่างวิธีการวัดแต่ละวิธีกับของคะแนนพัฒนาการที่เป็นเกณฑ์แต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน ยกเว้นข้อมูลวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) และมีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ที่พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าแตกต่างกัน ผู้วิจัยจึงได้ทำการตรวจสอบว่ากลุ่มวิธีคู่ใดที่มีค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการแตกต่างกัน โดยใช้การวิเคราะห์ด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น (Hieararchical Linear Model: HLM) 2 ระดับ ผลการวิเคราะห์มีรายละเอียดดังนี้ (รายละเอียดดังตารางที่ 4.26)

(1) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่า ค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 6 ค่า มีค่าเท่ากับ 0.612 ซึ่งแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ ($t = 3.973$, $df = 1$, $p = 0.180$) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ มีค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.747 และ 0.436 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($\chi^2 = 7.447$, $df = 1$, $p = 0.006$) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบมีค่าแตกต่างกัน

(2) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า ค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 7 ค่า มีค่าเท่ากับ 0.587 ซึ่งแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ ($t = 3.558$, $df = 1$, $p = 0.185$) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยของสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.747 และ 0.415 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($\chi^2 = 14.231$, $df = 1$, $p < 0.000$) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าแตกต่างกัน

(3) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า ค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 3 ค่า มีค่าเท่ากับ 0.424 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 42.486, df = 1, p = 0.017) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการเท่ากับ 0.436 และ 0.415 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2 = 1.771, df = 1, p = 0.180) สรุปได้ว่าค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าไม่แตกต่างกัน

ตารางที่ 4.26 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการของวิธีการวัดต่างๆกับคะแนนพัฒนาการที่เป็นเกณฑ์ ของข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการให้คะแนนความรู้บางส่วน (partial credit) จำแนกตามกลุ่มวิธีการวัด

กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
CTT- IRT	intercept, G_{oo}	0.612	0.154	1	3.973	0.180
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.196	0.039	1	7.447	0.006
	level R	0.139	0.019			
กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
CTT- วิธีที่พัฒนา	intercept, G_{oo}	0.587	0.165	1	3.558	0.185
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.224	0.050	1	14.231	0.000
	level R	0.17	0.029			
กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
IRT- วิธีที่พัฒนา	intercept, G_{oo}	0.423	0.010	1	42.486	0.017
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.009	0.000	1	1.771	0.180
	level R	0.013	0.000			

การทดสอบภายหลังของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการรายคู่

จากการวิเคราะห์พบว่าข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence) ที่มีการให้คะแนนความรู้บางส่วน (partial credit) ค่าเฉลี่ยสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการมีค่าแตกต่างกันระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กับกลุ่มวิธีที่พัฒนาขึ้นอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ดังนั้นผู้วิจัยจึงได้ทำการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการรายคู่เพื่อดูว่าค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการคู่ใดที่มีค่าแตกต่างกัน โดยใช้สูตรของ Hotelling (1940) ผลการทดสอบ พบว่า แต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมมีค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการสูงกว่าทุกวิธีในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและทุกวิธีในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 (α = 0.01, df= n-3, t=2.326) (รายละเอียดดังตารางที่ 4.27)

ตารางที่ 4.27 ค่าสถิติ t ที่ใช้ในการทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์เป็นรายคู่ของคะแนนพัฒนาการของข้อมูลชุดวัดซ้ำ 3 ครั้ง (equivalence 1) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit)

	วิธี CTT				
	DS	SC	NL	SR	SRR
วิธี IRT					
Lθ	7.896**	7.618**	4.006**	9.708**	31.194**
วิธีที่พัฒนา					
Sθ	8.222**	7.942**	4.274**	10.191**	31.567**
Vθ	8.499**	8.221**	4.587**	10.241**	32.177**

** p< 0.01

3) ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการภายในกลุ่มวิธี 3 กลุ่มวิธี คือภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยใช้สูตรของ Hotelling (1940) ได้ผลการทดสอบดังนี้ (รายละเอียดดังตารางที่ 4.28)

(1) ผลการวิเคราะห์ภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่าข้อมูลทุกชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน

พัฒนาการของวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์เกือบทุกวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 (α = 0.01, df = n-3, t = 2.326) ยกเว้นคะแนนพัฒนาการ NL กับ SC มีค่าไม่แตกต่างกันในข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิมและชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน คะแนนพัฒนาการจากวิธีการวัด SR กับ SC มีค่าไม่แตกต่างกันในข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน และคะแนนพัฒนาการวิธีการวัด SR กับ DS มีค่าไม่แตกต่างกันในข้อมูลทุกชุด ยกเว้นข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนานที่มีค่าแตกต่างกัน สำหรับข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน พบว่า ทุกวิธีการวัดมีค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนพัฒนาการกับวิธีเกณฑ์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.001

(2) ผลการวิเคราะห์ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่าข้อมูลทุกชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค คะแนนพัฒนาการจากวิธีการวัด MRMLC มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 (α = 0.01, df = n - 3, t = 2.326)

(3) ผลการวิเคราะห์ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่าข้อมูลทุกชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค คะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 (α = 0.01, df = n - 3, t = 2.326) ส่วนข้อมูลทุกชุดที่มีการตรวจให้คะแนนความรู้บางส่วน พบว่า คะแนนพัฒนาการจากวิธีการวัด Vθ และ Sθ มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ยกเว้นข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนานที่คะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์ต่ำกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 (α = 0.01, df = n - 3, t = 2.326)

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตารางที่ 4.28 ผลการเปรียบเทียบค่าสัมประสิทธิ์สหสัมพันธ์รายคู่ของคะแนนพัฒนาการจำแนกตามวิธีการวัด ชุดของข้อมูลและวิธีการตรวจให้คะแนน

ข้อมูลชุด		กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม					กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ				กลุ่มวิธีที่พัฒนา	
		DS	SC	NL	SR	SRR		Lθ	MRMLC		Sθ	Vθ
repeat		partial credit										
dichotomous	DS		*	*	*	*	Lθ		-	Sθ		NS
	SC	*		*	*	*	MRMLC	*		Vθ	*	
	NL	*	NS		*	*						
	SR	NS	*	*		*						
	SRR	*	*	*	*							
Counterbalance		partial credit										
dichotomous	DS		*	*	*	*	Lθ		-	Sθ		*
	SC	*		*	*	*	MRMLC	*		Vθ	*	
	NL	*	*		*	*						
	SR	*	NS	*		*						
	SRR	*	*	*	*							
equivalence 1		partial credit										
dichotomous	DS		*	*	*	*	Lθ		-	Sθ		NS
	SC	*		*	*	*	MRMLC	*		Vθ	*	
	NL	*	*		*	*						
	SR	NS	*	*		*						
	SRR	*	*	*	*							
equivalence 2		partial credit										
dichotomous	DS						Lθ			Sθ		
	SC	*					MRMLC	*		Vθ	*	
	NL	*	NS									
	SR	NS	*	*								
	SRR	*	*	*	*							

* หมายถึงค่าสัมประสิทธิ์สหสัมพันธ์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01

NS หมายถึงค่าสัมประสิทธิ์สหสัมพันธ์แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

4.1.2 ผลการวิเคราะห์ค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการ

ผลการวิเคราะห์ค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการ จะนำเสนอเป็น 3 ขั้นตอนคือ 1) ผลการประมาณค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการ 2) ผลการทดสอบความแตกต่างของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการระหว่างกลุ่มวิธี 3 กลุ่มวิธี 3) ผลการทดสอบความแตกต่างของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการภายในแต่ละกลุ่มวิธี โดยมีรายละเอียดดังนี้ดังนี้

1) ผลการประมาณค่าสหสัมพันธ์อันดับระหว่างคะแนนพัฒนาการของแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา กับคะแนนพัฒนาการที่เป็นเกณฑ์ โดยใช้สูตร Spearman's rank correlation พบว่าค่าสหสัมพันธ์อันดับระหว่างคะแนนพัฒนาการที่ได้จากวิธีการวัดแบบต่างๆกับคะแนนเกณฑ์ มีความสัมพันธ์กันทางบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทุกวิธี และส่วนใหญ่มีค่าค่อนข้างสูง (ค่าความสัมพันธ์อันดับอยู่ระหว่าง 0.3911 ถึง 0.9899) เมื่อพิจารณาแยกตามกลุ่มวิธีพบว่าในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ข้อมูลที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) คะแนนพัฒนาการ SR มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการจากวิธีเกณฑ์สูงที่สุด ส่วนข้อมูลที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) และข้อมูลที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) คะแนนพัฒนาการ DS มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงที่สุด ส่วนข้อมูลที่ได้จากการวัดซ้ำ 3 ครั้ง และ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1 และ equivalence 2) คะแนนพัฒนาการจากวิธีการวัด SRR มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงที่สุด ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบพบว่า ข้อมูลทุกชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) คะแนนพัฒนาการ จากวิธีการวัด MRMLC มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ ส่วนข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีคะแนนพัฒนาการที่ได้จากวิธีการวัดวิธีเดียวคือ Lθ ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาพบว่าคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่า Sθ ในกรณีที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และคะแนนพัฒนาการจากวิธีการวัด Sθ มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Vθ ในกรณีที่มีการตรวจให้คะแนน

ความรู้บางส่วน (partial credit) ยกเว้นในกรณีของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่คะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ

ในภาพรวม พบว่าคะแนนพัฒนาการจากวิธีการวัด DS มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงที่สุด สำหรับข้อมูลที่ได้จากการวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) และข้อมูลที่ได้จากการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) และคะแนนพัฒนาการจากวิธีการวัด SRR มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงที่สุด สำหรับข้อมูลที่ได้จากการวัดซ้ำ 3 และ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1 และ equivalence 2) ส่วนคะแนนพัฒนาการจากวิธีการวัด SR มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงที่สุด สำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) (รายละเอียดดังตารางที่ 4.29)

ตารางที่ 4.29 ค่าสหสัมพันธ์ระหว่างอันดับของคะแนนพัฒนาการที่ได้จากวิธีต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ จำแนกตามกลุ่มวิธีการวัด ชุดของข้อมูลและวิธีการตรวจให้คะแนน

	repeat		counterbalance		equivalence 1		equivalence 2
	dichotomous	partial credit	dichotomous	partial credit	dichotomous	partial credit	dichotomous
วิธีCTT							
DS	0.9575**	0.9899**	0.9503**	0.9497**	0.6964**	0.7033**	0.7035**
SC	0.7650**	0.7780**	0.8446**	0.8802**	0.6307**	0.6677**	0.5799**
NL	0.8455**	0.9148**	0.8648**	0.8901**	0.6401**	0.6691**	0.6151**
SR	0.9752**	0.9819**	0.9411**	0.9084**	0.7023**	0.7360**	0.7263**
SRR	0.9392**	0.8954**	0.7891**	0.6152**	0.9872**	0.9761**	0.7900**
วิธี IRT							
Lθ	0.7408**	0.7892**	0.7070**	0.7249**	0.5241**	0.4015**	0.5250**
MRMLC	0.9405**	-	0.9457**	-	0.6695**	-	0.6730**
วิธีที่พัฒนา							
Sθ	0.7173**	0.7826**	0.7042**	0.7399**	0.5066**	0.4093**	0.5170**
Vθ	0.7646**	0.7909**	0.7074**	0.6806**	0.5381**	0.3911**	0.5311**

** P < 0.01

2) ผลการทดสอบความแตกต่างของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการระหว่างกลุ่มวิธี 3 กลุ่มวิธี คือ กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยการ

ตรวจสอบว่ามีความแปรผันอันเนื่องมาจากกลุ่มวิธีการวัดหรือไม่ ในการวิจัยครั้งนี้ค่าสหสัมพันธ์อันดับที่นำมาวิเคราะห์มีจำนวนน้อย จึงอาจจะไม่เหมาะสมที่จะตรวจสอบค่าความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการโดยใช้การวิเคราะห์ด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น (Hiearchical Linear Model: HLM) แต่เนื่องจากยังไม่มีวิธีการวิเคราะห์วิธีอื่นที่เหมาะสมกว่าวิธีนี้ที่จะทำการตรวจสอบได้ว่ามีความแปรผันอันเนื่องมาจากกลุ่มวิธีการวัดหรือไม่ ดังนั้นผู้วิจัยจึงใช้การวิเคราะห์ด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น 2 ระดับ ซึ่งผลการวิเคราะห์มีรายละเอียดดังนี้ (รายละเอียดดังตารางที่ 4.30)

(1) ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) เมื่อตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 9 ค่า มีค่าเท่ากับ 0.836 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 17.865, df = 2, p = 0.000) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.896, 0.840 และ 0.740 ตามอันดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2 = 4.201, df = 2, p =0.120) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน ส่วนข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 8 ค่ามีค่าเท่ากับ 0.842 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 17.982, df = 2, p = 0.000) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.912, 0.789 และ 0.787 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (χ^2 = 6.007, df = 2, p = 0.048) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าแตกต่างกัน

(2) ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 9 วิธี มีค่าเท่ากับ 0.811 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 15.741, df = 2, p = 0.000) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธี

ของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.877, 0.826 และ 0.706 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2= 5.789, df = 2, p = 0.054) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน ส่วนข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 8 วิธี มีค่าเท่ากับ 0.785 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 14.564, df = 2, p = 0.000) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.849, 0.725 และ 0.710 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2= 2.637, df = 2, p = 0.267) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน

(3) ข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 9 ค่า มีค่าเท่ากับ 0.632 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 9.582, df = 2, p = 0.000) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.731, 0.597 และ 0.522 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของแต่ละกลุ่มวิธีทั้ง 3 กลุ่ม พบว่าแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2 = 4.822, df = 2, p = 0.088) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าไม่แตกต่างกัน ส่วนข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 8 ค่า มีค่าเท่ากับ 0.530 ซึ่งแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (t = 4.316, df = 2, p = 0.117) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.750, 0.402 และ 0.400 ตามลำดับ ผล

การทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (χ^2 = 22.350, df = 2, p = 0.000) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าแตกต่างกัน

(4) ข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับคะแนนพัฒนาการที่เป็นเกณฑ์รวม 9 ค่า มีค่าเท่ากับ 0.610 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 12.626, df = 2, p = 0.000) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.683, 0.599 และ 0.524 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 3 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (χ^2 = 6.347, df = 2, p = 0.041) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของแต่ละกลุ่มวิธีมีค่าแตกต่างกัน

ตารางที่ 4.30 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสหสัมพันธ์อันดับของอันดับคะแนนพัฒนาการของวิธีการวัดต่างๆกับอันดับคะแนนพัฒนาการที่เป็นเกณฑ์ จำแนกตามลักษณะของข้อมูล วิธีการตรวจให้คะแนน

ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
repeat	dichotomous	intercept, Goo	0.836	0.047	2	17.865	0.000
		random effect	SD	variance component	df	χ^2	P
		intercept, Uo	0.058	0.003	2	4.201	0.120
		level R	0.093	0.009			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
repeat	partial credit	intercept, Goo	0.842	0.047	2	17.982	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.062	0.004	2	6.007	0.048
		level R	0.074	0.006			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
counterbalance	dichotomous	intercept, Goo	0.811	0.052	2	15.741	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.071	0.005	2	5.789	0.054
		level R	0.088	0.008			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
counterbalance	partial credit	intercept, Goo	0.785	0.054	2	14.564	0.000
		random effect	SD	variance component	df	χ^2	P
		intercept, Uo	0.054	0.003	2	2.637	0.267
		level R	0.115	0.013			

ตารางที่ 4.30 (ต่อ)

ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
equivalence 1	dichotomous	intercept, Goo	0.632	0.066	2	9.582	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.085	0.007	2	4.822	0.088
		level R	0.125	0.016			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
equivalence 1	partial credit	intercept, Goo	0.530	0.123	2	4.316	0.117
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.195	0.038	2	22.350	0.000
		level R	0.115	0.013			
ข้อมูล	วิธีให้คะแนน	fixed effect	coefficient	SE	df	t	p
equivalence 2	dichotomous	intercept, Goo	0.610	0.048	2	12.626	0.000
		random effect	SD	variance component	df	χ^2	p
		intercept, Uo	0.067	0.005	2	6.347	0.041
		level R	0.081	0.007			

จากการทดสอบความแตกต่างของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดวิธีต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ระหว่างกลุ่มวิธี พบว่า มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ในข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิมและข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) และข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผู้วิจัยจึงได้ทำการทดสอบว่ากลุ่มวิธีคู่ใดที่มีค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการแตกต่างกัน โดยใช้การวิเคราะห์ด้วยโมเดลเชิงเส้นตรงระดับลดหลั่น (Hiearchical Linear Model: HLM) 2 ระดับ ผลการวิเคราะห์มีรายละเอียดดังนี้

2.1 ผลการทดสอบความแตกต่างของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดวิธีต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีรายละเอียดดังนี้ (รายละเอียดดังตารางที่ 4.31)

(1) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่า ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์รวม 6 ค่า มีค่าเท่ากับ 0.874 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($t = 15.375$, $df = 1$,

p = 0.022) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ มีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.912 และ 0.789 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2 = 1.971, df = 1, p = 0.157) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบมีค่าไม่แตกต่างกัน

(2) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์รวม 7 ค่า มีค่าเท่ากับ 0.856 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 13.669, df = 1, p = 0.024) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.912 และ 0.787 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (χ^2 = 4.313, df = 1, p = 0.035) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าแตกต่างกัน

(3) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์รวม 3 ค่า มีค่าเท่ากับ 0.424 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 305.982, df = 1, p = 0.000) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.789 และ 0.787 ตามลำดับ ผลการทดสอบค่าความแปร ปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2 = 0.207, df = 1, p > 0.500) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าไม่แตกต่างกัน

ตารางที่ 4.31 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ ของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการให้คะแนนความรู้บางส่วน (partial credit) จำแนกตามกลุ่มวิธีการวัด

กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
CTT- IRT	intercept, G_{oo}	0.874	0.057	1	15.375	0.022
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.057	0.003	1	1.970	0.157
	level R	0.085	0.007			
กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
CTT- วิธีที่พัฒนา	intercept, G_{oo}	0.856	0.063	1	13.669	0.024
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.077	0.006	1	4.313	0.035
	level R	0.076	0.006			
กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
IRT- วิธีที่พัฒนา	intercept, G_{oo}	0.788	0.003	1	305.982	0.000
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.001	0.000	1	0.207	>0.500
	level R	0.004	0.000			

2.2 ผลการทดสอบความแตกต่างของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดวิธีต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ระหว่างกลุ่มวิธีกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาของข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) มีรายละเอียดดังนี้ (รายละเอียดดังตารางที่ 4.32)

(1) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่า ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์รวม 6 ค่า มีค่าเท่ากับ 0.595 ซึ่งแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ ($t = 3.432$, $df = 1$, $p = 0.095$) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ มีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.750 และ 0.402 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนน

พัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (χ^2 = 9.465, df = 1, p = 0.003) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบมีค่าแตกต่างกัน

(2) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์รวม 7 ค่า มีค่าเท่ากับ 0.581 ซึ่งแตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (t = 3.330, df = 1, p = 0.096) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.750 และ 0.400 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (χ^2 = 15.587, df = 1, p = 0.000) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าแตกต่างกัน

(3) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์รวม 3 ค่า มีค่าเท่ากับ 0.401 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 75.784, df = 1, p = 0.005) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.402 และ 0.400 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2 = 0.013, df = 1, p > 0.500) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าไม่แตกต่างกัน

ตารางที่ 4.32 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ ของข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการให้คะแนนความรู้บางส่วน (partial credit) จำแนกตามกลุ่มวิธีการวัด

กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
CTT- IRT	intercept, G_{oo}	0.595	0.173	1	3.432	0.095
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.225	0.051	1	9.465	0.003
	level R	0.129	0.017			
กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
CTT- วิธีที่พัฒนา	intercept, G_{oo}	0.581	0.174	1	3.330	0.096
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.238	0.056	1	15.587	0.000
	level R	0.116	0.013			
กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
IRT- วิธีที่พัฒนา	intercept, G_{oo}	0.401	0.005	1	75.784	0.005
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.000	0.000	1	0.013	>0.500
	level R	0.009	0.000			

2.3 ผลการทดสอบความแตกต่างของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดวิธีต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ระหว่างกลุ่มวิธีกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาของข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) มีรายละเอียดดังนี้ (รายละเอียดดังตารางที่ 4.33)

(1) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่า ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์รวม 6 ค่า มีค่าเท่ากับ 0.655 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ($t = 16.269$, $df = 1$, $p = 0.021$) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ มีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.683 และ 0.599 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของ

คะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2 = 1.282, df = 1, p = 0.256) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบมีค่าไม่แตกต่างกัน

(2) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาขึ้น พบว่า ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์รวม 7 ค่า มีค่าเท่ากับ 0.609 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 7.728, df = 1, p = 0.045) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.683 และ 0.524 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (χ^2 = 7.176, df = 1, p = 0.007) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าแตกต่างกัน

(3) ผลการวิเคราะห์ระหว่างกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการจากทุกวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์รวม 3 ค่า มีค่าเท่ากับ 0.562 ซึ่งแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 (t = 14.984, df = 1, p = 0.022) โดยวิธีการวัดคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการเท่ากับ 0.599 และ 0.524 ตามลำดับ ผลการทดสอบค่าความแปรปรวนของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีทั้ง 2 กลุ่ม พบว่า แตกต่างจากศูนย์อย่างไม่มีนัยสำคัญทางสถิติ (χ^2 = 1.017, df = 1, p = 0.314) สรุปได้ว่าค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการของกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามีค่าไม่แตกต่างกัน

ตารางที่ 4.33 ผลการวิเคราะห์โมเดลศูนย์ (null model) ของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการของวิธีการวัดต่างๆกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์ ของข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการให้คะแนนแบบทวิวิภาค (dichotomous) จำแนกตามกลุ่มวิธีการวัด

กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
CTT- IRT	intercept, G_{oo}	0.655	0.040	1	16.269	0.021
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.029	0.001	1	1.282	0.256
	level R	0.089	0.008			
กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
CTT- วิธีที่พัฒนา	intercept, G_{oo}	0.609	0.079	1	7.728	0.045
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.102	0.011	1	7.176	0.007
	level R	0.076	0.006			
กลุ่มวิธี	fixed effect	coefficient	SE	df	t	p
IRT- วิธีที่พัฒนา	intercept, G_{oo}	0.562	0.037	1	14.984	0.022
	random effect	SD	variance component	df	χ^2	p
	intercept, U_o	0.007	0.000	1	1.017	0.314
	level R	0.074	0.006			

การทดสอบภายหลังของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการรายคู่

จากการวิเคราะห์ความแตกต่างของค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการระหว่างกลุ่มวิธี พบว่าข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการมีค่าแตกต่างกันระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการมีค่าแตกต่างกันระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 และข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ค่าเฉลี่ยสหสัมพันธ์อันดับของคะแนนพัฒนาการมีค่าแตกต่างกันระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กับกลุ่มวิธีของ

ทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ดังนั้นผู้วิจัยจึงได้ทำการทดสอบค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการเป็นรายคู่เพื่อดูว่าค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการคู่ใดที่มีค่าแตกต่างกัน โดยใช้สูตรของ Hotelling (1940) ผลการทดสอบ พบว่า ข้อมูลวัดซ้ำทุกชุด ในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมเกือบทุกวิธีมีค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการสูงกว่าทุกวิธีในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 (α = 0.01, df= n-3, t=2.326) ยกเว้น ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ที่ค่าสหสัมพันธ์อันดับของวิธีการวัด SC แตกต่างอย่างไม่มีนัยสำคัญทางสถิติจากค่าสหสัมพันธ์อันดับของวิธีการวัด Sθ และ Vθ และข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ที่ค่าสหสัมพันธ์อันดับของวิธีการวัด SC แตกต่างอย่างไม่มีนัยสำคัญทางสถิติจากค่าสหสัมพันธ์อันดับของวิธีการวัด Vθ และข้อมูลชุดวัดซ้ำ 3 ครั้ง ด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมทุกวิธีมีค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการสูงกว่าค่าสหสัมพันธ์อันดับของวิธีการวัด Lθ และสูงกว่าค่าสหสัมพันธ์อันดับของทุกวิธีในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ (รายละเอียดดังตารางที่ 4.34)

ตารางที่ 4.34 ค่าสถิติ t ในการทดสอบความแตกต่างของค่าสหสัมพันธ์อันดับเป็นรายคู่ของคะแนนพัฒนาการ จำแนกตามชุดของข้อมูลวัดซ้ำ

		วิธี CTT				
		DS	SC	NL	SR	SRR
repeat: partial credit	วิธีพัฒนา					
	Sθ	64.186**	-0.313	15.169**	40.511**	12.807**
	Vθ	60.498**	-0.839	13.477**	39.104**	11.350**
equivalence 1: partial credit	วิธี IRT					
	Lθ	8.229**	6.935**	6.925**	9.631**	47.817**
	วิธีที่พัฒนา					
	Sθ	8.061**	6.766**	6.752**	9.472**	47.449**
	Vθ	8.426**	7.133**	7.128**	9.816**	48.231**
equivalence 2: dichotomous	วิธีที่พัฒนา					
	Sθ	8.264**	2.437**	3.591**	9.348**	9.896**
	Vθ	7.775**	1.873	3.187**	8.830**	9.819**

** p< 0.01

3) ผลการทดสอบความแตกต่างของค่าสหสัมพันธ์อันดับของคะแนนพัฒนาการภายในกลุ่มวิธี 3 กลุ่มวิธี คือภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา โดยใช้สูตรของ Hotelling (1940 cited in Lindeman, Merenda and Gold, 1980) ได้ผลการทดสอบดังนี้ (รายละเอียดดังตารางที่ 4.35)

(1) ผลการวิเคราะห์ภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่าข้อมูลทุกชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค ค่าสหสัมพันธ์อันดับระหว่างอันดับของคะแนนพัฒนาการของวิธีการวัดกับอันดับของคะแนนพัฒนาการที่เป็นเกณฑ์เกือบทุกวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 (α = 0.01, df = n - 3, t = 2.326) ยกเว้นคะแนนพัฒนาการจากวิธีการวัด NL กับ SC และคะแนนพัฒนาการจากวิธีการวัด SR กับ DS ที่มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ในข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน สำหรับข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน พบว่า เกือบทุกวิธีการวัดมีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้น คะแนนพัฒนาการจากวิธีการวัด NL กับ SC และ คะแนนพัฒนาการจากวิธีการวัด NL กับ SR และ คะแนนพัฒนาการจากวิธีการวัด SC กับ SR ที่มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ในข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน และคะแนนพัฒนาการจากวิธีการวัด NL กับ SC มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ในข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน

(2) ผลการวิเคราะห์ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่าข้อมูลทุกชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค คะแนนพัฒนาการจากวิธีการวัด MRMLC มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01(α = 0.01, df= n-3, t=2.326)

(3) ผลการวิเคราะห์ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่าข้อมูลทุกชุดที่มีการตรวจให้คะแนนแบบทวิวิภาค คะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 (α = 0.01, df = n-3, t = 2.326) ยกเว้นข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน และข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนานที่คะแนนพัฒนาการจากวิธีการวัด Vθ และ Sθ มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนข้อมูลทุกชุดที่มีการตรวจให้คะแนนความรู้บางส่วน พบว่าคะแนนพัฒนาการจากวิธีการวัด Sθ มีค่าสหสัมพันธ์อันดับ

กับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Vθ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 (α = 0.01, df = n-3, t = 2.326) ยกเว้นข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิมและข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน ที่คะแนนพัฒนาการจากวิธีการวัด Vθ และ Sθ มีค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ตารางที่ 4.35 ผลการเปรียบเทียบค่าสหสัมพันธ์อันดับเป็นรายคู่ของคะแนนพัฒนาการจำแนกตามวิธีการวัด ชุดของข้อมูลและวิธีการตรวจให้คะแนน

ข้อมูลชุด		กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม					กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ				กลุ่มวิธีที่พัฒนา	
		DS	SC	NL	SR	SRR		Lθ	MRMLC		Sθ	Vθ
repeat		partial credit										
dichotomous	DS		*	*	*	*	Lθ		-	Sθ		NS
	SC	*		*	*	*	MRMLC	*		Vθ	*	
	NL	*	*		*	*						
	SR	*	*	*		*						
	SRR	*	*	*	*							
Counterbalance		partial credit										
dichotomous	DS		*	*	*	*	Lθ		-	Sθ		*
	SC	*		NS	NS	*	MRMLC	*		Vθ	NS	
	NL	*	*		NS	*						
	SR	*	*	*		*						
	SRR	*	*	*	*							
equivalence 1		partial credit										
dichotomous	DS		*	*	*	*	Lθ		-	Sθ		NS
	SC	*		NS	*	*	MRMLC	*		Vθ	*	
	NL	*	NS		*	*						
	SR	NS	*	*		*						
	SRR	*	*	*	*							
equivalence 2		partial credit										
dichotomous	DS						Lθ			Sθ		
	SC	*					MRMLC	*		Vθ	NS	
	NL	*	*									
	SR	*	*	*								
	SRR	*	*	*	*							

* หมายถึงค่าสหสัมพันธ์อันดับแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01

NS หมายถึงค่าสหสัมพันธ์อันดับแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

4.2 ผลการวิเคราะห์ค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการ

ในการวิเคราะห์ค่าความคลาดเคลื่อนของคะแนนพัฒนาการของแต่ละวิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา ผู้วิจัยทำการวิเคราะห์ค่าสถิติ 2 ค่า คือ 1) ค่าความลำเอียง (bias) ซึ่งหาจากค่าเฉลี่ยของความแตกต่างระหว่างคะแนนพัฒนาการที่ได้จากวิธีการวัดในแต่ละกลุ่มวิธีกับคะแนนพัฒนาการที่เป็นเกณฑ์ 2) RMSE (root mean squared error) ประมาณค่าจากค่ารากที่สองของค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างคะแนนพัฒนาการที่ได้จากวิธีการวัดในแต่ละกลุ่มวิธีกับคะแนนพัฒนาการที่เป็นเกณฑ์ โดยผู้วิจัยได้แปลงคะแนนพัฒนาการจากทุกวิธีให้เป็นคะแนนมาตรฐานเพื่อให้คะแนนพัฒนาการของทุกวิธีอยู่ในมาตร (scale) เดียวกัน แล้วจึงทำการวิเคราะห์ค่าความคลาดเคลื่อนของคะแนนพัฒนาการของแต่ละวิธี ซึ่งแยกตามลักษณะของแบบทดสอบ ดังรายละเอียดต่อไปนี้ (รายละเอียดดังตารางที่ 4.36)

4.2.1 ข้อมูลวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าคะแนนพัฒนาการจากทุกกลุ่มวิธีส่วนใหญ่มีค่าความลำเอียง (bias) เท่ากับ 0.0000 ยกเว้นคะแนนพัฒนาการจากวิธีการวัด NL และ MRMLC ที่มีค่าเท่ากับ 0.0006 และ 0.0596 ตามลำดับ ส่วนค่ารากที่สองของค่าเฉลี่ยกำลังที่สองของความคลาดเคลื่อน (RMSE) มีค่าอยู่ระหว่าง 0.2612 ถึง 0.9503 โดยคะแนนพัฒนาการจากวิธีการวัด DS และ SR มีค่า RMSE ต่ำที่สุดคือ 0.2612 และ 0.2886 ตามอันดับ และคะแนนพัฒนาการจากวิธีการวัด SRR มีค่า RMSE สูงที่สุดคือ 0.9503

ส่วนข้อมูลที่ตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการวิเคราะห์พบว่าคะแนนพัฒนาการจากทุกกลุ่มวิธีมีค่าความลำเอียง (bias) เท่ากับ 0.0000 ส่วนค่ารากที่สองของค่าเฉลี่ยกำลังที่สองของความคลาดเคลื่อน (RMSE) มีค่าอยู่ระหว่าง 0.1194 ถึง 1.2377 โดยคะแนนพัฒนาการจากวิธีการวัด DS และ SR มีค่า RMSE ต่ำที่สุดคือ 0.1194 และ 0.3288 ตามลำดับ และคะแนนพัฒนาการจากวิธีการวัด SRR มีค่า RMSE สูงที่สุดคือ 1.2377

4.2.2 ข้อมูลวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าคะแนนพัฒนาการจากทุกกลุ่มวิธีส่วนใหญ่มีค่าความลำเอียง (bias) เท่ากับ 0.0000 ยกเว้นคะแนนพัฒนาการจากวิธีการวัด MRMLC ที่มีค่าเท่ากับ 0.0740 ส่วนค่ารากที่สองของค่าเฉลี่ยกำลังที่สองของความคลาดเคลื่อน (RMSE) มีค่าอยู่ระหว่าง 0.3185 ถึง 1.0664 โดยคะแนนพัฒนาการจากวิธีการวัด DS และ MRMLC มีค่า RMSE ต่ำที่สุดคือ 0.3185 และ 0.3449 ตามอันดับ และคะแนนพัฒนาการจากวิธีการวัด SRR มีค่า RMSE สูงที่สุดคือ 1.0664

ส่วนข้อมูลที่ตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการวิเคราะห์พบว่าคะแนนพัฒนาการจากทุกกลุ่มวิธีมีค่าความลำเอียง (bias) เท่ากับ 0.0000 ส่วนค่ารากที่สองของค่าเฉลี่ยกำลังที่สองของความคลาดเคลื่อน (RMSE) มีค่าอยู่ระหว่าง 0.3048 ถึง 1.1937 โดย คะแนนพัฒนาการจากวิธีการวัด DS และ SC มีค่า RMSE ต่ำที่สุดคือ 0.3048 และ 0.4557 ตามอันดับ และคะแนนพัฒนาการจากวิธีการวัด SRR มีค่า RMSE สูงที่สุดคือ 1.1937

4.2.3 ข้อมูลวัดซ้ำด้วยแบบทดสอบคู่ขนาน (equivalence1) ที่มีการวัด 3 ครั้ง ข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) ผลการวิเคราะห์พบว่าคะแนนพัฒนาการจากวิธีการวัดแต่ละวิธีมีค่าความลำเอียง (bias) อยู่ระหว่าง – 0.0047 ถึง 0.0014 เกือบทุกวิธีมีค่า bias เป็นค่าลบ ยกเว้นคะแนนพัฒนาการจากวิธีการวัด NL และ SRR มีค่าเป็นบวก โดยคะแนนพัฒนาการจากวิธีการวัด MRMLC, NL, DS และ SC มีค่าของ bias ใกล้ศูนย์ และคะแนนพัฒนาการจากวิธีการวัด Sθ มีค่า bias ต่ำที่สุดคือ – 0.0047 ส่วนค่ารากที่สองของค่าเฉลี่ยกำลังที่สองของความคลาดเคลื่อน (RMSE) มีค่าอยู่ระหว่าง 0.2317 ถึง 1.0581 โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีค่า RMSE ต่ำที่สุดคือ 0.2317 รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด DS และSR ซึ่งมีค่า RMSE เท่ากับ 0.7455 และ 0.7615 ตามอันดับ และคะแนนพัฒนาการจากวิธีการวัด Sθ มีค่า RMSE สูงที่สุดคือ 1.0581

ส่วนข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit) ผลการวิเคราะห์พบว่าคะแนนพัฒนาการจากวิธีการวัดแต่ละวิธีมีค่าความลำเอียง (bias) อยู่ระหว่าง – 0.00032 ถึง 0.0075 โดยคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมมีค่า bias เป็นลบ มีค่าอยู่ระหว่าง –0.0043 ถึง – 0.0007 ส่วนคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบมีค่า bias เป็นบวก มีค่าอยู่ระหว่าง 0.0063 ถึง 0.0075 สำหรับค่ารากที่สองของค่าเฉลี่ยกำลังที่สองของความคลาดเคลื่อน (RMSE) มีค่าอยู่ระหว่าง 0.3070 ถึง 1.0902 โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีค่า RMSE ต่ำที่สุดคือ 0.3070 รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด SR มีค่า RMSEเท่ากับ 0.7045 คะแนนพัฒนาการจากวิธีการวัด Vθ มีค่า RMSE สูงที่สุดคือ 1.0902

4.2.4 ข้อมูลวัดซ้ำด้วยแบบทดสอบคู่ขนาน (equivalence2) ที่มีการวัด 5 ครั้ง ข้อมูลที่ตรวจให้คะแนนแบบทวิวิภาค (dichotomous) พบว่าคะแนนพัฒนาการจากวิธีการวัดแต่ละวิธีมีค่าความลำเอียง (bias) อยู่ระหว่าง – 0.0034 ถึง 0.0115 โดยคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมมีค่า bias เป็นบวก มีค่าอยู่ระหว่าง 0.0024 ถึง 0.0115 ส่วนคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบเกือบทุกวิธีมีค่าเป็นลบ คือมีค่าอยู่ระหว่าง - 0.0034 ถึง 0.0001 ยกเว้นคะแนนพัฒนาการจากวิธีการวัด MRMLC ที่มีค่าเป็นบวก 0.0095 และมีค่า bias สูงสุด สำหรับค่ารากที่สองของค่าเฉลี่ยกำลังที่สองของความคลาดเคลื่อน

(RMSE) มีค่าอยู่ระหว่าง 0.6894 ถึง 1.0481 โดยคะแนนพัฒนาการจากวิธีการวัด SRR มีค่า RMSE ต่ำที่สุดคือ 0.6894 รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด SR มีค่า RMSE เท่ากับ 0.7464 และคะแนนพัฒนาการจากวิธีการวัด Sθ มีค่า RMSE สูงที่สุดคือ 1.0481

สรุปโดยภาพรวม ผลการวิเคราะห์ค่าความคลาดเคลื่อนของคะแนนพัฒนาการระหว่างกลุ่มวิธี พบว่า ค่าความคลาดเคลื่อนของคะแนนพัฒนาการที่ประมาณค่าจากวิธีต่างๆในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมมีค่าต่ำกว่ากลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา ส่วนค่าความคลาดเคลื่อนของคะแนนพัฒนาการที่ประมาณค่าจากกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบมีค่าใกล้เคียงกับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา เมื่อวิเคราะห์ภายในกลุ่มวิธี พบว่า ภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม คะแนนพัฒนาการจากวิธีการวัด DS มีค่าความคลาดเคลื่อนต่ำที่สุด ยกเว้นในข้อมูลชุดวัดซ้ำ 3 ครั้ง และ 5 ครั้ง ด้วยแบบทดสอบคู่ขนานที่คะแนนพัฒนาการจากวิธีการวัด SRR มีค่าความคลาดเคลื่อนต่ำที่สุด ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบคะแนนพัฒนาการจากวิธีการวัด MRMLC มีค่าความคลาดเคลื่อนต่ำกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ และในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาข้อมูลเกือบทุกชุดคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าความคลาดเคลื่อนใกล้เคียงกับคะแนนพัฒนาการจากวิธีการวัด Sθ

ตารางที่ 4.36 ค่าความลำเอียงและค่ารากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน จำแนกตามชุดของข้อมูลและวิธีการให้คะแนน

กลุ่มวิธี	repeat						Counterbalance					
	dichotomous			partial credit			dichotomous			partial credit		
	N	bias	RMSE	N	bias	RMSE	N	bias	RMSE	N	bias	RMSE
วิธีCTT												
DS	635	0.0000	0.2612	635	0.0000	0.1194	581	0.0000	0.3185	581	0.0000	0.3048
SC	635	0.0000	0.6320	635	0.0000	0.5924	581	0.0000	0.5326	581	0.0000	0.4557
NL	634	0.0006	0.6323	635	0.0000	0.4986	581	0.0000	0.6189	581	0.0000	0.5280
SR	635	0.0000	0.2886	635	0.0000	0.3288	581	0.0000	0.5020	581	0.0000	0.6205
SRR	635	0.0000	0.9503	635	0.0000	1.2377	581	0.0000	1.0664	581	0.0000	1.1937
วิธี IRT												
Lθ	635	0.0000	0.6633	635	0.0000	0.6213	581	0.0000	0.7782	581	0.0000	0.7151
MRMLC	615	0.0596	0.3235				550	0.0740	0.3449			
วิธีที่พัฒนา												
Sθ	635	0.0000	0.8048	635	0.0000	0.6438	581	0.0000	0.8709	581	0.0000	0.7324
Vθ	635	0.0000	0.6508	635	0.0000	0.6205	581	0.0000	0.7792	581	0.0000	0.8193

ตารางที่ 4.36 (ต่อ)

กลุ่มวิธี	equivalence 1						equivalence 2		
	dichotomous			partial credit			dichotomous		
	N	BIAS	RMSE	N	BIAS	RMSE	N	BIAS	RMSE
วิธีCTT									
DS	378	-0.0007	0.7455	378	-0.0032	0.7558	362	0.0112	0.7674
SC	378	-0.0007	0.8252	378	-0.0032	0.7651	362	0.0063	0.9087
NL	378	0.0001	0.8811	378	-0.0024	0.8920	362	0.0064	0.9337
SR	378	-0.0018	0.7615	378	-0.0043	0.7045	362	0.0115	0.7464
SRR	378	0.0014	0.2317	378	-0.0007	0.3070	362	0.0024	0.6894
วิธี IRT									
Lθ	378	-0.0040	0.9375	378	0.0073	1.0625	362	-0.0007	0.9627
MRMLC	378	-0.0002	0.8063				362	0.0095	0.8237
วิธีที่พัฒนา									
Sθ	378	-0.0047	1.0581	378	0.0063	1.0731	362	-0.0034	1.0481
Vθ	378	-0.0040	0.9078	378	0.0075	1.0902	362	-0.0001	0.9486

ตอนที่ 5 ผลการเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการ

การเปรียบเทียบคุณภาพของคะแนนพัฒนาการที่ประมาณค่าจากกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา กับคะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่เป็นวิธีเกณฑ์ใน 2 ประเด็นคือ 1) ค่าความตรงตามเกณฑ์สัมพันธ์ซึ่งพิจารณาจากค่าความสัมพันธ์กับเกณฑ์ 2 วิธีคือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการที่ประมาณค่าจากวิธีการวัดใน 3 กลุ่มวิธีกับคะแนนพัฒนาการที่ประมาณค่าจากวิธีเกณฑ์ ค่าสหสัมพันธ์อันดับระหว่างอันดับของคะแนนพัฒนาการที่ประมาณค่าจากวิธีการวัดใน 3 กลุ่มวิธีกับอันดับของคะแนนพัฒนาการที่ประมาณค่าจากวิธีเกณฑ์ 2) ค่าความคลาดเคลื่อนของคะแนนพัฒนาการที่พิจารณาจากค่าความลำเอียงและค่ารากที่สองของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน ผลสรุปการวิเคราะห์มีดังนี้

5.1 ผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ของวิธีการวัดคะแนนพัฒนาการ

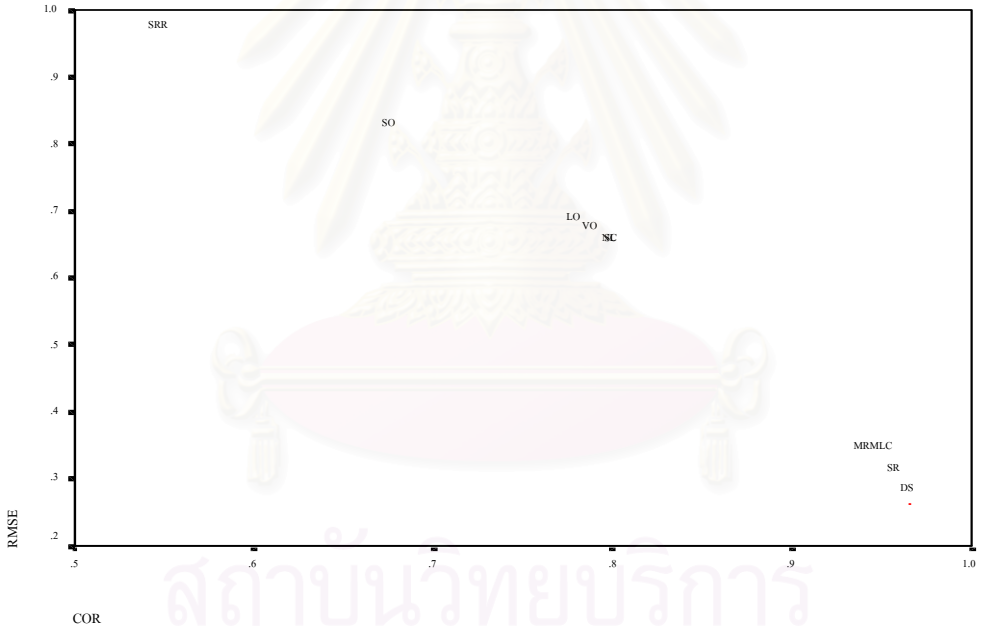
ผลการวิเคราะห์ค่าความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการในแต่ละกลุ่มวิธี เมื่อพิจารณาจากค่าความสัมพันธ์กับเกณฑ์ 2 วิธีคือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนพัฒนาการที่ประมาณค่าจากวิธีการวัดใน 3 กลุ่มวิธีกับคะแนนพัฒนาการที่ประมาณค่าจากวิธีเกณฑ์ และค่าสหสัมพันธ์อันดับระหว่างอันดับของคะแนนพัฒนาที่ประมาณค่าจากวิธีการวัดใน 3 กลุ่มวิธีกับอันดับของคะแนนพัฒนาการที่ประมาณค่าจากวิธีเกณฑ์ พบว่า กลุ่มวิธีของ

ทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีค่าความตรงตามเกณฑ์สัมพันธ์ไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ยกเว้นในข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน ที่มีการให้คะแนนความรู้บางส่วน ที่พบว่ากลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมมีค่าความตรงตามเกณฑ์สัมพันธ์สูงมากที่สุด รองลงมาคือกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา ตามลำดับ ภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม คะแนนพัฒนาการจากวิธีการวัด DS มีค่าความตรงตามเกณฑ์สัมพันธ์สูงที่สุด รองลงมาคือคะแนนพัฒนาการจากวิธีการวัด SR ยกเว้นในข้อมูลชุดวัดซ้ำ 3 ครั้ง และ 5 ครั้งด้วยแบบทดสอบคู่ขนาน ที่พบว่าคะแนนพัฒนาการจากวิธีการวัด SRR มีค่าความตรงตามเกณฑ์สัมพันธ์สูงที่สุด ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบคะแนนพัฒนาการจากวิธีการวัด MRMLC มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ ยกเว้นข้อมูลที่ตรวจให้คะแนนความรู้บางส่วนที่พบว่าคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าความตรงตามเกณฑ์สัมพันธ์แตกต่างจากคะแนนพัฒนาการจากวิธีการวัด Sθ อย่างไม่มีนัยสำคัญทางสถิติ และคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าความตรงตามเกณฑ์สัมพันธ์ต่ำกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ ในข้อมูลชุดวัดซ้ำแบบจตุรัสละตินที่มีการตรวจให้คะแนนความรู้บางส่วน

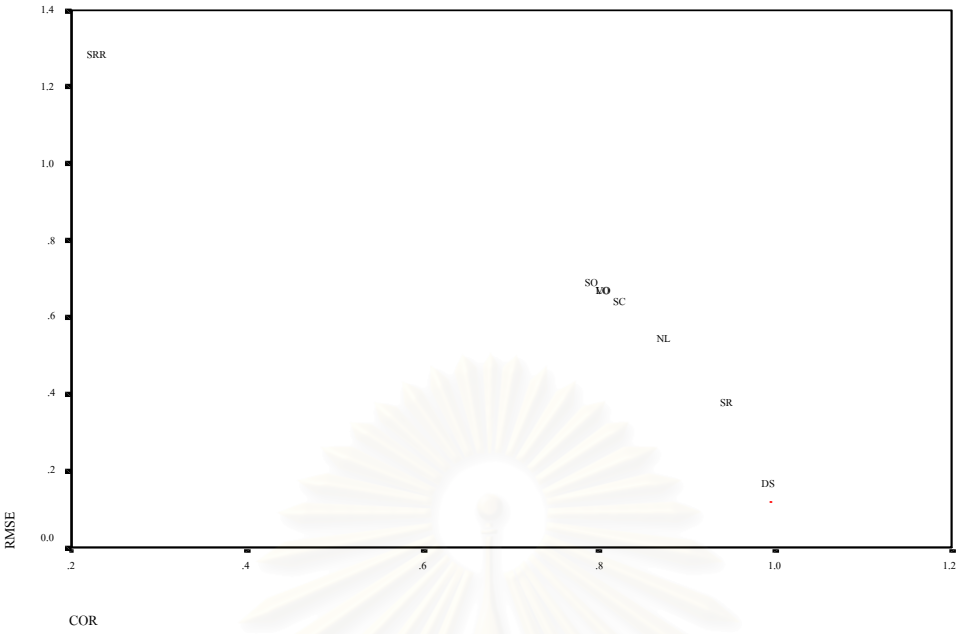
5.2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการ

ผลการวิเคราะห์ค่าความคลาดเคลื่อนของคะแนนพัฒนาการที่เบี่ยงเบนจากคะแนนพัฒนาการที่เป็นเกณฑ์ในแต่ละกลุ่มวิธี พบว่า กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมมีค่าความคลาดเคลื่อนต่ำที่สุด รองลงมาคือกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา ตามลำดับ ภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมคะแนนพัฒนาการจากวิธีการวัด DS มีค่าความคลาดเคลื่อนต่ำที่สุดรองลงมาคือคะแนนพัฒนาการจากวิธีการวัด SR และ SRR มีค่าความคลาดเคลื่อนสูงที่สุด ยกเว้นในข้อมูลชุดวัดซ้ำ 3 ครั้ง และ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน ที่พบว่าคะแนนพัฒนาการจากวิธีการวัด SRR มีค่าความคลาดเคลื่อนต่ำที่สุด ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบคะแนนพัฒนาการจากวิธีการวัด MRMLC มีค่าความคลาดเคลื่อนต่ำกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าความความคลาดเคลื่อนใกล้เคียงกับคะแนนพัฒนาการจากวิธีการวัด Sθ

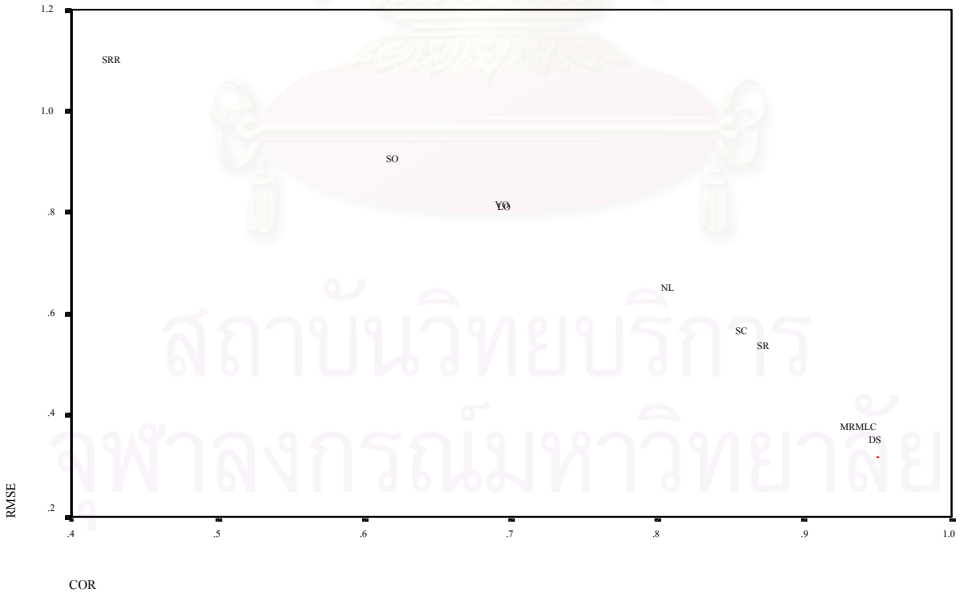
โดยสรุป ผลการเปรียบเทียบคุณภาพของคะแนนพัฒนาการที่ได้จากวิธีการวัดพัฒนาการระหว่างกลุ่มวิธี พบว่า คะแนนพัฒนาการของแต่ละกลุ่มวิธีมีคุณภาพไม่แตกต่างกัน สำหรับคุณภาพของคะแนนพัฒนาการภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมคะแนนพัฒนาการจากวิธีการวัด DS และ SR มีคุณภาพสูงกว่าคะแนนพัฒนาการที่ได้จากวิธีอื่นๆ สำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม และข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน ส่วนคะแนนพัฒนาการจากวิธีการวัด SRR มีคุณภาพสูงกว่าคะแนนพัฒนาการจากวิธีอื่นๆ สำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบคะแนนพัฒนาการจากวิธีการวัด MRMLC มีคุณภาพสูงกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ ภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาคะแนนพัฒนาการจากวิธีการวัด Vθ มีคุณภาพสูงกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค (รายละเอียดดังแผนภาพที่ 4.22 - 4.28)



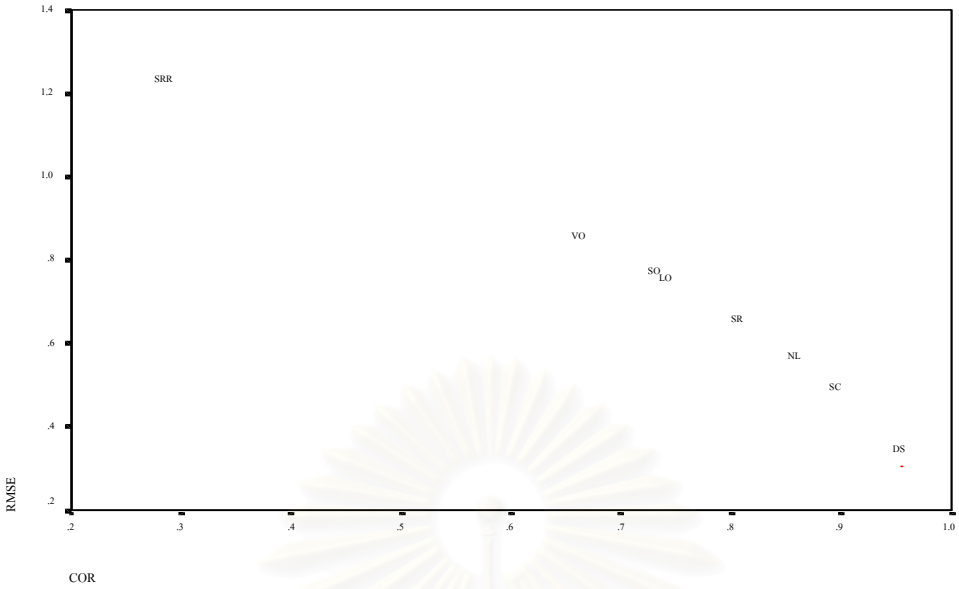
แผนภาพที่ 4.22 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)



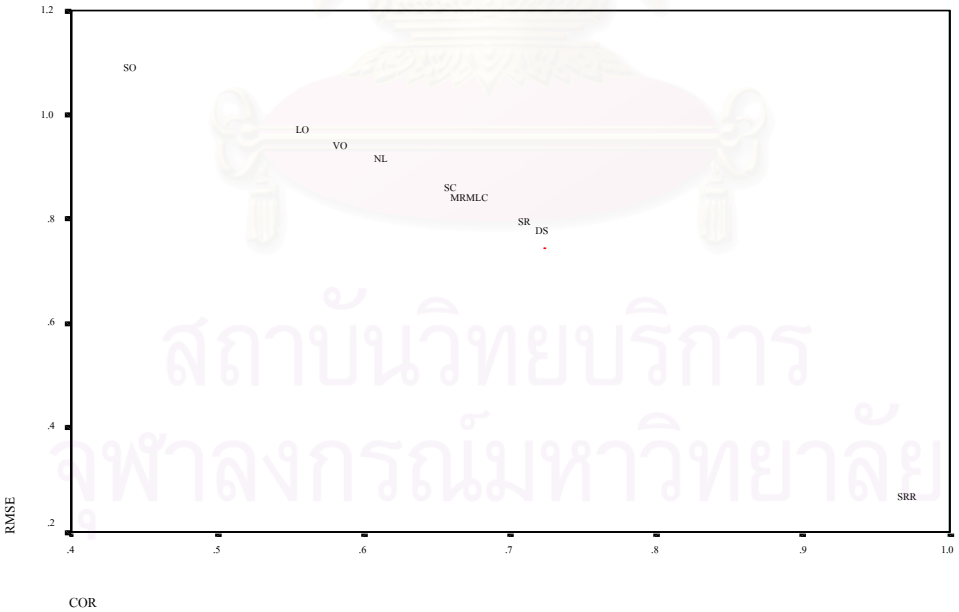
แผนภาพที่ 4.23 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit)



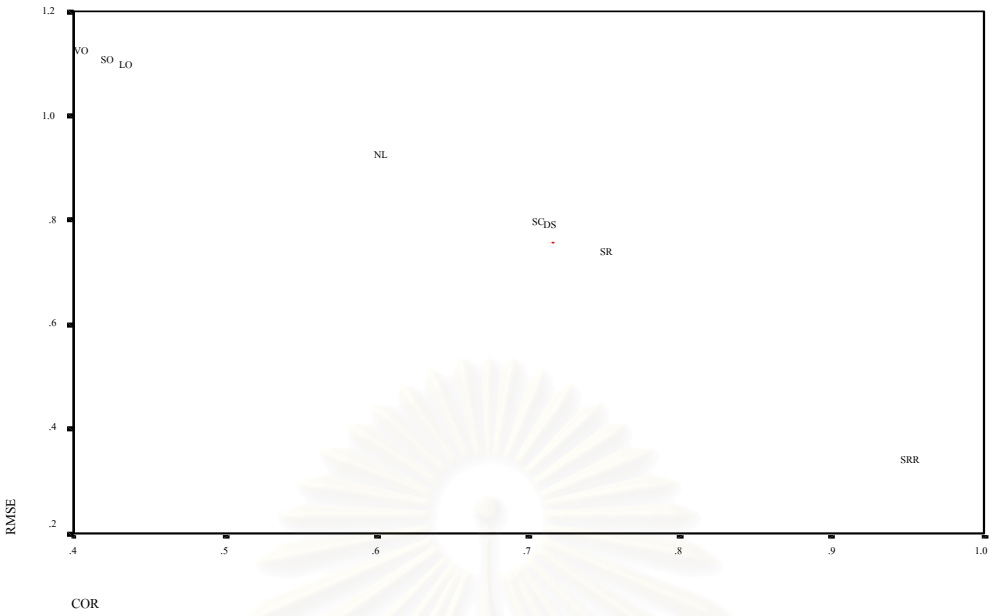
แผนภาพที่ 4.24 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)



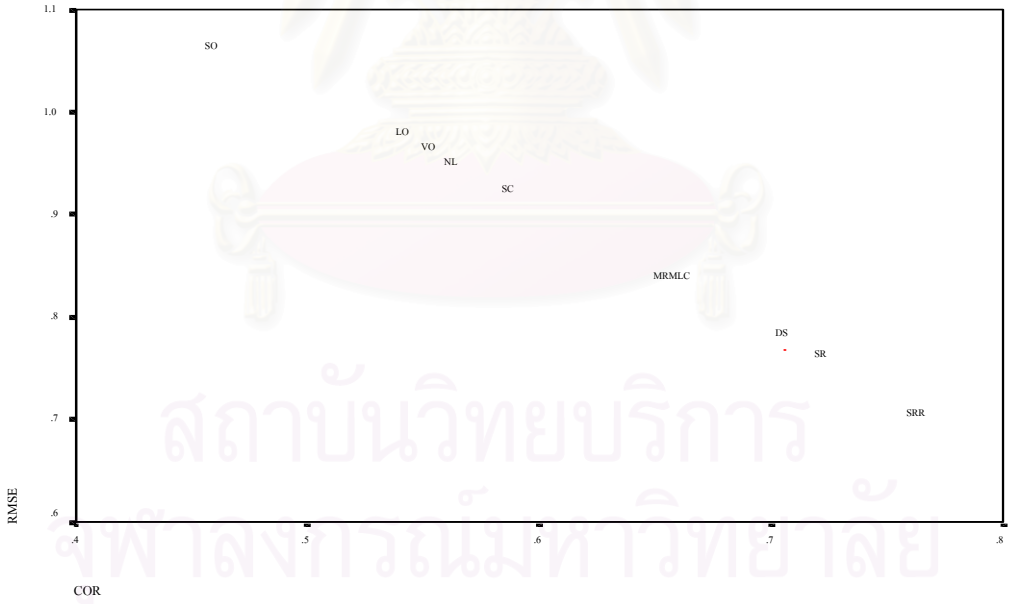
แผนภาพที่ 4.25 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit)



แผนภาพที่ 4.26 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 1) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)



แผนภาพที่ 4.27 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน (equivalence 2) ที่มีการตรวจให้คะแนนความรู้บางส่วน (partial credit)



แผนภาพที่ 4.28 ค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีของข้อมูลชุดวัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน (counterbalance 2) ที่มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous)

สรุปผลการศึกษาและการวิเคราะห์วิธีการวัดคะแนนพัฒนาการใน 3 กลุ่มวิธีดังนี้

วิธีวัด	จุดเด่น	จุดอ่อน
กลุ่มวิธี CTT 1. DS	1. ประมาณค่าง่ายมาก 2. ลักษณะการแจกแจงใกล้เคียงกับวิธีเกณฑ์มากที่สุด 3. ความตรงตามเกณฑ์สัมพันธ์สูงมาก 4. ความคลาดเคลื่อนต่ำมาก	1 .มีความสัมพันธ์ทางลบกับคะแนนครั้งแรก 2. คะแนนพัฒนาการที่เท่ากัน มีความหมายต่างกันถ้ามาจากคนที่มีความสามารถเริ่มต้นที่ต่างกัน
2. SC	1. ประมาณค่าง่ายมาก 2. ลักษณะการแจกแจงใกล้เคียงกับวิธีเกณฑ์ปานกลาง 3. ความตรงตามเกณฑ์สัมพันธ์สูง 4. ความคลาดเคลื่อนปานกลาง 5. สามารถเปรียบเทียบคะแนนพัฒนาการระหว่างวิชาที่มีมาตรการวัดต่างกันได้	1. ขึ้นกับความสามารถของกลุ่ม 2. พิสัยของคะแนน -4 ถึง 4 ทำให้เข้าใจความหมายของคะแนนค่อนข้างยาก
3. NL	1. ประมาณค่าง่าย ปานกลาง 2. ลักษณะการแจกแจงใกล้เคียงกับวิธีเกณฑ์ปานกลาง 3. ความตรงตามเกณฑ์สัมพันธ์สูง 4. ความคลาดเคลื่อนปานกลาง 5. สามารถเปรียบเทียบคะแนนพัฒนาการระหว่างวิชาที่มีมาตรการวัดต่างกันได้ 6. เหมาะกับคะแนนที่ครั้งแรกและครั้งหลังมีลักษณะเป็นสัดส่วนกัน	1. ไม่สามารถประมาณค่าได้สำหรับผู้สอบที่ได้คะแนนเป็นศูนย์ 2. คะแนนพัฒนาการที่ได้แปลความหมายยาก
4. SR	1. ประมาณค่าง่ายมาก 2. ลักษณะการแจกแจงใกล้เคียงกับวิธีเกณฑ์มากที่สุด 3. ความตรงตามเกณฑ์สัมพันธ์สูงมาก 4. ความคลาดเคลื่อนต่ำ 5. ไม่มีอิทธิพลเพดาน 6. แปลความหมายง่าย 7. เปรียบเทียบคะแนนพัฒนาการกันได้ระหว่างกลุ่ม	1. ไม่สามารถประมาณค่าได้สำหรับผู้สอบที่ได้คะแนนเต็มในการสอบก่อนเรียน
5. SRR	1. ไม่มีอิทธิพลเพดาน 2. เปรียบเทียบคะแนนพัฒนาการกันได้ระหว่างกลุ่ม	1. ลักษณะการแจกแจง ความตรงตามเกณฑ์สัมพันธ์ และความคลาดเคลื่อนขึ้นอยู่กับตำแหน่งของข้อมูลที่เป็นค่าสุดโต่ง 2. คำนวณค่ายาก หลายขั้นตอน
วิธี IRT 1. Lθ	1. ประมาณค่าง่าย 2. มีการนำค่าพารามิเตอร์ของข้อสอบมาร่วมในการประมาณค่าคะแนนพัฒนา 3. ลักษณะการแจกแจงใกล้เคียงกับวิธีเกณฑ์ค่อนข้างมาก 4. ความตรงตามเกณฑ์สูง	1. ความคลาดเคลื่อนค่อนข้างสูง 2. ต้องใช้กับผู้สอบจำนวนมาก

วิธีวัด	จุดเด่น	จุดอ่อน
2. MRMLC	1. ประมาณค่าง่าย 2. มีการนำค่าพารามิเตอร์ของข้อสอบมาร่วมในการประมาณค่าคะแนนพัฒนา 3. ลักษณะการแจกแจงใกล้เคียงกับวิธีเกณฑ์ค่อนข้างมาก 4. ความตรงตามเกณฑ์สัมพันธ์สูงมาก 5. ความคลาดเคลื่อนต่ำ	1. ต้องใช้ข้อมูลเป็นรายข้อคำถามทำให้เสียเวลา 2. ต้องใช้โปรแกรมเฉพาะ
วิธีที่พัฒนา 1. Sθ	1. ประมาณค่าง่าย 2. มีการนำค่าพารามิเตอร์ของข้อสอบมาร่วมในการประมาณค่าคะแนนพัฒนา 3. ความตรงตามเกณฑ์สัมพันธ์ปานกลาง	1. ลักษณะการแจกแจงต่างจากวิธีเกณฑ์มาก 2. ความคลาดเคลื่อนสูงมาก 3. ต้องใช้กับผู้สอบจำนวนมาก
2. Vθ	1. ประมาณค่าง่าย 2. มีการนำค่าพารามิเตอร์ของข้อสอบมาร่วมในการประมาณค่าคะแนนพัฒนา 3. ลักษณะการแจกแจงใกล้เคียงกับวิธีเกณฑ์มาก 4. ความตรงตามเกณฑ์สัมพันธ์ปานกลาง	1. ความคลาดเคลื่อนค่อนข้างสูง 2. ต้องใช้กับผู้สอบจำนวนมาก

สรุปผลการวิจัยในด้านการนำวิธีการวัดไปใช้ วิธีการวัดคะแนนพัฒนาการที่เหมาะสมที่จะให้ครูนำไปใช้ในโรงเรียนได้แก่ วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของคะแนนดิบ และวิธีการวัดคะแนนพัฒนาการสัมพัทธ์ เพราะเป็นวิธีที่ประมาณค่าคะแนนพัฒนาการได้ง่าย มีความตรงตามเกณฑ์สัมพันธ์สูง และมีค่าความคลาดเคลื่อนต่ำ สำหรับครูที่มีความสามารถในการวิเคราะห์ข้อมูลด้วยโปรแกรมสำเร็จรูปควรที่จะใช้วิธีวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง เพราะจะแก้ปัญหาในด้านคะแนนพัฒนาการที่เท่ากันมีความหมายไม่เหมือนกันถ้ามาจากผู้สอบที่มีความสามารถเริ่มต้นที่ไม่เท่ากัน นอกจากนั้นคุณภาพทั้งด้านค่าความตรงตามเกณฑ์สัมพันธ์และค่าความคลาดเคลื่อนอยู่ในเกณฑ์สูงใกล้เคียงกับวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของคะแนนดิบและวิธีการวัดคะแนนพัฒนาการสัมพัทธ์

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาวิธีการประมาณค่าคะแนนพัฒนาการและวิเคราะห์เปรียบเทียบคุณภาพของคะแนนพัฒนาการระหว่างกลุ่มวิธีการวัดคะแนนพัฒนาการ 3 กลุ่ม ได้แก่กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม 5 วิธี กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ 2 วิธี และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา 2 วิธี

กลุ่มวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมที่ใช้ในการวิจัยครั้งนี้มี 5 วิธีได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างระหว่างคะแนนดิบ 2) วิธีการวัดคะแนนพัฒนาการจากคะแนนมาตรฐาน 3) วิธีการวัดคะแนนพัฒนาการจากลอการิทึมของคะแนนดิบ 4) วิธีการวัดคะแนนพัฒนาการสัมพัทธ์ 5) วิธีการวัดคะแนนพัฒนาการส่วนที่เหลือเทียบกับศักยภาพ ส่วนกลุ่มวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบสนองข้อสอบมี 2 วิธี ได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความแตกต่างของความสามารถที่แท้จริง 2) วิธีการวัดคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (Multidimensional Rasch Model for Learning and Change: MRMLC) และกลุ่มวิธีการวัดคะแนนพัฒนาการของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนามี 2 วิธีได้แก่ 1) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา 2) วิธีการวัดคะแนนพัฒนาการจากความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน

กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 2 กลุ่ม กลุ่มละ 698 คน และ 637 คน ในภาคเรียนที่ 1 ปีการศึกษา 2543 จากโรงเรียนสังกัดกรมสามัญศึกษาเขตกรุงเทพมหานคร รวม 12 โรงเรียน

เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ในระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับที่มีลักษณะคู่ขนานกัน ซึ่งผู้วิจัยสร้างเอง เป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก มีจำนวนข้อสอบฉบับละ 66 ข้อ

ข้อมูลที่ใช้ในการวิจัยมี 3 ชุด 2 ชุดแรกเป็นข้อมูลปฐมภูมิ ที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ จากกลุ่มตัวอย่างที่เป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 ทั้ง 2 กลุ่ม การเก็บรวบรวมข้อมูลระยะยาวทำโดยการวัด 3 ครั้ง ห่างกันครั้งละ 4 สัปดาห์ ข้อมูลชุดแรกเป็นการวัดซ้ำด้วยแบบทดสอบฉบับเดิม ข้อมูลชุดที่สองเป็นการวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน ส่วนข้อมูลชุดที่สาม เป็นข้อมูลทุติยภูมิที่ได้จากฐานข้อมูลในงานวิจัยของสมถวิล วิจิตรวรรณา

(2543) ที่ได้จากคะแนนสอบวิชาคณิตศาสตร์ จากกลุ่มตัวอย่างที่เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ที่มีการวัดซ้ำ 5 ครั้ง ผู้วิจัยได้แยกวิเคราะห์ข้อมูล 2 แบบ คือ แบบวัดซ้ำ 3 ครั้ง และแบบวัดซ้ำ 5 ครั้ง ข้อมูลทั้ง 3 ชุด มีการตรวจให้คะแนนแบบทวิวิภาค (dichotomous) และตรวจให้คะแนนความรู้บางส่วน (partial credit) ยกเว้นข้อมูลชุดที่สามแบบวัดซ้ำ 5 ครั้งที่มีการตรวจให้คะแนนแบบทวิวิภาคเท่านั้น ดังนั้นในการวิจัยครั้งนี้ผู้วิจัยจึงมีฐานข้อมูลทั้งหมด 7 ชุด

การวิเคราะห์ข้อมูล สำหรับการประมาณค่าคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมใช้โปรแกรม SPSS-WIN 10.07 การประมาณค่าคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาใช้โปรแกรม BILOG 3 สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค และใช้โปรแกรม PARSCALE 2 สำหรับข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วน การประมาณค่าคะแนนพัฒนาการโดยใช้โมเดลราส์ชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลงใช้โปรแกรม LPCM-WIN 1.0 การประมาณค่าคะแนนพัฒนาการจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่เป็นวิธีเกณฑ์ใช้โปรแกรม LISREL 8.3 การวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการแต่ละวิธีใช้ค่าสถิติ 2 ประเภท ประเภทแรกคือ ความตรงตามเกณฑ์สัมพันธ์ ในการวิจัยครั้งนี้หาค่าความตรงตามเกณฑ์สัมพันธ์ ใช้ค่าสถิติสัมประสิทธิ์สหสัมพันธ์ของ Pearson และสหสัมพันธ์อันดับของ Spearman ประเภทที่สอง คือค่าความคลาดเคลื่อนในการประมาณค่าคะแนนพัฒนาการ ในการวิจัยครั้งนี้เป็นการประมาณค่าความคลาดเคลื่อนแบบค่าความลำเอียง (bias) และแบบค่ารากที่สองของค่าเฉลี่ยกำลังสองของคะแนนการวัดพัฒนาการที่เบี่ยงเบนจากเกณฑ์ โดยใช้โปรแกรม SPSS-WIN 10.07 สำหรับทดสอบความแตกต่างของค่าความตรงตามเกณฑ์สัมพันธ์ระหว่างกลุ่มวิธีใช้การวิเคราะห์พหุระดับด้วยโปรแกรม HLM - WIN 5.04 ส่วนการทดสอบความแตกต่างของค่าความตรงตามเกณฑ์สัมพันธ์ระหว่างวิธีการวัดคะแนนพัฒนาการภายในแต่ละกลุ่มวิธีใช้ค่าสถิติทดสอบของ Hotelling ด้วยโปรแกรม EXCEL

การนำเสนอผลสรุปการวิจัยแบ่งเป็น 4 ตอน ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติเบื้องต้นของข้อมูลทั้ง 3 ชุด ตอนที่ 2 ผลการวิเคราะห์เพื่อสร้างและตรวจสอบความตรงของโมเดลพัฒนาการที่มีตัวแปรแฝงว่าโมเดลสอดคล้องกลมกลืนกับคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์หรือไม่ เพื่อให้ได้คะแนนพัฒนาการที่ใช้เป็นเกณฑ์ ตอนที่ 3 ผลการวิเคราะห์ลักษณะการแจกแจงของคะแนนพัฒนาการทางคณิตศาสตร์ที่ได้จากวิธีการวัดของ 3 กลุ่มวิธีได้แก่ กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา เมื่อเทียบกับวิธีเกณฑ์ ตอนที่ 4 ผลการเปรียบเทียบคุณภาพของ

วิธีการวัดคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างกลุ่มวิธีและภายในกลุ่มวิธีการวัดทั้ง 3 กลุ่มวิธี

สรุปผลการวิจัย

ตอนที่ 1 ผลวิเคราะห์ค่าสถิติเบื้องต้นของข้อมูลทั้ง 3 ชุด

ข้อมูลทั้ง 3 ชุด ได้แก่ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม (repeat) ข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน (counterbalance) และข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน (equivalence) ที่มีการตรวจให้คะแนนแบบทวิวิภาค และตรวจให้คะแนนความรู้บางส่วน เมื่อนำมาวิเคราะห์เปรียบเทียบลักษณะการกระจายของข้อมูลจากการวัดซ้ำ พบว่า ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนที่ได้จากการวัดแต่ละครั้งมีค่าเพิ่มขึ้น ลักษณะการแจกแจงของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดในแต่ละครั้งเป็นแบบเบ้ขวา และมีลักษณะเบ้ขวาน้อยลงในครั้งที่ 3 และ 4 ส่วนครั้งที่ 5 มีแนวโน้มเบ้ขวามากขึ้น ลักษณะการแจกแจงของคะแนนที่ได้จากการวัดมีความโด่งสูงกว่าโค้งปกติ และค่าความโด่งลดลงในการวัดครั้งต่อมาจนมีลักษณะแบนกว่าโค้งปกติในการวัดครั้งสุดท้าย เมื่อพิจารณาจากค่าความคลาดเคลื่อนของค่าความเบ้และค่าความโด่ง พบว่า คะแนนจากการวัดทุกครั้งมีลักษณะการแจกแจงที่ใกล้เคียงกับโค้งปกติ

ผลการทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนผลสัมฤทธิ์ทางคณิตศาสตร์ที่ได้จากการวัดระหว่างครั้งที่วัด พบว่า ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม และข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน แบบที่มีการตรวจให้คะแนนแบบทวิวิภาคและแบบที่มีการตรวจให้คะแนนความรู้บางส่วนมีคะแนนเฉลี่ยจากการวัดแต่ละครั้งทั้ง 3 ครั้ง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 สำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน พบว่า คะแนนเฉลี่ยจากการวัดในแต่ละครั้งทั้ง 5 ครั้งมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นคะแนนเฉลี่ยจากการวัดในครั้งที่ 1 กับครั้งที่ 2 และครั้งที่ 4 กับครั้งที่ 5 ที่ไม่พบความแตกต่างอย่างมีนัยสำคัญทางสถิติ

ตอนที่ 2 ผลการวิเคราะห์เพื่อสร้างและตรวจสอบความตรงของโมเดลพัฒนาการที่มีตัวแปรแฝง

ในการวิจัยครั้งนี้ผู้วิจัยได้สร้างและตรวจสอบความตรงของโมเดลพัฒนาการที่มีตัวแปรแฝงทั้งหมด 5 โมเดล คือ โมเดลพัฒนาการที่ไม่มีความชัน (NSB) โมเดลพัฒนาการเชิงเส้น (LIN) โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์คงที่ (FIX) โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดค่าพารามิเตอร์อิสระ (FRE) และโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์อิสระและมีความแปรปรวนของค่าความคลาดเคลื่อนไม่เท่ากัน (UDV) การตรวจสอบความตรงดังกล่าวเป็นการตรวจสอบกับข้อมูลทั้ง 7 ฐานเพื่อให้ได้คะแนน

พัฒนาการมาใช้เป็นเกณฑ์ ผลการวิเคราะห์พบว่า โมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุด คือโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงและกำหนดค่าพารามิเตอร์อิสระแบบความแปร ปรวนของความคลาดเคลื่อนไม่เท่ากัน (latent growth curve model with free parameter and unequal disturbance variance model: UDV model) ในข้อมูล 5 ฐาน ยกเว้น 2 ฐาน ฐานแรก คือข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนานที่ได้จากการวัด 3 ครั้งและมีการตรวจให้คะแนนแบบทวิวิภาคที่พบว่าโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุดคือโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงและกำหนดค่าพารามิเตอร์อิสระ(latent growth curve model with free parameter model: FRE model) และฐานที่สองคือข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนานที่ได้จากการวัด 5 ครั้งและมีการตรวจให้คะแนนความรู้บางส่วนพบว่าโมเดลที่ทดสอบทั้ง 5 โมเดลไม่มีโมเดลใดที่สอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์

ตอนที่ 3 ผลการวิเคราะห์ลักษณะการแจกแจงของคะแนนพัฒนาการทางคณิตศาสตร์ที่ได้จากวิธีการวัดของ 3 กลุ่มวิธีเมื่อเทียบกับวิธีเกณฑ์

ผลการวิเคราะห์เปรียบเทียบคะแนนพัฒนาการทางคณิตศาสตร์ที่ได้จากวิธีการวัดทุกวิธีกับคะแนนพัฒนาการที่ใช้เป็นเกณฑ์ พบว่า มีลักษณะการแจกแจงใกล้เคียงกัน ยกเว้นคะแนนพัฒนาการจากวิธีการวัด SRR ข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม และข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน ที่มีลักษณะการแจกแจงต่างจากวิธีเกณฑ์มาก ผลการวิเคราะห์เปรียบเทียบคะแนนพัฒนาการทางคณิตศาสตร์ที่ได้จากวิธีการวัดทุกวิธีภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม พบว่าคะแนนพัฒนาการจากวิธีการวัด DS มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการที่เป็นเกณฑ์มากกว่าวิธีอื่น ผลการวิเคราะห์เปรียบเทียบคะแนนพัฒนาการทางคณิตศาสตร์ที่ได้จากวิธีการวัดทุกวิธีภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่าคะแนนพัฒนาการจากวิธีการวัด Lθ มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการที่เป็นเกณฑ์มากกว่าคะแนนพัฒนาการจากวิธีการวัด MRMLC ผลการวิเคราะห์เปรียบเทียบคะแนนพัฒนาการทางคณิตศาสตร์ที่ได้จากวิธีการวัดทุกวิธีภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่าคะแนนพัฒนาการจากวิธีการวัด Vθ มีลักษณะการแจกแจงใกล้เคียงกับคะแนนพัฒนาการที่เป็นเกณฑ์มากกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ

ตอนที่ 4 ผลการวิเคราะห์เปรียบเทียบคุณภาพของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างกลุ่มวิธีการวัดและภายในกลุ่มวิธีการวัด 3 กลุ่มวิธี

4.1 ผลการเปรียบเทียบคุณภาพของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างกลุ่มวิธี

4.1.1 ผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างกลุ่มวิธี

ผลการวิเคราะห์ค่าความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการที่ได้จากวิธีการวัด 9 วิธีในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่า คะแนนพัฒนาการที่ได้จากทุกวิธีการวัดในทุกกลุ่มวิธีมีความสัมพันธ์ทางบวกค่อนข้างสูงกับคะแนนพัฒนาการที่เป็นเกณฑ์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ซึ่งสอดคล้องกับสมมติฐานในข้อ 1.1 ยกเว้นคะแนนพัฒนาการจากวิธีการวัด SRR ในข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิมและข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน ที่มีความสัมพันธ์ค่อนข้างต่ำ

ผลการเปรียบเทียบค่าเฉลี่ยความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการระหว่างกลุ่มวิธีการวัด 3 กลุ่มวิธี พบว่า กลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีค่าเฉลี่ยความตรงตามเกณฑ์สัมพันธ์ของข้อมูลทั้ง 7 ฐาน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ยกเว้นในข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน ที่มีการให้คะแนนความรู้บางส่วน ซึ่งพบว่ากลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมมีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่ากลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาที่ระดับความมีนัยสำคัญ 0.05 ผลการวิเคราะห์ให้ผลขัดแย้งกับสมมติฐานในข้อ 1.2 ที่ว่า คะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม

4.1.2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างกลุ่มวิธีการวัด

ผลการเปรียบเทียบค่าความคลาดเคลื่อนของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างกลุ่มวิธีการวัด 3 กลุ่มวิธี พบว่า คะแนนพัฒนาการที่ได้จากวิธีการวัดภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมส่วนใหญ่มีค่าความคลาดเคลื่อนต่ำกว่ากลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา ยกเว้นคะแนนพัฒนาการจากวิธีการวัด MRMLC ที่มีค่าความคลาดเคลื่อนค่อนข้างต่ำและมีค่าใกล้เคียงกับคะแนนพัฒนาการที่ได้จากวิธีการวัดภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม ผลการวิเคราะห์ไม่สอดคล้องกับสมมติฐานในข้อที่ 2.1 ที่ว่า คะแนนพัฒนาการจากวิธีการวัดในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีค่าความคลาดเคลื่อนต่ำกว่าคะแนนพัฒนาการที่ได้จากวิธีการวัดในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม

4.2 ผลการเปรียบเทียบคุณภาพของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธี

4.2.1 ผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธี

ผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดแต่ละวิธีภายในกลุ่มวิธีทั้ง 3 กลุ่มวิธี พบว่า ในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมคะแนนพัฒนาการจากวิธีการวัด DS และ SR ในข้อมูลทุกชุดมีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการจากวิธีอื่นๆที่ระดับนัยสำคัญ 0.01 ยกเว้นในข้อมูลชุดวัดซ้ำ 3 ครั้ง และ 5 ครั้งด้วยแบบทดสอบคู่ขนาน ซึ่งพบว่าคะแนนพัฒนาการจากวิธีการวัด SRR มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการจากวิธีอื่นๆที่ระดับนัยสำคัญ 0.01 ซึ่งสอดคล้องกับสมมติฐานในข้อ 1.3 บางส่วน ที่ว่าคะแนนพัฒนาการจากวิธีการวัด DS และ SR มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการจากวิธีอื่นๆ ผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างกลุ่มวิธีการวัดภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่าคะแนนพัฒนาการจากวิธีการวัด MRMLC มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ ที่ระดับนัยสำคัญ 0.01 ซึ่งไม่สอดคล้องกับสมมติฐานในข้อ 1.4 ที่ว่าคะแนนพัฒนาการจากวิธีการวัด Lθ มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด MRMLC ผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่าคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ ที่ระดับนัยสำคัญ 0.01 ยกเว้นข้อมูลที่ตรวจให้คะแนนความรู้บางส่วนที่พบว่าคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าความตรงตามเกณฑ์สัมพันธ์แตกต่างจากคะแนนพัฒนาการจากวิธีการวัด Sθ อย่างไม่มีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับสมมติฐานในข้อ 1.5 บางส่วน ที่ว่าคะแนนพัฒนาการจากวิธีการวัด Vθ มีค่าความตรงตามเกณฑ์สัมพันธ์ไม่แตกต่างจากคะแนนพัฒนาการจากวิธีการวัด Sθ

ในการวิเคราะห์เปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ครั้งนี้มีประเด็นที่น่าสังเกต คือ ผลการวิเคราะห์ค่าสหสัมพันธ์อันดับแบบสเปียร์แมน จะให้ผลต่างจากผลการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน เมื่อวิเคราะห์ข้อมูล 2 ฐาน คือ ข้อมูลวัดซ้ำด้วยแบบทดสอบฉบับเดิมที่มีการตรวจให้คะแนนความรู้บางส่วน และในข้อมูลชุดวัดซ้ำ 5 ครั้งที่มีการตรวจให้คะแนนแบบทวิวิภาค ซึ่งพบว่า คะแนนพัฒนาการจากกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมมี

ค่าสหสัมพันธ์อันดับกับคะแนนพัฒนาการที่เป็นเกณฑ์สูงกว่าคะแนนพัฒนาการที่ได้จากกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

4.2.2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธี

ผลการเปรียบเทียบค่าความคลาดเคลื่อนของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธีทั้ง 3 กลุ่มวิธี พบว่า ภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิมคะแนนพัฒนาการจากวิธีการวัด DS และ SR ในข้อมูลทุกชุดมีค่าความคลาดเคลื่อนต่ำกว่าคะแนนพัฒนาการจากวิธีอื่นๆ และคะแนนพัฒนาการจากวิธีการวัด SRR มีค่าความคลาดเคลื่อนสูงที่สุด ยกเว้นในข้อมูลชุดวัดซ้ำ 3 ครั้ง และ 5 ครั้งด้วยแบบทดสอบคู่ขนาน ที่พบว่าคะแนนพัฒนาการจากวิธีการวัด SRR มีค่าความคลาดเคลื่อนต่ำที่สุด ซึ่งสอดคล้องกับสมมติฐานในข้อ 2.2 บางส่วน ที่ว่าคะแนนพัฒนาการจากวิธีการวัด DS และ SR มีค่าความคลาดเคลื่อนต่ำกว่าคะแนนพัฒนาการจากวิธีอื่นๆ ส่วนผลการเปรียบเทียบค่าความคลาดเคลื่อนของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่าคะแนนพัฒนาการจากวิธีการวัด MRMLC มีค่าความคลาดเคลื่อนต่ำกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ ซึ่งไม่สอดคล้องกับสมมติฐานในข้อ 2.3 ที่ว่าคะแนนพัฒนาการจากวิธีการวัด Lθ มีค่าความคลาดเคลื่อนต่ำกว่าคะแนนพัฒนาการจากวิธีการวัด MRMLC สำหรับผลการเปรียบเทียบค่าความคลาดเคลื่อนของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่าคะแนนพัฒนาการจากวิธีการวัด Vθ และ Sθ มีค่าความความคลาดเคลื่อนค่อนข้างใกล้เคียงกัน ซึ่งสอดคล้องกับสมมติฐานในข้อ 2.4 ที่ว่าคะแนนพัฒนาการจากวิธีการวัด Vθ และ Sθ มีค่าความความคลาดเคลื่อนไม่แตกต่างกัน

โดยภาพรวม ผลการเปรียบเทียบคุณภาพของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างกลุ่มวิธีการวัดทั้ง 3 กลุ่มวิธี พบว่า กลุ่มวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิม ทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีคุณภาพของวิธีการวัดคะแนนพัฒนาการแตกต่างกันอย่างไม่มีนัยสำคัญ ผลการเปรียบเทียบคุณภาพของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธี พบว่า ภายในกลุ่มวิธีการวัดตามทฤษฎีการทดสอบแบบดั้งเดิมคะแนนพัฒนาการจากวิธีการวัด DS และ SR มีคุณภาพสูงกว่าคะแนนพัฒนาการจากวิธีอื่นๆสำหรับข้อมูล 2 ชุดแรก และคะแนนพัฒนาการจากวิธีการวัด SRR มีคุณภาพสูงกว่าคะแนนพัฒนาการจากวิธีอื่นๆสำหรับข้อมูลชุดที่ 3 ผลการเปรียบเทียบคุณภาพของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธีการวัดตามทฤษฎีการตอบสนองข้อสอบ สำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาค

พบว่าคะแนนพัฒนาการจากวิธีการวัด MRMLC มีคุณภาพสูงกว่าคะแนนพัฒนาการจากวิธีการวัด Lθ และผลการเปรียบเทียบคุณภาพของคะแนนพัฒนาการทางคณิตศาสตร์ระหว่างวิธีการวัดภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนาพบว่าคะแนนพัฒนาการจากวิธีการวัด Vθ มีคุณภาพสูงกว่าคะแนนพัฒนาการจากวิธีการวัด Sθ เมื่อมีการตรวจให้คะแนนแบบทวิวิภาค (ดังสรุปผลในตารางที่ 5.1)

ตารางที่ 5.1 สรุปผลการเปรียบเทียบคุณภาพของคะแนนพัฒนาการระหว่างกลุ่มวิธีการวัดและภายในกลุ่มวิธีการวัด จำแนกตามชุดของข้อมูล และวิธีการตรวจให้คะแนน

ข้อมูลชุด	วิธีการให้คะแนน	ข้อมูลฐานที่	การเปรียบเทียบคุณภาพ			
			ระหว่างกลุ่มวิธี	ภายในกลุ่มวิธี		
				ทฤษฎีแบบดั้งเดิม	ทฤษฎีตอบสนองข้อสอบ	กลุ่มวิธีที่พัฒนา
วัดซ้ำด้วยแบบทดสอบฉบับเดิม	ทวิวิภาค	1	ไม่แตกต่าง	DS,SR	MRMLC	Vθ
	ความรู้บางส่วน	2	ไม่แตกต่าง	DS,SR	-	ไม่แตกต่าง
วัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน	ทวิวิภาค	3	ไม่แตกต่าง	DS,SR	MRMLC	Vθ
	ความรู้บางส่วน	4	ไม่แตกต่าง	DS,SR	-	ไม่แตกต่าง
วัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนาน	ทวิวิภาค	5	ไม่แตกต่าง	SRR	MRMLC	Vθ
	ความรู้บางส่วน	6	ทฤษฎีแบบดั้งเดิม	SRR	-	ไม่แตกต่าง
วัดซ้ำ 5 ครั้งด้วยแบบทดสอบคู่ขนาน	ทวิวิภาค	7	ไม่แตกต่าง	SRR	MRMLC	Vθ

การอภิปรายผลการวิจัย

การอภิปรายผลการวิจัยที่นำเสนอในตอนนี้แยกเป็น 3 ประเด็น สองประเด็นแรกเป็นการอภิปรายตามวัตถุประสงค์การวิจัยได้แก่ 1) ผลการพัฒนาวิธีการวัดคะแนนพัฒนาการ 2) ผลการวิเคราะห์เปรียบเทียบคุณภาพของคะแนนพัฒนาการ สำหรับประเด็นที่ 3 เป็นการอภิปรายเกี่ยวกับวิธีวิทยาที่ใช้ในการวิเคราะห์ ดังมีรายละเอียดต่อไปนี้

1. ผลการพัฒนาวิธีการวัดคะแนนพัฒนาการ

จากการวิเคราะห์คุณภาพของคะแนนพัฒนาการที่ได้จากวิธีการวัดคะแนนพัฒนาการ 3 กลุ่มวิธี คือกลุ่มวิธีของทฤษฎีการทดสอบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา รวมทั้งหมด 9 วิธี พบว่าคะแนนพัฒนาการจากวิธีการวัด DS และ SR ที่ได้จากวิธีการวัดในกลุ่มวิธีของทฤษฎีการทดสอบดั้งเดิมมีคุณภาพสูงกว่าคะแนนพัฒนาการที่ได้จากวิธีการวัดวิธีอื่นๆ สำหรับข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิมและข้อมูลชุดวัดซ้ำแบบจตุรัสละตินด้วยแบบทดสอบคู่ขนาน และพบว่าคะแนนพัฒนาการจาก

วิธีการวัด SRR มีคุณภาพสูงกว่าคะแนนพัฒนาการที่ได้จากวิธีการวัดวิธีอื่นๆ สำหรับข้อมูลชุดวัดซ้ำ 3 ครั้ง และ 5 ครั้ง ด้วยแบบทดสอบคู่ขนาน ในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่าคะแนนพัฒนาการจากวิธีการวัด MRMLC และ Vθ มีคุณภาพสูงกว่าคะแนนพัฒนาการที่ได้จากวิธีการวัดวิธีอื่นๆ ในข้อมูลทุกชุดที่มีการตรวจให้คะแนนแบบทววิภาค ผลการวิเคราะห์ขั้นต้นแสดงว่า คะแนนพัฒนาการจากวิธีการวัด DS, SR, SRR และ MRMLC มีคุณภาพเพียงพอที่จะสามารถนำไปใช้ประโยชน์ได้ แต่เมื่อพิจารณาถึงความสะดวกและความเหมาะสมในการปฏิบัติจริง พบว่าคะแนนพัฒนาการที่มีคุณภาพทั้งจากวิธีการวัด 4 วิธีนี้ มีวิธีการวัดคะแนนพัฒนาการเพียง 2 วิธีคือวิธีการวัด DS และ SR เป็นวิธีการวัดคะแนนพัฒนาการที่เหมาะสมสำหรับครูที่สุด เนื่องจากวิธีการวัดทั้ง 2 วิธีนี้เป็นวิธีที่ง่าย เก็บรวบรวมข้อมูลเพียง 2 ครั้ง ครูสามารถประมาณค่าคะแนนพัฒนาการจากคะแนนดิบโดยตรง ถึงแม้ว่าผลการวัดจะมีคุณภาพต่ำกว่าวิธีการวัด SRR ในข้อมูลชุดวัดซ้ำด้วยแบบทดสอบคู่ขนาน แต่ก็มีคุณภาพอยู่ในเกณฑ์ที่ไม่ต่ำมากทั้งในด้านความตรงตามเกณฑ์สัมพันธ์ (0.7061 - 0.7528) และค่าความคลาดเคลื่อน (0.7045 - 0.7615) จุดด้อยของวิธีการวัด SRR คือเป็นวิธีการประมาณค่าคะแนนพัฒนาการที่ค่อนข้างยุ่งยากเมื่อเปรียบเทียบกับวิธีการวัด DS และ SR และความยุ่งยากในการคำนวณไม่คุ้มค่ากับคุณภาพที่สูงขึ้นเพียงเล็กน้อย

2. ผลการวิเคราะห์เปรียบเทียบคุณภาพของคะแนนพัฒนาการ

2.1 ผลการวิเคราะห์เปรียบเทียบคุณภาพของคะแนนพัฒนาการระหว่างกลุ่มวิธี

2.1.1 ผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ระหว่างกลุ่มวิธี

ตามสมมติฐานวิจัยข้อ 1.2 ผู้วิจัยคาดว่า วิธีการวัดคะแนนพัฒนาการของกลุ่มวิธีทฤษฎีการตอบสนองข้อสอบ ควรมีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าวิธีการวัดคะแนนพัฒนาการของกลุ่มวิธีตามทฤษฎีการทดสอบแบบดั้งเดิม แต่ผลการวิจัยครั้งนี้ขัดแย้งกับสมมติฐานวิจัย กล่าวคือ ค่าความตรงตามเกณฑ์สัมพันธ์ของวิธีการวัดคะแนนพัฒนาการของทุกกลุ่มวิธีไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อวิเคราะห์กับข้อมูลทุกชุด ยกเว้น ข้อมูลชุดวัดซ้ำ 3 ครั้งด้วยแบบทดสอบคู่ขนานที่มีการตรวจให้คะแนนความรู้บางส่วนที่พบว่า กลุ่มวิธีตามทฤษฎีการทดสอบแบบดั้งเดิมมีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าวิธีการวัดคะแนนพัฒนาการของกลุ่มวิธีทฤษฎีการตอบสนองข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

ผลการวิจัยดังกล่าวนี้ข้างต้นอาจอธิบายได้ด้วยเหตุผล 2 ประการ ประการแรก กลุ่มวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมมีการประมาณค่าคะแนนพัฒนาการจากคะแนนดิบก่อนเรียนและหลังเรียนโดยตรง เช่นเดียวกับการประมาณค่าคะแนนพัฒนาการตามวิธีที่ใช้เป็นเกณฑ์ในการวิจัยครั้งนี้ (Meredith and Tisak, 1990; McArdle

and Hamagami, 1995) ซึ่งมีผลทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ที่เป็นค่าความตรงตามเกณฑ์สัมพันธ์มีค่าสูง ในขณะที่คะแนนพัฒนาการที่ได้จากกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบเป็นการประมาณค่าจากฟังก์ชั่นของคะแนนความสามารถที่แท้จริงของผู้สอบซึ่งขึ้นอยู่กับค่าพารามิเตอร์ของข้อสอบ (Embretson, 1991; May and Nicewander; 1998) ซึ่งอาจมีผลทำให้คะแนนพัฒนาการที่ได้มีความสัมพันธ์กับคะแนนพัฒนาการที่เป็นเกณฑ์ในระดับต่ำได้ ประการที่สองคะแนนพัฒนาการที่ใช้เป็นเกณฑ์ในการวิจัยครั้งนี้ได้มาจากการประมาณค่าด้วยโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงซึ่งมีการวัด 3 ครั้ง และ 5 ครั้ง โดย Stoolmiler (1995) อ้างว่าเป็นวิธีที่ดีที่สุดในการวัดคะแนนพัฒนาการ แต่คะแนนพัฒนาการที่ได้ตามวิธีนี้อาจจะไม่สอดคล้องและอาจจะเป็นคะแนนพัฒนาการคนละแบบกับคะแนนพัฒนาการที่เป็นความสามารถที่แท้จริงตามทฤษฎีการตอบสนองข้อสอบก็ได้

2.1.2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างกลุ่มวิธี

ตามสมมติฐานวิจัยข้อ 2.1 ผู้วิจัยคาดว่า วิธีการวัดคะแนนพัฒนาการของกลุ่มวิธีทฤษฎีการตอบสนองข้อสอบ ควรมีค่าความคลาดเคลื่อนต่ำกว่าวิธีการวัดคะแนนพัฒนาการของกลุ่มวิธีตามทฤษฎีการทดสอบแบบดั้งเดิม แต่ผลการวิจัยครั้งนี้ได้ผลขัดแย้งกับสมมติฐานวิจัย กล่าวคือ ค่าความคลาดเคลื่อนของวิธีการวัดคะแนนพัฒนาการของกลุ่มวิธีตามทฤษฎีการทดสอบแบบดั้งเดิมต่ำกว่าวิธีการวัดคะแนนพัฒนาการของกลุ่มวิธีทฤษฎีการตอบสนองข้อสอบ ผลการวิจัยที่ได้ดังกล่าวนี้อาจอธิบายได้ด้วยเหตุผล 2 ประการ ตามข้อ 2.1.1

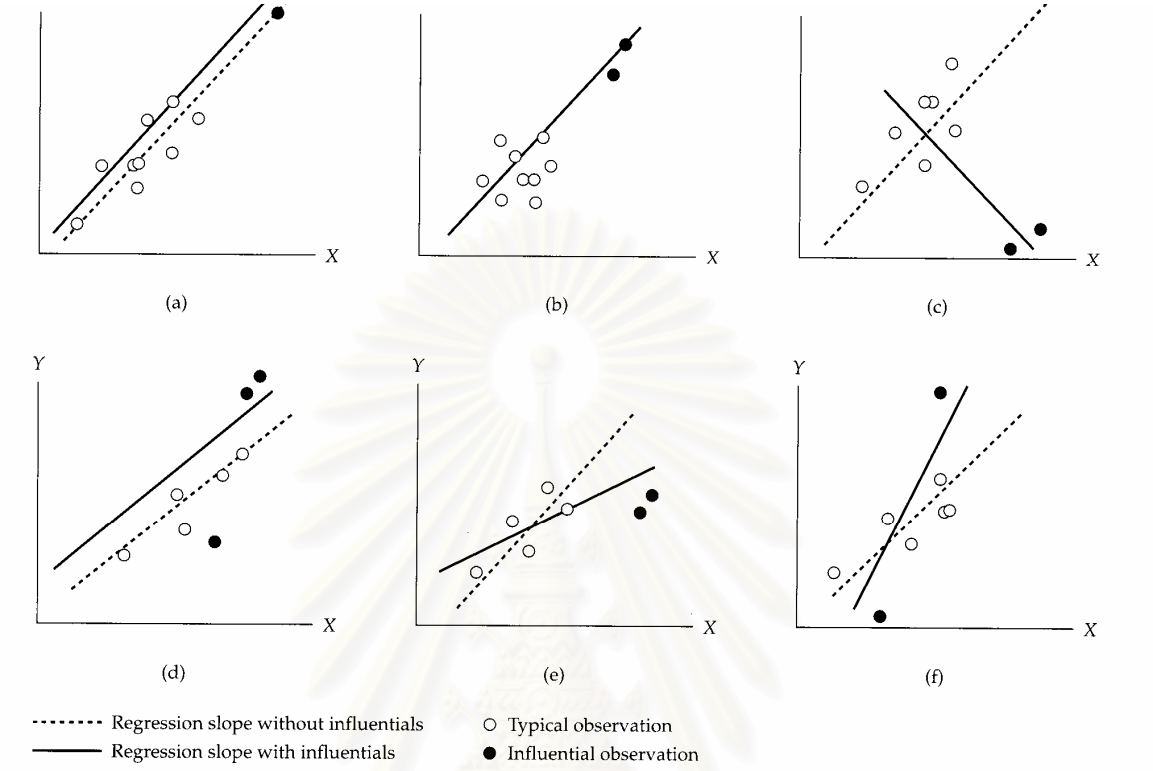
2.2 ผลการวิเคราะห์เปรียบเทียบคุณภาพของคะแนนพัฒนาการระหว่างวิธีการวัดภายในกลุ่มวิธี

2.2.1 ผลการเปรียบเทียบค่าความตรงตามเกณฑ์สัมพันธ์ระหว่างวิธีการวัดภายในกลุ่มวิธี

1) ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม

ผลการวิจัยสอดคล้องกับสมมติฐานวิจัยข้อ 1.3 เพียงบางส่วน ตามสมมติฐานวิจัยข้อ 1.3 ผู้วิจัยคาดว่าวิธีการวัด DS และ SR ควรมีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าวิธีการวัดคะแนนพัฒนาการวิธีอื่น แต่ผลการวิจัยพบว่าวิธีการวัด DS และ SR มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าวิธีการวัดคะแนนพัฒนาการวิธีอื่นเฉพาะข้อมูล 2 ชุดแรก และวิธีการวัด SRR มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าวิธีการวัดคะแนนพัฒนาการวิธีอื่นในข้อมูลชุดที่ 3 ทั้งนี้อาจเป็นเพราะว่าข้อมูล 2 ชุดแรกมีค่าสุดโต่งมากกว่าข้อมูลชุดที่ 3 และค่าสุดโต่งส่งผลต่อการประมาณค่าคะแนนพัฒนาการด้วยสมการทำนายดังแผนภาพ 5.1 ตามภาพจะเห็นได้ว่าจากภาพ (b), (c), (e) และ (f) มีค่าสุดโต่งซึ่งมีอิทธิพลทำให้กราฟสมการการทำนายเบี่ยงเบนไปจากกราฟสมการทำนายเมื่อไม่มีอิทธิพลจากค่าสุดโต่ง (Hair, Anderson, and Black, 1998; 185) แต่

อย่างไรก็ตามยังไม่อาจสรุปได้อย่างชัดเจนแน่ชัดว่าสาเหตุที่เป็นเช่นนี้เพราะลักษณะของข้อมูลที่ต่างกัน ควรจะต้องมีการวิจัยตรวจสอบโดยใช้วิธีการวิจัยที่มีการจำลอง (simulation) ข้อมูลกรณีที่มีค่าสุดโต่งลักษณะต่างๆกันหลายแบบ



แผนภาพที่ 5.1 ความสัมพันธ์ระหว่างตำแหน่งของค่าสุดโต่งกับสมการที่ใช้ในการทำนาย

2) ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ

ตามสมมติฐานวิจัยข้อ 1.4 ผู้วิจัยคาดว่าวิธีการวัด Lθ ควรมีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าวิธีการวัด MRMLC แต่ผลการวิจัยครั้งนี้ได้ผลขัดแย้งกับสมมติฐานวิจัย กล่าวคือวิธีการวัด Lθ มีค่าความตรงตามเกณฑ์สัมพันธ์ต่ำกว่าวิธีการวัด MRMLC ผลการวิจัยที่ได้ดังกล่าวนี้อาจอธิบายได้ว่าเนื่องมาจากวิธีการประมาณค่าพารามิเตอร์ของทั้ง 2 วิธีแตกต่างกัน วิธีการวัด Lθ ประมาณค่าพารามิเตอร์ด้วยวิธี EAP (expected a posteriori estimator) ส่วนวิธีการวัดMRMLC ประมาณค่าพารามิเตอร์ด้วยวิธี ML (maximum likelihood estimator) การประมาณค่าพารามิเตอร์ด้วยวิธี ML มีจุดอ่อนในด้านที่ไม่สามารถประมาณค่าพารามิเตอร์ของผู้ที่ได้คะแนนเป็นศูนย์และผู้ที่ได้คะแนนเต็ม (Mislevy and Bock, 1990: 1-13) ดังนั้นคะแนนพัฒนาการที่ประมาณค่าด้วยวิธีการวัด MRMLC จึงต้องมีการตัดคะแนนที่เป็นศูนย์และคะแนนเต็มออกไป จึงทำให้กลุ่มตัวอย่างที่เหลืออยู่ไม่มีอิทธิพลพื้นและอิทธิพลเพดาน คะแนนพัฒนาการจึงมีค่าใกล้เคียงกับคะแนนพัฒนาที่เป็นเกณฑ์ ทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ของวิธีการวัด

MRMLC มีค่าสูงกว่าค่าสัมประสิทธิ์สหสัมพันธ์ของวิธีการวัด Lθ ซึ่งมีอิทธิพลพื้นและอิทธิพลเพดานรวมอยู่ด้วย

3) ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา

ตามสมมติฐานวิจัยข้อ 1.5 ผู้วิจัยคาดว่าวิธีการวัด Vθ และ Sθ ควรมีค่าความตรงตามเกณฑ์สัมพันธ์ไม่แตกต่างกัน แต่ผลการวิจัยครั้งนี้ได้ผลสอดคล้องกับสมมติฐานวิจัยเพียงบางส่วน กล่าวคือวิธีการวัด Vθ มีค่าความตรงตามเกณฑ์สัมพันธ์สูงกว่าวิธีการวัด Sθ ในกรณีที่ข้อมูลมีการตรวจให้คะแนนแบบทวิวิภาค และวิธีการวัด Vθ กับ Sθ มีค่าความตรงตามเกณฑ์สัมพันธ์แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ เมื่อข้อมูลมีการตรวจให้คะแนนความรู้บางส่วน ผลการวิจัยที่ได้ดังกล่าวนี้อาจอธิบายได้ว่าเกิดจากวิธีการตรวจให้คะแนนต่างกันมีผลทำให้ลักษณะการแจกแจงของข้อมูลแตกต่างกัน การวิจัยครั้งนี้ พบว่าลักษณะการแจกแจงของข้อมูลทุกชุดจากการสอบทุกครั้งมีลักษณะค่อนข้างที่จะเบ้ขวา และลักษณะความเบ้ขวาลดลงในการวัดครั้งหลังแต่ปริมาณลักษณะการลดลงของความเบ้ในข้อมูลชุดที่มีการตรวจให้คะแนนแบบทวิภาคมีน้อยกว่าในข้อมูลชุดที่มีการตรวจให้คะแนนความรู้บางส่วน อย่างไรก็ตามข้อมูลในการวิจัยครั้งนี้ยังไม่อาจสรุปได้อย่างแน่ชัดว่าวิธีการวัด Vθ มีค่าความตรงตามเกณฑ์สัมพันธ์ไม่ต่างจากวิธีการวัด Sθ เนื่องจากวิธีการตรวจให้คะแนน ควรที่จะต้องมีการวิจัยตรวจสอบโดยใช้การวิจัยแบบจำลอง (simulation) ข้อมูลที่มีปริมาณการลดลงของความเบ้ในการวัดครั้งหลังๆด้วย

2.2.2 ผลการเปรียบเทียบค่าความคลาดเคลื่อนระหว่างวิธีการวัดภายในกลุ่มวิธี

1) ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม

ตามสมมติฐานวิจัยข้อ 2.2 ที่ผู้วิจัยคาดว่าวิธีการวัด DS และ SR ควรมีค่าความความคลาดเคลื่อนต่ำกว่าวิธีการวัดคะแนนพัฒนาการวิธีอื่น แต่ผลการวิจัยสอดคล้องกับสมมติฐานเพียงบางส่วนเท่านั้น กล่าวคือผลการวิจัยพบว่าวิธีการวัด DS และ SR มีค่าความความคลาดเคลื่อนต่ำกว่าวิธีการวัดคะแนนพัฒนาการวิธีอื่นเฉพาะข้อมูล 2 ชุดแรก ส่วนข้อมูลชุดที่ 2 ที่มีการตรวจให้คะแนนความรู้บางส่วนนั้นวิธีการวัด SR มีค่าความความคลาดเคลื่อนค่อนข้างสูงกว่าวิธีอื่น และสำหรับข้อมูลชุดที่ 3 วิธีการวัด SRR มีค่าความคลาดเคลื่อนต่ำกว่าวิธีการวัดคะแนนพัฒนาการวิธีอื่น ผลการวิจัยนี้อธิบายได้ด้วยเหตุผลเช่นเดียวกับข้อ 2.1.2 หัวข้อ1

2) ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ

ตามสมมติฐานวิจัยข้อ 2.3 ที่ผู้วิจัยคาดว่าวิธีการวัด Lθ ควรมีค่าความความคลาดเคลื่อนต่ำกว่าวิธีการวัด MRMLC แต่ผลการวิจัยครั้งนี้ได้ผลขัดแย้งกับสมมติฐาน กล่าวคือผลการวิจัยพบว่าวิธีการวัด Lθ ควรมีค่าความความคลาดเคลื่อนสูงกว่าวิธีการวัด MRMLC ผลการวิจัยนี้อธิบายได้ด้วยเหตุผลเช่นเดียวกับข้อ 2.1.2 หัวข้อ 2

3) ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา

ตามสมมติฐานวิจัยข้อ 2.4 ที่ผู้วิจัยคาดว่าวิธีการวัด Vθ และ Sθ ควรมีค่าความความคลาดเคลื่อนไม่แตกต่างกัน ผลการวิจัยครั้งนี้ค่อนข้างสอดคล้องกับสมมติฐาน กล่าวคือผลการวิจัยพบว่าวิธีการวัด Vθ ค่อนข้างที่จะมีค่าความคลาดเคลื่อนใกล้เคียงกับวิธีการวัด Sθ แต่มีแนวโน้มว่าสำหรับข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาคนั้นวิธีการวัด Vθ จะมีค่าความคลาดเคลื่อนต่ำกว่าวิธีการวัด Sθ ผลการวิจัยนี้อธิบายได้ด้วยเหตุผลเช่นเดียวกับข้อ 2.1.2 หัวข้อ 3

3. วิธีวิทยาที่ใช้ในการวิเคราะห์

3.1 ความสอดคล้องของวิธีการตรวจสอบความตรงตามเกณฑ์สัมพันธ์

ในการวิจัยครั้งนี้ได้ใช้วิธีการตรวจสอบความตรงตามเกณฑ์สัมพันธ์ด้วยค่าสถิติ 2 แบบ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ของ Pearson และค่าสหสัมพันธ์อันดับของ Spearman ค่าความตรงตามเกณฑ์สัมพันธ์ของค่าสถิติ 2 แบบควรจะให้ผลสอดคล้องกัน ในการวิจัยครั้งนี้พบว่าข้อมูลชุดวัดซ้ำด้วยแบบทดสอบฉบับเดิม ที่มีการตรวจให้คะแนนความรู้บางส่วนวิธีการวัด SRR มีค่าสัมประสิทธิ์สหสัมพันธ์ของ Pearson เท่ากับ 0.2329 และมีค่าสหสัมพันธ์อันดับของ Spearman เท่ากับ 0.8954 ที่เป็นเช่นนี้สามารถอธิบายได้ว่าเพราะคะแนนพัฒนาการที่ได้จากวิธีการวัด SRR ของผู้สอบแต่ละคนต่างกันมาก ในขณะที่คะแนนพัฒนาการจากวิธีเกณฑ์ของผู้สอบแต่ละคนต่างกันไม่มาก จึงทำให้คะแนนพัฒนาการไม่ค่อยสอดคล้องกับคะแนนพัฒนาการที่เป็นเกณฑ์ ทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ของ Pearson มีค่าต่ำ แต่เมื่อแปลงคะแนนพัฒนาการเป็นคะแนนอันดับ(rank) คะแนนพัฒนาการที่ไม่ค่อยต่างกัน หรือต่างกันมาก จะมีอันดับที่ไม่ต่างกัน ทำให้อันดับของคะแนนพัฒนาการมีความสอดคล้องกันสูงกับอันดับของคะแนนพัฒนาการจากวิธีเกณฑ์ ดังนั้นจึงทำให้สหสัมพันธ์อันดับมีค่าสูง

3.2 การตรวจสอบความแปรปรวนของค่าเฉลี่ยความตรงตามเกณฑ์สัมพันธ์ระหว่างกลุ่มวิธี

ในการวิจัยครั้งนี้ ได้ประยุกต์วิธีการวิเคราะห์พหุระดับมาใช้ในการตรวจสอบความแปรปรวนของค่าความตรงตามเกณฑ์สัมพันธ์ระหว่างกลุ่มวิธี ผลการวิจัยพบว่าให้ผลการวิเคราะห์ต่างกัน ประเด็นที่น่าสังเกตในการใช้การวิเคราะห์พหุระดับที่สามารถทดสอบสมมติฐานว่า ค่าเฉลี่ยรวมของค่าความตรงตามเกณฑ์สัมพันธ์มีค่าเท่ากับ 0 ด้วยสถิติทดสอบที และทดสอบสมมติฐานว่า ความแปรปรวนของค่าเฉลี่ยความตรงตามเกณฑ์สัมพันธ์มีค่าเท่ากับ 0 ด้วยสถิติทดสอบไค-สแควร์ ผลการวิจัยพบว่าการทดสอบค่าความตรงตามเกณฑ์สัมพันธ์ระหว่างกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กับกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ พบว่าค่าเฉลี่ยรวมของสัมประสิทธิ์สหสัมพันธ์ของทุกวิธีการวัดมีค่าเท่ากับ 0.612 มีค่าไม่แตกต่างจากศูนย์อย่างมี

นัยสำคัญทางสถิติที่ระดับ 0.05 ในขณะที่การทดสอบความแปรปรวนของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างกลุ่มวิธีการวัดของทฤษฎีการตอบสนองข้อสอบ กับกลุ่มวิธีการวัดของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา พบว่าค่าเฉลี่ยรวมของสัมประสิทธิ์สหสัมพันธ์ของทุกวิธีการวัดมีค่าเท่ากับ 0.423 มีค่าแตกต่างจากศูนย์อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 ผลการวิเคราะห์แสดงให้เห็นว่า ค่าเฉลี่ยในการทดสอบชุดแรกมีค่าสูงถึง 0.612 แต่ยังไม่มีความแตกต่างจากศูนย์อย่างมีนัยสำคัญ แต่ค่าเฉลี่ยในการทดสอบชุดหลังมีค่าเพียง 0.423 กลับมีความแตกต่างจากศูนย์อย่างมีนัยสำคัญ ผลการวิเคราะห์ดังกล่าวสามารถอธิบายได้ว่าความถูกต้องในการประมาณค่าเฉลี่ยรวมของสัมประสิทธิ์สหสัมพันธ์ขึ้นอยู่กับขนาดของกลุ่มตัวอย่าง ถ้ามีจำนวนของกลุ่มวิธีในการวัดคะแนนพัฒนาการมากกลุ่มวิธีความถูกต้องในการประมาณค่าจะถูกต้องมากกว่าเมื่อมีจำนวนกลุ่มวิธีน้อยกลุ่มวิธี (Bryk and Raudenbush, 1992: 69) กล่าวคือการประมาณค่าเฉลี่ยรวมของสัมประสิทธิ์สหสัมพันธ์จะถูกต้องยิ่งขึ้นถ้ามีจำนวนกลุ่มวิธีในการวัดคะแนนพัฒนาการเป็นจำนวนมาก ดังนั้นจึงอาจเป็นได้ว่า ค่าเฉลี่ยรวมของค่าสัมประสิทธิ์สหสัมพันธ์ของกลุ่มวิธีการวัดของทฤษฎีการตอบสนองข้อสอบและกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา มีค่าเท่ากับ 0.423 นั้นอาจมีค่าไม่แตกต่างศูนย์ เพราะจำนวนวิธีการวัดในที่นี้มีเพียง 3 วิธีที่นำมาวิเคราะห์เท่านั้น

ข้อเสนอแนะ

1. **ข้อเสนอแนะในเชิงนโยบาย**

ผลการวิจัยครั้งนี้พบว่า ผลสัมฤทธิ์ทางคณิตศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 ในโรงเรียนโครงการขยายโอกาสการศึกษาขั้นพื้นฐาน และระดับชั้นมัธยมศึกษาปีที่ 3 โรงเรียนในสังกัดกรมสามัญศึกษา เขตกรุงเทพมหานคร มีพัฒนาการค่อนข้างต่ำ ส่วนใหญ่จะมีพัฒนาการทางคณิตศาสตร์ต่ำกว่าค่าเฉลี่ย ดังนั้นหน่วยงานที่เกี่ยวข้องน่าจะได้หาทางแก้ไข เพราะวิชาคณิตศาสตร์เป็นวิชาที่สำคัญในการที่จะพัฒนาเยาวชนของประเทศไทยให้มีความสามารถและมีศักยภาพที่จะแข่งขันกับประเทศอื่นได้

2. **ข้อเสนอแนะในการนำไปปฏิบัติ**

ตามพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 การวัดและประเมินการศึกษาต้องมีการวัดพัฒนาการของผู้เรียนในทุกด้าน ดังนั้นการวัดและประเมินการศึกษาในทุกระดับการศึกษาจึงจำเป็นต้องคำนึงถึงวิธีการวัดคะแนนพัฒนาการที่สามารถวัดพัฒนาการได้ถูกต้องสอดคล้องกับสภาพที่แท้จริง และสามารถที่จะใช้ได้เหมาะสมกับบริบทของสถานศึกษา เนื้อหาวิชา ความรู้ความสามารถของครู อาจารย์ เนื่องจากผลการวิจัยครั้งนี้สรุปได้ว่าวิธีการวัด DS, SR MRMLC และ SRR เป็นวิธีที่ประมาณค่าคะแนนพัฒนาการได้มีคุณภาพใกล้เคียงกันและ

ในจำนวน 4 วิธี วิธีการวัด DS และ SR เป็นวิธีการวัดที่สะดวกและเหมาะสมในการปฏิบัติสำหรับครูดังนั้นผู้วิจัยขอเสนอแนะวิธีการวัดคะแนนพัฒนาการวิธีต่างๆดังนี้

2.1 ในระดับประถมศึกษา โรงเรียนควรที่จะใช้วิธีการวัดคะแนนพัฒนาการของนักเรียนในทุกๆด้าน ทั้งด้านผลสัมฤทธิ์ทางการเรียน หรือด้านการวัดพฤติกรรมการเรียนรู้ โดยจัดให้มีการวัด 2 ครั้ง ก่อนและหลังเรียน ด้วยแบบทดสอบหรือแบบประเมินพฤติกรรมฉบับเดิม เนื่องจากในสถาพของโรงเรียนระดับประถมศึกษาที่ครูมีภาระงานหนักมากต้องทำหน้าที่สอนทุกรายวิชาและต้องอยู่ดูแลนักเรียนตลอดเวลา และนักเรียนยังเป็นเด็กเล็ก และระยะเวลาการสอบทั้ง 2 ครั้งที่ห่างกันจึงทำให้ไม่มีผลในเรื่องของการจำข้อสอบได้ วิธีการวัดคะแนนพัฒนาการที่เหมาะสมที่สุดคือวิธีการวัด DS และ SR เพราะมีความตรงตามเกณฑ์สัมพันธ์สูง และมีค่าความคลาดเคลื่อนต่ำ และทั้ง 2 วิธีนี้เป็นวิธีการวัดคะแนนพัฒนาการที่ง่ายไม่ต้องใช้สถิติขั้นสูง

2.2 ในระดับมัธยมศึกษา โรงเรียนควรที่จะใช้วิธีการวัดพัฒนาการของนักเรียนในทุกๆด้าน ทั้งด้านผลสัมฤทธิ์ทางการเรียน หรือด้านการวัดพฤติกรรมการเรียนรู้ โดยจัดให้มีการวัด 2 ครั้ง ก่อนและหลังเรียน ด้วยการจัดสอบแบบจตุรัสละตินด้วยแบบทดสอบหรือแบบประเมินพฤติกรรมคู่ขนาน เช่นถ้าวัดนักเรียนเพียง 1 ห้อง อาจแบ่งนักเรียนในห้องออกเป็น 2 กลุ่ม ครั้งแรกนักเรียนกลุ่มที่ 1, 2 สอบข้อสอบชุดที่ 1, 2 ตามลำดับ ในการสอบครั้งที่ 2 นักเรียนกลุ่มที่ 1, 2 สอบข้อสอบชุดที่ 2, 1 ตามลำดับ เพื่อแก้ปัญหาในด้านการจำข้อสอบได้เพราะในระดับมัธมศึกษานักเรียนโตแล้วถ้าใช้แบบทดสอบฉบับเดิมจะทำให้เกิดความคลาดเคลื่อนในการวัดเนื่องมาจากการจำข้อสอบได้ และเกิดความเบื่อหน่ายในการทำข้อสอบซ้ำเดิมหรือเกิดจากการคุ้นเคยกับข้อสอบเดิม (Embretson,1991) ดังนั้นจึงควรใช้วิธีการวัดคะแนนพัฒนาการด้วยวิธีการวัด DS และ SR สำหรับข้อสอบและแบบวัดด้านทักษะต่างๆ และเนื่องจากในปัจจุบันครูในโรงเรียนมัธยมศึกษาส่วนใหญ่มีศักยภาพในด้านการใช้คอมพิวเตอร์ จึงควรใช้วิธีการวัด MRMLC สำหรับข้อสอบที่วัดคุณลักษณะต่างๆของนักเรียนที่มีการให้คะแนนแบบทวิวิภาค เพราะวิธีการวัด MRMLC สามารถประมาณค่าด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูปที่ไม่ยุ่งยาก และมีจุดเด่นของวิธีนี้คือเป็นการวัดคุณลักษณะที่แท้จริงที่เปลี่ยนแปลงไป โดยการนำค่าพารามิเตอร์ของข้อสอบและความสามารถของคนสอบมาประมาณค่าคุณลักษณะที่แท้จริงที่เปลี่ยนแปลงไป นอกจากนั้นทั้งวิธีการวัด DS, SR และ MRMLC ต่างก็ให้คะแนนพัฒนาการที่มีค่าความตรงตามเกณฑ์สัมพันธ์สูงและค่าความคลาดเคลื่อนต่ำ

2.3 ในระดับที่สูงกว่ามัธยมศึกษา ควรมีการวัดคะแนนพัฒนาการของผู้เรียน ทั้งด้านผลสัมฤทธิ์ทางการเรียน หรือด้านการวัดพฤติกรรมการเรียนรู้ โดยจัดให้มีการวัด 2 ครั้ง ก่อนและหลังเรียน ด้วยการจัดสอบแบบจตุรัสละตินด้วยแบบทดสอบหรือแบบประเมินพฤติกรรมคู่ขนานเพื่อป้องกันปัญหาในด้านการจำข้อสอบได้ และความเบื่อหน่ายในการทำข้อสอบซ้ำ

(Embretson,1991) อีกทั้งการศึกษาในระดับที่สูงกว่ามัธยมศึกษานั้นมุ่งเน้นที่จะวัดคุณลักษณะที่เปลี่ยนแปลงไปมากกว่าที่จะวัดทักษะที่สูงขึ้น รวมทั้งตัวอาจารย์เป็นผู้ที่มีศักยภาพในด้านการใช้คอมพิวเตอร์ ดังนั้นจึงควรใช้วิธีการวัดคะแนนพัฒนาการด้วยวิธีการวัด MRMLC เพราะวิธีการวัด MRMLC สามารถประมาณค่าด้วยโปรแกรมคอมพิวเตอร์สำเร็จรูปที่ไม่ยุ่งยาก และมีจุดเด่นของวิธีนี้คือเป็นการวัดคุณลักษณะที่แท้จริงที่เปลี่ยนแปลงไป โดยการนำค่าพารามิเตอร์ของข้อสอบและความสามารถของคนสอบมาประมาณค่าคุณลักษณะที่แท้จริงที่เปลี่ยนแปลงไป ประเด็นที่สำคัญคือวิธีการวัด MRMLC ให้คะแนนพัฒนาการที่มีค่าความตรงตามเกณฑ์สัมพันธ์สูง และค่าความคลาดเคลื่อนต่ำ

2.4 ในระดับหน่วยงานนโยบาย เช่น กรมวิชาการ ซึ่งได้มีการจัดการตรวจสอบคุณภาพของนักเรียนในระดับต่างๆทั่วประเทศเพื่อรักษามาตรฐานการศึกษาเป็นประจำทุกปีอยู่แล้ว เช่น ในระดับประถมศึกษาปีที่ 3 และ 6 ระดับมัธยมศึกษาปีที่ 3 และ 6 จึงควรจะมีการออกแบบเครื่องมือในการวัดให้เป็นแบบวัดคุณลักษณะเดียวกันแต่มีระดับของความซับซ้อนยิ่งขึ้นทุกปี โดยมีข้อสอบร่วมประมาณร้อยละ 20 เพื่อที่จะได้นำผลการสอบนั้นมาใช้ในการวัดคะแนนพัฒนาการของผู้เรียนด้วย ทำให้สามารถทราบพัฒนาการของนักเรียนในระดับกลุ่มโรงเรียน ระดับจังหวัด ระดับภาค และระดับประเทศ โดยการนำผลการวัดเหล่านั้นมาหาคะแนนพัฒนาการโดยใช้โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงหรือวิธีการวัด MRMLC จะทำให้หน่วยงานระดับนโยบายมีข้อมูลด้านพัฒนาการของนักเรียนเพิ่มขึ้นที่จะช่วยให้ผู้บริหารระดับนโยบายสามารถที่จะตัดสินใจกำหนดทิศทางในการพัฒนาการเรียนรู้ของนักเรียนในแต่ละระดับให้มีคุณภาพและมาตรฐานยิ่งขึ้น

3. ข้อเสนอแนะในการทำวิจัย

จากการวิจัยครั้งนี้ ผู้วิจัยขอเสนอแนะประเด็นที่ควรทำวิจัยต่อไปดังนี้

3.1 ในการวิจัยครั้งนี้ผู้วิจัยใช้โมเดลพัฒนาการเชิงเส้นตรงที่มีตัวแปรแฝงเป็นวิธีเกณฑ์โดยตรวจสอบความสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์รวม 7 ฐาน ด้วยโมเดล 5 รูปแบบ ซึ่งพบว่าผลการวิเคราะห์ข้อมูลชุดวัดซ้ำ 5 ครั้ง ด้วยแบบทดสอบคู่ขนานที่มีการตรวจให้คะแนนความรู้บางส่วนไม่มีโมเดลใดสอดคล้องกลมกลืนกับข้อมูลเชิงประจักษ์ ในเชิงวิชาการยังมีโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงแบบอื่นอีก ดังนั้นในการวิจัยครั้งต่อไปน่าจะเพิ่มโมเดลในการวิเคราะห์ เช่น โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงที่กำหนดพารามิเตอร์อิสระและมีความแปรปรวนของค่าความคลาดเคลื่อนไม่เท่ากันและมีความสัมพันธ์กัน และโมเดลพหุระดับมาใช้เป็นวิธีเกณฑ์ด้วย

3.2 ข้อมูลในการวัดคะแนนพัฒนาการโดยทั่วไปมักจะพบว่ามีค่าสุดโต่งอยู่ด้วย และในการวิจัยครั้งนี้พบว่าคะแนนสุดโต่งอาจมีอิทธิพลต่อการประมาณค่าคะแนนพัฒนาการ น่าจะมี

การวิจัยเปรียบเทียบคะแนนพัฒนาการในประเด็นที่ข้อมูลมีค่าสุดโต่งทางบวก ข้อมูลที่มีค่าสุดโต่งทางลบ และไม่มีค่าสุดโต่ง และขนาดของค่าสุดโต่งที่จะมีผลกระทบต่อคะแนนพัฒนาที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝง โมเดลพหุระดับ เป็นต้น

3.3 การวัดคะแนนพัฒนาการครั้งนี้เป็นการวัดคะแนนพัฒนาการด้านผลสัมฤทธิ์ทางการเรียนทางคณิตศาสตร์ที่มีการวัดซ้ำ 3 ครั้ง ส่วนการวัดซ้ำ 5 ครั้งมีเพียงข้อมูลชุดเดียว จึงน่าจะมีการวิจัยเปรียบเทียบคะแนนพัฒนาการของวิธีการวัดในกลุ่มวิธีของทฤษฎีการทดสอบแบบดั้งเดิม กลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบ และกลุ่มวิธีของทฤษฎีการตอบสนองข้อสอบที่ผู้วิจัยพัฒนา สำหรับการวัดซ้ำ 3 ครั้ง 4 ครั้ง และ 5 ครั้ง กรณีที่มีข้อมูลไม่สมบูรณ์ และวัดซ้ำในระยะเวลาที่ต่างกัน

3.4 การวัดคะแนนพัฒนาการผลสัมฤทธิ์ทางคณิตศาสตร์ครั้งนี้ผู้วิจัยไม่ได้แยกวิเคราะห์โมเดลจำแนกตามเพศ ดังนั้นในการวิจัยครั้งต่อไปน่าจะได้มีการวิเคราะห์จำแนกโมเดลตามเพศ ตามสหวิทยาเขตและตรวจสอบความไม่แปรเปลี่ยนของโมเดลโดยใช้เทคนิคพหุกลุ่ม เพื่อตรวจสอบความไม่แปรเปลี่ยนของโมเดล

3.5 ผลการวิจัยครั้งนี้ยังมีประเด็นที่ไม่สามารถจะสรุปได้หลายประเด็น เช่น รูปแบบของการวัดซ้ำมีผลต่อวิธีการวัดคะแนนพัฒนาการหรือไม่ ลักษณะของการแจกแจงของข้อมูลมีผลต่อวิธีการวัดคะแนนพัฒนาการหรือไม่ เป็นต้น ซึ่งควรจะได้มีการวิจัยต่อ โดยใช้วิธีการวิจัยแบบจำลอง (simulate) ข้อมูล

3.6 ในการวิจัยครั้งนี้ได้นำวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของคะแนนดิบมาเทียบกับวิธีการวัดวิธีอื่นๆโดยใช้โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงเป็นวิธีเกณฑ์ ซึ่งน่าจะมีการวิจัยต่อ โดยการนำวิธีการวัดคะแนนพัฒนาการจากความแตกต่างของคะแนนดิบและโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงมาเปรียบเทียบคุณภาพของคะแนนพัฒนาการโดยใช้วิธีการวิจัยแบบจำลอง (simulate) ข้อมูล

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายการอ้างอิง

ภาษาไทย

นงลักษณ์ วิรัชชัย. (2542). **โมเดลลิสเรล: สถิติวิเคราะห์สำหรับการวิจัย**. กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

ประสิทธิ์ ไชยกาล. (2539). **การเปรียบเทียบประสิทธิภาพระหว่างโมเดลลิสเรล 3 แบบ ที่ใช้ในการศึกษาตัวแปรที่สัมพันธ์กับการเปลี่ยนแปลงในระยะยาวของผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

พรทิพย์ ไชยโส. (2534). **การพัฒนาสูตรการให้คะแนนแบบสอบเลือกตอบสำหรับความรู้บางส่วนของผู้ตอบ: การประยุกต์ใช้วิธีการของอาร์โนลด์และวิธีการของแฮมดาน**. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

มนต์ทิวา ไชยแก้ว. (2542). **การเปรียบเทียบการเปลี่ยนแปลงระยะยาวโดยใช้โมเดลประยุกต์โค้งพัฒนาการที่มีตัวแปรแฝง เมื่ออัตราการขาดหายของข้อมูล ช่วงเวลาการวัด และจำนวนครั้งที่วัดแตกต่างกัน**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

วินิจ เทือกทอง. (2537). **การศึกษาเปรียบเทียบประสิทธิภาพของการคำนวณคะแนนเพิ่มวิธีต่าง ๆ ด้วยระเบียบวิธีการมอนติคาร์โล**. วิทยานิพนธ์ปริญญามหาบัณฑิต มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.

ศิริชัย กาญจนวาสี. (2540). **ทฤษฎีการวัดและประเมิน**. เอกสารประกอบการสอนวิชา 2702639. เอกสารเย็บเล่ม.

ศิริชัย กาญจนวาสี. (2543). **การประเมินการเรียนรู้: ข้อเสนอแนะเชิงนโยบาย**. กรุงเทพมหานคร: บริษัทอมรินทร์พริ้นติ้งแอนด์พับลิชชิ่งจำกัด.

ศิริชัย กาญจนวาสี. (2544). **ทฤษฎีการทดสอบแบบดั้งเดิม**(CLASSICAL TEST THEORY). กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี, ทวีวัฒน์ ปิตยานนท์ และดิเรก ศรีสุโข. (2537). **การเลือกใช้สถิติที่เหมาะสมสำหรับการวิจัย**. กรุงเทพมหานคร: โรงพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

สมถวิล วิจิตรวรรณา. (2543). **การเปรียบเทียบผลที่ได้จากการวัดพัฒนาการตัวแปรเอกนามและพหุนามจากการใช้ 3 โมเดล คือ โมเดลพัฒนาการที่มีตัวแปรแฝง โมเดลพหุระดับ และโมเดลกึ่งซิมเพลกซ์**. ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

สามัญศึกษา, กรม. (2543). **พื้นที่บริการรับนักเรียนของสหวิทยาเขตและโรงเรียนสังกัดกรมสามัญศึกษาในกรุงเทพมหานคร ปีการศึกษา 2543**. กรุงเทพมหานคร: โฆสิตการพิมพ์.

อิทธิพงษ์ ตั้งสกุลเรืองไล. (2541). **การเปรียบเทียบประสิทธิภาพของโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง 4 รูปแบบในการศึกษาการเปลี่ยนแปลงระยะยาวของการพัฒนาทางกายและผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนประถมศึกษา**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

อรุณี อ่อนสวัสดิ์. (2537). **การพัฒนาวิธีการวัดการเปลี่ยนแปลงการเรียนรู้**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

เอื้อมพร หลินเจริญ. (2539). **การพัฒนาโมเดลลิสเรลในการศึกษาการเปลี่ยนแปลงคุณลักษณะทางวิทยาศาสตร์ของนักเรียน**. วิทยานิพนธ์ปริญญามหาบัณฑิต ภาควิชาวิจัยการศึกษา บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.

ภาษาอังกฤษ

Andersen, E. B. (1977). Sufficient statistics and latent trait models. **Psychometrika** 42: 69-81.

Bock, R. D. (1997). A brief history of item response theory. **Educational Measurement : Issues and Practice** 16: 21-33.

Boonruangrutana, S. (1978). What happens when correlated errors are introduced?. **Educational and Psychological Test Bureau Srinakharinwirot University** 21:1-15.

Bryk, A. S. and Raudenbush, S. W. (1992). **Hierarchical Linear Models.** California: Sage Publications.

Bryk, A. S., Raudenbush, S. W. and Congdon, R. T. (1994). **HLM: Hierarchical Linear Modeling with the HLM/2L and HLM/3L Programs.** Chicago: Scientific Software.

Burr, J. A. and Nesselroade, J. R. (1990). Change measurement. In A. V. Eye (Ed.), **Statistical Methods in Longitudinal Research Volume I and II: Principles and Structuring Change,** pp. 1-34. Boston: Academic Press.

Chen, S. K.; Hou, L. and Dodd, B. G. (1998). A comparison of maximum likelihood estimation and expected a posteriori estimation in CAT using the partial credit model. **Educational and Psychological Measurement** 58 : 569-595.

Daniel, W. W. (1995). **Biostatistics: A Foundation for Analysis in the Health Sciences.** (6^{th} ed.), NY: John Wiley & Sons.

Embretson, S. E. (1991a). Implications of a multidimensional latent trait model for measuring change. In L. M. Collins and J. L. Horn (Eds.), **Best Methods for the Analysis of Change**, pp. 184-197. Washington DC: American Psychological Association.

Embretson, S. E. (1991b). A multidimensional latent trait model for measuring learning and change. **Psychometrika** 56 : 495-515.

Embretson, S. E. (1995). A measurement model for linking individual learning to processes and knowledge: Application to mathematical reasoning. **Journal of Educational Measurement** 32 : 277-294.

Embretson, S. E. (1997). Structured ability models in test designed from cognitive theory. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective Measurement IV: Theory into Practice**, pp. 223-236. Greenwich, CT: A blex.

Embretson, S. E. (1999). Issues in the measurement of cognitive abilities. In S. E. Embretson and S. L. Hershberger (Eds.), **The new rules of measurement** , pp. 1 -16. Mahwah, NJ: Lawrence Erlbaum Associates.

Fischer, G. H. (1989). An IRT-based model for dichotomous longitudinal data. **Psychometrika** 54: 599-624.

Fischer, G. H. (1995a). Some neglected problems in IRT. Psychometrika. 60: 459-487.

Fischer, G. H. (1995b). Linear Logistic Models for Change. In G. H. Fischer and I. W. Molenaar. (Eds.), **Rasch Models Foundations, Recent Developments, and Applications**, pp. 157 - 180. NY.: Springer - Verlag.

Fischer, G. H. (1997). Structural rasch models: some theory, applications, and software, In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective Measurement IV: Theory into Practice**, pp. 185-207. Greenwich, CT: A blex.

Fischer, G. H. and Parzer, P. (1991). An extension of the rating scale model with an application to the measurement of change. **Psychometrika** 56: 637-651.

Fischer, G. H. and Ponocny, I. (1994). An extension of the partial credit model with and application to the measurement of change. **Psychometrika** 59: 177-192.

Hair, J. F. Jr., Anderson, R. E., Tatham, R. L. and Black, W. C. (1998). **Multivariate data analysis**.(5 edition). NJ: Prentice-Hall.

Hambleton, R. K. and Swaminathan, H. (1985). **Item Response Theory: Principles and Applications**. Norwell, MA: Kluwer Academic Publishers.

Kirk, R. E. (1995). **Experimental Design: Procedures for the Behavioral Sciences**.(third edition). California: Books/Cole Publishing Company.

Lindeman, R. H., Merenda, P. F. and Gold, R. Z. (1980). **Introduction to Bivariate and Multivariate Analysis**. Glenview,IL: Scott, Foresman.

Lord, F. M. (1980). **Applications of Item Response Theory to Practical Testing Problems**. NJ: Lawrence Eribaum Associates.

May, K. and Nicewander, W. A. (1998). Measuring change conventionally and adaptively. **Educational and Psychological Measurement** 58: 882-897.

Masters, G. N. and Wright, B. D. (1984). The essential process in a family of measurement models. **Psychometrika** 49: 529-54.

McArdle, J. J. & Hamagami, F. (1991, 1995). Modeling incomplete longitudinal and cross-ssection data using latent growth structural models. In L. M Collins & J. L. Horn (Eds). **Best Methods for The Analysis of Change**, pp. 149 - 163. Washington DC: American Psychological Association.

Mislevy, R. J. and Bock, R. D. (1990). **BILOG 3 : Item Analysis and Test Scoring with Binary Logistic Models**. Chicago: Scientific Software.

Muraki, E. and Bock, R. D. (1993). **PARSCALE 2: IRT Based Test Scoring and Item Analysis for Graded Open-ended Exercises and Performance Tasks**. Chicago: Scientific Software.

Pike, G. R. (1991). Using structural equation models with latent variables to study student growth and development. **Research in Higher Education** 32: 499-523.

Ponocny, I. And Ponocny, E. S. (1997). Applications of the program LPCM in the field of measuring change. In M. Wilson, G. Engelhard & K. Draney (Eds.), **Objective Measurement IV: Theory into Practice**, pp. 209-222. Greenwich, CT: A blex.

Raykov, T. (1993). A structural equation model for measuring residualized change and discerning patterns of growth of decline. **Applied Psychological Measurement** 17: 53-71.

Raykov, T. (1994). Studying correlates and predictors of longitudinal change using structural equation modeling. **Applied Psychological Measurement** 18: 63-77.

SPSS Inc. (1998). **SPSS Base 8.0: Applications Guide.** IL: SPSS .

Stevens, J. (1996). **Applied multivariate statistics for the social sciences.** (third edition).New Jersey: Lawrence Erlbaum Associates.

Stoolmiller, M. (1994). Antisocial behavior, delinquent peer association, and unsupervised wandering for boys: growth and change from childhood to early adolescence. **Multivariate Behavioral Research** 29: 263-288.

Stoolmiller, M. (1995). Using latent growth curve models to study developmental processes. In John M. Gottman (Ed.), **The Analysis of Change**, pp. 103-138. Mahwah, NJ: Lawrence Erlbaum Associates.

Stoolmiller, M. and Muthen, B. (1997). Latent variable modeling of longitudinal and multilevel substance use data. **Multivariate Behavioral Research** 32: 275-318.

Tisak, J. and Tisak, M. S. (1996). Longitudinal models of reliability and validity: A latent curve appoarch. **Applied Psychological Measurement** 20: 275-288.

Wilcox, R. R. (1996). **Statistics for the Social Sciences.** San Diego: Academic Press.

Willett J. B. (1994). Change, Measurement of. In Torsten Husen and T. N. Postlethwaite (Eds.), **The international Encyclopedia of Education.** (2nd Ed.), 2. pp. 671-678. NY: Elsevier Science.

Willett, J. B. and Rogosa, D. R. (1985). Understanding correlates of change by modeling individual differences in growth. **Psychometrika** 50: 203-228.

Willett, J. B. and Sayer, A. G. (1994). Using covariance structure analysis to detect correlates and predictors of individual change over time. **Psychological Bulletin** 116: 363-381.

Wright, B. D. and Masters, G. N. (1982). **Rating scale analysis: Rasch measurement.** Chicago: MESA PRESS.

ภาคผนวก

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ก

รายชื่อผู้เชี่ยวชาญ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

รายชื่อผู้เชี่ยวชาญ

1. อาจารย์ทองสุข ทับเจริญ อาจารย์ 2 ระดับ 7 โรงเรียนศึกษานารี วุฒิการศึกษาปริญญาโท สาขาสถิติการศึกษา (คม.) ประสบการณ์การสอนคณิตศาสตร์ 24 ปี
2. อาจารย์อุทุมพร ชินตะวัน อาจารย์ 2 ระดับ 7 โรงเรียนมัธยมวัดสิงห์ วุฒิการศึกษาปริญญาโท สาขาการสอนคณิตศาสตร์ (วทม.) ประสบการณ์การสอนคณิตศาสตร์ 27 ปี
3. อาจารย์สุวรรณี ถาวรกูล อาจารย์ 2 ระดับ 7 โรงเรียนวัดนวลนรดิศ วุฒิการศึกษาปริญญาโท สาขาการสอนคณิตศาสตร์ (กศ.ม.) ประสบการณ์การสอนคณิตศาสตร์ 27 ปี
4. อาจารย์สิริมาศ สิทธิหล่อ อาจารย์ 2 ระดับ 7 โรงเรียนสตรีศรีสุริโยทัย วุฒิการศึกษาปริญญาโทสาขาการวัดและประเมินผลการศึกษา (คม.) ประสบการณ์การสอนคณิตศาสตร์ 22 ปี
5. อาจารย์สุวรรณ สถลชา อาจารย์ 2 ระดับ 7 โรงเรียนสตรีศรีสุริโยทัย วุฒิการศึกษาปริญญาโท สาขาการวัดและประเมินผลการศึกษา (คม.) ประสบการณ์การสอนคณิตศาสตร์ 24 ปี
6. อาจารย์ลัดดา ด่านวิริยะกุล อาจารย์ 2 ระดับ 7 หัวหน้างานวัดผล โรงเรียนราชวินิตบางแค ปานขำ วุฒิการศึกษาปริญญาโทสาขาสถิติการศึกษา (คม.) ประสบการณ์การสอนคณิตศาสตร์ 22 ปี
7. อาจารย์ประทิน พงศ์อิศวรานันท์ อาจารย์ 2 ระดับ 7 หัวหน้างานวัดผล โรงเรียนเตรียมอุดมศึกษาพัฒนาการ วุฒิการศึกษาปริญญาโทสาขาการวัดและประเมินผลการศึกษา (คม.) ประสบการณ์การสอนคณิตศาสตร์ 25 ปี
8. ผู้ช่วยศาสตราจารย์ชัยศักดิ์ ชั่งใจ โรงเรียนสาธิตจุฬาฯฝ่ายมัธยม วุฒิการศึกษาปริญญาโท สาขาการวัดและประเมินผลการศึกษา (คม.) ประสบการณ์การสอนคณิตศาสตร์ 22 ปี
9. ดร.รังสรรค์ มณีเล็ก เจ้าหน้าที่วิเคราะห์นโยบายและแผน 8 ว. สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ วุฒิการศึกษาปริญญาเอกสาขาการทดสอบและการวัดผลทางการศึกษา (กศ.ด.)วุฒิการศึกษาปริญญาโทสาขาการวัดและประเมินผลการศึกษา (คม.) ประสบการณ์เป็นศึกษานิเทศก์ทางด้านการวัดและประเมินผล 5 ปี นักวิชาการการศึกษา13 ปี
10. ดร.กฤษณีย์ อุทุมพร ศึกษานิเทศก์ 8 กรมอาชีวศึกษา วุฒิการศึกษาปริญญาเอกสาขาวิจัยและพัฒนาหลักสูตร (กศ.ด.) วุฒิการศึกษาปริญญาโทสาขาการวัดและประเมินผลการศึกษา (คม.) ประสบการณ์เป็นครูสอนภาษาอังกฤษ 5 ปี เป็นศึกษานิเทศก์ทางด้านการวัดและประเมินผล 17 ปี

ผู้เชี่ยวชาญหมายเลข 1-7 เป็นผู้ให้น้ำหนักความสำคัญของจุดประสงค์การเรียนรู้ของเนื้อหาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 และผู้เชี่ยวชาญหมายเลข 4-10 เป็นผู้ตรวจสอบความสอดคล้องของข้อสอบกับจุดประสงค์การเรียนรู้ และตรวจสอบข้อสอบที่วัดความรู้บางส่วน

ภาคผนวก ข

รายชื่อโรงเรียนกลุ่มตัวอย่าง

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1. โรงเรียนศึกษานารี
2. โรงเรียนมัธยมวัดสิงห์
3. โรงเรียนสุวรรณารามวิทยาคม
4. โรงเรียนพระโขนงพิทยาลัย
5. โรงเรียนเตรียมอุดมศึกษาพัฒนาการ
6. โรงเรียนพุทธจักรวิทยา
7. โรงเรียนวัดสุทธิวราราม
8. โรงเรียนยานนาเวศวิทยาคม
9. โรงเรียนนนทรีวิทยา
10. โรงเรียนวัดรางบัว
11. โรงเรียนราชวินิตบางแคปานขำ
12. โรงเรียนปัญญาวรคุณ
13. โรงเรียนเทพศิรินทร์
14. โรงเรียนวัดสระเกศ
15. โรงเรียนไตรมิตรวิทยาลัย
16. โรงเรียนประดู่ในทรงธรรม
17. โรงเรียนวัดนวลนรดิศ
18. โรงเรียนจันทร์ประดิษฐารามวิทยาคม

หมายเหตุ โรงเรียนหมายเลข 1-6 เป็นกลุ่มตัวอย่างที่ใช้ในการทดลองเครื่องมือ

โรงเรียนหมายเลข 7-12 เป็นกลุ่มตัวอย่างที่ใช้เก็บข้อมูลจริงในข้อมูลชุดวัดซ้ำชุดที่ 1

โรงเรียนหมายเลข 13-18 เป็นกลุ่มตัวอย่างที่ใช้เก็บข้อมูลจริงในข้อมูลชุดวัดซ้ำชุดที่ 2

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ค

ตัวอย่างคำสั่งและผลการวิเคราะห์ข้อมูลโมเดลพัฒนาการเชิงเส้นโค้ง ที่มีตัวแปรแฝงด้วยโปรแกรม LISREL 8.30 for student

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

DATE: 12/ 8/2001
TIME: 8:45

L I S R E L 8.30

BY

Karl G. Jöreskog & Dag Sörbom

This program is published exclusively by
Scientific Software International, Inc.
7383 N. Lincoln Avenue, Suite 100
Lincolnwood, IL 60712, U.S.A.
Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
Copyright by Scientific Software International, Inc., 1981-2000
Use of this program is subject to the terms specified in the
Universal Copyright Convention.
Website: www.ssicentral.com

The following lines were read from file F:\╠╚╦╤╤╬╣~1\REPEAT\LISREL\RE1D.SPL:

```
THE UNEQUAL DISTURBANCE VARIANCE MODEL
FOR REPEAT TEST 3 TIMES AND SCORED BY DICHOTOMOUS METHOD
DA NG=1 NI=4 NO=635 MA=MM
LA
SCOR1 SCOR2 SCOR3 CONST
KM
1.000
0.571 1.000
0.492 0.683 1.000
0.000 0.000 0.000 1.000
ME
6.9991 7.9795 10.1150 1.0000
SD
3.2743 4.1503 5.3518 0.0000
SE
4 1 2 3/
MO NY=4 NE=11 BE=FU,FI PS=SY,FI LY=FU,FI TE=ZE
LE
'1.CONST' '2.SCOR1' '3.SCOR2' '4.SCOR3' '5.ER-S1' '6.ER-S2'
'7.ER-S3' '8.INITIAL' '9.GROWTH' '10.ER-INI' '11.ER-GRO'
MA LY
1 0 0 0 0 0 0 0 0 0 0
0 1 0 0 0 0 0 0 0 0 0
0 0 1 0 0 0 0 0 0 0 0
0 0 0 1 0 0 0 0 0 0 0
MA BE
0 0 0 0 0 0 0 0 0 0 0
0 0 0 0 1 0 0 1 0.0000 0 0
0 0 0 0 0 1 0 1 0.4805 0 0
0 0 0 0 0 0 1 1 1.5195 0 0
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0
 8 0 0 0 0 0 0 0 0 1 0
 1 0 0 0 0 0 0 0 0 0 1
 0 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0 0
 MA PS
 1
 0 0
 0 0 0
 0 0 0 0
 0 0 0 0 1
 0 0 0 0 0 1
 0 0 0 0 0 0 1
 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 0
 0 0 0 0 0 0 0 0 0 1
 0 0 0 0 0 0 0 0 0 1 1
 FR BE 2 5 BE 3 6
 FR BE 3 9
 FR BE 8 10 BE 9 11
 FR BE 9 1 BE 8 1
 FR PS 10 11
 FR PS 1 1
 OU NS RS SE TV PC SS FS MI  AD=OFF TO ND=4 IT=500
```

THE UNEQUAL DISTURBANCE VARIANCE MODEL

Number of Input Variables	4
Number of Y - Variables	4
Number of X - Variables	0
Number of ETA - Variables	11
Number of KSI - Variables	0
Number of Observations	635

THE UNEQUAL DISTURBANCE VARIANCE MODEL

Moment Matrix to be Analyzed

	CONST	SCOR1	SCOR2	SCOR3
CONST	1.0000			
SCOR1	6.9991	59.7084		
SCOR2	7.9795	63.6088	80.8974	
SCOR3	10.1150	79.4174	95.8831	130.9550

THE UNEQUAL DISTURBANCE VARIANCE MODEL

Parameter Specifications

BETA

1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.CONST	0	0	0	0	0	0
2.SCOR1	0	0	0	0	1	0
3.SCOR2	0	0	0	0	0	2
4.SCOR3	0	0	0	0	0	0
5.ER-S1	0	0	0	0	0	0
6.ER-S2	0	0	0	0	0	0
7.ER-S3	0	0	0	0	0	0
8.INITIA	4	0	0	0	0	0
9.GROWTH	6	0	0	0	0	0
10.ER-IN	0	0	0	0	0	0
11.ER-GR	0	0	0	0	0	0

BETA

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
1.CONST	0	0	0	0	0
2.SCOR1	0	0	0	0	0
3.SCOR2	0	0	3	0	0
4.SCOR3	0	0	0	0	0
5.ER-S1	0	0	0	0	0
6.ER-S2	0	0	0	0	0
7.ER-S3	0	0	0	0	0
8.INITIA	0	0	0	5	0
9.GROWTH	0	0	0	0	7
10.ER-IN	0	0	0	0	0
11.ER-GR	0	0	0	0	0

PSI

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	8					
2.SCOR1	0	0				
3.SCOR2	0	0	0			
4.SCOR3	0	0	0	0		
5.ER-S1	0	0	0	0	0	
6.ER-S2	0	0	0	0	0	0
7.ER-S3	0	0	0	0	0	0
8.INITIA	0	0	0	0	0	0
9.GROWTH	0	0	0	0	0	0
10.ER-IN	0	0	0	0	0	0
11.ER-GR	0	0	0	0	0	0

PSI

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
7.ER-S3	0				
8.INITIA	0	0			
9.GROWTH	0	0	0		
10.ER-IN	0	0	0	0	
11.ER-GR	0	0	0	9	0

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

THE UNEQUAL DISTURBANCE VARIANCE MODEL

Number of Iterations = 3

LISREL Estimates (Maximum Likelihood)

LAMBDA-Y

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
CONST	1.0000	- -	- -	- -	- -	- -
SCOR1	- -	1.0000	- -	- -	- -	- -
SCOR2	- -	- -	1.0000	- -	- -	- -
SCOR3	- -	- -	- -	1.0000	- -	- -

LAMBDA-Y

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
CONST	- -	- -	- -	- -	- -
SCOR1	- -	- -	- -	- -	- -
SCOR2	- -	- -	- -	- -	- -
SCOR3	- -	- -	- -	- -	- -

BETA

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	- -	- -	- -	- -	- -	- -
2.SCOR1	- -	- -	- -	- -	1.8715 (0.1436) 13.0292	- -
3.SCOR2	- -	- -	- -	- -	- -	2.6240 (0.0866) 30.3092
4.SCOR3	- -	- -	- -	- -	- -	- -
5.ER-S1	- -	- -	- -	- -	- -	- -
6.ER-S2	- -	- -	- -	- -	- -	- -
7.ER-S3	- -	- -	- -	- -	- -	- -

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

8.INITIA	6.9849 (0.1281) 54.5307	- -	- -	- -	- -	- -
9.GROWTH	2.0587 (0.1220) 16.8795	- -	- -	- -	- -	- -
10.ER-IN	- -	- -	- -	- -	- -	- -
11.ER-GR	- -	- -	- -	- -	- -	- -

BETA

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
1.CONST	- -	- -	- -	- -	- -
2.SCOR1	- -	1.0000	- -	- -	- -
3.SCOR2	- -	1.0000	0.5035 (0.0374) 13.4760	- -	- -
4.SCOR3	1.0000	1.0000	1.5195	- -	- -
5.ER-S1	- -	- -	- -	- -	- -
6.ER-S2	- -	- -	- -	- -	- -
7.ER-S3	- -	- -	- -	- -	- -
8.INITIA	- -	- -	- -	2.6943 (0.1239) 21.7431	- -
9.GROWTH	- -	- -	- -	- -	2.7560 (0.1024) 26.9237
10.ER-IN	- -	- -	- -	- -	- -
11.ER-GR	- -	- -	- -	- -	- -

Covariance Matrix of ETA

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	1.0000					
2.SCOR1	6.9849	59.5502				
3.SCOR2	8.0214	63.7623	81.3626			
4.SCOR3	10.1130	79.3295	96.0976	130.9333		

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

5.ER-S1	- -	1.8715	- -	- -	1.0000	
6.ER-S2	- -	- -	2.6240	- -	- -	1.0000
7.ER-S3	- -	- -	- -	1.0000	- -	- -
8.INITIA	6.9849	56.0476	63.7623	79.3295	- -	- -
9.GROWTH	2.0587	15.3221	21.2802	33.3029	- -	- -
10.ER-IN	- -	2.6943	2.8705	3.2259	- -	- -
11.ER-GR	- -	0.3420	1.7297	4.5297	- -	- -

Covariance Matrix of ETA

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
7.ER-S3	1.0000				
8.INITIA	- -	56.0476			
9.GROWTH	- -	15.3221	11.8334		
10.ER-IN	- -	2.6943	0.3499	1.0000	
11.ER-GR	- -	0.3420	2.7560	0.1269	1.0000

PSI

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	1.0000 (0.0562) 17.8045					
2.SCOR1	- -	- -				
3.SCOR2	- -	- -	- -			
4.SCOR3	- -	- -	- -	- -		
5.ER-S1	- -	- -	- -	- -	1.0000	
6.ER-S2	- -	- -	- -	- -	- -	1.0000
7.ER-S3	- -	- -	- -	- -	- -	- -
8.INITIA	- -	- -	- -	- -	- -	- -
9.GROWTH	- -	- -	- -	- -	- -	- -
10.ER-IN	- -	- -	- -	- -	- -	- -
11.ER-GR	- -	- -	- -	- -	- -	- -

PSI

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
7.ER-S3	1.0000				
8.INITIA	- -	- -			

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

9.GROWTH	- -	- -	- -		
10.ER-IN	- -	- -	- -	1.0000	
11.ER-GR	- -	- -	- -	0.1269 (0.0679) 1.8684	1.0000

Squared Multiple Correlations for Structural Equations

1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
- -	1.0000	1.0000	1.0000	- -	- -

Squared Multiple Correlations for Structural Equations

7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
- -	1.0000	1.0000	- -	- -

Squared Multiple Correlations for Reduced Form

1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
- -	- -	- -	- -	- -	1.0000

Squared Multiple Correlations for Reduced Form

7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
-5.9849	1.0000	1.0000	1.0000	1.0000

Squared Multiple Correlations for Y - Variables

CONST	SCOR1	SCOR2	SCOR3
1.0000	1.0000	1.0000	1.0000

Goodness of Fit Statistics

Degrees of Freedom = 1
Minimum Fit Function Chi-Square = 0.3580 (P = 0.5496)
Normal Theory Weighted Least Squares Chi-Square = 0.3576 (P = 0.5498)
Estimated Non-centrality Parameter (NCP) = 0.0
90 Percent Confidence Interval for NCP = (0.0 ; 4.9278)

Minimum Fit Function Value = 0.0005646
Population Discrepancy Function Value (F0) = 0.0
90 Percent Confidence Interval for F0 = (0.0 ; 0.007773)
Root Mean Square Error of Approximation (RMSEA) = 0.0
90 Percent Confidence Interval for RMSEA = (0.0 ; 0.08816)
P-Value for Test of Close Fit (RMSEA < 0.05) = 0.7773

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Expected Cross-Validation Index (ECVI) = 0.02997
90 Percent Confidence Interval for ECVI = (0.02997 ; 0.03774)
ECVI for Saturated Model = 0.03155
ECVI for Independence Model = 5.8489

Chi-Square for Independence Model with 6 Degrees of Freedom = 3700.2149
Independence AIC = 3708.2149
Model AIC = 18.3576
Saturated AIC = 20.0000
Independence CAIC = 3730.0294
Model CAIC = 67.4402
Saturated CAIC = 74.5362

Normed Fit Index (NFI) = 0.9999
Non-Normed Fit Index (NNFI) = 1.0010
Parsimony Normed Fit Index (PNFI) = 0.1667
Comparative Fit Index (CFI) = 1.0000
Incremental Fit Index (IFI) = 1.0002
Relative Fit Index (RFI) = 0.9994

Critical N (CN) = 11752.7878

Root Mean Square Residual (RMR) = 0.1792
Standardized RMR = 0.002738
Goodness of Fit Index (GFI) = 0.9997
Adjusted Goodness of Fit Index (AGFI) = 0.9972
Parsimony Goodness of Fit Index (PGFI) = 0.09997

THE UNEQUAL DISTURBANCE VARIANCE MODEL

Fitted Covariance Matrix

	CONST	SCOR1	SCOR2	SCOR3
CONST	1.0000			
SCOR1	6.9849	59.5502		
SCOR2	8.0214	63.7623	81.3626	
SCOR3	10.1130	79.3295	96.0976	130.9333

Fitted Residuals

	CONST	SCOR1	SCOR2	SCOR3
CONST	- -			
SCOR1	0.0142	0.1582		
SCOR2	-0.0419	-0.1535	-0.4651	
SCOR3	0.0020	0.0880	-0.2145	0.0217

Summary Statistics for Fitted Residuals

Smallest Fitted Residual = -0.4651
Median Fitted Residual = 0.0010
Largest Fitted Residual = 0.1582

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Stemleaf Plot

```
- 4|7
- 2|1
- 0|5400
  0|1296
```

Standardized Residuals

	CONST	SCOR1	SCOR2	SCOR3
CONST	- -			
SCOR1	0.5980	0.5980		
SCOR2	-0.5980	-0.5980	-0.5980	
SCOR3	0.5980	0.5980	-0.5980	0.5980

Summary Statistics for Standardized Residuals

```
Smallest Standardized Residual =   -0.5980
  Median Standardized Residual =    0.2990
 Largest Standardized Residual =    0.5980
```

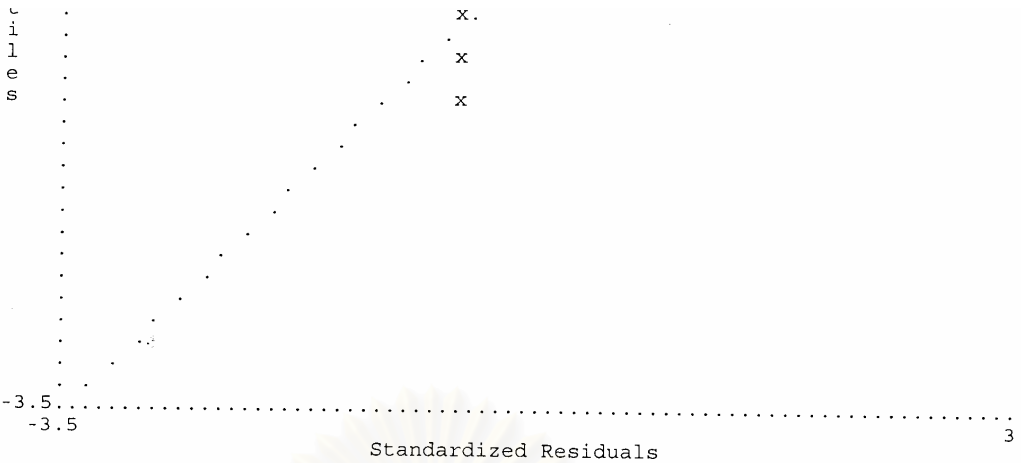
Stemleaf Plot

```
- 0|6666
- 0|0
  0|
  0|66666
```

THE UNEQUAL DISTURBANCE VARIANCE MODEL

Qplot of Standardized Residuals

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย



THE UNEQUAL DISTURBANCE VARIANCE MODEL

Modification Indices and Expected Change

Modification Indices for LAMBDA-Y

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
CONST	- -	0.3576	0.3576	0.3575	0.3576	0.3576
SCOR1	0.3576	0.3576	0.3576	0.3576	- -	- -
SCOR2	0.3576	0.3576	0.3576	0.3576	- -	- -
SCOR3	0.3576	0.3576	0.3576	0.3576	0.3576	0.3576

Modification Indices for LAMBDA-Y

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
CONST	0.3576	- -	- -	- -	- -
SCOR1	0.3576	0.3576	- -	0.3576	0.3576
SCOR2	0.3576	0.3576	- -	0.3576	0.3576
SCOR3	0.3576	0.3576	- -	0.3576	0.3576

Expected Change for LAMBDA-Y

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
CONST	- -	0.0148	-0.0050	0.1049	0.0278	-0.0132
SCOR1	0.1387	0.0250	0.0250	0.0257	- -	- -
SCOR2	-0.0927	-0.0167	-0.0167	-0.0172	- -	- -

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

SCOR3	0.2798	0.0638	0.0312	0.0536	-0.4492	0.2142

Expected Change for LAMBDA-Y

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
CONST	0.1049	- -	- -	- -	- -
SCOR1	-0.8408	0.0250	- -	-0.2606	-0.1856
SCOR2	0.5622	-0.0167	- -	0.1743	0.1241
SCOR3	-0.8483	0.0504	- -	-0.5259	-0.3745

Standardized Expected Change for LAMBDA-Y

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
CONST	- -	0.1145	-0.0455	1.1999	0.0278	-0.0132
SCOR1	0.1387	0.1927	0.2253	0.2945	- -	- -
SCOR2	-0.0927	-0.1289	-0.1506	-0.1969	- -	- -
SCOR3	0.2798	0.4923	0.2811	0.6131	-0.4492	0.2142

Standardized Expected Change for LAMBDA-Y

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
CONST	0.1049	- -	- -	- -	- -
SCOR1	-0.8408	0.1870	- -	-0.2606	-0.1856
SCOR2	0.5622	-0.1250	- -	0.1743	0.1241
SCOR3	-0.8483	0.3773	- -	-0.5259	-0.3745

Modification Indices for BETA

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	- -	0.3576	0.3576	0.3576	0.3576	0.3576
2.SCOR1	0.3576	0.3576	0.3576	0.3576	- -	- -
3.SCOR2	0.3576	0.3576	0.3576	0.3576	- -	- -
4.SCOR3	0.3576	0.3576	0.3576	0.3576	0.3576	0.3576
5.ER-S1	0.3576	0.3576	0.3576	0.3576	- -	- -
6.ER-S2	0.3576	0.3576	0.3576	0.3576	- -	- -
7.ER-S3	0.3576	0.3576	0.3576	0.3576	0.3576	0.3576
8.INITIA	- -	0.3576	0.3576	0.3574	0.3576	0.3576
9.GROWTH	- -	0.3576	0.3576	0.3577	0.3576	0.3576
10.ER-IN	- -	0.3576	0.3576	0.3574	0.3576	0.3576
11.ER-GR	- -	0.3576	0.3576	0.3577	0.3576	0.3576

Modification Indices for BETA

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
1.CONST	0.3576	- -	- -	- -	- -
2.SCOR1	0.3576	0.3576	- -	0.3576	0.3576
3.SCOR2	0.3576	0.3576	- -	0.3576	0.3576
4.SCOR3	0.3576	0.3576	- -	0.3576	0.3576
5.ER-S1	0.3576	0.3576	- -	0.3576	0.3576
6.ER-S2	0.3576	0.3576	- -	0.3576	0.3576

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

7.ER-S3	0.3576	0.3576	- -	0.3576	0.3576
8.INITIA	0.3576	- -	- -	- -	- -
9.GROWTH	0.3576	- -	- -	- -	- -
10.ER-IN	0.3576	- -	- -	- -	- -
11.ER-GR	0.3576	- -	- -	- -	- -

Expected Change for BETA

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	- -	0.0396	-0.0135	0.2798	0.0741	-0.0353
2.SCOR1	0.1387	0.0250	0.0250	0.0257	- -	- -
3.SCOR2	-0.0927	-0.0167	-0.0167	-0.0172	- -	- -
4.SCOR3	0.2798	0.0638	0.0312	0.0536	-0.4492	0.2142
5.ER-S1	0.0741	0.0133	0.0133	0.0138	- -	- -
6.ER-S2	-0.0353	-0.0064	-0.0064	-0.0066	- -	- -
7.ER-S3	0.2798	0.0638	0.0312	0.0536	-0.4492	0.2142
8.INITIA	- -	-0.2400	0.0816	-1.6960	-0.4492	0.2142
9.GROWTH	- -	-0.1580	0.0537	-1.1167	-0.2957	0.1410
10.ER-IN	- -	-0.0891	0.0303	-0.6295	-0.1667	0.0795
11.ER-GR	- -	-0.0573	0.0195	-0.4052	-0.1073	0.0512

Expected Change for BETA

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
1.CONST	0.2798	- -	- -	- -	- -
2.SCOR1	-0.8408	0.0250	- -	-0.2606	-0.1856
3.SCOR2	0.5622	-0.0167	- -	0.1743	0.1241
4.SCOR3	-0.8483	0.0504	- -	-0.5259	-0.3745
5.ER-S1	-0.4492	0.0133	- -	-0.1393	-0.0992
6.ER-S2	0.2142	-0.0064	- -	0.0664	0.0473
7.ER-S3	-0.8483	0.0504	- -	-0.5259	-0.3745
8.INITIA	-1.6966	- -	- -	- -	- -
9.GROWTH	-1.1165	- -	- -	- -	- -
10.ER-IN	-0.6297	- -	- -	- -	- -
11.ER-GR	-0.4051	- -	- -	- -	- -

Standardized Expected Change for BETA

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	- -	0.0051	-0.0015	0.0245	0.0741	-0.0353
2.SCOR1	0.0180	0.0004	0.0004	0.0003	- -	- -
3.SCOR2	-0.0103	-0.0002	-0.0002	-0.0002	- -	- -
4.SCOR3	0.0245	0.0007	0.0003	0.0004	-0.0393	0.0187
5.ER-S1	0.0741	0.0017	0.0015	0.0012	- -	- -
6.ER-S2	-0.0353	-0.0008	-0.0007	-0.0006	- -	- -
7.ER-S3	0.2798	0.0083	0.0035	0.0047	-0.4492	0.2142
8.INITIA	- -	-0.0042	0.0012	-0.0198	-0.0600	0.0286
9.GROWTH	- -	-0.0060	0.0017	-0.0284	-0.0859	0.0410
10.ER-IN	- -	-0.0115	0.0034	-0.0550	-0.1667	0.0795
11.ER-GR	- -	-0.0074	0.0022	-0.0354	-0.1073	0.0512

Standardized Expected Change for BETA

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
1.CONST	0.2798	- -	- -	- -	- -
2.SCOR1	-0.1090	0.0004	- -	-0.0338	-0.0241
3.SCOR2	0.0623	-0.0002	- -	0.0193	0.0138
4.SCOR3	-0.0741	0.0006	- -	-0.0460	-0.0327
5.ER-S1	-0.4492	0.0018	- -	-0.1393	-0.0992
6.ER S2	0.2142	-0.0009	- -	0.0664	0.0473
7.ER-S3	-0.8483	0.0067	- -	-0.5259	-0.3745
8.INITIA	-0.2266	- -	- -	- -	- -
9.GROWTH	-0.3246	- -	- -	- -	- -
10.ER-IN	-0.6297	- -	- -	- -	- -
11.ER-GR	-0.4051	- -	- -	- -	- -

Modification Indices for PSI

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	- -					
2.SCOR1	0.3576	- -				
3.SCOR2	0.3576	- -	- -			
4.SCOR3	0.3576	0.3576	0.3576	0.3576		
5.ER-S1	0.3576	- -	- -	0.3576	- -	
6.ER-S2	0.3576	- -	- -	0.3576	- -	- -
7.ER-S3	0.3576	0.3576	0.3576	0.3576	0.3576	0.3576
8.INITIA	- -	0.3576	0.3576	0.3576	0.3576	0.3576
9.GROWTH	- -	0.3576	0.3576	0.3576	0.3576	0.3576
10.ER-IN	- -	0.3576	0.3576	0.3576	0.3576	0.3576
11.ER-GR	- -	0.3576	0.3576	0.3576	0.3576	0.3576

Modification Indices for PSI

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
7.ER-S3	0.3576				
8.INITIA	0.3576	- -			
9.GROWTH	0.3576	- -	- -		
10.ER-IN	0.3576	- -	- -	- -	
11.ER-GR	0.3576	- -	- -	- -	- -

Expected Change for PSI

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	- -					
2.SCOR1	0.1387	- -				
3.SCOR2	-0.0927	- -	- -			
4.SCOR3	0.2798	-0.8408	0.5622	-1.6966		
5.ER-S1	0.0741	- -	- -	-0.4492	- -	
6.ER-S2	-0.0353	- -	- -	0.2142	- -	- -
7.ER-S3	0.2798	-0.8408	0.5622	-0.8483	-0.4492	0.2142
8.INITIA	- -	-0.8408	0.5622	-1.6966	-0.4492	0.2142
9.GROWTH	- -	-0.5533	0.3700	-1.1165	-0.2957	0.1410
10.ER-IN	- -	-0.3121	0.2087	-0.6297	-0.1667	0.0795

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

11.ER-GR	- -	-0.2008	0.1342	-0.4051	-0.1073	0.0512

Expected Change for PSI

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
7.ER-S3	-1.6966				
8.INITIA	-1.6966	- -			
9.GROWTH	-1.1165	- -	- -		
10.ER-IN	-0.6297	- -	- -	- -	
11.ER-GR	-0.4051	- -	- -	- -	- -

Standardized Expected Change for PSI

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	- -					
2.SCOR1	0.0180	- -				
3.SCOR2	-0.0103	- -	- -			
4.SCOR3	0.0245	-0.0095	0.0054	-0.0130		
5.ER-S1	0.0741	- -	- -	-0.0393	- -	
6.ER-S2	-0.0353	- -	- -	0.0187	- -	- -
7.ER-S3	0.2798	-0.1090	0.0623	-0.0741	-0.4492	0.2142
8.INITIA	- -	-0.0146	0.0083	-0.0198	-0.0600	0.0286
9.GROWTH	- -	-0.0208	0.0119	-0.0284	-0.0859	0.0410
10.ER-IN	- -	-0.0404	0.0231	-0.0550	-0.1667	0.0795
11.ER-GR	- -	-0.0260	0.0149	-0.0354	-0.1073	0.0512

Standardized Expected Change for PSI

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
7.ER-S3	-1.6966				
8.INITIA	-0.2266	- -			
9.GROWTH	-0.3246	- -	- -		
10.ER-IN	-0.6297	- -	- -	- -	
11.ER-GR	-0.4051	- -	- -	- -	- -

Modification Indices for THETA-EPS

	CONST	SCOR1	SCOR2	SCOR3
CONST	- -			
SCOR1	0.3576	- -		
SCOR2	0.3576	- -	- -	
SCOR3	0.3576	0.3576	0.3576	0.3576

Expected Change for THETA-EPS

	CONST	SCOR1	SCOR2	SCOR3
CONST	- -			
SCOR1	0.0520	- -		
SCOR2	-0.0348	- -	- -	
SCOR3	0.1049	-0.8408	0.5622	-1.6966

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

Maximum Modification Index is 0.36 for Element (11, 4) of BETA

Covariance Matrix of Parameter Estimates

	BE 2,5	BE 3,6	BE 3,9	BE 8,1	BE 8,10	BE 9,1
BE 2,5	0.0206					
BE 3,6	-0.0019	0.0075				
BE 3,9	-0.0016	0.0004	0.0014			
BE 8,1	0.0009	-0.0002	-0.0008	0.0164		
BE 8,10	-0.0093	0.0003	0.0006	-0.0003	0.0154	
BE 9,1	-0.0005	0.0001	0.0004	-0.0018	0.0002	0.0149
BE 9,11	-0.0050	0.0004	0.0009	-0.0005	0.0023	0.0003
PS 1,1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PS 11,10	0.0057	-0.0004	-0.0008	0.0004	-0.0039	-0.0003

Covariance Matrix of Parameter Estimates

	BE 9,11	PS 1,1	PS 11,10
BE 9,11	0.0105		
PS 1,1	0.0000	0.0032	
PS 11,10	-0.0027	0.0000	0.0046

THE UNEQUAL DISTURBANCE VARIANCE MODEL

Correlation Matrix of Parameter Estimates

	BE 2,5	BE 3,6	BE 3,9	BE 8,1	BE 8,10	BE 9,1
BE 2,5	1.0000					
BE 3,6	-0.1515	1.0000				
BE 3,9	-0.2994	0.1347	1.0000			
BE 8,1	0.0492	-0.0221	-0.1643	1.0000		
BE 8,10	-0.5212	0.0266	0.1252	-0.0206	1.0000	
BE 9,1	-0.0292	0.0131	0.0975	-0.1170	0.0122	1.0000
BE 9,11	-0.3369	0.0461	0.2407	-0.0395	0.1802	0.0235
PS 1,1	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
PS 11,10	0.5883	-0.0599	-0.3101	0.0509	-0.4620	-0.0302

Correlation Matrix of Parameter Estimates

	BE 9,11	PS 1,1	PS 11,10
BE 9,11	1.0000		
PS 1,1	0.0000	1.0000	
PS 11,10	-0.3839	0.0000	1.0000

THE UNEQUAL DISTURBANCE VARIANCE MODEL

Factor Scores Regressions

ETA

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

	CONST	SCOR1	SCOR2	SCOR3
1.CONST	1.0000	0.0000	0.0000	0.0000
2.SCOR1	0.0000	1.0000	0.0000	0.0000
3.SCOR2	0.0000	0.0000	1.0000	0.0000
4.SCOR3	0.0000	0.0000	0.0000	1.0000
5.ER-S1	-0.7914	0.2667	-0.0926	-0.0325
6.ER-S2	-0.3958	-0.1298	0.3305	-0.1333
7.ER-S3	-0.1458	-0.0174	-0.0508	0.0667
8.INITIA	1.4812	0.5009	0.1733	0.0608
9.GROWTH	-0.8789	-0.3182	-0.0806	0.5742
10.ER-IN	-2.0427	0.1859	0.0643	0.0226
11.ER-GR	-1.0659	-0.1155	-0.0292	0.2083

THE UNEQUAL DISTURBANCE VARIANCE MODEL

Standardized Solution

LAMBDA-Y

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
CONST	1.0000	- -	- -	- -	- -	- -
SCOR1	- -	7.7169	- -	- -	- -	- -
SCOR2	- -	- -	9.0201	- -	- -	- -
SCOR3	- -	- -	- -	11.4426	- -	- -

LAMBDA-Y

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
CONST	- -	- -	- -	- -	- -
SCOR1	- -	- -	- -	- -	- -
SCOR2	- -	- -	- -	- -	- -
SCOR3	- -	- -	- -	- -	- -

BETA

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	- -	- -	- -	- -	- -	- -
2.SCOR1	- -	- -	- -	- -	0.2425	- -
3.SCOR2	- -	- -	- -	- -	- -	0.2909
4.SCOR3	- -	- -	- -	- -	- -	- -
5.ER-S1	- -	- -	- -	- -	- -	- -
6.ER-S2	- -	- -	- -	- -	- -	- -
7.ER-S3	- -	- -	- -	- -	- -	- -
8.INITIA	0.9330	- -	- -	- -	- -	- -
9.GROWTH	0.5985	- -	- -	- -	- -	- -
10.ER-IN	- -	- -	- -	- -	- -	- -
11.ER-GR	- -	- -	- -	- -	- -	- -

BETA

7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.CONST	- -	- -	- -	- -	- -
2.SCOR1	- -	0.9701	- -	- -	- -
3.SCOR2	- -	0.8300	0.1920	- -	- -
4.SCOR3	0.0874	0.6543	0.4568	- -	- -
5.ER-S1	- -	- -	- -	- -	- -
6.ER-S2	- -	- -	- -	- -	- -
7.ER-S3	- -	- -	- -	- -	- -
8.INITIA	- -	- -	- -	0.3599	- -
9.GROWTH	- -	- -	- -	- -	0.8012
10.ER-IN	- -	- -	- -	- -	- -
11.ER-GR	- -	- -	- -	- -	- -

Correlation Matrix of ETA

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	1.0000					
2.SCOR1	0.9051	1.0000				
3.SCOR2	0.8893	0.9160	1.0000			
4.SCOR3	0.8838	0.8984	0.9311	1.0000		
5.ER-S1	- -	0.2425	- -	- -	1.0000	
6.ER-S2	- -	- -	0.2909	- -	- -	1.0000
7.ER-S3	- -	- -	- -	0.0874	- -	- -
8.INITIA	0.9330	0.9701	0.9442	0.9260	- -	- -
9.GROWTH	0.5985	0.5772	0.6858	0.8461	- -	- -
10.ER-IN	- -	0.3491	0.3182	0.2819	- -	- -
11.ER-GR	- -	0.0443	0.1918	0.3959	- -	- -

Correlation Matrix of ETA

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
7.ER-S3	1.0000				
8.INITIA	- -	1.0000			
9.GROWTH	- -	0.5950	1.0000		
10.ER-IN	- -	0.3599	0.1017	1.0000	
11.ER-GR	- -	0.0457	0.8012	0.1269	1.0000

PSI

	1.CONST	2.SCOR1	3.SCOR2	4.SCOR3	5.ER-S1	6.ER-S2
1.CONST	1.0000					
2.SCOR1	- -	- -				
3.SCOR2	- -	- -	- -			
4.SCOR3	- -	- -	- -	- -		
5.ER-S1	- -	- -	- -	- -	1.0000	
6.ER-S2	- -	- -	- -	- -	- -	1.0000
7.ER-S3	- -	- -	- -	- -	- -	- -
8.INITIA	- -	- -	- -	- -	- -	- -
9.GROWTH	- -	- -	- -	- -	- -	- -
10.ER-IN	- -	- -	- -	- -	- -	- -
11.ER-GR	- -	- -	- -	- -	- -	- -

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

PSI

	7.ER-S3	8.INITIA	9.GROWTH	10.ER-IN	11.ER-GR
7.ER-S3	1.0000				
8.INITIA	- -	- -			
9.GROWTH	- -	- -	- -		
10.ER-IN	- -	- -	- -	1.0000	
11.ER-GR	- -	- -	- -	0.1269	1.0000

The Problem used 21456 Bytes (= 0.0% of Available Workspace)

Time used: 0.109 Seconds

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ง

ตัวอย่างคำสั่งและผลการวิเคราะห์ข้อมูลเพื่อประมาณค่าความสามารถที่แท้จริงของข้อมูลที่มีการตรวจให้คะแนนแบบทวิวิภาคด้วยโปรแกรม BILOG 3

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
        THE PERSON ABILITY:FILE DIRE1.BLG :USING DICHOTOMOUS FIRST
       TIME RESPONSE / 22 ITEMS FOR REPEATED TEST WITH 4 CHOICES
>COMMENT  DATA FROM STUDENT 635 PERSAON
            6 SCHOOLS
>GLOBAL     DFName ='C:\BILOG3\DISCORE1.dat', NPARM=3, logistic, SAVE;
>SAVE       score='C:\BILOG3\DIRE1.sco', PARME='C:\BILOG3\DIRE1.PAR';
>LENGTH     NITEMS=22;
>INPut      NTOtal=22, NALT=2, NIDch=6,kFName ='C:\BILOG3\DISCORE1.dat',
            oFName ='C:\BILOG3\DISCORE1.dat';
(6A1,2X,22A1)
>TEST       TNAme=MATH;
>CALib      TPRIOR, SPRIOR, GPRIOR, FLOAT, cycles=100, NEWTON=2, crit=0.001;
>score      RSCtype=4;
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

DOS-BILOG V3.04

MAXIMUM LIKELIHOOD ITEM ANALYSIS AND TEST SCORING: LOGISTIC MODEL

DISTRIBUTED BY

SCIENTIFIC SOFTWARE, INC.
1369 NEITZEL ROAD
MOORESVILLE, INDIANA 46158

(800)247-6113 or (317)831-6336

PROGRAM COPYRIGHT HELD BY SCIENTIFIC SOFTWARE, INC. 1985-90
ALL RIGHTS RESERVED.

1

*** LOGISTIC MODEL ITEM ANALYSER ***

*** PHASE 1 ***

THE PERSON ABILITY:FILE DIRE1.BLG :USING DICHOTOMOUS FIRST
TIME RESPONSE / 22 ITEMS FOR REPEATED TEST WITH 4 CHOICES

>COMMENT DATA FROM STUDENT 635 PERSAON
6 SCHOOLS

>GLOBAL DFName ='C:\BILOG3\DISCORE1.dat', NPARM=3, logistic, SAVE;

GLOBAL PARAMETERS
==========

NUMBER OF SUBTESTS	1
CASE WEIGHTING	NONE EMPLOYED
ITEM RESPONSE MODEL	3 PARAMETER LOGISTIC
	LOGIT METRIC (I.E., D = 1.0)

>SAVE score='C:\BILOG3\DIRE1.sco', PARME='C:\BILOG3\DIRE1.PAR';

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
>LENGTH        NITEMS=22;

SUBTEST LENGTHS:
     22


>INPut         NTOtal=22, NALT=2, NIDch=6,kFName ='C:\BILOG3\DISCORE1.dat',

               oFName ='C:\BILOG3\DISCORE1.dat';


DATA INPUT SPECIFICATIONS
=========================

NUMBER OF FORMAT CARDS                          1
NUMBER OF ITEMS IN INPUT STREAM                22
NUMBER OF RESPONSE ALTERNATIVES                 2

NUMBER OF SUBJECT ID CHARACTERS                 6
SUBJECT DATA INPUT OPTION
                                    SINGLE-SUBJECT DATA, NO CASE WEIGH

MAXIMUM SAMPLE SIZE FOR ITEM CALIBRATION     1000
ALL SUBJECTS INCLUDED IN RUN

FORMAT CARD FOR INPUT IS

         (6A1,2X,22A1)


FILE ASSIGNMENT AND DISPOSITION
===============================

[INPUT FILES]

SUBJECT DATA INPUT FILE             C:\BILOG3\DISCORE1.DAT
                                    SINGLE-SUBJECT DATA, NO CASE WEIGH

CORRECT-RESPONSE KEY FILE           C:\BILOG3\DISCORE1.DAT

OMITTED RESPONSE KEY FILE           C:\BILOG3\DISCORE1.DAT


[OUTPUT FILES]

ITEM PARAMETERS FILE                C:\BILOG3\DIRE1.PAR

SUBJECT SCALE-SCORE FILE            C:\BILOG3\DIRE1.SCO


[SCRATCH FILES]
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
BILOG SYSTEM BINARY DATA FILE                DIRE1.MFL
CALIBRATION BINARY DATA FILE                 DIRE1.CFL
ESTIMATED COVARIANCE FILE                    DIRE1.VFL
TEMPORARY FILE                               DIRE1.T02
TEMPORARY FILE                               DIRE1.T03
TEMPORARY FILE                               DIRE1.T14
TEMPORARY FILE                               DIRE1.T99

>TEST          TNAme=MATH;

ANSWER KEY:
   1    MATH         11111111111111111111

OMIT KEY:
   1    MATH         00000000000000000000

OBSERVATION     1  WEIGHT:      1.0000  ID : 111101

 SUBTEST    1    MATH

     TRIED     RIGHT
    22.000     9.000

 ITEM        1      2      3      4      5      6      7      8      9     10
 TRIED     1.0    1.0    1.0    1.0    1.0    1.0    1.0    1.0    1.0    1.0
 RIGHT     1.0   -1.0   -1.0    1.0    1.0    1.0    1.0   -1.0    1.0    1.0

 ITEM       11     12     13     14     15     16     17     18     19     20
 TRIED     1.0    1.0    1.0    1.0    1.0    1.0    1.0    1.0    1.0    1.0
 RIGHT     1.0    1.0   -1.0   -1.0   -1.0   -1.0   -1.0   -1.0   -1.0   -1.0

 ITEM       21     22
 TRIED     1.0    1.0
 RIGHT    -1.0   -1.0

OBSERVATION     2  WEIGHT:      1.0000  ID : 111102

 SUBTEST    1    MATH

     TRIED     RIGHT
    22.000    10.000
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ITEM	1	2	3	4	5	6	7	8	9	10
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	1.0	-1.0	1.0	1.0	1.0	-1.0	1.0	1.0	1.0	1.0

ITEM	11	12	13	14	15	16	17	18	19	20
TRIED	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
RIGHT	1.0	-1.0	-1.0	-1.0	-1.0	-1.0	1.0	-1.0	-1.0	-1.0

ITEM	21	22
TRIED	1.0	1.0
RIGHT	-1.0	-1.0

635 OBSERVATIONS READ FROM FILE: C:\BILOG3\DISCORE1.DAT
635 OBSERVATIONS WRITTEN TO FILE: DIRE1.MFL

1

CLASSICAL ITEM STATISTICS FOR SUBTEST MATH

ITEM	NAME	NUMBER TRIED	NUMBER RIGHT	PERCENT	LOGIT	ITEM*TEST CORRELATION PEARSON	BISERIAL
1	0001	635.0	340.0	.535	.14	.266	.334
2	0002	635.0	180.0	.283	-.93	.147	.195
3	0003	635.0	165.0	.260	-1.05	.248	.335
4	0004	635.0	307.0	.483	-.07	.253	.317
5	0005	635.0	282.0	.444	-.22	.192	.242
6	0006	635.0	278.0	.438	-.25	.202	.255
7	0007	635.0	195.0	.307	-.81	.231	.304
8	0008	635.0	241.0	.380	-.49	.206	.263
9	0009	635.0	155.0	.244	-1.13	.171	.234
10	0010	635.0	229.0	.361	-.57	.349	.448
11	0011	635.0	178.0	.280	-.94	.201	.267
12	0012	635.0	98.0	.154	-1.70	.029	.045
13	0013	635.0	163.0	.257	-1.06	.142	.192
14	0014	635.0	185.0	.291	-.89	.085	.112
15	0015	635.0	163.0	.257	-1.06	.139	.189
16	0016	635.0	214.0	.337	-.68	.250	.323
17	0017	635.0	191.0	.301	-.84	.256	.337
18	0018	635.0	134.0	.211	-1.32	.208	.293
19	0019	635.0	157.0	.247	-1.11	.264	.360
20	0020	635.0	302.0	.476	-.10	.228	.286
21	0021	635.0	136.0	.214	-1.30	.055	.077
22	0022	635.0	147.0	.231	-1.20	.121	.167

5864 BYTES OF WORKSPACE USED OF 208000 AVAILABLE IN PHASE-1

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
                    ***  LOGISTIC MODEL ITEM ANALYSER  ***

                                *** PHASE 2  ***

          THE PERSON ABILITY:FILE DIRE1.BLG :USING DICHOTOMOUS FIRST
        TIME RESPONSE / 22 ITEMS FOR REPEATED TEST WITH 4 CHOICES


 >CALib        TPRIOR, SPRIOR, GPRIOR, FLOAT, cycles=100, NEWTON=2, crit=0.001

 CALIBRATION PARAMETERS
 ======================

 MAXIMUM NUMBER OF EM CYCLES:                  100

 MAXIMUM NUMBER OF NEWTON CYCLES:               2

 CONVERGENCE CRITERION:                        .0010

 SUBJECT DISTRIBUTION:                         NORMAL PRIOR

 PLOT EMPIRICAL VS. FITTED ICC'S:              NO

 DATA HANDLING:                                RIGHT/WRONG/NOT PRESENTED DATA,
                                               IN CORE

 PRIOR DISTRIBUTION ON ASYMPTOTES:             YES

 PRIOR DISTRIBUTION ON SLOPES:                 YES

 PRIOR DISTRIBUTION ON THRESHOLDS:             YES

 SOURCE OF ITEM HYPERPARAMETERS:               PROGRAM DEFAULTS
                                               HYPERPARAMETERS WILL BE UPDATED
                                               EACH CYCLE
1
                          ******************************

                              CALIBRATION OF SUBTEST

                                      MATH

                          ******************************
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

METHOD OF SOLUTION
==========

EM CYCLES (MAXIMUM OF 100)
FOLLOWED BY NEWTON-RAPHSON STEPS (MAXIMUM OF 2)

QUADRATURE POINTS AND PRIOR WEIGHTS:

	1	2	3	4	5
POINT	-.4000E+01	-.3111E+01	-.2222E+01	-.1333E+01	-.4444E+00
WEIGHT	.1190E-03	.2805E-02	.3002E-01	.1458E+00	.3213E+00

	6	7	8	9	10
POINT	.4444E+00	.1333E+01	.2222E+01	.3111E+01	.4000E+01
WEIGHT	.3213E+00	.1458E+00	.3002E-01	.2805E-02	.1190E-03

PRIOR DISTRIBUTIONS ON ITEM PARAMETERS
(THRESHOLDS, NORMAL; SLOPES, LOG-NORMAL; GUESSING, BETA)

	THRESHOLDS		SLOPES		ASYMPTOTES	
ITEM	MU	SIGMA	MU	SIGMA	ALPHA	BETA
0001	.000	2.000	.000	.500	11.00	11.00
0002	.000	2.000	.000	.500	11.00	11.00
0003	.000	2.000	.000	.500	11.00	11.00
0004	.000	2.000	.000	.500	11.00	11.00
0005	.000	2.000	.000	.500	11.00	11.00
0006	.000	2.000	.000	.500	11.00	11.00
0007	.000	2.000	.000	.500	11.00	11.00
0008	.000	2.000	.000	.500	11.00	11.00
0009	.000	2.000	.000	.500	11.00	11.00
0010	.000	2.000	.000	.500	11.00	11.00
0011	.000	2.000	.000	.500	11.00	11.00
0012	.000	2.000	.000	.500	11.00	11.00
0013	.000	2.000	.000	.500	11.00	11.00
0014	.000	2.000	.000	.500	11.00	11.00
0015	.000	2.000	.000	.500	11.00	11.00
0016	.000	2.000	.000	.500	11.00	11.00
0017	.000	2.000	.000	.500	11.00	11.00
0018	.000	2.000	.000	.500	11.00	11.00
0019	.000	2.000	.000	.500	11.00	11.00
0020	.000	2.000	.000	.500	11.00	11.00
0021	.000	2.000	.000	.500	11.00	11.00
0022	.000	2.000	.000	.500	11.00	11.00

DATA FOR FIRST TWO OBSERVATIONS (TT=RIGHT, TF=WRONG, FF=NOT PRESENTED)

OBSERVATION 1 WEIGHT = 1.000
TT TF TF TT TT TT TT TF TT TT TT TT TF TF TF TF TF TF TF TF TF TF

OBSERVATION 2 WEIGHT = 1.000
TT TF TT TT TT TF TT TT TT TT TT TF TF TF TF TF TT TF TF TF TF TF

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
[EM STEP]

         -2 LOG LIKELIHOOD =     20009.5472

CYCLE  1:   LARGEST CHANGE =   3.78212

         -2 LOG LIKELIHOOD =     16535.2726
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =      6.60272     15.39728
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =          .45028       .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =         2.05432      2.00000

CYCLE  2:   LARGEST CHANGE =   1.04939

         -2 LOG LIKELIHOOD =     16367.9595
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =      5.77313     16.22687
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =          .53397       .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =         1.86391      2.00000

CYCLE  3:   LARGEST CHANGE =   1.42256

         -2 LOG LIKELIHOOD =     16347.5511
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =      5.53135     16.46865
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =          .56552       .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =         1.81194      2.00000

CYCLE  4:   LARGEST CHANGE =   3.48310

         -2 LOG LIKELIHOOD =     16332.5327
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =      5.44253     16.55747
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =          .60032       .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =         1.78344      2.00000

CYCLE  5:   LARGEST CHANGE =    .63438

         -2 LOG LIKELIHOOD =     16333.9651
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =      5.45676     16.54324
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =          .63244       .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =         1.79448      2.00000

CYCLE  6:   LARGEST CHANGE =    .88115

         -2 LOG LIKELIHOOD =     16328.8046
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =      5.47129     16.52871
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =          .64352       .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =         1.78496      2.00000

CYCLE  7:   LARGEST CHANGE =    .35254

         -2 LOG LIKELIHOOD =     16330.0068
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =      5.49415     16.50585
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =          .66402       .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =         1.80271      2.00000

CYCLE  8:   LARGEST CHANGE =    .16019
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =        .75069      .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       1.79931     2.00000

CYCLE 48:   LARGEST CHANGE =    .00691

         -2 LOG LIKELIHOOD =    16329.0946
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =    5.63626    16.36374
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =        .75157      .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       1.79934     2.00000

CYCLE 49:   LARGEST CHANGE =    .03079

         -2 LOG LIKELIHOOD =    16327.9029
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =    5.63240    16.36760
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =        .74852      .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       1.79862     2.00000

CYCLE 50:   LARGEST CHANGE =    .03495

         -2 LOG LIKELIHOOD =    16329.1402
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =    5.63341    16.36659
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =        .75200      .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       1.79910     2.00000

CYCLE 51:   LARGEST CHANGE =    .04057

         -2 LOG LIKELIHOOD =    16327.9060
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =    5.63194    16.36806
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =        .74827      .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       1.79858     2.00000

CYCLE 52:   LARGEST CHANGE =    .02038

         -2 LOG LIKELIHOOD =    16327.9273
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =    5.63300    16.36700
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =        .74980      .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       1.79898     2.00000

CYCLE 53:   LARGEST CHANGE =    .00886

         -2 LOG LIKELIHOOD =    16327.9172
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =    5.63494    16.36506
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =        .75087      .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       1.79910     2.00000

CYCLE 54:   LARGEST CHANGE =    .00694

         -2 LOG LIKELIHOOD =    16329.0951
UPDATED PRIOR ON ASYMPTOTES; ALPHA & BETA =    5.63635    16.36365
UPDATED PRIOR ON LOG SLOPES; MEAN & SD =        .75174      .50000
UPDATED PRIOR ON THRESHOLDS; MEAN & SD =       1.79914     2.00000

CYCLE 55:   LARGEST CHANGE =    .03079

         -2 LOG LIKELIHOOD =    16327.9039
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
                    ***  LOGISTIC MODEL ITEM ANALYSER  ***

                              *** PHASE 3  ***

          THE PERSON ABILITY:FILE DIRE1.BLG :USING DICHOTOMOUS FIRST
         TIME RESPONSE / 22 ITEMS FOR REPEATED TEST WITH 4 CHOICES


 >score        RSCtype=4;


 PARAMETERS FOR SCORING, RESCALING, AND TEST AND ITEM INFORMATION

 METHOD OF SCORING SUBJECTS:                  EXPECTATION A POSTERIORI
                                              (EAP; BAYES ESTIMATES)
 TYPE OF PRIOR:                               NORMAL
 SCORES WRITTEN TO FILE                       C:\BILOG3\DIRE1.SCO
 SCORES WRITTEN TO FILE                       DIRE1.PH3
 TYPE OF RESCALING:                           CENTER WITH RESPECT TO
                                              ESTIMATED LATENT DISTRIBUTION
 ITEM AND TEST INFORMATION:                   NONE REQUESTED

                  QUAD    PRIOR     PRIOR    RESCALING CONSTANTS
 TEST     NAME   POINTS    MEAN    STN DEV    SCALING    LOCATION
 ------------------------------------------------------------------
    1  MATH        10       .000     1.000      1.000        .000
1

                          ******************************

                                     SCORING

                          ******************************


 EAP SUBJECT ESTIMATION, SUBTEST:MATH

 QUADRATURE POINTS AND PRIOR WEIGHTS:

                  1           2           3           4           5
 POINT      -.4000E+01  -.3111E+01  -.2222E+01  -.1333E+01  -.4444E+00
 WEIGHT      .1190E-03   .2805E-02   .3002E-01   .1458E+00   .3213E+00

                  6           7           8           9          10
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

POINT	.4444E+00	.1333E+01	.2222E+01	.3111E+01	.4000E+01
WEIGHT	.3213E+00	.1458E+00	.3002E-01	.2805E-02	.1190E-03

CORRELATIONS AMONG SUBTEST SCORE ESTIMATES:

	MATH
MATH	1.000

MEANS AND STANDARD DEVIATIONS OF SCORE ESTIMATES:

TEST: MATH
MEAN: -.008
S.D.: .727

MARGINAL LATENT DISTRIBUTION FOR SUBTEST MATH
MEAN = -.008 S.D. = .963

	1	2	3	4	5
POINT	-.4000E+01	-.3111E+01	-.2222E+01	-.1333E+01	-.4444E+00
WEIGHT	.1037E-03	.2682E-02	.2997E-01	.1467E+00	.3229E+00

	6	7	8	9	10
POINT	.4444E+00	.1333E+01	.2222E+01	.3111E+01	.4000E+01
WEIGHT	.3188E+00	.1536E+00	.2112E-01	.4062E-02	.1024E-03

1

RESCALING DONE WITH RESPECT TO SUBJECT DISTRIBUTION

SUBTEST	SCALING CONSTANT	LOCATION CONSTANT
MATH	1.038	.008

1

SUBJECT WEIGHT	IDENTIFICATION SUBTEST	TRIED	RIGHT	PERCENT	ABILITY	S.E.	MARGINAL PROB
111101							
1.00	MATH	22	9	.4091	.8795	.6171	.000001
111102							
1.00	MATH	22	10	.4545	1.3136	.2905	.000002
111104							
1.00	MATH	22	9	.4091	1.1820	.4453	.000005
111105							
1.00	MATH	22	14	.6364	1.6957	.4514	.000001
111106							
1.00	MATH	22	13	.5909	1.4035	.2671	.000001
111107							
1.00	MATH	22	12	.5455	1.1271	.4873	.000000
111108							
1.00	MATH	22	15	.6818	1.8461	.4690	.000000
111109							

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1.00 112640	MATH	22	5	.2273	-.4382	.7979	.000008
1.00 112641	MATH	22	7	.3182	-.0633	.7265	.000002
1.00 112642	MATH	22	7	.3182	-.1578	.7426	.000001
1.00 112643	MATH	22	7	.3182	-.1419	.7324	.000001
1.00 112644	MATH	22	2	.0909	-.5810	.7356	.000135
1.00 113101	MATH	22	12	.5455	1.3669	.2620	.000002
1.00 113102	MATH	22	14	.6364	1.5006	.3396	.000001
1.00 113103	MATH	22	15	.6818	1.9704	.4512	.000001
1.00 113104	MATH	22	16	.7273	1.4376	.3053	.000000

1

SUBTEST MATH ; RESCALED ITEM PARAMETERS

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	DISPERSN S.E.	ASYMPTOTE S.E.
0001	-1.872	3.326	.563	.301	.317
	1.046*	1.341*	1.108*	.121*	.056*
0002	-3.498	1.580	2.214	.633	.228
	1.011*	.541*	1.293*	.217*	.029*
0003	-2.568	1.634	1.571	.612	.139
	.690*	.428*	.825*	.160*	.037*
0004	-1.432	2.170	.660	.461	.253
	.705*	.589*	.753*	.125*	.065*
0005	-2.547	2.035	1.252	.492	.322
	1.046*	.713*	1.170*	.172*	.046*
0006	-3.156	2.033	1.552	.492	.354
	1.118*	.685*	1.271*	.166*	.035*
0007	-4.041	2.892	1.397	.346	.211
	1.761*	1.258*	1.925*	.150*	.030*
0008	-3.605	2.343	1.539	.427	.298
	1.251*	.800*	1.400*	.146*	.031*
0009	-4.530	2.128	2.129	.470	.202
	1.362*	.772*	1.609*	.171*	.023*
0010	-1.793	2.020	.887	.495	.145
	.490*	.335*	.529*	.082*	.044*
0011	-2.594	1.374	1.888	.728	.181
	.701*	.376*	.891*	.199*	.038*
0012	-8.647	2.926	2.956	.342	.151
	3.511*	1.354*	3.892*	.158*	.015*
0013	-4.672	1.742	2.681	.574	.231
	1.382*	.639*	1.737*	.210*	.023*
0014	-6.745	2.403	2.807	.416	.280

0017	-3.167	2.028	1.561	.493	.204
	.993*	.641*	1.142*	.156*	.033*
0018	-4.717	2.531	1.864	.395	.158
	1.467*	.926*	1.667*	.145*	.021*
0019	-3.635	2.236	1.625	.447	.162
	1.169*	.775*	1.336*	.155*	.028*
0020	-2.985	1.953	1.528	.512	.391
	1.173*	.755*	1.352*	.198*	.037*
0021	-5.933	1.781	3.332	.562	.206
	1.858*	.673*	2.302*	.212*	.018*
0022	-4.084	1.334	3.062	.750	.204
	1.099*	.456*	1.546*	.256*	.024*

PARAMETER	MEAN	STN DEV
ASYMPTOTE	.232	.070
SLOPE	2.101	.504
LOG(SLOPE)	.716	.236
THRESHOLD	1.876	.766

MEAN & SD OF LATENT DISTRIBUTION AFTER RESCALING: .000 1.000

7164 BYTES OF WORKSPACE USED OF 208000 AVAILABLE IN PHASE-3

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก จ

ตัวอย่างคำสั่งและผลการวิเคราะห์ข้อมูลเพื่อประมาณค่าความสามารถที่แท้จริง ของข้อมูลที่มีการตรวจให้คะแนนความรู้บางส่วนด้วยโปรแกรม PARSCALE 2

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
        THE PERSON ABILITY:FILE PRE1.PSL :USING PARSCALE FIRST
       TIME RESPONSE / 22 ITEMS FOR REPEATED TEST WITH 4 CHOICE
>COMMENT  DATA FROM STUDENT 635 PERSONS
                  6 SCHOOLS
>FILes       DFName ='C:\PSL2\pre1.dat', SAVE;
>SAVE         SCOre='C:\PSL2\pre1.sco', INFormation='C:\PSL2\pre1.inf',
              PARM='C:\PSL2\pre1.par';
>INPut        NTEst=1, LENght=22,NIDch=6,NTOtal=22,MAXcat=4;
(6A1,2X,22A1)
>TEST         TNAme='MATH1', NBLock=1, ITEmes=(1(1)22);
>BLOck        BNAme=test1, NITems=22, NCAt=4, ORIginal=(0,1,2,3),
              MODIfied=(0,1,2,3);
>CALib        PARtial, LOGistic, SCAle=1.7, NQPts=25, CYCles=(100,1,1,1,1),
              CRIT=0.001;
>SCOre        EAP, PRInt, NAME=EAP;
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

111101	1	1.00		
1 EAP	2.64	1.00	2.0089	0.5800
111102	2	1.00		
1 EAP	2.00	1.00	0.7328	0.4067
111104	3	1.00		
1 EAP	2.42	1.00	1.4579	0.4863
111105	4	1.00		
1 EAP	2.37	1.00	1.7876	0.4092
111106	5	1.00		
1 EAP	2.43	1.00	1.5554	0.4679
111107	6	1.00		
1 EAP	2.33	1.00	1.6504	0.4153
111108	7	1.00		
1 EAP	2.56	1.00	1.7730	0.4285
111109	8	1.00		
1 EAP	2.83	1.00	2.4358	0.5806
111110	9	1.00		
1 EAP	2.50	1.00	1.4437	0.4319
111111	10	1.00		
1 EAP	3.00	1.00	3.1336	0.5089
111112	11	1.00		
1 EAP	2.85	1.00	2.5064	0.5538
111113	12	1.00		
1 EAP	2.53	1.00	2.0590	0.4796
111114	13	1.00		
1 EAP	2.95	1.00	3.1947	0.4827
111115	14	1.00		

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ภาคผนวก ฉ

ตัวอย่างผลการวิเคราะห์ข้อมูลเพื่อประมาณค่าความสามารถที่แท้จริงของข้อมูล ด้วยโมเดล MRMLC ด้วยโปรแกรม LPCM-WIN 1.0

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
*******************************************************************************
c:\lpcmwin\thesis\lpcmc1.lis on  Tuesday 4. September 2001 at  13:11:51
Results for data file c:\lpcmwin\thesis\lpcmc1.dat
obtained by means of program LpcM-WiN 1.0

*******************************************************************************
Calculation of a Linear Logistic Test Model
Total Sample
*******************************************************************************

581 Persons
  1 Treatment Group
 22 Items (out of 22 Items)
  2 Basic Parameter Estimates
  1 Iterations
Convergence reached in 0.0000 seconds

Total Log-Likelihood          -6452.738169

          Basic   Standard
           Parameters     Errors         z-Value         Significance

 1       0.0000           0.0415          0.0000         n.s.
 2       0.0000           0.0415          0.0000         n.s.
*******************************************************************************
Calculation of a Rasch Model
Total Sample
*******************************************************************************

581 Persons
  1 Treatment Group
 22 Items (out of 22 Items)
 21 Item Parameter Estimates (1 Item(s) was (were) set to zero for normalizati
 17 Iterations
Convergence reached in 0.1100 seconds

Total Log-Likelihood          -6236.047151

          Item                    Standard
        Parameters        Errors

  1          0    (Normalization Condition)
  2     -0.8019                  0.1014
  3     -1.0163                  0.1028
  4     -0.0391                  0.1224
  5     -0.3355                  0.1040
  6     -0.7272                  0.1008
  7     -0.6293                  0.1013
  8     -1.1419                  0.1063
  9     -1.0696                  0.1043
 10     -0.4772                  0.1045
 11     -1.3896                  0.1086
 12     -1.4096                  0.1085
 13     -0.8523                  0.1016
 14     -0.8271                  0.1015
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

```
 15      -0.8523                  0.1016
 16      -1.3306                  0.1105
 17      -0.8271                  0.1015
 18      -1.4500                  0.1277
 19      -0.6212                  0.1015
 20      -0.4377                  0.1048
 21      -1.1056                  0.1054
 22      -1.3015                  0.1098

                        ************************************
                          Conditional Likelihood-Ratio-Test
                         Hypothesis 0 against Hypothesis 1
                        ************************************

Andersen chi_2 =                  433.3820
Degrees of Freedom =              19
chi_2 at alpha(5 percent) =       30.1348         (Wilson-Hilferty approximatio
chi_2 at alpha(1 percent) =       36.2160         (Wilson-Hilferty approximatio

                        ************************************

                        Reproduced Itemparameter

                        ************************************

         LLTM (rescaled)          Rasch Model

 1       0.0000                    0.0000
 2       0.0000                   -0.8019
 3       0.0000                   -1.0163
 4       0.0000                   -0.0391
 5       0.0000                   -0.3355
 6       0.0000                   -0.7272
 7       0.0000                   -0.6293
 8       0.0000                   -1.1419
 9       0.0000                   -1.0696
 10      0.0000                   -0.4772
 11      0.0000                   -1.3896
 12      0.0000                   -1.4096
 13      0.0000                   -0.8523
 14      0.0000                   -0.8271
 15      0.0000                   -0.8523
 16      0.0000                   -1.3306
 17      0.0000                   -0.8271
 18      0.0000                   -1.4500
 19      0.0000                   -0.6212
 20      0.0000                   -0.4377
 21      0.0000                   -1.1056
 22      0.0000                   -1.3015

Pearson Correlation between parameters =        0.0000
```

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

1	4	11	1.4835e+00	-1.5738e-02	3.1792e+00
2	7	7	1.0000e-09	-1.4696e+00	1.4696e+00
3	0	10	N.E.	1.2252e+00	N.E.
4	5	7	4.6113e-01	-1.0790e+00	2.0725e+00
5	15	11	-6.8378e-01	-2.0223e+00	6.1828e-01
6	10	12	3.4451e-01	-9.6351e-01	1.6717e+00
7	4	5	2.8100e-01	-1.4544e+00	2.0867e+00
8	0	6	N.E.	3.3654e-01	N.E.
9	11	9	-3.5348e-01	-1.7025e+00	9.7347e-01
10	7	5	-4.6113e-01	-2.0724e+00	1.0790e+00
11	2	6	1.3228e+00	-5.8709e-01	3.7781e+00
12	3	3	1.0000e-09	-2.1576e+00	2.1577e+00
13	3	7	1.0845e+00	-6.1030e-01	3.0520e+00
14	1	4	1.5347e+00	-9.1586e-01	5.5254e+00
15	0	2	N.E.	-1.6628e+00	N.E.
16	4	7	7.4215e-01	-8.5845e-01	2.4843e+00
17	0	5	N.E.	3.4164e-02	N.E.
18	0	3	N.E.	-8.4739e-01	N.E.
19	9	13	6.9342e-01	-6.1887e-01	2.0451e+00

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ประวัติผู้วิจัย

นางอวยพร เรืองตระกูล เกิดเดือนตุลาคม พ.ศ. 2496 กรุงเทพมหานคร สำเร็จปริญญาการศึกษาบัณฑิต วิชาเอกคณิตศาสตร์ วิชาโทเคมี จากมหาวิทยาลัยศรีนครินทรวิโรฒปทุมวัน ในปีการศึกษา 2520 และได้ปริญญาครุศาสตรมหาบัณฑิต สาขาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2526 รับราชการครั้งแรกที่โรงเรียนสิงหราชพิทยาคม กรมสามัญศึกษา กระทรวงศึกษาธิการ ในปี พ.ศ. 2521 และโอนมาเป็นอาจารย์ประจำภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปี พ.ศ. 2534 และได้เข้าศึกษาต่อในหลักสูตรปริญญาครุศาสตรดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ในปีการศึกษา 2540

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย