

การศึกษาค้นคว้าของการวัดการเปลี่ยนแปลงน้ำหนักตัวซ้ำสองครั้ง ภายใต้เงื่อนไข
ความยาวของแบบทดสอบที่ต่างกัน และ ขนาดกลุ่มตัวอย่างไม่เท่ากัน

ปริญญานิพนธ์

ของ

จรรยา สิงห์ทอง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

สิงหาคม 2550

การศึกษามูลของการวัดการเปลี่ยนแปลงหัวข้อแบบวัดซ้ำสองครั้ง ภายใต้เงื่อนไข
ความยาวของแบบทดสอบที่ต่างกัน และ ขนาดกลุ่มตัวอย่างไม่เท่ากัน

ปริญญานิพนธ์

ของ

จรรยา สิงห์ทอง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุขบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

สิงหาคม 2550

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัย

จาก

มหาวิทยาลัยศรีนครินทรวิโรฒ

ประเภททุนงบประมาณแผ่นดิน ประจำปีงบประมาณ 2550

ประกาศคุณูปการ

ปริญญานิพนธ์นี้สำเร็จได้ด้วยความกรุณาอย่างสูงจากคณะกรรมการควบคุมปริญญานิพนธ์ คือ รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนันตพงษ์ ประธานควบคุมปริญญานิพนธ์ รองศาสตราจารย์ ดร.ผจงจิต อินทสุวรรณ และผู้ช่วยศาสตราจารย์ ดร.สุรัชย์ มีชาญ กรรมการควบคุมปริญญานิพนธ์ ที่ได้ให้คำปรึกษา ข้อชี้แนะในการจัดทำงานวิจัยนี้ทุกขั้นตอน และแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่ ตลอดจนให้กำลังใจเสมอมา ผู้วิจัยซาบซึ้งและขอกราบขอบพระคุณในความกรุณาสมาเสมอของท่าน ทั้งสาม ด้วยความเคารพอย่างสูง

ขอกราบขอบพระคุณกรรมการที่แต่งตั้งเพิ่มเติมทั้งสองท่าน คือ ผู้ช่วยศาสตราจารย์ ดร.องอาจ นัยพัฒน์ และ อาจารย์ ดร.รังสรรค์ มณีเล็ก ที่ให้ความกรุณาเป็นกรรมการในการสอบปริญญานิพนธ์ และให้ข้อเสนอแนะที่เป็นประโยชน์เพิ่มเติมทำให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณ Dr. Leo Stam แห่ง Scientific Software International (SSI) และ อาจารย์ ดร.สุนิสา จัยม่วงศรี มหาวิทยาลัยสุโขทัยธรรมาธิราช ที่ให้ความกรุณาช่วยเหลือแนะนำการวิเคราะห์ คะแนนมวลความรู้ด้วยโปรแกรม BILOGMG-3 และ Dr. Mathilda, du Toit แห่ง Scientific Software International (SSI) ที่กรุณาให้ความรู้และความเข้าใจเกี่ยวกับการวิเคราะห์คะแนนสอบก่อนเรียนจริง และอัตราพัฒนาการจริงรายบุคคลด้วยโปรแกรม HLM 6.02

ขอขอบพระคุณผู้เชี่ยวชาญทุกท่านที่ให้ความช่วยเหลือในการตรวจสอบคุณภาพของเครื่องมือ ขอขอบพระคุณผู้อำนวยการศูนย์บริการการศึกษานอกโรงเรียนเขตและอำเภอที่ให้ความอนุเคราะห์ นักศึกษาเพื่อทดลองใช้เครื่องมือทั้งสองครั้ง ในการพัฒนาเครื่องมือให้มีคุณภาพสำหรับการวิจัย

ขอขอบพระคุณผู้อำนวยการศูนย์การศึกษานอกโรงเรียนกรุงเทพมหานคร 1 ที่ให้การสนับสนุน ผู้วิจัยในการศึกษาต่อ และกรุณาให้ความอนุเคราะห์กลุ่มตัวอย่างในการวิจัย ขอขอบคุณผู้บริหาร สถานศึกษา ข้าราชการ พนักงานราชการ ครูศูนย์การเรียนรู้ชุมชน และ ขอขอบใจนักศึกษาที่เป็นกลุ่ม ตัวอย่างทุกคน ที่ให้ความร่วมมือเป็นอย่างดีในการเก็บรวบรวมข้อมูลเพื่อการวิจัย

ขอขอบคุณอาจารย์ ดร. หทัยรัตน์ มาประณีต ภาควิชาสังคมวิทยา คณะสังคมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ที่กรุณาตรวจความถูกต้องของบทคัดย่อภาษาอังกฤษ

ท้ายสุดผู้วิจัยขอกราบขอบพระคุณ บิดา มารดา และขอบคุณ ญาติพี่น้องและเพื่อนๆ ทุกคนที่ ให้ทั้งกำลังใจและกำลังใจที่ดีเยี่ยมตลอดระยะเวลาที่ศึกษาและทำวิจัย

จรรยา สิงห์ทอง

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง.....	1
ความมุ่งหมายของการวิจัย.....	13
ความสำคัญของการวิจัย.....	14
ขอบเขตของการวิจัย.....	15
กรอบแนวคิดในการวิจัย.....	16
นิยามศัพท์เฉพาะ.....	19
สมมติฐานของการวิจัย.....	21
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	23
ทฤษฎีการทดสอบแบบมาตรฐานเดิม.....	23
การวัดการเปลี่ยนแปลงตามทฤษฎีการทดสอบแบบมาตรฐานเดิม.....	24
โมเดลพระระดับสำหรับการวัดการเปลี่ยนแปลง.....	39
ทฤษฎีการตอบข้อสอบ.....	43
การวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบ.....	47
คะแนนมวลความรู้.....	64
งานวิจัยที่เกี่ยวข้อง.....	71
3 วิธีดำเนินการวิจัย.....	87
การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง.....	87
การพัฒนาเครื่องมือที่ใช้ในการวิจัย.....	89
การเก็บรวบรวมข้อมูล.....	97
การจัดกระทำและการวิเคราะห์ข้อมูล.....	98

สารบัญ (ต่อ)

บทที่	หน้า
4 ผลการวิเคราะห์ข้อมูล.....	108
การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ.....	109
การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลง.....	123
การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง.....	144
การศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง.....	151
5 สรุป อภิปราย และข้อเสนอแนะ.....	155
สรุปผลการวิจัย.....	158
อภิปรายผลการวิจัย.....	164
ข้อเสนอแนะ.....	169
บรรณานุกรม.....	173
ภาคผนวก.....	183
ภาคผนวก ก.....	184
ภาคผนวก ข.....	212
ภาคผนวก ค.....	232
ประวัติย่อผู้วิจัย.....	257

บัญชีตาราง

ตาราง	หน้า
1 จำนวนนักศึกษาที่เป็นกลุ่มตัวอย่าง.....	88
2 ผลการวิเคราะห์เนื้อหา เรื่อง สมการเชิงเส้นตัวแปรเดียว ระดับมัธยมศึกษาตอนต้น.....	90
3 จำนวนนักศึกษาที่ตอบแบบทดสอบแต่ละฉบับในการทดลองใช้เครื่องมือครั้งที่ 1.....	92
4 จำนวนนักศึกษาที่ตอบแบบทดสอบในการทดลองใช้เครื่องมือครั้งที่ 2.....	93
5 ผลการวิเคราะห์องค์ประกอบเชิงสำรวจของแบบทดสอบคณิตศาสตร์.....	94
6 ค่าพารามิเตอร์ของข้อสอบเป็นรายข้อของแบบทดสอบ ตามทฤษฎีการตอบข้อสอบ.....	185
7 ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้ง.....	110
8 ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้ง ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน.....	187
9 ค่าสถิติ Kolmogorov-Smirnov ของคะแนนสอบแต่ละครั้ง.....	118
10 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลง ภายใต้เงื่อนไขความยาวของแบบทดสอบต่างกันและขนาดกลุ่มตัวอย่างไม่เท่ากัน.....	120
11 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และ เชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 40 ข้อ..	189
12 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และ เชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 30 ข้อ..	189
13 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 20 ข้อ...	189
14 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 40 ข้อ...	190
15 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 30 ข้อ....	190
16 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 20 ข้อ....	190
17 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 40 ข้อ....	191

บัญชีตาราง (ต่อ)

ตาราง	หน้า
18 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 30 ข้อ...	191
19 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 20 ข้อ.....	191
20 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 40 ข้อ.....	192
21 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 30 ข้อ....	192
22 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 20 ข้อ.....	192
23 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการสอบสามครั้ง ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบคณิตศาสตร์ที่มีความยาวต่างกัน.....	121
24 การวิเคราะห์ความแตกต่างของค่าเฉลี่ยของคะแนนสอบซ้ำสามครั้ง ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน.....	122
25 การเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนสอบซ้ำสามครั้ง ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน.....	123
26 ค่าสัมประสิทธิ์การถดถอยระดับบุคคลของโมเดลพหุระดับสามแบบ ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน.....	124
27 ผลการทดสอบอิทธิพลคงที่และอิทธิพลสุ่มของโมเดลพหุระดับสามแบบ ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน.....	126
28 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 40 ข้อ.....	193
29 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 30 ข้อ.....	194
30 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 20 ข้อ.....	195

บัญชีตาราง (ต่อ)

ตาราง	หน้า
31 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 40 ข้อ.....	196
32 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 30 ข้อ.....	197
33 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 20 ข้อ.....	198
34 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 40 ข้อ.....	199
35 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 30 ข้อ.....	200
36 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 20 ข้อ.....	201
37 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 40 ข้อ.....	202
38 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 30 ข้อ.....	203
39 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบคณิตศาสตร์จำนวน 20 ข้อ.....	204
40 ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี.....	133
41 ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงแต่ละวิธี ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วย แบบทดสอบที่มีความยาวต่างกัน.....	205
42 ค่าสถิติ Kolmogorov-Smirnov ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจาก แต่ละวิธี.....	143
43 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ได้ มาจากแต่ละวิธี ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน...	145

บัญชีตาราง (ต่อ)

ตาราง	หน้า
44 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ได้มาจากแต่ละวิธีกับวิธี โมเดลพหุระดับ ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน...	148
45 ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลง ระหว่างแต่ละวิธีกับวิธีโมเดลพหุระดับ ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบ ที่มีความยาวต่างกัน.....	207
46 ผลการทดสอบความแตกต่างรายคู่ของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการ เปลี่ยนแปลงที่ได้มาจากแต่ละวิธีกับวิธีโมเดลพหุระดับ ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน.....	209
47 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความลำเอียงของคะแนน การเปลี่ยนแปลงที่ได้จากแต่ละวิธี ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มี ความยาวต่างกัน.....	152
48 สรุปผลการศึกษาการวัดการเปลี่ยนแปลงแต่ละวิธี ภายใต้เงื่อนไขความยาวของ แบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน.....	163

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 ลูกบาศก์ข้อมูล ที่มีผู้สอบ 4 คน สอบข้อสอบ 3 ข้อ ที่มีการสอบ 2 ครั้ง.....	49
2 แผนภาพโมเดล MRMLC.....	61
3 จำนวนองค์ประกอบในแบบทดสอบคณิตศาสตร์.....	95
4 ค่าฟังก์ชันสารสนเทศของแบบทดสอบคณิตศาสตร์ในแต่ละระดับความสามารถ	96
5 ความสัมพันธ์ระหว่างคะแนนเฉลี่ยกับช่วงเวลาในการสอบซ้ำสามครั้ง ของ กลุ่มตัวอย่างจำนวน 1,085 คน 500 คน 300 คน และ 100 คน ด้วย แบบทดสอบคณิตศาสตร์ที่มีความยาว 40 ข้อ 30 ข้อ และ 20 ข้อ.....	119
6 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการ เปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี ของกลุ่มตัวอย่างครั้งที่ 1,085 คน 500 คน 300 คน และ 100 คน เมื่อความยาวของแบบทดสอบไม่เท่ากัน.....	146
7 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนก่อนเรียนกับคะแนนการเปลี่ยนแปลง ที่ประมาณค่ามาจากแต่ละวิธีของกลุ่มตัวอย่างไม่เท่ากัน เมื่อความยาว ของแบบทดสอบครั้งที่ 40 ข้อ 30 ข้อ และ 20 ข้อ.....	146
8 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลง ของกลุ่มตัวอย่างที่ ครั้งที่ 1,085 คน 500 คน 300 คน และ 100 คน เมื่อความยาวของ แบบทดสอบที่ต่างกัน.....	149
9 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลง ของกลุ่มตัวอย่าง ที่ไม่เท่ากัน เมื่อความยาวของแบบทดสอบครั้งที่ 40 ข้อ 30 ข้อ และ 20 ข้อ ...	149
10 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของคะแนนการเปลี่ยนแปลง ที่ได้จากแต่ละวิธี ของกลุ่มตัวอย่างครั้งที่ 1,085 คน 500 คน 300 คน และ 100 คน เมื่อความยาวของแบบทดสอบ 40 ข้อ 30 ข้อ และ 20 ข้อ.....	153
11 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของคะแนนการเปลี่ยนแปลง ที่ได้จากแต่ละวิธี ของกลุ่มตัวอย่างที่ไม่เท่ากัน เมื่อความยาวของแบบทดสอบ ครั้งที่ 40 ข้อ 30 ข้อ และ 20 ข้อ.....	153

การศึกษานผลของการวัดการเปลี่ยนแปลงน้ำหนักวิธีแบบวัดซ้ำสองครั้ง ภายใต้เงื่อนไข
ความยาวของแบบทดสอบที่ต่างกัน และ ขนาดกลุ่มตัวอย่างไม่เท่ากัน

บทคัดย่อ
ของ
จรรยา สิงห์ทอง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุขบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา
สิงหาคม 2550

จรรยา สิงห์ทอง. (2550). การศึกษาผลของการวัดการเปลี่ยนแปลงน้ำหนักวิธีแบบวัดซ้ำสองครั้ง ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และ ขนาดกลุ่มตัวอย่างไม่เท่ากัน. ปริญญาานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม: รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนตพงษ์, รองศาสตราจารย์ ดร. ผจญจิต อินทสุวรรณ, ผู้ช่วยศาสตราจารย์ ดร. สุรัชย์ มีชาญ

การวิจัยครั้งนี้เป็นการศึกษาผลของการวัดการเปลี่ยนแปลงน้ำหนักวิธีแบบวัดซ้ำสองครั้ง คือ 1) การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ 2) การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ 3) การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง 4) การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และ 5)การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ วิธีประมาณค่าที่ใช้เป็นเกณฑ์ คือ การวัดการเปลี่ยนแปลงด้วยวิธีโมเดลพระระดับ ภายใต้เงื่อนไข ความยาวของแบบทดสอบ 3 ขนาด คือ 40, 30 และ 20 ข้อ และ ขนาดกลุ่มตัวอย่าง 4 ขนาด คือ 1,000, 500, 300 และ 100 คน โดยพิจารณาผลของการวัดการเปลี่ยนแปลงจาก ค่าความสัมพันธ์ของคะแนนการเปลี่ยนแปลงและประสิทธิภาพของการวัดคะแนนการเปลี่ยนแปลง

เครื่องมือที่ใช้ในการวิจัยในครั้งนี้ เป็นแบบทดสอบคณิตศาสตร์ เรื่องสมการเชิงเส้นตัวแปรเดียว ระดับมัธยมศึกษาตอนต้น ซึ่งผู้วิจัยสร้างขึ้น มีลักษณะเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ ด้วยการสุ่มแบบแบ่งชั้น โดยใช้มวลความรู้เป็นชั้นและข้อสอบเป็นหน่วยในการสุ่ม เพื่อจัดเป็นแบบทดสอบที่มีความยาว 30 ข้อ และ 20 ข้อ

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ เป็นนักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 ระดับมัธยมศึกษาตอนต้น ของศูนย์บริการการศึกษานอกโรงเรียนเขต ในสังกัดศูนย์การศึกษา นอกโรงเรียนกรุงเทพมหานคร 1 ที่ได้มาด้วยวิธีการสุ่มแบบหลายขั้นตอน จำนวน 1,085 คน แล้วสุ่มตัวอย่างอย่างง่าย เพื่อให้กลุ่มตัวอย่างมีขนาด 500 คน, 300 คน และ 100 คน

ผลการวิจัยสรุปได้ดังนี้

การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงไม่เป็นลบเข้าใกล้ศูนย์ ส่วนคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์กับวิธีโมเดลพระระดับมีความสัมพันธ์กันทางบวกอยู่ในระดับที่สูงกว่าวิธีอื่น ในทุกเงื่อนไขความยาวของแบบทดสอบ และขนาดกลุ่มตัวอย่าง

การวัดคะแนนการเปลี่ยนแปลงทุกวิธีมีความเที่ยงตรงระดับเดียวกันในการประมาณค่าคะแนนการเปลี่ยนแปลง การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์มีความเชื่อมั่นสูงกว่าวิธีอื่น ในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง

RESULTS OF FIVE TWO-WAVE CHANGE MEASUREMENT APPROACHES IN
DIFFERENT TEST LENGTHS AND UNEQUAL SAMPLE SIZES

AN ABSTRACT

BY

JANYA SINGHATHONG

Presented in Partial Fulfillment of the Requirements for the
Doctor of Education Degree in Testing and Measurement
at Srinakharinwirot University

August 2007

Janya Singhathong. (2007). *Results of Five Two-wave Change Measurements in Different Test Lengths and Unequal Sample Sizes*. Dissertation, Ed.D. (Testing and Measurement). Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee: Assoc. Prof. Dr. Boonchird Pinyoanuntapong, Assoc. Prof. Dr. Pachongchit Intasuwan, Asst. Prof. Dr. Surachai Meechan.

The purpose of this research was to investigate the results of five two-wave change measurements: 1) different score, 2) relative change score, 3) latent change score, 4) different domain score, and 5) relative different domain score measurements. Multilevel model was employed to compare the qualities of these five measurements under different conditions of three test lengths (40, 30 and 20 items) and four unequal sample sizes (1,085, 500, 300 and 100 students).

The tool used in this research was a mathematics test for lower secondary level about linear equation with one variable. It comprised 40 items and four alternative choices in each item. These 40 items were drawn twice by using stratified random sampling in order to form 30-item and 20-item tests respectively.

The samples used in the study were 1,085 non-formal education students, of 2001 Basic Education Curriculum at the lower secondary level, under the jurisdiction of Bangkok Non-formal Education Center 1. They were obtained by using multi-stage random sampling. The simple random sampling was also used to arrange 500- student, 300-student, and 100-student sample sizes respectively.

The results of the research were as follows:

The different domain score and relative different domain score measurements had a non-negative and near-zero correlation with pretest and change score. The change score obtained from the relative different score measurement with multilevel model had positive correlation which was higher than other measurements in all different conditions of test lengths and unequal sample sizes.

All change score measurements had the same level of validity. However, the reliability level of relative change score measurement was higher than others in all different conditions of test lengths and unequal sample sizes.

บทที่ 1

บทนำ

ภูมิหลัง

การวัดการเปลี่ยนแปลงของผู้เรียนเป็นสิ่งสำคัญสำหรับการจัดการศึกษา ทั้งผู้บริหารสถานศึกษา ครูและผู้เกี่ยวข้องกับการศึกษาทุกฝ่ายต่างมีความคาดหวังให้ผู้เรียนเกิดการเปลี่ยนแปลงพฤติกรรมเพื่อให้เป็นไปตามที่พึงประสงค์ คือ เกิดความเจริญงอกงามและการพัฒนาขึ้น ซึ่งจัดได้ว่าเป็นหัวใจสำคัญของการจัดการศึกษาตามพระราชบัญญัติการศึกษาแห่งชาติ พุทธศักราช 2542 และที่แก้ไขเพิ่มเติม (ฉบับที่ 2) พ.ศ. 2545 มาตรา 6 และมาตรา 26 ดังนั้น ครูและผู้เกี่ยวข้องทางการศึกษาจึงได้นำการวัดประเมินผล (Assessment) มาบูรณาการกับการจัดกิจกรรมการเรียนรู้ เพื่อที่จะได้เห็นการพัฒนาหรือความก้าวหน้าหรือการเปลี่ยนแปลงของผู้เรียนได้ตลอดเวลา จะเห็นได้ว่าการตรวจสอบพฤติกรรมการเรียนรู้ของบุคคลนั้นจึงมีความสำคัญและจำเป็นอย่างยิ่งต่อการจัดการศึกษาของชาติ ซึ่งผลจากการตรวจสอบพฤติกรรมจะช่วยให้ทราบถึงการเปลี่ยนแปลงของบุคคลทั้งในทางที่เจริญขึ้นและทางที่เสื่อมลง เพื่อเป็นสารสนเทศในการตัดสินใจของครูและนักเรียนในการส่งเสริมปรับปรุงพฤติกรรมการเรียนรู้ของผู้เรียนให้สามารถพัฒนาไปตามศักยภาพของแต่ละบุคคล และยังเป็นสารสนเทศในการตัดสินใจของผู้บริหารในด้านการบริหารการศึกษาก็ด้วย

การทดสอบ (Testing) เป็นการวัดผลอย่างหนึ่งที่ใช้แบบทดสอบเป็นเครื่องมือ ซึ่งการทดสอบสามารถแบ่งตามลักษณะการแปลความหมาย (Interpretation) ของคะแนนที่สอบได้หรือคะแนนสังเกตที่วัดได้ออกเป็นสามประเภทใหญ่ (Henning. 1987: Online; Paula. 2001: Online; Hansen. 2005: Online; O'Grince. 2005: Online) ได้แก่ การทดสอบอิงกลุ่ม (Norm-referenced testing) ซึ่งเป็นการทดสอบที่แปลความหมายของคะแนนจากการทดสอบด้วยแบบทดสอบที่เรียกว่าแบบทดสอบอิงกลุ่มแล้วนำไปเทียบกับคะแนนของผู้สอบคนอื่นในกลุ่ม เพื่อบอกว่าผู้สอบที่ได้คะแนนนี้จะอยู่ลำดับที่เท่าไรของกลุ่ม การทดสอบอิงเกณฑ์ (Criterion-referenced testing) ซึ่งเป็นการทดสอบที่แปลความหมายของคะแนนจากการทดสอบด้วยแบบทดสอบอิงเกณฑ์แล้วนำไปเทียบกับเกณฑ์ (Criteria) หรือมาตรฐาน (Standard) เพื่อบอกว่าผู้สอบมีความรอบรู้หรือสามารถทำอะไรได้ และการทดสอบอิงพัฒนาการ (Growth-referenced testing) ซึ่งเป็นการทดสอบที่แปลความหมายด้วยการนำคะแนนที่ได้จากแบบทดสอบไปเทียบกับคะแนนสอบครั้งแรก เพื่อบอกว่าผู้สอบมีพัฒนาการเรียนรู้หรือมีการเปลี่ยนแปลงไปเท่าไร

การทดสอบอิงพัฒนาการ เป็นการวัดการเปลี่ยนแปลงที่มีกระบวนการวัดความแตกต่างของ พฤติกรรมการเรียนรู้ที่เกิดขึ้นระหว่างก่อนการเรียนรู้กับหลังการเรียนรู้ ซึ่งนักจิตวิทยาได้ศึกษาและ พัฒนาวิธีการวัดการเปลี่ยนแปลงมาอย่างต่อเนื่องเป็นระยะเวลายาวนานนับตั้งแต่ปี ค.ศ. 1924 โดยธอร์นไคค์ (Heidenberg. 2000: 1; citing Thorndike. 1924. Journal of Experimental Psychology. pp. 225-232) ที่มุ่งปรับปรุงวิธีการวัดการเปลี่ยนแปลงรายบุคคลหรือกลุ่มที่มีคุณภาพ ให้สามารถวัด การเปลี่ยนแปลงได้ถูกต้อง (Accuracy) แม่นยำ (Precision) มีความคลาดเคลื่อน (Error) จากคะแนน การเปลี่ยนแปลงจริงน้อยที่สุด จากการศึกษาเอกสารและงานวิจัยพบว่า การวัดการเปลี่ยนแปลงมี พัฒนาการเริ่มต้นจากการวัดด้วยข้อมูลที่มีการวัดซ้ำสองครั้ง ต่อมานักจิตวิทยาได้พัฒนาวิธีการวัดการ เปลี่ยนแปลงอย่างต่อเนื่อง จนกระทั่งเป็นการวัดการเปลี่ยนแปลงระยะยาวที่มีการเก็บรวบรวมข้อมูล หลายครั้ง จึงทำให้นักจิตวิทยาได้นำจำนวนครั้ง (Wave) ในการเก็บรวบรวมข้อมูลมาเป็นเกณฑ์ในการ แบ่งวิธีการวัดการเปลี่ยนแปลงร่วมกับลักษณะของวิธีการวัดและผลของการวัด

วิธีการวัดการเปลี่ยนแปลงมีสองแนวคิด ได้แก่ แนวคิดแรกเป็นวิธีการวัดแบบมาตรฐานเดิม (Classical measurement) นักจิตวิทยาได้คำนวณคะแนนการเปลี่ยนแปลงเชิงปริมาณแบบมาตรฐานเดิม ที่มีผลการวัดเป็นปริมาณการเปลี่ยนแปลงรายบุคคล สถิติที่ใช้ไม่ซับซ้อน และมีการเก็บรวบรวมข้อมูล สองครั้ง (Two-wave) โดยผลที่ได้จากการวัดการเปลี่ยนแปลงวิธีแรกนี้เรียกว่า คะแนนความแตกต่าง (Difference score) ซึ่งได้จากผลต่างระหว่างคะแนนดิบหรือคะแนนสังเกตในการสอบสองครั้ง ต่อมา นักจิตวิทยาได้พัฒนาวิธีการวัดการเปลี่ยนแปลงเชิงปริมาณวิธีอื่นๆ ขึ้นมา เรียกผลการวัดการเปลี่ยนแปลง นี้ว่า คะแนนเพิ่ม (Gain score) และ แนวคิดที่สองเป็นวิธีการแนวใหม่ (Modern methods) ซึ่งนักจิตวิทยา ใช้โมเดลตามแนวคิดที่ได้รับการพัฒนาขึ้นในระยะหลัง ส่วนใหญ่มีการเก็บรวบรวมข้อมูลหลายครั้ง (Multiwave) หรือ เป็นการวัดการเปลี่ยนแปลงระยะยาว ผลการวัดการเปลี่ยนแปลงเชิงปริมาณนี้มัก จะเรียกว่า คะแนนพัฒนาการ (Growth score) แต่โดยทั่วไปแล้วคะแนนประเภทนี้ทั้งที่เป็นผลจากการวัด การเปลี่ยนแปลงแบบมาตรฐานเดิม (Classic) และการวัดการเปลี่ยนแปลงแนวใหม่ (Modern) จะเรียกว่า คะแนนการเปลี่ยนแปลง (Change score)

วิธีการวัดการเปลี่ยนแปลงตามแบบมาตรฐานเดิม (Classical measurement) พบว่ามี หลากหลายวิธี เป็นการวัดการเปลี่ยนแปลงที่มีการเก็บรวบรวมข้อมูลสองครั้ง คือ ก่อนเรียนและหลังเรียน ในระยะแรกใช้คะแนนดิบหรือคะแนนสังเกต (Observed scores) รวม ซึ่งเป็นคะแนนที่ได้มาจากจำนวนข้อ ที่ตอบถูกต้องทั้งหมดในแบบทดสอบของผู้สอบ อันเป็นแนวคิดตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical test theory) คือ คะแนนสอบก่อนเรียนและคะแนนสอบหลังเรียนจะอยู่ในรูปของโมเดลเชิงบวก (Additive model) โดยคะแนนสอบก่อนเรียนประกอบไปด้วยคะแนนสองส่วน คือ คะแนนจริง (True score) กับคะแนนความคลาดเคลื่อน (Error score) ($X=T+E$) ส่วนคะแนนหลังเรียนประกอบไปด้วย

คะแนนสามส่วน คือ คะแนนจริงที่เป็นคะแนนสอบครั้งแรก คะแนนการเปลี่ยนแปลงที่เป็นคะแนนการเปลี่ยนแปลงจริง และความคลาดเคลื่อน ($Y = T_x + C + E$) ที่มีข้อตกลงเบื้องต้นว่า ความคลาดเคลื่อนจะเป็นตัวแปรเชิงสุ่ม (Random variable) มีค่าเฉลี่ยเป็นศูนย์ และความแปรปรวนของความคลาดเคลื่อนทั้งสองครั้งเท่ากัน โดยค่าความคลาดเคลื่อนทั้งสองค่าเป็นอิสระต่อกัน และจะไม่มีความสัมพันธ์กับคะแนนจริง ซึ่งผู้วิจัยศึกษาประเด็นที่สำคัญและสามารถสรุปได้ดังต่อไปนี้

การวัดคะแนนความแตกต่างอย่างง่ายหรือความแตกต่างระหว่างคะแนนดิบ (Difference scores) เป็นวิธีแรกที่ได้รับการพัฒนาขึ้นและได้รับความนิยมมาจนถึงปัจจุบัน เป็นการคำนวณคะแนนการเปลี่ยนแปลงจากผลต่างของคะแนนดิบหรือคะแนนสังเกตระหว่างการสอบครั้งแรกหรือก่อนเรียน (X) และการสอบครั้งหลังหรือหลังเรียน (Y) ซึ่งเป็นวิธีที่คำนวณง่าย สะดวกและสามารถแปลความหมายได้ค่อนข้างชัดเจน จึงเป็นวิธีที่เหมาะสมสำหรับการวัดการเปลี่ยนแปลงที่มีการวัดซ้ำสองครั้ง (นงลักษณ์ วิรัชชัย. 2542: 264; อ้างอิงจาก Rogosa. 1995. The Analysis of Change.)

อย่างไรก็ตามนักจิตวิทยา ได้แก่ ลอร์ด (Lord. 1950: 429); เบอริทเทอร์ (Embretson. 1991: 496; citing Bereiter. 1963. Problems in Measuring Change. pp. 3-20), ครอนบาค และเฟอร์บี้ (Cronbach; & Furby. 1970: 68-80), ลินน์ และสไลนด์ (Linn; & Slinde. 1977: 121-150), เมย์ และไนส์วันเดอร์ (May; & Nicewander. 1998: 882-897), ทราบบ (Traub. 1994: 127-133), วิลเลทและเซเยอร์ (Willett; & Sayer. 1994: 363-365), เปรียเลอร์ และราเวน (Prieler; & Raven. 2002: 3-4), นงลักษณ์ วิรัชชัย. (2542: 259-261) และ อรุณี อ่อนสวัสดิ์. (2537: 24-28) พบว่าวิธีนี้มีจุดอ่อนที่สำคัญสี่ประการ คือ 1) มีค่าสัมประสิทธิ์สหสัมพันธ์เชิงลบระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง หมายถึง ผู้สอบที่ได้คะแนนสอบก่อนเรียนสูงมีแนวโน้มที่จะได้คะแนนการเปลี่ยนแปลงน้อย ส่วนผู้สอบที่ได้คะแนนสอบก่อนเรียนต่ำมีแนวโน้มที่จะได้คะแนนการเปลี่ยนแปลงมาก ทำให้เกิดความไม่เป็นธรรมต่อผู้สอบที่ได้คะแนนสอบก่อนเรียนสูง 2) ความเชื่อมั่นของคะแนนการเปลี่ยนแปลงมีค่าต่ำ และต่ำกว่าความเชื่อมั่นของคะแนนสอบก่อนเรียนและหลังเรียน 3) คะแนนการเปลี่ยนแปลงที่มีผลมาจากอิทธิพลคะแนนสูงสุด (Ceiling effect) หมายถึง ผู้สอบที่ได้คะแนนสอบก่อนเรียนสูง มีโอกาสน้อยมากที่จะได้คะแนนการเปลี่ยนแปลงสูง เพราะคะแนนสอบก่อนเรียนที่ทำได้เกือบเต็ม อาจเป็นเพราะแบบทดสอบที่ใช้ในการสอบก่อนเรียนและหลังเรียนซึ่งเป็นแบบทดสอบฉบับเดิมหรือคู่ขนานกันมีจำนวนข้อสอบจำกัด จุดอ่อนประเด็นนี้อาจเป็นสาเหตุของจุดอ่อนประเด็นแรกด้วย คือ ทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงเป็นลบ (อรุณี อ่อนสวัสดิ์. 2537: 1) และ 4) การวัดคะแนนการเปลี่ยนแปลงโดยไม่คำนึงถึงคะแนนสอบก่อนเรียนหรือพื้นความรู้เดิม (Floor effect) ทำให้มีปัญหาในการแปลความหมายคะแนนการเปลี่ยนแปลง เช่น นักเรียนสองคนได้คะแนนการเปลี่ยนแปลงเท่ากัน แต่มีคะแนนสอบก่อนเรียนต่างกัน ซึ่งปัญหาเหล่านี้เป็นปัญหาของการวัดการ

เปลี่ยนแปลงด้วยวิธีการวัดคะแนนความแตกต่างอย่างง่ายหรือความแตกต่างระหว่างคะแนนดิบ ถ้าในทางปฏิบัติเราละเลย ไม่ใส่ใจต่อปัญหาเหล่านี้แล้วจะทำให้การแปลความหมายของคะแนนการเปลี่ยนแปลงผิดพลาด ไม่ถูกต้องได้

ดังนั้น ในเวลาต่อมาจึงมีนักจิตวิทยาหลายท่านได้พัฒนาวิธีการวัดการเปลี่ยนแปลงตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical test theory) เพื่อแก้ปัญหาต่างๆ ข้างต้นขึ้นอีกมากมาย ได้แก่ การวัดคะแนนเพิ่มที่แท้จริง (True gain) (Lord. 1956; & Mc Namar. 1958) การวัดคะแนนแตกต่างของส่วนที่เหลือ (Residualized differences) (Manning; & DuBois. 1958; Cronbach; & Furby. 1970) การวัดคะแนนการเปลี่ยนแปลงที่เป็นอิสระจากการวัดครั้งแรกที่เป็นฐาน (The base-free measurement of change) (Tucker; Damarin; & Messick. 1966) การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ (Relative change score) (Kanjanaawasee. 1989) การวัดคะแนนการเปลี่ยนแปลงมาตรฐาน (Standardized measure of change) ของทอร์นควิสต์ (Tornquist) (Burr; & Nesselroade. 1990) การวัดการเปลี่ยนแปลงการเรียนรู้ (อรุณี อ่อนสวัสดิ์. 2537 และ 2538) การวัดการเปลี่ยนแปลงวิธีเทียบส่วน (วินิจ เทือกทอง. 2537) การวัดการเปลี่ยนแปลงโดยใช้ตำแหน่งเปอร์เซ็นต์ไทล์ (Change in percentile rank) และ วิธีที่ใช้ร้อยละ (Percent change) (Russell. 2001)

การวัดคะแนนแตกต่างของส่วนที่เหลือ (Residualized differences) (Manning; & DuBois. 1958; Mc Namar. 1958; Cronbach; & Furby. 1970) และการวัดคะแนนการเปลี่ยนแปลงที่เป็นอิสระจากการวัดครั้งแรกที่เป็นฐาน (the base-free measurement of change) (Tucker; Damarin; & Messick. 1966) เป็นวิธีการวัดการเปลี่ยนแปลงที่มีแนวคิดคล้ายคลึงกัน คือ การนำหลักการวิเคราะห์การถดถอย (Regression) มาใช้ โดยการวัดคะแนนแตกต่างของส่วนที่เหลือของครอนบาคและเฟอร์บี้ (Cronbach; & Furby) ใช้คะแนนดิบ ซึ่งเมนิงและดুবอยส์ (Manning; & DuBois) ได้พัฒนาโดยใช้คะแนนเบี่ยงเบน ส่วนคะแนนการเปลี่ยนแปลงที่เป็นอิสระจากการวัดครั้งแรกที่เป็นฐานใช้คะแนนจริง ทั้งสองวิธีนี้สามารถกำจัดความสัมพันธ์ลวงระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงได้ แต่พบว่าเป็นวิธีที่ยากแก่การแปลความหมายเนื่องจากค่าสัมประสิทธิ์การถดถอยที่ใช้นี้ผันแปรไปตามขนาดกลุ่มตัวอย่างทำให้มีค่าความคลาดเคลื่อนรวมอยู่ด้วย ซึ่งคะแนนการเปลี่ยนแปลงที่ได้จากวิธีนี้สามารถบอกได้แต่เพียงว่าใครมีการเปลี่ยนแปลงมากกว่าหรือน้อยกว่าที่คาดหวัง ในขณะที่การวัดคะแนนเพิ่มที่แท้จริง (True gain) (Lord. 1956) พัฒนาโดยใช้หลักการวิเคราะห์การถดถอยพหุ (Multiple regression) ซึ่งทั้งสามวิธีนี้มีจุดเด่นที่คำนึงถึงอิทธิพลพื้นฐานเดิม แต่ก็ยังพบว่ามีจุดอ่อนเกี่ยวกับอิทธิพลคะแนนสูงสุด และมีความยุ่งยากในการคำนวณ

เมื่อเปรียบเทียบวิธีการวัดคะแนนแตกต่างของส่วนที่เหลือกับการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ พบว่าการวัดคะแนนแตกต่างของส่วนที่เหลือเป็นวิธีที่มีความเชื่อมั่นสูงกว่าการวัด

คะแนนความแตกต่างระหว่างคะแนนดิบ นั่นคือ เมื่อผู้สอบสองคนได้คะแนนสอบหลังเรียนเท่ากัน แต่คะแนนสอบก่อนเรียนแตกต่างกัน เมื่อนำมาคำนวณคะแนนการเปลี่ยนแปลงด้วยวิธีนี้จะได้คะแนนการเปลี่ยนแปลงที่แตกต่างกัน แต่ถ้าคะแนนสอบก่อนเรียนและหลังเรียนมีความสัมพันธ์กันสูงแล้ว ความเชื่อมั่นของคะแนนการเปลี่ยนแปลงจะต่ำเช่นเดียวกับการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ ส่วนการวัดคะแนนการเปลี่ยนแปลงที่เป็นอิสระจากการวัดครั้งแรกที่เป็นฐานมีแนวโน้มที่จะมีความคลาดเคลื่อนของคะแนนการเปลี่ยนแปลงมากกว่าการวัดคะแนนความแตกต่างระหว่างคะแนนดิบและการวัดคะแนนแตกต่างของส่วนที่เหลือ ซึ่งวิธีเหล่านี้มีวิธีคำนวณที่ยุ่งยากและซับซ้อน ในปัจจุบันจึงไม่นิยมนำมาใช้

ต่อมาในปี ค.ศ. 1985 ทอร์นควิสท์ (Tornquist. 1985) (นงลักษณ์ วิรัชชัย. 2542. 263; อ้างอิงจาก Burr; & Nesselroade. 1990. Statistical Methods in Longitudinal Research Volume 1: Principles and Structural Change) ได้พัฒนาการวัดคะแนนการเปลี่ยนแปลงมาตรฐาน (Standardized measure of change) ขึ้น โดยกล่าวว่าเป็นวิธีการวัดการเปลี่ยนแปลงที่คะแนนมีการแจกแจงแบบปกติ และเป็นวิธีที่เหมาะสมสำหรับการวัดซ้ำสองครั้งที่ใช้เครื่องมือวัดที่มีมาตรวัดที่แตกต่างกัน แต่พบว่ามีข้อจำกัดเกี่ยวกับความแปรปรวนของคะแนนการวัดทั้งสองครั้งต้องเท่ากัน หากไม่เท่ากันแล้ว ค่าคะแนนการเปลี่ยนแปลงจะมีความลำเอียง โดยมีค่าสูงเกินกว่าที่ควรจะเป็น ต่อจากนั้น ศิริชัย กาญจนวาสี ได้พัฒนาวิธีการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ (Relative change score) (ศิริชัย กาญจนวาสี. 2538. 144-149. อ้างอิงจาก Kanjanawasee. 1989. Alternative Strategies for Policy Analysis: An Assessment of School Effects on Student's Cognitive and Affective Mathematics Outcomes in Lower Secondary School in Thailand. pp. 50-51) โดยคำนวณคะแนนการเปลี่ยนแปลงสัมพัทธ์จากอัตราส่วนของคะแนนเพิ่มกับศักยภาพที่แต่ละคนควรได้ ซึ่งเป็นวิธีที่มีจุดเด่นที่สามารถลดอิทธิพลคะแนนสูงสุดและเป็นการคำนึงถึงความสามารถก่อนเรียน

รัสเซล (Russell. 2001: Online) ได้กล่าวถึงการวัดการเปลี่ยนแปลงที่นิยมใช้ในการประเมินสถานศึกษาของ Council of Chief State School Officers (1998) คือ การวัดการเปลี่ยนแปลงโดยใช้ตำแหน่งเปอร์เซ็นต์ไทล์ (Change in percentile rank) และ วิธีที่ใช้ร้อยละ (Percent change) โดยการวัดการเปลี่ยนแปลงที่ใช้ตำแหน่งเปอร์เซ็นต์ไทล์ เป็นการคำนวณคะแนนการเปลี่ยนแปลงจากคะแนนความแตกต่างระหว่างตำแหน่งเปอร์เซ็นต์ไทล์ในการสอบก่อนเรียนกับหลังเรียน เป็นวิธีที่คำนวณง่าย แต่คะแนนการเปลี่ยนแปลงที่ได้ขึ้นอยู่กับตำแหน่งเปอร์เซ็นต์ไทล์ของกลุ่ม จึงไม่สามารถบอกปริมาณการเปลี่ยนแปลงที่แท้จริงได้ บอกได้แต่เพียงว่านักเรียนคนนี้ได้คะแนนสอบหลังเรียนมากกว่าคนอื่นเปลี่ยนแปลงไปจากการสอบก่อนเรียนกี่คน ส่วนการวัดการเปลี่ยนแปลงโดยใช้ค่าร้อยละนั้น เป็น

การคำนวณคะแนนการเปลี่ยนแปลงจากผลต่างระหว่างร้อยละในการสอบก่อนเรียนกับหลังเรียน ทำให้เกิดคะแนนการเปลี่ยนแปลงของกลุ่มผู้สอบที่มีคะแนนสอบก่อนเรียนต่ำไม่ถูกต้อง คือ ผู้สอบกลุ่มนี้จะมีคะแนนการเปลี่ยนแปลงสูงกว่าผู้สอบกลุ่มที่มีผลการสอบก่อนเรียนสูง ในขณะที่ได้คะแนนผลต่างระหว่างคะแนนดิบเท่ากัน และเป็นวิธีที่ยากแก่การแปลความหมาย

สำหรับในประเทศไทย อรุณี อ่อนสวัสดิ์ (2537: 70-77; 2538: 1-15) ได้พัฒนาวิธีการวัดการเปลี่ยนแปลงการเรียนรู้โดยอาศัยแนวคิดตามทฤษฎีการเรียนรู้ของบลูมและอิทธิพลคะแนนสูงสุด เป็นการพัฒนาวิธีการของลอว์ดกับแนวคิดทฤษฎีการเรียนรู้ของบลูม ได้วิธีวัดการเปลี่ยนแปลงการเรียนรู้จำนวนแปดวิธี ที่คำนวณโดยใช้คะแนนจริงสี่วิธีและคะแนนดิบสี่วิธี เมื่อนำทั้งแปดวิธีที่พัฒนาขึ้นไปเปรียบเทียบคุณภาพกับวิธีอื่นๆ พบว่าวิธีที่พัฒนาขึ้นทั้งหมดมีค่าความเชื่อมั่นและค่าความคลาดเคลื่อนไม่แตกต่างจากวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ และวิธีการวัดคะแนนเพิ่มที่แท้จริงของลอว์ด แต่วิธีของอรุณีมีการคำนวณที่ค่อนข้างยุ่งยากมากกว่า และในปีเดียวกันนี้ วินิจ เทือกทอง (2537: 20 -25) ได้พัฒนาวิธีการวัดการเปลี่ยนแปลงทั้งหมดหกวิธี โดยการประยุกต์วิธีดั้งเดิม วิธีเศษเหลือ และวิธีของลอว์ด ดังนี้ 1) วิธีของศิริชัย เป็นการนำคะแนนเพิ่มจากทั้งสามวิธีมาเทียบกับศักยภาพของผู้เข้าสอบตามวิธีของศิริชัย และ 2) วิธีเทียบส่วนร้อยละ เป็นการนำคะแนนเพิ่มจากทั้งสามวิธีมาเทียบกับคะแนนสอบครั้งแรกของแต่ละคนที่ใช้เป็นฐาน แล้วนำทั้งหกวิธีที่พัฒนาขึ้นไปเปรียบเทียบความคลาดเคลื่อนในการวัดกับวิธีดั้งเดิม วิธีเศษเหลือ วิธีของลอว์ด พบว่าในกรณีที่ไม่นับถึงอิทธิพลคะแนนสูงสุดนั้น วิธีดั้งเดิมคือ วิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ และวิธีของลอว์ด มีค่าความคลาดเคลื่อนในการวัดน้อยที่สุด ส่วนในกรณีที่ไม่นับถึงอิทธิพลคะแนนสูงสุดแล้ว วิธีเศษเหลือเทียบกับศักยภาพของผู้สอบมีค่าความคลาดเคลื่อนน้อยที่สุด ในปีต่อมา อรุณี อ่อนสวัสดิ์ (2538) ได้พัฒนาวิธีการวัดการเปลี่ยนแปลงเพิ่มเติมที่ไม่นับถึงอิทธิพลคะแนนสูงสุด โดยใช้สัมประสิทธิ์ถดถอยอย่างง่ายและสัมประสิทธิ์ถดถอยพหุที่แตกต่างกันจำนวนสี่แนวคิด รวมแปดวิธี แต่ยังไม่มีการตรวจสอบคุณภาพ

จากพัฒนาการของวิธีการวัดการเปลี่ยนแปลงตามแนวคิดการวัดแบบมาตรฐานเดิม (Classical measurement) ซึ่งเป็นการวัดซ้ำสองครั้งนั้น พบจุดอ่อนของการใช้วิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบที่ใช้อธิบายการเปลี่ยนแปลงการเรียนรู้ ทำให้นักวิจัยวิทยาได้แสวงหาวิธีการวัดการเปลี่ยนแปลงในแนวทางอื่นเพื่อแก้ปัญหาหรือข้อจำกัดต่างๆ อีกมากมายหลายวิธีตามแนวคิดการวัดการเปลี่ยนแปลงแบบมาตรฐานเดิมดังที่กล่าวมาแล้วข้างต้น แต่ก็ไม่สามารถให้ข้อสรุปที่ชัดเจนได้ว่าวิธีการวัดการเปลี่ยนแปลงที่พัฒนาขึ้นมาอย่างต่อเนื่องนั้นมีคุณภาพสูงกว่าวิธีการวัดความแตกต่างระหว่างคะแนนดิบ

นอกจากนี้นักวิจัยวิทยายังได้แสวงหาวิธีการวัดการเปลี่ยนแปลงการเรียนรู้อีกแนวคิดหนึ่งคือ วิธีการแนวใหม่ (Modern methods) ซึ่งเป็นการวัดการเปลี่ยนแปลงที่นักวิจัยวิทยาเชื่อว่าการวัดการเปลี่ยนแปลงที่

มีการเก็บรวบรวมข้อมูลหลายครั้งสามารถอธิบายการเปลี่ยนแปลงความสามารถหรือคุณลักษณะที่ต้องการวัดของบุคคลได้ชัดเจนกว่าการเก็บรวบรวมข้อมูลสองครั้ง เนื่องจากผลของการวัดแสดงให้เห็นถึงแบบแผนหรือแนวโน้มของการเปลี่ยนแปลงที่อาจจะไม่เป็นเส้นตรงเหมือนการวัดสองครั้ง ดังเช่น โมเดลโค้งพัฒนาการ (Growth curves model) โมเดลพหุระดับ (Multilevel model) และ ทฤษฎีการตอบข้อสอบ (Item response theory) เป็นต้น

การวัดการเปลี่ยนแปลงแนวโน้มในรูปของโมเดลโค้งพัฒนาการ (Growth curves model) เป็นโมเดลในทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical test theory) โดยได้นำโมเดลการวัด (Measurement model) ในรูปสมการโครงสร้างเชิงเส้น (Linear structural equation) มาใช้ (Tisak; & Meridith. 1987) เป็นโมเดลที่นำความคลาดเคลื่อนในการวัด (Error of measurement) มาร่วมวิเคราะห์ด้วย ซึ่งส่งผลต่อการประมาณค่าพารามิเตอร์ได้ถูกต้องมากยิ่งขึ้น และเป็นโมเดลที่สามารถทดสอบความสัมพันธ์ของตัวแปรทั้งหมดโดยยอมให้ความคลาดเคลื่อนในการวัดเป็นอิสระต่อกันหรือมีความสัมพันธ์กันได้ ตลอดจนเป็นโมเดลที่สามารถทดสอบโครงสร้างขององค์ประกอบเดียวกันที่ถูกวัดในช่วงเวลาที่ต่างกันได้ ต่อมา เรย์คอฟ (Raykov. 1994) ได้พัฒนาโมเดลและเสนอแนวคิดเกี่ยวกับคะแนนสังเกตของการวัดบุคคลในครั้งหนึ่งนั้นเป็นคะแนนประกอบ (Composite score) ที่ประกอบด้วยสามองค์ประกอบ คือ ความสามารถเริ่มต้น (Initial factor) การเปลี่ยนแปลงทั้งหมด (Overall change factor) และ ความคลาดเคลื่อนในการวัด (Error of measurement factor) และนักวิจัยวิทยาได้มีการพัฒนาโมเดลอย่างต่อเนื่อง จนกระทั่งได้โมเดลที่มีความยืดหยุ่น สามารถปรับโมเดลให้สอดคล้องกับโค้งพัฒนาการรูปแบบต่างๆ ได้สี่แบบ คือ โมเดลพัฒนาการเชิงเส้นโค้ง โมเดลพัฒนาการที่กำหนดพารามิเตอร์คงที่ โมเดลพัฒนาการเชิงเส้นตรงและโมเดลพัฒนาการที่ไม่มีความชัน ซึ่งโมเดลดังกล่าวเป็นที่นิยมใช้กัน แต่ไม่สามารถประมาณค่าอัตราพัฒนาการรายบุคคลได้ จะต้องนำไปจัดกระทำต่อจึงจะสามารถประมาณค่าได้

ส่วนโมเดลพหุระดับ (Multilevel model) หรือที่เรียกชื่อแตกต่างกัน เช่น โมเดลเชิงเส้นระดับลดหลั่น (Hierarchical linear model; HLM) หรือ โมเดลสมการโครงสร้างพหุระดับ (Multi-level structural equation model) หรือ โมเดลการถดถอยสัมประสิทธิ์แบบสุ่ม (Random-coefficient regression model) เป็นต้น (Bry; & Raudenbush. 1986, 2002; Little, T.D.; Schnabel, K.U.; & Baumert, J. 2000) เป็นโมเดลที่ใช้กับข้อมูลที่มีลักษณะหลายระดับและซ้อนกันเป็นระดับลดหลั่น โดยใช้เทคนิคเอชแอลเอ็มซึ่งเป็นเทคนิคที่ได้รับการยอมรับว่าเป็นเทคนิคที่เหมาะสมที่สุดในการวิเคราะห์ข้อมูลพหุระดับ เป็นเทคนิคที่พัฒนามาจากสถิติหลายประเภท คือ เทคนิคการวิเคราะห์ความแปรปรวน โมเดลผสม (Mixed-model ANOVA) สัมประสิทธิ์การถดถอยแบบสุ่ม (Regression with random coefficients) โมเดลส่วนประกอบความแปรปรวนร่วม (Covariance component models) และการ

ประมาณค่าในโมเดลเชิงเส้นด้วยวิธีของเบส์ (Bayesian estimation for linear models) จึงเป็นเทคนิคที่ให้ผลการวิเคราะห์ที่มีความคงเส้นคงวาและน่าเชื่อถือกว่าเทคนิคกำลังสองน้อยที่สุดแบบสองสมการ (Bry; & Raudenbush. 1986, 2002; ศิริชัย กาญจนวาสี. 2548; วราภรณ์ วิหคโต. 2536)

ต่อมานักวิจัยวิทยาได้ประยุกต์โมเดลพหุระดับมาใช้ในการวัดการเปลี่ยนแปลง ซึ่งเป็นโมเดลที่เหมาะสมที่สุดโมเดลหนึ่งในการวัดปริมาณการเปลี่ยนแปลงเป็นรายบุคคลและกลุ่ม มีกระบวนการวิเคราะห์ที่คำนึงถึงหน่วยในการวิเคราะห์ที่แตกต่างกัน ทำให้การสรุปผลระหว่างระดับมีความคลาดเคลื่อนต่ำ ส่งผลให้การสรุปผลได้ถูกต้องมากยิ่งขึ้น และยังเป็นโมเดลที่มีผู้นิยมใช้กันมากในการวัดการเปลี่ยนแปลง ดังตัวอย่างงานวิจัยของ ไบรย์และเรเดนบิช (Bry; & Raudenbush. 1987); เบอริชินัล และแอปเพิลบาร์ม (Burchinal; & Appelbaum. 1991); ชิม (Shim. 1995); ไชว เบนเลอร์ และเพนท์ (Chou; Bentler; & Pentz.1998); มา และมา (Ma; & Ma. 2004); แม็คโคช (McCoach.; et al. 2006); วีระศักดิ์ คำล้าน (2540); สุภารัตน์ เรือจันทิก (2542); และ สมถวิล วิจิตรรรรณา (2543) เป็นต้น จากผลการวิจัยดังกล่าว พบว่าผลการวัดปริมาณการเปลี่ยนแปลงรายบุคคลที่ได้มีความถูกต้องสูง เชื่อถือได้ มีความคลาดเคลื่อนต่ำ และเป็นที่ยอมรับใช้ ผู้วิจัยจึงจะใช้โมเดลพหุระดับในการประมาณค่าเป็นเกณฑ์สำหรับการศึกษาการวัดการเปลี่ยนแปลงครั้งนี้

จากการศึกษาวิธีการวัดการเปลี่ยนแปลงแนวใหม่ดังกล่าวข้างต้นนี้พบว่า เป็นวิธีวัดที่มีโมเดลที่ยุ่งยากซับซ้อน และต้องเก็บรวบรวมข้อมูลอย่างน้อยสามครั้ง จึงจะให้ค่าที่เป็นปริมาณการเปลี่ยนแปลงที่เชื่อถือได้ (สมถวิล วิจิตรรรรณา. 2543: 13; อ้างอิงจาก Raykov. 1994. Applied Psychological Measurement. pp. 63-77) ซึ่งเป็นการเพิ่มภาระงานแก่ผู้ปฏิบัติในสถานศึกษาเป็นอันมาก ยากแก่การนำมาใช้ปฏิบัติงาน และในขณะเดียวกันตั้งแต่ปี 1976 นักวิจัยได้นำทฤษฎีการตอบข้อสอบมาใช้ในการวัดการเปลี่ยนแปลง

ทฤษฎีการตอบข้อสอบ (Item response theory) เป็นการวัดที่ข้อสอบแต่ละข้อแทนด้วยฟังก์ชันเพิ่มขึ้นทางเดียว (Monotonically increasing function) เป็นวิธีที่ประมาณค่าพารามิเตอร์โดยใช้แบบแผนคำตอบของผู้สอบ ซึ่งค่าพารามิเตอร์มีคุณสมบัติของการไม่แปรเปลี่ยน (Invariance) โดยค่าพารามิเตอร์ความยากของข้อสอบจะไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ จึงมีค่าคงที่ในการสอบซ้ำๆ ดังนั้น ทฤษฎีการตอบข้อสอบจึงสามารถนำมาใช้วัดการเปลี่ยนแปลง เนื่องจากการวัดการเปลี่ยนแปลงเกิดจากความแตกต่างของการประมาณค่าความสามารถของคนที่ทำข้อสอบที่มีคุณลักษณะไม่แปรเปลี่ยนไปตามการแจกแจงความสามารถของผู้สอบ ประกอบกับคุณลักษณะเด่นอีกประการหนึ่งของทฤษฎีการตอบข้อสอบคือการกำหนดค่าความคลาดเคลื่อนมาตรฐานในการวัดของผู้สอบแต่ละคนไม่จำเป็นต้องเท่ากัน ส่งผลให้การประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของคนมีความถูกต้องมากยิ่งขึ้น การวัดการเปลี่ยนแปลงตามทฤษฎีนี้มีทั้งที่เก็บรวบรวมข้อมูลสองครั้งและมากกว่าสองครั้ง ได้แก่

โมเดลคุณลักษณะแฝงโลจิสติกแบบเส้นตรงที่ผ่อนคลายข้อตั้ง (Linear logistic latent trait model with relaxed assumptions; LLRA) ของ ฟิชเชอร์ (Fischer. 1976) การประยุกต์ใช้โค้งการตอบข้อสอบของโค้งพัฒนาการ (Growth model) ของบอค (Bock. 1976) โมเดลราซพหุมิติ (Multidimensional rasch models) ของแอนเดอร์สัน (Anderson. 1985) และ เอ็มเบรตัน (Embretson. 1991) การเปลี่ยนแปลงความสามารถแฝง เสนอโดยเมย์ และไนส์วันเดอร์ (May; & Nicewander. 1998) การวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบของอวยพร เรื่องตระกูล (2544) และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ (Heidenberg. 2000) ซึ่งผู้วิจัยศึกษาประเด็นที่สำคัญและสามารถสรุปได้ดังต่อไปนี้

ฟิชเชอร์ (Fischer. 1976) ได้พัฒนาโมเดลคุณลักษณะแฝงโลจิสติกแบบเส้นตรงที่ผ่อนคลายข้อตั้ง (Linear logistic latent trait model with relaxed assumptions; LLRA) และได้ให้แนวคิดเกี่ยวกับการวัดการเปลี่ยนแปลงว่า เมื่อเวลาเปลี่ยนไป การเปลี่ยนแปลงการเรียนรู้ของบุคคลจะขึ้นอยู่กับ การเปลี่ยนแปลงค่าพารามิเตอร์ของข้อสอบมากกว่าการเปลี่ยนแปลงค่าพารามิเตอร์ความสามารถของผู้สอบ จึงได้พัฒนาโมเดลนี้ขึ้นมาเพื่อใช้กับการวัดสองครั้ง ที่มีข้อมูลเป็นแบบสองค่า (Dichotomous) และสามารถใช้ได้กับข้อมูลที่เป็นทั้งมิติเดียวและหลายมิติ ซึ่งขัดแย้งกับความคิดของบอค (Embretson. 1991: 495 - 515; citing Bock. 1976. Advances in psychological and educational measurement. pp. 75-96) ที่ได้เสนอความคิดเกี่ยวกับการวัดการเปลี่ยนแปลงที่คำนวณมาจากผลต่างระหว่างค่าพารามิเตอร์ความสามารถของผู้สอบ เรียกว่า การประยุกต์ใช้โค้งการตอบข้อสอบของโค้งพัฒนาการ (Growth model) โดยการหาเส้นโค้งของโมเดลทฤษฎีการตอบข้อสอบที่สอดคล้องกลมกลืนกับเส้นโค้งพัฒนาการ แต่พบว่าค่าความเชื่อมั่นของการวัดแต่ละครั้งมีค่าต่ำมาก ต่อมา แอนเดอร์สัน (Anderson. 1985: 3-16) และ เอ็มเบรตัน (Embretson. 1991: 495-515) ได้พัฒนาโมเดลราซพหุมิติ (Multidimensional rasch models) ซึ่งเป็นโมเดลที่ใช้กับข้อมูลแบบสองค่า (Dichotomous) โดยวัดได้มากกว่าสองครั้ง มีข้อตั้งเบื้องต้นว่า ในเวลาที่เปลี่ยนไปความสามารถของผู้สอบต้องเพิ่มขึ้นจากเดิมเสมอ และยังได้จำแนกความสามารถที่ประมาณได้ออกเป็นสองลักษณะ คือ ความสามารถที่เป็นผล (Effective ability) และความสามารถในการเรียนรู้หรือความสามารถที่เพิ่มขึ้น (Learning ability or Modifiability) โดยความสามารถที่เป็นผลนั้นเป็นศักยภาพแฝงของผู้สอบในการสอบแต่ละครั้ง ซึ่งความสามารถที่เป็นผลในการวัดครั้งแรกจะเป็นความสามารถเริ่มต้น ส่วนความสามารถที่เป็นผลในครั้งต่อไปจะประกอบไปด้วยความสามารถเริ่มต้นและความสามารถในการเรียนรู้หรือความสามารถที่เพิ่มขึ้น ซึ่งอาจจะเป็นหนึ่งค่าหรือมากกว่า และในเวลาต่อมาได้มีการพัฒนาโมเดลสำหรับการวัดที่มากกว่าสองครั้ง

แต่เนื่องจากโมเดลการวัดตามทฤษฎีการตอบข้อสอบที่กล่าวมาแล้วนั้นค่อนข้างยุ่งยากซับซ้อนยากแก่การเข้าใจ ทำให้ไม่ได้รับความนิยมเท่าที่ควร ต่อมาเทคโนโลยีคอมพิวเตอร์มีความเจริญก้าวหน้า

มากยิ่งขึ้น นักวิจัยทางจิตวิทยาจึงได้หันกลับมาพิจารณาวิธีการวัดการเปลี่ยนแปลงโดยประยุกต์ทฤษฎีการตอบข้อสอบมาใช้ในการเก็บรวบรวมข้อมูลสองครั้ง ดังเช่น เมย์และไนส์วันเดอร์ (May; & Nicewander. 1998: 882-897) ได้ศึกษาการประมาณค่าการเปลี่ยนแปลงความสามารถแฝง (Latent change) ที่คำนวณมาจากความแตกต่างระหว่างความสามารถก่อนเรียนและหลังเรียนที่ประมาณค่ามาจากการใช้โมเดลโลจิสติกแบบสามพารามิเตอร์ในการประมาณค่าพารามิเตอร์ด้วยวิธี EAP เป็นการวิจัยเพื่อเปรียบเทียบผลจากการวัดคะแนนการเปลี่ยนแปลงความสามารถแฝงกับการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ พบว่าการวัดคะแนนการเปลี่ยนแปลงความสามารถแฝงนี้สามารถลดอิทธิพลของปฏิสัมพันธ์ระหว่างความยากของข้อสอบกับระดับความสามารถเริ่มต้นของผู้สอบที่มีต่อคะแนนการเปลี่ยนแปลงได้ โดยการนำค่าพารามิเตอร์ข้อสอบมาร่วมในการประมาณค่าความสามารถ เป็นวิธีที่ง่ายในการคำนวณ และสามารถเปรียบเทียบคะแนนการเปลี่ยนแปลงระหว่างคน และวิชาที่แตกต่างกันได้ แต่มีข้อจำกัด คือ ต้องใช้กับกลุ่มตัวอย่างที่มีขนาดใหญ่ มีความสัมพันธ์เป็นลบระหว่างความสามารถเริ่มต้นกับคะแนนการเปลี่ยนแปลง มีอิทธิพลคะแนนสูงสุด และมีอิทธิพลพื้นความรู้เดิม คือ ไม่คำนึงถึงคะแนนสอบก่อนเรียน ทำให้การแปลความคะแนนการเปลี่ยนแปลงจึงไม่ถูกต้องนัก

ส่วน ไฮเดนเบิร์ก (Heidenberg. 2000) ได้ประยุกต์คะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ (IRT Domain scores) มาศึกษาการลดค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง ด้วยการจำลองข้อมูลจากการสอบสองครั้ง ภายใต้เงื่อนไข ความยาวของแบบทดสอบ ขนาดกลุ่มตัวอย่าง ปริมาณของคะแนนเปลี่ยนแปลงจริง (True change) ลักษณะการแจกแจงของคะแนนเปลี่ยนแปลงจริง และจำนวนพารามิเตอร์ของโมเดลโลจิสติก ซึ่งพบว่า คะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบสามารถลดค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าวได้ทุกเงื่อนไข

สำหรับในประเทศไทย อวยพร เรืองตระกูล. (2544: 82-85) ได้พัฒนาวิธีการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบโดยประยุกต์ใช้วิธีคะแนนการเปลี่ยนแปลงสัมพัทธ์ตามแนวคิดของศิริชัย กาญจนวาสี (2532) และวินิจ เทือกทอง (2537) มาใช้ในการวัดการเปลี่ยนแปลงสำหรับการวัดสองครั้ง ซึ่งเป็นวิธีที่คำนวณได้ง่าย โดยวัดการเปลี่ยนแปลงจากความสามารถที่แท้จริงสัมพัทธ์สองวิธี คือ 1) เมื่อเทียบกับศักยภาพพัฒนาการ ซึ่งหมายถึง ผลต่างของความสามารถสูงสุดกับความสามารถที่แท้จริงก่อนเรียน และ 2) เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน ซึ่งหมายถึง ผลรวมของความสามารถสูงสุดกับความสามารถที่แท้จริงก่อนเรียน แล้วนำไปเปรียบเทียบกับวิธีอื่น พบว่าการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ และการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ เป็นวิธีที่มีความเชื่อมั่นสูง และความคลาดเคลื่อนต่ำ จึงมีความเหมาะสมสำหรับการนำไปปฏิบัติในระดับสถานศึกษา

จากพัฒนาการวิธีการวัดการเปลี่ยนแปลงในทฤษฎีการตอบข้อสอบ ตามวิธีการแนวใหม่ (Modern methods) ดังที่กล่าวมาแล้วข้างต้นพบว่ามีสองแนวคิดใหญ่ๆ คือ แนวคิดแรก เชื่อว่าคะแนนการเปลี่ยนแปลงเกิดจากการเปลี่ยนแปลงค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าได้ในเวลาที่แตกต่างกัน ในขณะที่ค่าพารามิเตอร์ความสามารถของผู้สอบคงที่ ได้แก่ วิธีของฟิชเชอร์ (Fischer) ส่วนแนวคิดที่สอง เชื่อว่าคะแนนการเปลี่ยนแปลงเกิดจากการเปลี่ยนแปลงค่าพารามิเตอร์ความสามารถของผู้สอบในเวลาที่แตกต่างกัน ในขณะที่มีค่าพารามิเตอร์ของข้อสอบคงที่ ได้แก่ วิธีของบอค (Bock) แอนเดอร์สันและเอ็มเบรตัน (Anderson; & Embretson) เมย์และไนส์วันเดอร์ (May; & Nicewander) และ ไฮเดนเบิร์ก (Heidenberg)

ในการศึกษาครั้งนี้เป็นการประยุกต์ทฤษฎีการตอบข้อสอบมาใช้ในการวัดการเปลี่ยนแปลงตามแนวคิดที่สอง คือ คะแนนการเปลี่ยนแปลงเกิดจากการเปลี่ยนแปลงค่าพารามิเตอร์ความสามารถของผู้สอบในเวลาที่แตกต่างกัน ในขณะที่มีค่าพารามิเตอร์ของข้อสอบคงที่ ด้วยการนำคะแนนมวลความรู้ที่ประมาณค่าตามทฤษฎีการตอบข้อสอบ (IRT Domain scores) มาใช้ในการวัดการเปลี่ยนแปลง

คะแนนมวลความรู้ (Domain scores; π) เป็นคะแนนตัวแทนมวลความรู้ที่ได้จากการสอบด้วยแบบทดสอบที่ประกอบไปด้วยข้อสอบที่มีการวัดจุดประสงค์การเรียนรู้เดียวกันที่นิยามไว้อย่างชัดเจน ซึ่งเดิมนักวิจัยศึกษาใช้คะแนนมวลความรู้ในการประมาณค่าความสามารถจริงของผู้สอบ และคำนวณหาคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ ในทฤษฎีการทดสอบแบบมาตรฐานเดิม ต่อมานักวิจัยศึกษาได้นำหลักการของทฤษฎีการตอบข้อสอบมาใช้วัดคะแนนมวลความรู้ ซึ่ง แฮมเบิลตัน สวามินาธาน และโรเจอร์ (Hambleton; Swaminathan; & Rogers. 1991: 85) ได้ให้นิยามว่าเป็นคะแนนที่คาดหวัง (Expected scores) จากข้อสอบทุกข้อในมวลความรู้ ซึ่งเป็นคะแนนแปลงของคะแนนความสามารถ θ ไปเป็นคะแนนจริง (True score) ที่ไม่เป็นเส้นตรง (Nonlinear) เมื่อโมเดลมีความสอดคล้องกลมกลืนกับข้อมูล คะแนนจริงเป็นผลรวมของโค้งลักษณะข้อสอบ (Item characteristic curve) ณ ความสามารถ θ ซึ่งก็คือ โค้งลักษณะแบบทดสอบ (Test characteristic curve) นั่นเอง คะแนนจริงเป็นมาตรเดียวกันกับคะแนนจำนวนข้อที่ตอบถูก ดังนั้น คะแนนมวลความรู้จึงเกิดจากค่าสัดส่วนของคะแนนจริงที่ตอบถูก (True proportion correct ; τ) หารด้วยจำนวนข้อคำถาม n นั่นคือ

$$\pi = \frac{1}{n} \sum_{j=1}^n P_j(\theta) = \frac{\tau}{n} \text{ ที่มีค่าอยู่ระหว่างศูนย์ถึงหนึ่งหรืออยู่ในเทอมของร้อยละระหว่างร้อยละศูนย์ถึงร้อยละ}$$

นักวิจัยศึกษาส่วนใหญ่ได้นำคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบไปใช้ในการศึกษาการรับรู้ของผู้เรียนโดยการใช้คะแนนมวลความรู้ในการกำหนดคะแนนจุดตัดของแบบทดสอบ เช่น แฮมเบิลตัน และกรูเจเตอร์ (Hambleton; & De Gruijter. 1983: 362); เคน (Kane. 1987); สคัลซ์โคเลน และไนส์วันเดอร์ (Schulz; Kolen; & Nicewander. 1999); และ บอค ทิสเซน และซิมโมสกี (Bock; Thissen; & Zimowski. 1997: 197-209) ซึ่ง บอค ทิสเซน และซิมโมสกี ได้ให้ข้อเสนอแนะใน

การนำคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ (IRT Domain scores) ไปใช้ในการวัดการเปลี่ยนแปลง เพื่อลดหรือกำจัดการบิดเบือนเกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงได้ ต่อมา ไฮเดนเบิร์ก (Heidenberg, 2000: Abstract) ได้ศึกษาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงของการวัดการเปลี่ยนแปลงด้วยการใช้คะแนนมวลความรู้ที่ประมาณค่ามาจากทฤษฎีการตอบข้อสอบ โดยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล ภายใต้เงื่อนไขต่างๆ ได้แก่ ความยาวของแบบทดสอบ ปริมาณของคะแนนเพิ่มจริง ขนาดกลุ่มตัวอย่าง และลักษณะการแจกแจงค่าความยากของแบบทดสอบก่อนเรียนและหลังเรียน พบว่า คะแนนการเปลี่ยนแปลงกับคะแนนสอบก่อนเรียนมีค่าสัมประสิทธิ์สหสัมพันธ์อย่างไม่น่าสำคัญ ในทุกเงื่อนไข แต่ไม่ปรากฏผลการศึกษาความถูกต้อง แม่นยำของคะแนนการเปลี่ยนแปลงด้วยผลต่างระหว่างคะแนนมวลความรู้ ซึ่งคะแนนการเปลี่ยนแปลงที่ได้นั้นอาจมีความแตกต่างหรือไม่แตกต่างจากคะแนนการเปลี่ยนแปลงที่คำนวณด้วยวิธีอื่นๆ จึงเป็นประเด็นที่ผู้วิจัยสนใจ

การวัดคะแนนการเปลี่ยนแปลงด้วยการนำคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบมาใช้ของไฮเดนเบิร์ก เป็นวิธีที่สามารถลดปัญหาค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงได้ (Bock; Thissen; & Zimowski, 1997) และยังสามารถแก้ปัญหาค่าความสามารถที่มีค่าเป็นลบ (Heidenberg, 2000) ช่วยทำให้การแปลความหมายของคะแนนการเปลี่ยนแปลงหรือการพัฒนาการเรียนรู้ได้ง่าย ตรงไปตรงมาและชัดเจนขึ้น แต่ก็ยังพบว่าเป็นวิธีที่ไม่คำนึงถึงอิทธิพลคะแนนสูงสุดและพื้นความรู้เดิม ดังนั้น ผู้วิจัยจึงได้ปรับขยายวิธีการวัดการเปลี่ยนแปลงที่ใช้คะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบของไฮเดนเบิร์ก (Heidenberg) โดยใช้วิธีการประมาณค่าพารามิเตอร์ด้วย EAP (Expected a posteriori) ซึ่งเป็นการประมาณค่าด้วยวิธีของเบย์ (Bayesian estimation) โดยการนำค่าการแจกแจงเริ่มต้นของพารามิเตอร์ (Prior Distribution; $f(\theta, b, a, c)$) มาร่วมวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ด้วยการทำให้ฟังก์ชันการแจกแจงภายหลังของพารามิเตอร์ $f(\theta, b, a, c/u)$ มีค่าสูงสุด สุชาดา กิระนันท์ (2545: 90) ได้กล่าวว่าเมื่อจำนวนผู้เข้าสอบและจำนวนข้อสอบมีขนาดเล็กแล้ว การประมาณค่าพารามิเตอร์ด้วยวิธีเบย์ จะช่วยลดความซับซ้อนของกระบวนการประมาณค่าความสามารถของบุคคล ซึ่งทำให้การประมาณค่าพารามิเตอร์ได้ถูกต้องมากยิ่งขึ้น ดังที่เมย์และไนส์วันเดอร์ (May; & Nicewander) ใช้ในการศึกษาวิจัยคะแนนการเปลี่ยนแปลง นอกจากนี้ยังได้นำแนวคิดเกี่ยวกับการแก้ปัญหาอิทธิพลคะแนนสูงสุดและพื้นความรู้เดิม ของ ศิริชัย กาญจนวาสี มาใช้ในการศึกษาครั้งนี้โดยนำผลต่างของคะแนนมวลความรู้มาเทียบกับศักยภาพของผู้สอบ ซึ่งจะเรียกวิธีการดังกล่าวนี้ว่า การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์

จากการศึกษาเอกสารและงานวิจัยพบว่า มีการวัดการเปลี่ยนแปลงที่น่าสนใจหาวิธี สามารถจัดเป็นสองกลุ่มตามทฤษฎีการทดสอบ คือ กลุ่มที่หนึ่งเป็นกลุ่มวิธีตามทฤษฎีการทดสอบแบบมาตรฐานเดิม

สองวิธี ได้แก่ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ และการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ ซึ่งการคำนวณด้วยวิธีเหล่านี้เป็นวิธีที่ไม่ยุ่งยากซับซ้อน ง่ายต่อการทำความเข้าใจและการแปลผล เป็นวิธีที่ประมาณค่าคะแนนการเปลี่ยนแปลงที่มีความเชื่อมั่นสูงและมีความคลาดเคลื่อนต่ำ ทั้งยังใช้กลุ่มตัวอย่างขนาดเล็ก ทำให้เสียค่าใช้จ่ายน้อย ส่วนกลุ่มที่สองเป็นกลุ่มวิธีตามทฤษฎีการตอบข้อสอบสามวิธี ได้แก่ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ เป็นวิธีที่น่าค่าพารามิเตอร์ข้อสอบมาช่วยในการประมาณค่าความสามารถ เป็นวิธีที่ง่ายในการคำนวณเช่นกัน แต่ต้องใช้กับกลุ่มตัวอย่างขนาดใหญ่ ซึ่งมีนักวิจัยพยายามหาทางพยายามเปรียบเทียบผลของการวัดการเปลี่ยนแปลงภายใต้สถานการณ์ต่างๆ ทั้งการจำลองข้อมูล (Simulation) และใช้ข้อมูลเชิงประจักษ์ (Empirical data) เพื่อหาข้อสรุปว่าวิธีใดสามารถวัดคะแนนการเปลี่ยนแปลงได้แม่นยำมากกว่า โดยความแม่นยำสามารถพิจารณาได้จากค่าสัมประสิทธิ์สหสัมพันธ์และค่าความคลาดเคลื่อนของคะแนนการเปลี่ยนแปลงระหว่างวิธี

จากเหตุผลดังกล่าวข้างต้น ผู้วิจัยจึงมีความสนใจที่จะศึกษาผลของการวัดคะแนนการเปลี่ยนแปลงห้าวิธีแบบวัดซ้ำสองครั้ง ภายใต้เงื่อนไขเกี่ยวกับความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน เพื่อตรวจสอบว่าผลของการวัดการเปลี่ยนแปลงในแต่ละวิธีแตกต่างกันหรือไม่ วิธีใดจะให้คะแนนการเปลี่ยนแปลงได้สอดคล้องมากที่สุดกับวิธีโมเดลพหุระดับ ซึ่งเป็นวิธีประมาณค่าที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ โดยผู้วิจัยได้พิจารณาผลของการวัดการเปลี่ยนแปลงจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง ค่าความสัมพัทธ์ของคะแนนการเปลี่ยนแปลง ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าของคะแนนการเปลี่ยนแปลง และค่าความลำเอียงของคะแนนการเปลี่ยนแปลง

ความมุ่งหมายของการวิจัย

ความมุ่งหมายของการวิจัยครั้งนี้เป็นการศึกษาผลของการวัดการเปลี่ยนแปลงห้าวิธี คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ โดยการวัดการเปลี่ยนแปลงด้วยวิธีโมเดลพหุระดับ เป็นวิธีประมาณค่าที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ ดังนี้

1. ศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง โดยศึกษาสองประเด็น ดังนี้

1.1 ศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษาและวิธีโมเดลพหุระดับ ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

1.2 เปรียบเทียบความสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

2. ศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง โดยศึกษาสองประเด็น ดังนี้

2.1 ศึกษาความเชื่อมั่นของการวัดการเปลี่ยนแปลง โดยพิจารณาจากความคลาดเคลื่อนมาตรฐานในการประมาณค่าคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

2.2 ศึกษาความเที่ยงตรงของการวัดการเปลี่ยนแปลง โดยพิจารณาจากค่าความลำเอียงของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

ความสำคัญของการวิจัย

การวิจัยในครั้งนี้เป็นการศึกษาผลของการวัดการเปลี่ยนแปลงจำนวนห้าวิธี คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ ว่าวิธีวัดใดให้ผลการวัดการเปลี่ยนแปลงที่เชื่อถือได้ และมีความสอดคล้องมากที่สุดกับวิธีโมเดลพหุระดับ ซึ่งเป็นการประมาณค่าที่ใช้เป็นเกณฑ์

ผลจากการศึกษานี้จะทำให้ได้หลักฐานเชิงประจักษ์เพื่อยืนยันประสิทธิภาพของวิธีการวัดการเปลี่ยนแปลงเชิงปริมาณแบบวัดซ้ำสองครั้งที่ใช้ในการศึกษาพัฒนาการรายบุคคล ในเนื้อหาวิชาคณิตศาสตร์ เรื่อง สมการเชิงเส้นตัวแปรเดียว ของนักศึกษาการศึกษานอกโรงเรียน ระดับมัธยมศึกษาตอนต้น ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน และทำให้มีทางเลือกในการวัดการเปลี่ยนแปลงเพิ่มขึ้น ที่สามารถประมาณค่าคะแนนการเปลี่ยนแปลงใกล้เคียงหรือสอดคล้องกับคะแนนการเปลี่ยนแปลงจริงของแต่ละบุคคลและยังเป็นการกระตุ้นให้ครูผู้สอนตระหนักถึงความสำคัญของคะแนนการเปลี่ยนแปลงรายบุคคลตามพระราชบัญญัติการศึกษาแห่งชาติ พุทธศักราช 2542 และที่แก้ไขเพิ่มเติม (ฉบับที่ 2) พ.ศ. 2545 ในการนำไปใช้ร่วมกับคะแนนสอบปลายภาค

เพื่อการตัดสินใจผลการเรียนของนักเรียน นักศึกษา ได้ถูกต้อง แม่นยำมากยิ่งขึ้น เป็นสารสนเทศเกี่ยวกับค่าประมาณคะแนนการเปลี่ยนแปลงรายบุคคลที่เป็นประโยชน์ต่อครูผู้สอนในการพัฒนาการเรียนรู้นักเรียน ตลอดจนเป็นประโยชน์ต่อสถานศึกษาในการประเมินคุณภาพการพัฒนาการศึกษาของนักศึกษา สำหรับการรับรองการประกันคุณภาพสถานศึกษา และยังสามารถนำวิธีการวัดการเปลี่ยนแปลงที่เชื่อถือได้ไปใช้ในการฝึกอบรมได้อีกด้วย

นอกจากนี้ยังได้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนในเนื้อหาวิชาคณิตศาสตร์ เรื่อง สมการเชิงเส้นตัวแปรเดียว ของนักศึกษาการศึกษานอกโรงเรียน ระดับมัธยมศึกษาตอนต้น ที่ผ่านการตรวจสอบและพัฒนาคุณภาพตามทฤษฎีการทดสอบแบบมาตรฐานเดิมและทฤษฎีการตอบข้อสอบ อันเป็นแนวทางในการสร้างและพัฒนาแบบทดสอบในเนื้อหาวิชา และระดับอื่นต่อไปด้วย

ขอบเขตของการวิจัย

ประชากรที่ใช้ในการวิจัย

ประชากรในการวิจัยครั้งนี้ คือ นักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาระดับพื้นฐาน พุทธศักราช 2544 ระดับมัธยมศึกษาตอนต้น ของศูนย์บริการการศึกษานอกโรงเรียนเขต จำนวน 12 แห่ง ซึ่งเป็นสถานศึกษาในสังกัดศูนย์การศึกษานอกโรงเรียนกรุงเทพมหานคร 1 ในภาคเรียนที่ 2 ปีการศึกษา 2549 มีจำนวน 7,555 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาระดับพื้นฐาน พุทธศักราช 2544 ระดับมัธยมศึกษาตอนต้น ของศูนย์บริการการศึกษานอกโรงเรียนเขต จำนวน 7 แห่ง ซึ่งเป็นสถานศึกษาในสังกัดศูนย์การศึกษานอกโรงเรียนกรุงเทพมหานคร 1 ในภาคเรียนที่ 2 ปีการศึกษา 2549 มีจำนวน 1,085 คน

ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ ได้แก่

1.1 การวัดการเปลี่ยนแปลง จำนวน 5 วิธี คือ

1.1.1 การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ

1.1.2 การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์

1.1.3 การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง

1.1.4 การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้

1.1.5 การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์

1.2 ความยาวของแบบทดสอบ จำนวน 3 ขนาด คือ 40 ข้อ, 30 ข้อ และ 20 ข้อ

1.3 ขนาดกลุ่มตัวอย่าง จำนวน 4 ขนาด คือ 1,000 คน, 500 คน, 300 คน และ 100 คน

2. ตัวแปรตาม ได้แก่ ผลของการวัดการเปลี่ยนแปลง

กรอบแนวคิดในการวิจัย

ในการศึกษาครั้งนี้ ผู้วิจัยกำหนดกรอบแนวคิดไว้ 2 ประเด็น ดังนี้

1. วิธีการวัดการเปลี่ยนแปลง

การรายงานผลพัฒนาการเรียนรู้รายบุคคล รายชั้นเรียน รายสถานศึกษาทั้งในประเทศและต่างประเทศพบว่า ส่วนใหญ่นิยมใช้การรายงานผลพัฒนาการเรียนรู้ด้วยคะแนนความแตกต่างระหว่างคะแนนดิบ แต่เนื่องจากการวัดการเปลี่ยนแปลงด้วยวิธีดังกล่าวมีข้อจำกัดเกี่ยวกับผลจากอิทธิพลคะแนนสูงสุด ซึ่งส่งผลทำให้เกิดปัญหาเกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงมีค่าเป็นลบ และไม่คำนึงถึงอิทธิพลพื้นฐานความรู้เดิมทำให้มีผลต่อการแปลความหมายของคะแนนการเปลี่ยนแปลงที่เท่ากันของผู้สอบที่มีความสามารถก่อนเรียนแตกต่างกัน ตลอดจนคะแนนการเปลี่ยนแปลงยังมีความเชื่อมั่นต่ำอีกด้วย ต่อมา บอค ทิสเซน และซิมโวกสกี (Bock; Thissen; & Zimowski. 1997) และไฮเดนเบิร์ก (Heidenberg. 2000: 1-25) ได้ศึกษาและวิจัยพบว่าการใช้คะแนนมวลความรู้ที่ประมาณค่าตามทฤษฎีการตอบข้อสอบเป็นการลดความสัมพันธ์ที่เป็นลบระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงได้จริง จึงได้เสนอแนะให้นำคะแนนมวลความรู้ที่ประมาณค่าตามทฤษฎีการตอบข้อสอบมาใช้ในการวัดการเปลี่ยนแปลงจากการสอบก่อนเรียนและหลังเรียน โดยไม่ต้องคำนึงถึงลักษณะการแจกแจงความสามารถเริ่มต้นของผู้สอบ ดังนั้น ผู้วิจัยจึงได้ปรับขยายวิธีการวัดการเปลี่ยนแปลงของไฮเดนเบิร์ก (Heidenberg. 2000) โดยการนำวิธีการประมาณค่าพารามิเตอร์ด้วย EAP ซึ่งเป็นวิธีของเบย์ ตามแนวคิดของ เมย์ และไนส์วันเดอร์ (May; & Nicewander) และคำนึงถึงอิทธิพลคะแนนสูงสุดและพื้นฐานความรู้เดิมโดยการประยุกต์แนวคิดของศิริชัย กาญจนวาสี (2538) มาใช้ ซึ่งเรียกวิธีที่ปรับขยายนี้ว่า *การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์* เพื่อนำมาศึกษาผลของการวัดการเปลี่ยนแปลงเมื่อเทียบกับวิธีอื่นๆ ที่เป็นวิธีที่นิยมใช้ หรือ เป็นวิธีที่มีค่าความคลาดเคลื่อนในการวัดต่ำ หรือเป็นวิธีที่ใช้เป็นฐานความคิดในการปรับขยายวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ในครั้งนี้

2. ปัจจัยที่ส่งผลต่อคะแนนการเปลี่ยนแปลง

ผลการวิจัยของ ฟิชเชอร์ (Fischer. 2003: 3-26) พบว่า มีหลายปัจจัยที่ส่งผลต่อความแม่นยำในการวัดคะแนนการเปลี่ยนแปลง เช่น ความยาวของแบบทดสอบ ลักษณะการแจกแจงค่าพารามิเตอร์ข้อสอบ ลักษณะการแจกแจงความสามารถก่อนเรียน ลักษณะการแจกแจงคะแนนการเปลี่ยนแปลง ขนาด

กลุ่มตัวอย่าง จำนวนครั้งในการเก็บรวบรวมข้อมูล ระยะห่างในการเก็บรวบรวมข้อมูลแต่ละครั้ง และการขาดหายของกลุ่มตัวอย่าง เป็นต้น ส่วนไฮเดนเบิร์ก (Heidenberg. 2000) พบว่าเงื่อนไขเกี่ยวกับลักษณะการแจกแจงค่าความยากของแบบทดสอบ ความยาวของแบบทดสอบ ขนาดกลุ่มตัวอย่าง ปริมาณคะแนนการเปลี่ยนแปลงจริง ลักษณะการแจกแจงคะแนนการเปลี่ยนแปลงจริง และจำนวนพารามิเตอร์ในโมเดลโลจิสติก มีผลต่อค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงกับคะแนนสอบก่อนเรียนทุกกรณี

ในขณะที่ฮูลิน ดราสโกว และพาร์สัน (Hulin; Drasgow; & Parsons. 1983?: 99) ได้กล่าวถึงปัจจัยที่ส่งผลต่อการประมาณค่าพารามิเตอร์ คือ ความยาวของแบบทดสอบ ขนาดกลุ่มตัวอย่าง และจำนวนพารามิเตอร์ของโมเดลโลจิสติก จากผลการวิจัยดังกล่าวพบว่า ทั้งความยาวของแบบทดสอบ และขนาดกลุ่มตัวอย่าง เป็นปัจจัยสำคัญที่ส่งผลต่อความแม่นยำในการวัดคะแนนการเปลี่ยนแปลง ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง และการประมาณค่าพารามิเตอร์ ดังนั้นในการวิจัยครั้งนี้จึงศึกษาปัจจัยเกี่ยวกับ ความยาวของแบบทดสอบ และขนาดกลุ่มตัวอย่างที่แตกต่างกัน ซึ่งจะกล่าวแต่ละประเด็นดังนี้

2.1 ความยาวของแบบทดสอบ ผู้วิจัยกำหนดความยาวของแบบทดสอบโดยใช้เกณฑ์ของเวลาที่ใช้ในการดำเนินการสอบและความนิยมใช้ในเชิงปฏิบัติ ดังนักวิจัยวิทยาได้กล่าวไว้ คือ นักวัดผลการศึกษาได้กำหนดความยาวแบบทดสอบไว้หลายขนาด โดยเฉพาะแบบทดสอบที่ใช้วัดซ้ำสองครั้ง คือ ก่อนเรียนและหลังเรียนในระยะเวลาที่ใกล้เคียงกันแล้วทำให้นักศึกษาเกิดการเปลี่ยนแปลงนั้น อาจเกิดจากการจำข้อสอบได้ของผู้สอบและเพื่อไม่ให้เกิดอิทธิพลคะแนนสูงสุด ดังนั้นจำนวนข้อสอบจึงควรมีมากพอ ส่วนนอวิก และเลวิส (Norvick; & Lewis. 1974: 31) ได้กำหนดความยาวของแบบทดสอบอิงเกณฑ์ ไว้ว่า ควรมีจำนวนข้อสอบอย่างน้อย 20 ข้อ สอดคล้องกับ ฮาลาดินา และรอยด์ (Haladyna; & Roid. 1983: 280) ที่กล่าวไว้ว่า แบบทดสอบที่มีจำนวน 20-30 ข้อ มีความเหมาะสมในการประมาณค่าคะแนนมวลความรู้มากที่สุดและให้ค่าการประมาณความสามารถที่แม่นยำ แต่ถ้าเพิ่มจำนวนข้อสอบมากกว่า 30 ข้อ จะทำให้สิ้นเปลืองค่าใช้จ่ายสูงเกินไป

นอกจากนี้ ฮูลิน ดราสโกว และพาร์สัน (Hulin; Drasgow; & Parsons. 1983?: 101-105) ได้เสนอแนะว่าการใช้โมเดลโลจิสติกแบบสามพารามิเตอร์เพื่อประมาณค่าพารามิเตอร์ความสามารถด้วยวิธีความเป็นไปได้สูงสุดให้ได้ผลถูกต้อง (Accuracy) นั้น ควรใช้แบบทดสอบที่มีจำนวนข้อ 60 ข้อ กับกลุ่มตัวอย่างจำนวนไม่น้อยกว่า 200 คน หรือถ้าใช้แบบทดสอบที่มีจำนวนข้อน้อยกว่า 30 ข้อ จะต้องใช้กลุ่มตัวอย่างจำนวนไม่น้อยกว่า 1,000 คน และได้กล่าวถึงความสัมพันธ์ระหว่างความยาวของแบบทดสอบกับขนาดกลุ่มตัวอย่างไว้ว่า ถ้าใช้แบบทดสอบที่มีจำนวนข้อเพิ่มขึ้นเท่าตัวแล้ว สามารถลดจำนวนผู้สอบลงได้เท่าตัวในการประมาณค่าพารามิเตอร์ความสามารถให้ถูกต้อง ในขณะที่ ลอร์ด

(Hulin; Drasgow; & Parsons. 1983?: 100, 139; citing Lord. 1968: Educational and Psychological Measurement. pp. 989-1020) กล่าวว่าการประมาณค่าพารามิเตอร์ด้วยวิธีความเป็นไปได้สูงสุด (Maximum likelihood) ในโมเดลโลจิสติกแบบสามพารามิเตอร์ให้มีความถูกต้องนั้น ควรใช้ข้อสอบประมาณ 50 ข้อ กับกลุ่มผู้เข้าสอบอย่างน้อย 1,000 คน

ส่วนฮัทเทน (Hutten. 1981: 4799-A) ได้กล่าวถึง ความสัมพันธ์ของค่าความสามารถของผู้สอบจำนวน 1,000 คน ในโมเดลโลจิสติกแบบสามพารามิเตอร์ ระหว่างแบบทดสอบที่มีจำนวน 40 ข้อ กับ 20 ข้อ มีความสัมพันธ์กันสูง เท่ากับ 0.87 โดยที่ ซามิไจมา (Samejima. 1994: 328) พบว่า ความแม่นยำในการประมาณค่าความสามารถของผู้สอบจำนวน 800 คน ที่มีพิสัยของตำแหน่งความสามารถที่แตกต่างกันกับจำนวนข้อสอบในแบบทดสอบ คือ พิสัยของตำแหน่งความสามารถของผู้สอบ -3.0 ถึง +3.0, -2.6 ถึง +2.6, -2.2 ถึง +2.2 และ -1.8 ถึง +1.8 ที่ประมาณค่าด้วยวิธีความเป็นไปได้สูงสุดได้ถูกต้องนั้น ต้องใช้แบบทดสอบที่มีจำนวนข้อสอบที่แตกต่างกัน คือ 2,000, 300, 100 และ 50 ข้อ ตามลำดับ ส่วนอีเบล (Ebel. 1972: 205) ได้เสนอแบบทดสอบเลือกตอบแบบ 4 ตัวเลือกมีความเหมาะสมในจำนวน 50 ข้อ ที่ใช้เวลาสอบประมาณ 1 ชั่วโมง ซึ่งในทางปฏิบัติแล้วข้อสอบที่ใช้ในการทดสอบนักศึกษาการศึกษานอกโรงเรียนมีลักษณะเป็นแบบเลือกตอบ 4 ตัวเลือก และมีจำนวนข้อสอบอยู่ระหว่าง 20-30 ข้อ

ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงสร้างข้อสอบ 4 ตัวเลือก โดยกำหนดความยาวของแบบทดสอบเป็น 3 ขนาด คือ 40 ข้อ 30 ข้อ และ 20 ข้อ โดยการสุ่มแบบแบ่งชั้น (Stratified random sampling) ด้วยการสุ่มผลการสอบของนักศึกษาจากแบบทดสอบที่มีข้อสอบจำนวน 40 ข้อ ไปเป็นผลการสอบของแบบทดสอบ 2 ฉบับ ตามสัดส่วนความสำคัญของแต่ละเนื้อหาในแต่ละมวลความรู้ย่อย ได้แก่ ฉบับ 30 ข้อ และฉบับ 20 ข้อ เพื่อศึกษาผลของการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษา ด้วยแบบทดสอบที่มีความยาวแตกต่างกันว่าสามารถประมาณค่าคะแนนการเปลี่ยนแปลงของผู้สอบได้สอดคล้องกันหรือไม่แตกต่างกันเพียงใด

2.2 ขนาดกลุ่มตัวอย่าง ผู้วิจัยได้กำหนดขนาดกลุ่มตัวอย่างโดยใช้หลักการของความเป็นตัวแทนกลุ่มขนาดใหญ่ กลางและเล็ก และขนาดกลุ่มตัวอย่างที่มีผลต่อความแม่นยำ ถูกต้องในการประมาณค่าพารามิเตอร์ ดังผลการศึกษาของ ฮูลิน ดรอสโกว และพาร์สัน (Hulin; Drasgow; & Parsons. 1983?: 101-105) ได้เสนอแนะว่าการใช้โมเดลโลจิสติกแบบสามพารามิเตอร์เพื่อประมาณค่าพารามิเตอร์ความสามารถให้ได้ผลถูกต้องนั้น ควรใช้กลุ่มตัวอย่างประมาณ 1,000 คน กับแบบทดสอบที่มีจำนวนข้อสอบน้อยกว่า 30 ข้อ สอดคล้องกับ ลอร์ด (Hulin; Drasgow; & Parsons. 1983?: 100, 139. citing Lord. 1968. Educational and Psychological Measurement. pp. 989-1020) และ ฮัทเทน (Hutten. 1981: 4799-A) ที่กล่าวไว้ว่าถ้าเลือกใช้โมเดลโลจิสติกแบบสามพารามิเตอร์จะประมาณค่าพารามิเตอร์

ความสามารถได้ดีที่สุดเมื่อขนาดกลุ่มตัวอย่าง 1,000 คนขึ้นไป แต่ถ้าเลือกใช้โมเดลโลจิสติกแบบหนึ่งหรือสองพารามิเตอร์ควรมีกกลุ่มตัวอย่างประมาณ 300-500 คน

สภาพการดำเนินงานจัดการศึกษานอกโรงเรียนตามหลักสูตรการศึกษาขั้นพื้นฐาน ในแต่ละกลุ่มที่จัดกิจกรรมการเรียนรู้มีนักศึกษาประมาณ 80 -120 คน และในแต่ละศูนย์การเรียนรู้ชุมชนมีนักศึกษาประมาณ 80-500 คน ดังนั้นในการศึกษาค้างนี้จึงใช้กลุ่มตัวอย่าง 4 ขนาด คือ 1,000 คน 500 คน 300 คน และ 100 คน ซึ่งเป็นตัวแทนกลุ่มตัวอย่างที่มีขนาดใหญ่ ปานกลาง และเล็ก ตามลำดับ โดยการสุ่มผลการสอบของนักศึกษาจำนวน 1,000 คน ไปเป็นผลการสอบของกลุ่มตัวอย่าง 3 ขนาด คือ 500 คน 300 คน และ 100 คน เพื่อศึกษาผลของการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับกลุ่มตัวอย่างที่มีจำนวนผู้สอบแตกต่างกันว่าสามารถประมาณค่าคะแนนการเปลี่ยนแปลงของผู้สอบได้สอดคล้องกันหรือแตกต่างกันเพียงใด

นิยามศัพท์เฉพาะ

1. **คะแนนการเปลี่ยนแปลง** หมายถึง คะแนนที่แตกต่างไปจากเดิมที่เกิดจากการสอบซ้ำสองครั้ง คือก่อนเรียนและหลังเรียน ซึ่งในการวิจัยครั้งนี้เป็นคะแนนการเปลี่ยนแปลงเนื้อหาวิชาคณิตศาสตร์ เรื่อง สมการเชิงเส้นตัวแปรเดียว ของนักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาขั้นพื้นฐาน ระดับมัธยมศึกษาตอนต้น ที่ใช้การวัดการเปลี่ยนแปลงจำนวนห้าวิธี คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ โดยวิธีประมาณค่าที่ใช้เป็นเกณฑ์ในการศึกษาค้างนี้ คือ โมเดลพหุระดับ

2. **คะแนนมวลความรู้** หมายถึง คะแนนที่คาดหวังของการตอบข้อสอบในประชากรข้อสอบได้ถูกต้อง โดยคำนวณจากค่าเฉลี่ยของความน่าจะเป็นในการตอบข้อสอบที่ได้มาจากประชากรข้อสอบที่กำหนดหรือนิยามไว้อย่างชัดเจน สำหรับประมาณค่าความสามารถของผู้สอบ ตามแนวคิดของ บอค ทิสเซน และซิมโมสกี (Bock; Thissen; & Zimsky. 1997) ในทฤษฎีการตอบข้อสอบ

3. **ผลของการวัดการเปลี่ยนแปลง** หมายถึง คะแนนการเปลี่ยนแปลงที่ประมาณค่าด้วยการวัดการเปลี่ยนแปลงที่แตกต่างกันห้าวิธี

4. **ประสิทธิภาพของการวัดการเปลี่ยนแปลง** หมายถึง ความเชื่อมั่น และความเที่ยงตรงของการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษาจำนวนห้าวิธี โดยพิจารณาจากค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และค่าความลำเอียง (BIAS) ตามลำดับ

5. การวัดการเปลี่ยนแปลง หมายถึง วิธีคำนวณคะแนนการเปลี่ยนแปลงที่ผู้วิจัยสนใจศึกษา จำนวนหัวข้อวิธี โดยมีวิธีโมเดลพหุระดับ เป็นวิธีประมาณค่าที่ใช้เป็นเกณฑ์ ดังนี้

5.1 วิธีวัดคะแนนความแตกต่างระหว่างคะแนนดิบ หมายถึง วิธีการวัดคะแนนการเปลี่ยนแปลงด้วยการนำคะแนนดิบหลังเรียนลบด้วยคะแนนดิบก่อนเรียน ($X_2 - X_1$)

5.2 วิธีวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ หมายถึง วิธีการวัดคะแนนการเปลี่ยนแปลงด้วยร้อยละของอัตราส่วนระหว่างผลต่างระหว่างคะแนนดิบกับผลต่างระหว่างคะแนนเต็มกับคะแนนดิบก่อนเรียน $\left(\left(\frac{X_2 - X_1}{F - X_1} \right) 100 \right)$

5.3 วิธีวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง หมายถึง วิธีการวัดคะแนนการเปลี่ยนแปลงด้วยความสามารถหลังเรียนลบด้วยความสามารถก่อนเรียน โดยใช้วิธีการประมาณค่าพารามิเตอร์ด้วย EAP (Expected a posteriori) ($\theta_2 - \theta_1$)

5.4 วิธีวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ หมายถึง วิธีการวัดคะแนนการเปลี่ยนแปลงด้วยคะแนนมวลความรู้หลังเรียนลบด้วยคะแนนมวลความรู้ก่อนเรียนที่ประมาณค่าตามทฤษฎีการตอบข้อสอบ ($\pi_2 - \pi_1$)

5.5 วิธีวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ หมายถึง วิธีการวัดคะแนนการเปลี่ยนแปลงที่ผู้วิจัยปรับขยายขึ้นมา เป็นค่าร้อยละของอัตราส่วนระหว่างผลต่างระหว่างคะแนนมวลความรู้กับผลต่างระหว่างคะแนนเต็มร้อยละกับคะแนนมวลความรู้ก่อนเรียนที่ประมาณค่าตามทฤษฎีการตอบข้อสอบ $\left(\left(\frac{\pi_2 - \pi_1}{100 - \pi_1} \right) 100 \right)$

6. วิธีโมเดลพหุระดับ หมายถึง วิธีคำนวณคะแนนการเปลี่ยนแปลงโดยใช้คะแนนจริง ซึ่งเป็นวิธีที่นำค่าความคลาดเคลื่อนมาร่วมวิเคราะห์ ทำให้ผลการวิเคราะห์คะแนนการเปลี่ยนแปลงเชื่อถือได้ ในการวิจัยครั้งนี้ได้ประยุกต์ใช้โมเดลพหุระดับ ที่เรียกว่า โมเดลวัดการเปลี่ยนแปลงแบบพหุระดับ โดยวิเคราะห์ข้อมูลด้วยโปรแกรม HLM (Hierarchical linear and nonlinear model) แบบ 2 ระดับ ของเราดินบุช ไบรย์ และคองดอน (Raudenbush; Bry; & Congdon. 2000) ซึ่งในการศึกษาครั้งนี้ ผู้วิจัยใช้คะแนนการเปลี่ยนแปลงที่ได้จากการประมาณค่าวิธีนี้เป็นเกณฑ์

7. ความยาวของแบบทดสอบ หมายถึง จำนวนข้อสอบในแบบทดสอบที่ใช้ในการสอบทั้งก่อนเรียนและหลังเรียน ในเนื้อหาวิชาคณิตศาสตร์ เรื่องสมการเชิงเส้นตัวแปรเดียว ระดับมัธยมศึกษาตอนต้น หลักสูตรการศึกษานอกโรงเรียน พุทธศักราช 2544 ในการศึกษาค้นคว้าครั้งนี้ผู้วิจัยกำหนดความยาวของแบบทดสอบออกเป็น 3 ขนาด คือ 40 ข้อ 30 ข้อ และ 20 ข้อ โดยการสร้างแบบทดสอบที่มีจำนวน 40 ข้อ แล้วสุ่มข้อสอบแบบแบ่งชั้น โดยมีมวลความรู้ย่อยเป็นชั้น และข้อสอบเป็นหน่วยการสุ่ม จัดให้เป็น

แบบทดสอบฉบับที่มีจำนวน 30 ข้อ และ 20 ข้อ ตามลำดับ เพื่อให้ได้ข้อสอบที่มีความครอบคลุมเนื้อหา โดยพิจารณาสัดส่วนความสำคัญของแต่ละมวลความรู้ย่อย

8. ขนาดกลุ่มตัวอย่าง หมายถึง จำนวนนักศึกษาที่เข้าสอบก่อนเรียน ระหว่างเรียนและหลังเรียน ซึ่งเป็นคนเดียวกัน ในการศึกษาครั้งนี้ผู้วิจัยกำหนดขนาดกลุ่มตัวอย่างเป็น 4 ขนาด คือ 1,000 คน 500 คน 300 คน และ 100 คน โดยการเก็บรวบรวมข้อมูลกับกลุ่มตัวอย่างจำนวน 1,000 คน แล้วสุ่มผู้สอบอย่างง่าย จัดให้มีขนาดกลุ่มตัวอย่าง 500 คน 300 คน และ 100 คน ตามลำดับ

9. แบบทดสอบก่อนเรียนและหลังเรียน หมายถึง แบบทดสอบที่ใช้สอบก่อนเรียนและหลังเรียนเนื้อหาวิชาคณิตศาสตร์ เรื่อง สมการเชิงเส้นตัวแปรเดียว ระดับมัธยมศึกษาตอนต้น หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 ตามหลักเกณฑ์และวิธีการจัดการศึกษานอกโรงเรียน ที่สร้างเป็นแบบทดสอบอิงเกณฑ์โดยใช้เทคโนโลยีการสร้างแบบทดสอบ คือ ลักษณะเฉพาะข้อสอบ (Item Specification) ตามแนวคิดของ สงบ ลักษณะ (2525)

สมมติฐานของการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยนำผลการวัดการเปลี่ยนแปลงที่สนใจศึกษาจำนวนห้าวิธี คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มาเปรียบเทียบกัน เมื่อวิธีโมเดลพหุระดับเป็นวิธีประมาณค่าที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ โดยพิจารณาจากค่าความสัมพันธ์ของคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง ค่าความสัมพันธ์ของคะแนนการเปลี่ยนแปลง ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าคะแนนการเปลี่ยนแปลง และค่าความลำเอียงของคะแนนการเปลี่ยนแปลง ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน โดยผู้วิจัยได้ตั้งสมมติฐาน ดังนี้

1. คะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีความสัมพันธ์แบบเพียร์สันไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่าวิธีอื่น โดย

1.1 แบบทดสอบที่มีจำนวนข้อสอบมากจะมีค่าความสัมพันธ์ไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่าแบบทดสอบที่มีจำนวนข้อสอบน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง

1.2 กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีค่าความสัมพันธ์ไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่ากลุ่มตัวอย่างที่มีจำนวนผู้สอบน้อยกว่า ในทุกความยาวของแบบทดสอบ

2. คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนน
มวลดความรู้สัมพัทธ์ มีความสัมพันธ์แบบเพียร์สันทางบวกกับวิธีโมเดลพระระดับสูงกว่าวิธีอื่น โดย

2.1 แบบทดสอบที่มีจำนวนข้อสอบมากจะมีค่าความสัมพันธ์สูงกว่าแบบทดสอบที่มี
จำนวนข้อน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง

2.2 กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีค่าความสัมพันธ์สูงกว่ากลุ่มตัวอย่างที่มีจำนวน
ผู้สอบน้อยกว่า ในทุกความยาวของแบบทดสอบ

3. การวัดคะแนนความแตกต่างระหว่างคะแนนมวลดความรู้สัมพัทธ์ มีความเชื่อมั่นในการ
ประมาณค่าคะแนนการเปลี่ยนแปลงสูงกว่าวิธีอื่น โดย

3.1 แบบทดสอบที่มีจำนวนข้อสอบมากจะมีความเชื่อมั่นในการประมาณค่าสูงกว่าแบบ
ทดสอบที่มีจำนวนข้อน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง

3.2 กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีความเชื่อมั่นในการประมาณค่าสูงกว่ากลุ่ม
ตัวอย่างที่มีจำนวนผู้สอบที่น้อยกว่า ในทุกความยาวของแบบทดสอบ

4. การวัดคะแนนความแตกต่างระหว่างคะแนนมวลดความรู้สัมพัทธ์ มีความเที่ยงตรงในการ
ประมาณค่าคะแนนการเปลี่ยนแปลงสูงกว่าวิธีอื่น โดย

4.1 แบบทดสอบที่มีจำนวนข้อสอบมากจะมีความเที่ยงตรงในการประมาณค่าสูงกว่าแบบ
ทดสอบที่มีจำนวนข้อน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง

4.2 กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีความเที่ยงตรงในการประมาณค่าสูงกว่ากลุ่ม
ตัวอย่างที่มีจำนวนผู้สอบที่น้อยกว่า ในทุกความยาวของแบบทดสอบ

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาแนวคิด ทฤษฎีต่างๆ จากเอกสารและงานวิจัยที่เกี่ยวข้อง และได้นำเสนอตามหัวข้อต่อไปนี้

1. ทฤษฎีการทดสอบแบบมาตรฐานเดิม
2. การวัดการเปลี่ยนแปลงตามทฤษฎีการทดสอบแบบมาตรฐานเดิม
3. โมเดลพหุระดับสำหรับการวัดการเปลี่ยนแปลง
4. ทฤษฎีการตอบข้อสอบ
5. การวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบ
6. คะแนนมวลความรู้
7. งานวิจัยที่เกี่ยวข้อง

ทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory)

ทฤษฎีการทดสอบแบบมาตรฐานเดิมเริ่มต้นจากการศึกษาค้นคว้าของ ชาร์ล สเปียร์แมน (Charles Spearman. 1904) เกี่ยวกับค่าสัมประสิทธิ์สหสัมพันธ์ที่ปรับแก้ด้วยความคลาดเคลื่อนในการวัดสำหรับการประมาณค่าความเชื่อมั่น และได้รับการพัฒนาอย่างต่อเนื่อง ซึ่งทฤษฎีการทดสอบแบบมาตรฐานเดิมนี้อธิบายความแม่นยำของการวัด โดยยึดถือข้อตกลงคุณสมบัติคู่ขนานหรือความเท่าเทียมกันเป็นสำคัญ ในการวัดภายใต้เงื่อนไขการวัดที่เท่าเทียมกัน (Equivalent) เช่น การวัดด้วยแบบทดสอบคู่ขนาน (Paralleled test) นั้น มีข้อตกลงว่า คะแนนสอบที่ได้ค่าเฉลี่ยเท่ากัน ความแปรปรวนเท่ากัน ความแปรปรวนร่วมเท่ากันและไม่มีปฏิสัมพันธ์ (Interaction) ระหว่างผู้สอบกับแบบทดสอบรวม คะแนนที่สอบได้หรือค่าสังเกต (Observe score: X) ของผู้สอบคนหนึ่งๆ จะประกอบด้วย 2 ส่วน คือ ส่วนที่เป็นคะแนนจริง (True score: T) ซึ่งเป็นค่าความคาดหวังของคะแนนที่ได้จากความสามารถที่แท้จริงของผู้สอบเมื่อสอบด้วยแบบทดสอบคู่ขนาน และส่วนที่เป็นความคลาดเคลื่อน (Error score: E) ซึ่งเกิดจากแหล่งความคลาดเคลื่อนเพียงแหล่งเดียวเท่านั้น คือ ความคลาดเคลื่อนเชิงสุ่ม (Random error) โดยที่คะแนนจริงกับความคลาดเคลื่อนมีความเป็นอิสระต่อกัน คะแนนสังเกตหรือคะแนนที่สอบได้สามารถเขียนเป็นสมการได้ดังนี้

$$X = T + E$$

ดังนั้น ความแปรปรวนของคะแนนสังเกต (σ_x^2) จึงเป็นผลรวมของความแปรปรวนของคะแนนจริง (σ_t^2) กับความแปรปรวนของความคลาดเคลื่อน (σ_e^2) โดยมีสมการ ดังนี้

$$\sigma_x^2 = \sigma_t^2 + \sigma_e^2$$

สำหรับความคลาดเคลื่อนซึ่งได้ค่าสัมประสิทธิ์สหสัมพันธ์แบบ Intraclass จะประมาณค่าได้จากกำลังสองของความสัมพันธ์ระหว่างคะแนนสังเกตกับคะแนนจริง ซึ่งเท่ากับอัตราส่วนระหว่างความแปรปรวนของคะแนนจริงกับค่าความแปรปรวนของคะแนนสังเกต ได้ดังสมการต่อไปนี้

$$\begin{aligned} \rho_{xx'} &= \rho_{xt} \\ &= \sigma_t^2 / \sigma_x^2 \end{aligned}$$

ต่อมาได้มีการพัฒนาดัชนีค่าความเชื่อมั่นตามลักษณะของการแบ่งแบบทดสอบเป็นฉบับย่อยที่ใช้ระดับของความเป็นคู่ขนานของการวัดในแต่ละฉบับย่อยเป็นเกณฑ์ และขนาดความยาวของแบบทดสอบย่อย เช่น ความเชื่อมั่นแบบแบ่งครึ่งฉบับ โดย Spearman-Brown (1910) ความสัมพันธ์ระหว่างคะแนนข้อสอบกับคะแนนแบบทดสอบทั้งฉบับที่ให้คะแนนเป็นแบบสองค่า โดย Kuder-Richardson (1936 -7) การหาความเชื่อมั่นด้วยการวิเคราะห์ความแปรปรวน โดย Hoyt เป็นต้น

การวัดการเปลี่ยนแปลงตามทฤษฎีการทดสอบแบบมาตรฐานเดิม

การวัดการเปลี่ยนแปลงตามทฤษฎีการทดสอบแบบมาตรฐานเดิมเป็นวิธีการวัดการเปลี่ยนแปลงที่เริ่มวิเคราะห์จากคะแนนสังเกตหรือคะแนนดิบจากการสอบสองครั้ง คือในช่วงก่อนเรียนและหลังเรียน หรือก่อนและหลังการได้รับการจัดกระทำ (Treatment) ซึ่งไม่มีการประมาณค่าความคลาดเคลื่อนในการวัด ต่อมานักจิตวิทยาได้พัฒนาวิธีการที่คำนึงถึงค่าความคลาดเคลื่อนในการวัด ในการวัด การเปลี่ยนแปลงที่จะนำเสนอต่อไปนี้เป็น การวัดซ้ำสองครั้ง (Two-wave) โดยมีผู้พัฒนาวิธีการไว้หลายวิธีสามารถสรุปวิธีการที่สำคัญได้ ดังนี้

1. การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ (Difference scores)
2. การวัดคะแนนการเปลี่ยนแปลงส่วนที่เหลือ (Residual score)
3. การวัดคะแนนการเปลี่ยนแปลงที่เป็นอิสระจากการวัดครั้งแรกที่เป็นฐาน (Base – free measure of change)
4. การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ (Relative change score)
5. การวัดคะแนนเพิ่มที่แท้จริง (Estimated true gain score)
6. การวัดการเปลี่ยนแปลงการเรียนรู้ ของ อรุณี อ่อนสวัสดิ์
7. การวัดคะแนนเพิ่ม ของ วินิจ เทือกทอง

8. การวัดคะแนนการเปลี่ยนแปลงของตำแหน่งเปอร์เซ็นต์ไทล์ (Change in percentile rank)
9. การวัดคะแนนการเปลี่ยนแปลงเป็นร้อยละ (Percent change)

1. การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ (Difference scores)

ลอร์ด (Traub. 1994: 127-133; citing Lord. 1956. Educational and Psychological Measurement. 421-437; 1958.; 1963. Problems in measuring change.) ได้กล่าวถึงการวัดคะแนนความแตกต่างอย่างง่ายหรือความแตกต่างระหว่างคะแนนดิบ (Difference scores) เป็นวิธีแรกที่ได้รับการพัฒนาขึ้นโดยคำนวณคะแนนการเปลี่ยนแปลงจากผลต่างของคะแนนดิบหรือคะแนนสังเกตระหว่างการวัดก่อนเรียน (X) และการวัดหลังเรียน (Y) มีข้อตกลงเบื้องต้นว่าคะแนนดิบจากการวัดแต่ละครั้งประกอบด้วยคะแนนจริง (true score) และความคลาดเคลื่อนในการวัด (Measurement error) โดยที่ความคลาดเคลื่อนในการวัดของคะแนนการวัดหลังเรียนและก่อนเรียนมีค่าเท่ากัน

$$D_i = Y_i - X_i$$

เมื่อ D_i หมายถึง คะแนนการเปลี่ยนแปลง

Y_i หมายถึง คะแนนดิบผลการวัดหลังเรียน

X_i หมายถึง คะแนนดิบผลการวัดก่อนเรียน

การวัดคะแนนความแตกต่างระหว่างคะแนนดิบนี้เป็นวิธีที่คำนวณง่าย สะดวกและสามารถแปลความหมายได้ค่อนข้างชัดเจน จึงเป็นวิธีที่เหมาะสมสำหรับการวัดการเปลี่ยนแปลงที่มีการวัดซ้ำสองครั้ง แต่ก็ยังพบจุดอ่อนของการวัดการเปลี่ยนแปลงด้วยวิธีนี้

จุดอ่อนของวิธีการวัดความแตกต่างระหว่างคะแนนดิบ

นักจิตวิทยาหลายท่านพบว่าวิธีนี้มีจุดอ่อนที่สำคัญ 4 ประการ คือ (Lord. 1950: 429; Embretson. 1991: 496; citing Bereiter. 1963. Problems in Measuring Change. pp. 3-20; Cronbach; & Furby. 1970: 68-80; Linn; & Slinde. 1977: 121-150; May; & Nicewander. 1998: 882-897; Traub. 1994: 127-133; Willett; & Sayer. 1994: 363-365; Prieler; & Raven. 2002: 3-4; นงลักษณ์ วิรัชชัย. 2542: 259-261; อรุณี อ่อนสวัสดิ์. 2537: 24-28)

1. มีค่าสัมประสิทธิ์สหสัมพันธ์เป็นลบระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง หมายถึง ผู้สอบที่ได้คะแนนสอบก่อนเรียนสูงมีแนวโน้มที่จะได้คะแนนการเปลี่ยนแปลงน้อย ส่วนผู้สอบที่ได้คะแนนสอบก่อนเรียนต่ำมีแนวโน้มที่จะได้คะแนนการเปลี่ยนแปลงมาก ซึ่งในสภาพที่ควรจะเป็นก็คือ คะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงควรมีค่าสัมประสิทธิ์สหสัมพันธ์เป็นบวก เช่น

แบบทดสอบฉบับหนึ่งมีคะแนนเต็ม 100 คะแนน ผู้สอบคนแรกสอบก่อนเรียนได้คะแนน 70 คะแนน จะได้คะแนนการเปลี่ยนแปลงสูงสุดหรือศักยภาพในการทำคะแนนจากการวัดหลังเรียนสูงสุดเท่ากับ 30 คะแนน ส่วนผู้สอบคนที่สองสอบก่อนเรียนได้คะแนน 40 คะแนน จะได้คะแนนการเปลี่ยนแปลงสูงสุดหรือศักยภาพในการทำคะแนนจากการวัดหลังเรียนสูงสุดเท่ากับ 60 คะแนน จะเห็นได้ว่าผู้สอบคนแรก ซึ่งเป็นผู้ที่มีคะแนนการสอบก่อนเรียนสูงกว่าได้คะแนนการเปลี่ยนแปลงสูงสุดหรือศักยภาพในการทำคะแนนจากการวัดหลังเรียนสูงสุดน้อยกว่าคนที่สอง ทั้งๆ ที่ผู้สอบคนแรกควรจะได้คะแนนการเปลี่ยนแปลงมากกว่าคนที่สอง

2. ความเชื่อมั่นของคะแนนการวัดการเปลี่ยนแปลงมีค่าต่ำ เนื่องจากค่าสัมประสิทธิ์ของความเชื่อมั่นในคะแนนการเปลี่ยนแปลง เป็นอัตราส่วนระหว่างความแปรปรวนของคะแนนการเปลี่ยนแปลงจริงกับความแปรปรวนของคะแนนการเปลี่ยนแปลงสังเกต ($\rho_D^2 = \sigma_{T_D}^2 / \sigma_{X_D}^2$) โดยที่ความแปรปรวนของคะแนนการเปลี่ยนแปลงทั้งสองคือความแปรปรวนของคะแนนการเปลี่ยนแปลงจริงและความแปรปรวนของคะแนนการเปลี่ยนแปลงสังเกตต่างก็ประกอบด้วยฟังก์ชันของความแปรปรวนของคะแนนการวัดก่อนเรียน หลังเรียนและความแปรปรวนร่วมของผลการวัดทั้งสองครั้ง ($\sigma_{T_D}^2 = \sigma_{T_Y}^2 + \sigma_{T_X}^2 - 2\sigma_{T_{XY}}$) ดังนั้น ความเชื่อมั่นของคะแนนการเปลี่ยนแปลงจึงขึ้นอยู่กับค่าความสัมพันธ์ของความแปรปรวนคะแนนจริงจากการวัดก่อนเรียนและหลังเรียน และความแปรปรวนร่วมของผลการวัดก่อนเรียนและหลังเรียนด้วย นั่นคือ ถ้าค่าความสัมพันธ์ระหว่างคะแนนการวัดก่อนเรียนกับหลังเรียนสูงมากขึ้นแล้วค่าความเชื่อมั่นของคะแนนการเปลี่ยนแปลงก็จะต่ำลง ($\rho_{yx} = \sigma_{yx} / \sigma_y \sigma_x$) ซึ่งในสภาพที่ควรจะเป็นแล้วค่าความสัมพันธ์ระหว่างคะแนนการวัดก่อนเรียนกับหลังเรียนนั้นควรสูงทั้งนี้เพราะผู้สอบที่ได้คะแนนสอบก่อนเรียนสูงย่อมได้คะแนนหลังเรียนสูงด้วย

3. คะแนนการเปลี่ยนแปลงที่มีผลมาจากอิทธิพลคะแนนสูงสุด (Ceiling Effect) หมายถึง ผู้สอบที่ได้คะแนนการสอบก่อนเรียนสูง มีโอกาสน้อยมากที่จะได้คะแนนการเปลี่ยนแปลงสูง เพราะคะแนนสอบก่อนเรียนที่ทำได้เกือบเต็ม อาจเป็นเพราะวิธีการวัดการเปลี่ยนแปลงนี้ได้กำหนดแบบทดสอบที่ใช้ในการวัดก่อนเรียนและหลังเรียนจะต้องเป็นแบบทดสอบฉบับเดียวกันหรือคู่ขนานกัน ซึ่งจุดอ่อนประเด็นนี้อาจเป็นสาเหตุของจุดอ่อนประเด็นแรกด้วย คือ ทำให้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงเป็นลบ

4. การวัดคะแนนการเปลี่ยนแปลงโดยไม่คำนึงถึงคะแนนการวัดก่อนเรียนหรือพื้นความรู้เดิม (Floor Effect) ทำให้มีปัญหาในการแปลความคะแนนการเปลี่ยนแปลง เช่น นักเรียนคนแรกได้คะแนนสอบก่อนเรียน 10 คะแนน หลังเรียน 30 คะแนน ส่วนคนที่สองได้คะแนนสอบก่อนเรียน 70 คะแนน หลังเรียน 90 คะแนน เมื่อวัดการเปลี่ยนแปลงตามการวัดคะแนนความแตกต่างระหว่างคะแนนดิบแล้ว พบว่า นักเรียนทั้งสองคนได้คะแนนการเปลี่ยนแปลงเท่ากัน คือ 20 คะแนน ทั้งที่สภาพที่ควรจะเป็นนักเรียนคนที่สองควรมีการเปลี่ยนแปลงหรือการพัฒนามากกว่านักเรียนคนแรก เนื่องจาก

คะแนนการเปลี่ยนแปลง 20 คะแนน (จาก 70 คะแนนไปเป็น 90 คะแนน) ของนักเรียนคนที่สองเป็นคะแนนที่ต้องใช้ความสามารถสูงกว่าคะแนนการเปลี่ยนแปลง 20 คะแนน (จาก 10 คะแนนไปเป็น 30 คะแนน) ของนักเรียนคนแรก

ซึ่งปัญหาเหล่านี้เป็นปัญหาของการวัดการเปลี่ยนแปลงด้วยวิธีการวัดคะแนนความแตกต่างอย่างง่ายหรือความแตกต่างระหว่างคะแนนดิบ ถ้าในทางปฏิบัติเราละเลย ไม่ใส่ใจต่อปัญหาเหล่านี้จะทำให้การแปลความหมายของการเปลี่ยนแปลงผิดพลาด ไม่ถูกต้องแม่นยำ ต่อมา นักวิจัยวิทยาได้พัฒนาการวัดการเปลี่ยนแปลงเพื่อแก้ไขข้อบกพร่องเหล่านี้ของการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ

2. การวัดคะแนนการเปลี่ยนแปลงส่วนที่เหลือ (Residual Score)

เมนนิ่ง และดูบอยส์ (Manning; & DuBois. 1958: 191-194; 1962: 287-361) ได้เสนอวิธีคำนวณคะแนนการเปลี่ยนแปลงส่วนที่เหลือ โดยพิจารณาจากผลต่างระหว่างคะแนนสอบหลังเรียนกับคะแนนทำนายผลสอบหลังเรียนด้วยคะแนนสอบก่อนเรียน

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

เมื่อ R_i หมายถึง คะแนนการเปลี่ยนแปลงส่วนที่เหลือ

r_{xy} หมายถึง สัมประสิทธิ์สหสัมพันธ์ระหว่าง X กับ Y

ต่อมา ครอนบาค และเฟอร์บี้. (Cronbach; & Furby. 1970: 74) ได้พัฒนาวิธีการวัดคะแนนการเปลี่ยนแปลงของส่วนที่เหลือ (Residualized differences) เพิ่มเติม โดยใช้หลักการเดิมแต่ไม่ใช้ค่าคะแนนสอบก่อนเรียนและหลังเรียนที่เบี่ยงเบนไปจากคะแนนเฉลี่ยของคะแนนการวัดแต่ละครั้งมาพิจารณาร่วมด้วย เป็นการให้การวิเคราะห์การถดถอยเชิงเส้นโดยให้คะแนนสอบหลังเรียน (Y) เป็นตัวแปรตามและคะแนนสอบก่อนเรียน (X) เป็นตัวแปรต้น เป็นวิธีที่คำนึงถึงคะแนนสอบก่อนเรียนจึงได้นำมาเป็นฐานการคิดเพื่อทำนายคะแนนสอบหลังเรียน ซึ่งสามารถแก้ปัญหาพื้นความรู้เดิม (Floor effect) ดังนั้น เมื่อผู้สอบสองคนได้คะแนนสอบหลังเรียนเท่ากัน แต่คะแนนสอบก่อนเรียนแตกต่างกัน เมื่อนำมาคำนวณคะแนนการเปลี่ยนแปลงด้วยวิธีนี้จะได้คะแนนการเปลี่ยนแปลงที่ต่างกัน คือ คนที่ได้คะแนนสอบก่อนเรียนสูงจะได้คะแนนการเปลี่ยนแปลงสูงกว่าคนที่ได้คะแนนสอบก่อนเรียนต่ำ ซึ่งเป็นวิธีที่มีความเชื่อมั่นสูงกว่าวิธีการวัดความแตกต่างระหว่างคะแนนดิบ และสามารถกำจัดความสัมพันธ์ลวงระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงได้ แต่ถ้าคะแนนสอบก่อนเรียนและหลังเรียน

มีความสัมพันธ์กันสูงแล้ว ความเชื่อมั่นของคะแนนการเปลี่ยนแปลงจะต่ำเช่นเดียวกับวิธีการวัดความแตกต่างระหว่างคะแนนดิบ และยากแก่การแปลความหมายเนื่องจากค่าสัมประสิทธิ์การถดถอยที่ใช้นี้ผันแปรตามขนาดกลุ่มตัวอย่างทำให้มีความคลาดเคลื่อนรวมอยู่ด้วย ซึ่งคะแนนการเปลี่ยนแปลงที่ได้จากวิธีนี้สามารถบอกได้แต่เพียงว่าใครมีการเปลี่ยนแปลงมากกว่าหรือน้อยกว่าที่คาดหวังเท่านั้น

3. การวัดคะแนนการเปลี่ยนแปลงที่เป็นอิสระจากการวัดครั้งแรกที่เป็นฐาน (Base – Free Measure of Change)

ทักเกอร์ ดามาริน และ เมสสิก. (Tucker; Damarin; & Messick. 1966: 457-473) ได้เสนอวิธีคำนวณคะแนนการเปลี่ยนแปลงที่เป็นอิสระจากคะแนนการวัดก่อนเรียน โดยพิจารณาจากผลต่างระหว่างคะแนนดิบที่ได้จากการวัดหลังเรียนกับคะแนนทำนายผลสอบหลังเรียนด้วยคะแนนจริงจากคะแนนสอบก่อนเรียน

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

เมื่อ B_i หมายถึง คะแนนการเปลี่ยนแปลงที่เป็นอิสระจากการวัดครั้งแรกที่เป็นฐาน

$R_{xx'}$ หมายถึง ค่าความเชื่อมั่นของเครื่องมือวัดผลก่อนเรียน

$S_{Tx,Ty}$ หมายถึง ค่าความแปรปรวนร่วมระหว่างคะแนนจริงของ X และ Y

S_{Tx}^2 หมายถึง ค่าความแปรปรวนของคะแนนจริง X

เป็นวิธีที่สามารถกำจัดความสัมพันธ์ลวงระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง แต่ความคลาดเคลื่อนของคะแนนการเปลี่ยนแปลงที่เกิดขึ้นจากวิธีนี้มีแนวโน้มที่จะมีค่ามากกว่าวิธีการวัดความแตกต่างระหว่างคะแนนดิบ และวิธีคะแนนส่วนที่เหลือ

4. การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ (Relative Change Score)

ศิริชัย กาญจนวาสี (2538. 144-149; อ้างอิงจาก Kanjanawasee. 1989: Alternative strategies for policy analysis: An assessment of school effects on student's cognitive and affective mathematics outcomes in lower secondary school in Thailand. pp. 50-51) ได้เสนอวิธีนี้ซึ่งคะแนนการเปลี่ยนแปลงสัมพัทธ์ เป็นคะแนนความมอกงามเชิงสัมพัทธ์ที่คำนวณได้จากสัดส่วนของคะแนนการเปลี่ยนแปลงที่เกิดขึ้นจริงกับคะแนนการเปลี่ยนแปลงที่น่าจะเกิดขึ้นได้ทั้งหมด หมายถึง ผลต่าง

ระหว่างคะแนนดิบจากการวัดหลังเรียนกับก่อนเรียนหารด้วยผลต่างของคะแนนเต็มกับคะแนนดิบจากการวัดก่อนเรียน

$$S_i = \{(Y - X)/(F - X)\}100$$

เมื่อ S_i หมายถึง คะแนนการเปลี่ยนแปลงสัมพัทธ์

F หมายถึง คะแนนสูงสุดที่เป็นไปได้ (คะแนนเต็ม)

จะเห็นว่าคะแนนการเปลี่ยนแปลงที่ได้จากวิธีการนี้เป็นการลดอิทธิพลคะแนนสูงสุด (Ceiling Effect) และเป็นการคำนึงถึงความสามารถก่อนเรียนหรืออิทธิพลพื้นฐานความรู้เดิมด้วย ซึ่งเป็นคะแนนการเปลี่ยนแปลงที่แต่ละคนควรจะได้หรือศักยภาพของแต่ละคนควรได้มาเป็นตัวหาร ซึ่งศิริชัย กาญจนวาสี ได้วิจัยเพื่อการค้นหาดัชนีในการวัดตัวแปรเกี่ยวกับการประเมินอิทธิพลของโรงเรียนที่มีต่อความสามารถและเจตคติทางคณิตศาสตร์ของนักเรียนมัธยมศึกษาตอนต้น

ต่อมาในปี 2548 ธีระ อาชวเมธี; กมลวรรณ ตังธนาภานนท์; และ สมหวัง พิธิยานุวัฒน์ (Teera Archwamety; Kamonwan Tangdhanakanond; & Somwung Pitayanuwat. 2005: 213-227) ได้พัฒนาวิธีการคำนวณคะแนนเพิ่มจากวิธีคะแนนการเปลี่ยนแปลงสัมพัทธ์ของ ศิริชัย กาญจนวาสี โดยเรียกว่า Balanced Relative Gain Score ด้วยสูตร

$$\text{Balanced Relative Gain Score} = \frac{Y_2 - Y_1}{F - (Y_2 + Y_1)/2} 100$$

5. การวัดคะแนนเพิ่มที่แท้จริง (Estimated True Gain Score)

ลอร์ด (Traub. 1994. 135; citing Lord. 1956. Educational and Psychological Measurement. 421-437; 1958.; Educational and Psychological Measurement. 437-454.; 1963. Problems in measuring change.) และ แม็ก เนมาร์ (Mc Namar. 1958: 47-55) ได้เสนอวิธีคำนวณคะแนนเพิ่มที่แท้จริง โดยใช้การวิเคราะห์การถดถอยเชิงพหุ (Multiple Regression Analysis) ด้วยการทำนายผลต่างระหว่างคะแนนจริงจากการวัดหลังเรียนกับก่อนเรียน ที่ใช้คะแนนดิบจากการวัดหลังเรียนและก่อนเรียนเป็นตัวพยากรณ์

$$\begin{aligned} V_i &= K + W_x X_i + W_y Y_i \\ &= Y + b_{v_x, y} (X_i - X) + b_{v_y, x} (Y_i - Y) \end{aligned}$$

เมื่อ V_i หมายถึง คะแนนเพิ่มที่แท้จริง

K หมายถึง ค่าคงที่

W_x หมายถึง สัมประสิทธิ์ถดถอยพหุสำหรับการทำนาย ($T_y - T_x$) ด้วย X

W_y หมายถึง สัมประสิทธิ์ถดถอยพหุสำหรับการทำนาย ($T_y - T_x$) ด้วย Y

$b_{v_x,y}$ หมายถึง สัมประสิทธิ์ถดถอยพหุสำหรับการทำนาย ($T_y - T_x$) ด้วย X

เมื่อควบคุม Y

$b_{v_y,x}$ หมายถึง สัมประสิทธิ์ถดถอยพหุสำหรับการทำนาย ($T_y - T_x$) ด้วย

Y เมื่อควบคุม X

การวัดคะแนนเพิ่มที่แท้จริง ใช้หลักการวิเคราะห์การถดถอย (Regression) ที่มีคะแนนดิบก่อนเรียน (X) และคะแนนดิบหลังเรียน (Y) เป็นตัวแปรต้นในการประมาณค่าความสามารถเพิ่มที่แท้จริงรายบุคคลได้ดี คือ คำนึงถึงพื้นฐานความรู้เดิม ทำให้การแปลความหมายของคนที่มีคะแนนการเปลี่ยนแปลงที่เท่ากันแต่มีความสามารถก่อนเรียนแตกต่างกันได้ถูกต้อง นั่นคือ เมื่อผู้สอบสองคนมีคะแนนการเปลี่ยนแปลงเท่ากันแล้วคนที่มีความสามารถก่อนเรียนสูงกว่าจะได้คะแนนการเปลี่ยนแปลงมากกว่าด้วย แต่พบว่ายังเป็นวิธีที่มีจุดอ่อนเกี่ยวกับอิทธิพลคะแนนสูงสุด ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ มีความยุ่งยากในการคำนวณ และคะแนนที่ได้ขึ้นอยู่กับกลุ่มตัวอย่าง

6. การวัดการเปลี่ยนแปลงการเรียนรู้ (อรุณี อ่อนสวัสดิ์, 2537: 70-77; 2538: 1-15)

อรุณี อ่อนสวัสดิ์ ได้พัฒนาวิธีวัดการเปลี่ยนแปลงการเรียนรู้ที่มีการสอบสองครั้ง คือก่อนเรียนและหลังเรียน โดยอาศัยแนวคิดตามทฤษฎีการเรียนรู้ของบลูมและอิทธิพลคะแนนสูงสุด ซึ่งแนวคิดการเปลี่ยนแปลงการเรียนรู้ เป็นฟังก์ชันของพื้นฐานความรู้เดิม การเรียนการสอนและอิทธิพลคะแนนสูงสุด (Ceiling effect) ทำให้ได้แนวทางการปฏิบัติเป็น 2 แนวทาง คือ 1) วัดด้วยค่าประมาณคะแนนจริง และ 2) วัดด้วยคะแนนดิบ ซึ่ง อรุณี อ่อนสวัสดิ์ (2537: 62) ได้เสนอวิธีวัดการเปลี่ยนแปลงการเรียนรู้แปดวิธีดังนี้

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

$$W_{2.1} = \frac{(S_x^2 R_{xx'} + S_y^2 R_{yy'} + 2r_{xy}S_xS_y)^{\frac{1}{2}}}{S_y} \frac{r_{DX}r_{DY} - R_{XX'}r_{DY}}{\sqrt{R_{DD}(R_{XX'}R_{YY'} - r_{XY}^2)}}$$

ซึ่งอรุณี อ่อนสวัสดิ์ ได้นำวิธีที่พัฒนาขึ้นมาทั้งแปดวิธีไปเปรียบเทียบกับวิธีอื่น พบว่าวิธีที่พัฒนาขึ้นทั้งแปดวิธีมีค่าความเชื่อมั่นและค่าความคลาดเคลื่อนไม่แตกต่างจากวิธีการวัดคะแนนความแตกต่างอย่างง่ายหรือการวัดคะแนนความแตกต่างระหว่างคะแนนดิบและวิธีการวัดคะแนนเพิ่มที่แท้จริงของลอร์ด แต่วิธีของอรุณีมีการคำนวณที่ค่อนข้างยุ่งยากมากกว่า

ต่อมาอรุณี อ่อนสวัสดิ์ ได้พัฒนาวิธีการวัดการเปลี่ยนแปลงขึ้นมาอีกแปดวิธีจากสี่แนวคิด ที่คิดจากค่าประมาณคะแนนจริงและคิดจากค่าคะแนนดิบ แต่ยังไม่มีการพิสูจน์ โดยมีแนวคิดทั่วไปในการวัดการเปลี่ยนแปลง คือ ความแตกต่างระหว่างคะแนนจริง เกิดจากการเปลี่ยนแปลงเนื่องจากพื้นฐานความรู้เดิม การเปลี่ยนแปลงเนื่องจากการเรียนการสอนและลบอิทธิพลคะแนนสูงสุดออกไป ดังแนวคิดทั้งสี่ของ อรุณี อ่อนสวัสดิ์ (2538: 13-15) ต่อไปนี้

แนวคิดที่ 1 อิทธิพลคะแนนสูงสุดขึ้นกับคะแนนจริงหลังเรียน และถือว่าผู้ที่ได้คะแนนหลังเรียนเท่ากันจะได้รับอิทธิพลคะแนนสูงสุดเท่ากัน โดยใช้สัมประสิทธิ์ถดถอยอย่างง่าย ที่ได้จากคะแนนจริง ความแตกต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียน ถดถอยลงบนคะแนนจริงความแตกต่างระหว่างคะแนนเต็มและคะแนนหลังเรียน จึงได้วิธีวัดสองวิธี ที่คิดจากค่าประมาณคะแนนจริงในสูตร 1 และคิดจากค่าคะแนนดิบในสูตร 2 คือ

$$1. \quad V1A_i = (1 - W_2) \hat{T}_{Yi} - (1 + W_1) \hat{T}_{Xi}$$

$$2. \quad V1B_i = (1 - W_2) Y_i - (1 + W_1) X_i$$

แนวคิดที่ 2 อิทธิพลคะแนนสูงสุด ขึ้นกับคะแนนจริงหลังเรียนและคะแนนจริงก่อนเรียน เนื่องจากผู้ที่ได้คะแนนหลังเรียนเท่ากันไม่จำเป็นต้องได้รับอิทธิพลคะแนนสูงสุดเท่ากัน ตัวแปรอิทธิพลจึงมีสองตัว คือ 1) สัมประสิทธิ์ถดถอยอย่างง่าย ที่ได้จากคะแนนความแตกต่างระหว่างคะแนนจริงหลังเรียนและก่อนเรียน ถดถอยลงบนคะแนนจริงความแตกต่างระหว่างคะแนนเต็มและคะแนนหลังเรียน และ 2) ใช้สัมประสิทธิ์ถดถอยอย่างง่าย ที่ได้จากคะแนนความแตกต่างระหว่างคะแนนจริงก่อนเรียน

และหลังเรียน ถดถอยลงบนคะแนนจริง ความแตกต่างระหว่างคะแนนเต็มและคะแนนสอบก่อนเรียน จึงได้วิธีวัดสองวิธี ที่คิดจากค่าประมาณคะแนนจริงในสูตร 3 และคิดจากค่าคะแนนดิบในสูตร 4 คือ

$$3. \quad V2A_i = (1 - W_2) \hat{T}_{Yi} - (1 + W_1 - W_3) \hat{T}_{Xi}$$

$$4. \quad V2B_i = (1 - W_2) Y_i - (1 + W_1 - W_3) X_i$$

แนวคิดที่ 3 อิทธิพลคะแนนสูงสุดขึ้นอยู่กับคะแนนจริงหลังเรียนและคะแนนจริงก่อนเรียน เนื่องจากผู้ที่ได้คะแนนหลังเรียนเท่ากันไม่จำเป็นต้องได้รับอิทธิพลคะแนนสูงสุดเท่ากัน จึงใช้สัมประสิทธิ์ถดถอยพหุที่ได้จากคะแนนความแตกต่างระหว่างคะแนนจริงหลังเรียนและก่อนเรียน ถดถอยลงบนคะแนนจริง ความแตกต่างระหว่างคะแนนเต็มและคะแนนหลังเรียน เมื่อควบคุมคะแนนจริงก่อนเรียน จึงได้วิธีวัดสองวิธี ที่คิดจากค่าประมาณคะแนนจริงในสูตร 5 และคิดจากค่าคะแนนดิบในสูตร 6 คือ

$$5. \quad V3A_i = (1 - W_{2.1}) \hat{T}_{Yi} - (1 + W_1) \hat{T}_{Xi}$$

$$6. \quad V3B_i = (1 - W_{2.1}) Y_i - (1 + W_1) X_i$$

แนวคิดที่ 4 อิทธิพลคะแนนสูงสุดขึ้นกับคะแนนจริงหลังเรียนและคะแนนจริงก่อนเรียน เนื่องจากผู้ที่ได้คะแนนหลังเรียนเท่ากันไม่จำเป็นต้องได้รับอิทธิพลคะแนนสูงสุดเท่ากัน แต่อิทธิพลคะแนนสูงสุดน่าจะเกิดขึ้นตามสัดส่วนของคะแนนจริงก่อนเรียน กล่าวคือ ถ้าคะแนนจริงหลังเรียนเท่ากัน ผู้ที่ได้คะแนนจริงก่อนเรียนต่ำกว่าจะได้รับอิทธิพลคะแนนสูงสูดน้อยกว่าผู้ที่ได้คะแนนจริงก่อนเรียนสูงกว่า จึงใช้สัมประสิทธิ์ถดถอยอย่างง่ายที่ได้จากคะแนนความแตกต่างระหว่างคะแนนจริงก่อนเรียนและหลังเรียน ถดถอยลงบนคะแนนจริง ความแตกต่างระหว่างคะแนนเต็มและคะแนนหลังเรียน หารด้วยคะแนนจริงก่อนเรียน จึงได้วิธีวัดสองวิธี ที่คิดจากค่าประมาณคะแนนจริงในสูตร 7 และคิดจากค่าคะแนนดิบในสูตร 8 คือ

$$7. \quad V4A_i = (1 - W_2 / \hat{T}_{Xi}) \hat{T}_{Yi} - (1 + W_1) \hat{T}_{Xi} + W_2 F / \hat{T}_{Xi}$$

$$8. \quad V4B_i = (1 - W_2 / X_i) Y_i - (1 + W_1) X_i + W_2 F / X_i$$

โดยที่ W_1 คือ สัมประสิทธิ์ถดถอย $T_y - T_x$ ลงบน T_x

$$W_1 = r_{Dx} S_D / R_{xx} S_x$$

ส่วนสัมประสิทธิ์อื่นๆ มีค่าเหมือนที่กล่าวมาแล้ว

7. การวัดคะแนนเพิ่มด้วยวิธีเทียบส่วน (วินิจ เทือกทอง. 2537: 21-24)

วินิจ เทือกทอง ได้พัฒนาวิธีการวัดคะแนนเพิ่มโดยประยุกต์วิธีดั้งเดิม วิธีเศษเหลือ และวิธีของลอว์ด์เข้ากับวิธีของศิริชัย และวิธีเทียบส่วนร้อย ดังนี้

7.1 วิธีของศิริชัย เป็นวิธีที่เกิดจากการประยุกต์วิธีดั้งเดิม วิธีเศษเหลือ และวิธีของลอร์ด เข้ากับวิธีของศิริชัยโดยการนำคะแนนเพิ่มจากวิธีทั้งสามมาเทียบกับศักยภาพของผู้เข้าสอบตามวิธีของศิริชัย ดังนั้น การคำนวณหาค่าคะแนนเพิ่มด้วยแนวคิดนี้จึงมีสามวิธี คือ

7.1.1 คะแนนเพิ่มวิธีดั้งเดิมเทียบกับศักยภาพของผู้เข้าสอบ

$$\begin{aligned} \text{SRG} &= (\text{RG}/\text{F} - \text{X})100 \\ &= (\text{Y} - \text{X}/\text{F} - \text{X})100 \end{aligned}$$

7.1.2 คะแนนเพิ่มวิธีเศษเหลือเทียบกับศักยภาพของผู้เข้าสอบ

$$\begin{aligned} \text{SDG} &= (\text{DG}/\text{F} - \text{X})100 \\ &= (\text{Y} - b_w \text{X}/\text{F} - \text{X})100 \end{aligned}$$

7.1.3 คะแนนเพิ่มวิธีของลอร์ดเทียบกับศักยภาพของผู้เข้าสอบ

$$\begin{aligned} \text{SLG} &= (\text{LG}/\text{F} - \text{X})100 \\ &= (\bar{G} + b_{\text{GX.Y}}(\text{X} - \bar{X}) + b_{\text{GY.X}}(\text{Y} - \bar{Y})/\text{F} - \text{X})100 \end{aligned}$$

7.2 วิธีเทียบส่วนร้อย เป็นวิธีที่เกิดจากการประยุกต์วิธีดั้งเดิม วิธีเศษเหลือ และวิธีของลอร์ดเข้ากับวิธีเทียบส่วนร้อย โดยการนำคะแนนเพิ่มจากวิธีทั้งสามมาเทียบกับคะแนนสอบก่อนเรียนของแต่ละคน ซึ่งใช้เป็นฐาน ดังนั้น การคำนวณหาค่าคะแนนเพิ่มด้วยแนวคิดนี้จึงมีสามวิธี คือ

7.2.1 คะแนนเพิ่มวิธีดั้งเดิมเทียบกับคะแนนสอบก่อนเรียน

$$\begin{aligned} \text{PRG} &= (\text{RG}/\text{X})100 \\ &= (\text{Y} - \text{X}/\text{X})100 \end{aligned}$$

7.2.2 คะแนนเพิ่มวิธีเศษเหลือเทียบกับคะแนนสอบก่อนเรียน

$$\begin{aligned} \text{PDG} &= (\text{DG}/\text{X})100 \\ &= (\text{Y} - b_w \text{X}/\text{X})100 \end{aligned}$$

7.2.3 คะแนนเพิ่มวิธีของลอร์ดเทียบกับคะแนนสอบก่อนเรียน

$$\begin{aligned} \text{PLG} &= (\text{LG}/\text{X})100 \\ &= (\bar{G} + b_{\text{GX.Y}}(\text{X} - \bar{X}) + b_{\text{GY.X}}(\text{Y} - \bar{Y})/\text{X})100 \end{aligned}$$

เมื่อ Y หมายถึง คะแนนดิบจากการวัดหลังเรียน

X หมายถึง คะแนนดิบจากการวัดก่อนเรียน

$\bar{G}$ หมายถึง คะแนนเฉลี่ยของกลุ่มตัวอย่าง

b_w หมายถึง สัมประสิทธิ์ของการถดถอย

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

วินิจ เพื่อทบทวน ได้นำวิธีการทั้งหมดหกวิธี ที่พัฒนาขึ้นไปเปรียบเทียบความคลาดเคลื่อนในการวัดของวิธีดั้งเดิม วิธีพิเศษเหลือ วิธีของลอร์ด พบว่าในกรณีที่ไม่คำนึงถึงอิทธิพลคะแนนสูงสุดวิธีดั้งเดิม คือ วิธีการวัดคะแนนความแตกต่างอย่างง่ายหรือความแตกต่างระหว่างคะแนนดิบ และวิธีของลอร์ด มีค่าความคลาดเคลื่อนในการวัดน้อยที่สุด ส่วนในกรณีที่ไม่คำนึงถึงอิทธิพลคะแนนสูงสุดวิธีพิเศษเหลือเทียบกับศักยภาพของผู้สอบมีค่าความคลาดเคลื่อนน้อยที่สุด

8. การวัดการเปลี่ยนแปลงของตำแหน่งเปอร์เซ็นต์ไทล์ (Change in Percentile Rank)

(Russell. 2001: Online)

รัสเซล (Russell) ได้เสนอการเปลี่ยนแปลงที่เน้นการเพิ่มขึ้นหรือการลดลงของตำแหน่งเปอร์เซ็นต์ไทล์เฉลี่ย (Mean percentile ranking) เป็นวิธีที่ได้รับความนิยมเนื่องจากคำนวณได้ง่าย แต่พบจุดบกพร่องของวิธีนี้สองประการ คือ 1) การคำนวณค่าตำแหน่งเปอร์เซ็นต์ไทล์เฉลี่ย (Mean percentile rank) ขึ้นอยู่กับตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) รายบุคคล ซึ่งทำให้การประมาณค่าเฉลี่ยของกลุ่มไม่ถูกต้อง และ 2) ตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) ที่แตกต่างกันของช่วงที่ไม่เท่ากัน ทำให้ไม่สามารถบอกปริมาณการเปลี่ยนแปลงที่แท้จริงได้ ซึ่งกลาส และฮอบคิน (Russell. 2001: Online; citing Glass; & Hopkins. 1984. Statistical Methods in Education and Psychology.) ได้อธิบายถึงจุดบกพร่องทั้งสองประการนี้ คือ ประการแรก การนำค่าเฉลี่ยของการจัดอันดับ (Averaging ranks) ไปใช้เป็นค่าเฉลี่ยการจัดอันดับของกลุ่ม (Mean group rank) ไม่ถูกต้อง เนื่องจากตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) เป็นการวัดข้อมูลอันดับ (Ordinal) ซึ่งปริมาณการวัดคุณลักษณะแสดงการเปลี่ยนแปลงด้วยการเพิ่มขึ้นของแต่ละอันดับหนึ่งหน่วย เนื่องจากช่วงระหว่างตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) แต่ละตำแหน่งไม่เท่ากัน (Unequal intervals) ดังนั้น การใช้ตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) ของแต่ละบุคคลเพื่อพิจารณาตัดสินนั้นย่อมได้ผลในการประมาณค่าเฉลี่ยของกลุ่มไม่ถูกต้อง ดังตัวอย่างการจัดอันดับและเวลาที่ใช้ของรถด่วนห้าขบวน

จะเห็นว่าการจัดอันดับและเวลาที่ใช้ของรถด่วน (Sprinter) ห้าขบวนสอดคล้องกัน ในตัวอย่างนี้คุณลักษณะที่วัดคือ ความเร็วที่วิ่ง โดยที่การจัดอันดับเฉลี่ย (Mean rank) ของกลุ่มรถด่วนเป็นอันดับที่สาม จากการบันทึกข้อมูลความเร็วของรถด่วนที่ถึงเส้นชัยเป็นลำดับที่สามมีความเร็วเฉลี่ยช้ากว่า

ความเร็วเฉลี่ยของกลุ่มมาก เนื่องจากการแบ่งช่วงระยะเวลาออกจากกันไม่เท่ากัน (Unequal time intervals separating) การใช้การจัดอันดับอย่างง่ายสำหรับการพิจารณาตัดสินผลการจัดอันดับเฉลี่ยในการประมาณค่าจึงไม่ถูกต้องสำหรับความเร็วที่วิ่งโดยเฉลี่ยของกลุ่ม

อันดับ	เวลาที่ใช้
1	10.2
2	10.3
3	10.5
4	11.2
5	12.4
อันดับเฉลี่ย	เวลาเฉลี่ย
3	10.9

ที่มา: Russell, M. (2001). Online.

การแก้ปัญหาวิธีหนึ่งเกี่ยวกับการคำนวณทางคณิตศาสตร์จากการจัดอันดับ (Ordinal ranks) คือ ตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank ; PR) แต่ละตำแหน่ง เพื่อพิจารณาตัดสินค่าเฉลี่ยของกลุ่ม และการพิจารณาตัดสินตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile rank) ที่สอดคล้องกับคะแนนเฉลี่ยของกลุ่ม ดังตารางต่อไปนี้แสดงการใช้คะแนนมาตรฐานของนักเรียน (Standard scores) หรือคะแนนเทียบเท่าในโค้งปกติ (Normal Curve Equivalent <NCE> scores) เพื่อบรรณค่าเฉลี่ยของกลุ่มและค้นหาตำแหน่งเปอร์เซ็นต์ไทล์ที่สอดคล้องกับค่าเฉลี่ยอื่นๆ ที่ให้ค่าเฉลี่ยของตำแหน่งเปอร์เซ็นต์ไทล์ที่ประมาณค่าแทนบุคคลมากกว่าผลสัมฤทธิ์ทางการเรียนเฉลี่ยของกลุ่ม

คะแนนมาตรฐาน คะแนนเทียบเท่าในโค้งปกติ และตำแหน่งเปอร์เซ็นต์ไทล์ ของนักเรียนเกรด 3 ของการสอบด้วยแบบทดสอบวัดทักษะพื้นฐานด้านภาษาของ Iowa จะเห็นว่าตำแหน่งเปอร์เซ็นต์ไทล์เฉลี่ยที่ 27.2 ตรงกับคะแนนมาตรฐานที่มีค่าใกล้เคียงคะแนนเฉลี่ยที่แท้จริงสำหรับกลุ่มนักเรียนมีค่าเท่ากับ 159.3 ซึ่งตรงกับตำแหน่งเปอร์เซ็นต์ไทล์ที่ใกล้เคียงกับ 21 ในตัวอย่างนี้ ตำแหน่งเปอร์เซ็นต์ไทล์เฉลี่ยประมาณค่าผลสัมฤทธิ์ทางภาษาเฉลี่ยของกลุ่มเกินจริงและชี้ให้เห็นว่านักเรียนโดยเฉลี่ยอยู่สูงกว่าคะแนนมาตรฐานเฉลี่ยจำนวน 6 เปอร์เซ็นต์ไทล์

คนที่	คะแนนมาตรฐาน	คะแนนเทียบเท่าในโค้งปกติ	อันดับเปอร์เซ็นต์ไทล์
1	127	1	1
2	143	10	3
3	151	23	10
4	161	35	24
5	164	39	30

ที่มา: Russell, M. (2001). Online.

จุดอ่อนประการที่สองของวิธีนี้ก็คือ ความแตกต่างระหว่างเปอร์เซ็นต์ไทล์เฉลี่ยเป็นคะแนนลง ซึ่งได้อธิบายไว้ว่า เมื่อตำแหน่งเปอร์เซ็นต์ไทล์เฉลี่ยถูกใช้คำนวณคะแนนมาตรฐานของนักเรียน หรือ คะแนนเทียบเท่าในโค้งปกติ คะแนนความแตกต่างระหว่างตำแหน่งเปอร์เซ็นต์ไทล์เฉลี่ยของกลุ่มทำให้เกิดความผิดพลาด เพราะปริมาณการเปลี่ยนแปลงที่เท่ากันแทนด้วยปริมาณพัฒนาการที่เท่ากันของทุก ครั้งการสอบบนมาตรวัดที่เป็นเปอร์เซ็นต์ไทล์ ในความเป็นจริงตำแหน่งเปอร์เซ็นต์ไทล์ที่เบี่ยงเบนห่างออกไปจากค่าเฉลี่ยมากขึ้น คะแนนของนักเรียนคนหนึ่งต้องเพิ่มขึ้นมากกว่าการเพิ่มตำแหน่งเปอร์เซ็นต์ไทล์ของเขา ความสัมพันธ์นี้เป็นผลโดยตรงของการแจกแจงคะแนนภายใต้โค้งปกติ ซึ่งในโค้งปกติจำนวนที่ไม่ได้สัดส่วนของคะแนนประชากรมีค่าใกล้เคียงค่าเฉลี่ย จะเกิดการเปลี่ยนแปลงในคะแนนการทดสอบของบุคคลที่มีค่าใกล้เคียงเล็กน้อยและเกิดการเปลี่ยนแปลงใหญ่มากในความสัมพันธ์ของอันดับของเขา กับผู้สอบคนอื่นๆ เมื่อเปรียบเทียบกับ การเปลี่ยนแปลงที่เท่ากันที่อยู่ค่าปลายๆ ของการกระจาย

ตัวอย่างในแบบทดสอบวัดทักษะพื้นฐานของ Iowa (ITBS) ฉบับย่อที่เป็นด้านภาษา คะแนนมาตรฐานของนักเรียนเกรด 3 ต้องเพิ่มขึ้น 7 คะแนน ที่เกี่ยวกับการเลื่อนจากตำแหน่งเปอร์เซ็นต์ไทล์ที่ 10 ไปยังเปอร์เซ็นต์ไทล์ที่ 20 ในขณะที่การเลื่อนตำแหน่งเปอร์เซ็นต์ไทล์จากตำแหน่งที่ 50 ไปยังตำแหน่งที่ 60 คะแนนมาตรฐานของนักเรียนจำเป็นต้องเพิ่มเพียง 4 คะแนน เท่านั้น ในทำนองเดียวกัน การเพิ่ม 10 คะแนนในตำแหน่งเปอร์เซ็นต์ไทล์ที่ 10 มีค่าเบี่ยงเบนมาตรฐานประมาณ 0.44 ในขณะที่คะแนนเพิ่มขึ้น 10 คะแนนของตำแหน่งเปอร์เซ็นต์ไทล์ 50 มีค่าเบี่ยงเบนมาตรฐานเพียง 0.25 เท่านั้น การเปลี่ยนแปลงการปฏิบัติงานหรือผลการสอบของนักเรียนนี้มากเกินไปจนความจริงหรือน้อยเกินความจริง นั้นขึ้นอยู่กับตำแหน่งเปอร์เซ็นต์ไทล์ของนักเรียนหรือกลุ่มนักเรียนนั่นเอง

9. การวัดคะแนนการเปลี่ยนแปลงเป็นร้อยละ (Percent Change) (Russell. 2001: 4-5)

รัสเซล (Russell) ได้อธิบายการวัดคะแนนการเปลี่ยนแปลงเป็นร้อยละ ไว้ว่าเป็นคะแนนสอบหลังเรียนที่เบี่ยงเบนหรือห่างออกไปจากคะแนนสอบก่อนเรียน เช่น โรงเรียนเอกชนแห่งหนึ่งได้รายงานคะแนนการสอบได้มีการพัฒนา ปรับปรุง ร้อยละ 20–30 ดังนั้นเพื่อให้ได้คะแนนพัฒนาการตามที่กล่าวจึงนำคะแนนการสอบในปีปัจจุบันหารด้วยคะแนนการสอบในปีที่ผ่านมา จะได้การเปลี่ยนแปลงเป็นร้อยละ วิธีนี้เน้นที่การเปลี่ยนแปลงคะแนนมาตรฐานหรือ คะแนนเทียบเท่าในโค้งปกติให้เป็นร้อยละซึ่งเป็นวิธีที่ดีกว่าวิธีตำแหน่งเปอร์เซ็นต์ไทล์ อย่างไรก็ตาม วิธีนี้กำหนดให้คะแนนที่ใช้เพื่อพิจารณาตัดสินการเปลี่ยนแปลงเป็นร้อยละซึ่งเป็นมาตรวัดข้อมูลที่เป็นอัตราส่วน (Ratio scale) ทั้งคะแนนมาตรฐานและคะแนนเทียบเท่าในโค้งปกติต่างก็เป็นข้อมูลการวัดที่เป็นอันตรภาค (Interval) ในขณะที่วิธีที่ใช้ตำแหน่งเปอร์เซ็นต์ไทล์เป็นข้อมูลการวัดที่เป็นอันดับ (Ordinal) อย่างชัดเจน ดังที่ กลาส และฮอบคิน (Russell. 2001: Online; citing Glass; & Hopkins. 1984. Statistical Methods in Education and Psychology.) อธิบายได้อย่างสมบูรณ์ชัดเจนมากกว่าที่กล่าวถึงข้อมูลการวัดที่เป็นอัตราส่วนหรือข้อมูลการวัดที่เป็นอันตรภาค ส่วนข้อมูลการวัดที่เป็นอันดับนั้นไม่มีความหมาย

การวัดคะแนนการเปลี่ยนแปลงเป็นร้อยละจะให้ค่าที่ไม่เป็นจริง เป็นค่าลงตาไปกับคะแนนสอบก่อนเรียนหรือ ผลการปฏิบัติงานเริ่มต้นน้อย เช่น โรงเรียนสองแห่งมีความรู้ของนักเรียนเฉลี่ยเพิ่มขึ้น โดยมีคะแนนเฉลี่ยสูงขึ้นเท่ากับ 5 คะแนนทั้งสองแห่ง โรงเรียน A มีคะแนนเฉลี่ยเพิ่มขึ้นจาก 20 คะแนนไปเป็น 25 คะแนน ในขณะที่โรงเรียน B มีคะแนนเฉลี่ยเพิ่มขึ้นจาก 50 คะแนนไปเป็น 55 คะแนน แม้ว่าปริมาณการเปลี่ยนแปลงเท่ากัน แต่วิธีการเปลี่ยนแปลงเป็นร้อยละสามารถบอกได้ว่า โรงเรียน A ปรับปรุงไป ร้อยละ 25 ในขณะที่โรงเรียน B ปรับปรุงไปเพียงร้อยละ 10 เท่านั้น

วิลเลท (Russell. 2001: Online; citing Willet. 1988. Review of Research in Education. pp. 345-422) ได้พบว่า การเปลี่ยนแปลงทุกวิธีมีปัญหาความเชื่อมั่นของคะแนนการเปลี่ยนแปลงต่ำ แต่วิธีที่ต้องการคือวิธีที่ประมาณค่าคะแนนการเปลี่ยนแปลงที่ได้มาตรฐาน เป็นคะแนนที่มีความหมายเหมือนกันในทุกคะแนนที่สอบบนมาตรวัด และสามารถเปรียบเทียบระหว่างแบบทดสอบและระดับชั้นเรียนได้ โดยมีค่าความเชื่อมั่นพอประมาณ

การวัดการเปลี่ยนแปลงแต่ละวิธีตามแนวคิดการวัดการเปลี่ยนแปลงแบบมาตรฐานเดิม (Classical measurement) ที่เป็นทฤษฎีการทดสอบแบบมาตรฐานเดิมนั้นมีทั้งจุดเด่นและจุดด้อยสามารถสรุปได้ดังนี้

วิธี	ผู้พัฒนาและปีที่พัฒนา	จุดเด่น	จุดด้อย
1. การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ (Difference score)	-	เป็นวิธีที่มีเหตุผลตรงไปตรงมา คำนวณและแปลความได้ง่าย	1. มีความสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงกับคะแนนสอบก่อนเรียนเป็นลบ 2. มีความเชื่อมั่นต่ำ 3. มีอิทธิพลคะแนนสูงสุด 4. มีอิทธิพลพื้นฐานความรู้เดิม
2. การวัดคะแนนการเปลี่ยนแปลงส่วนที่เหลือ (Residual score)	Manning; & Dubois. 1958; Cronbach; & Furby, 1970	1. กำจัดความสัมพันธ์เชิงลบระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง 2. มีการนำเอาคะแนนพื้นฐานความรู้เดิมหรือคะแนนสอบก่อนเรียนมาคำนวณ	ค่าสัมประสิทธิ์การถดถอยที่ใช้นี้มีค่าความคลาดเคลื่อนรวมอยู่ด้วยทำให้ยากแก่การแปลความหมาย
3. การวัดคะแนนการเปลี่ยนแปลงที่เป็นอิสระจากการวัดครั้งแรกที่เป็นฐาน (Base – free measure of change)	Tucker; Damarin; & Messick. 1966	ปรับให้คะแนนการเปลี่ยนแปลงไม่สัมพันธ์กับคะแนนจริงก่อนเรียน	ยากในการแปลความ ไม่สามารถบอกปริมาณการเปลี่ยนแปลงของแต่ละบุคคล แม้บอกได้ว่าใครมีการเปลี่ยนแปลงมากหรือน้อยกว่าที่ทำนายได้จากคะแนนจริงก่อนเรียนเพียงใด
4. การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ (Relative change score)	ศิริชัย กาญจนวาสี (Kanjana-wasee) 1989	ขจัดอิทธิพลคะแนนสูงสุด โดยการใช้อิทธิพลคะแนนสูงสุดหรือศักยภาพของผู้เข้าสอบมาคำนวณ	ไม่มีวิธีประมาณค่าความเชื่อมั่น เนื่องจากไม่ใช่โมเดลเชิงบวก
5. การวัดคะแนนการเปลี่ยนแปลงแท้จริง (Estimated true gain score)	Lord. 1956; Mc Namar. 1958	1. ขจัดอิทธิพลคะแนนสูงสุด 2. สามารถอธิบายได้ว่าใครมีการเปลี่ยนแปลงมากหรือน้อยกว่าได้	1. มีความยุ่งยาก ซับซ้อน 2. มีอิทธิพลคะแนนสูงสุด 3. ค่าสัมประสิทธิ์การถดถอยที่ใช้มีค่าความคลาดเคลื่อนรวมอยู่ด้วย

วิธี	ผู้พัฒนาและ ปีที่พัฒนา	จุดเด่น	จุดด้อย
6. การวัดการเปลี่ยนแปลงการเรียนรู้	อรุณี อ่อนสวัสดิ์. 2537 และ 2538	1. กำจัดอิทธิพลคะแนน สูงสุด 2. คำนึงถึงคะแนนพื้นฐานความรู้ เดิม	มีความยุ่งยากซับซ้อน
7. การวัดคะแนนเพิ่ม ของวินิจ	วินิจ เทือกทอง. 2537	แก้ไขอิทธิพลคะแนนสูงสุด ด้วยการนำคะแนนการเปลี่ยนแปลงด้วยวิธีต่างๆ ไปเทียบ กับคะแนนสอบก่อนเรียน และศักยภาพของผู้สอบ	ไม่มีวิธีประมาณค่าความ เชื่อมั่น เนื่องจากไม่ใช่ โมเดลเชิงบวก
8. การวัดการ เปลี่ยนแปลงของ ตำแหน่งเปอร์เซ็นต์ไทล์ (Change in percentile rank)	Russell. 2001	เป็นการวัดข้อมูลที่เป็นอันดับ (ordinal)	ขึ้นอยู่กับตำแหน่งเปอร์เซ็นต์ไทล์ รายบุคคล ซึ่งทำให้การ ประมาณค่าเฉลี่ยของกลุ่ม ไม่ถูกต้อง
9. การวัดคะแนนการ เปลี่ยนแปลงเป็นร้อยละ (Percent change)	Russell. 2001	เป็นการวัดข้อมูลที่เป็น อันดับ และ ข้อมูลที่เป็น อัตราส่วน	จะให้ค่าที่ไม่เป็นจริง เมื่อ คะแนนสอบก่อนเรียนต่ำ

โมเดลพหุระดับสำหรับการวัดการเปลี่ยนแปลง

โมเดลพหุระดับ (Multilevel model) ที่เรียกชื่อแตกต่างกันหลายชื่อตามผลงานวิชาการของนักวิจัย ได้แก่ โมเดลเชิงเส้นระดับลดหลั่น (Hierarchical linear model; HLM) (Lindley; & Smith. 1972; Bry; & Raudenbush. 1992) หรือ โมเดลการถดถอยสัมประสิทธิ์แบบสุ่ม (Random-coefficient regression model) (Rosenberg. 1973) หรือ โมเดลส่วนประกอบความแปรปรวนร่วม (Covariances components model) (Longford. 1987) หรือโมเดลอิทธิพลผสมหรือโมเดลอิทธิพลสุ่ม (Mixed-effects model หรือ Random-effect model) (Laird; & Ware. 1982) โมเดลเชิงเส้นพหุระดับ (Multi-level linear model) (Mason. 1983; & Goldstein. 1987) มีแนวคิดในการนำโครงสร้างระดับข้อมูลมาพิจารณาในการวิเคราะห์ โดยข้อมูลมีลักษณะที่ซ้อนกันและเป็นระดับลดหลั่น (Hierarchical nested data) หรือเป็นข้อมูลหลายระดับ (Multi-level data) เช่น นักเรียนเป็นหน่วยย่อยในห้องเรียน ห้องเรียนเป็นหน่วยย่อยของโรงเรียน เป็นต้น

นักจิตวิทยาได้พัฒนาโมเดลอย่างต่อเนื่องจนกระทั่ง ปี 1987 ไบรย์ และเรเดนบิวช (Bry; & Raudenbush) ได้พัฒนาโมเดลพหุระดับในการวิเคราะห์การเปลี่ยนแปลงรายบุคคล จากการออกแบบการวัดในหลายช่วงเวลา (Multiple-time-point design) เพื่อศึกษาสภาพการเปลี่ยนแปลงรายบุคคล (Individual status) โดยใช้โปรแกรมเฮซแอลเอ็มในการประมาณค่าพารามิเตอร์ต่างๆ ในการอธิบายการเปลี่ยนแปลงรายบุคคล ซึ่งนักจิตวิทยาได้ยอมรับว่าโปรแกรมเฮซแอลเอ็มเป็นเทคนิคที่เหมาะสมที่สุดเทคนิคหนึ่งในการวัดการเปลี่ยนแปลงรายบุคคล โดยใช้โมเดลเชิงเส้นพหุระดับ (Hierarchical linear model) มาประยุกต์ใช้ในการวัดการเปลี่ยนแปลง จากการวัดหลายครั้ง (Multiple-time-point) เพื่อศึกษาโครงสร้างของพัฒนาการรายบุคคล

เทคนิคเฮซแอลเอ็มนี้พัฒนามาจากสถิติหลายประเภท คือ เทคนิคการวิเคราะห์ความแปรปรวนโมเดลผสม (Mixed-model ANOVA), สัมประสิทธิ์การถดถอยแบบสุ่ม (Regression with random coefficients), โมเดลส่วนประกอบความแปรปรวนร่วม (Covariance component models) และการประมาณค่าในโมเดลเชิงเส้นด้วยวิธีของเบย์ส์ (Bayesian estimation for linear models) จึงเป็นเทคนิคที่ให้ผลการวิเคราะห์ที่มีความคงเส้นคงวาและน่าเชื่อถือกว่าเทคนิคกำลังสองน้อยที่สุด (Bry; & Raudenbush. 1986; ศิริชัย กาญจนวาสี. 2542; วราภรณ์ วิหคโต. 2536) ซึ่งในการศึกษาค้นคว้าครั้งนี้ได้ใช้โมเดลพหุระดับ ด้วยโมเดลวัดการเปลี่ยนแปลงแบบพหุระดับ 2 ระดับ เป็นวิธีประมาณค่าที่ใช้เป็นเกณฑ์ (Bryk; & Raudenbush. 1987; Williamson, Appelbaum; & Epanchin. 1991; Rogosa; & Saner. 1995)

ในการวิเคราะห์สภาพเริ่มต้นและอัตราพัฒนาการ นั้นประกอบไปด้วยรูปแบบของโมเดล ดังนี้

ระดับที่ 1 โมเดลการสังเกตซ้ำ (repeated-observations model) เป็นโมเดลการเปลี่ยนแปลงภายในของบุคคล

$$Y_i = \pi_0 + \pi_1(\text{TIME}) + E$$

เมื่อ π_0 เป็นค่าจุดตัด (intercept) ซึ่งเป็นความสามารถเริ่มต้นของคนที่ i เมื่อค่าตัวแปรทำนายคือ เวลาที่มีค่าเท่ากับ 0

π_1 เป็นค่าอัตราพัฒนาการ (growth rate) ของคนที่ i เมื่อเวลาผ่านไป 1 ช่วงเวลาที่กำหนด
TIME เป็นระยะเวลาในการวัดของคนที่ i

ระดับที่ 2 โมเดลระดับบุคคล (person-level model) สามารถเขียนอยู่ในรูปสมการได้ดังนี้

$$\pi_0 = \beta_{00} + R_0$$

$$\pi_1 = \beta_{10} + R_1$$

- เมื่อ β_{00} เป็นค่าเฉลี่ยความสามารถเริ่มต้นของกลุ่ม
 R_0 เป็นอิทธิพลสุ่มของความสามารถเริ่มต้นของคนที่ i
 β_{10} เป็นค่าเฉลี่ยอัตราพัฒนาการของกลุ่ม
 R_1 เป็นอิทธิพลสุ่มของอัตราพัฒนาการของคนที่ i

ขั้นตอนการวิเคราะห์โมเดลพหุระดับสำหรับการวัดการเปลี่ยนแปลง

ขั้นแรก ตรวจสอบความแปรปรวนของค่าพารามิเตอร์ของตัวแปรตาม โดยในสมการไม่มีตัวแปรทำนายใดๆ เพื่อพิจารณาและตรวจสอบว่าตัวแปรตามมีความแปรปรวนในแต่ละบุคคลหรือไม่ จะทำให้ทราบว่าเป็นการเพียงพอที่จะค้นหาตัวแปรทำนายที่มีอิทธิพลในขั้นต่อไปหรือไม่ ซึ่งเป็นการวิเคราะห์ในชั้นโมเดลไร้เงื่อนไขอย่างสมบูรณ์ (Fully unconditional model) หรือโมเดลศูนย์ (Null model) โดยใช้สถิติที (t - test) ทดสอบอิทธิพลคงที่ (Fixed effect); $H_0 : \beta_{00} = 0$ ถ้าไม่เป็นศูนย์ แสดงว่า จุดตัด (Intercept) และตัวแปรทำนายส่งผลต่อตัวแปรตาม (Y) และใช้ไค-สแควร์ (Chi - Square) ทดสอบอิทธิพลสุ่ม (Random effect); $H_0 : \text{Var}(\pi_0) = \text{Var}(R_0) = 0$ ถ้าไม่เป็นศูนย์ แสดงว่า มีความแปรปรวนของคะแนนการสอบซึ่งเป็นตัวแปรตามระหว่างนักเรียน จึงสมเหตุสมผลที่จะหาตัวแปรทำนายมาอธิบายความแปรปรวนดังกล่าว มีโมเดลการวิเคราะห์ ดังนี้

ระดับ 1

$$Y_t = \pi_0 + E$$

ระดับ 2

$$\pi_0 = \beta_{00} + R_0$$

ต่อจากนั้นจึงนำตัวแปรทำนายที่สามารถทำนายความแปรปรวนของคะแนนการสอบซึ่งเป็นตัวแปรตามเข้าไปในสมการ ในที่นี้ตัวแปรทำนายคือช่วงเวลา เป็นการวิเคราะห์ในชั้นโมเดลไร้เงื่อนไข (Unconditional model) หรือโมเดลอย่างง่าย (Simple model) มีโมเดลการวิเคราะห์ ดังนี้

ระดับ 1

$$Y_t = \pi_0 + \pi_1(\text{TIME}) + E$$

ระดับ 2

$$\pi_0 = \beta_{00} + R_0$$

$$\pi_1 = \beta_{10} + R_1$$

ในการวิเคราะห์ในชั้นโมเดลไร้เงื่อนไข (Unconditional model) นี้ มีวัตถุประสงค์ในการวิเคราะห์เพื่อตรวจสอบค่าอิทธิพลคงที่และอิทธิพลสุ่ม ดังนี้

1. ทดสอบอิทธิพลคงที่ (Fixed effects) ได้แก่ ค่าเฉลี่ยของความสามารถเริ่มต้น และค่าเฉลี่ยอัตราพัฒนาการ ว่าเป็นศูนย์หรือไม่ ($H_0: \beta_{00} = 0; \beta_{10} = 0$) โดยใช้สถิติที (t - test) ถ้าไม่เป็นศูนย์ แสดงว่า ค่าเฉลี่ยของความสามารถเริ่มต้น (Mean initial status; β_{00}) และ ค่าเฉลี่ยอัตราพัฒนาการ (Mean growth rate; β_{10}) มีอิทธิพลต่อตัวแปรตาม (Y) หรือไม่

2. ทดสอบอิทธิพลสุ่ม (Random effects) ได้แก่ ความแปรปรวนของค่าเฉลี่ยของความสามารถเริ่มต้น และค่าเฉลี่ยอัตราพัฒนาการ ว่าเป็นศูนย์หรือไม่ ($H_0: \text{Var}(\pi_0) = \text{Var}(R_0) = 0; \text{Var}(\pi_1) = \text{Var}(R_1) = 0$) โดยใช้สถิติไค-สแควร์ (Chi-Square) ถ้าไม่เป็นศูนย์ แสดงว่า ความแปรปรวนของค่าพารามิเตอร์ของค่าเฉลี่ยของความสามารถเริ่มต้น (π_0) และค่าเฉลี่ยอัตราพัฒนาการ (π_1) มีความผันแปรระหว่างหน่วยหรือระหว่างบุคคล นั่นคือมีความแปรปรวนระหว่างบุคคล

ขั้นที่ 2 ตรวจสอบความเชื่อมั่นของค่าพารามิเตอร์ความสามารถเริ่มต้นและอัตราพัฒนาการ (Reliability of assessments of initial status and change) ถ้าการประมาณค่าพารามิเตอร์ ไม่มีความเชื่อมั่นจะไม่สามารถสรุปได้ว่า มีความสัมพันธ์ในค่าพารามิเตอร์ทั้งสองกับตัวแปรในระดับบุคคล ความเชื่อมั่นในการประมาณค่าพารามิเตอร์สามารถคำนวณได้จากอัตราส่วนของความแปรปรวนของค่าพารามิเตอร์จริงกับความแปรปรวนที่สังเกตได้ทั้งหมด ในทฤษฎีการทดสอบแบบมาตรฐานเดิม ดังนี้

$$\rho_k = \text{Var}(\pi_k) / \text{Var}(\hat{\pi}_k) \\ = \tau_{kk} / (v_{kk} + \tau_{kk})$$

เมื่อความแปรปรวนที่สังเกตได้ทั้งหมดได้แก่ $\text{Var}(\hat{\pi}_k)$ ในการประมาณค่าพารามิเตอร์พัฒนาการ K จะประกอบไปด้วย ความแปรปรวนของการสุ่ม (Sampling variance) กับความแปรปรวนของพารามิเตอร์ (Parameter variance) ดังนี้

$$\text{Var}(\hat{\pi}_k) = \text{Var}(\hat{\pi}_k / \pi_k) + \text{Var}(\pi_k) \\ = v_{kk} + \tau_{kk}$$

$k = 0, 1, \dots, K-1$ ของการประมาณค่าพารามิเตอร์

ρ_k เป็นค่าความเชื่อมั่นของการประมาณค่าพารามิเตอร์พัฒนาการของแต่ละคน; π_k

โมเดลพหุระดับสำหรับการวัดการเปลี่ยนแปลง เป็นที่นิยมใช้ทั้งในการวัดปริมาณการเปลี่ยนแปลงรายบุคคลและกลุ่ม รวมทั้งศึกษาปัจจัยที่ส่งผลต่อการเปลี่ยนแปลงในข้อมูลที่มีระดับลดหลั่นกัน เช่น ระดับบุคคล ระดับชั้นเรียน ระดับสถานศึกษา เป็นต้น ดังตัวอย่างงานวิจัยของ ไบรย์และเรอาเดนบิวช (Bry; & Raudenbush. 1987); เบอริชินัลและแอปเพิลบาร์ม (Burchinal; & Appelbaum. 1991); ชิม (Shim. 1995); ไชว เบนเลอร์ และเพนซ์ (Chou; Bentler; & Pentz.1998); มา และมา (Ma; & Ma.

2004); แม็คโคช (McCoach.; et al. 2006); วีระศักดิ์ คำล้าน (2540); สุภารัตน์ เรือจันทิก (2542); และ สมถวิล วิจิตรวรรณ (2543) เป็นต้น

ทฤษฎีการตอบข้อสอบ (Item Response Theory)

ทฤษฎีการตอบข้อสอบ (Item Response Theory; IRT) ซึ่งนิยมเรียกว่า IRT เป็นทฤษฎีการวัดที่สามารถอธิบายความสัมพันธ์ระหว่างผลการตอบข้อสอบของผู้เรียนที่เราสามารถสังเกตได้ กับคุณลักษณะภายใน หรือความสามารถที่แท้จริงของผู้เรียนที่เราไม่สามารถสังเกตวัดได้โดยตรง มีข้อตกลงเบื้องต้น 2 ประการ คือ 1) คุณลักษณะหรือคุณลักษณะแฝงหรือความสามารถ เป็นตัวทำนายการตอบข้อสอบของผู้เรียน และ 2) ความสัมพันธ์ระหว่างการตอบข้อสอบของผู้เรียนกับความสามารถอธิบายได้ด้วยฟังก์ชันที่มีลักษณะเพิ่มขึ้นแบบเพิ่มขึ้นหรือลดลงทางเดียว (Monotonic) ที่เรียกว่า ฟังก์ชันคุณลักษณะข้อสอบ (Item characteristic function) หรือ โค้งลักษณะข้อสอบ (Item characteristic curve; ICC) จะเห็นได้ว่า ถ้าระดับความสามารถของผู้เรียนสูงขึ้นแล้วความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องจะสูงขึ้นด้วย

โมเดลของการตอบข้อสอบสามารถเขียนด้วยฟังก์ชันลักษณะข้อสอบในเชิงคณิตศาสตร์ที่แตกต่างกัน ซึ่งมีหลายรูปแบบตามจำนวนพารามิเตอร์ ได้แก่ โมเดล 1, 2 หรือ 3 พารามิเตอร์ เมื่อเลือกใช้โมเดลการตอบข้อสอบได้อย่างเหมาะสม จะช่วยให้สามารถวิเคราะห์ข้อสอบ เพื่อทราบค่าสถิติของข้อสอบที่ไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ทำให้ประมาณค่าความสามารถที่แท้จริงของผู้เรียนได้ และได้ทราบค่าสารสนเทศของข้อสอบและแบบทดสอบ การประยุกต์ทฤษฎีการตอบข้อสอบ สำหรับการวัดผลทางการเรียน จึงเป็นประโยชน์อย่างยิ่งต่อการพัฒนาและปรับปรุงคุณภาพการเรียนการสอน และการวัดประเมินผลการเรียนที่ถูกต้อง

ทฤษฎีการตอบข้อสอบมีจุดเด่นที่น่าสนใจคือ เมื่อโมเดลการตอบข้อสอบมีความสอดคล้องกลมกลืนกับข้อมูลแล้ว การประมาณค่าความสามารถของผู้เรียนจะไม่ขึ้นอยู่กับข้อสอบและกลุ่มผู้เรียน หรือค่าพารามิเตอร์ของข้อสอบและความสามารถจะไม่แปรเปลี่ยน นั่นคือค่าพารามิเตอร์ของข้อสอบแต่ละข้อไม่แปรเปลี่ยน (Invariant of Item Parameter) จึงสามารถใช้ข้อสอบกับกลุ่มผู้สอบต่างกลุ่มได้ (Person Free) และจะใช้ข้อสอบข้อใดก็ได้ (Item Free) ในการวิเคราะห์ข้อสอบเป็นรายข้อตามทฤษฎีการตอบข้อสอบนี้ เป็นจุดเด่นที่น่าสนใจของทฤษฎีนี้ แต่มีข้อจำกัด คือ ต้องใช้กลุ่มตัวอย่างขนาดใหญ่

ข้อตกลงเบื้องต้น (Assumption) ในการใช้ทฤษฎีการตอบข้อสอบ มีดังนี้

1. **ความเป็นมิติเดียว (Unidimensionality)** หมายความว่า ถ้าความสามารถของผู้เรียนมี

อยู่ทั้งหมด K ด้าน ซึ่งเรียกว่า Latent Space โดยที่ความสามารถแต่ละด้านต่างก็ส่งผลต่อการตอบข้อสอบ (Test performance) ต่างๆ ที่รวมกันเป็นแบบทดสอบ ถ้าผลการตอบข้อสอบหรือคะแนนของผู้เรียนนั้นสามารถอธิบายได้เพียงความสามารถเดียวก็อาจเรียกข้อตกลงนี้ว่ามีความเป็นมิติเดียว ซึ่งเกี่ยวข้องกับความเป็นอิสระเฉพาะที่ในการตอบข้อสอบ (Local Independence) ซึ่งหมายถึง ความน่าจะเป็นในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกต้องหรือผิด จะไม่มีผลกระทบต่อคำตอบข้อสอบข้ออื่นๆ หรือกล่าวในเชิงคณิตศาสตร์ได้ว่า ความเป็นอิสระเฉพาะที่ (Local Independence) ในการตอบข้อสอบ หมายถึง ความน่าจะเป็นที่จะตอบข้อสอบได้ถูกต้องทั้งหมดจะเท่ากับผลคูณของความน่าจะเป็นการตอบข้อสอบถูกเป็นรายข้อ ดังสูตร ซึ่งอาจตรวจสอบโดยใช้วิธีการวิเคราะห์องค์ประกอบ (Factor analysis)

$$\begin{aligned}\text{Prob}(U_1, U_2, \dots, U_n | \theta) &= P(U_1 | \theta) P(U_2 | \theta) \dots P(U_n | \theta) \\ &= \prod_{i=1}^n P(U_i | \theta)\end{aligned}$$

2. **ฟังก์ชันลักษณะข้อสอบ (Item Characteristic Function)** เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบข้อนั้นได้ถูกต้องกับระดับความสามารถ ซึ่งโมเดลการตอบข้อสอบเป็นโมเดลโลจิสติก (Logistic) แบบ 1, 2, และ 3 พารามิเตอร์ จากการตอบข้อสอบแบบสองค่า (Dichotomous) ซึ่งมี 3 โมเดล ดังนี้

2.1 โมเดลโลจิสติกแบบ 1 พารามิเตอร์ (One-Parameter logistic model)

โมเดลโลจิสติกแบบ 1 พารามิเตอร์นี้เป็นโมเดลการตอบข้อสอบที่นิยมใช้มากที่สุดโมเดลหนึ่งที่ค่าความยากของข้อสอบ (b) เป็นจุดบนมาตราความสามารถที่ความน่าจะเป็นในการตอบถูกเป็น .50 ถ้าค่าความยาก (b) มาก ค่าความสามารถ (θ) จะมากด้วย หมายความว่า ถ้ามีจำนวนผู้เรียนร้อยละ 50 อยู่ทางขวามือของมาตราแล้ว ข้อสอบข้อนั้นจะเป็นข้อสอบที่ยาก ถ้าแปลงค่าความสามารถให้มีค่าเฉลี่ยเท่ากับ 0 และ ค่าเบี่ยงเบนมาตรฐานเท่ากับ 1 ค่าความยาก (b) อยู่ระหว่าง -2.0 ถึง 2.0 นั่นคือ ข้อคำถามที่มีค่าความยาก (b) เข้าใกล้ -2.0 จะเป็นข้อคำถามที่ง่าย ในขณะที่ ค่าความยาก (b) เข้าใกล้ 2.0 จะเป็นข้อคำถามที่ยาก มีสมการดังนี้

$$P_i(\theta) = \frac{e^{i(\theta-b_i)}}{1 + e^{i(\theta-b_i)}} \quad ; i = 1, 2, \dots, n$$

2.2 โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (Two-parameter logistic model)

ลอร์ด (Hambleton; Swaminathan; & Roger. 1991: 14; citing Lord. 1952. Psychometric Monograph.) เป็นคนแรกในการพัฒนาโมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์ ที่มีลักษณะการกระจายเป็นโค้งปกติสะสม (Normal ogive) ต่อมา เบอรัณบอม (Hambleton, Swaminathan, & Roger. 1991: 14; citing Birnbuam. 1968. Statistical theories of mental test scores.) ใช้ฟังก์ชันโลจิสติกแบบ 2 พารามิเตอร์ แทนโค้งปกติสะสม (Normal ogive) แบบ 2 พารามิเตอร์ ซึ่งมีลักษณะแบบโค้งลักษณะข้อสอบ หรือฟังก์ชันลักษณะข้อสอบ มีความสะดวกต่อการใช้งานมากกว่าฟังก์ชันโค้งปกติสะสม (Normal ogive) เนื่องจากโค้งปกติสะสม (Normal ogive) มีค่า Integrate และ เบอรัณบอม (Bimbuam) ได้พัฒนาโค้งลักษณะข้อสอบ (Item characteristic curves ; ICC) ของโมเดลโลจิสติกแบบ 2 พารามิเตอร์ เป็นโมเดลที่คล้ายกับโมเดลโลจิสติกแบบ 1 พารามิเตอร์ ต่างกันที่จำนวนพารามิเตอร์มี 2 ตัว โดยมีพารามิเตอร์อำนาจจำแนกเพิ่มขึ้นมา และ D แทน ค่าคงที่ มีค่าเท่ากับ 1.7 ซึ่งค่าความน่าจะเป็นในการตอบข้อสอบได้ถูก เมื่อกำหนดค่า θ ให้ ; $P(\theta)$ ของโค้งปกติสะสม (Normal ogive) แบบ 2 พารามิเตอร์ ต่างจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์อยู่ 0.1 ในทุกๆ ค่าของ θ

พารามิเตอร์อำนาจจำแนก (a) เรียกว่า พารามิเตอร์อำนาจจำแนกของข้อสอบ โดยที่ค่าอำนาจจำแนก (a) เป็น สัดส่วนกับความชันของโค้งลักษณะข้อสอบ (ICC) ณ จุดความยาก (b) บนมาตรความสามารถ เป็นค่าที่ใช้จำแนกระดับความสามารถของผู้เรียนที่ต่างกัน ถ้ามีความชันมากจะจำแนกได้ดี ถ้ามีความชันน้อยจะจำแนกได้ไม่ดี ในทางทฤษฎีแล้ว พารามิเตอร์อำนาจจำแนกนี้อยู่บนมาตรตั้งแต่ $-\infty$ ถึง ∞ ถ้าข้อสอบมีค่าอำนาจจำแนกติดลบจะถูกตัดทิ้ง ค่าความยาก (a) มีค่าอยู่ระหว่าง 0 ถึง 2 สามารถเขียนเป็นสมการได้ดังนี้

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad ; i = 1, 2, \dots, n$$

2.3 โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (Three-parameter logistic model)

เป็นโมเดลที่พัฒนามาจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์ เพื่อความเหมาะสมกับการทดสอบที่มีอิทธิพลของการเดาเข้ามาเกี่ยวข้อง โดยเพิ่มค่าพารามิเตอร์ขึ้นอีกตัวหนึ่งคือ ค่าการเดา c เป็น Pseudo-chance-level จะทำให้ค่าแอสิมโทต (Asymtote) ด้านต่ำ ใน ICC ไม่เป็นศูนย์ ซึ่งเขียนเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} \quad ; i = 1, 2, \dots, n$$

คุณสมบัติของการไม่แปรเปลี่ยน (Property of Invariance)

ลักษณะเด่นของ IRT คือ คุณสมบัติการไม่แปรเปลี่ยนทั้งของพารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถ ดังรายละเอียดต่อไปนี้

1. คุณสมบัติการไม่แปรเปลี่ยนของพารามิเตอร์ข้อสอบ เมื่อโมเดลมีความสอดคล้องกลมกลืนกับข้อมูล จะได้โค้งลักษณะข้อสอบ (Item Characteristic Curve; ICC) ที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ซึ่งคุณสมบัตินี้เป็นที่รู้จักกันดี คือ โมเดลการถดถอยเชิงเส้น (Linear regression model)

2. คุณสมบัติการไม่แปรเปลี่ยนพารามิเตอร์ความสามารถ หมายความว่า ค่าความสามารถของผู้เรียนจะไม่ขึ้นอยู่กับชุดของข้อสอบ

คุณสมบัติการไม่แปรเปลี่ยนของพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถเป็นคุณสมบัติที่สำคัญของทฤษฎีการตอบข้อสอบและการนำไปใช้เกี่ยวกับการวัดผลการศึกษา เช่น การเทียบคะแนน (Equating) การจัดทำคลังข้อสอบ (Item banking) ความลำเอียงของข้อสอบ หรือ การทำหน้าที่ต่างกันของข้อสอบ (Item bias หรือ Differential item functioning) และ การทดสอบแบบปรับเหมาะ (Adaptive testing)

โมเดลอื่นๆ (Other Promising Models)

เป็นโมเดลที่ใช้กับข้อมูลที่ให้คะแนนไม่เป็นแบบสองค่า (Dichotomous) แต่เป็นข้อมูลที่ให้คะแนนแบบหลายค่า (Polytomous) ได้แก่โมเดลดังต่อไปนี้

1. Nominal Response Model ซึ่ง บอค (Hambleton; Swaminathan; & Roger. 1991: 26; citing Bock. 1972. Psychometrika.) เป็นผู้พัฒนาขึ้นจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์ โดยให้ความน่าจะเป็นในการตอบคำถามของผู้ตอบจะเลือกตอบ k ตัวเลือก จากตัวเลือกทั้งหมด m ตัว ในข้อที่ i ที่มีเงื่อนไขว่าผลรวมของความน่าจะเป็นในการตอบทุกตัวเลือก (m ตัว) ของผู้ตอบที่ระดับความสามารถที่กำหนดให้มีค่าเท่ากับหนึ่ง ซึ่งเขียนเป็นโมเดลทางคณิตศาสตร์ได้ ดังนี้

$$P_{ik}(\theta) = \frac{e^{a_{ik}(\theta - b_{ik}^*)}}{\sum_{h=1}^m e^{a_{ih}(\theta - b_{ih}^*)}} \quad ; i = 1, 2, \dots, n \quad ; k = 1, 2, \dots, m$$

2. Graded-Response Model ซึ่ง ซามิไจมา (Hambleton; Swaminathan; & Roger. 1991: 26; citing Samejima. 1969. Psychometric Monograph.) เป็นผู้พัฒนาขึ้น โดยมีข้อตกลงเบื้องต้นเกี่ยวกับการตอบคำถามข้อที่ i ของผู้ตอบซึ่งแบ่งออกเป็น m_{i+1} ลำดับ (Categories) โดยให้คะแนนแต่ละลำดับเป็น $0, 1, 2, \dots, m_i$ เช่น มาตราวัดแบบลิเคิร์ต (Likert) หรือการให้ 4 ลำดับคะแนนในการเขียนตอบ เป็นการหาความน่าจะเป็นในการตอบคำถามแต่ละรายการจากการตอบลำดับอื่นๆ ของผู้ตอบแต่ละคน ซึ่งเขียนเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้

$$P_{x_i}(\theta) = P_{x_i}^*(\theta) - P_{x_i+1}^*(\theta)$$

3. Binomial Trails Model ซึ่ง ซาฟริท คอสตา และ โคเฮน (Hambleton; Swaminathan; & Roger. 1991: 26-27; citing Safrit Costa & Cohen. 1989. Research Quarterly for Exercise and Sport.) และ สเปรย์ (Hambleton, Swaminathan, & Roger. 1991: 27; citing Spray. 1990. Research Quarterly for Exercise and Sport.) ศึกษาความสามารถด้านตัวเลขและความเข้าใจในการอ่านกับความสมบูรณ์ของร่างกาย ความสามารถในการโยนลูกบาสเก็ตบอลเป็นการประเมินทางด้านทักษะ ค่าพารามิเตอร์ของข้อคำถามในโมเดลก็คือ การทดลองแต่ละครั้งก็คือข้อคำถามแต่ละข้อ การประมาณค่าของข้อคำถามจึงหาได้ง่าย โดยที่การทดลองจะต้องเป็นอิสระ ตัวอย่างเป็นการประยุกต์ข้อมูลการโยนลูกบาสเก็ตบอล (จำนวนของการโยนลง) θ จะเป็นความสามารถในการโยนลูกบาสเก็ตบอลลง ซึ่งเขียนเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้

$$P_{x_i}(\theta) = \binom{n}{x} P(\theta)^x Q(\theta)^{n-x}$$

4. Poisson Counts Model เป็นโมเดลที่ความน่าจะเป็นในการทำงานเสร็จเรียบร้อยภายใน 1 นาที ซึ่งเขียนเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้

$$P(X = x | \theta, b) = \frac{e^{x(\theta-b)}}{x! e^{e(\theta-b)}}$$

การวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบ

การวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบ เป็นวิธีการวัดแนวใหม่ที่ใช้โมเดลตามแนวคิดที่ได้รับการพัฒนาขึ้นในระยะหลัง ซึ่งจะมีลักษณะของวิธีการวัด การใช้สถิติและผลของการวัดที่แตกต่างไปจากการวัดแบบมาตรฐานเดิม การใช้โมเดลการตอบข้อสอบสำหรับการวิเคราะห์การออกแบบการวัดซ้ำ ได้มีการเผยแพร่มาประมาณ 20 ปีที่แล้ว (Fischer.1983,1994,1995; Fischer; & Formann. 1982: 397-416; Rost. 1989,1996) โมเดลเหล่านี้สามารถแก้ปัญหาบางประการของการวัดการเปลี่ยนแปลงแบบมาตรฐานเดิม (Glück; & Spiel. 1997: Online; citing Bereiter. 1963. Problems in Measuring Change.) ข้อดีที่สำคัญที่สุดประการหนึ่งของโมเดลการตอบข้อสอบ คือ เป็นการกระทำบนพื้นฐานข้อสอบเดี่ยว (Single-item basis) ที่เป็นอิสระจากข้อสอบข้ออื่นในแบบทดสอบ ซึ่งหมายถึง มีความเป็นอิสระเฉพาะที่นั่นเอง

การวัดการเปลี่ยนแปลงนี้ มีผู้พัฒนาวิธีการไว้หลายแบบ สามารถสรุปได้ ดังนี้

1. โมเดลคุณลักษณะแฝงโลจิสติกแบบเส้นตรงที่ผ่อนคลายข้อตกลงเบื้องต้น (Linear logistic latent trait model with relaxed assumptions; LLRA) (Fischer. 1976: 97-110)
2. โมเดลราชพหุมิติ (Multidimensional Rasch models) (Andersen. 1985: 3-16; & Embretson. 1991: 495-515)
3. การเปลี่ยนแปลงความสามารถแฝง (Latent change) (May; & Nicewander. 1998: 882-897)
4. การวัดคะแนนพัฒนาการความสามารถที่แท้จริงสัมพันธ์ อวยพร เรืองตระกูล. (2544: 82-85) พัฒนาขึ้น เป็นวิธีการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบโดยใช้แนวคิดของ ศิริชัย กาญจนวาสีและวินิจ เทือกทอง
5. การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ (IRT Domain scores) (Heidenberg. 2000)

1. โมเดลคุณลักษณะแฝงโลจิสติกแบบเส้นตรงที่ผ่อนคลายข้อตกลงเบื้องต้น (Linear logistic latent trait model with relaxed assumptions; LLRA)

ฟิชเชอร์ (Fischer. 1976: 97-110) ได้เสนอแนวคิดตั้งแต่ปี 1973 เป็นต้นมา เป็นการเสนอโมเดลการตอบข้อสอบทั้งสิ้น 5 โมเดล เพื่อใช้วัดการเปลี่ยนแปลงโดยกล่าวถึงพื้นฐานเชิงทฤษฎีของโมเดลรวมทั้งข้อดีและข้อบกพร่อง ดังนี้

- 1.1 โมเดลการทดสอบโลจิสติกเชิงเส้นตรง (Linear Logistic Test Model : LLTM)
- 1.2 โมเดลโลจิสติกเชิงเส้นตรงที่ผ่อนปรนข้อตกลงเบื้องต้น (Linear Logistic Model with relaxed Assumptions : LLRA)
- 1.3 โมเดลโลจิสติกเชิงเส้นตรงที่มีลักษณะกึ่งผ่อนปรนข้อตกลงเบื้องต้น (Hybrid-LLRA : HLLRA)
- 1.4 โมเดลมาตราประมาณค่าเชิงเส้นตรง (Linear Rating Scale Model ; LRSM)
- 1.5 โมเดลการให้คะแนนบางส่วนเชิงเส้นตรง (Linear Partial Credit Model ; LPCM)

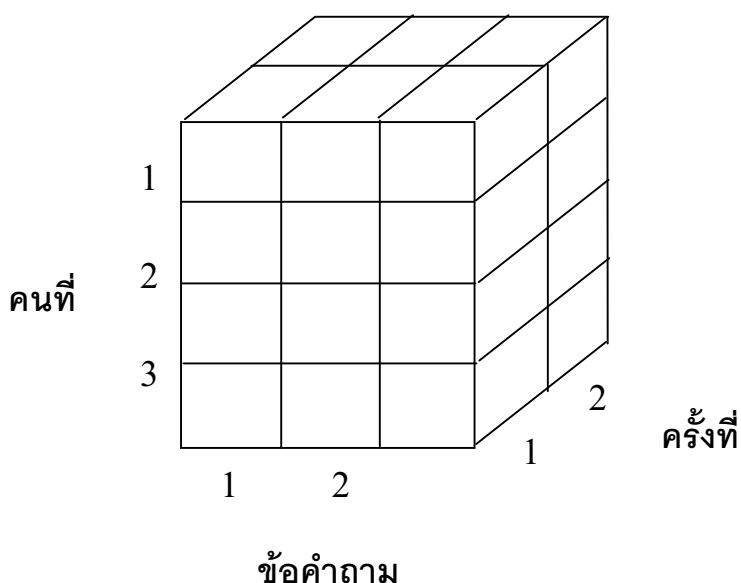
ข้อตกลงเบื้องต้นที่เข้มงวดในการใช้ทฤษฎีการตอบข้อสอบ คือ ความเป็นมิติเดียว (Unidimensionality) ของแบบทดสอบ ซึ่งเป็นการยากที่ผู้สร้างข้อสอบจะทำได้สำเร็จ ในบางโมเดลจึงไม่สนใจกับข้อตกลงเบื้องต้นข้อนี้ ในการวิเคราะห์ที่มีวัตถุประสงค์เพื่อการศึกษาปริมาณการเปลี่ยนแปลงของกลุ่มย่อย ที่ไม่ใช่ต้องการศึกษาคะแนนการเปลี่ยนแปลงภายในบุคคล (Intraindividual changes) ดังนั้น ข้อตกลงเบื้องต้นเกี่ยวกับความเป็นมิติเดียวจึงไม่มีความจำเป็น เช่น ถ้าการเปลี่ยนแปลงในคน 2 กลุ่ม (กลุ่มทดลองและกลุ่มควบคุม) ถูกเปรียบเทียบกัน ความเป็นเอกพันธ์ของข้อสอบจึงไม่มีความ

จำเป็น เพราะการเปลี่ยนแปลงในค่าความยากของข้อสอบถูกประเมินแยกกันในแต่ละข้อ นั่นแสดงถึงข้อจำกัดข้อหนึ่งของวิธีการวัดการเปลี่ยนแปลงตามโมเดลการตอบข้อสอบ 3 โมเดลต่อไปนี้จะได้แก่ Linear Logistic Test Model (LLTM), Linear Logistic Model with relaxed Assumptions (LLRA) และ Hybrid-LLRA (HLLRA)

โมเดลเหล่านี้ได้รับแนวคิดมาจากโมเดลราช (โมเดลโลจิสติกแบบ 1 พารามิเตอร์ และโมเดลราช) แต่โมเดล LLTM มีความขัดแย้งกับ LLRA และ HLLRA ตรงที่ฟิสเซอร์ไม่ได้พัฒนาขึ้นมาเพื่อการวัดการเปลี่ยนแปลงตั้งแต่เริ่มต้น ซึ่งทั้ง 3 โมเดลได้นำไปประยุกต์ใช้กับข้อมูลที่เป็นแบบสองค่าหรือทวิภาค (Dichotomous) หรือ กลุ่มลำดับ (Ordered categorical)

ความคิดพื้นฐานของโมเดลการตอบข้อสอบสำหรับการวัดการเปลี่ยนแปลง

การใช้โมเดลราชวิเคราะห์เมตริกซ์ข้อมูล 2 มิติ เมื่อข้อสอบเป็นมิติเดียว (Unidimensionality) แล้วการประมาณค่าพารามิเตอร์มีความเที่ยงตรง โดยลำดับที่และความแตกต่างของความยากระหว่างข้อสอบเท่ากันทุกข้อของทั้งฉบับ นั่นคือมีช่วงความยากของข้อสอบเท่ากันทุกช่วง สำหรับการวัดการเปลี่ยนแปลงนั้นขึ้นอยู่กับลูกบาศก์ข้อมูล (Data cube) ซึ่งประกอบไปด้วยผู้สอบ (Subject ; N) มากกว่าหนึ่งคน ข้อสอบ (Item ; I) มากกว่าหนึ่งข้อ และจำนวนครั้งที่สอบ (Time point ; T) มากกว่าหนึ่งครั้ง ดังภาพประกอบ 1



ภาพประกอบ 1 ลูกบาศก์ข้อมูล ที่มีผู้สอบ 4 คน สอบข้อสอบ 3 ข้อ ที่มีการสอบ 2 ครั้ง

ที่มา: Glück; & Spiel. (1997). Online.

ในทางปฏิบัติการออกแบบการทดลอง จะพบว่าผู้สอบ (Subject ; N) อาจถูกแบ่งเป็นกลุ่มย่อยๆ ได้อีก เช่น กลุ่มทดลองและกลุ่มควบคุม ดังนั้นเพื่อใช้โมเดลการตอบข้อสอบสำหรับการวัดการเปลี่ยนแปลง จึงต้องจัดโครงสร้างข้อมูล 3 มิติใหม่ให้อยู่ในรูปของเมทริกซ์ 2 มิติ ซึ่งสามารถจัดทำได้ 3 วิธีที่แตกต่างกัน ดังนี้

วิธีที่ 1 ถ้าข้อสอบแต่ละข้อที่ใช้วัดเป็นมิติเดียว (Unidimensionality) แล้ว มาตรวัดจึงมีลักษณะเป็นเอกพันธ์ (Homogeneous) เมื่อนำไปสอบจำนวน t ครั้ง จะสามารถจัดโครงสร้างของข้อมูลเป็นเมทริกซ์ 2 มิติได้ดังเช่น ถ้าผู้สอบ (N) 4 คน แต่ละคนจะสอบด้วยข้อสอบจำนวน $k \times t = 3 \times 2 = 6$ ข้อ ซึ่ง $k \times t$ ข้อนี้ เรียกว่า “ข้อสอบเสมือนจริง” (Virtual items) ของเมทริกซ์ข้อมูล ดังตัวอย่างลูกบาศก์ข้อมูลที่ถูกจัดโครงสร้างข้อมูลให้อยู่ในรูปเมทริกซ์ 2 มิติในกรณีที่มาตรวัดเป็นเอกพันธ์ (Homogeneous) โดยการเปลี่ยนแปลงเกิดจากการเปรียบเทียบค่าพารามิเตอร์ความยากของข้อสอบของการสอบครั้งที่ t นั่นคือ ถ้าข้อสอบในครั้งที่ 2 ง่ายกว่าครั้งที่ 1 อย่างมีนัยสำคัญแล้ว แสดงว่า ความสามารถของผู้สอบมีการพัฒนา ปรับปรุงอย่างแน่นอน ซึ่งค่าพารามิเตอร์ข้อสอบจะประมาณค่าแยกกลุ่มผู้สอบย่อยๆ ดังกล่าว คือ กลุ่มทดลองและกลุ่มควบคุม ดังนั้นความอาจเป็นไปได้ประการแรกในการทดสอบสมมติฐานเฉพาะก็คือ กลุ่มทดลองมีการพัฒนาปรับปรุงความสามารถมากกว่ากลุ่มควบคุมหรือไม่ วิธีการนี้เป็นพื้นฐานของการวัดการเปลี่ยนแปลงด้วย Linear Logistic Test Model (LLTM)

	ข้อ 1 ครั้ง 1	ข้อ 2 ครั้ง 1	ข้อ 3 ครั้ง 1	ข้อ 1 ครั้ง 2	ข้อ 2 ครั้ง 2	ข้อ 3 ครั้ง 2
คนที่ 1	X_{111}	X_{121}	X_{131}	X_{112}	X_{122}	X_{132}
คนที่ 2	X_{211}	X_{221}	X_{231}	X_{212}	X_{222}	X_{232}
คนที่ 3	X_{311}	X_{321}	X_{331}	X_{312}	X_{322}	X_{332}
คนที่ 4	X_{411}	X_{421}	X_{431}	X_{412}	X_{422}	X_{432}

ที่มา: Glück; & Spiel. (1997). Online.

วิธีที่ 2 ความอาจเป็นไปได้ของลูกบาศก์ข้อมูลที่ถูกจัดโครงสร้างข้อมูลใหม่ให้อยู่ในรูปเมทริกซ์ 2 มิติ โดยไม่คำนึงถึงความเป็นมิติเดียว (Unidimensionality) ของข้อสอบในแบบทดสอบ เมทริกซ์ที่สร้างขึ้นมานั้น คอลัมน์ (Column) แสดงครั้งที่สอบ และแถว (rows) แสดงถึงผู้สอบเสมือนจริง “Virtual subjects” ที่มี $N \times k$ แถว เมื่อนำข้อมูลชุดเดิมจากลูกบาศก์มาสร้างเมทริกซ์ 2 มิติด้วยวิธีนี้ได้ $4 \times 3 = 12$ แถวหรือคน โดยแต่ละคนสอบข้อสอบ 2 ข้อ นั่นคือ มี 2 คอลัมน์ ดังตัวอย่างเมทริกซ์ข้อมูลในกรณีที่ข้อสอบเป็นวิวิธพันธ์ (Heterogeneous)

ดังนั้นการเปลี่ยนแปลงในค่าพารามิเตอร์ความยากของข้อสอบสะท้อนถึงการเปลี่ยนแปลงในความสามารถของผู้สอบ (Subject's ability) โดยการจำแนกผู้สอบออกเป็นกลุ่มย่อย คือ ผู้สอบเสมือนจริงกับผู้สอบจริง ซึ่งเป็นวิธีที่ใช้ทดสอบปริมาณการเปลี่ยนแปลงที่เท่ากันข้ามกลุ่มผู้สอบย่อย และสำหรับข้อสอบที่แตกต่างกัน วิธีนี้ได้นำไปใช้ใน Linear Logistic Model with relaxed Assumptions (LLRA)

	ครั้งที่ 1	ครั้งที่ 2
คนที่ 1 ข้อที่ 1	X_{111}	X_{112}
คนที่ 2 ข้อที่ 1	X_{211}	X_{212}
คนที่ 3 ข้อที่ 1	X_{311}	X_{312}
คนที่ 4 ข้อที่ 1	X_{411}	X_{412}
คนที่ 1 ข้อที่ 2	X_{121}	X_{122}
คนที่ 2 ข้อที่ 2	X_{221}	X_{222}
คนที่ 3 ข้อที่ 2	X_{321}	X_{322}
คนที่ 4 ข้อที่ 2	X_{421}	X_{422}
คนที่ 1 ข้อที่ 3	X_{131}	X_{132}
คนที่ 2 ข้อที่ 3	X_{231}	X_{232}
คนที่ 3 ข้อที่ 3	X_{331}	X_{332}
คนที่ 4 ข้อที่ 3	X_{431}	X_{432}

ที่มา: Glück; & Spiel. (1997). Online.

วิธีที่ 3 ความอาจเป็นไปได้ของลูกบาศก์ข้อมูลที่จัดโครงสร้างข้อมูลใหม่ให้อยู่ในรูปเมทริกซ์ 2 มิติ วิธีนี้อยู่ระหว่างกลางของสองวิธีข้างต้น ซึ่งเป็นการวัดการเปลี่ยนแปลงเชิงคุณภาพระหว่างครั้งของการสอบ อาจทำได้ด้วยการจัดกลุ่ม/จัดชั้นเรียนของผู้สอบตามแบบแผนการตอบ (Response patterns) ของแต่ละคน นั่นคือในทั้งสองวิธีข้างต้นเป็นการวัดการเปลี่ยนแปลงเชิงปริมาณภายในชั้นเรียนหนึ่งหรือการเปลี่ยนแปลงเชิงคุณภาพเป็นรายบุคคลในชั้นเรียน และเพื่อจะวิเคราะห์ดังกล่าวจึงจัดโครงสร้างข้อมูลใหม่โดยการสอบ 2 ครั้ง ซึ่งแต่ละครั้งกลุ่มสอบกลุ่มย่อยมีความเป็นอิสระในการตอบข้อสอบ ผู้สอบอาจเลือกตอบข้อสอบข้อเดียวกันก็ได้ ดังนั้นผลการจัดโครงสร้างข้อมูลจึงมีข้อสอบจำนวน k ข้อ และผู้สอบเสมือนจริงเท่ากับ $t * n$ คน วิธีนี้ได้นำไปใช้ใน Mixed Rasch Model (MRM) ซึ่งต่อไปนี้เป็นรายละเอียดของแต่ละโมเดล

1.1 โมเดลการทดสอบโลจิสติกเชิงเส้นตรง (Linear Logistic Test Model : LLTM)

การวัดการเปลี่ยนแปลงที่ใช้แบบทดสอบที่มีลักษณะเป็นเอกพันธ์ (Homogeneous) ถ้าใช้ชุดข้อสอบที่มีลักษณะเป็นมิติเดียวจำนวน k ข้อ ในการสอบทุกครั้ง การวัดการเปลี่ยนแปลงโดยใช้ Linear Logistic Test Model (LLTM) เป็นการทดสอบนัยสำคัญของผลและความแตกต่างระหว่างผลในการออกแบบการทดลองหรือกึ่งทดลอง โดยทั่วไป LLTM อธิบายความแตกต่างระหว่างข้อสอบที่เป็นเอกพันธ์แบบราช (Rasch-homogeneous) ได้นำค่าพารามิเตอร์ความยากเข้าไปรวมเป็นเส้นตรงซึ่งเป็นค่าพารามิเตอร์พื้นฐานสำหรับการวัดการเปลี่ยนแปลงด้วยวิธีนี้ที่กำหนดค่าพารามิเตอร์ความยากของข้อสอบในแบบทดสอบของกลุ่มผู้สอบในครั้งที่ 2 ; t_2 เป็นผลรวมค่าพารามิเตอร์ความยากของข้อสอบในครั้งที่ 1 บวกกับค่าพารามิเตอร์ของผลเฉพาะ 1 กลุ่ม หรือมากกว่า 1 กลุ่ม หรือค่าพารามิเตอร์ของผลทั่วไป ซึ่งมีขั้นตอนดำเนินการทดสอบโมเดลที่มีความสอดคล้องกลมกลืนกับข้อมูล ดังนี้

ขั้นตอนที่ 1 โมเดลกึ่งสมบูรณ์ (Quasi-saturated model) การประมาณค่าพารามิเตอร์ความยากของข้อสอบที่แยกจากข้อสอบเสมือนจริง (Virtual-item) ทุกข้อ นั่นคือ สำหรับข้อสอบทุกข้อ (Every item) $\times$ จำนวนครั้งที่สอบ (Time point) ในทุกกลุ่มผู้สอบย่อย โมเดลนี้สามารถอธิบายได้อย่างเพียงพอ

ขั้นตอนที่ 2 การทดสอบอัตราส่วนความเป็นไปได้ (Likelihood ratio tests) เป็นการทดสอบความยากของข้อสอบเสมือนจริง (Virtual-item) เพราะว่า ผลรวมของค่าพารามิเตอร์ความยากข้อสอบในการสอบครั้งที่ 1 บวกกับค่าพารามิเตอร์ผลกลุ่มเฉพาะหรือกลุ่มทั่วไป

ขั้นตอนที่ 3 เป็นการทดสอบข้อกำหนดหรือสมมติฐาน

ในกรณีที่โมเดลมีความสอดคล้องกลมกลืนแล้ว ค่าพารามิเตอร์ในผลของกลุ่มควบคุมหรือค่าพารามิเตอร์ที่เทียบเท่าจะทำให้ค่าพารามิเตอร์มีค่าลดลงอย่างมีนัยสำคัญ กระบวนการที่เป็นขั้นตอนนี้ได้ใช้กระบวนการ Stepwise เพื่อให้เกิดโมเดลที่ประหยัดที่สุด ที่มีความสอดคล้องกลมกลืนกับข้อมูลและมีประสิทธิภาพ

1.2 โมเดลโลจิสติกเชิงเส้นตรงที่ผ่อนปรนข้อตกลงเบื้องต้น (Linear Logistic Model with relaxed Assumptions : LLRA)

LLRA (Fischer. 1976, 1983, 1987, 1989, 1996; Fischer & Formann, 1982: 397-416) เป็นโมเดลที่ขัดแย้งกับ LLTM ซึ่ง LLRA พัฒนาขึ้นมาสำหรับการวัดการเปลี่ยนแปลงโดยเฉพาะ การใช้ผู้สอบเสมือนจริงในการประมาณค่าการเปลี่ยนแปลงของกลุ่มเฉพาะ แม้ว่าข้อสอบในแบบทดสอบไม่สามารถวัดมิติแฝงที่มีลักษณะมิติเดียว ในกรณีค่าปลายๆ สุดของข้อสอบทุกข้ออาจวัดความสามารถที่ต่างกัน ซึ่งสมการโมเดลของ LLRA คือ

$$p(x_{vi1}=1) = \frac{\exp(\theta_{vi})}{1 + \exp(\theta_{vi})}$$

$$p(x_{vi2}=1) = \frac{\exp(\theta_{vi} + \delta_v)}{1 + \exp(\theta_{vi} + \delta_v)}$$

เมื่อ $p(x_{vi1}=1)$ หมายถึง ความน่าจะเป็นของคนที่ j ในการตอบข้อสอบข้อที่ i ถูก ในการสอบครั้งที่ 1

$p(x_{vi2}=1)$ หมายถึง ความน่าจะเป็นของคนที่ j ในการตอบข้อสอบข้อที่ i ถูก ในการสอบครั้งที่ 2

θ_{vi} หมายถึง ความสามารถของคนที่ v ในการวัดด้วยข้อสอบข้อที่ i

δ_v หมายถึง ผลรวมการเปลี่ยนแปลงทั้งหมดของคนที่ v ในการสอบสองครั้ง

โดยทั่วไป δ_v เหมือนผลรวมของค่าพารามิเตอร์ของกลุ่มเฉพาะ (กลุ่มที่ได้รับทรีทเมนต์) และ ความเป็นอิสระของการเปลี่ยนแปลงของทรีทเมนต์ เช่น กระบวนการพัฒนา ผลของการสอบซ้ำ)

ดังนั้นค่าพารามิเตอร์การเปลี่ยนแปลงรายบุคคลถูกกำหนดขึ้นใน LLRA สำหรับทุกคนและทุกข้อ ความหมายและการประมาณค่าพารามิเตอร์ที่คงที่ (Stable) มีความเป็นไปได้สำหรับกลุ่มของผู้สอบที่แทนด้วยผู้สอบกลุ่มเดียวกัน (Single subjects) ข้อตกลงเบื้องต้นที่ค่อนข้างเข้มงวดของการเปลี่ยนแปลงที่เท่ากันสำหรับข้อสอบทุกข้อ หมายถึง สมการโมเดลถูกทำให้ลดความรุนแรงลงโดยความจริงที่มีการทดสอบสมมติฐานเกี่ยวกับความแตกต่างของการเปลี่ยนแปลงระหว่างชุดย่อยของข้อสอบ

LLRA ใช้กระบวนการทดสอบอัตราส่วนความเป็นไปได้แบบ Stepwise (Stepwise likelihood ratio testing procedure) สำหรับการหาโมเดลที่ดีที่สุด ประหยัดที่สุด เหมือนใน LLTM โดยเริ่มที่โมเดลกึ่งสมบูรณ์ (Quasi-saturated model) เช่นกัน ที่กำหนดข้อมูล เพื่อให้โมเดลสุดท้ายมีความสอดคล้องกลมกลืนกับข้อมูล ซึ่งได้อธิบายถึงการประยุกต์ใช้ LLRA ของ แบร์ลิสซ์ (Glück; & Spiel. 1997. Online; citing Barisch. 1989. Einstellung zur Epilepsie und Einstellungsänderung durch Information.)

การประยุกต์ใช้ LLRA ในการวิเคราะห์ปริมาณการเปลี่ยนแปลงสำหรับข้อสอบข้อหนึ่งทีประมาณค่าจากความถี่ของแบบแผนการตอบ (Response pattern) ที่มีลักษณะเป็น (0,1) หมายถึง ผู้สอบทำข้อสอบข้อนี้ไม่ได้ในการสอบครั้งที่ 1 แต่ทำได้ในการสอบครั้งที่ 2 ส่วน (1,0) หมายถึง ผู้สอบทำข้อสอบข้อนี้ได้ในการสอบครั้งที่ 1 แต่ทำไม่ได้ในการสอบครั้งที่ 2 ถ้าแบบแผนใดแบบแผนหนึ่งในสองแบบแผนนี้ปรากฏขึ้นนานๆ ครั้ง หรือไม่ปรากฏเลย จึงจะใช้การประมาณค่าพารามิเตอร์ที่เรียกว่าแบบ Hybrid LLRA (HLLRA)

1.3 โมเดลโลจิสติกเชิงเส้นตรงที่มีลักษณะกึ่งผ่อนปรนข้อตกลงเบื้องต้น (Hybrid-LLRA : HLLRA)

โมเดลนี้ใช้ข้อสอบเป็นคู่ที่มีลักษณะเป็นมิติเดียว เป็นโมเดลที่ LLRA ที่เป็นมาตรฐานไม่สามารถนำไปประยุกต์ใช้ได้กับข้อมูลที่มีการเปลี่ยนแปลงในทิศทางใดทิศทางหนึ่งที่เห็นได้อย่างชัดเจน ดังนั้น Hybrid-LLRA (HLLRA) เป็นวิธีทางเลือกหนึ่ง โดยใช้ข้อสอบที่เป็นคู่ที่มีลักษณะเป็นเอกพันธ์กัน และใช้ข้อสอบที่มีความยากแตกต่างกันในครั้งที่ 2 นั่นก็คือ ข้อสอบที่ง่ายกว่านำไปใช้สอบในครั้งที่ 1 ส่วนข้อที่มีความยากมากกว่านำไปใช้สอบในครั้งที่ 2 โดยที่ข้อสอบทั้งสองวัดมิติแฝงที่เป็นมิติเดียวกัน ดังนั้นการเปลี่ยนแปลงความสามารถระหว่าง t_1 และ t_2 สามารถทำให้เท่ากันได้ด้วยการเปลี่ยนแปลงค่าความยากของข้อสอบเหมือนใน LLRA ที่เป็นมาตรฐาน นั่นคือ รูปแบบของคู่ข้อสอบ 2 ข้อที่มีลักษณะเป็นมิติเดียวสามารถนำไปใช้ทดสอบได้มากกว่าหนึ่งครั้ง

HLLRA สามารถทดสอบสมมติฐานเช่นเดียวกับ LLRA ที่เป็นมาตรฐาน สามารถอธิบายปริมาณการเปลี่ยนแปลงของกลุ่มที่แตกต่างกัน ผลของแนวโน้มการเปลี่ยนแปลงที่ปรากฏในกลุ่มทั้งหมด (เพราะการสอบซ้ำ) ไม่สามารถแก้ไขหรือทำให้เหมาะสมได้ ดังเช่น ผลเหล่านี้มีความแตกต่างของค่าความยากข้อสอบเข้าไปปะปนรวมอยู่ด้วย ซึ่งจุดหลักของความสนใจในการออกแบบการทดลองและกึ่งทดลองโดยทั่วไปเน้นที่ความแตกต่างของกลุ่ม ไม่ใช่การเปลี่ยนแปลงรายบุคคล ซึ่งอาจจะไม่เป็นปัญหาแต่ถ้าการประมาณค่าความยากข้อสอบกับกลุ่มตัวอย่างที่อิสระจากกัน ผลของแนวโน้มสามารถแก้ไขหรือทำให้เหมาะสมได้ด้วยความแตกต่างของค่าความยากข้อสอบที่แท้จริง ข้อตกลงเบื้องต้นของ HLLRA มีความยืดหยุ่นน้อยกว่า LLRA ที่เป็นมาตรฐานมาก เพราะความเป็นเอกพันธ์ของคู่ของข้อสอบค่อนข้างเข้มงวดมากกว่า

Hybrid-LLRA ได้กำหนดข้อตกลงเบื้องต้นของความเป็นมิติเดียวของแบบทดสอบสำหรับทุกข้อในแบบทดสอบย่อย “Visu-motor skills” โดยการประยุกต์ใช้ข้อสอบที่เป็นคู่จำนวน 3 คู่ ที่มี ความยากของข้อสอบแตกต่างกัน คู่ของข้อสอบ 3 คู่ สามารถจัดให้อยู่ในเทอมของ “Hybrid items” ใน Hybrid LLRA ได้ดังต่อไปนี้

ข้อสอบคู่ที่	ข้อคำถามในการสอบครั้งที่ 1	ข้อคำถามในการสอบครั้งที่ 2
1	ข้อที่ 2 กระโดด	ข้อที่ 1 ยืนขาเดียว
2	ข้อที่ 3 ทำตามแบบโครงสร้างของอิฐ	ข้อที่ 6 ถือแม่แรงแล้วกระโดด
3	ข้อที่ 4 ถือแก้วน้ำ	ข้อที่ 5 วาดวงกลมวงใหม่

ที่มา: Glück; & Spiel. (1997). Online.

ในการประยุกต์ใช้ครั้งนี้ได้แบ่งผู้สอบซึ่งเป็นเด็กออกเป็น 2 กลุ่ม โดยใช้เกณฑ์การแบ่ง คือ ทักษะการมีปฏิสัมพันธ์ทางสังคม (SIS) ของครู โดย กลุ่มที่ 1 เป็นเด็กกลุ่มของครูที่มีปฏิสัมพันธ์ทางสังคมไม่ดี จำนวน 23 คน และกลุ่มที่ 2 เป็นเด็กของครูที่มีทักษะปฏิสัมพันธ์ทางสังคมดี จำนวน 71 คน

ผู้สอบข้อสอบที่เป็นแบบ Hybrid นี้มีจำนวนทั้งสิ้น 94 คน ดังนั้นเมทริกซ์ข้อมูลมีเท่ากับ $3 \times 94 = 282$ แถว ในขั้นตอนแรก โมเดลศูนย์แบ่งผู้สอบเสมือนจริงจำนวน 282 คน เป็นกลุ่มย่อย 6 กลุ่ม (ข้อสอบแบบ Hybrid จำนวน 3 คู่ และกลุ่มผู้สอบ 2 กลุ่มที่แบ่งตามทักษะการมีปฏิสัมพันธ์ทางสังคมของครู) ค่าพารามิเตอร์ในแต่ละกลุ่มย่อยมีแนวโน้มถูกรวมเพื่อการเปลี่ยนแปลงทั่วไประหว่างการสอบ t_1 กับ t_2 ค่าพารามิเตอร์ผล(ทรีทเมนต์) ที่เพิ่มเข้ามาถูกใส่ลงไปในโมเดลที่อธิบายด้วยการเปลี่ยนแปลงที่เพิ่มขึ้นซึ่งรวมผลที่คาดหวังในกลุ่มที่ 2 ดังในตารางต่อไป ซึ่งแสดงเมทริกซ์การออกแบบสำหรับโมเดลศูนย์ สังเกตว่าเมทริกซ์นี้เป็นการจัดลำดับของคอลัมน์ที่สมบูรณ์

ในขั้นต่อไป ค่าพารามิเตอร์ผล (ทรีทเมนต์) ทั้งหมดถูกเอาออก (โมเดลหนึ่ง) ซึ่งเป็นค่าพารามิเตอร์แนวโน้ม 3 ค่าที่อธิบายการเปลี่ยนแปลงทั้งหมด ซึ่งแสดงให้เห็นถึงสมมติฐานที่ไม่มีความแตกต่างในการเปลี่ยนแปลงระหว่างกลุ่มที่ SIS ของครูดีกับไม่ดี (ตัวอย่างโมเดล 2 ใน LLTM) ซึ่งสามารถแสดงเมทริกซ์การออกแบบของโมเดลศูนย์ที่กึ่งสมบูรณ์ ด้วยข้อสอบที่เป็นแบบ Hybrid ได้ดังต่อไปนี้

		τ_{HI1}	η_{HI1}	τ_{HI2}	η_{HI2}	τ_{HI3}	η_{HI3}
Group 1	HI1/t1	0	0	0	0	0	0
	HI1/t2	1	0	0	0	0	0
	HI2/t1	0	0	0	0	0	0
	HI2/t2	0	0	1	0	0	0
	HI3/t1	0	0	0	0	0	0
	HI3/t2	0	0	0	0	1	0
Group 2	HI1/t1	0	0	0	0	0	0
	HI1/t2	1	1	0	0	0	0
	HI2/t1	0	0	0	0	0	0
	HI2/t2	0	0	1	1	0	0
	HI3/t1	0	0	0	0	0	0
	HI3/t2	0	0	0	0	1	1

ที่มา: Glück; & Spiel. (1997). Online.

โมเดลนี้ไม่สามารถลดความสอดคล้องกลมกลืนลงอย่างมีนัยสำคัญ ซึ่งพิจารณาจากการทดสอบอัตราส่วนความเป็นไปได้ในผลการทดสอบอัตราส่วนความเป็นไปได้ของ LLRA

โมเดล	จำนวนพารามิเตอร์	log L	χ^2	จำนวนต่างของพารามิเตอร์	$P(\chi^2)$
0	6	-77.898	-	-	-
1	3	-79.730	3.665	3	0.300
2	1	-79.793	3.791	5	0.590
3	0	-86.643	17.491	6	0.008

ที่มา: Glück; & Spiel. (1997). Online.

ในโมเดลสองเห็นได้อย่างชัดเจน ซึ่งประกอบไปด้วยค่าพารามิเตอร์แนวโน้มค่าหนึ่งในข้อสอบทั้ง 3 ข้อ แสดงว่าในสมมติฐานนั้นพบว่าปริมาณการเปลี่ยนแปลงของทุกกลุ่มและในทุกข้อมีค่าเท่ากันซึ่งข้อสอบทุกข้อเป็นแบบ Hybrid ทั้ง 3 ข้อ เมื่อโมเดลนี้มีความสอดคล้องกลมกลืนแล้ว สมมติฐานที่ตั้งไว้ว่า “ไม่มีการเปลี่ยนแปลง” แสดงว่าการเปลี่ยนแปลงในความยากของข้อสอบมีความแตกต่างกันมากกว่าความยากระหว่างข้อสอบที่ใช้ในครั้งที่ 1 และครั้งที่ 2 ดังนั้น สมมติฐานที่ว่า “ไม่มีการเปลี่ยนแปลง” จึงหมายถึงว่า ความแตกต่างระหว่างความสามารถของเด็กในครั้งที่ t1 และ t2 เท่ากับความแตกต่างของค่าความยากระหว่างข้อสอบที่ใช้ในการสอบครั้งที่ t1 และ t2

1.4 โมเดลมาตรฐานค่าเชิงเส้นตรง (Linear Rating Scale Model ; LRSM)

ฟิชเชอร์; และพาร์เซอร์ (Fischer; & Parzer. 1991: 637-651) ได้เสนอโมเดลนี้ขึ้นในปี 1991 โดยได้ขยายมาตรฐานค่า (Rating Scale) ของ แอนดริช (Andrich. 1978: 561-573) ออกไปให้อยู่ในรูปของ LLTM โดยกระจายค่าพารามิเตอร์ของแต่ละขั้นหรือกลุ่ม (Category) ออกไปในรูปแบบของผลรวมของพารามิเตอร์อิทธิพล โดยมีสูตรดังนี้

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

K_{vi} แทน ค่าในขั้นหรือกลุ่ม (Category) ของคนที่ v และข้อที่ i (มีค่าเท่ากับ 0,1,2,...,h,...,m)

$$\begin{aligned} \theta_v & \text{แทน ความสามารถของคนที่ } v \\ \beta_i & \text{แทน ค่าความยากของข้อที่ } i \\ \omega_0, \dots, \omega_m & \text{แทน ค่าพารามิเตอร์ของชั้น (Category) และ} \\ \beta_i &= \sum_{j=1}^n q_{ij} \alpha_j + c; \quad i = 1, 2, \dots, t \\ \text{โดยที่ } q_{ij} & \text{เป็นน้ำหนักของพารามิเตอร์พื้นฐานที่ } j \text{ ของข้อที่ } i \\ \alpha_j & \text{เป็นพารามิเตอร์พื้นฐานที่ } j \end{aligned}$$

LRSM ใช้กับการวัดซ้ำ 2 ครั้ง ด้วยข้อสอบชุดเดิมที่เป็นมาตรฐานค่า (Rating scale) ที่วัดมิติเดียว (Unidimension) โดยใช้ CML(Conditional Maximum Likelihood) ในการประมาณค่าพารามิเตอร์ด้วยกระบวนการแบบย้อนซ้ำด้วยวิธี Newton-Raphson Algorithm และทดสอบสมมติฐานด้วย CLR

1.5 โมเดลการให้คะแนนบางส่วนเชิงเส้นตรง (Linear Partial Credit Model; LPCM)

Partial Credit Model (Fischer; & Ponocny. 1994: 177-192) ได้คิดพิจารณาภายใต้ข้อตกลงเบื้องต้นของการแยกองค์ประกอบเชิงเส้นตรง (Linear decomposition) ของพารามิเตอร์ข้อสอบและค่าพารามิเตอร์ของลำดับชั้นการตอบ (Item x Category parameters) ; δ_{ih} ในค่าพารามิเตอร์พื้นฐาน (Basic parameters) ; α_j โมเดลนี้เรียกว่า “Linear partial credit model” โดยใช้วิธีเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional maximum likelihood algorithm; CML) สำหรับประมาณค่า α_j ซึ่งขึ้นอยู่กับ a) การเกิดซ้ำของฟังก์ชันผลรวม (Combinatorial functions) ที่เกี่ยวข้องกัน และ b) การใช้วิธีที่กึ่งนิวตัน (Quasi-Newton) ที่เรียกกันว่าวิธี Broyden-Fletcher-Goldfarb-Shanno (BFGS) ที่ให้ผลลัพธ์ที่คงที่แน่นอนและหลีกเลี่ยงการคำนวณโดยตรงของ the Hesse matrix ซึ่งเป็นกระบวนการตามลำดับของเมทริกซ์ที่นิยามทางบวกอย่างแน่นอน B_k เมื่อ $k = 1, 2, \dots$ ลงไปสู่เมทริกซ์ความแปรปรวน-ความแปรปรวนร่วมแอสิมโทติก (Asymptotic variance-covariance matrix) ของ α_j

โมเดลการประมาณค่า (Rating scale model; RSM) ของ แอนดริช (Andrich. 1978: 561-573) เป็นกรณีเฉพาะในตระกูลของโมเดลราชพหุวิภาค (Polytomous Rasch models) เป็นการก่อให้เกิดความนิยมชมชอบใน โครงสร้างเชิงเส้น (Linear structure) ของค่าพารามิเตอร์ข้อสอบ “Linear rating scale model; LRSM” ถูกนำไปประยุกต์ใช้กับข้อมูลที่ใช้ในการรักษาผู้ป่วย อธิบายอาการป่วยของแต่ละความน่าจะเป็นของผลการตอบในเทอมของค่าพารามิเตอร์ข้อสอบหนึ่งข้อ คือ β_i , ค่าพารามิเตอร์ของกลุ่มคำตอบหนึ่งกลุ่ม คือ ω_h และ ค่าพารามิเตอร์ของคนหนึ่งคน คือ θ_v ที่มีข้อจำกัดมากสำหรับความสอดคล้องกลมกลืนที่ดีกับข้อมูลที่ยู่ยากซับซ้อน การนำเสนอนี้จึงขึ้นอยู่กับการนำเสนอ “Partial

credit model” (PCM; Anderson. 1983; Master. 1982) ซึ่งกำหนดให้ค่าพารามิเตอร์อิสระหนึ่งค่าในการรวมแต่ละข้อคำถาม $I \times$ กลุ่ม C (Each item $I \times$ Category C) และมีความยืดหยุ่นมากกว่าโครงสร้างเชิงเส้น (Linear structure) ของค่าพารามิเตอร์ถูกนำเสนออีกครั้ง การขยาย PCM นี้ปรากฏให้เห็นเป็นผลก่อนเรียนในการเปลี่ยนแปลงเล็กน้อยกว่าเมื่อเปรียบเทียบกับ LRSM เพราะว่าสูตรบางสูตรดูคล้ายกัน แต่การตรวจสอบได้ละเอียดกว่าจะทำให้ความแตกต่างมีนัยสำคัญ 1) LPCM มีโครงสร้างทั่วไปมากกว่า เพื่อกำหนดขึ้นเป็นสูตรและทดสอบเชิงประจักษ์ ซึ่ง LRSM เป็นกรณีเฉพาะของ LPCM และ 2) LPCM มีค่าพารามิเตอร์มากกว่ามาก ดังนั้นการประมาณค่าที่ไม่สำคัญเหล่านั้นก่อให้เกิดปัญหาและต้องการวิธีการทางคณิตศาสตร์ใหม่ๆ เพื่อการคำนวณการแก้ปัญหาวีธีเป็นไปได้อย่างสูงสุดแบบมีเงื่อนไข (Conditional maximum likelihood; CML) ดังตัวอย่างต่อไปนี้

แบบประเมินตนเองซึ่งใช้ประเมินครั้งเดียว ประกอบไปด้วยข้อคำถาม 30 ข้อ (k) ในแต่ละข้อมีกลุ่มคำตอบ 5 กลุ่ม ($m+1$) ($C_0, C_1, \dots, C_m$; $m = 4$) LPCM ประกอบด้วยข้อคำถาม $(km-1) \times$ ค่าพารามิเตอร์กลุ่มคำตอบ (Category parameters) เช่น ค่าพารามิเตอร์ 119 ค่า ที่ถูกประมาณค่าด้วยวิธี CML ใน LRSM ส่วนของเส้นตรงของค่าพารามิเตอร์ฟังก์ชันเอ็กโพเนนเชียล (Exponential function) คือ $h(\theta_v + \beta_i) + \omega_h$ ดังนั้น ค่าพารามิเตอร์ข้อคำถาม คือ $k-1 = 29$ บวกกับค่าพารามิเตอร์กลุ่มคำตอบ คือ $m-1 = 3$ นั่นก็คือ ค่าพารามิเตอร์ที่ถูกประมาณค่ามี 32 ค่า ในตัวอย่างนี้ PCM จะมีค่าประมาณ 3.7 เท่ากว่าๆ ของค่าพารามิเตอร์ที่ได้จาก LRSM กระบวนการแบบย้อนซ้ำด้วยวิธี Newton-Raphson ที่ใช้โดย ฟิชเชอร์ และพาร์เซอร์ (Fischer; & Parzer. 1991) ขึ้นอยู่กับอนุพันธ์แบ่งส่วนอันดับที่สอง (Second-order partial derivatives) เป็นหลัก และ matrix inversion การประมาณค่าเพิ่มขึ้นโดยกำลังสามของจำนวนพารามิเตอร์ การคำนวณตัวเลขที่สอดคล้องกันใน PCM จะทำซ้ำทั้งหมด 3.7^3 ประมาณ 51 ครั้ง ซึ่งเสียค่าใช้จ่ายมาก จึงมีการพัฒนา กฎ CML ที่มีประสิทธิภาพมากกว่าสำหรับ LPCM

PCM นิยามด้วยสมการ ดังนี้

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

เมื่อ $i = 1, \dots, k$

$h = 0, \dots, m$

C_h = ชั้นของคำตอบ

A_{vih} = คำตอบแบบทวิภาคของคนที่ S_v ในข้อที่ I_i

m_{i+1} = จำนวนชั้นคำตอบของข้อที่ I_i , $C_0, \dots, C_m$

β_{ih} = ความง่ายของชั้นคำตอบของข้อที่ I_i

θ_v = ตำแหน่งคุณลักษณะแฝงที่เป็นมิติเดียวของคน S_v

PCM ถูกพิจารณาภายใต้ข้อจำกัดของเส้นตรง

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

เมื่อ $i = 1, \dots, k$ และ $h = 1, \dots, m_i$

$c =$ ค่าคงที่

$\alpha_j, j = 1, \dots, p$ ซึ่งเป็นค่าพารามิเตอร์พื้นฐานของผลการวัดด้วยข้อคำถาม

$w_{ihj} =$ น้ำหนักของ α_j

สมการ PCM ทั้งสองรวมกัน จะเรียกว่า “Linear partial credit model” (LPCM) ค่าพารามิเตอร์พื้นฐาน (Basic parameters ; α_j) สามารถแปลความหมายได้หลากหลายขึ้นอยู่กับการประยุกต์ใช้ ซึ่งอาจแทนผลลัพธ์ของการกระทำทางความคิด (Cognition operation) หรือ ผลลัพธ์ของการเรียนรู้หรือ ทริทเมนต์ของคุณลักษณะหรือ ผลลัพธ์ของการติดต่อสื่อสารหรือ การโฆษณาประชาสัมพันธ์ ในการวัดเจตคติโดยใช้ข้อคำถามที่เป็นแบบลิเคิร์ต

จากที่กล่าวมาสามารถสรุปรูปแบบของการออกแบบการวิจัยและเงื่อนไขในการประยุกต์ใช้โมเดลต่างๆ ตามแนวคิดของฟิชเชอร์ (Fischer) ได้ ดังนี้

โมเดล	เงื่อนไขสำหรับการประยุกต์ใช้
1. Linear Logistic Test Model (LLTM)	เป็นมิติเดียว (ข้อสอบเป็นเอกพันธ์แบบราช)
2. Linear Logistic Model with relaxed Assumptions (LLRA)	ไม่มี
3. Hybrid Linear Logistic Model with relaxed Assumptions (HLLRA)	เป็นมิติเดียวของทุกข้อคำถามของการสอบหลายครั้ง (2 ครั้ง หรือ มากกว่า)
4. Linear Rating Scale Model (LRSM)	เป็นมิติเดียว สอบ 2 ครั้ง โดยใช้ข้อคำถามเดิมที่เป็นแบบมาตราประมาณค่า
5. Linear Partial Credit Model (LPCM)	เป็นมิติเดียว สอบ 2 ครั้ง โดยใช้ข้อคำถามเดิมที่เป็นการให้คะแนนบางส่วน

2. โมเดลราชพหุมิติสำหรับการวัดการเปลี่ยนแปลงการเรียนรู้ (Multidimensional Rasch model for learning and change : MRMLC)

เอมเบรตสัน (Embretson) ได้พัฒนาโมเดลนี้ขึ้นเมื่อปี ค.ศ. 1989 (Embretson. 1991: 495-515) โมเดลคุณลักษณะแฝงถูกนำเสนอสำหรับการวัดความสามารถแบบซ้ำที่ขึ้นอยู่กับหลายมิติของ

การเปลี่ยนแปลง โครงสร้างอย่างง่ายถูกกำหนดขึ้นมาเพื่อเชื่อมโยงการปฏิบัติการข้อสอบภายใต้เงื่อนไขการวัดหรือจำนวนครั้งในการวัดสำหรับความสามารถเริ่มต้นและความสามารถเพิ่มขึ้นหนึ่งหรือหลายความสามารถที่แทนด้วยความแตกต่างรายบุคคลในการเปลี่ยนแปลง ภายใต้เงื่อนไขการวัดที่กำหนดให้ค่าอำนาจจำแนกข้อสอบมีค่าเท่ากัน โมเดลนี้อยู่ในตระกูลของโมเดลราชพหุมิติ (Multidimensional Rasch models) ซึ่งประมาณค่าพารามิเตอร์ข้อสอบและความสามารถด้วยความเป็นไปได้สูงสุด (Maximum likelihood; ML)

โมเดลราชพหุมิติสำหรับการวัดการเปลี่ยนแปลงการเรียนรู้ ขึ้นอยู่กับทฤษฎีการตอบข้อสอบ (Item Response Theory) โมเดลจำแนก Wiener simplex pattern ที่แสดงความเกี่ยวพันของความสามารถเริ่มต้น (Initial ability) และ ความสามารถเพิ่มขึ้นหนึ่งหรือหลายความสามารถ (Modifiability) จากคำตอบที่ได้จากการวัดหลายครั้งได้สำเร็จ โดยที่โมเดลดังกล่าวอยู่ในตระกูลโมเดลราชพหุมิติ (Multidimensional Rasch Models. 1961) ที่ค่าอำนาจจำแนกข้อสอบมีค่าเท่ากันในการวัดแต่ละครั้ง แม้ว่าโมเดลอาจจะถูกประยุกต์ใช้จริงและเห็นได้ชัดเจนกับบางสาขาซึ่งโมเดลการตอบข้อสอบ (IRT model) มีความเหมาะสมในการวัดความสามารถในการเรียนรู้

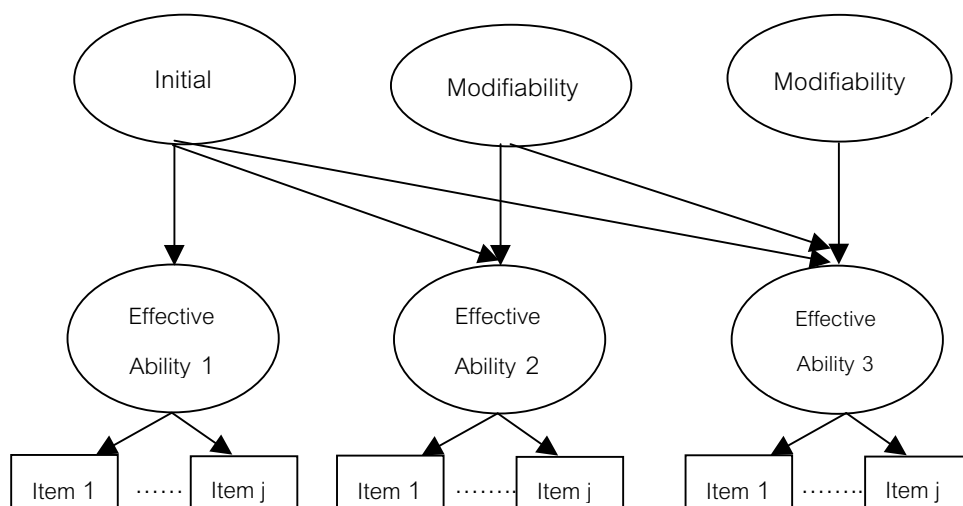
ในปี 1991 เอ็มเบรตสัน (Embretson) เสนอแนวคิดในการใช้โมเดลราชศัพท์ระดับวัดคะแนน พัฒนาการที่ได้จากการวัดตั้งแต่ 2 ครั้งขึ้นไปกับข้อมูลแบบทวิภาค (Dichotomous) โดยนิยามว่าความสามารถที่ถูกวัดมี 2 ประการ คือ ความสามารถที่เป็นผล (Effective ability) และความสามารถในการเรียนรู้หรือความสามารถที่เพิ่มขึ้น (Learning ability or modifiability) ซึ่งความสามารถที่เป็นผลนี้เป็นศักยภาพแฝงของผู้เรียนที่อยู่ภายใต้เงื่อนไขของเวลาในการวัด โดยเป็นแบบแผนการวัดซ้ำ 2-5 ครั้ง สำหรับชุดข้อมูลที่มีความแปรปรวนของความสามารถที่ลดลงหรือเป็นงานที่ทำซ้ำด้วยความสามารถคงเดิม ไม่ควรใช้โมเดลนี้ โมเดล MRMLC มีข้อตกลงเบื้องต้นว่าความสามารถที่เพิ่มขึ้นของผู้เรียนระหว่างเวลาที่มีโครงสร้าง เป็นแบบ Wiener Simplex กล่าวคือ ความสามารถของผู้เรียนจะมีแต่เพิ่มขึ้น ไม่มีลดลงจากเดิม

จากโมเดลราชพื้นฐาน

$$P(X_{ij} = 1 | \theta_j, b_i) = \frac{\exp(\theta_j - b_i)}{1 + \exp(\theta_j - b_i)}$$

θ_j คือ ความสามารถของคนที่ j ที่จะตอบข้อที่ i ถูก

b_i คือ ค่าความยากของข้อคำถามข้อที่ i



ภาพประกอบ 2 แผนภาพโมเดล MRMLC

ที่มา: Collins L.M., & Horn J.L. (1995). Best Methods for the Analysis of Change. p. 189.

โมเดลราชพหุมิติ (Multidimensional Rasch model) ของ แอนเดอร์เซน (Andersen. 1985: 3-16) สำหรับการสอบซ้ำด้วยข้อสอบชุดเดิมกับกลุ่มตัวอย่าง โดยมีโมเดลดังนี้

$$P(X_{i(k)j} = 1 | \theta_{jk}^*, b_i) = \frac{\exp(\theta_{jk}^* - b_i)}{1 + \exp(\theta_{jk}^* - b_i)}$$

เมื่อ θ_{jk}^* เป็นความสามารถของคนที่ j ในการสอบครั้งที่ k

b_i เป็นความยากของข้อสอบ

ความยากของข้อสอบมีความคงที่ทุกครั้งที่สอบ แต่ความสามารถที่เกี่ยวข้องกันขึ้นอยู่กับจำนวนครั้งที่สอบ ดังนั้น จำนวนครั้งที่สอบจึงนำมาใช้บรรยายความสามารถที่แตกต่างกัน แอนเดอร์เซน (Andersen) พัฒนาโมเดลสำหรับวิเคราะห์โครงสร้างคุณลักษณะแฝงระยะยาว (Model for longitudinal latent structure analysis) และได้สูตรสัมประสิทธิ์สหสัมพันธ์แฝงของความสามารถระหว่างครั้งในการสอบ แม้ว่าโมเดลของ แอนเดอร์เซน (Andersen) มีความเหมาะสมทำให้เข้าใจผลที่เกิดจากเวลา หรือทริทเมนต์ที่กระจายบนความสามารถ แต่โมเดลไม่สามารถคำนวณค่าพารามิเตอร์การเปลี่ยนแปลงเป็นรายบุคคล ดังนั้น ความสามารถในการโมเดลแอนเดอร์เซน (Andersen) เป็นการสอบเฉพาะ และไม่สะท้อนถึงความแตกต่างในการเปลี่ยนแปลงเป็นรายบุคคลที่มีการสอบหลายครั้ง

การวัดความแตกต่างรายบุคคลในการพัฒนาถูกตรวจสอบด้วยทฤษฎีการตอบข้อสอบ (Item response theory) ดังเช่น วิธีที่ต้องการโค้งที่มีความสอดคล้องกลมกลืนกับข้อมูลดังในการประยุกต์ใช้โมเดลการตอบข้อสอบ (Item response model) ของ บอค (Embretson. 1991: 497; citing Bock. 1976: Advance in psychological and educational measurement. pp. 75-96) สำหรับตรวจสอบความสอดคล้องกลมกลืนของโค้งพัฒนาการทางจิตใจเป็นรายบุคคล โค้งอาจจะสอดคล้องกลมกลืนด้วยองค์ประกอบแนวโน้มออโธกอนอล (Orthogonal trend components) หรือโดยการประมาณค่าพารามิเตอร์ในโค้งพัฒนาการเอ็กโปเนนเชียลและไฮเพอร์โบลิก (Exponential and Hyperbolic growth curves) (Embretson. 1991: 497; citing Keats. 1983. Principals of modern psychological measurement. pp. 81-101) อย่างไรก็ตาม Bock สังเกตว่าโค้งพัฒนาการทางจิตใจอาจยากที่จะมีความสอดคล้องกลมกลืน เพราะว่าความเชื่อมั่นของค่าสังเกตรายบุคคลไม่มีประสิทธิภาพสำหรับทดสอบสมมติฐานเกี่ยวกับแนวโน้ม (Trend) ต่อมาการวัดซ้ำที่มีอยู่เป็นการพอเพียงสำหรับทดสอบองค์ประกอบแนวโน้มที่มีลำดับที่สูงขึ้น (Higher-order trend components)

3. การวัดการเปลี่ยนแปลงความสามารถแฝง (Latent change) (May; & Nicewander. 1998: 882-297)

เมย์และไนส์วันเดอร์ (May; & Nicewander) ได้เสนอวิธีการวัดการเปลี่ยนแปลงความสามารถแฝงโดยใช้ทฤษฎีการตอบข้อสอบ โดยใช้สัญลักษณ์ d_θ

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

เมื่อ θ_1, θ_2 คือ ความสามารถที่แท้จริงก่อนเรียนและหลังเรียนที่ประมาณค่าด้วยวิธี EAP

4. การวัดคะแนนพัฒนาการความสามารถที่แท้จริงสัมพัทธ์ (อวยพร เรื่องตระกูล. 2544: 84-85)

อวยพร เรื่องตระกูล ได้พัฒนาวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการตอบข้อสอบ 2 วิธี โดยใช้แนวคิดของศิริชัย กาญจนวาสี (2532) และวินิจ เทือกทอง (2537) ดังนี้

4.1 วิธีการวัดคะแนนพัฒนาการความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับศักยภาพการพัฒนา (Relative true ability of growth and potential ratio method) พัฒนาโดยใช้แนวคิดของศิริชัย กาญจนวาสี (2532) ใช้สัญลักษณ์ S_θ

$$S_\theta = \frac{\theta_2 - \theta_1}{\text{Max} - \theta_1}$$

4.2 วิธีการวัดคะแนนพัฒนาการความสามารถที่แท้จริงสัมพัทธ์เมื่อเทียบกับความสามารถที่แท้จริงก่อนเรียน (Relative true ability of growth and initial ability method) พัฒนาโดยใช้แนวคิดของวินิจ เทือกทอง (2537) ใช้สัญลักษณ์ V_θ

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

เมื่อ θ_1, θ_2 คือ ความสามารถที่แท้จริงก่อนเรียนและหลังเรียน

Max คือ ความสามารถสูงสุด

5. การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ (Heidenberg. 2000)

ไฮเดนเบิร์ก (Heidenberg) ได้พัฒนาวิธีการวัดการเปลี่ยนแปลงโดยใช้คะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ (IRT domain scores) เพื่อศึกษาความสัมพันธ์ระหว่างคะแนนเปลี่ยนแปลงกับคะแนนสอบก่อนเรียน โดยใช้สัญลักษณ์ D_θ

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

เมื่อ π_1, π_2 คือ คะแนนมวลความรู้ก่อนเรียนและหลังเรียนที่ประมาณค่าตามแนวคิดของบอค (Bock) ในทฤษฎีการตอบข้อสอบ

จากการศึกษาการวัดการเปลี่ยนแปลงตามทฤษฎีการตอบข้อสอบ ซึ่งเป็นกลุ่มวิธีหนึ่งในการวัดการเปลี่ยนแนวใหม่ (Modern methods) สามารถสรุปในลักษณะเปรียบเทียบ ได้ดังนี้

วิธี	ผู้พัฒนา	พื้นฐานความคิด	ข้อตกลงเบื้องต้น	ผลที่ได้
1.โมเดลคุณลักษณะ แฝงโลจิสติกแบบ เส้นตรงที่ผ่อนคลาย ข้อตกลงเบื้องต้น (LLRA)	Fischer (1976)	โมเดลราช	1. ค่าการเปลี่ยนแปลงต้อง ไปในทิศทางบวกเท่านั้น 2. ใช้กับข้อมูลที่เป็น ทวิภาค(dichotomous)	คะแนนการ เปลี่ยนแปลง เป็นกลุ่ม

วิธี	ผู้พัฒนา	พื้นฐานความคิด	ข้อตกลงเบื้องต้น	ผลที่ได้
2. โมเดลราชพหุมิติ สำหรับการวัดการ เปลี่ยนแปลงการ เรียนรู้ (MRMLC)	Andersen (1985)	โมเดลราช และ โมเดลวิเคราะห์ โครงสร้าง	ใช้ข้อคำถามชุดเดิมใน การสอบซ้ำ	คะแนนการ เปลี่ยนแปลง เป็นกลุ่ม
	Embretson (1989)	โมเดลราช และ โมเดลทั้งโครงสร้าง วีเนอร์ซิมเพิลิกซ์	1. ใช้กับข้อมูลที่เป็น ทวิภาค (dichotomous) 2. การวัดผู้เรียนต้องใช้ ความสามารถเพิ่มขึ้น แบบสะสมตั้งแต่ 2 ครั้ง ขึ้นไป	1. คะแนนการ เปลี่ยนแปลงเป็น รายบุคคล 2. คุณภาพของ ข้อสอบ (ความ ยาก)
3. การวัดการเปลี่ยน แปลงความสามารถ แฝง (d_0)	May; & Nicewander (1998)	โมเดลการตอบข้อ สอบ และการวัดความ แตกต่าง	1. เครื่องมือวัดมิติเดียว (Unidimensionality) 2. ประมาณค่าพารามิเตอร์ ด้วยวิธี EAP	คะแนนการ เปลี่ยนแปลง เป็นรายบุคคล
4. การวัดคะแนน พัฒนาการความ สามารถที่แท้จริง สัมพัทธ์ (S_0, V_0)	อวยพร เรืองตระกูล (2544)	แนวคิดของ ศิริชัย กาญจนวาลี (2532) และแนวคิดของ วินิจ เทือกทอง (2537)	1. เครื่องมือวัดมิติเดียว (Unidimensionality) 2. ประมาณค่าพารามิเตอร์ด้วยวิธี EAP	คะแนนการ เปลี่ยนแปลง เป็นรายบุคคล
5. การวัดคะแนน ความแตกต่าง ระหว่างคะแนน มวลความรู้ (D_0)	Heidenberg (2000)	โมเดลการตอบ ข้อสอบและการวัด ความแตกต่าง ระหว่างคะแนนจริง	เครื่องมือวัดมิติเดียว (Unidimensionality)	คะแนนการ เปลี่ยนแปลงเป็น รายบุคคล

คะแนนมวลความรู้ (Domain scores)

คะแนนมวลความรู้ (Domain scores) เป็นคะแนนตัวแทนมวลความรู้ที่ได้จากการสอบด้วยแบบทดสอบที่ประกอบไปด้วยข้อสอบที่มีการวัดจุดประสงค์การเรียนรู้เดียวกันที่นิยามไว้อย่างชัดเจน ซึ่งเดิมนักวิชาการศึกษาใช้คะแนนมวลความรู้ในการประมาณค่าความสามารถจริงของผู้สอบ และคำนวณหาคะแนนจุดตัดของแบบทดสอบอิงเกณฑ์ ในทฤษฎีการทดสอบแบบมาตรฐานเดิม ต่อมา นักวิชาการศึกษาได้นำหลักการของทฤษฎีการตอบข้อสอบมาใช้วัดคะแนนมวลความรู้ ซึ่งผู้วิจัยได้สรุปไว้ดังนี้

คะแนนมวลความรู้ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม

คะแนนมวลความรู้ (Domain scores) ตามทฤษฎีการทดสอบแบบมาตรฐานเดิมเป็นคะแนนที่ได้จากจำนวนข้อที่ตอบแบบทดสอบได้ถูกต้อง (X) ของข้อสอบที่สุ่มมาจากประชากรข้อสอบเพื่อไปประมาณค่าคะแนนจริงในการวัดแบบอิงเกณฑ์ (Criterion-referenced measurement) เนื่องจากแบบทดสอบอิงเกณฑ์ สร้างตามจุดประสงค์การเรียนรู้โดยแยกเป็นฉบับย่อย ส่วนใหญ่จะมีจำนวนข้อน้อย จึงจำเป็นต้องมีการประมาณค่าคะแนนมวลความรู้เพื่อจะได้ทราบความสามารถแท้จริงของผู้สอบ การประมาณค่าคะแนนมวลความรู้ตามทฤษฎีการทดสอบแบบมาตรฐานเดิมของการวัดแบบอิงเกณฑ์ นั้น มี 5 วิธี (Hambleton.; et al. 1978: 5 -9) ดังนี้

1. คะแนนสัดส่วนจำนวนตอบข้อสอบถูก (Proportion – correct score estimate) คำนวณจากสัดส่วนระหว่างจำนวนข้อสอบที่ตอบถูกทั้งหมดในฉบับกับจำนวนข้อสอบทั้งฉบับ (x/n) เป็นวิธีที่ง่าย แต่พบว่าถ้าแบบทดสอบมีจำนวนข้อสอบน้อยมากเกินไปแล้ว คะแนนมวลความรู้ที่คำนวณด้วยวิธีนี้จะมีความเชื่อถือได้ต่ำ

2. โมเดลแบบมาตรฐานเดิม II (Classical model II estimate) เป็นการประมาณค่าคะแนนจริงจากการถดถอยที่ใช้แนวคิดของเคลลี (Kelley. 1927) ที่ใช้คะแนนสอบของผู้สอบกับคะแนนเฉลี่ยของกลุ่มผู้สอบ ต่อมาแจ๊คสัน (Jackson. 1972) ได้พัฒนาวิธีการดังกล่าวมาใช้สำหรับข้อมูลแบบสองค่า (Binary data) และอาศัยหลักการแปลงคะแนนตามวิธีของฟรีแมน-ตุ๊กกี (Freeman-Tukey) ซึ่งเป็นคะแนนมวลความรู้ที่แปลงมา วิธีนี้พบว่าแบบทดสอบควรมีข้อสอบอย่างน้อย 8 ข้อ จึงจะทำให้คะแนนสอบที่แปลงมา มีการแจกแจงเป็นแบบโค้งปกติ

3. โมเดลเบย์ส์ (Bayesian model II estimate) พัฒนาขึ้นโดยโนวิก เลวิส และแจ๊คสัน (Hambleton.; et al. 1978: 7; citing Novick; Lewis; & Jackson. 1973. Psychometrika.) เป็นการนำสารสนเทศเดิม คือ พื้นความรู้เดิมของผู้เรียนมาร่วมคิดคำนวณ เพื่อให้การประมาณค่าคะแนนมวลความรู้ถูกต้องมากยิ่งขึ้น

4. ค่าเฉลี่ยกลุ่มย่อย (Marginal mean estimate) พัฒนาขึ้นโดยเลวิส แวง และ โนวิก (Lewis; Wang; & Novick. 1973) มีข้อตกลงเบื้องต้นว่า คะแนนมวลความรู้ไม่ขึ้นอยู่กับพื้นความรู้เดิม

5. กึ่งเบย์ส์ (Quasi bayesian estimate) เป็นวิธีแบบโมเดลเบย์ส์กับวิธีค่าเฉลี่ยกลุ่มย่อย ที่คะแนนมวลความรู้ขึ้นอยู่กับพื้นความรู้เดิม

ซึ่งแฮมเบิลตัน ฮัทเทน และสวามินาธาน (Hambleton.; et al. 1978. 9; citing Hambleton; Hutten; & Swaminathan. 1976. Journal of Experimental Education.) ได้เปรียบเทียบคุณภาพของวิธีการประมาณค่าคะแนนมวลความรู้ทั้ง 5 วิธี ด้วยการจำลองข้อมูล ภายใต้เงื่อนไขต่างๆ ได้แก่ ขนาดกลุ่มตัวอย่าง ความยาวของแบบทดสอบ ความเป็นเอกพันธ์ของการแจกแจงคะแนนมวลความรู้ คุณลักษณะ

ของสารสนเทศเดิม และคะแนนจุดตัด พบว่า วิธีการที่ใช้ในกระบวนการประมาณค่าของเบส์เป็นวิธีที่ดีที่สุดสำหรับการประมาณค่าคะแนนมวลความรู้ในทุกเงื่อนไข เนื่องจากมีค่าความคลาดเคลื่อนน้อยที่สุด

คะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ

บอค ทิสเซน และซิมโมสกี (Bock; Thissen; & Zimowski. 1997: 197-211) ได้เสนอวิธีการที่ใช้ทฤษฎีการตอบข้อสอบในการประมาณค่าคะแนนมวลความรู้ (Domain score) สำหรับแบบทดสอบเพื่อการรอบรู้ในเนื้อหา โดยให้ความหมายของคะแนนมวลความรู้ (Domain score) หมายถึง สัดส่วนของมวลความรู้ที่รอบรู้ของบุคคลที่ได้ร้อยละเท่าไรของมวลความรู้ทั้งหมด ดังนั้น มวลความรู้แทนด้วยคลังข้อสอบ (Bank of items) ที่รู้ค่าพารามิเตอร์แล้ว โดยที่การประมาณค่าคะแนนจำนวนที่ตอบข้อสอบถูกต้องที่คาดหวัง (Expected number correct score) และ คะแนนร้อยละของการตอบถูก (Percent-correct score) สามารถคำนวณได้จากสูตร

$$d(\hat{\theta}) = \frac{\sum_{j=1}^n w_j P_j(\hat{\theta})}{\sum_{j=1}^n w_j}$$

$$D^*(\hat{\theta}) = 100d(\hat{\theta})$$

เมื่อ w_j เป็นน้ำหนักของการสุ่ม

$P_j(\hat{\theta})$ เป็นความน่าจะเป็นของการตอบข้อสอบถูก ด้วยความสามารถ $\hat{\theta}$ โดยมีค่าความคลาดเคลื่อนมาตรฐานของการวัดคะแนนร้อยละมวลความรู้ คือ

$$SE(D) = \frac{100SE(\hat{\theta})}{\sum_{j=1}^n w_j} \sqrt{\left[w_j^2 \left[\frac{\partial P_j(\hat{\theta})}{\partial \theta} \right]^2 \right]_{\hat{\theta}}}$$

เมื่อ $\left[\frac{\partial P_j(\hat{\theta})}{\partial \theta} \right]_{\hat{\theta}}$ เป็นอนุพันธ์ของฟังก์ชันการตอบข้อสอบข้อที่ j ด้วยความ

สามารถ θ ที่ประมาณค่าด้วย $\hat{\theta}$

$SE(\hat{\theta})$ เป็นความคลาดเคลื่อนมาตรฐานของคะแนนความสามารถตามทฤษฎีการตอบข้อสอบ

ในขณะที่ แฮมเบลตัน สวามินาธาน และโรเจอร์ (Hambleton; Swaminathan; & Roger. 1991: 84-88) ได้กล่าวถึง คะแนนมวลความรู้ เป็นคะแนนแปลงจากความสามารถ (Ability) เป็นคะแนนจริง (True score) ซึ่งได้อธิบายไว้ ดังนี้

ให้ X เป็นคะแนนที่เกิดจากจำนวนข้อสอบที่ตอบถูก (Number-right score) ที่นิยามดังนี้ $X = \sum_{j=1}^n U_j$ เมื่อ U_j มีค่าเป็น 1 หรือ 0 ที่เป็นคำตอบของผู้เรียนที่ตอบข้อคำถามข้อที่ j ถ้าให้ τ เป็นคะแนนจริง (True score) ดังนั้น

$$\tau = E(X) = E\left(\sum_{j=1}^n U_j\right)$$

ซึ่ง

$$E\left(\sum_{j=1}^n U_j\right) = \sum_{j=1}^n E(U_j)$$

ถ้า Y เป็นตัวแปรสุ่มที่มีค่า y_1 และ y_2 ด้วยความน่าจะเป็น P_1 และ P_2 ดังนั้น

$$E(Y) = y_1 P_1 + y_2 P_2$$

ดังนั้น U_j มีค่าเป็น 1 ด้วยค่าความน่าจะเป็น $P_j(\theta)$ และมีค่าเป็น 0 ด้วยความน่าจะเป็น $Q_j(\theta) = 1 - P_j(\theta)$ นั่นก็คือ

$$E(U_j) = 1 \cdot P_j(\theta) + 0 \cdot Q_j(\theta) = P_j(\theta)$$

ดังนั้น

$$\tau = \sum_{j=1}^n P_j(\theta)$$

นั่นคือ คะแนนจริง (True score) ของผู้เรียนที่มีความสามารถ θ คือ ผลรวมของโค้งลักษณะข้อสอบ (Item characteristic curve) ซึ่งคะแนนจริง คือ τ เรียกว่า โค้งลักษณะแบบทดสอบ (Test characteristic curve) เพราะว่าเป็นผลรวมโค้งลักษณะข้อสอบ ซึ่งในลักษณะที่มีความเข้มงวดแล้ว ความสัมพันธ์ข้างต้นจะเป็นจริงก็ต่อเมื่อโมเดลการตอบข้อสอบ (Item response model) มีความสอดคล้องกลมกลืนกับข้อมูล ดังจะเห็นได้จาก ค่าคะแนนจริง τ ที่เป็นตัวบ่งชี้ถึงค่า $\tau|\theta$ นั่นคือ

$$\tau|\theta = \sum_{j=1}^n P_j(\theta)$$

ซึ่งสามารถใช้ τ แทน $\tau|\theta$ ได้ คะแนนจริง τ และ θ มีความสัมพันธ์กันแบบทางเดียว (Monotonic) นั่นคือ คะแนนจริงอาจถูกแปลงจากค่า θ ในรูปที่ไม่ใช่เส้นตรง (Nonlinear) ในเมื่อ $P_j(\theta)$ มีค่าอยู่ระหว่าง 0 ถึง 1 และ τ มีค่าอยู่ระหว่าง 0 ถึง n ด้วยเหตุนี้ ค่า τ จึงอยู่บนมาตรเดียวกันกับคะแนนที่ได้จากจำนวนข้อสอบที่ตอบถูก (Number-right score) การประมาณค่าจาก θ ไป

เป็นค่า τ มีประโยชน์สำหรับการรายงานค่าความสามารถ โดยรายงานค่า τ ที่มีค่าอยู่ระหว่าง 0 ถึง n แทนการรายงานค่า θ

ค่า τ เป็นสัดส่วนของคะแนนจริงที่ตอบถูก (True proportion correct) หรือ คะแนนมวลความรู้ (Domain score) ที่เกิดจากการหาค่า τ ด้วยจำนวนข้อสอบ n ดังนั้น

$$\pi = \frac{1}{n} \sum_{j=1}^n P_j(\theta)$$

โดย $-\infty < \theta < \infty$ และ $0 < \pi < 1$ หรือ อยู่ในเทอมของร้อยละซึ่งมีค่าอยู่ระหว่าง 0% ถึง 100% ขีดจำกัดล่างของ π ในโมเดลโลจิสติกแบบหนึ่งพารามิเตอร์และสองพารามิเตอร์ มีค่าเป็น 0 ส่วนในโมเดลโลจิสติกแบบสามพารามิเตอร์ เมื่อ θ เข้าใกล้ $-\infty$ แล้ว $P_j(\theta)$ เข้าใกล้ค่า c_j ที่แอดซิมโทตล่าง ดังนั้นขีดจำกัดล่างของ π คือ $\sum c_j/n$ และขีดจำกัดล่างของ τ คือ $\sum c_j$

ความสำคัญของการแปลงค่า θ ไปเป็นคะแนนจริง (True score) หรือ คะแนนมวลความรู้ (Domain score) มีดังนี้

1. เป็นการกำจัดค่าความสามารถที่เป็นลบ
2. ช่วยในการแปลความหมายของคะแนนได้ง่ายและชัดเจนขึ้น เนื่องจากการแปลงคะแนนให้อยู่ในมาตราที่มีช่วงอยู่ระหว่าง 0 ถึง 1 หรือ 0% ถึง 100%

ตัวอย่างการคำนวณคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ โดย แฮมเบิลตัน สวามินาธาน และโรเจอร์ (Hambleton; Swaminathan; & Roger. 1991: 86-87) ข้อสอบ 5 ข้อที่มีค่าพารามิเตอร์ที่คำนวณด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ ดังนี้

Item	b	a	c
1	-2.00	0.80	0.00
2	-1.00	1.00	0.00
3	0.00	1.20	0.10
4	1.00	1.50	0.15
5	2.00	2.00	0.20

ที่มา: Hambleton, Swaminathan, & Roger. (1991). Fundamentals of Item Response Theory. p. 86.

ขั้นตอนการคำนวณค่ามวลความรู้ตามทฤษฎีการตอบข้อสอบ มีดังนี้

1. คำนวณค่าความน่าจะเป็นของคำตอบที่ตอบถูกโดยใช้โมเดลโลจิสติกแบบสามพารามิเตอร์
ความสามารถ $\theta = -3, -2, -1, 0, 1, 2, 3$ คือ ค่า $P(\theta)$

2. รวมค่าความน่าจะเป็นทั้ง 5 ข้อ ในแต่ละความสามารถ นั่นก็คือ $\sum_{j=1}^5 P_j(\theta)$

3. คะแนนมวลความรู้ มีค่าเท่ากับผลรวมความน่าจะเป็นหารด้วยจำนวนข้อ คือ 5 เท่ากับ $\sum_j P_j(\theta) / 5$ ซึ่งผลแสดงความสัมพันธ์ระหว่างคะแนนมวลความรู้กับความสามารถ ดังนี้

Theta	P1	P2	P3	P4	P5	SUM P	Domain Scores
-3	0.2042	0.0323	0.1020	0.1500	0.2000	0.6885	0.1377
-2	0.5000	0.1151	0.1150	0.1504	0.2000	1.0804	0.2161
-1	0.7958	0.5000	0.2036	0.1552	0.2000	1.8545	0.3709
0	0.9382	0.8455	0.5500	0.2116	0.2009	2.7462	0.5492
1	0.9834	0.9677	0.8964	0.5750	0.2258	3.6484	0.7297
2	0.9957	0.9939	0.9850	0.9384	0.6000	4.5131	0.9026
3	0.9989	0.9989	0.9980	0.9948	0.9742	4.9648	0.9930

ที่มา: Hambleton., Swaminathan., & Roger. (1991). Fundamentals of Item Response Theory. p. 87.

คะแนนมวลความรู้ในแบบทดสอบพหุมิติหรือหลายมิติ (Multidimensional Domains)

การประมาณค่าคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ ได้ขยายขอบข่ายไปยังกรณีหลายมิติ ซึ่งในผลสัมฤทธิ์ทางการเรียนของนักเรียนไม่สามารถอธิบายความสามารถในลักษณะของมิติเดียว (Single dimension) แต่เป็นการตอบข้อสอบที่เป็นแบบพหุมิติหรือหลายมิติ (Multidimensional) ซึ่งมีลักษณะคล้ายกับโมเดลมิติเดียว (Unidimensional models) แต่เป็นการรวมองค์ประกอบอื่นๆ ในการตอบข้อสอบซึ่งเป็นการเพิ่มมิติ นักเรียนแต่ละคนมีความสามารถในการแต่ละมิติในการตอบข้อสอบหรือแบบฝึกได้ถูกต้อง โดยในแต่ละแบบฝึกหรือข้อสอบมีค่าพารามิเตอร์ที่ได้กำหนดน้ำหนักความสำคัญของแต่ละมิติไว้เฉพาะ ซึ่งค่าความสามารถสอดคล้องกับคะแนนองค์ประกอบของผู้สอบ และค่าพารามิเตอร์ข้อสอบสอดคล้องกับค่าน้ำหนักองค์ประกอบของแบบฝึกหรือข้อสอบ เช่น สมมติฟังก์ชันการตอบข้อสอบ

หนึ่งที่มีการแจกแจงแบบปกติ คือ $P_j(\theta) = \phi(y_j)$ เมื่อ y เป็นค่าเบี่ยงเบนมาตรฐาน ดังนั้น ค่าเบี่ยงเบนในโมเดล m มิติ คือ

$$y_j = \gamma_j + \alpha_{j1}\theta_1 + \alpha_{j2}\theta_2 + \dots + \alpha_{jm}\theta_m.$$

เมื่อ γ_j เป็นค่าความยากของข้อสอบ

α เป็นค่าน้ำหนักองค์ประกอบ (Factor loading)

โมเดลเหล่านี้เป็นจริงทั้งข้อสอบที่เป็นแบบสองค่า (Binary) (Bock; & Aitkin. 1981) และ แบบหลายค่า (Muraki; & Carlson. 1995) การประมาณค่าพารามิเตอร์อาจเกิดจากการวิเคราะห์องค์ประกอบข้อสอบด้วยทฤษฎีการตอบข้อสอบที่ให้สารสนเทศที่สมบูรณ์ (Full-information IRT item factor analysis)

ชุดของข้อสอบที่เป็นพหุมิติมีลักษณะคล้ายกับการรวมกลุ่มแบบฝึกที่เหมือนกันกับกลุ่มย่อยของนักเรียนในประชากรที่ตอบแตกต่างกัน เช่น แบบฝึก แบบทดสอบความรู้เรื่องคำที่ขึ้นอยู่กับการอ่านออกเสียงภาษาอังกฤษที่สุ่มจากพจนานุกรมที่เป็นการรวมคำที่ปรากฏมากที่สุดในบทความของการอ่านออกเขียนได้และเรื่องแต่งขึ้นที่ได้รับความนิยม เช่น คำตรงข้ามของคำร้อยแก้วที่ปรากฏในหนังสือพิมพ์หรือสื่ออื่นๆ ซึ่งสามารถคำนวณคะแนนมวลความรู้โดยใช้โมเดลพหุมิติ (Multidimensional models) เหมือนการคำนวณชุดข้อสอบแบบมิติเดียว (Unidimensional) แต่มีคุณสมบัติในการประมาณค่าคะแนนมวลความรู้ดีกว่า สเทาท์ (Stout. 1990) ได้กล่าวว่าโดยทั่วไปแล้ว แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนได้พิจารณาความเป็นมิติเดียวจากมิติที่เด่น (Dominant) เพียงมิติเดียวของแบบทดสอบ แม้ว่า จะพบว่าแบบทดสอบเป็นแบบพหุมิติแต่ก็ได้แบ่งเนื้อหาและรูปแบบออกเป็นส่วนๆ เพื่อจัดเป็นแบบทดสอบย่อยที่มีลักษณะเป็นมิติเดียว

คะแนนมวลความรู้ในข้อคำถามปลายเปิด (Open-Ended Items)

การนำคะแนนมวลความรู้ไปใช้กับการวัดและประเมินผลที่ใช้แบบทดสอบอัตนัย (Essays) และ ข้อสอบปลายเปิด (Open-ended items) หรือการให้เกรดการปฏิบัติงานของนักเรียน ซึ่งฟังก์ชันของการให้คะแนนเหล่านี้ให้สารสนเทศจากการตอบข้อสอบของนักเรียนได้มากกว่าการให้คะแนนจากการให้คะแนนแบบสองค่า คือ ถูกให้ 1 คะแนน ผิดให้ 0 คะแนน

คะแนนมวลความรู้ตามทฤษฎีการทดสอบแบบมาตรฐานเดิมของการให้คะแนนหลายค่า นั้นเป็นจำนวนหรือสัดส่วนของคำตอบที่ปรากฏในแต่ละกลุ่มคำตอบ ดังโมเดลของ มูรากิ และซามิไจมา (Muraki. 1992.; Samejima. 1969) ซึ่งกระบวนการประมาณค่าความสามารถของผู้สอบมีลักษณะคล้ายกับกระบวนการประมาณค่าพารามิเตอร์ของข้อสอบที่ให้คะแนนแบบสองค่า (Binary-scores items) ดังนั้น ถ้า $P_j(k|\theta)$ เป็นความน่าจะเป็นที่ k ให้โดยโมเดลที่ใช้ในการประมาณค่าจากคำตอบของ

ผู้สอบ ณ ความสามารถที่ θ ในข้อสอบข้อที่ j ของมวลความรู้จะเกิดขึ้นในกลุ่มคำตอบ k เมื่อ $k = 1, 2, \dots, m$ ดังนั้น จำนวนของคำตอบที่ถูกประมาณค่าจากผู้สอบในกลุ่มคำตอบ คือ $d(k, \hat{\theta}) = \sum_{j=1}^n P(k/\hat{\theta})$ และ ร้อยละของมวลความรู้ คือ $D(k, \hat{\theta}) = 100d/n$ ในทำนองเดียวกัน ถ้า $\sum_k P_j(k/\hat{\theta}) = 1$ คะแนนมวลความรู้ของการให้คะแนนหลายค่า คือ $R(\hat{\theta}) = \sum_{j=1}^n P(k, \hat{\theta})/n$

งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ

ลิน และ ชุง (Lin; & Hsiung. 1994: 331-359) ศึกษาการประมาณค่ามวลความรู้ด้วยวิธีเบส์ (EABEs; Simple empirical approximate Bayes estimators) 2 แบบ คือ $\tilde{d}_N(x)$ และ $\tilde{\delta}_N(x)$ ของคะแนนแบบทดสอบที่มีการแจกแจงเป็นแบบไบโนเมียลและไฮเปอร์จีโอเมทริกซ์ ด้วยการเปรียบเทียบผลที่เกิดขึ้นในสถานการณ์จำลองข้อมูลกับข้อมูลจริง โดยข้อมูลจริงเป็นผลจากการสอบเข้าเรียนในระดับวิทยาลัยในไต้หวัน ของนักเรียนระดับมัธยมศึกษาจำนวน 73,595 คน ด้วยแบบทดสอบที่เป็นแบบเลือกตอบจำนวน 50 ข้อ ที่มีการให้คะแนนแบบทวิภาค ส่วนข้อมูลจำลอง เป็นการจำลองคะแนนสอบจากแบบทดสอบที่มีความยาวต่างกัน 3 แบบ คือ 10, 15 และ 20 ข้อ ที่สุ่มมาจากคลังข้อสอบ 50 และ 100 ข้อ โดยที่แบบทดสอบจำนวน 50 ข้อ มีการแจกแจงเป็นแบบโค้งเดียว (Unimodal) ส่วนแบบทดสอบจำนวน 100 ข้อ มีการแจกแจงเป็นแบบโค้งเดียว (Unimodal) และสองโค้ง (Bimodal) แบบไม่ใส่คั่นซึ่งเป็นการแจกแจงแบบไบโนเมียล ด้วยการสุ่มกลุ่มตัวอย่างที่มีจำนวนอยู่ระหว่าง 132 ถึง 209 คน จากประชากรผู้สอบ 50,000 คน แบบใส่คั่น ซึ่งเป็นการแจกแจงแบบไฮเปอร์จีโอเมทริกซ์

ผลการศึกษาพบว่า ตัวประมาณค่าทั้ง $\tilde{d}_N(x)$ และ $\tilde{\delta}_N(x)$ ขึ้นอยู่กับความยาวของแบบทดสอบ ขนาดกลุ่มตัวอย่างและอัตราส่วนระหว่างความยาวแบบทดสอบกับจำนวนข้อสอบในมวลความรู้ โดยตัวประมาณค่า $\tilde{\delta}_N(x)$ มีค่าความความคลาดเคลื่อนเฉลี่ยกำลังสองน้อยกว่า $\tilde{d}_N(x)$ ในคะแนนมวลความรู้ที่มีการแจกแจงเป็นแบบโค้งเดียว (Unimodal distribution) ที่มีความยาวของแบบทดสอบน้อยข้อ หรือข้อสอบในแบบทดสอบมีความเป็นวิวิธพันธ์ (Heterogeneity) สูง ส่วนตัวประมาณค่า $\tilde{d}_N(x)$ มีค่าความความคลาดเคลื่อนเฉลี่ยกำลังสองน้อยกว่า $\tilde{\delta}_N(x)$ ในคะแนนมวลความรู้ที่มีการแจกแจงแบบสองโค้ง (Bimodal distribution) และแบบทดสอบมีลักษณะเป็นเอกพันธ์ (Homogeneous)

บอค และ ทิสเซน (Bock; Thissen; & Zimowski. 1997: 198-200; citing Bock; & Thissen. 1990. One hundred-word spelling test.) ศึกษาความแม่นยำในการประมาณค่าความสามารถด้วยคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบจากแบบทดสอบที่มีความยาวแตกต่างกัน โดยการสุ่ม

คำศัพท์ 100 คำ จากหนังสือที่เลขานุการต้องใช้ ประกอบด้วย 25,000 คำ แล้วนำคำศัพท์ไปอ่านเพื่อ บันทึกเทปในสองลักษณะคือ เป็นคำและเป็นประโยคที่มีคำศัพท์ที่สุ่มมา ให้ผู้สอบคือนักศึกษาระดับมหาวิทยาลัยเขียนคำศัพท์จากการฟังคำหรือประโยคละ 2 ครั้ง ภายใน 10 วินาที นำมาตรวจให้คะแนนแบบทวิภาค (Dichotomous) ประเมินค่าพารามิเตอร์ข้อสอบด้วยโมเดลโลจิสติกแบบสองพารามิเตอร์ด้วยโปรแกรม BILOG-MG ที่พัฒนาขึ้นโดย ซิมโมสกี และคนอื่นๆ (Zimowski.; et al. 1996.) พบว่ามีข้อสอบที่ดีจำนวน 89 ข้อ ($P > 0.01$) ต่อจากนั้นสุ่มคำศัพท์จาก 100 คำ เพื่อนำมาจัดชุดแบบทดสอบจำนวนฉบับละ 15 และ 20 ข้อ แล้วประมาณค่าความสามารถของผู้สอบจำนวน 1,000 คน ด้วย 3 วิธี ได้แก่ 1) คะแนนจำนวนข้อที่ตอบถูก (Number-correct score) 2) คะแนนร้อยละที่ตอบถูกของมวลความรู้ตามทฤษฎีการตอบข้อสอบ (IRT domain percent-correct score) ที่ประมาณค่าด้วยวิธีความเป็นไปได้สูงสุด (Maximum likelihood) และ 3) คะแนนร้อยละที่ตอบถูกของมวลความรู้ตามทฤษฎีการตอบข้อสอบ (IRT domain percent-correct score) ที่ประมาณค่าด้วยวิธีเบย์ส (Bayes)

ผลการศึกษาพบว่า การประมาณค่าความสามารถด้วยคะแนนร้อยละที่ตอบถูกของมวลความรู้ตามทฤษฎีการตอบข้อสอบมีความแม่นยำมากกว่าคะแนนจำนวนข้อที่ตอบถูก (Number-correct score) และพบว่า การประมาณค่าความสามารถด้วยคะแนนร้อยละที่ตอบถูกของมวลความรู้ตามทฤษฎีการตอบข้อสอบด้วยวิธีเบย์สมีความแม่นยำมากกว่าการใช้วิธีความเป็นไปได้สูงสุด (Maximum likelihood) ทั้งในแบบทดสอบที่มีจำนวน 15 และ 20 ข้อ จึงได้เสนอแนะให้ใช้การประมาณค่าความสามารถด้วยวิธีของเบย์ส เนื่องจากมีความคงที่ ณ ค่าความสามารถที่อยู่ปลายๆ ของการกระจายมากกว่า

บอค ทิสเซน และ ซิมโมสกี (Bock; Thissen; & Zimowski. 1997: 204-207) ศึกษาความแม่นยำของคะแนนมวลความรู้ที่ประมาณค่าตามทฤษฎีการตอบข้อสอบกับทฤษฎีการทดสอบแบบมาตรฐานเดิมในแบบทดสอบพหุมิติ (Multidimensional) โดยใช้ข้อมูลจากการสอบแบบทดสอบวิชาวิทยาศาสตร์ทั่วไป จำนวน 57 ข้อ กับผู้ใหญ่วัยต้นจำนวน 2,902 คน ซึ่ง ซิมโมสกี และบอค (Zimowski; & Bock: 1987) ได้วิเคราะห์องค์ประกอบของแบบทดสอบด้วยการหมุนแกนแบบแวกซ์แม็ก ด้วยโปรแกรม TESTFACT พบว่ามี 3 องค์ประกอบ ได้แก่ ชีววิทยา ฟิสิกส์ และ เคมี มีค่าความแปรปรวนร้อยละ 45.7, 3.7 และ 1.8 ตามลำดับ ต่อจากนั้นได้สุ่มข้อสอบจาก 57 ข้อ มาจัดชุดแบบทดสอบจำนวน 2 ฉบับ ที่มีความยาว 15 และ 20 ข้อ ประมาณค่าความสามารถด้วยคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบที่ใช้วิธีเบย์ส และตามทฤษฎีการทดสอบแบบมาตรฐานเดิม โดยคำนวณค่าความสามารถจากแบบทดสอบแต่ละฉบับทั้งในลักษณะที่เป็นมิติเดียว (Unidimensional) และพหุมิติ (Multidimensional)

ผลการศึกษาพบว่า การประมาณค่าความสามารถด้วยคะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบมีความแม่นยำกว่าทฤษฎีการทดสอบแบบมาตรฐานเดิมทั้งในแบบทดสอบที่มีจำนวนข้อ 15 ข้อ และ 20 ข้อ โดยค่าประมาณความสามารถจากโมเดลการตอบข้อสอบที่เป็นมิติเดียว

(Unidimensional) ที่แยกข้อสอบตามเนื้อหาชีววิทยา ฟิสิกส์ และเคมี มีความแม่นยำในการประมาณค่ามากกว่าการประมาณค่าแบบทดสอบทั้งฉบับด้วยมิติเดียว

พอมเมอริช ไนส์วันเดอร์ และฮาวันสัน (Pommerrich; Nicewander; & Hauson. 1999: 199-216) ศึกษาความถูกต้องและความคงเส้นคงวาของวิธีการประมาณค่าคะแนนมวลความรู้ 7 วิธี คือ EAP1, EAP2, MLE1, MLE2, MU, EM และ OBS ด้วยการจำลองข้อมูลของแบบทดสอบ 9 ฉบับ ซึ่งมี 6 ฉบับ (A2, B2, C2, A3, B3 และ C3) เป็นแบบทดสอบที่ใช้มวลความรู้อิงเนื้อหา ส่วนอีก 3 ฉบับ (A1, B1 และ C1) เป็นแบบทดสอบการปฏิบัติของมวลความรู้ โดยกำหนดลักษณะการแจกแจงของค่าพารามิเตอร์ข้อสอบมี 3 ลักษณะ ดังนี้ ฉบับที่ A1, A2 และ A3 มีค่าเป็น $N(.5,1)$ ฉบับที่ B1, B2 และ B3 มีค่าเป็น $N(0,1)$ และฉบับที่ C1, C2 และ C3 มีค่าเป็น $N(-.5,1)$ ขนาดความยาวของแบบทดสอบที่มีสัดส่วนของการได้มาจากมวลความรู้ที่คงที่ คือ จำนวนข้อน้อยกว่าในคลัง 6 เท่า มี 3 ขนาด คือ 5, 10 และ 20 ข้อ จากขนาดของมวลความรู้ 30, 60 และ 120 ข้อ กับผู้สอบต่อโรงเรียนคือ 25, 50 และ 100 คน ในโรงเรียน 100 แห่ง วิเคราะห์ค่าความถูกต้องจากความแตกต่างสมบูรณ์เฉลี่ย (Average absolute difference; ABSDIF) ซึ่งค่า ABSDIF ควรมีค่าน้อยกว่า .05 ส่วนค่าความคงเส้นคงวาวิเคราะห์จากความเบี่ยงเบนมาตรฐานเฉลี่ย (Average standard deviation)

ผลการวิจัยพบว่า วิธีการประมาณค่าคะแนนมวลความรู้มีความเหมาะสมสำหรับการวัดความถูกต้องโดยค่า ABSDIF มีค่าน้อยกว่า .05 และการวัดความคงเส้นคงวา ทุกจำนวนผู้สอบในแต่ละโรงเรียน ได้แก่ วิธี EM, EAP1, EAP2, MLE1 และ MU ส่วนวิธี OBS และ MLE2 ไม่เหมาะสมในการวัดความถูกต้องและความคงเส้นคงวา โดยที่วิธี EM เป็นวิธีที่เหมาะสมที่สุดทุกสถานการณ์

ซูลซ์ โคลน และไนส์วันเดอร์ (Schulz; Kolen; & Nicewander. 1999: 347-362) ศึกษาความเชื่อมั่น ความคงเส้นคงวาในการจัดกลุ่มผู้สอบและความคลาดเคลื่อนในการจัดกลุ่มผู้สอบ ของการให้คะแนนมวลความรู้ที่ประมาณค่าด้วยทฤษฎีการตอบข้อสอบ ที่เรียกว่า NC-level Scoring และการให้คะแนนแบบกัตแมนสองแบบ คือ จำนวนของระดับที่รอบรู้ (Number of Level mastered; NLM) และจำนวนระดับที่รอบรู้ติดต่อกันสูงสุด (Highest contiguous level mastered) ใช้แบบทดสอบคณิตศาสตร์ประยุกต์ (Applied Mathematics Tests) ของ ACT Inc. จัดเป็น 3 ฉบับที่คู่ขนานกัน ในแต่ละฉบับมี 5 ระดับทักษะๆ ละ 6 ข้อ รวมแบบทดสอบฉบับละ 30 ข้อ โดยในแต่ละระดับมีคลังข้อสอบ 18 ข้อ นำไปใช้กับกลุ่มตัวอย่างที่เป็นนักเรียนระดับมัธยมศึกษาตอนต้นและมัธยมศึกษาตอนปลาย มีจำนวนอยู่ระหว่าง 1,996 – 2,046 คนต่อฉบับ ให้คะแนนแบบสองค่า วิเคราะห์ค่าพารามิเตอร์ข้อสอบด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์

ผลการวิจัยพบว่า การให้คะแนนมวลความรู้ที่ประมาณค่าด้วยทฤษฎีการตอบข้อสอบที่เรียกว่า NC-level Scoring ให้ความเชื่อมั่นสูงกว่า ความคงเส้นคงวาในการจัดกลุ่มสูงกว่า และความคลาดเคลื่อนในการจัดกลุ่มต่ำกว่า เมื่อเทียบกับการให้คะแนนแบบกัตแมนทั้งสองแบบ

ซูลซ์ ลี และมุลเลน (Schulz; Lee; & Mullen. 2005: 1-26) ศึกษาการพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนของนักเรียนเกรด 8 ที่มีความสามารถระดับสูงและระดับต่ำ จำนวน 159 คน ในปี 2000 โดยใช้แบบทดสอบระดับชาติ (National Assessment of Educational Progress; NAEP) ซึ่งประกอบไปด้วย 3 เรื่อง ดังนี้ 1) คณิตศาสตร์ (Mathematics) มีมวลความรู้ 5 เรื่อง ได้แก่ ตัวเลข (Number sense) การวัด (Measurement) ภูมิศาสตร์ (Geometry) การวิเคราะห์ข้อมูล (Data analysis) และพีชคณิต (Algebra) 2) กระบวนการทางพลังขับเคลื่อน (Power process) 3 เรื่อง ได้แก่ เหตุผล (Reasoning) การเชื่อมโยง (Connections) และการติดต่อสื่อสาร (Communication) 3) กระบวนการทางความสามารถ (Ability process) 3 เรื่อง ได้แก่ การแก้ปัญหา (Problem solving) ความรู้ในกระบวนการ (Procedural knowledge) และ ความเข้าใจในความคิดรวบยอด (Conceptual understanding) โดยแบบทดสอบเป็นประเภทเลือกตอบ และตอบคำถามสั้นๆ โดยให้คะแนนแบบสองค่า ให้คะแนนบางส่วน และให้คะแนนแบบหลายค่า โดยที่แบบทดสอบวัดกระบวนการทางการขับเคลื่อนและกระบวนการทางความสามารถ เป็นแบบทดสอบเลือกตอบ ตอบคำถามสั้นๆ ที่ให้คะแนนแบบหลายค่า วิเคราะห์ข้อสอบด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ ส่วนแบบทดสอบเลือกตอบและตอบคำถามสั้นๆ ที่ให้คะแนนสองค่า วิเคราะห์ข้อสอบด้วยโมเดลโลจิสติกแบบสองพารามิเตอร์ ข้อสอบจัดเรียงตามความยากเป็นระดับ

ผลการวิจัยพบว่า เมื่อกำหนดเกณฑ์การรอบรู้ที่ร้อยละ 65 ข้อสอบแบบให้คะแนนหลายค่ามีความยากมากกว่าข้อสอบที่ให้คะแนนแบบสองค่า ในส่วนของข้อสอบคณิตศาสตร์ พบว่า ข้อสอบในมวลความรู้เรื่องการวัดมีความยากมากที่สุด ในขณะที่ข้อสอบในมวลความรู้เรื่องตัวเลขง่ายที่สุด ในข้อสอบกระบวนการทางความสามารถ พบว่า ข้อสอบในมวลความรู้เรื่องการแก้ปัญหามีความยากมากที่สุด ในขณะที่ข้อสอบในมวลความรู้เรื่องความรู้ในกระบวนการง่ายที่สุด ต่อมาได้จัดมวลความรู้ใหม่โดยใช้กระบวนการที่ผ่านผู้เชี่ยวชาญหลักสูตร และคณะกรรมการพิจารณาองค์ประกอบมวลความรู้ได้ข้อสอบคณิตศาสตร์จำนวน 14 ข้อ สามารถจัดออกเป็น 3 ระดับ ได้แก่ ขั้นพื้นฐาน (Basic) ขั้นทักษะ (Proficient) และ ขั้นยาก (Advance) พบว่า ทุกมวลความรู้ในคณิตศาสตร์มีโค้งลักษณะของคะแนนมวลความรู้ไม่ตัดกัน และเมื่อใช้เกณฑ์การรอบรู้ที่ร้อยละ 65 หรือสูงกว่า พบว่า ในแบบทดสอบชุดขั้นพื้นฐานมีขอบเขตของคะแนนมวลความรู้ที่คาดหวังอยู่ระหว่างร้อยละ 15-84 ในแบบทดสอบชุดขั้นทักษะมีขอบเขตของคะแนนมวลความรู้ที่คาดหวังอยู่ระหว่างร้อยละ 30-94 และ ในแบบทดสอบชุดขั้นยากมีขอบเขตของคะแนนมวลความรู้ที่คาดหวังอยู่ระหว่างร้อยละ 51-98

งานวิจัยที่เกี่ยวกับการวัดการเปลี่ยนแปลง

งานวิจัยต่างประเทศ

เอ็มเบรทสัน (Embretson. 1991: 495-515) ใช้โมเดลราซพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (Multidimensional Rasch model for learning and change; MRMLC) ในการจำลองข้อมูลและการทดลองจริง สำหรับการจำลองข้อมูลนั้น ได้มีการจำลองข้อมูลข้อสอบ 2 ชุดที่มีจำนวนข้อเท่ากันทั้งในแบบทดสอบก่อนเรียนและหลังเรียน จำนวนฉบับละ 24 ข้อ โดยมี 4 ข้อ ที่เป็นข้อสอบรวมทั้งในแบบทดสอบก่อนเรียนและหลังเรียน โดยกำหนดให้ทุกค่าของคะแนนเพิ่มในการสอบครั้งที่สองมีค่าเท่ากับ 0.385 ($\theta_2 = \theta_1 + 0.385$) เพื่อตรวจสอบการประมาณค่าความสามารถของผู้สอบจำนวน 1,000 คน ด้วยโมเดล MRMLC ที่ประมาณค่าด้วย CML

ผลการวิจัยพบว่าการประมาณค่าพารามิเตอร์ข้อสอบมีค่าใกล้เคียงกัน มีค่าอำนาจจำแนกค่าการเดาและค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.99, 0.00 และ 0.99 ตามลำดับ แล้วนำไปหาค่าความสามารถของผู้สอบจำนวน 1,931 คน ด้วยกระบวนการแบบย้อนซ้ำด้วยวิธี Multivariate Newton-Raphson ที่ระดับ .01 ได้ค่าสัมประสิทธิ์สหสัมพันธ์และค่าการเดาของความสามารถเริ่มต้น เท่ากับ 1.02 และ .00 ตามลำดับ ค่าสัมประสิทธิ์สหสัมพันธ์และค่าการเดาของความสามารถในการเรียนรู้เพิ่ม เท่ากับ .99 และ .02 ตามลำดับ ซึ่งมีค่าใกล้เคียง 1.00 และ .00 ตามลำดับ สัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดความสามารถและการประมาณค่ามีค่าสูงทั้งสองค่า คือ ค่าความสามารถเริ่มต้น ($r=.91$) และ ค่าความสามารถในการเรียนรู้เพิ่ม ($r=.86$) และได้ตรวจสอบผลการจำลองข้อมูลด้วยการทดลองจริง โดยใช้แบบทดสอบมิติสัมพันธ์ (SLAT) จำนวน 76 ข้อ สอบ 3 ครั้ง แต่ละครั้ง 24 ข้อ และมี 4 ข้อ เป็นข้อสอบเพิ่มในแบบทดสอบก่อนเรียนที่ใช้เป็นข้อสอบรวมเพื่อเชื่อมสำหรับการประมาณค่าระหว่างกลุ่ม ผลการวิจัยพบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง MRMLC กับคะแนนมาตรฐานที่ได้จากการวัดการเปลี่ยนแปลงแบบมาตรฐานเดิมของคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงครั้งที่ 1 และ 2 มีค่าสูงทุกค่า (0.994, 0.953 และ 0.949) ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างโมเดลของ Anderson กับคะแนนมาตรฐานที่ได้จากการวัดการเปลี่ยนแปลงแบบมาตรฐานเดิมของคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงครั้งที่ 1 และ 2 มีค่าต่ำกว่า (0.994, 0.598 และ 0.389)

ฟิชเชอร์ และพาร์เซอร์ (Fischer; & Parzer. 1991: 637-651) ใช้โมเดลมาตราการประมาณค่าเชิงเส้นตรง (Linear rating scale model; LRSM) ในการศึกษาการเปลี่ยนแปลง โดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional maximum likelihood; CML) ด้วยกระบวนการแบบย้อนซ้ำวิธี Newton Raphson algorithm ในการประมาณค่าพารามิเตอร์และใช้อัตราส่วนความเป็นไปได้ (Likelihood ratio) ทดสอบสมมติฐานที่ว่า ไม่มีการเปลี่ยนแปลง โดยใช้ทั้งการจำลองข้อมูล (Simulate) และการทดลองจริง (Empirical) ได้จำลองข้อมูล 2 ชุด มาเป็นกลุ่มตัวอย่างที่สร้างมาจากการให้คะแนนแบบสองค่าด้วยโมเดลราซ จำนวน 100 ข้อ ที่มีค่าพารามิเตอร์ที่มีการแจกแจงเป็นแบบสม่ำเสมอ (Uniform) มีค่าอยู่ระหว่าง -2.0 ถึง + 2.0 มีผู้ตอบ 10,000 คน ส่วนข้อมูลอีกชุดหนึ่งเป็นการให้คะแนนแบบ 5 ค่า จำนวน 30 ข้อ มีค่าพารามิเตอร์ที่มีการแจกแจงเป็นแบบสม่ำเสมอ (Uniform)

มีค่าอยู่ระหว่าง -2.0 ถึง $+2.0$ ผลการวิจัยพบว่า การวัดคะแนนการเปลี่ยนแปลงจากข้อมูลที่ได้คะแนนแบบห้าค่ามีความแม่นยำมากกว่า

สำหรับข้อมูลในการทดลองจริงได้ทดลองกับคนไข้ที่มีความผิดปกติทางจิตของคลินิกในมหาวิทยาลัยเวียนนา จำนวน 87 คน แบ่งออกเป็น 2 กลุ่ม กลุ่มทดลองให้ยาสำหรับการรักษาและเข้ารับการฝึกอบรม 7 สัปดาห์ จำนวน 37 คน กลุ่มควบคุมให้ยาสำหรับรักษาเพียงอย่างเดียว จำนวน 50 คน แล้วให้คนไข้ประเมินอาการของตนเอง ในแบบประเมินที่เป็นมาตรวัดแบบ 4 ระดับ จำนวน 24 ข้อ ประเมินครั้งที่หนึ่งเมื่อมารักษาครั้งแรก และประเมินครั้งที่ 2 เมื่อเวลาผ่านไป 1 ปี ผลการวิจัยพบว่า กลุ่มทดลองไม่มีการเปลี่ยนแปลงเกิดขึ้นจากการสอบสองครั้ง และมีปฏิสัมพันธ์ระหว่างลักษณะของคนไข้กับการจัดกระทำ (Treatment) คือ การให้ยาในการรักษา

ฟิชเชอร์ และพอนอคนี (Fischer; & Ponocny. 1994: 177-192) ใช้โมเดลการให้คะแนนความรู้บางส่วนเชิงเส้นตรง (Linear partial credit model; LPCM) มาใช้วัดการเปลี่ยนแปลงแทนโมเดลมาตราการประมาณค่าเชิงเส้นตรง (Linear rating scale model; LRS) โดยที่ LRS เป็นกรณีเฉพาะของ LPCM ใช้ความเป็นไปได้แบบมีเงื่อนไข (Conditional maximum likelihood; CML) ด้วยวิธี Broyden-Fletcher Goldfarb Shanno (Quasi-Newton) ในการประมาณค่าพารามิเตอร์ เป็นการทดลองกับกลุ่มตัวอย่างเดิมที่ใช้ทดลองด้วยการใช้โมเดลมาตราการประเมินเชิงเส้นตรง (Linear rating scale model; LRS) ได้มีการปรับการให้คะแนนในแบบประเมินตนเองให้เป็นคะแนนเป็นส่วนที่ไม่เท่ากัน

ผลการวิจัยพบว่า กลุ่มทดลองไม่มีการเปลี่ยนแปลงในการอาการป่วย ซึ่งได้ผลการวิจัยสอดคล้องกับการใช้โมเดลมาตราการประมาณค่าเชิงเส้นตรง (Linear rating scale model; LRS)

เอมเบรตสัน (Embretson. 1995: 227-294) ใช้โมเดลราชพหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลง (MRMLC) โดยการนำกระบวนการและโครงสร้างของความรู้เข้าไปในโมเดลราช เรียกว่า MRMLC⁺ เป็นโมเดลที่หาคะแนนการเปลี่ยนแปลงรายบุคคลได้ พิจารณาทั้งความสามารถที่เพิ่มขึ้นระยะเวลา และจากพัฒนาการในกระบวนการเรียนและความรู้ที่เพิ่มขึ้น ซึ่งอยู่ในรูปของค่าพารามิเตอร์ของข้อสอบ เป็นการทดลองกับพนักงานใหม่ของ Air force จำนวน 578 คน ที่ได้รับการอบรมขั้นพื้นฐาน 3 สัปดาห์ แล้วใช้แบบทดสอบการแก้ปัญหาทางคณิตศาสตร์ของ SAT สอบ 3 ครั้งที่ใช้แบบทดสอบไม่ใช่ฉบับเดิม จำนวน 30 ข้อ โดยแบบทดสอบก่อนเรียนและหลังเรียนมีจำนวนฉบับละ 10 ข้อ ที่เป็นเรื่องอัตราส่วน 7 ข้อ และอัตราส่วนที่เพิ่มตัวแปรและจำนวนสมการให้มากขึ้นอีก 3 ข้อ ส่วนแบบทดสอบระหว่างเรียนเป็นเรื่องอัตราส่วน 7 ข้อ มี 3 ข้อที่เป็นข้อสอบรวม ใช้เวลาในการสอบข้อละ 3 นาที แต่ทั้งนี้ให้ใช้เวลาทั้งหมดในกระบวนการสอบภายใน 1 ชั่วโมง จัดสอบแบบจัตุรัสละติน (Latin square)

ผลการวิจัยพบว่า โดยภาพรวมข้อมูลมีความสอดคล้องกลมกลืนกับโมเดล สำหรับการทำนายค่าความยากของข้อสอบเป็นอย่างดีเมื่อเปรียบเทียบกับค่าสัมประสิทธิ์สหสัมพันธ์ นั่นคือ ตัวแปรต่างๆ

ได้แก่ จำนวนคำในโจทย์ปัญหา ระดับความสามารถในการอ่าน จำนวนตัวแปร มีค่าสัมประสิทธิ์ - สหสัมพันธ์สูงกับค่าความยากที่ได้จากโมเดลราช

กลัซ และสเปล (Glück; & Spiel. 1997: 1-18) ได้ทดสอบคุณภาพของทักษะปฏิสัมพันธ์ทางสังคม (Social interaction skills; SIS) ของครูโรงเรียนอนุบาลมีอิทธิพลต่อการพัฒนาทักษะการรับรู้ (Visu-motor) หรือไม่ ดังนั้นจึงเปรียบเทียบเด็ก 2 กลุ่ม ระหว่างกลุ่มนักเรียนที่ครูมีทักษะปฏิสัมพันธ์ทางสังคมไม่ดีกับกลุ่มนักเรียนที่ครูมีทักษะปฏิสัมพันธ์ทางสังคมดี จำนวน 23 และ 71 คน ตามลำดับ ซึ่งจะนำไปพัฒนานักเรียนที่มีครูที่มีทักษะปฏิสัมพันธ์ทางสังคมดี (กลุ่มที่ 2) ข้อมูลจากการทดสอบแบบทดสอบฉบับหนึ่งมีข้อคำถาม 12 ข้อ จากการสอบครั้งที่ t1 จำนวน 6 ข้อ ใน และในการสอบครั้งที่ t2 อีก 6 ข้อ ในโมเดลเชื่อมต่อ (Anchor model) ก็คือโมเดลศูนย์ หรือโมเดลกึ่งสมบูรณ์ (Model 0; Quasi-saturated model) การประมาณค่าพารามิเตอร์ความยากข้อสอบแยกจากกันทั้งสองกลุ่มและทั้งสองครั้ง ในขั้นต่อไปโมเดลหนึ่ง (Model 1) ค่าพารามิเตอร์มีข้อจำกัดรุนแรง ข้อสอบ 24 ข้อจากโมเดลศูนย์ถูกนำมาเป็นตัวอย่างเพื่อทำเหมือนเดิม เพราะว่าการรวมผลบวกของค่าความยากข้อสอบ 6 ข้อในการสอบครั้งที่ 1 และค่าพารามิเตอร์ผลกลุ่มเฉพาะอีก 2 กลุ่ม ข้อจำกัดนี้แสดงให้เห็นถึงข้อตกลงเบื้องต้นที่การเปลี่ยนแปลงในการสอบหลายครั้งเป็นของกลุ่มเฉพาะ ในจำนวนข้อสอบที่เท่ากัน

ผลการวิจัยพบว่า โมเดลหนึ่ง มีความสอดคล้องกลมกลืนอย่างไม่มีนัยสำคัญ จึงใช้กระบวนการ Stepwise สำหรับลดค่าพารามิเตอร์ต่อไป ซึ่งโมเดลศูนย์ให้ผลที่ดีกว่า

ในโมเดลสอง (Model 2) เป็นการประมาณค่าพารามิเตอร์การเปลี่ยนแปลงร่วมกันทั้งสองกลุ่ม แทนที่ค่าพารามิเตอร์ผลกลุ่มเฉพาะ การเปรียบเทียบโมเดลนี้กับโมเดลศูนย์ ซึ่งเป็นการทดสอบสมมติฐานหลักของการวิเคราะห์นี้ คือ การลดลงอย่างมีนัยสำคัญของความสอดคล้องกลมกลืนระหว่างข้อมูลกับโมเดล ซึ่งหมายความว่า ปริมาณการเปลี่ยนแปลงของสองกลุ่มนั้นมีความแตกต่างกันอย่างมีนัยสำคัญ โมเดลนี้ไม่มีความสอดคล้องกลมกลืนกับข้อมูล ซึ่งพบว่าโมเดลหนึ่งให้ผลที่ดีกว่าในกลุ่มสองกลุ่มแสดงปริมาณการเปลี่ยนแปลงที่เท่ากัน ดังนั้น ทักษะปฏิสัมพันธ์ทางสังคมของครูโรงเรียนอนุบาลจึงไม่มีผลต่อการพัฒนาทักษะการรับรู้ (Visu-motor) ของเด็ก

ในโมเดลสาม (Model 3) ซึ่งเป็นโมเดลสุดท้ายนั้น ได้กำจัดค่าพารามิเตอร์การเปลี่ยนแปลงทั้งหมดออก ถ้าสมมติไม่มีการเปลี่ยนแปลงของการสอบ 2 ครั้ง โมเดลนี้แสดงความสอดคล้องกลมกลืนกับข้อมูลอย่างมีนัยสำคัญ เพราะฉะนั้น โมเดลสองเป็นโมเดลที่ประหยัดที่สุดที่อธิบายข้อมูลได้อย่างเพียงพอ นั่นคือ เด็กมีการปรับปรุงระหว่างการสอบทั้งสองครั้งอย่างมีนัยสำคัญ โดยเป็นอิสระจากทักษะปฏิสัมพันธ์ทางสังคมของครู การเปลี่ยนแปลงระหว่างการสอบครั้งที่ t1 กับ t2 ถูกอธิบายด้วยการใช้ค่าพารามิเตอร์ของผลเพียงค่าเดียว

แบร์สซ์ (Glück; & Spiel. 1997: Online; citing Barisch. 1989. Einstellung zur Epilepsie und Einstellungsänderung durch Information.) ประยุกต์ใช้ LLRA เพื่อหาปริมาณการเปลี่ยนแปลง

สำหรับข้อสอบข้อหนึ่งที่ประมาณค่ามาจากความถี่ของแบบแผนการตอบ (Response pattern) ที่มีลักษณะเป็น (0,1) หมายถึง ผู้สอบทำข้อสอบข้อนี้ได้ในการสอบครั้งที่ 1 แต่ทำไม่ได้ในการสอบครั้งที่ 2 และ (1,0) หมายถึง ผู้สอบทำข้อสอบข้อนี้ได้ในการสอบครั้งที่ 1 แต่ทำไม่ได้ในการสอบครั้งที่ 2 ถ้าแบบแผนใดแบบแผนหนึ่งในสองแบบแผนนี้ปรากฏขึ้นนานๆ ครั้ง หรือไม่ปรากฏเลย ซึ่งผลการประมาณค่าพารามิเตอร์อาจเป็นไปได้ในกรณีเช่นนี้ ซึ่งควรใช้โมเดล Hybrid LLRA แทน ซึ่งค่าการเปลี่ยนแปลงต้องเป็นไปในทิศทางบวกเท่านั้น

ผลการวิจัยพบว่า ไม่มีเด็กที่มีการกระทำลดลง นั่นคือ ไม่มีเด็กที่สามารถทำได้ในการสอบครั้งที่ 1 แต่ทำไม่ได้ในการสอบครั้งที่ 2 ซึ่ง LLRA ไม่สามารถนำมาใช้กับข้อมูลที่มีการเปลี่ยนแปลงทางลบได้

กลück และสเปิล (Glück; & Spiel. 1997: Online) ประยุกต์ใช้ Hybrid LLRA ซึ่งมีข้อตกลงเบื้องต้นของความเป็นมิติเดียวสำหรับทุกข้อในแบบทดสอบย่อย “Visu-motor skills” ได้จัดข้อสอบ 3 คู่ ที่สอดคล้องกับความแตกต่างของความยาก ซึ่งคู่ของข้อสอบ 3 ข้อ จะจัดในเทอมของ “Hybrid items” ตามการสอบครั้งที่ 1 และ 2 ดังต่อไปนี้ ข้อสอบคู่ที่ 1 ได้แก่ ข้อที่ 2 (กระโดด) กับ ข้อที่ 1 (ยืนขาเดียว) ข้อสอบคู่ที่ 2 ได้แก่ ข้อที่ 3 (ทำตามแบบโครงสร้างของอิฐ) กับ ข้อที่ 6 (ถือแม่แรงแล้วกระโดด) และ ข้อสอบคู่ที่ 3 ได้แก่ ข้อที่ 4 (ถือแก้วน้ำ) กับ ข้อที่ 5 (วาดวงกลมวงใหม่) โดยแบ่งผู้สอบออกเป็น 2 กลุ่มที่สอดคล้องกับทักษะปฏิสัมพันธ์ทางสังคม (SIS) ของครู โดยกลุ่มที่ 1 เป็นเด็กกลุ่มที่มีครูมีทักษะปฏิสัมพันธ์ไม่ดี จำนวน 23 คน และกลุ่มที่ 2 เป็นเด็กที่มีครูที่มีทักษะ ปฏิสัมพันธ์ทางสังคมดี จำนวน 71 คน จำนวนผู้ตอบข้อสอบทุกข้อที่เป็นแบบ Hybrid มี 94 คน โมเดลนี้ไม่สามารถลดความสอดคล้องกลมกลืนกับข้อมูลลงอย่างมีนัยสำคัญ

ผลการวิจัยพบว่า ในโมเดลสอง ซึ่งประกอบไปด้วยค่าพารามิเตอร์แน้มค่าหนึ่งในข้อคำถามทั้ง 3 ข้อ แสดงว่าสมมติฐานเป็นปริมาณการเปลี่ยนแปลงที่เท่ากันปรากฏขึ้นในทุกกลุ่มและในทุกข้อสอบแบบ Hybrid ทั้ง 3 ข้อ เมื่อโมเดลนี้มีความสอดคล้องกลมกลืนกับข้อมูล สมมติฐานที่ตั้งคือ “ไม่มีการเปลี่ยนแปลง” แสดงว่าการเปลี่ยนแปลงในความยากของข้อสอบมีความแตกต่างกันมากกว่าความยากระหว่างข้อสอบที่ใช้ในการสอบครั้งที่ 1 และครั้งที่ 2 ดังนั้น “ไม่มีการเปลี่ยนแปลง” จึงหมายถึง ความแตกต่างระหว่างความสามารถของเด็กในการสอบครั้งที่ t1 และ t2 เท่ากับความแตกต่างของความยากระหว่างข้อสอบที่ใช้ในการสอบครั้งที่ t1 และ t2

เมย์ และไนส์วันเดอร์ (May; & Nicewander. 1998: 882-897) ศึกษาวิธีการเปลี่ยนแปลงความสามารถแฝง (Latent change) ซึ่งเป็นการเปลี่ยนแปลงที่แท้จริง โดยเสนอการวัดคะแนนการเปลี่ยนแปลงจากผลต่างระหว่างค่าความสามารถตามทฤษฎีการตอบข้อสอบ แบบสามพารามิเตอร์ ในการสอบสองครั้ง คือ ก่อนเรียนและหลังเรียน ($d_s = \theta_2 - \theta_1$) โดยใช้การจำลองข้อมูล (Simulate) โดยได้กำหนดเงื่อนไข ดังต่อไปนี้ 1) โมเดลเชิงเส้นตรง 3 โมเดล คือ โมเดลการเปลี่ยนแปลงคงที่ ($\theta_2 = \theta_1 + 1$) โมเดลการเปลี่ยนแปลงเพิ่มขึ้น ($\theta_2 = 1.2\theta_1 + 1$) และโมเดลการเปลี่ยนแปลงลดลง ($\theta_2 = 0.8\theta_1 + 1$) 2)

ผู้สอบ 3 กลุ่มจำแนกตามระดับความสามารถจริง คือ เก่ง ปานกลาง และอ่อน และ 3) การทดสอบสองแบบ คือ แบบดั้งเดิม และแบบทดสอบปรับเหมาะ โดยแบบทดสอบดั้งเดิมแบ่งตามลักษณะการแจกแจงค่าพารามิเตอร์ข้อสอบเป็น 4 แบบ คือ สมมาตร (Uniform), ปกติ (Symmetric), ง่าย (Easy) และ ยาก (Difficulty) ให้แต่ละแบบมีจำนวน 50 ข้อ ส่วนการทดสอบปรับเหมาะได้จัดคลังข้อสอบเป็น 2 แบบ คือ คลังข้อสอบที่ดีและคลังข้อสอบที่ไม่ดี แล้วใช้การคำนวณคะแนนการเปลี่ยนแปลงด้วยวิธีการเปลี่ยนแปลงความสามารถแฝงที่นำเสนอเพื่อเปรียบเทียบกับคะแนนการเปลี่ยนแปลงที่ได้จากผลต่างระหว่างคะแนนดิบ

ผลการวิจัยพบว่า คะแนนการเปลี่ยนแปลงที่มาจากผลต่างระหว่างคะแนนดิบ ในการสอบครั้งที่ 1 และครั้งที่ 2 ของแบบทดสอบทุกแบบและทุกกลุ่มความสามารถของผู้สอบ และคะแนนการเปลี่ยนแปลงทุกคะแนนบิดเบือนจากคะแนนการเปลี่ยนแปลงที่แท้จริง โดยเฉพาะอย่างยิ่งในผู้สอบกลุ่มเก่งจะได้คะแนนการเปลี่ยนแปลงที่ต่ำกว่าความเป็นจริงมากกว่ากลุ่มอ่อนและกลุ่มปานกลาง สำหรับกลุ่มอ่อนนั้นจะได้คะแนนการเปลี่ยนแปลงต่ำกว่าความเป็นจริงมากในกรณีข้อสอบยาก สำหรับคะแนนการเปลี่ยนแปลงที่ได้จากผลต่างระหว่างคะแนนความสามารถแฝงหรือการเปลี่ยนแปลงความสามารถแฝงที่วิเคราะห์ตามทฤษฎีการตอบข้อสอบ ด้วยวิธี EAP(θ) พบว่า คะแนนการเปลี่ยนแปลงที่ได้มีค่าใกล้เคียงกับคะแนนการเปลี่ยนแปลงที่แท้จริงเกือบทุกกรณี ยกเว้นในกรณีผู้สอบกลุ่มเก่งที่ใช้แบบทดสอบง่ายและกลุ่มอ่อนที่ใช้แบบทดสอบยาก จะได้คะแนนความแตกต่างที่ต่ำกว่าความเป็นจริงเล็กน้อย ซึ่งมีความน้อยเมื่อเทียบกับคะแนนการเปลี่ยนแปลงที่ได้จากผลต่างระหว่างคะแนนดิบ สำหรับแบบทดสอบที่ใช้แบบทดสอบปรับเหมาะนั้น คะแนนการเปลี่ยนแปลงที่ได้จากวิธี EAP(θ) จะมีค่าใกล้เคียงกับคะแนนการเปลี่ยนแปลงที่แท้จริงทุกกรณี

ไฮเดนเบิร์ก (Heidenberg, 2000: Abstract) ได้ศึกษาความสัมพันธ์ระหว่างค่าพารามิเตอร์ข้อสอบกับการแจกแจงค่าความสามารถของคะแนนสอบก่อนเรียน โดยการจำลองข้อมูลด้วยเทคนิคมอนติคาร์โล โดยใช้คะแนนมวลความรู้ที่วิเคราะห์ตามทฤษฎีการตอบข้อสอบในการศึกษาการวัดการเปลี่ยนแปลงภายใต้เงื่อนไขต่างๆ ได้แก่ ลักษณะการแจกแจงของค่าพารามิเตอร์แบบทดสอบสองแบบ คือ แบบสมมาตร (Uniform) และ แบบปกติ (Normal) ความยาวของแบบทดสอบ (30, 40, 50 และ 60 ข้อ) ขนาดกลุ่มตัวอย่าง (200, 500 และ 2,000 คน) ปริมาณของคะแนนการเปลี่ยนแปลงจริง (True change) (0.3, 0.7, และ 1.0 หน่วย) ลักษณะการแจกแจงของคะแนนเปลี่ยนแปลงจริง เป็นแบบปกติ $\{N(.7, .301), N(.7, .426), N(.7, .546) \text{ และ } N(.7, 0)\}$ และจำนวนพารามิเตอร์ของโมเดลโลจิสติก (1PL, 2PL และ 3PL) เพื่อศึกษาความสัมพันธ์ที่สัมพันธ์ระหว่างคะแนนเพิ่มกับคะแนนสอบก่อนเรียน

ผลการวิจัยพบว่า คะแนนสอบก่อนเรียนกับคะแนนเพิ่มมีความสัมพันธ์กันอย่างไม่มีนัยสำคัญในทุกเงื่อนไข แสดงว่าความลำเอียงที่เกิดจากความสัมพันธ์เชิงลบของคะแนนเพิ่มกับคะแนนสอบ

ก่อนเรียนมีค่าลดลงอย่างมาก ซึ่งเป็นกระบวนการที่มีประสิทธิภาพที่อาจสร้างแบบทดสอบเพื่อวัดคะแนนการเปลี่ยนแปลงโดยไม่ต้องคำนึงถึงลักษณะการแจกแจงของความสามารถเริ่มต้น

ฟิชเชอร์ (Fischer, 2003: 3-26) ศึกษาความแม่นยำของคะแนนการเปลี่ยนแปลงที่เกิดจากผลต่างระหว่างคะแนนดิบ โดยอธิบายในลักษณะของช่วงความเชื่อมั่นบนมาตราวัดคุณลักษณะแฝงและความน่าจะเป็นอย่างมีนัยสำคัญของคะแนนเพิ่มภายใต้สมมติฐานกลาง (H_0) ที่ว่า “ไม่มีการเปลี่ยนแปลง” เป็นการเปรียบเทียบ 2 วิธี คือ Asymptotic normal distribution of the maximum likelihood estimator และ Exact conditional distribution of the gain scores ใช้กับคะแนนดิบที่ได้จากจำนวนข้อสอบที่ตอบถูกต้องในการสอบสองครั้ง

ผลการวิจัยพบว่า ความแม่นยำของคะแนนการเปลี่ยนแปลงทั้งสองวิธีขึ้นอยู่กับหลายองค์ประกอบ ได้แก่ ความยาวของแบบทดสอบ การแจกแจงของค่าพารามิเตอร์ข้อสอบในแบบทดสอบ การให้คะแนนแบบทวิภาคหรือพหุภาค ลักษณะของแบบทดสอบ เช่น เป็นมาตราวัดประมาณค่า (Rating scale) หรือ เป็นการให้คะแนนบางส่วน (Partial credit model) รวมทั้งการกำหนดสมมติฐานเลือก (H_1) เป็นแบบข้างเดียว (One-sided) หรือ สองข้าง (Two-sided)

งานวิจัยภายในประเทศ

อรุณี อ่อนสวัสดิ์ (2537) พัฒนาวีธีวัดการเปลี่ยนแปลงการเรียนรู้ในสถานการณ์ที่มีการสอบก่อนเรียนและหลังเรียน และเปรียบเทียบคุณภาพของวิธีวัดการเปลี่ยนแปลงการเรียนรู้ที่พัฒนาขึ้นกับวิธีหาความแตกต่างระหว่างคะแนนจริง วิธีหาความแตกต่างระหว่างความสามารถ วิธีหาผลต่างระหว่างคะแนนดิบ และวิธีวัดการเปลี่ยนแปลงการเรียนรู้ของลอว์รี ในการพัฒนาผู้วิจัยได้อาศัยทฤษฎีการเรียนรู้ของบลูมและคำนึงถึงอิทธิพลคะแนนสูงสุด 4 แนวคิด แต่ละแนวคิดมีทางปฏิบัติเป็น 2 แนวทาง แนวทางที่หนึ่ง วัดด้วยค่าประมาณคะแนนจริง ส่วนอีกแนวทางหนึ่ง วัดด้วยคะแนนดิบ ซึ่งได้เสนอวิธีวัดการเปลี่ยนแปลงการเรียนรู้ได้ 8 วิธี ใช้เทคนิคมอนติคาร์โลในการจำลองปัญหาการวิจัย ในสภาพการเรียนรู้ 3 ลักษณะ (แบบรอบรู้ แบบทั่วไป และแบบกึ่งรอบรู้) ความสัมพันธ์ระหว่างความสามารถก่อนเรียนและหลังเรียน 3 ระดับ (0.4 0.6 และ 0.8) กลุ่มตัวอย่าง 3 ขนาด (30 คน 50 คน และ 100 คน) และแบบทดสอบมีความยาว 3 ระดับ (30 ข้อ 60 ข้อ และ 120 ข้อ)

ผลการวิจัยพบว่า 1) วิธีวัดการเปลี่ยนแปลงการเรียนรู้แต่ละวิธี มีความตรงเชิงเกณฑ์สัมพันธ์สูงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในเกือบทุกแบบแผนการวิจัย 2) วิธีวัดการเปลี่ยนแปลงการเรียนรู้แต่ละวิธี จำแนกอันดับที่ได้สอดคล้องกับวิธีเกณฑ์ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทุกแบบแผนการวิจัยและทุกลักษณะการเรียนรู้ 3) ทุกวิธีวัดที่ผู้วิจัยพัฒนา มีความเชื่อมั่นสูงกว่าวิธีหาผลต่างระหว่างคะแนนดิบ และวิธีวัดการเปลี่ยนแปลงการเรียนรู้ของลอว์รีเกินกว่า 74 ครั้งของการกระทำซ้ำ 100 ครั้ง 4) วิธีวัดการเปลี่ยนแปลงการเรียนรู้ที่อิทธิพลคะแนนสูงสุดเป็นผลเนื่องจากสัดส่วนของคะแนนจริงของ

ความแตกต่างระหว่างคะแนนเต็ม และคะแนนหลังเรียนกับคะแนนจริงก่อนเรียน ซึ่งทำการวัดด้วยการประมาณคะแนนจริง มีความคลาดเคลื่อนมาตรฐานต่ำกว่าวิธีหาผลต่างระหว่างคะแนนดิบ เกินกว่า 98 ครั้งของการกระทำซ้ำ 100 ครั้ง

วินิจ เทือกทอง (2537) เปรียบเทียบประสิทธิภาพของการคำนวณคะแนนเพิ่ม 5 วิธี คือ วิธีดั้งเดิม วิธีเศษเหลือ วิธีของลอร์ด วิธีของศิริชัย และวิธีเทียบส่วนร้อย ตามลักษณะการแจกแจงของคะแนนการวัดก่อนเรียนและคะแนนการวัดหลังเรียน ซึ่งมีรูปแบบการแจกแจงที่แตกต่างกัน ตามสถานการณ์ดังต่อไปนี้ ค่าคะแนนเพิ่ม ค่าสหสัมพันธ์ระหว่างคะแนนการวัดก่อนเรียนกับคะแนนเพิ่ม ขนาดตัวอย่าง ความเบ้ของข้อมูล ความโด่งของข้อมูล การพิจารณาค่าประสิทธิภาพนั้นจะพิจารณาจากค่าความคลาดเคลื่อนที่น้อยที่สุด การดำเนินการวิจัยครั้งนี้ใช้การจำลองสถานการณ์ ด้วยเทคนิคมอนติคาร์โล โดยการคำนวณหาค่าคะแนนเพิ่มของแต่ละวิธีการ ตามแต่ละสถานการณ์จำนวน 1,000 ครั้ง

ผลการวิจัยพบว่า เมื่อไม่ได้คำนึงถึงเพดานของคะแนนการสอบวัดแล้ว วิธีการคำนวณหาค่าคะแนนเพิ่มด้วยวิธีดั้งเดิม หรือวิธีของลอร์ด มีค่าความคลาดเคลื่อนน้อยที่สุด แต่เมื่อมีการปรับแก้เพดานของคะแนนการสอบวัด วิธีการคำนวณหาค่าคะแนนเพิ่มด้วยวิธีเศษเหลือเทียบกับศักยภาพของผู้เรียนมีค่าความคลาดเคลื่อนน้อยที่สุด

ประสิทธิ์ ไชยกาล (2539) เปรียบเทียบประสิทธิภาพของโมเดลอิสระ 3 แบบ ที่ใช้ในการศึกษาตัวแปรที่สัมพันธ์กับการเปลี่ยนแปลงในระยะยาว ของผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ได้แก่ โมเดลอิสระที่มีการวัดการเปลี่ยนแปลงในรูปโมเดลพื้นฐานการวิเคราะห์องค์ประกอบระยะยาว โมเดลอิสระที่มีการวัดการเปลี่ยนแปลงในรูปโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้ตัวเดียว และโมเดลอิสระที่มีการวัดการเปลี่ยนแปลงในรูปโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้หลายตัว กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ นักเรียนชั้นประถมศึกษาปีที่ 6 ของโรงเรียนในสังกัดกรุงเทพมหานคร จำนวน 606 คน การดำเนินการเก็บข้อมูลเป็นลักษณะการเก็บข้อมูลระยะยาว โดยวัดตัวแปรด้านผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ และเจตคติต่อพฤติกรรมการเรียนคณิตศาสตร์ 3 ครั้ง และวัดตัวแปรด้านความถนัดทางคณิตศาสตร์ 1 ครั้ง

ผลการวิจัยสรุปได้ว่า โมเดลอิสระที่มีการวัดการเปลี่ยนแปลงในโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้หลายตัว มีประสิทธิภาพสูงสุดเพราะโมเดลมีความคลาดเคลื่อนต่ำที่สุด และสามารถให้ผลการทดสอบที่สามารถบ่งชี้ได้ว่า ความสัมพันธ์ของตัวแปรที่สัมพันธ์กับการเปลี่ยนแปลงมีความไม่แปรเปลี่ยน รองลงไปคือ โมเดลอิสระที่มีการวัดการเปลี่ยนแปลงในรูปโมเดลพื้นฐานการวิเคราะห์องค์ประกอบระยะยาวและโมเดลอิสระที่มีการวัดการเปลี่ยนแปลงในรูปโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้ตัวเดียว ตามลำดับ สำหรับโมเดลการวัดการเปลี่ยนแปลงระยะยาว พบว่า โมเดลการวัดการเปลี่ยนแปลงในรูปโมเดลการวิเคราะห์องค์ประกอบระยะยาวที่วัดด้วยตัวบ่งชี้หลายตัว มีประสิทธิภาพสูงสุด เพราะโมเดลสามารถประมาณค่าพารามิเตอร์ที่บ่งชี้การ

เปลี่ยนแปลงที่เกิดขึ้นตลอดช่วงเวลาได้ และโมเดลมีความคลาดเคลื่อนต่ำที่สุดด้วย รองลงไปคือ โมเดล การวัดการเปลี่ยนแปลงในรูปโมเดลพื้นฐานการวิเคราะห์องค์ประกอบระยะยาว ตามลำดับ

เอื้อมพร หลินเจริญ (2539) พัฒนาโมเดลการศึกษาการเปลี่ยนแปลงคุณลักษณะทางวิทยาศาสตร์ ของนักเรียน และทำการทดสอบความไม่แปรเปลี่ยนระหว่างเวลา ความไม่แปรเปลี่ยนระหว่างกลุ่มของ โมเดลที่ผู้วิจัยพัฒนาขึ้น โมเดลประกอบด้วยตัวแปรสังเกตได้ 14 ตัวแปร และตัวแปรแฝง 4 ตัวแปร ได้แก่ ลักษณะนักเรียน สภาพแวดล้อมทางครอบครัว คุณลักษณะทางวิทยาศาสตร์ก่อนเรียน และ คุณลักษณะวิทยาศาสตร์หลังเรียน กลุ่มตัวอย่างประกอบด้วยนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่อยู่ใน โรงเรียนสังกัดกรมสามัญศึกษา ในกรุงเทพมหานคร และในจังหวัดกำแพงเพชรที่อยู่ในอำเภอเมืองและ นอกเขตอำเภอเมืองจำนวนทั้งสิ้น 450 คน

ผลการทดสอบพบว่า โมเดลการศึกษาการเปลี่ยนแปลงคุณลักษณะทางวิทยาศาสตร์ที่ผู้วิจัยได้ พัฒนาขึ้น มีความสอดคล้องกับข้อมูลเชิงประจักษ์ ผลการทดสอบไค-สแควร์ มีค่า 46.90 ที่องศาอิสระ เท่ากับ 52 GFI เท่ากับ .99 P เท่ากับ .67 ผลการทดสอบความไม่แปรเปลี่ยนระหว่างเวลาของโมเดล การวัดคุณลักษณะทางวิทยาศาสตร์ พบว่าโมเดลมีความไม่แปรเปลี่ยนระหว่างเวลา แสดงให้เห็นว่า การวัดการเปลี่ยนแปลงทั้ง 2 ครั้ง เป็นการวัดตัวแปรแฝงตัวเดียวกัน ส่วนผลการทดสอบความไม่ แปรเปลี่ยนระหว่างกลุ่มพบว่าโมเดลยังไม่สอดคล้องกับข้อมูลเชิงประจักษ์

ผลการศึกษาและเปรียบเทียบความแตกต่างของคะแนนการเปลี่ยนแปลงระหว่างกลุ่มด้วยสถิติ วิเคราะห์ความแปรปรวนแบบทางเดียวและการทดสอบภายหลัง พบว่า ค่ามัชฌิมเลขคณิตของคะแนน การเปลี่ยนแปลงคุณลักษณะทางวิทยาศาสตร์ของนักเรียนที่ศึกษาในกรุงเทพมหานคร สูงกว่านักเรียน ที่ศึกษาในจังหวัดกำแพงเพชรที่อยู่นอกเขตอำเภอเมือง อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

วีระศักดิ์ คำล้าน (2540) สร้างโมเดลโค้งพัฒนาการของการเปลี่ยนแปลงระยะยาวของ ผลสัมฤทธิ์ทางการเรียนคำศัพท์ภาษาอังกฤษ และวิเคราะห์การเปลี่ยนแปลงระยะยาวแบบพหุระดับของ ผลสัมฤทธิ์ทางการเรียนคำศัพท์ภาษาอังกฤษ โดยการประยุกต์ใช้โปรแกรม เอชแอลเอ็ม ศึกษา กับ นักเรียนชั้นประถมศึกษาปีที่ 6 จำนวน 603 คน เครื่องมือที่ใช้ได้แก่ แบบวัดผลสัมฤทธิ์ทางการเรียน แบบ วัดเจตคติ แบบวัดแรงจูงใจ และแบบวัดเชาวน์ปัญญา สแตนด์ดาร์ด โปรเกรสซีฟ

ผลการวิจัยที่สำคัญ พบว่า โมเดลโค้งพัฒนาการการเปลี่ยนแปลงระยะยาวที่สร้างขึ้นเป็น โมเดลพัฒนาการแบบควอดราติก และกระบวนการวิเคราะห์โดยใช้โปรแกรมเอชแอลเอ็ม สามารถ ประมาณค่าพารามิเตอร์สถานภาพเริ่มต้น อัตราพัฒนาการและอัตราเร่งได้อย่างครบถ้วน โดยผลการ วิเคราะห์โมเดลโค้งพัฒนาการของการเปลี่ยนแปลงระยะยาวของผลสัมฤทธิ์ทางการเรียนคำศัพท์ ภาษาอังกฤษ เป็นโมเดลพัฒนาการแบบควอดราติก สามารถอธิบายความแปรปรวนของการเปลี่ยนแปลง ของผลสัมฤทธิ์คำศัพท์ภาษาอังกฤษ ได้สูงกว่าโมเดลพัฒนาการเชิงเส้น นักเรียนมีอัตราพัฒนาการใน การเรียนรู้คำศัพท์เฉลี่ย .38 คำต่อ 2 สัปดาห์ และมีอัตราเร่งเท่ากับ -0.02 ความเที่ยงในการประมาณค่า

สถานภาพเริ่มต้น อัตราพัฒนาการ และอัตราเร่งเท่ากับ 0.99, 0.58 และ 0.43 ตามลำดับ สถานภาพเริ่มต้นมีความสัมพันธ์ทางบวกกับอัตราพัฒนาการและอัตราเร่ง และตัวแปรเชาวน์ปัญญากับแรงจูงใจส่งอิทธิพลต่อสถานภาพเริ่มต้นอย่างมีนัยสำคัญ ตัวแปรเจตคติและตัวแปรคัมมีเพศหญิงส่งอิทธิพลต่ออัตราเร่งอย่างมีนัยสำคัญ

อิทธิพงษ์ ตั้งสกุลเรืองไธ (2541) ประยุกต์ใช้และตรวจสอบความตรงของโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง 4 รูป คือ โมเดลโค้งพัฒนาการเชิงเส้นที่มีตัวแปรแฝงและกำหนดค่าพารามิเตอร์อิสระ โมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงและกำหนดค่าพารามิเตอร์คงที่ โมเดลพัฒนาการเชิงเส้นตรง และโมเดลพัฒนาการพื้นฐานที่ไม่มีค่าความชัน ในการศึกษาการเปลี่ยนแปลงระยะยาวของพัฒนาการด้านผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์และพัฒนาการทางกายภาพด้านน้ำหนักและส่วนสูงของนักเรียนประถมศึกษา และเปรียบเทียบประสิทธิภาพของโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง 4 รูปแบบโดยใช้ดัชนี 4 ประเภท เป็นเกณฑ์ในการเปรียบเทียบ ได้แก่ ค่าสถิติไค-สแควร์ ดัชนีวัดระดับความกลมกลื่น ดัชนีรากของกำลังสองเฉลี่ยของเศษเหลือ และค่าความคลาดเคลื่อนในรูปคะแนนมาตรฐานสูงสุด กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ นักเรียนระดับประถมศึกษาของโรงเรียนในสังกัดกรุงเทพมหานคร จำนวน 406 คน ในสังกัดสำนักงานการประถมศึกษาจังหวัดพิษณุโลก จำนวน 592 คน เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์และแบบบันทึกการตรวจสุขภาพของนักเรียน ข้อมูลสำหรับการวิจัยเป็นข้อมูลระยะยาว 5 ครั้ง และผู้วิจัยวิเคราะห์ข้อมูลโดยใช้การวิเคราะห์โมเดลลิสเรลแบบมีตัวแปรแฝง

ผลการวิจัยสรุปได้ว่า โค้งพัฒนาการผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ของนักเรียนมีลักษณะเป็นเส้นโค้งพาราโบลาคว่ำ ในขณะที่น้ำหนักและส่วนสูงของนักเรียนมีลักษณะเป็นเส้นโค้งพาราโบลาหงาย ผลจากการประยุกต์ใช้โมเดลโค้งพัฒนาการที่มีตัวแปรแฝงและกำหนดค่าพารามิเตอร์อิสระชนิดความแปรปรวนของความคลาดเคลื่อนไม่เท่าเทียมกันมีประสิทธิภาพในการศึกษาพัฒนาการเกี่ยวกับการวัดการเปลี่ยนแปลงในระยะยาวของพัฒนาการด้านผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์และพัฒนาการทางกายภาพด้านน้ำหนักและส่วนสูงได้ดีที่สุด

มนต์ทิศา ไชยแก้ว (2542) เปรียบเทียบผลการวิเคราะห์การเปลี่ยนแปลงระยะยาวที่ใช้โมเดลประยุกต์โค้งพัฒนาการที่มีตัวแปรแฝงระหว่างโมเดลที่มีข้อมูลครบสมบูรณ์จากการวัดระยะยาว 5 ครั้ง กับโมเดลที่มีการวัดไม่ครบสมบูรณ์และโมเดลที่มีข้อมูลขาดหาย โมเดลการวัดไม่ครบสมบูรณ์ประกอบด้วยโมเดลที่มีการวัด 3 ครั้ง รวม 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 ชุดที่เหลือเป็นข้อมูลอัตราการเข้าเรียนในระดับประถมศึกษา มัธยมศึกษาและอุดมศึกษา จากประเทศต่างๆ จำนวน 125, 118 และ 112 ประเทศ เก็บข้อมูลในปี ค.ศ. 1975, 1980, 1985, 1990 และ 1995 ตามลำดับ การวิเคราะห์ความกลมกลืนของโมเดลประยุกต์โค้งพัฒนาการที่มีตัวแปรแฝงกับข้อมูลประจักษ์ทั้ง 4 ชุดใช้โปรแกรมลิสเรล เวอร์ชัน 8.10 ด้วยรูปแบบคำสั่ง สัญลักษณ์แรม และ โปรแกรมอีคิวเอส สำหรับวินโดว เวอร์ชัน 5.7 ปี

ผลการวิจัยสรุปได้ว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการเปลี่ยนแปลงระยะยาวของโมเดลที่มีการวัดไม่ครบสมบูรณ์กับโมเดลที่มีการวัดครบสมบูรณ์ส่วนใหญ่มีค่าสูงกว่า .70 โดยโมเดลที่วัดได้ 4 ครั้งมีค่าสัมประสิทธิ์สหสัมพันธ์สูงกว่าโมเดลที่วัดได้ 3 ครั้ง และโมเดลที่มีช่วงระยะเวลาที่วัดนานมีค่าสัมประสิทธิ์สหสัมพันธ์สูงกว่าโมเดลที่มีช่วงระยะเวลาที่วัดสั้น สำหรับค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยระหว่างการเปลี่ยนแปลงระยะยาวจากโมเดลที่มีการวัดครบสมบูรณ์กับโมเดลที่วัดได้ 3 ครั้ง มีค่าสูงกว่าโมเดลที่วัดได้ 4 ครั้ง ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการเปลี่ยนแปลงระยะยาวของโมเดลที่มีการวัดครบสมบูรณ์กับโมเดลที่มีอัตราการขาดหายของข้อมูลต่ำมีค่าสูงกว่าโมเดลที่มีอัตราการขาดหายสูง และค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยระหว่างการเปลี่ยนแปลงระยะยาวจากโมเดลที่มีการวัดครบสมบูรณ์กับโมเดลที่มีอัตราการขาดหายต่ำมีค่าสูงกว่าโมเดลที่มีอัตราการขาดหายสูง

สุภารัตน์ เรือจันทิก (2542) เปรียบเทียบผลการวิเคราะห์การเปลี่ยนแปลงระยะยาวของความรู้สิ่งแวดล้อมระหว่างโมเดลเอชแอลเอ็ม และโมเดลลิสเรล ตัวอย่างประชากรคือ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2541 ในจังหวัดศรีสะเกษ จำนวน 509 คน เครื่องมือที่ใช้ได้แก่ แบบวัดความรู้สิ่งแวดล้อม แบบวัดเจตคติ และแบบสอบถามประสพการณ์ในการร่วมกิจกรรมสิ่งแวดล้อม

ผลการวิจัยสามารถสรุปได้ว่า ผลการวิเคราะห์ตัวแปรที่มีอิทธิพลต่อสถานภาพเริ่มต้นในโมเดลทั้งสอง พบว่า ตัวแปรประสพการณ์ในการเข้าร่วมกิจกรรมเกี่ยวกับสิ่งแวดล้อมมีอิทธิพลต่อสถานภาพเริ่มต้นอย่างมีนัยสำคัญ ตัวแปรเพศไม่มีอิทธิพลต่อสถานภาพเริ่มต้น ส่วนตัวแปรเจตคติต่อสิ่งแวดล้อมที่วิเคราะห์ด้วยโมเดลเอชแอลเอ็มมีอิทธิพลต่อสถานภาพเริ่มต้นอย่างมีนัยสำคัญ แต่ตัวแปรเจตคติต่อสิ่งแวดล้อมที่วิเคราะห์ด้วยโมเดลลิสเรลไม่มีอิทธิพลต่อสถานภาพเริ่มต้น สำหรับผลการวิเคราะห์ตัวแปรที่มีอิทธิพลต่ออัตราการเปลี่ยนแปลงทั้งสองโมเดล พบว่า ตัวแปรเจตคติต่อสิ่งแวดล้อมมีอิทธิพลต่ออัตราการเปลี่ยนแปลงของความรู้ด้านสิ่งแวดล้อมอย่างมีนัยสำคัญ ส่วนตัวแปรเพศและตัวแปรประสพการณ์ในการเข้าร่วมกิจกรรมเกี่ยวกับสิ่งแวดล้อมไม่มีอิทธิพลต่ออัตราการเปลี่ยนแปลง

ในการวิเคราะห์ความสัมพันธ์ระหว่างสถานภาพเริ่มต้นและอัตราการเปลี่ยนแปลงทั้งสองโมเดล พบว่า สถานภาพเริ่มต้นมีความสัมพันธ์ทางบวกกับอัตราการเปลี่ยนแปลงอย่างมีนัยสำคัญ แต่พบว่าการวิเคราะห์การเปลี่ยนแปลงระยะยาวด้วยโมเดลลิสเรลสามารถอธิบายการเปลี่ยนแปลงระยะยาวได้ดีกว่าโมเดลเอชแอลเอ็ม เนื่องจากโมเดลมีความคลาดเคลื่อนที่ต่ำกว่า

สมถวิล วิจิตรวรรณ (2543) เปรียบเทียบประสิทธิภาพของโมเดลการวัดการเปลี่ยนแปลงระยะยาวชนิดตัวแปรเดียวและชนิดตัวแปรพหุ ระหว่างโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง โมเดลพหุระดับ และโมเดลกึ่งซิมเพิลิกซ์ที่มีตัวแปรแฝงพัฒนาการ เป็นการวัดคุณลักษณะทางคณิตศาสตร์ 2 คุณลักษณะ คือ การคิดคำนวณและการแก้โจทย์ปัญหาทางคณิตศาสตร์ กับนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 469 คน เก็บข้อมูล 5 ครั้ง ห่างกันครั้งละ 3 สัปดาห์ โดยจัดทำแบบทดสอบที่เป็นคู่ขนาน 5 ฉบับ และวิเคราะห์ค่าสถิติเบื้องต้นของข้อมูล โดยใช้โปรแกรม SPSS 9.0 และโปรแกรม EQS 5.7B ในการวิเคราะห์โมเดลพหุระดับ ชนิดตัวแปรเดียวการคิดคำนวณ และโมเดลพหุระดับชนิดตัวแปรเดียวการแก้โจทย์ปัญหา ใช้โปรแกรม HLM 4.01 เพื่ออธิบายพัฒนาการของแต่ละคุณลักษณะ ส่วนการวิเคราะห์โมเดลพหุระดับ ชนิดตัวแปรพหุ ใช้โปรแกรม EQS 5.7B เพื่อหาความสัมพันธ์ระหว่างพัฒนาการของคุณลักษณะการคิดคำนวณกับการแก้โจทย์ปัญหา โดยคงค่าประมาณพารามิเตอร์ชนิดตัวแปรเดียวไว้ สำหรับโมเดลโค้งพัฒนาการที่มีตัวแปรแฝงชนิดตัวแปรเดียวและตัวแปรพหุ และโมเดลกึ่งซิมเพิลิกซ์ที่มีตัวแปรแฝงพัฒนาการชนิดตัวแปรเดียวและตัวแปรพหุ วิเคราะห์ด้วยโปรแกรม EQS 5.7B

ผลการวิจัยพบว่า โมเดลโค้งพัฒนาการที่มีตัวแปรแฝงเป็นโมเดลมีประสิทธิภาพสอดคล้องกับข้อมูลสูงสุด รองลงมาคือ โมเดลกึ่งซิมเพิลิกซ์ที่มีตัวแปรแฝงพัฒนาการและโมเดลพหุระดับ ทั้งในการวัดการเปลี่ยนแปลงระยะยาวชนิดตัวแปรเดียว คุณลักษณะการคิดคำนวณที่มีแบบแผนพัฒนาการเป็นเส้นตรง คุณลักษณะการแก้โจทย์ปัญหาที่มีแบบแผนพัฒนาการไม่ใช่เส้นตรง และการวัดการเปลี่ยนแปลงระยะยาวชนิดตัวแปรพหุของคุณลักษณะการคิดคำนวณและการแก้โจทย์ปัญหา เนื่องจากโมเดลพหุระดับเป็นโมเดลที่ประมาณค่าความแปรปรวนของความคลาดเคลื่อนในการวัด 5 ครั้งเท่ากัน จึงสร้างโมเดลโค้งพัฒนาการที่มีตัวแปรแฝงที่กำหนดประมาณค่าความแปรปรวนของความคลาดเคลื่อนในการวัด 5 ครั้งเท่ากัน และโมเดลกึ่งซิมเพิลิกซ์ที่มีตัวแปรแฝงพัฒนาการที่กำหนดประมาณค่าความแปรปรวนของความคลาดเคลื่อนในการวัด 5 ครั้งเท่ากัน ปรากฏว่าทุกโมเดลที่กำหนดประมาณค่าความแปรปรวนของความคลาดเคลื่อนในการวัด 5 ครั้งเท่ากัน เป็นโมเดลที่มีความสอดคล้องกับข้อมูลต่ำสุด

อวยพร เรืองตระกูล (2544) พัฒนาและวิเคราะห์เปรียบเทียบคุณภาพคะแนนพัฒนาการระหว่างวิธีการวัดตามทฤษฎีการทดสอบแบบมาตรฐานเดิม 5 วิธี ตามทฤษฎีการตอบข้อสอบ 2 วิธี และวิธีที่ผู้วิจัยพัฒนา 2 วิธี ได้พัฒนาแบบทดสอบแบบเลือกตอบ 4 ตัวเลือก เพื่อวัดความรู้บางส่วน จำนวน 66 ข้อ ในวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 3 ฉบับที่คู่ขนานกัน แต่ละฉบับมีค่าความเชื่อมั่นเท่ากับ 0.913, 0.918 และ 0.928 ตามลำดับ เพื่อใช้ในการวัดซ้ำ 3 ครั้ง ห่างกันครั้งละ 4 สัปดาห์ ข้อมูลชุดแรกเป็นการวัดซ้ำด้วยแบบทดสอบฉบับเดิม ข้อมูลชุดที่สองเป็นการวัดซ้ำแบบจัดสุ่มละตินด้วยแบบทดสอบคู่ขนาน ส่วนข้อมูลชุดที่สามเป็นข้อมูลทุติยภูมิซึ่งเป็นฐานข้อมูลจากงานวิจัยของสมถวิล วิจิตรวรรณ กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 2 กลุ่ม ๆ ละ 698 และ 637 คน ข้อมูลทั้งสามชุดมีการตรวจให้คะแนนแบบทวิภาค

(Dichotomous) และตรวจให้คะแนนความรู้อย่างบางส่วน (Partial credit) ยกเว้นข้อมูลชุดที่สามที่มีการตรวจให้คะแนนแบบทวิภาคเท่านั้น

การวิเคราะห์ข้อมูลเป็นการประมาณค่าคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบมาตรฐานเดิมโดยใช้โปรแกรม SPSS-WIN 10.07 ตามทฤษฎีการตอบข้อสอบ โดยใช้โปรแกรม BILOG 3 เมื่อข้อมูลมีการตรวจให้คะแนนแบบทวิภาค และโปรแกรม PARSCALE 2 เมื่อข้อมูลมีการตรวจให้คะแนนความรู้อย่างบางส่วน ส่วนคะแนนพัฒนาการที่ประมาณค่าจากโมเดลราชส์พหุมิติสำหรับการเรียนรู้และการเปลี่ยนแปลงใช้โปรแกรม LPCM-WIN 1.0 และคะแนนพัฒนาการที่ประมาณค่าจากโมเดลพัฒนาการเชิงเส้นโค้งที่มีตัวแปรแฝงใช้โปรแกรม LISREL 8.3 สำหรับการวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการด้านความเที่ยงตรงตามเกณฑ์สัมพัทธ์ด้วยค่าสถิติความสัมพัทธ์แบบเพียร์สัน และสเปียร์แมน และด้านความคลาดเคลื่อนของคะแนนพัฒนาการใช้โปรแกรม SPSS-WIN 10.07 ส่วนการเปรียบเทียบคุณภาพของวิธีการวัดคะแนนพัฒนาการด้านความเที่ยงตรงตามเกณฑ์สัมพัทธ์ระหว่างกลุ่มวิธีใช้การวิเคราะห์ด้วยโมเดลเชิงเส้นระดับลดหลั่นโดยใช้โปรแกรม HLM-WIN 5.04 ส่วนภายในกลุ่มวิธีใช้สถิติทดสอบ Hotelling โดยใช้โปรแกรม EXCEL

ผลการวิจัยพบว่า เมื่อเปรียบเทียบผลการประมาณค่าคะแนนพัฒนาการระหว่าง 3 กลุ่มวิธีพบว่า กลุ่มวิธีของทฤษฎีการทดสอบแบบมาตรฐานเดิม กลุ่มวิธีของทฤษฎีการตอบข้อสอบและกลุ่มวิธีของทฤษฎีการตอบข้อสอบที่ผู้วิจัยพัฒนา มีคุณภาพไม่แตกต่างกัน ยกเว้นข้อมูลชุดที่ 3 ที่ไม่มีการตรวจให้คะแนนความรู้อย่างบางส่วน ส่วนผลการเปรียบเทียบภายในกลุ่มวิธีการวัดด้วยสถิติทดสอบ t ของ Hotelling (1940) พบว่า คะแนนพัฒนาการจากวิธีการวัดส่วนใหญ่มีค่าความเที่ยงตรงตามเกณฑ์สัมพัทธ์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยพบว่าภายในกลุ่มวิธีของทฤษฎีการทดสอบแบบมาตรฐานเดิม วิธีการวัด DS และ SR มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูล 2 ชุดแรก และวิธีการวัด SSR มีคุณภาพสูงกว่าวิธีอื่นสำหรับข้อมูลชุดที่ 3 ผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบข้อสอบ พบว่าวิธีการวัด MRMLC มีค่าคุณภาพสูงกว่าวิธีการวัด L_θ และผลการเปรียบเทียบภายในกลุ่มวิธีของทฤษฎีการตอบข้อสอบที่ผู้วิจัยพัฒนา พบว่าวิธีการวัด V_θ มีคุณภาพสูงกว่า S_θ เมื่อมีการตรวจให้คะแนนแบบทวิภาค

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการวัดการเปลี่ยนแปลงนั้น สามารถสรุปได้ว่าการวิจัยที่ผ่านมา มี 2 ลักษณะ คือ 1) การศึกษาเพื่อพัฒนาวิธีการวัดปริมาณการเปลี่ยนแปลง เพื่อนำไปใช้ในการศึกษาความก้าวหน้าและพัฒนาการรายบุคคลและกลุ่ม เช่น เอ็มเบรตสัน (Embretson), ฟิชเชอร์ (Fischer), กลักซ์ และ สเปล (Glück; & Spiel), เมย์ และ ไนส์วันเดอร์ (May; & Nicewander) และ ไฮเดนเบอร์ก (Heidenberg) เป็นต้น และ 2) การศึกษาปัจจัยหรือองค์ประกอบใดที่มีอิทธิพลหรือมีความสัมพันธ์กับการเปลี่ยนแปลง เช่น ชิม (Shim), คิริชัย กาญจนวาสี และประสิทธิ์ ไชยกาล เป็นต้น

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาเชิงประจักษ์ที่มีจุดมุ่งหมายเพื่อศึกษาผลของการวัดการเปลี่ยนแปลงรายบุคคลที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษาจำนวนห้าวิธี คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ โดยพิจารณาจากค่าความสัมพันธ์ของคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง ความสัมพันธ์ของคะแนนการเปลี่ยนแปลง ความคลาดเคลื่อนในการประมาณค่าคะแนนการเปลี่ยนแปลง และความลำเอียงของคะแนนการเปลี่ยนแปลง ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน ในขอบข่ายเนื้อหาวิชาคณิตศาสตร์ เรื่อง สมการเชิงเส้นตัวแปรเดียว กับนักศึกษาการศึกษาการศึกษานอกโรงเรียน ระดับมัธยมศึกษาตอนต้น ซึ่งผู้วิจัยได้ดำเนินการตามขั้นตอน ดังนี้

1. การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง
2. การพัฒนาเครื่องมือที่ใช้ในการศึกษา
3. การเก็บรวบรวมข้อมูล
4. การจัดกระทำและการวิเคราะห์ข้อมูล

การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง

ประชากร

ประชากรในการศึกษาครั้งนี้ คือ นักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาขั้นพื้นฐาน ระดับมัธยมศึกษาตอนต้น ของศูนย์บริการการศึกษานอกโรงเรียนเขต จำนวน 12 แห่ง ในสังกัดศูนย์การศึกษา นอกโรงเรียนกรุงเทพมหานคร 1 ภาคเรียนที่ 2 ปีการศึกษา 2549 มีจำนวน 7,555 คน

กลุ่มตัวอย่าง

การกำหนดขนาดกลุ่มตัวอย่าง ได้คำนวณหาขนาดกลุ่มตัวอย่างจากสูตร (ศิริชัย กาญจนวาสี; ทวีวัฒน์ ปิตยานนท์; และดิเรก ศรีสุข. 2544: 24) ที่ระดับความเชื่อมั่น 95% เมื่อยอมให้ความคลาดเคลื่อนของการประมาณค่าเกิดขึ้นได้ในระดับ $\pm 10\%$ ดังนี้

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

เมื่อ $n_{\hat{\mu}}$ คือ ขนาดของกลุ่มตัวอย่างที่ต้องการ

N คือ จำนวนประชากร

Z คือ ค่าความเชื่อมั่นที่กำหนด ซึ่งผู้วิจัยกำหนดค่าที่ 95% ค่า Z มีค่าเท่ากับ 1.96

σ^2 คือ ค่าความแปรปรวนของคะแนนสอบวิชาคณิตศาสตร์ของประชากร

E คือ ค่าความคลาดเคลื่อนที่ผู้วิจัยยอมรับได้สำหรับการสรุปผล ซึ่งผู้วิจัยกำหนดไว้ที่

0.10 ของ σ

จากการคำนวณขนาดตัวอย่างที่เหมาะสมจากประชากร 7,555 คน ได้กลุ่มตัวอย่าง 367 คน แต่เพื่อให้ครอบคลุมตัวแปรอิสระ คือ ขนาดกลุ่มตัวอย่าง ผู้วิจัยจึงเพิ่มขนาดกลุ่มตัวอย่างในการวิจัยครั้งนี้ เป็น 1,000 คน โดยใช้วิธีการสุ่มแบบหลายขั้นตอน (Multistage Random Sampling) ซึ่งมีขั้นตอนการสุ่มดังนี้

ขั้นที่ 1 สุ่มเขตที่อยู่ในความรับผิดชอบของศูนย์การศึกษาออกโรงเรียนกรุงเทพมหานคร 1 ซึ่งมีทั้งหมด 12 เขต (สบข.) ผู้วิจัยสุ่มอย่างง่ายมา 60% ได้จำนวน 7 เขต

ขั้นที่ 2 สุ่มศูนย์การเรียนชุมชนจากแต่ละเขตที่ได้รับการสุ่มจากขั้นที่ 1 มาแล้ว ด้วยการสุ่มอย่างง่ายเขตละ 3 ศูนย์การเรียนชุมชน ได้ศูนย์การเรียนชุมชนทั้งสิ้น 21 ศูนย์การเรียนชุมชน

ขั้นที่ 3 สุ่มกลุ่มเรียนจากแต่ละศูนย์การเรียนชุมชนที่ได้รับการสุ่มจากขั้นที่ 2 มาแล้ว ด้วยการสุ่มอย่างง่าย ศูนย์การเรียนชุมชนละ 1 กลุ่ม ได้จำนวนนักศึกษาที่เป็นกลุ่มตัวอย่าง จำนวน 1,499 คน ซึ่งได้สอบครั้งแรก ต่อมาในระหว่างการสอนครั้งที่สองและสาม นักศึกษาขาดหายไปจากการทดสอบครั้งแรก จึงทำให้นักศึกษาที่เป็นกลุ่มตัวอย่างมีการเปลี่ยนแปลงในแต่ละครั้งของการสอบ ผู้วิจัยจึงคัดเลือกเฉพาะนักศึกษาที่เข้าสอบทั้งสามครั้ง ได้จำนวนนักศึกษาที่เป็นกลุ่มตัวอย่าง จำนวน 1,085 คน (คิดเป็นร้อยละ 72.38) ปรากฏผลดังตาราง 1

ตาราง 1 จำนวนนักศึกษาที่เป็นกลุ่มตัวอย่าง

ศูนย์บริการการศึกษา- นอกโรงเรียนเขต (สบข.)	จำนวนนักศึกษาที่สอบ			จำนวนนักศึกษาที่ สอบทั้งสามครั้ง
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3	
1. สบข. คลองสาน	204	169	150	149
2. สบข. จอมทอง	199	184	158	157
3. สบข. บางกอกใหญ่	173	140	127	120

ตาราง 1 (ต่อ)

ศูนย์บริการการศึกษา- นอกโรงเรียนเขต (ศบข.)	จำนวนนักศึกษาที่สอบ			จำนวนนักศึกษาที่ สอบทั้งสามครั้ง
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3	
4. ศบข. ภาษีเจริญ	244	224	187	185
5. ศบข. ราษฎร์บูรณะ	282	207	164	143
6. ศบข. หนองแขม	260	253	247	234
7. ศบข. บางแค	137	135	97	97
รวม	1,499	1,312	1,130	1,085

การพัฒนาเครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการศึกษาครั้งนี้ คือ แบบทดสอบวิชาคณิตศาสตร์ เนื้อหาเรื่อง สมการเชิงเส้นตัวแปรเดียว ระดับมัธยมศึกษาตอนต้น หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 ที่ผู้วิจัยสร้างขึ้นเป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก โดยมีขั้นตอนการสร้างและพัฒนาแบบทดสอบ ดังนี้

1. ศึกษาหลักสูตร และเอกสารงานวิจัยที่เกี่ยวข้องเพื่อเป็นแนวทางในการสร้างแบบทดสอบ วิเคราะห์มาตรฐานการเรียนรู้ กรอบสาระการเรียนรู้ รายละเอียดสาระการเรียนรู้และผลการเรียนรู้ที่คาดหวัง เพื่อกำหนดเนื้อหา มวลความรู้และมวลความรู้ย่อย ซึ่งได้ดำเนินการดังนี้

1.1 วิเคราะห์เนื้อหาวิชาคณิตศาสตร์ เรื่องสมการเชิงเส้นตัวแปรเดียว ระดับมัธยมศึกษาตอนต้น กำหนดเป็นขอบเขตเนื้อหาใหญ่ และวิเคราะห์เนื้อหาย่อยในแต่ละเนื้อหาใหญ่

1.2 วิเคราะห์ผลการเรียนรู้ที่คาดหวังเฉพาะวิชาของเนื้อหาวิชาคณิตศาสตร์ เรื่องสมการเชิงเส้นตัวแปรเดียว เพื่อหาพฤติกรรมหลัก ซึ่งเป็นพฤติกรรมใหญ่ (Conceptual objective) ที่ต้องการวัดโดยใช้แนวการแบ่งพฤติกรรมด้านพุทธิพิสัย (Cognitive domain) ตามแนวคิดของบลูม

1.3 วิเคราะห์ความสัมพันธ์ของผลที่ได้จากข้อ 1.1 และ 1.2 เพื่อกำหนดขอบเขตของมวลความรู้และมวลความรู้ย่อย กำหนดตารางวิเคราะห์ข้อสอบ

2. ตรวจสอบความเหมาะสมของการกำหนดมวลความรู้ย่อยในแต่ละมวลความรู้ ภายในขอบเขตเนื้อหาที่กำหนด โดยให้ผู้เชี่ยวชาญการสอนวิชาคณิตศาสตร์ไม่น้อยกว่า 5 ปี จำนวน 10 ท่าน (ดังรายชื่อในภาคผนวก ข) เป็นผู้พิจารณาตัดสินความครอบคลุมของมวลความรู้ย่อยในมวลความรู้ที่กำหนดขึ้น ตามวิธีการตัดสินความสอดคล้องของ โรวินลลี และแฮมเบิลตัน (Hambleton.; et al. 1978: 34; citing Rovinelli; &

Hambleton. 1977. Journal for Educational Research.) ถ้าค่าเฉลี่ยของคะแนนการตัดสินของผู้เชี่ยวชาญในแต่ละมวลความรู้ย่อยมีค่าสูงกว่า 0.50 จะแสดงว่ามวลความรู้ย่อยที่กำหนดขึ้นมีความครอบคลุมมวลความรู้ในขอบเขตเนื้อหานั้นๆ

ผู้วิจัยนำมวลความรู้ย่อยไปให้ผู้เชี่ยวชาญพิจารณา แล้วนำคะแนนความคิดเห็นมาคำนวณค่า IOC จะได้ค่าคะแนนเฉลี่ยจากผลการตัดสินของผู้เชี่ยวชาญอยู่ระหว่าง 0.80 ถึง 1.00 ซึ่งแสดงว่ามวลความรู้ย่อยที่กำหนดขึ้นทุกข้อมีความครอบคลุมมวลความรู้ในขอบเขตเนื้อหาสมการเชิงเส้นตัวแปรเดียว

3. พิจารณาน้ำหนักความสำคัญของมวลความรู้ย่อย เพื่อใช้เป็นเกณฑ์ในการกำหนดจำนวนข้อสอบที่เหมาะสมของแต่ละมวลความรู้ย่อย โดยคิดจากมวลความรู้หลักเท่ากับร้อยละ 100 โดยใช้ผู้เชี่ยวชาญการสอนวิชาคณิตศาสตร์ไม่น้อยกว่า 5 ปี จำนวน 10 ท่าน (ดังรายชื่อในภาคผนวก ข) ปรากฏผลดังตาราง 2

ตาราง 2 ผลการวิเคราะห์เนื้อหา เรื่องสมการเชิงเส้นตัวแปรเดียว ระดับมัธยมศึกษาตอนต้น

มวลความรู้	มวลความรู้ย่อย	น้ำหนัก ความสำคัญ
1. ความรู้พื้นฐานเกี่ยวกับ สมการ	1.1 นักศึกษาสามารถระบุตัวแปรได้	5
	1.2 นักศึกษาสามารถระบุประโยคภาษาหรือประโยค สัญลักษณ์หรือสมการได้	5
2. ความสามารถในการคิด คำนวณเพื่อหาคำตอบของ สมการ โดยใช้การแทนค่า ตัวแปรในสมการ	2.1 นักศึกษาสามารถระบุจำนวนที่กำหนดให้ได้ว่า เป็นคำตอบของสมการ	10
	2.2 นักศึกษาสามารถระบุจำนวนที่กำหนดให้ได้ว่า เป็นคำตอบของสมการ	5
3. ความสามารถในการคิด คำนวณหาคำตอบของสมการ โดยใช้สมบัติของการเท่ากัน	3.1 นักศึกษาสามารถหาคำตอบของสมการโดยใช้ สมบัติการเท่ากันของการบวกได้	10
	3.2 นักศึกษาสามารถหาคำตอบของสมการโดยใช้ สมบัติการเท่ากันของการคูณได้	10
	3.3 นักศึกษาสามารถหาคำตอบของสมการโดยใช้ สมบัติการเท่ากันทั้งสองสมบัติได้	10

ตาราง 2 (ต่อ)

มวลความรู้	มวลความรู้ย่อย	น้ำหนัก ความสำคัญ
4. ความสามารถในการแก้ สมการเชิงเส้นตัวแปรเดียว จากประโยคสัญลักษณ์	4.1 นักศึกษาสามารถแก้สมการเชิงเส้นตัวแปรเดียวได้ 4.2 นักศึกษาสามารถตรวจสอบคำตอบของสมการ เชิงเส้นตัวแปรเดียวได้	12.5 12.5
5. ความสามารถในการเปลี่ยน ประโยคระหว่างประโยคภาษา กับประโยคสัญลักษณ์	5.1 นักศึกษาสามารถระบุประโยคสัญลักษณ์ที่เขียน มาจากประโยคภาษาที่กำหนดให้ได้	7.5
6. ความสามารถในการแก้ โจทย์ปัญหาสมการเชิงเส้น ตัวแปรเดียว	6.1 นักศึกษาสามารถแก้โจทย์ปัญหาสมการเชิงเส้น ตัวแปรเดียวได้	12.5
รวม		100

4. สร้างลักษณะเฉพาะข้อสอบ (Item specification) ตามแนวคิดของสงบ ลักษณะ (บุญเชิด ภิญญอนันตพงษ์. 2526: 81-83; อ้างอิงจาก สงบ ลักษณะ. 2525. หน้า 24-25) จากมวลความรู้ย่อยที่กำหนดไว้ เพื่อกำหนดโครงสร้างของข้อสอบรายข้อ คือ กำหนดกรอบลักษณะคำถาม ที่มาของตัวลงแต่ละตัว ทำการตรวจสอบความเหมาะสมของลักษณะเฉพาะข้อสอบโดยผู้เชี่ยวชาญด้านการสอนวิชาคณิตศาสตร์และ วัสดุ จำนวน 10 ท่าน (ดังรายชื่อในภาคผนวก ข) แล้วนำข้อเสนอแนะมาปรับปรุงและพัฒนา

5. สร้างข้อสอบตามโครงสร้างของลักษณะเฉพาะข้อสอบ จำนวน 2 เท่าของจำนวนที่ใช้ในการ ศึกษา ซึ่งในการวิจัยครั้งนี้ต้องการใช้ข้อสอบจำนวน 40 ข้อ ดังนั้นผู้วิจัยจึงสร้างข้อสอบจำนวน 80 ข้อ

6. จัดข้อสอบเป็น 2 ฉบับๆ ละ 40 ข้อ ทั้งนี้เพื่อให้แบบทดสอบมีจำนวนข้อสอบที่เหมาะสมต่อ เวลาและความตั้งใจของผู้สอบ ซึ่งอาจจะมีผลต่อการใช้ความสามารถอย่างเต็มที่ในการตอบข้อสอบของ ผู้สอบ โดยใช้วิธีการสุ่มข้อสอบที่สร้างจากลักษณะเฉพาะข้อสอบตามมวลความรู้ย่อยเดียวกันออกเป็น 2 กลุ่ม ให้ได้จำนวนข้อตามที่กำหนดน้ำหนักความสำคัญไว้ในข้อที่ 3 แล้วนำข้อสอบไปให้ผู้เชี่ยวชาญการสอน คณิตศาสตร์และวัสดุ จำนวน 10 ท่าน (ดังรายชื่อในภาคผนวก ข) ตรวจสอบความเที่ยงตรงเชิงเนื้อหา ความ เป็นปรนัยของข้อคำถาม ตามวิธีการตัดสินความสอดคล้องของโรวินสกีและแฮมเบิลตัน (Hambleton.; et al.

1978: 34; citing Rovinelli; & Hambleton. 1977. Journal for Educational Research.) ถ้าค่าเฉลี่ยของคะแนนการตัดสินของผู้เชี่ยวชาญในแต่ละข้อมีค่าสูงกว่า 0.50 จะแสดงว่าข้อสอบข้อนั้นวัดตามคุณลักษณะเฉพาะข้อสอบที่กำหนดขึ้นในแต่ละมวลความรู้ย่อย และนำมาปรับปรุง พัฒนา

ผู้วิจัยนำข้อสอบไปให้ผู้เชี่ยวชาญพิจารณา แล้วนำคะแนนความคิดเห็นมาคำนวณค่า IOC ได้ค่าคะแนนเฉลี่ยจากผลการตัดสินของผู้เชี่ยวชาญอยู่ระหว่าง 0.20 ถึง 1.00 และ 0.80 ถึง 1.00 ในแบบทดสอบฉบับที่ 1 และ ฉบับที่ 2 ตามลำดับ (ดังผลที่ปรากฏในภาคผนวก ข) ข้อสอบที่สร้างขึ้นและมีค่าคะแนนเฉลี่ยจากผลการตัดสินของผู้เชี่ยวชาญน้อยกว่า 0.50 มีจำนวน 1 ข้อ ในแบบทดสอบฉบับที่ 1 ได้ตัดข้อสอบข้อดังกล่าวทิ้งและสร้างข้อสอบขึ้นมาใหม่ พร้อมทั้งได้ปรับปรุงข้อสอบข้อที่มีค่าคะแนนเฉลี่ยจากผลการตัดสินของผู้เชี่ยวชาญเท่ากับ 0.80 ของทั้งสองฉบับ ซึ่งมีจำนวนฉบับละ 12 ข้อ ทั้งข้อคำถามและตัวเลือก แล้วนำไปให้ผู้เชี่ยวชาญพิจารณาอีกครั้งก่อนนำไปทดลองใช้

7. ทดลองใช้แบบทดสอบครั้งที่ 1 โดยการนำแบบทดสอบที่ปรับปรุงแก้ไขแล้วไปทดลองใช้กับ นักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 ระดับมัธยมศึกษาตอนต้น ที่ลงทะเบียนเรียนวิชาคณิตศาสตร์ในภาคเรียนที่ 1 ปีการศึกษา 2549 ฉบับที่ 1 จำนวน 209 คน และฉบับที่ 2 จำนวน 197 คน โดยการแจกแบบทดสอบสลับกันในแต่ละห้องสอบ เพื่อให้แบบทดสอบแต่ละฉบับได้รับการทดสอบจากผู้สอบที่มีกิจกรรมการเรียนรู้เหมือนกัน ปรากฏผลดังตาราง 3

ตาราง 3 จำนวนนักศึกษาที่ตอบแบบทดสอบแต่ละฉบับในการทดลองใช้เครื่องมือครั้งที่ 1

สถานศึกษา	จำนวนนักศึกษา (คน)		รวมนักศึกษา (คน)
	ฉบับที่ 1	ฉบับที่ 2	
1. ศูนย์บริการการศึกษานอกโรงเรียนเขตพญาไท	52	52	104
2. ศูนย์บริการการศึกษานอกโรงเรียนเขตดินแดง	70	69	139
3. ศูนย์บริการการศึกษานอกโรงเรียนอำเภอบางบัวทอง	87	76	163
รวม	209	197	406

8. วิเคราะห์หาคุณลักษณะทางสถิติรายข้อตามทฤษฎีการทดสอบแบบมาตรฐานเดิม เพื่อหาค่าความยากและค่าอำนาจจำแนกของข้อสอบ ด้วยโปรแกรม B-Index Win 6.5 (สาคร แสงผึ้ง. 2546) ได้ค่าความยากของข้อสอบอยู่ระหว่าง 0.23 ถึง 0.78 และระหว่าง 0.28 ถึง 0.78 ในแบบทดสอบฉบับที่ 1

และ 2 ตามลำดับ ค่าอำนาจจำแนกอยู่ระหว่าง 0.17 ถึง 0.58 และระหว่าง 0.13 ถึง 0.57 ในแบบทดสอบฉบับที่ 1 และ 2 ตามลำดับ มีข้อสอบที่มีคุณลักษณะทางสถิติไม่อยู่ในเกณฑ์ คือ ค่าอำนาจจำแนกต่ำกว่า 0.20 จำนวน 10 ข้อ และ 13 ข้อ ในแบบทดสอบฉบับที่ 1 และ 2 ตามลำดับ จึงคัดเลือกข้อสอบเพื่อให้ได้ข้อสอบจำนวน 50 ข้อ ตามเกณฑ์คุณลักษณะทางสถิติของข้อสอบ (บุญเชิด ภิญญอนันตพงษ์. 2526: 55) ด้านความเที่ยงตรงเชิงเนื้อหาตั้งแต่ 0.50 เป็นต้นไป และค่าอำนาจจำแนกตั้งแต่ 0.20 เป็นต้นไป รวมทั้งพิจารณาเวลาที่พอเหมาะสำหรับการทำแบบทดสอบแต่ละฉบับ ซึ่งเป็นเวลาที่นักศึกษาส่วนใหญ่ประมาณร้อยละ 90 ทำข้อสอบเสร็จ คือ 1 ชั่วโมง 30 นาที และนำไปทดลองใช้ครั้งที่ 2

9. ทดลองใช้แบบทดสอบครั้งที่ 2 โดยการนำแบบทดสอบที่ปรับปรุงแก้ไขแล้วจากการทดลองใช้ครั้งที่ 1 จำนวน 50 ข้อ ไปสอบกับนักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 ระดับมัธยมศึกษาตอนต้น ที่ลงทะเบียนเรียนวิชาคณิตศาสตร์ ในภาคเรียนที่ 1 ปีการศึกษา 2549 จำนวน 1,034 คน ปรากฏผลดังตาราง 4

ตาราง 4 จำนวนนักศึกษาที่ตอบแบบทดสอบในการทดลองใช้เครื่องมือครั้งที่ 2

สถานศึกษา	จำนวนนักศึกษา (คน)
1. ศูนย์บริการการศึกษานอกโรงเรียนเขตบางพลัด	117
2. ศูนย์บริการการศึกษานอกโรงเรียนเขตภาษีเจริญ	109
3. ศูนย์บริการการศึกษานอกโรงเรียนเขตลาดพร้าว	60
4. ศูนย์บริการการศึกษานอกโรงเรียนเขตสวนหลวง	94
5. ศูนย์บริการการศึกษานอกโรงเรียนเขตดินแดง	42
6. ศูนย์บริการการศึกษานอกโรงเรียนอำเภอเมืองนนทบุรี	122
7. ศูนย์บริการการศึกษานอกโรงเรียนอำเภอบางใหญ่	148
8. ศูนย์บริการการศึกษานอกโรงเรียนอำเภอปากเกร็ด	132
9. ศูนย์บริการการศึกษานอกโรงเรียนอำเภอบางบัวทอง	127
10. ศูนย์บริการการศึกษานอกโรงเรียนอำเภอบ้านแพ้ว	83
รวม	1,034

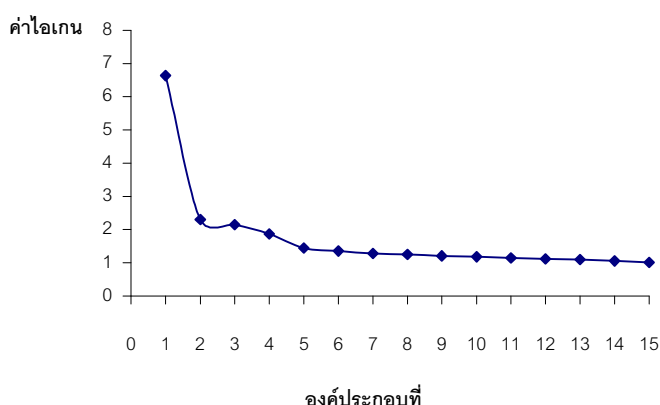
10. ตรวจสอบความเป็นมิติเดียว (Unidimensional) ของแบบทดสอบ วิเคราะห์คุณลักษณะทางสถิติรายข้อและแบบทดสอบตามทฤษฎีการตอบข้อสอบ

10.1 ตรวจสอบความเป็นมิติเดียว (Unidimensional) ของแบบทดสอบ เพื่อให้เป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ ผู้วิจัยได้ตรวจสอบจำนวนองค์ประกอบในแบบทดสอบด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory factor analysis; EFA) แบบ Principal component analysis (PCA) โดยใช้เกณฑ์ค่าไอเกนตั้งแต่ 1 ขึ้นไป นับเป็น 1 องค์ประกอบ วิเคราะห์ค่า KMO หรือ MSA (Kaiser-Meyer-Olkin measure of sampling adequacy) และสถิติทดสอบ Bartlett (Bartlett's test of sphericity) รวมทั้งผลที่ได้จาก Scree plot ด้วยโปรแกรม SPSS และการวิเคราะห์เพื่อยืนยันองค์ประกอบ (Confirmatory factor analysis; CFA) ด้วยโปรแกรม LISREL ปรากฏผลการวิเคราะห์ดังตาราง 5 และภาพประกอบ 3

ตาราง 5 ผลการวิเคราะห์องค์ประกอบเชิงสำรวจของแบบทดสอบคณิตศาสตร์

ค่าไอเกน	ร้อยละของ ความแปรปรวน	ความแปรปรวนสะสม	ค่า KMO	ค่า Bartlett
6.63	13.26	13.26	0.85	$\chi^2=8588.82$ $df=1225$ $p=.00$
2.30	4.61	17.87		
2.15	4.30	22.17		
1.87	3.74	25.91		
1.44	2.89	28.79		
1.35	2.70	31.49		
1.27	2.54	34.04		
1.24	2.49	36.53		
1.20	2.41	38.93		
1.18	2.35	41.28		
1.14	2.28	43.57		
1.12	2.24	45.80		
1.10	2.19	47.99		
1.06	2.11	50.10		
1.01	2.02	52.12		

ผลการวิเคราะห์ตามตาราง 5 พบว่า ค่า KMO หรือ MSA มีค่าเท่ากับ 0.85 ซึ่งมีค่ามากกว่า 0.80 ตามที่คิมและมุลเลอร์เสนอไว้ (นงลักษณ์ วิรัชชัย. 2542: 149; อ้างอิงจาก Kim; & Mueller. 1978: 54) ซึ่งแสดงว่าตัวแปรต่างๆ มีความสัมพันธ์กัน ดังนั้นข้อมูลในแบบทดสอบมีความสัมพันธ์กันสูง จึงมีความเหมาะสมที่จะวิเคราะห์องค์ประกอบ โดยแบบทดสอบฉบับนี้มีจำนวนองค์ประกอบที่มีค่าไอเกนตั้งแต่ 1 ขึ้นไป จำนวน 15 องค์ประกอบ มีค่าเท่ากับ 6.63, 2.30, 2.15, 1.87, 1.44, 1.35, 1.27, 1.24, 1.20, 1.18, 1.14, 1.12, 1.10, 1.06 และ 1.01 ตามลำดับ



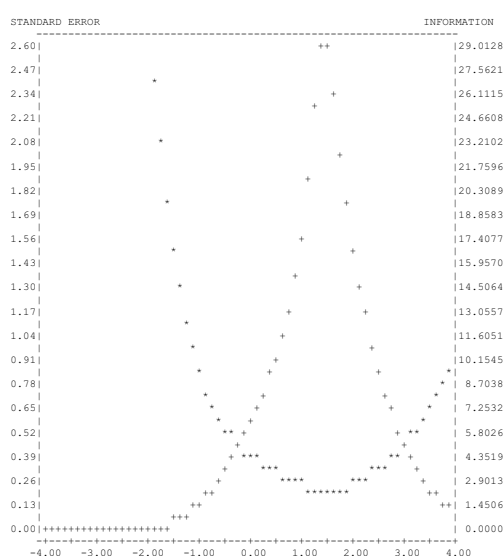
ภาพประกอบ 3 จำนวนองค์ประกอบในแบบทดสอบคณิตศาสตร์

ผลจากการสร้างกราฟตามภาพประกอบ 3 พบว่า ค่าไอเกนจากองค์ประกอบที่ 1 ในแบบทดสอบลดลงอย่างรวดเร็ว ในขณะที่ค่าไอเกนจากองค์ประกอบที่ 2 ลดลงเล็กน้อยมาก เมื่อพิจารณาอัตราส่วนระหว่างองค์ประกอบที่ 1 และองค์ประกอบที่ 2 ในแบบทดสอบคณิตศาสตร์แล้ว มีค่าเท่ากับ 2.85 และเพื่อเป็นการยืนยันว่าแบบทดสอบมีหนึ่งองค์ประกอบจริง จึงได้ทำการวิเคราะห์องค์ประกอบเชิงยืนยัน โดยแปลงตัวแปรคะแนนให้เป็นตัวแปรสุ่มแบบปกติ พบว่าแบบทดสอบมีค่าไค-สแควร์เท่ากับ 2580.04 ค่า p เท่ากับ 0.00 มีชั้นแห่งความเป็นอิสระเท่ากับ 1148 และเมื่อหาค่าไค-สแควร์ด้วยชั้นแห่งความเป็นอิสระได้ค่าเท่ากับ 2.25 แสดงว่าโมเดลไม่สอดคล้องกับข้อมูล แต่ค่าอื่นๆ ได้แก่ RMSEA, NNFI, GFI, AGFI และ CFI มีค่าเท่ากับ 0.04, 0.92, 0.90, 0.89 และ 0.93 ตามลำดับ ซึ่งค่าดังกล่าวอยู่ในเกณฑ์ที่ยอมรับได้ จึงสรุปได้ว่าแบบทดสอบนี้มีหนึ่งองค์ประกอบหลักจริงหรือมีองค์ประกอบที่เด่นชัด (Dominant) เพียงหนึ่งองค์ประกอบ และแสดงให้เห็นว่าแบบทดสอบนี้มีความเที่ยงตรงเชิงโครงสร้าง

10.2 วิเคราะห์คุณลักษณะทางสถิติรายข้อตามทฤษฎีการตอบข้อสอบ ด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ โดยใช้โปรแกรม BILOG-MG3 ได้ค่าความยากอยู่ระหว่าง -0.41 ถึง 2.24 มีค่าเฉลี่ย

เท่ากับ 1.15 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.64 ค่าอำนาจจำแนกอยู่ระหว่าง 0.95 ถึง 6.62 มีค่าเฉลี่ยเท่ากับ 1.87 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.89 ค่าการเดาอยู่ระหว่าง 0.13 ถึง 0.35 มีค่าเฉลี่ยเท่ากับ 0.23 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 0.05 มีข้อสอบที่มีคุณลักษณะทางสถิติรายข้อตามเกณฑ์ของวอร์ม (Warm, 1978: 52-53) ที่กำหนดค่าความยาก (b) ให้อยู่ระหว่าง -2.50 ถึง +2.50 ส่วนค่าอำนาจจำแนก (a) มากกว่า 0.80 ค่าการเดา (c) น้อยกว่า 0.30 จำนวน 44 ข้อ และได้ค่าสารสนเทศข้อสอบแต่ละข้ออยู่ระหว่าง 0.16 ถึง 7.57 จะเห็นได้ว่าค่าฟังก์ชันสารสนเทศของข้อสอบจะมีค่าสูงเมื่อค่าอำนาจจำแนกสูง เนื่องจากฟังก์ชันสารสนเทศของข้อสอบประกอบด้วยค่าพารามิเตอร์ความยาก ค่าอำนาจจำแนก และค่าการเดา ต่อจากนั้นจึงคัดเลือกข้อสอบที่มีคุณลักษณะทางสถิติรายข้อตามเกณฑ์จำนวน 40 ข้อ เพื่อใช้ในการศึกษาครั้งนี้โดยคัดเลือกข้อสอบให้ได้น้ำหนักตามตารางน้ำหนักความสำคัญของมวลความรู้ย่อยที่ได้จากข้อ 3 ด้วยการสุ่มแบบแบ่งชั้น (Stratified random sampling) โดยใช้มวลความรู้ย่อยเป็นชั้น และข้อสอบเป็นหน่วยในการสุ่ม (Sampling unit) ปรากฏผลการวิเคราะห์ดังตาราง 6 ในภาคผนวก ก

10.3 วิเคราะห์คุณลักษณะทางสถิติของแบบทดสอบตามทฤษฎีการตอบข้อสอบด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ โดยใช้โปรแกรม BILOG-MG3 ได้ค่าฟังก์ชันสารสนเทศของแบบทดสอบคณิตศาสตร์มีค่าสูงสุดที่ระดับความสามารถ 1.38 เท่ากับ 29.01 และค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์ต่ำสุดเท่ากับ 0.19 และเพื่อให้เห็นภาพรวมของแบบทดสอบจึงได้นำค่าฟังก์ชันสารสนเทศของแบบทดสอบในแต่ละระดับความสามารถมาเขียนเป็นกราฟ ดังภาพประกอบ 4



ภาพประกอบ 4 ค่าฟังก์ชันสารสนเทศของแบบทดสอบคณิตศาสตร์ในแต่ละระดับความสามารถ

การเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลสำหรับการศึกษาผลของการวัดการเปลี่ยนแปลงในครั้งนี้ ผู้วิจัยได้กำหนดเก็บข้อมูลแบบวัดซ้ำสามครั้งกับผู้สอบคนเดิม คือ ก่อนเรียน ระหว่างเรียนและหลังเรียน โดยแต่ละครั้งห่างกัน 3 สัปดาห์ ตั้งแต่วันที่ 12 พฤศจิกายน 2549 ถึง วันที่ 11 กุมภาพันธ์ 2550 ในการดำเนินการสอบทั้งสามครั้ง ผู้วิจัยได้ใช้ระยะเวลาเก็บรวบรวมข้อมูลทั้งสิ้น 3 เดือน ด้วยแบบทดสอบวิชาคณิตศาสตร์ เนื้อหาเรื่องสมการเชิงเส้นตัวแปรเดียว กับนักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาขั้นพื้นฐาน ระดับมัธยมศึกษาตอนต้น ของศูนย์บริการการศึกษานอกโรงเรียนเขต 7 แห่ง ซึ่งเป็นสถานศึกษาในสังกัดศูนย์การศึกษานอกโรงเรียนกรุงเทพมหานคร 1 จำนวน 1,499 คน ตามขั้นตอน ดังนี้

1. ทำหนังสือจากบัณฑิตวิทยาลัยและติดต่อขอความร่วมมือจากผู้อำนวยการศูนย์การศึกษานอกโรงเรียนกรุงเทพมหานคร 1 ที่เป็นกลุ่มตัวอย่าง เพื่อขอความร่วมมือสถานศึกษาในสังกัด เกี่ยวกับการลงทะเบียนเรียนวิชาคณิตศาสตร์ของกลุ่มตัวอย่าง ซึ่งแจ้งวัตถุประสงค์การเก็บข้อมูล ลักษณะเครื่องมือที่ใช้ จำนวนกลุ่มตัวอย่าง ระยะเวลาและจำนวนครั้งในการเก็บข้อมูล รวมทั้งขอเสนอแนะในการนำคะแนนการเปลี่ยนแปลงรายบุคคลที่ได้ไปใช้ในการวัดผลระหว่างภาค

2. ประสานงานและชี้แจงการดำเนินการสอบแก่ผู้อำนวยการศูนย์บริการการศึกษานอกโรงเรียนเขต ที่เป็นกลุ่มตัวอย่าง พร้อมทั้งชี้แจงการดำเนินการสอบกับครูศูนย์การเรียนชุมชนที่เป็นครูของกลุ่มตัวอย่าง

3. ดำเนินการสอบโดยนำแบบทดสอบที่มีความยาว 40 ข้อ ไปทดสอบกับกลุ่มตัวอย่าง จำนวน 1,499 คน ก่อนลงมือสอบผู้วิจัยหรือครูศูนย์การเรียนชุมชนซึ่งเป็นผู้ช่วยในการเก็บรวบรวมข้อมูลชี้แจงให้นักศึกษาตั้งใจทำข้อสอบ และแจ้งให้ทราบว่าต้องสอบเนื้อหาเรื่องสมการเชิงเส้นตัวแปรเดียวนี้นับสามครั้ง เพื่อนำคะแนนไปเป็นคะแนนสอบระหว่างภาค และผลการสอบครั้งที่สองจะนำไปเป็นคะแนนสอบย่อย (quiz) ซึ่งจะทำได้ผลการสอบที่น่าเชื่อถือได้ ในการสอบแต่ละครั้งมีระยะห่างกัน 3 สัปดาห์ ดังนี้

ครั้งแรก สอบก่อนเรียนเรื่องสมการเชิงเส้นตัวแปรเดียว ตั้งแต่วันที่ 12 พฤศจิกายน 2549 ถึงวันที่ 24 ธันวาคม 2549

ครั้งที่สอง สอบระหว่างเรียนเรื่องสมการเชิงเส้นตัวแปรเดียว ตั้งแต่วันที่ 3 ธันวาคม 2549 ถึงวันที่ 21 มกราคม 2550

ครั้งที่สาม สอบหลังจากเรียนเรื่องสมการเชิงเส้นตัวแปรเดียวเสร็จเรียบร้อยแล้ว ตั้งแต่วันที่ 24 ธันวาคม 2549 ถึงวันที่ 11 กุมภาพันธ์ 2550

การจัดกระทำและการวิเคราะห์ข้อมูล

ในการวิจัยครั้งนี้มีขั้นตอนการจัดกระทำและวิเคราะห์ข้อมูล ดังต่อไปนี้

1. การวิเคราะห์ข้อมูลในการทดลองเพื่อสร้างเครื่องมือ ดังนี้

1.1 ค่าความเที่ยงตรงเชิงเนื้อหา ใช้ดัชนีความสอดคล้องของโรบินลลีและแฮมเบิลตัน

(Hambleton.; et al. 1978: 34; citing Rovinelli; & Hambleton. 1977. Journal for Educational Research.) ถ้าค่าเฉลี่ยของคะแนนการตัดสินของผู้เชี่ยวชาญมีค่าเกิน 0.50 จะแสดงว่ามวลดความรู้อยู่ที่กำหนดขึ้นมีความครอบคลุมมวลดความรู้ในขอบเขตเนื้อหาและข้อสอบที่สร้างขึ้นวัดตามคุณลักษณะเฉพาะข้อสอบที่กำหนดขึ้นในแต่ละมวลดความรู้อยู่ โดยใช้สูตร ดังนี้ (บุญเชิด ภิญโญอนันตพงษ์. 2546: 89)

$$IOC = \frac{\sum R}{N}$$

เมื่อ R เป็นจำนวนผู้เชี่ยวชาญที่ตัดสินว่าสิ่งที่ตรวจสอบสอดคล้องกัน

N เป็นจำนวนผู้เชี่ยวชาญทั้งหมด

1.2 ค่าความยาก ใช้ค่าเฉลี่ยคะแนนรายข้อ (Item mean) คือ

$$P = \frac{\sum X}{N}$$

เมื่อ X เป็นผู้สอบที่ตอบถูก

N เป็นจำนวนผู้สอบ

1.3 ค่าอำนาจจำแนก ใช้ดัชนีเบรนนอน หรือ ดัชนีบี (Brennan Index) คือ (บุญเชิด ภิญโญอนันตพงษ์. 2526: 108; อ้างอิงจาก Crehan. 1974. Journal of Educational Measurement: 257)

$$B = \frac{U}{N_1} - \frac{L}{N_2}$$

เมื่อ N_1 เป็นจำนวนผู้สอบที่สอบได้คะแนนสูงกว่าคะแนนจุดตัด

N_2 เป็นจำนวนผู้สอบที่สอบได้คะแนนต่ำกว่าคะแนนจุดตัด

U เป็นจำนวนผู้สอบในกลุ่ม N_1 ที่ตอบข้อสอบถูก

L เป็นจำนวนผู้สอบในกลุ่ม N_2 ที่ตอบข้อสอบถูก

ผู้วิจัยได้หาคะแนนจุดตัดแบบที่มีการสอบกับผู้สอบกลุ่มเดียวและครั้งเดียวจากสูตรของบุญเชิด ภิญโญนนตพงษ์ ซึ่งเป็นการกำหนดคะแนนจุดตัดถาวรไว้ล่วงหน้าได้ โดยกำหนดจุดตัดถาวร คือ ค่ากึ่งกลางระหว่างคะแนนเต็ม (k) กับคะแนนขอบเขตบนของคะแนนการเดาเฉลี่ย ดังสูตร (บุญเชิด ภิญโญนนตพงษ์. 2535: 18-19)

$$C = \frac{k(A+1) + 3\sqrt{k(A+1)}}{2A}$$

เมื่อ k เป็นจำนวนข้อสอบในแบบทดสอบ

A เป็นจำนวนตัวเลือกของข้อสอบเลือกตอบ

จากสูตรดังกล่าวข้างต้นสามารถกำหนดคะแนนจุดตัดของแบบทดสอบที่มีความยาว 40 ข้อ ไว้ล่วงหน้าได้ที่ 29 คะแนน คิดเป็นร้อยละ 72.50

1.4 ตรวจสอบความเป็นมิติเดียว (Unidimensionality) ของแบบทดสอบด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis; EFA) โดยสกัดองค์ประกอบด้วยวิธีวิเคราะห์องค์ประกอบหลัก (Principal component analysis; PCA) พิจารณาค่าความแปรปรวนสูงสุดจากองค์ประกอบหลักตัวแรกมีสัดส่วนมากกว่าองค์ประกอบอื่นๆ มาก โดยใช้โปรแกรม SPSS และวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis; CFA) เป็นการทดสอบความสอดคล้องระหว่างข้อมูลกับโมเดล ที่กำหนดให้มีตัวแปรแฝงเพียงตัวเดียวในตัวแปรสังเกตทั้งหมด โดยใช้เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วมในการวิเคราะห์ด้วยโปรแกรมลิสเรล (LISREL)

1.5 วิเคราะห์ค่าความยาก อำนาจจำแนก การเดาและค่าฟังก์ชันสารสนเทศของข้อสอบตามทฤษฎีการตอบข้อสอบด้วยโปรแกรม BILOG-MG3 ที่ประมาณค่าด้วยวิธีเบส

1.6 หาค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์; $SE(\theta)$ ค่าฟังก์ชันสารสนเทศของแบบทดสอบ; $I(\theta)$ ด้วยโปรแกรม BILOG-MG3 โดยใช้สูตร ดังนี้ (Hambleton; Swaminathan; & Roger. 1991: 94)

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

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

$$\text{เมื่อ } I_i(\theta) = \sum_{i=1}^n \frac{P_i'(\theta)^2}{P_i(\theta)Q_i(\theta)}$$

$I_i(\theta)$ เป็นฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อ

$P_i(\theta)$ เป็นความน่าจะเป็นที่ผู้สอบ ณ ระดับความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$Q_i(\theta)$ เป็นค่า $1 - P_i(\theta)$

$P'_i(\theta)$ เป็นความชันของโค้งลักษณะข้อสอบ ณ ระดับความสามารถ θ ของข้อสอบข้อที่ i

2. การวิเคราะห์ข้อมูลในการวิจัย ดังนี้

2.1 การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ

2.2 การวิเคราะห์คะแนนการเปลี่ยนแปลง

2.3 การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง

2.4 การศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง

รายละเอียดของการวิเคราะห์ข้อมูลมีดังนี้

2.1 การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ

ผู้วิจัยได้นำผลการตอบข้อสอบของนักศึกษาจำนวน 1,085 คน ที่สอบด้วยแบบทดสอบคณิตศาสตร์ จำนวน 40 ข้อ มาสุ่มด้วยการสุ่มอย่างง่าย ตามเงื่อนไขที่สนใจศึกษา จากทั้งสองเงื่อนไขและระดับของเงื่อนไข เพื่อสุ่มผลการตอบของนักศึกษาออกเป็น 3 ขนาดกลุ่มตัวอย่าง คือ กลุ่มตัวอย่างที่มีจำนวน 500 คน, 300 คน และ 100 คน และสุ่มข้อสอบจาก 40 ข้อ ด้วยการสุ่มแบบแบ่งชั้น โดยมีมวลความรู้ย่อยเป็นชั้น และข้อสอบเป็นหน่วยในการสุ่ม เพื่อเป็นแบบทดสอบที่มีความยาว 30 ข้อ และ 20 ข้อ สามารถจัดข้อมูลสำหรับการศึกษาในครั้งนี้ได้ทั้งหมด 12 ชุด (4X3) ดังนี้

เงื่อนไขที่สนใจศึกษา		ความยาวของแบบทดสอบ		
		40 ข้อ	30 ข้อ	20 ข้อ
ขนาดกลุ่มตัวอย่าง	1,000 คน	X	X	X
	500 คน	X	X	X
	300 คน	X	X	X
	100 คน	X	X	X

ข้อมูลสำหรับการศึกษาในครั้งนี้ดังกล่าวข้างต้น เป็นผลการตอบข้อสอบของนักศึกษาที่เป็นกลุ่มตัวอย่าง 4 ขนาด คือ กลุ่มตัวอย่างที่มีจำนวน 1,085 คน, 500 คน, 300 คน และ 100 คน โดยการสุ่มขนาด 500 คน, 300 คน และ 100 คน จากขนาดกลุ่มตัวอย่างที่สอบ 1,085 คน ด้วยการสุ่มอย่างง่าย ซึ่งแต่ละกลุ่มตัวอย่างสอบด้วยแบบทดสอบที่มีความยาว 3 ขนาด คือ แบบทดสอบที่มีความยาว 40 ข้อ, 30 ข้อ และ 20 ข้อ ซึ่งความยาวของแบบทดสอบ 30 และ 20 ข้อ ได้มาจากการสุ่มแบบแบ่งชั้น โดยมีมวลความรู้ย่อยเป็นชั้น และข้อสอบเป็นหน่วยในการสุ่ม ซึ่งจะทำให้ได้ข้อสอบตามน้ำหนักความสำคัญของแต่ละมวลความรู้ เพื่อวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบแต่ละครั้ง ดังนี้

2.1.1 วิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบแต่ละครั้ง เป็นการตรวจสอบลักษณะโดยทั่วไปของข้อมูล ได้แก่ คะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด คะแนนต่ำสุด พิสัย มัธยฐานฐานนิยม ความเบ้ ความโด่ง และความคลาดเคลื่อนมาตรฐาน รวมทั้งตรวจสอบลักษณะการแจกแจง ด้วยโปรแกรม SPSS

2.1.2 ตรวจสอบแบบแผนการเปลี่ยนแปลง โดยใช้การวิเคราะห์การถดถอย ด้วยโปรแกรม SPSS

2.1.3 วิเคราะห์ความสัมพันธ์ระหว่างคะแนนสอบแต่ละครั้ง โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ด้วยโปรแกรม SPSS

2.1.4 วิเคราะห์ความแตกต่างของค่าเฉลี่ยของคะแนนสอบทั้งสามครั้ง โดยใช้วิธีการวิเคราะห์ความแปรปรวนแบบวัดซ้ำ (Repeated measure analysis of variance) ด้วยโปรแกรม SPSS

2.2 การวิเคราะห์คะแนนการเปลี่ยนแปลง

ผู้วิจัยได้นำผลการตอบของกลุ่มตัวอย่าง 4 ขนาด คือ กลุ่มตัวอย่างที่มีจำนวน 1,085 คน, 500 คน, 300 คน และ 100 คน โดยแต่ละกลุ่มสอบด้วยแบบทดสอบที่มีจำนวนข้อสอบ 3 ขนาด คือ แบบทดสอบที่มีความยาว 40 ข้อ, 30 ข้อ และ 20 ข้อ มาประมาณค่าคะแนนการเปลี่ยนแปลงด้วยการวัดการเปลี่ยนแปลงทำวิธีที่สนใจและวิธีโมเดลพหุระดับ ซึ่งเป็นวิธีที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ และวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี ดังนี้

2.2.1 วิธีการวัดการเปลี่ยนแปลงที่สนใจศึกษา

ในการวิจัยครั้งนี้เป็นการศึกษาผลของการวัดการเปลี่ยนแปลงที่มีการวัดซ้ำสองครั้ง ซึ่งจะใช้ผลการตอบข้อสอบของผู้สอบในครั้งแรกและครั้งที่สามสำหรับการวิเคราะห์คะแนนการเปลี่ยนแปลง มีทำวิธีดังนี้

2.2.1.1 การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ เป็นการหาผลต่างระหว่างคะแนนดิบก่อนเรียนกับหลังเรียน จากสูตร $X_2 - X_1$ ด้วยโปรแกรม MS-EXCEL

2.2.1.2 การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ เป็นการหาร้อยละของอัตราส่วนระหว่างผลต่างระหว่างคะแนนดิบกับศักยภาพของผู้สอบ จากสูตร $\left(\frac{X_2 - X_1}{F - X_1} \right) 100$ ด้วยโปรแกรม MS-EXCEL

2.2.1.3 การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง เป็นการหาผลต่างระหว่างความสามารถก่อนเรียนกับหลังเรียน จากสูตร $\theta_2 - \theta_1$ ด้วยโปรแกรม MS-EXCEL โดยหาค่าความสามารถก่อนเรียนและหลังเรียนด้วยวิธีการประมาณค่า EAP (Expected a posteriori) ในโปรแกรม BILOG-MG3 เนื่องจากผลการศึกษา (Allen; & Yen. 1987; Mislevy; & Stocking. 1989; Swaminathan; & Griffor. 1988; Muraki; & Engelhard, 1985. รัตนา ศรีเหรียญ. 2539; & วิชุดา บัวคง. 2533; สุชาติ กิระนันท์. 2545) พบว่า เมื่อข้อสอบจำนวนน้อย ข้อสอบที่มีความยากมาก กลุ่มตัวอย่างมีขนาดเล็ก และโดยเฉพาะการประมาณค่าความสามารถผู้สอบที่มีความสามารถปานกลางและต่ำ ควรใช้วิธีการประมาณค่าพารามิเตอร์ด้วยค่าภายหลังที่คาดหวัง (Expected a posteriori; EAP) ของเบส์ จะทำให้การประมาณค่าพารามิเตอร์ความสามารถได้ถูกต้อง น่าเชื่อถือมากกว่าวิธีอื่น ซึ่งในโปรแกรม BILOG-MG3 สามารถเลือกใช้วิธีดังกล่าวประมาณค่าพารามิเตอร์ได้

2.2.1.4 การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ เป็นการหาผลต่างระหว่างคะแนนมวลความรู้ก่อนเรียนกับหลังเรียนตามทฤษฎีการตอบข้อสอบ จากสูตร $\pi_2 - \pi_1$ ด้วยโปรแกรม MS-EXCEL โดยมีขั้นตอนการหาค่าคะแนนมวลความรู้ของผลการสอบแต่ละครั้ง ด้วยโปรแกรม BILOG-MG3 ดังนี้

2.2.1.4.1 หาค่าพารามิเตอร์ข้อสอบทั้งสามค่า คือ ค่าความยาก อำนาจจำแนก และการเดาที่ได้จากการประมาณค่าด้วยวิธี EAP (Expected a posteriori)

2.2.1.4.2 นำค่าพารามิเตอร์ข้อสอบที่ทราบแล้วมาจัดรูปแบบ ตามแบบ 4F10.5 แล้วนำผลการตอบข้อสอบทุกข้อของผู้สอบทุกคนมาวิเคราะห์หาค่าอีกครั้ง เพื่อประมาณค่าความสามารถของผู้สอบแต่ละคนในรูปคะแนนมวลความรู้ ด้วยสูตร $\pi = \frac{1}{n} \sum_{j=1}^n P_j(\theta)$

2.2.1.5 การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ เป็นการหาร้อยละของอัตราส่วนระหว่างผลต่างระหว่างคะแนนมวลความรู้กับศักยภาพของผู้สอบ จากสูตร

$\left(\frac{\pi_2 - \pi_1}{100 - \pi_1} \right) 100$ ด้วยโปรแกรม MS-EXCEL โดยมีขั้นตอนการหาค่าคะแนนมวลความรู้เช่นเดียวกับ

ข้อ 2.2.1.4.1 – 2.2.1.4.2 ดังกล่าวข้างต้น

2.2.2 วิธีโมเดลพหุระดับ เป็นวิธีการประมาณค่าคะแนนการเปลี่ยนแปลงที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ เป็นการนำข้อมูลจากการสอบซ้ำสามครั้งมาวิเคราะห์ตามขั้นตอนต่อไปนี้

2.2.2.1 หารูปแบบความสัมพันธ์ที่เหมาะสมสำหรับข้อมูล จึงได้ตรวจสอบความสัมพันธ์ของตัวแปรระหว่างคะแนนจากการสอบสามครั้งกับเวลา ว่าความสัมพันธ์อยู่ในรูปแบบใด อยู่ในรูปเชิงเส้นหรือไม่ โดยใช้เทคนิคการวิเคราะห์ความถดถอย ด้วยโปรแกรม SPSS

2.2.2.2 คำนวณคะแนนการเปลี่ยนแปลง ซึ่งโมเดลพหุระดับ (Multilevel model) สำหรับการวัดในครั้งนี้เป็นโมเดลพหุระดับ แบบ 2 ระดับ ด้วยโปรแกรม HLM ดังนี้

ระดับที่ 1 เป็นโมเดลวัดซ้ำ (Repeated-observation model) ซึ่งแบ่งเป็น 2 ขั้นตอนย่อย คือ

1.1 วิเคราะห์ชั้นโมเดลไร้เงื่อนไขแบบสมบูรณ์ (Fully unconditional model) หรือโมเดลศูนย์ (Null model) เพื่อตรวจสอบความแปรปรวนของตัวแปรตามในหน่วย (Within subjects) และระหว่างหน่วย (Between subjects)

1.2 วิเคราะห์ชั้นโมเดลไร้เงื่อนไข (Unconditional model) หรือ โมเดลอย่างง่าย (Simple model) ผลการวิเคราะห์จะได้ค่าพารามิเตอร์ต่างๆ คือ ค่าความสามารถเริ่มต้น อัตราพัฒนาการ ความเชื่อมั่นในการประมาณค่าพารามิเตอร์ และความสัมพันธ์ระหว่างอัตราพัฒนาการกับความสามารถเริ่มต้น โดยในขั้นนี้วิเคราะห์ไฟล์ส่วนที่เหลือระดับ 1 (Residual files level-2) เพื่อวิเคราะห์คะแนนจริงของผลการสอบแต่ละครั้งเป็นรายคน

ระดับที่ 2 โมเดลระดับบุคคล (Person-level model) เป็นการวิเคราะห์ชั้นโมเดลสมมติฐาน (Hypothetical model) เพื่อหาความสัมพันธ์ระหว่างปัจจัยที่ส่งผลต่อความสามารถเริ่มต้นและอัตราพัฒนาการ ซึ่งในการศึกษาครั้งนี้ปัจจัยที่ส่งผลต่อความสามารถเริ่มต้นและอัตราพัฒนาการ คือ จำนวนครั้งในการสอบหรือเวลา โดยในขั้นตอนนี้วิเคราะห์ไฟล์ส่วนที่เหลือระดับ 2 (Residual files level-2) เพื่อวิเคราะห์ความสามารถเริ่มต้นและอัตราพัฒนาการเป็นรายบุคคลที่เป็นคะแนนจริง

ในขั้นนี้เริ่มต้นการวิเคราะห์ด้วยการตรวจสอบแบบแผนการเปลี่ยนแปลง พิจารณาความสอดคล้องกลมกลืนของโมเดลกับข้อมูลเชิงประจักษ์ ซึ่งจะตรวจสอบกับโมเดลพหุระดับ 3 รูปแบบ ซึ่งเป็นโมเดลเชิงเส้นตรงตัวแปรตามเป็นฟังก์ชันที่ขึ้นอยู่กับค่าเฉลี่ยของเวลาในการวัด คือ

2.1 รูปแบบโมเดลเชิงเส้นตรงตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model)

2.2 รูปแบบโมเดลเชิงเส้นตรงตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยของตัวแปรต้น (Group centered model)

2.3 รูปแบบโมเดลเชิงเส้นตรงตัวแปรตามเป็นฟังก์ชันที่ยืดค่าเฉลี่ยรวมของตัวแปรต้น

(Grand centered model)

ซึ่งแต่ละรูปแบบโมเดลจะให้ค่าคะแนนเริ่มต้นหรือคะแนนสอบก่อนเรียนต่างกัน เพื่อนำรูปแบบโมเดลที่สอดคล้องกลมกลืนกับข้อมูลมากที่สุด เป็นรูปแบบโมเดลที่สามารถอธิบายความแปรปรวนได้สูงสุด โดยพิจารณาจากค่าสัมประสิทธิ์การพยากรณ์ (R^2) สูงสุด มาใช้ในการประมาณค่าคะแนนเริ่มต้นและคะแนนการเปลี่ยนแปลงสำหรับเป็นคะแนนเกณฑ์

2.2.3 วิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงรายบุคคล เป็นการตรวจสอบลักษณะโดยทั่วไปของข้อมูล ได้แก่ คะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด คะแนนต่ำสุด มัธยฐาน ฐานนิยม ความเบ้ ความโด่ง และความคลาดเคลื่อนมาตรฐาน รวมทั้งตรวจสอบลักษณะการแจกแจง ด้วยโปรแกรม SPSS

2.3 การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง

ผู้วิจัยศึกษาความสัมพันธ์ของคะแนนการวัดการเปลี่ยนแปลงโดยการศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง และเปรียบเทียบความสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ตามเงื่อนไขความยาวของแบบทดสอบที่ต่างกันและขนาดกลุ่มตัวอย่างไม่เท่ากัน โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ดังนี้

2.3.1 การศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง เป็นการศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษาและวิธีโมเดลพหุระดับ ตามเงื่อนไขความยาวของแบบทดสอบที่ต่างกันและขนาดกลุ่มตัวอย่างไม่เท่ากัน โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ด้วยโปรแกรม SPSS

2.3.2 การเปรียบเทียบความสัมพันธ์ของคะแนนการเปลี่ยนแปลง เป็นการเปรียบเทียบความสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ตามเงื่อนไขความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน มีขั้นตอนดังนี้

2.3.2.1 หาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างคะแนนการเปลี่ยนแปลงที่ได้จากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ

2.3.2.2 ทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โดยนำค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันมาแปลงเป็นคะแนนมาตรฐาน (Z) ของฟิชเชอร์ แล้วนำมาทดสอบ โดยใช้สถิติไค-สแควร์ (Chi-Square; χ^2) ดังนี้ (Meng; Rosenthal; & Rubin. 1992: 173)

$$\chi^2(k-1) = \frac{(N-3) \sum_i (Z_{ri} - \bar{Z}_r)^2}{(1-r_x)h}$$

โดยที่

$$h = 1 + \frac{\bar{r}^2}{1-r^2}(1-f)$$

$$f = \frac{1-r_x}{2(1-r^2)}$$

เมื่อ Z_r เป็นคะแนนมาตรฐานที่แปลงมาจากค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ของฟิชเชอร์

r เป็นค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

r_x เป็นค่ามัธยฐาน (Median) ของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการ

เปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงระหว่างแต่ละวิธี
ที่สนใจศึกษา

N เป็นจำนวนผู้สอบในกลุ่มตัวอย่าง

k เป็นจำนวนค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่ต้องการเปรียบเทียบ

2.3.2.3 ถ้าพบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่นำมาทดสอบในข้อ 2.3.2.2 มีความแตกต่างกันอย่างมีนัยสำคัญ นำค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันมาทดสอบความแตกต่างรายคู่ โดยใช้สถิติ Z ดังนี้ (Meng; Rosenthal; & Rubin. 1992: 173)

$$Z = (Z_{r1} - Z_{r2}) \sqrt{\frac{N-3}{2(1-r_x)h}}$$

เมื่อ Z_r เป็นคะแนนมาตรฐานที่แปลงมาจากค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ของฟิชเชอร์

r เป็นค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

- r_x เป็นค่ามัธยฐาน (Median) ของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงระหว่างแต่ละวิธีที่สนใจศึกษา
- N เป็นจำนวนผู้สอบในกลุ่มตัวอย่าง
- k เป็นจำนวนค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่ต้องการเปรียบเทียบ

2.4 การศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง

ผู้วิจัยศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลงโดยการศึกษาความเชื่อมั่นและความเที่ยงตรงของการวัดการเปลี่ยนแปลง โดยพิจารณาจากค่าความคลาดเคลื่อนของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ที่แปลงมาอยู่ในมาตรเดียวกันแล้ว ตามเงื่อนไขความยาวของแบบทดสอบที่ต่างกันและขนาดกลุ่มตัวอย่างไม่เท่ากัน ซึ่งได้แก่ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard error of estimate) และค่าความลำเอียง (BIAS)

ในการศึกษาครั้งนี้ ผู้วิจัยจะแปลงคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษาให้อยู่ในมาตรเดียวกัน เพื่อสามารถเปรียบเทียบกันได้ โดยแปลงเป็นคะแนนมาตรฐาน (Z-score) ต่อจากนั้นจึงนำคะแนนมาตรฐานมาจัดกระทำต่อไป ดังนี้

2.4.1 วิเคราะห์ความเชื่อมั่นของการวัดการเปลี่ยนแปลง เป็นการพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard error of estimate) ด้วยการหาค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root mean squared of error; RMSE) ซึ่งค่า RMSE เป็นวิธีประเมินความคงที่ (Stability) ในการประมาณค่าพารามิเตอร์ที่ให้ค่าที่บอกถึงขนาดของความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ ถ้าค่า RMSE ต่ำเข้าใกล้ศูนย์ หมายความว่า วิธีวัดคะแนนการเปลี่ยนแปลงที่สนใจศึกษามีความเชื่อมั่นสูงในการประมาณค่า โดยมีสูตร ดังนี้ (Bock; Thissen; & Zimowski. 1997: 199)

$$RMSE = \left[\sum_i^N (\theta_{com(i)} - \theta_{cri(i)})^2 / N \right]^{\frac{1}{2}}$$

เมื่อ $\theta_{com(i)}$ เป็นคะแนนการเปลี่ยนแปลงจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษาของคนที i

$\theta_{cri(i)}$ เป็นคะแนนการเปลี่ยนแปลงจากวิธีโมเดลพหุระดับของคนที i ซึ่งใช้เป็นเกณฑ์

N เป็นจำนวนผู้สอบทั้งหมด

2.4.2 วิเคราะห์ความเที่ยงตรงของการวัดการเปลี่ยนแปลง เป็นการพิจารณาจากค่าความลำเอียง (BIAS) ซึ่งเป็นค่าความลำเอียงเฉลี่ยที่คิดเครื่องหมาย (Average signed bias; ASB) ซึ่งเป็นการบอกทิศทางของความลำเอียง ค่า BIAS ต่ำเข้าใกล้ศูนย์ หมายความว่า วิธีวัดคะแนนการเปลี่ยนแปลงที่สนใจศึกษามีความเที่ยงตรงสูงในการประมาณค่า โดยมีสูตร ดังนี้ (Nering. 1995: 124)

$$BIAS = \sum_{i=1}^N (\theta_{com(i)} - \theta_{cri(i)}) / N$$

เมื่อ $\theta_{com(i)}$ เป็นคะแนนการเปลี่ยนแปลงจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจ
ศึกษาของคนที่ i

$\theta_{cri(i)}$ เป็นคะแนนการเปลี่ยนแปลงจากวิธีโมเดลพหุระดับของคนที่ i ซึ่งใช้เป็นเกณฑ์

N เป็นจำนวนผู้สอบทั้งหมด

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาผลของการวัดการเปลี่ยนแปลงรายบุคคลที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษาจำนวนห้าวิธี คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ โดยพิจารณาจากค่าความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง ความสัมพันธ์ของคะแนนการเปลี่ยนแปลง ความคลาดเคลื่อนในการประมาณค่าคะแนนการเปลี่ยนแปลง และค่าความลำเอียงของคะแนนการเปลี่ยนแปลง ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน ซึ่งผู้วิจัยจะนำเสนอผลการวิเคราะห์ข้อมูลเป็น 4 ตอน ดังนี้

- ตอนที่ 1 การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ
- ตอนที่ 2 การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลง
- ตอนที่ 3 การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง
- ตอนที่ 4 การศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง

ในการนำเสนอผลการวิเคราะห์ข้อมูล เพื่อให้เกิดความเข้าใจในผลของการวิจัยตรงกัน ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้แทนค่าสถิติและตัวแปรต่างๆ ดังนี้

N	แทน	ขนาดกลุ่มตัวอย่าง
K	แทน	ความยาวของแบบทดสอบ
T	แทน	ครั้งที่ในการสอบ
X	แทน	คะแนนดิบ
θ	แทน	ค่าความสามารถ
π	แทน	คะแนนมวลความรู้
$\bar{X}$	แทน	ค่าเฉลี่ยเลขคณิตของคะแนนดิบ
Mdn	แทน	ค่ามัธยฐาน
Mo	แทน	ค่าฐานนิยม
Max	แทน	ค่าสูงสุด
Min	แทน	ค่าต่ำสุด

R	แทน	ค่าพิสัย
S	แทน	ส่วนเบี่ยงเบนมาตรฐาน
S.E.	แทน	ความคลาดเคลื่อนมาตรฐาน
Sk	แทน	ค่าความเบ้ของการแจกแจงความถี่ของคะแนน
Ku	แทน	ค่าความโด่งของการแจกแจงความถี่ของคะแนน
MED	แทน	ค่าความแตกต่างสูงสุดของความน่าจะเป็นสะสมที่เกิดขึ้นจริงของข้อมูลกับความน่าจะเป็นสะสมแบบโค้งปกติ
K-S	แทน	ค่าสถิติของ Komolgorov-Smirnov
r	แทน	ค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน
MD	แทน	ค่าเฉลี่ยของคะแนนการเปลี่ยนแปลง
D	แทน	วิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ
S	แทน	วิธีการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์
d_0	แทน	วิธีการวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง
D_0	แทน	วิธีการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้
DS_0	แทน	วิธีการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์
HLM	แทน	วิธีการวัดคะแนนการเปลี่ยนแปลงด้วยโมเดลพหุระดับ
Var(INT)	แทน	ความแปรปรวนของคะแนนเริ่มต้นหรือคะแนนสอบก่อนเรียน
Var(SLP)	แทน	ความแปรปรวนของคะแนนการเปลี่ยนแปลง
Cor(INT,SLP)	แทน	สหสัมพันธ์ระหว่างคะแนนเริ่มต้นกับคะแนนการเปลี่ยนแปลง
R^2	แทน	สัมประสิทธิ์การพยากรณ์

ตอนที่ 1 การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ

ผู้วิจัยได้แบ่งการวิเคราะห์ในตอนนี้ออกเป็น 4 ขั้นตอน คือ 1) การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบแต่ละครั้ง 2) การตรวจสอบแบบแผนการเปลี่ยนแปลง 3) การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนสอบแต่ละครั้ง และ 4) การวิเคราะห์ความแตกต่างของค่าเฉลี่ยของคะแนนสอบทั้งสามครั้ง ซึ่งการวิเคราะห์ดังกล่าวเป็นการวิเคราะห์ข้อมูลเบื้องต้นเพื่ออธิบายคุณลักษณะของคะแนนดิบที่ได้จากการสอบทั้งสามครั้งของกลุ่มตัวอย่างที่ไม่เท่ากัน คือ 1,085 คน, 500 คน, 300 คน และ 100 คน ด้วยแบบทดสอบคณิตศาสตร์ เรื่องสมการเชิงเส้นตัวแปรเดียว ที่มีจำนวนข้อสอบที่ต่างกัน คือ 40 ข้อ, 30 ข้อ และ 20 ข้อ โดยมีรายละเอียดดังนี้

1.1 การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบแต่ละครั้ง

การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบแต่ละครั้งทั้งสามครั้ง ผู้วิจัยตรวจสอบลักษณะโดยทั่วไปของข้อมูล ได้แก่ คะแนนเฉลี่ย ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด คะแนนต่ำสุด พิสัย มัธยฐาน ฐานนิยม ความเบ้ และความโด่ง โดยได้จำแนกข้อมูลออกตามความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ปรากฏผลการวิเคราะห์ดังตาราง 7 (และรายละเอียดการวิเคราะห์ปรากฏดังตาราง 8 ในภาคผนวก ก)

ตาราง 7 ค่าสถิติพื้นฐานของคะแนนการสอบซ้ำสามครั้ง

N	K	T1				T2				T3			
		$\bar{X}$	S	Sk	Ku	$\bar{X}$	S	Sk	Ku	$\bar{X}$	S	Sk	Ku
1,085 คน	40 ข้อ	13.77	6.74	1.37	1.59	19.20	9.50	0.79	-0.58	23.48	9.98	0.19	-1.29
	30 ข้อ	9.96	5.26	1.33	1.54	14.03	7.58	0.78	-0.65	17.14	7.80	0.20	-1.30
	20 ข้อ	7.02	3.64	1.10	0.92	9.90	4.95	0.64	-0.70	11.91	5.19	0.11	-1.26
500 คน	40 ข้อ	13.93	6.79	1.34	1.39	19.26	9.42	0.80	-0.55	23.58	9.95	0.16	-1.29
	30 ข้อ	10.06	5.36	1.32	1.38	14.23	7.50	0.79	-0.62	17.41	7.79	0.17	-1.28
	20 ข้อ	7.16	3.67	1.07	0.75	9.88	4.90	0.66	-0.62	11.99	5.18	0.10	-1.26
300 คน	40 ข้อ	13.87	7.11	1.23	0.87	19.36	9.90	0.75	-0.68	24.10	10.08	0.10	-1.35
	30 ข้อ	10.09	5.46	1.22	0.96	14.21	7.78	0.75	-0.74	17.57	7.83	0.15	-1.40
	20 ข้อ	7.06	3.74	1.01	0.42	9.95	5.12	0.61	-0.79	12.20	5.17	0.03	-1.27
100 คน	40 ข้อ	13.33	6.06	1.49	1.94	19.44	8.38	0.91	-0.08	23.16	8.49	0.24	-1.00
	30 ข้อ	9.57	4.75	1.39	1.73	14.40	6.95	0.89	-0.30	16.96	6.70	0.27	-0.98
	20 ข้อ	6.83	3.32	1.15	0.97	9.99	4.54	0.73	-0.42	11.93	4.59	0.03	-1.16

ผลการวิเคราะห์ตามตาราง 7 สามารถสรุปได้ ดังนี้

1.1.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 40 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 23.48, 19.20 และ 13.77 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 23.00, 16.00 และ 12.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 14.00, 11.00 และ 10.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มี

ค่าเท่ากับ 9.98 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 9.50 และครั้งที่ 1 มีค่าเท่ากับ 6.74

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่าคะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.37 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.79 และครั้งที่ 3 มีค่าเท่ากับ 0.19 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 1.59 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.58 และ -1.29 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.2 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 30 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 17.14, 14.03 และ 9.96 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 17.00, 11.00 และ 8.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 10.00, 8.00 และ 7.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 7.80 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 7.58 และครั้งที่ 1 มีค่าเท่ากับ 5.26

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่าคะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.33 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.78 และครั้งที่ 3 มีค่าเท่ากับ 0.20 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 1.54 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.65 และ -1.30 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.3 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 20 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของ

คะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 11.91, 9.90 และ 7.02 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 12.00, 8.00 และ 6.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 20.00, 6.00 และ 5.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 5.19 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 4.95 และครั้งที่ 1 มีค่าเท่ากับ 3.64

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่าคะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.10 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.64 และครั้งที่ 3 มีค่าเท่ากับ 0.11 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 0.92 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.70 และ -1.26 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.4 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 40 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และ ครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 23.58, 19.26 และ 13.93 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 23.00, 16.00 และ 12.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 39.00, 15.00 และ 10.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 9.95 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 9.42 และครั้งที่ 1 มีค่าเท่ากับ 6.79

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่าคะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.34 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.80 และครั้งที่ 3 มีค่าเท่ากับ 0.16 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 1.39 ส่วนคะแนนการสอบครั้งที่

2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.55 และ -1.29 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.5 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 30 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และ ครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 17.41, 14.23 และ 10.06 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 17.00, 12.00 และ 8.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 10.00, 9.00 และ 7.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 7.79 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 7.50 และ ครั้งที่ 1 มีค่าเท่ากับ 5.36

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่า คะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.32 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.79 และครั้งที่ 3 มีค่าเท่ากับ 0.17 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 1.38 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.62 และ -1.28 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.6 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 20 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และ ครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 11.99, 9.88 และ 7.16 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 12.00, 8.00 และ 6.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 20.00, 6.00 และ 5.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 5.18 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 4.90 และครั้งที่ 1 มีค่าเท่ากับ 3.67

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่าคะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.07 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.66 และครั้งที่ 3 มีค่าเท่ากับ 0.10 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 0.75 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.62 และ -1.26 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.7 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 40 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และ ครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 24.10, 19.36 และ 13.87 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 23.50, 16.00 และ 12.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 40.00, 13.00 และ 9.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 10.08 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 9.90 และครั้งที่ 1 มีค่าเท่ากับ 7.11

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่าคะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.23 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.75 และครั้งที่ 3 มีค่าเท่ากับ 0.10 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 0.87 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.68 และ -1.35 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.8 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 30 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และ ครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 17.57, 14.21 และ 10.09 คะแนน ตามลำดับ มีค่า

มัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 17.50, 12.00 และ 9.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 30.00, 9.00 และ 6.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 7.83 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 7.78 และ ครั้งที่ 1 มีค่าเท่ากับ 5.46

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่าคะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.22 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.75 และครั้งที่ 3 มีค่าเท่ากับ 0.15 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 0.96 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.74 และ -1.40 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.9 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 20 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และ ครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 12.20, 9.95 และ 7.06 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 12.00, 8.00 และ 6.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 20.00, 8.00 และ 6.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 5.17 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 5.12 และครั้งที่ 1 มีค่าเท่ากับ 3.74

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่าคะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.01 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.61 และครั้งที่ 3 มีค่าเท่ากับ 0.03 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 0.42 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.79 และ -1.27 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.10 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 40 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 23.16, 19.44 และ 13.33 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 24.00, 17.00 และ 11.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 28.00, 12.00 และ 10.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 8.49 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 8.38 และครั้งที่ 1 มีค่าเท่ากับ 6.06

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่า คะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.49 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.91 และครั้งที่ 3 มีค่าเท่ากับ 0.24 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 1.94 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.08 และ -1.00 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.11 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 30 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และ ค่าฐานนิยมของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 16.96, 14.40 และ 9.57 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 18.00, 12.00 และ 8.00 คะแนน ตามลำดับ และมีค่าฐานนิยมของคะแนนสอบแต่ละครั้งเท่ากับ 21.00, 9.00 และ 7.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 6.70 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 6.95 และครั้งที่ 1 มีค่าเท่ากับ 4.75

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่า คะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.39 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่า

เท่ากับ 0.89 และครั้งที่ 3 มีค่าเท่ากับ 0.27 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 1.73 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.30 และ -0.98 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

1.1.12 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้งของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 20 ข้อ พบว่า ค่ากลางของคะแนนสอบแต่ละครั้ง คือ คะแนนเฉลี่ย และค่ามัธยฐานของคะแนนจากการสอบซ้ำสามครั้งมีแนวโน้มเพิ่มขึ้น โดยที่ค่ากลางของคะแนนจากการสอบครั้งที่ 3 มีค่าสูงสุด รองลงมาเป็นค่ากลางของคะแนนจากการสอบครั้งที่ 2 และ ครั้งที่ 1 คือ มีคะแนนเฉลี่ยของคะแนนสอบแต่ละครั้งเท่ากับ 11.93, 9.99 และ 6.83 คะแนน ตามลำดับ มีค่ามัธยฐานของคะแนนสอบแต่ละครั้งเท่ากับ 13.00, 9.00 และ 6.00 คะแนน ตามลำดับ ส่วนค่าฐานนิยมของคะแนนจากการสอบแต่ละครั้ง มีค่าไม่เป็นระบบ คือมีค่าเท่ากับ 14.00, 6.00 และ 7.00 คะแนน ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้น โดยส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 3 มีค่าเท่ากับ 4.59 รองลงมาเป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 4.54 และครั้งที่ 1 มีค่าเท่ากับ 3.32

เมื่อพิจารณาลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้และความโด่ง พบว่า คะแนนจากการสอบทั้งสามครั้งมีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยคะแนนจากการสอบครั้งที่ 1 มีความเบ้ขวามากที่สุด คือ มีค่าเท่ากับ 1.15 รองลงมาเป็นคะแนนการสอบครั้งที่ 2 มีค่าเท่ากับ 0.73 และครั้งที่ 3 มีค่าเท่ากับ 0.03 ตามลำดับ ในขณะที่คะแนนการสอบครั้งที่ 1 มีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ 0.97 ส่วนคะแนนการสอบครั้งที่ 2 และ 3 มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ คือ มีค่าความโด่งเท่ากับ -0.42 และ -1.16 ตามลำดับ โดยคะแนนจากการสอบครั้งที่ 3 มีลักษณะการแจกแจงที่แบนกว่าคะแนนการสอบครั้งที่ 2

แต่อย่างไรก็ตาม ผู้วิจัยได้ตรวจสอบลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากการสอบซ้ำสามครั้ง ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกันและขนาดกลุ่มตัวอย่างไม่เท่ากัน ว่ามีลักษณะการแจกแจงเป็นแบบโค้งปกติหรือไม่ โดยการเปรียบเทียบความน่าจะเป็นสะสมของคะแนนสอบกับการแจกแจงแบบโค้งปกติ ด้วยการนำข้อมูลดังกล่าวมาทดสอบโดยใช้สถิติ Kolmogorov-Smirnov Test ปรากฏผลการวิเคราะห์ดังตาราง 9

ตาราง 9 ค่าสถิติ Kolmogorov-Smirnov ของคะแนนสอบแต่ละครั้ง

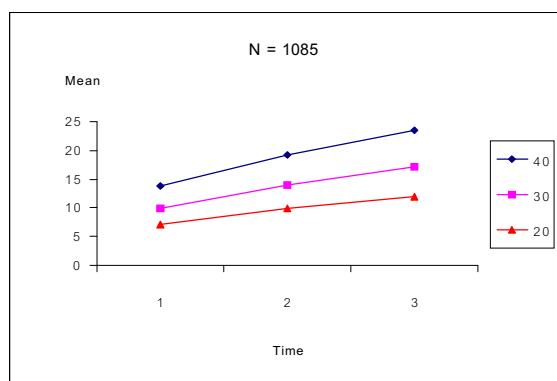
N	K	T1			T2			T3		
		MED	K-S	p	MED	K-S	p	MED	K-S	p
1,085 คน	40 ข้อ	0.18	6.00**	0.00	0.16	5.26**	0.00	0.13	4.11**	0.00
	30 ข้อ	0.18	5.91**	0.00	0.17	5.47**	0.00	0.13	4.24**	0.00
	20 ข้อ	0.18	5.94**	0.00	0.16	5.21**	0.00	0.12	3.78**	0.00
500 คน	40 ข้อ	0.20	4.41**	0.00	0.17	3.69**	0.00	0.12	2.65**	0.00
	30 ข้อ	0.18	4.10**	0.00	0.17	3.78**	0.00	0.12	2.62**	0.00
	20 ข้อ	0.19	4.13**	0.00	0.15	3.42**	0.00	0.11	2.55**	0.00
300 คน	40 ข้อ	0.18	3.14**	0.00	0.17	2.91**	0.00	0.12	2.01**	0.00
	30 ข้อ	0.19	3.20**	0.00	0.19	3.20**	0.00	0.14	2.41**	0.00
	20 ข้อ	0.18	3.07**	0.00	0.16	2.75**	0.00	0.12	2.09**	0.00
100 คน	40 ข้อ	0.17	1.73**	0.00	0.16	1.58	0.02	0.12	1.22	0.10
	30 ข้อ	0.17	1.74**	0.00	0.20	2.00**	0.00	0.15	1.48	0.03
	20 ข้อ	0.20	1.98**	0.00	0.16	1.56	0.02	0.14	1.37	0.04

** p<.01

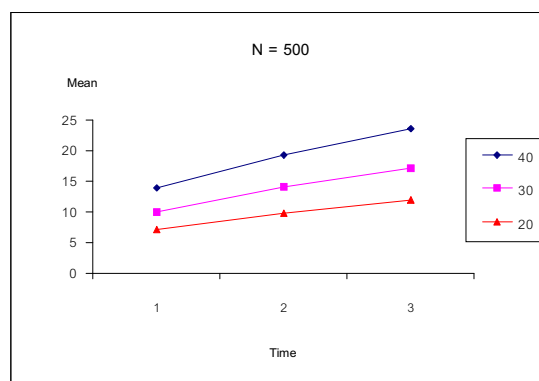
ผลการวิเคราะห์ตามตาราง 9 พบว่า คะแนนสอบทั้งสามครั้ง ในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง มีลักษณะการแจกแจงไม่เป็นโค้งปกติ นั่นคือ คะแนนสอบทั้งสามครั้งส่วนใหญ่มีลักษณะการแจกแจงแบบเบ้ขวา ยกเว้น คะแนนการสอบของนักศึกษาขนาดกลุ่มตัวอย่าง 100 คน จากการสอบครั้งที่ 2 ในแบบทดสอบที่มีความยาว 40 ข้อ และ 20 ข้อ และจากการสอบครั้งที่ 3 ในทุกความยาวของแบบทดสอบ ที่พบว่า มีลักษณะการแจกแจงเป็นโค้งปกติ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

1.2 การตรวจสอบแบบแผนการเปลี่ยนแปลง

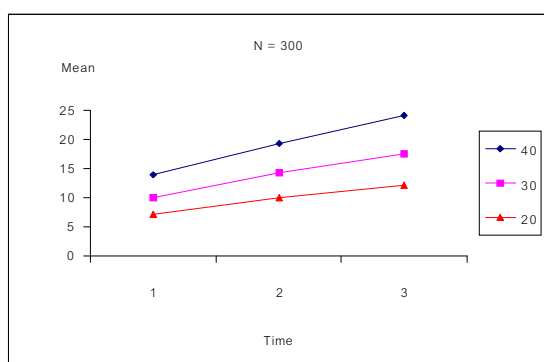
ผู้วิจัยนำค่าเฉลี่ยของคะแนนการสอบวิชาคณิตศาสตร์ของนักศึกษาการศึกษานอกโรงเรียนระดับมัธยมศึกษาตอนต้น ที่เป็นกลุ่มตัวอย่างที่ได้จากการสอบซ้ำสามครั้ง ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกันและขนาดกลุ่มตัวอย่างไม่เท่ากัน มาสร้างกราฟเพื่อพิจารณาลักษณะการแจกแจงว่าเป็นรูปแบบใด มีลักษณะสอดคล้องกับเส้นตรงหรือเส้นโค้ง ได้กราฟดังภาพประกอบ 5 (ก-ง)



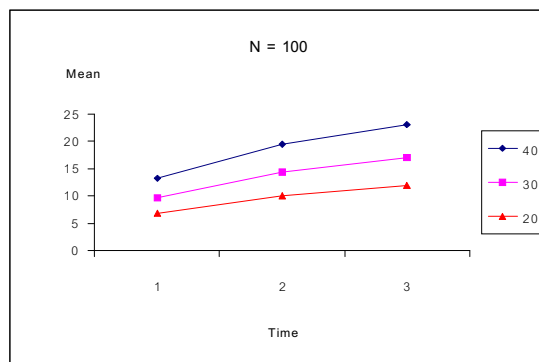
5 (ก)



5 (ข)



5 (ค)



5 (ง)

ภาพประกอบ 5 (ก-ง) แสดงความสัมพันธ์ระหว่างคะแนนเฉลี่ยกับช่วงเวลาในการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่างจำนวน 1,085 คน, 500 คน, 300 คน และ 100 คน ด้วยแบบทดสอบที่มีความยาว 40 ข้อ, 30 ข้อ และ 20 ข้อ

ผลจากการสร้างกราฟลักษณะการแจกแจงของข้อมูลตามภาพประกอบ 5 พบว่า ข้อมูลทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่างแนวโน้มมีลักษณะเป็นเส้นตรง ผู้วิจัยจึงได้ทำการทดสอบการถดถอย ด้วยโปรแกรม SPSS เพื่อตรวจสอบแบบแผนการเปลี่ยนแปลงว่ามีลักษณะเป็นเส้นตรงหรือเส้นโค้ง ปรากฏผลการวิเคราะห์ดังตาราง 10 (และรายละเอียดของผลการวิเคราะห์ดังตาราง 11 – 22 ในภาคผนวก ก)

ตาราง 10 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลง ภายใต้เงื่อนไขความยาวของแบบทดสอบต่างกันและขนาดกลุ่มตัวอย่างไม่เท่ากัน

ขนาด กลุ่มตัวอย่าง (คน)	ความยาวของ แบบทดสอบ (ข้อ)	แบบแผนการเปลี่ยนแปลง เชิงเส้นโค้ง				แบบแผนการเปลี่ยนแปลง เชิงเส้นตรง			
		R ²	SE	F	Sig	R ²	SE	F	Sig
1,085	40	0.22	8.56	464.15**	0.00	0.22	8.56	927.89**	0.00
	30	0.20	6.76	413.28**	0.00	0.20	6.76	826.56**	0.00
	20	0.21	4.49	432.90**	0.00	0.21	4.49	865.09**	0.00
500	40	0.23	8.47	227.86**	0.00	0.23	8.47	456.02**	0.00
	30	0.21	6.71	203.44**	0.00	0.21	6.71	407.16**	0.00
	20	0.22	4.46	208.17**	0.00	0.22	4.45	416.62**	0.00
300	40	0.24	8.76	141.28**	0.00	0.24	8.76	282.32**	0.00
	30	0.22	6.84	126.45**	0.00	0.22	6.84	252.44**	0.00
	20	0.23	4.53	136.98**	0.00	0.23	4.53	273.32**	0.00
100	40	0.24	7.61	47.05**	0.00	0.24	7.61	93.41**	0.00
	30	0.22	6.14	41.05**	0.00	0.22	6.13	81.50**	0.00
	20	0.23	4.12	44.50**	0.00	0.23	4.12	88.56**	0.00

** p<.01

ผลการวิเคราะห์ตามตาราง 10 พบว่า ข้อมูลทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่างมีค่าสัมประสิทธิ์การตัดสินใจ (Coefficient of Determination; R^2) ของแบบแผนการเปลี่ยนแปลงเชิงเส้นตรงและเชิงเส้นโค้งเท่ากันในแต่ละเงื่อนไข ซึ่งหมายถึงว่า สัดส่วนหรือร้อยละที่ตัวแปรต้น คือ เวลา สามารถอธิบายการเปลี่ยนแปลงของคะแนนการสอบจากแบบแผนการเปลี่ยนแปลงเชิงเส้นตรงและเชิงเส้นโค้งได้เท่ากันในแต่ละเงื่อนไข นั่นคือข้อมูลทุกเงื่อนไขมีความเหมาะสมที่จะวิเคราะห์ด้วยการถดถอยเชิงเส้นตรง ดังนั้นสมการโพลิโนเมียลมีกำลังเป็นหนึ่ง เนื่องจากผู้วิจัยใช้หลักการเลือกลำดับโพลิโนเมียลที่ต่ำสุดที่จะอธิบายความสัมพันธ์ระหว่างตัวแปรต้น (เวลา) และตัวแปรตาม (คะแนนสอบ) ได้ดีที่สุด ดังนั้นผู้วิจัยจึงเลือกที่จะทำการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง

1.3 การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนสอบแต่ละครั้ง

ผู้วิจัยนำคะแนนรวมของผู้สอบแต่ละคนทั้งสามครั้งมาวิเคราะห์โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ด้วยโปรแกรม SPSS ปรากฏผลการวิเคราะห์ดังตาราง 23

ตาราง 23 ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการสอบซ้ำสามครั้งของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน

N	K	T	40 ข้อ			30 ข้อ			20 ข้อ		
			1	2	3	1	2	3	1	2	3
1,085 คน	ข้อ	1	1.00			30	1.00		20	1.00	
		2	0.75**	1.00		ข้อ	0.73**	1.00	ข้อ	0.69**	1.00
		3	0.64**	0.76**	1.00		0.61**	0.75**	1.00	0.57**	0.71**
500 คน	ข้อ	1	1.00			30	1.00		20	1.00	
		2	0.76**	1.00		ข้อ	0.74**	1.00	ข้อ	0.70**	1.00
		3	0.65**	0.75**	1.00		0.62**	0.74**	1.00	0.56**	0.72**
300 คน	ข้อ	1	1.00			30	1.00		20	1.00	
		2	0.77**	1.00		ข้อ	0.76**	1.00	ข้อ	0.62**	1.00
		3	0.66**	0.77**	1.00		0.63**	0.76**	1.00	0.53**	0.70**
100 คน	ข้อ	1	1.00			30	1.00		20	1.00	
		2	0.68**	1.00		ข้อ	0.65**	1.00	ข้อ	0.53**	1.00
		3	0.59**	0.77**	1.00		0.55**	0.76**	1.00	0.43**	0.69**

** p<.01

ผลการวิเคราะห์ตามตาราง 23 พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบแต่ละครั้ง ในเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง มีความสัมพันธ์ทางบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 มีค่าตั้งแต่ 0.43 ถึง 0.77 โดยค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบครั้งที่ 2 กับครั้งที่ 3 มีค่าสูงสุด มีค่าตั้งแต่ 0.69 ถึง 0.77 รองลงมาคือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบครั้งที่ 1 กับครั้งที่ 2 มีค่าตั้งแต่ 0.53 ถึง 0.77 และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบครั้งที่ 1 กับครั้งที่ 3 มีค่าต่ำสุด มีค่าตั้งแต่ 0.43 ถึง 0.66 ทุกเงื่อนไขความยาวของแบบทดสอบและทุกขนาดกลุ่มตัวอย่าง

เมื่อความยาวของแบบทดสอบเพิ่มมากขึ้นในทุกขนาดกลุ่มตัวอย่างจะมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบแต่ละครั้งเพิ่มมากขึ้น และเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นในความยาวแบบทดสอบจำนวน 20 ข้อ พบว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบแต่ละครั้งเพิ่มมากขึ้นด้วยเช่นกัน สำหรับขนาดกลุ่มตัวอย่างที่เพิ่มขึ้นในความยาวของแบบทดสอบจำนวน 40 ข้อ และ 30 ข้อ มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบแต่ละครั้งเพิ่มขึ้นเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นจากจำนวน 100 คน เป็น 300 คน ส่วนขนาดกลุ่มตัวอย่าง 500 คน และ 1,085 คน มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบแต่ละครั้งค่อนข้างใกล้เคียงกัน

1.4 การวิเคราะห์ความแตกต่างของค่าเฉลี่ยของคะแนนสอบทั้งสามครั้ง

ผู้วิจัยทดสอบความแตกต่างของค่าเฉลี่ยของคะแนนสอบทั้งสามครั้ง โดยวิเคราะห์ความแปรปรวนทางเดียวแบบวัดซ้ำ (Repeated measure analysis of variance) เพื่อทดสอบความแตกต่างระหว่างช่วงเวลาการสอบซ้ำ ปรากฏผลการวิเคราะห์ดังตาราง 24

ตาราง 24 การวิเคราะห์ความแตกต่างของค่าเฉลี่ยของคะแนนสอบซ้ำสามครั้ง ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน

N	K	Source	SS	df	MS	F	P
1085 คน	40 ข้อ	Between (time)	51418.73	2	25709.37	1059.16**	0.00
		Within	52624.60	2168	24.27		
	30 ข้อ	Between (time)	28118.73	2	14059.37	883.77**	0.00
		Within	34489.27	2168	15.91		
	20 ข้อ	Between (time)	13068.94	2	6534.47	845.62**	0.00
		Within	16753.07	2168	7.73		
500 คน	40 ข้อ	Between (time)	23384.27	2	11692.13	493.07**	0.00
		Within	23665.74	998	23.71		
	30 ข้อ	Between (time)	12926.17	2	6463.09	413.99**	0.00
		Within	15580.50	998	15.61		
	20 ข้อ	Between (time)	5853.17	2	2926.59	382.59**	0.00
		Within	7634.16	998	7.65		
300 คน	40 ข้อ	Between (time)	15736.04	2	7868.02	328.46**	0.00
		Within	14324.62	598	23.95		
	30 ข้อ	Between (time)	8428.67	2	4214.33	275.71**	0.00
		Within	9140.66	598	15.29		
	20 ข้อ	Between (time)	3978.93	2	1989.46	263.15**	0.00
		Within	4521.07	598	7.56		
100 คน	40 ข้อ	Between (time)	4926.65	2	2463.32	124.34**	0.00
		Within	3922.69	198	19.81		
	30 ข้อ	Between (time)	2816.49	2	1408.24	101.46**	0.00
		Within	2748.18	198	13.88		
	20 ข้อ	Between (time)	1325.31	2	662.65	83.39**	0.00
		Within	1573.36	198	7.95		

** p<.01

ผลการวิเคราะห์ตามตาราง 22 พบว่า ค่าสถิติทดสอบเอฟ (F-test) ที่ใช้ทดสอบความแตกต่างระหว่างช่วงเวลาการสอบมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงให้เห็นว่า ช่วงเวลาในการวัดแต่ละครั้งมีผลทำให้คะแนนสอบคณิตศาสตร์ของนักศึกษาแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง จึงได้ทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffe') ปรากฏผลการวิเคราะห์ดังตาราง 25

ตาราง 25 การเปรียบเทียบค่าเฉลี่ยรายคู่ของคะแนนสอบซ้ำสามครั้ง ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน

N	Compair	40 ข้อ			30 ข้อ			20 ข้อ		
		MD	S.E.	Sig.	MD	S.E.	Sig.	MD	S.E.	Sig.
1085 คน	1 กับ 2	5.44	0.19	0.00	4.07	0.16	0.00	2.87	0.11	0.00
	2 กับ 3	4.28	0.21	0.00	3.11	0.17	0.00	2.01	0.12	0.00
	1 กับ 3	9.71	0.23	0.00	7.18	0.19	0.00	4.88	0.13	0.00
500 คน	1 กับ 2	5.33	0.28	0.00	4.06	0.23	0.00	2.72	0.16	0.00
	2 กับ 3	4.32	0.31	0.00	3.11	0.25	0.00	2.11	0.17	0.00
	1 กับ 3	9.65	0.34	0.00	7.17	0.28	0.00	4.83	0.20	0.00
300 คน	1 กับ 2	5.49	0.37	0.00	4.12	0.29	0.00	2.89	0.21	0.00
	2 กับ 3	4.74	0.39	0.00	3.36	0.31	0.00	2.24	0.22	0.00
	1 กับ 3	10.23	0.44	0.00	7.48	0.35	0.00	5.14	0.24	0.00
100 คน	1 กับ 2	6.11	0.61	0.00	4.83	0.53	0.00	3.16	0.40	0.00
	2 กับ 3	3.72	0.58	0.00	2.56	0.48	0.00	1.94	0.36	0.00
	1 กับ 3	9.83	0.69	0.00	7.39	0.57	0.00	5.10	0.44	0.00

ผลการวิเคราะห์ตามตาราง 25 พบว่า ค่าเฉลี่ยของคะแนนจากการสอบครั้งที่ 1 มีค่าต่ำกว่าค่าเฉลี่ยของคะแนนจากการสอบครั้งที่ 2 และครั้งที่ 3 และค่าเฉลี่ยของคะแนนจากการสอบครั้งที่ 2 มีค่าต่ำกว่าค่าเฉลี่ยของคะแนนจากการสอบครั้งที่ 3 อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง

ตอนที่ 2 การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลง

ผู้วิจัยได้แบ่งการวิเคราะห์ในตอนนี้ออกเป็น 2 ขั้นตอน คือ 1) การตรวจสอบความสอดคล้องของข้อมูลเชิงประจักษ์กับโมเดลพหุระดับ และ 2) การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลง

โดยสุ่มผลการตอบของผู้สอบตามเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง แล้วนำข้อมูลที่ได้ทั้งหมดจำนวน 12 ชุด มาวิเคราะห์ โดยมีรายละเอียดดังนี้

2.1 ตรวจสอบความสอดคล้องของข้อมูลเชิงประจักษ์กับโมเดลพหุระดับ

การวิเคราะห์โมเดลพหุระดับด้วยโปรแกรม HLM นี้ ผู้วิจัยได้สร้างและทำการตรวจสอบโมเดลตามประเภทของตัวแปรตามซึ่งขึ้นอยู่กับตัวแปรต้น ในที่นี้คือ ตัวแปรเวลา จำนวน 3 แบบ คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ปรากฏผลการวิเคราะห์ดังตาราง 26-27 (และรายละเอียดการวิเคราะห์ปรากฏดังตาราง 28-39 ในภาคผนวก ก)

ตาราง 26 ค่าสัมประสิทธิ์การทำนาย (R^2) ระดับบุคคลของโมเดลพหุระดับสามแบบ ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

N	K	uncentered	group centered	grand centered
		model	model	model
1,085 คน	40 ข้อ	0.61	0.61	0.61
	30 ข้อ	0.56	0.56	0.56
	20 ข้อ	0.55	0.55	0.55
500 คน	40 ข้อ	0.61	0.61	0.61
	30 ข้อ	0.56	0.56	0.56
	20 ข้อ	0.57	0.57	0.57
300 คน	40 ข้อ	0.63	0.63	0.63
	30 ข้อ	0.58	0.58	0.58
	20 ข้อ	0.56	0.56	0.56
100 คน	40 ข้อ	0.61	0.61	0.61
	30 ข้อ	0.56	0.56	0.56
	20 ข้อ	0.54	0.54	0.54

ผลการวิเคราะห์ตามตาราง 26 พบว่า การวิเคราะห์การเปลี่ยนแปลงของคะแนนคณิตศาสตร์ของนักศึกษาการศึกษานอกโรงเรียน ระดับมัธยมศึกษาตอนต้น โดยใช้โมเดลพัฒนาการเชิงเส้นตรง (Linear growth model) แบบสองระดับ ด้วยสามโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ของทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง มีประสิทธิภาพของโมเดลเท่ากัน แสดงว่า โมเดลทั้งสามสามารถอธิบายการเปลี่ยนแปลงของคะแนนคณิตศาสตร์ของนักศึกษาการศึกษานอกโรงเรียน ระดับมัธยมศึกษาตอนต้นได้เหมาะสมเท่ากัน คือ ในกลุ่มตัวอย่างจำนวน 1,085 คน แบบทดสอบจำนวน 40 ข้อ, 30 ข้อ และ 20 ข้อ มีค่าเท่ากับ 0.61, 0.56 และ 0.55 ตามลำดับ กลุ่มตัวอย่าง 500 คน แบบทดสอบจำนวน 40 ข้อ, 30 ข้อ และ 20 ข้อ มีค่าเท่ากับ 0.61, 0.56 และ 0.57 ตามลำดับ กลุ่มตัวอย่าง 300 คน แบบทดสอบจำนวน 40 ข้อ, 30 ข้อ และ 20 ข้อ มีค่าเท่ากับ 0.63, 0.58 และ 0.56 ตามลำดับ และกลุ่มตัวอย่าง 100 คน แบบทดสอบจำนวน 40 ข้อ, 30 ข้อ และ 20 ข้อ มีค่าเท่ากับ 0.61, 0.56 และ 0.54 ตามลำดับ

ตาราง 27 ผลการทดสอบอิทธิพลคงที่และอิทธิพลสุ่มของโมเดลพหุระดับสามแบบ ภายใต้เงื่อนไข
ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

N	K	Para- meter	Uncentered model				Group centered model / Grand centered model			
			Coefficient	t-ratio	Variance Component	χ^2	Coefficient	t-ratio	Variance Component	χ^2
1,085 คน	40	Initial	13.97	64.25**	35.48	3520.39**	18.82	78.42**	56.18	10729.94**
		Change	4.86	41.46**	5.41	1703.53**	4.86	41.46**	5.41	1703.53**
		Error			18.94				18.94	
	30	Initial	10.12	59.30**	20.95	3215.07**	13.71	73.19**	33.81	9683.68**
		Change	3.59	38.16**	3.20	1626.54**	3.59	38.16**	3.20	1626.54**
		Error			12.79				12.79	
	20	Initial	7.17	61.94**	9.37	3048.01**	9.61	78.20**	14.33	8589.54**
		Change	2.44	37.28**	1.55	1625.42**	2.44	37.28**	1.55	1625.42**
		Error			6.21				6.21	
500 คน	40	Initial	14.11	43.80**	36.45	1680.05**	18.93	53.60**	56.22	5053.40**
		Change	4.83	28.35**	5.25	782.38**	4.83	28.35**	5.25	782.38**
		Error			18.48				18.48	
	30	Initial	10.22	40.15**	21.98	1554.07**	13.80	49.95**	34.02	4581.85**
		Change	3.59	26.10**	3.20	754.93**	3.59	26.10**	3.20	754.93**
		Error			12.47				12.47	
	20	Initial	7.27	42.28**	9.99	1541.94**	9.68	53.46**	14.48	4277.92**
		Change	2.41	24.79**	1.87	823.94**	2.41	24.79**	1.87	823.94**
		Error			5.74				5.74	
300 คน	40	Initial	14.00	32.02**	41.87	1106.28**	19.12	40.30**	61.35	3255.72**
		Change	5.12	23.25**	5.23	466.92**	5.12	23.25**	5.23	466.92**
		Error			18.61				18.61	
	30	Initial	10.21	30.18**	24.19	1011.01**	13.95	38.03**	36.32	2971.58**
		Change	3.74	21.31**	3.15	453.32**	3.74	21.31**	3.15	453.32**
		Error			12.19				12.19	
	20	Initial	7.17	31.49**	10.36	896.40**	9.74	40.59**	15.19	2488.82**
		Change	2.57	21.03**	1.36	430.06**	2.57	21.03**	1.36	430.06**
		Error			6.22				6.22	
100 คน	40	Initial	13.78	21.88**	25.39	272.57**	18.69	27.30**	41.07	801.05**
		Change	4.92	14.22**	3.27	136.23**	4.92	14.22**	3.27	136.23**
		Error			17.38				17.38	
	30	Initial	9.95	19.69**	15.39	249.04**	13.64	25.20**	25.25	714.53**
		Change	3.70	12.93**	2.07	132.65**	3.70	12.93**	2.07	132.65**
		Error			12.18				12.18	
	20	Initial	7.03	20.54**	6.25	212.52**	9.58	27.17**	10.26	564.94**
		Change	2.56	11.97**	1.29	137.94**	2.56	11.97**	1.29	137.94**
		Error			6.54				6.54	

** p<.01

ผลการวิเคราะห์ตามตาราง 27 สามารถสรุปได้ ดังนี้

1. กลุ่มตัวอย่าง 1,085 คน สอบด้วยแบบทดสอบจำนวน 40 ข้อ ผลการทดสอบอิทธิพลคงที่ พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 13.97 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 18.82 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 4.86 ผลการทดสอบอิทธิพลสุ่ม พบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 35.48 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 56.18 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 5.41 และ 18.94 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

2. กลุ่มตัวอย่าง 1,085 คน สอบด้วยแบบทดสอบจำนวน 30 ข้อ ผลการทดสอบอิทธิพลคงที่ พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 10.12 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 13.71 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 3.59 ผลการทดสอบอิทธิพลสุ่ม พบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 20.95 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 33.81 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 3.20 และ 12.79 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

3. กลุ่มตัวอย่าง 1,085 คน สอบด้วยแบบทดสอบจำนวน 20 ข้อ ผลการทดสอบอิทธิพลคงที่ พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 7.17 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 9.61 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 2.44 ผลการทดสอบอิทธิพลสุ่มพบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 9.37 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 14.33 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 1.55 และ 6.21 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

4. กลุ่มตัวอย่าง 500 คน สอบด้วยแบบทดสอบจำนวน 40 ข้อ ผลการทดสอบอิทธิพลคงที่ พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 14.11 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 18.93 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 4.83 ผลการทดสอบอิทธิพลสุ่มพบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 36.45 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 56.22 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 5.25 และ 18.48 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

5. กลุ่มตัวอย่าง 500 คน สอบด้วยแบบทดสอบจำนวน 30 ข้อ ผลการทดสอบอิทธิพลคงที่ พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 10.22 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชัน

ที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 13.80 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 3.59 ผลการทดสอบอิทธิพลสุ่มพบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 21.98 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 34.02 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 3.20 และ 12.47 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

6. กลุ่มตัวอย่าง 500 คน สอดด้วยแบบทดสอบจำนวน 20 ข้อ ผลการทดสอบอิทธิพลคงที่พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 7.27 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 9.68 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 2.41 ผลการทดสอบอิทธิพลสุ่มพบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 9.99 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 14.48 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 1.87 และ 5.74 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

7. กลุ่มตัวอย่าง 300 คน สอดด้วยแบบทดสอบจำนวน 40 ข้อ ผลการทดสอบอิทธิพลคงที่พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 14.00 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 19.12 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 5.12 ผลการทดสอบอิทธิพลสุ่มพบว่า

8. กลุ่มตัวอย่าง 300 คน สอบด้วยแบบทดสอบจำนวน 30 ข้อ ผลการทดสอบอภิปิพดคทที่พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 10.21 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 13.95 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 3.74 ผลการทดสอบอภิปิพดลุ่มพบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 24.19 ซึ่งมีความน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 36.32 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 3.15 และ 12.19 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

9. กลุ่มตัวอย่าง 300 คน สอบด้วยแบบทดสอบจำนวน 20 ข้อ ผลการทดสอบอิทธิพลคงที่พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 7.17 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 9.74 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากันคือ มีค่าเท่ากับ 2.57 ผลการทดสอบอิทธิพลสุ่มพบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 10.36 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชัน

ที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือมีค่าเท่ากับ 15.19 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือมีค่าเท่ากับ 1.36 และ 6.22 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

10. กลุ่มตัวอย่าง 100 คน สอบด้วยแบบทดสอบจำนวน 40 ข้อ ผลการทดสอบอิทธิพลคงที่พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 13.78 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 18.69 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 4.92 ผลการทดสอบอิทธิพลสุ่มพบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 25.39 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 41.07 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 3.27 และ 17.38 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

11. กลุ่มตัวอย่าง 100 คน สอบด้วยแบบทดสอบจำนวน 30 ข้อ ผลการทดสอบอิทธิพลคงที่พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 9.95 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 13.64 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 3.70 ผลการทดสอบอิทธิพลสุ่มพบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 15.39 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 25.25 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของ

ความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 2.07 และ 12.18 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

12. กลุ่มตัวอย่าง 100 คน สอบด้วยแบบทดสอบจำนวน 20 ข้อ ผลการทดสอบอิทธิพลคงที่ พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 7.03 ซึ่งมีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) มีคะแนนเริ่มต้นเฉลี่ยเท่ากัน คือ 9.58 ส่วนคะแนนการเปลี่ยนแปลงเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากันคือ มีค่าเท่ากับ 2.56 ผลการทดสอบอิทธิพลสุ่มพบว่า ความแปรปรวนของคะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าเท่ากับ 6.25 ซึ่งมีค่าน้อยกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ซึ่งมีค่าความแปรปรวนของคะแนนเริ่มต้นเท่ากัน คือ มีค่าเท่ากับ 10.26 ความแปรปรวนของคะแนนการเปลี่ยนแปลงและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน คือ มีค่าเท่ากับ 1.29 และ 6.54 ตามลำดับ ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

ผลการวิเคราะห์ข้อมูลดังกล่าวข้างต้น จึงทำให้ผู้วิจัยเลือกที่จะนำคะแนนเริ่มต้นและอัตราพัฒนาการที่ได้จากการวิเคราะห์คะแนนการเปลี่ยนแปลงด้วยโมเดลพหุระดับ แบบสองระดับ ที่ใช้โมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มาใช้ในการศึกษาตามวัตถุประสงค์ของการวิจัยในครั้งนี้ ด้วยเหตุผลว่าโมเดลทั้งสาม คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model), โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยกลุ่มของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) สามารถอธิบายคะแนนการเปลี่ยนแปลงที่เกิดเวลาของข้อมูลทั้ง 12 ชุด ได้เหมาะสมเท่ากัน ซึ่งโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) ประมาณค่าคะแนนเริ่มต้นจากคะแนนที่เส้นการทำนายเฉลี่ยตัดแกนตั้ง ซึ่งเป็นคะแนน ณ ตำแหน่งเวลาสอบครั้งแรกหรือเริ่มต้น คือ t เท่ากับ 0

2.2 การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงแต่ละวิชา

การวิเคราะห์ข้อมูลในตอนนี้ ผู้วิจัยนำคะแนนจากการสอบครั้งที่หนึ่งและครั้งที่สามของนักศึกษาแต่ละคนมาวิเคราะห์คะแนนการเปลี่ยนแปลงด้วยวิธีการวัดการเปลี่ยนแปลงที่สนใจศึกษาจำนวนห้าวิชา

คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และ การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ โดยมีคะแนนจากการสอบทั้งสาม ครั้งของนักศึกษาแต่ละคนมาวิเคราะห์คะแนนการเปลี่ยนแปลงด้วยวิธีโมเดลพหุระดับ ซึ่งเป็นวิธีการประมาณค่าคะแนนการเปลี่ยนแปลงที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ ต่อจากนั้นผู้วิจัยได้คำนวณค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธี ซึ่งเป็นการตรวจสอบลักษณะโดยทั่วไปของข้อมูล ได้แก่ คะแนนเฉลี่ย ความคลาดเคลื่อนมาตรฐานของคะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด คะแนนต่ำสุด มัธยฐาน ฐานนิยม ความเบ้ และความโด่ง โดยจำแนกข้อมูลออกตามความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ปรากฏผลการวิเคราะห์ ดังตาราง 40 (และรายละเอียดของผลการวิเคราะห์ดังตาราง 41 ในภาคผนวก ก)

ตาราง 40 ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี

N	วิธีวัด	40 ข้อ				30 ข้อ				20 ข้อ			
		$\bar{X}$	S	Sk	Ku	$\bar{X}$	S	Sk	Ku	$\bar{X}$	S	Sk	Ku
1,085 คน	D	9.71	7.71	0.33	-0.62	7.18	6.19	0.30	-0.64	4.88	4.32	0.20	-0.48
	S	40.58	33.07	0.17	-0.82	39.13	34.57	0.13	-0.82	40.89	36.13	0.04	-0.92
	d_0	0	0.87	0.13	-0.31	0	0.85	0.21	-0.23	0	0.87	0.26	-0.18
	D_0	9.05	14.05	1.21	1.07	9.70	15.95	1.47	1.75	7.68	13.94	1.47	1.88
	RD_0	13.56	22.51	0.72	2.20	13.89	24.15	0.78	2.49	11.38	21.88	0.69	4.32
	HLM	4.86	1.82	0.24	-1.20	3.59	1.41	0.32	-1.17	2.44	0.92	0.12	-1.18
500 คน	D	9.65	7.61	0.32	-0.65	7.17	6.14	0.32	-0.60	4.83	4.35	0.24	-0.46
	S	40.53	32.99	0.09	-0.52	39.13	34.53	0.07	-0.56	40.66	36.42	0.06	-0.92
	d_0	0	0.84	0.21	-0.37	0	0.83	0.30	-0.22	0.00	0.87	0.36	-0.36
	D_0	8.50	13.78	1.30	1.81	8.13	14.38	1.46	1.59	7.63	13.70	1.54	1.76
	RD_0	12.67	22.22	0.48	3.46	11.66	23.05	0.39	4.63	11.57	21.31	1.25	1.28
	HLM	4.83	1.77	0.21	-1.17	3.59	1.37	0.25	-1.15	2.41	1.00	0.05	-1.06
300 คน	D	10.23	7.62	0.26	-0.71	7.48	6.07	0.21	-0.84	5.14	4.23	0.07	-0.49
	S	43.51	32.85	0.16	-1.12	41.63	34.04	0.14	-1.10	43.37	35.92	-0.08	-0.75
	d_0	0.00	0.81	0.12	-0.54	0.00	0.81	0.22	-0.67	0.00	0.84	0.25	-0.55
	D_0	8.61	13.59	1.13	0.86	9.08	15.10	1.36	1.17	7.44	13.73	1.35	1.61
	RD_0	13.44	22.41	0.97	0.94	13.38	24.13	0.75	2.21	11.29	21.65	0.97	1.41
	HLM	5.12	1.72	0.06	-1.29	3.74	1.35	0.16	-1.31	2.57	0.85	0.04	-1.21

ตาราง 40 (ต่อ)

N	วิธีวัด	40 ข้อ				30 ข้อ				20 ข้อ			
		$\bar{X}$	S	Sk	Ku	$\bar{X}$	S	Sk	Ku	$\bar{X}$	S	Sk	Ku
100	D	9.83	6.92	0.37	-0.44	7.39	5.71	0.31	-0.41	5.11	4.27	0.01	-0.84
คน	S	38.55	28.10	0.30	-0.43	37.53	30.44	0.18	-0.12	39.74	33.81	-0.12	-0.73
	d_0	0.00	0.92	0.14	-0.19	0.00	0.97	-0.02	0.22	0.00	1.01	-0.04	-0.49
	D_0	8.18	13.66	1.36	1.77	9.01	15.60	1.42	1.64	7.87	13.47	1.51	1.58
	RD_0	11.95	21.41	1.17	1.72	12.75	23.54	1.26	1.67	11.54	20.03	1.52	1.62
	HLM	4.92	1.45	0.41	-0.78	3.70	1.15	0.48	-0.79	2.56	0.83	0.20	-0.99

ผลการวิเคราะห์ตามตาราง 40 สามารถสรุปได้ ดังนี้

2.2.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 40 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 9.71, 9.00 และ 5.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 40.58, 36.00 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -1.00 และ 0.38 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 9.05, 5.40 และ 34.27 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 13.56, 7.74 และ 68.60 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 4.86, 4.65 และ 3.17 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.23 และ 7.71 ตามลำดับ วิธี S มีค่าเท่ากับ 1.00 และ 33.08 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0 และ 0.87 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 0.43 และ 14.05 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 0.68 และ 22.51 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.06 และ 1.82 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากทุกวิธีมีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D, S, d_0 , D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.33, 0.17, 0.13, 1.21, 0.72 และ 0.24 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D, S, d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.62, -0.82, -0.31 และ -1.20 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 1.07 และ 2.20 ตามลำดับ

2.2.2 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 30 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 7.18, 7.00 และ 4.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 39.13, 34.78 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -0.01 และ 0.18 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 9.70, 3.06 และ 45.43 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 13.89, 4.20 และ 74.83 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 3.59, 3.38 และ 6.15 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.19 และ 6.19 ตามลำดับ วิธี S มีค่าเท่ากับ 1.05 และ 34.57 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.03 และ 0.85 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 0.48 และ 15.95 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 0.73 และ 24.15 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.04 และ 1.41 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากทุกวิธีมีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D, S, d_0 , D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.30, 0.13, 0.21, 1.47, 0.78 และ 0.32 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D, S, d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.64, -0.82, -0.23 และ -1.17 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 1.75 และ 2.49 ตามลำดับ

2.2.3 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 20 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 4.88, 5.00 และ 6.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 40.89, 37.50 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -0.08 และ -0.60 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 7.68, 3.15 และ 6.97 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 11.38, 4.42 และ 24.11 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 2.44, 2.34 และ 1.77 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.13 และ 4.32 ตามลำดับ วิธี S มีค่าเท่ากับ 1.10 และ 36.13 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.03 และ

0.87 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 0.42 และ 13.94 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 0.66 และ 21.88 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.03 และ 0.92 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากทุกวิธีมีความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D , S , d_0 , D_0 , RD_0 และ HLM มีความเบ้เท่ากับ 0.20, 0.04, 0.26, 1.47, 0.69 และ 0.12 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D , S , d_0 และ HLM มีความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีความโด่งเท่ากับ -0.48, -0.92, -0.18 และ -1.18 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีความโด่งเท่ากับ 1.88 และ 4.32 ตามลำดับ

2.2.4 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 40 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 9.65, 8.50 และ 5.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 40.53, 36.36 และ 0 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -0.03 และ 0.38 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 8.50, 5.21 และ -2.91 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 12.67, 7.25 และ 68.85 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 4.83, 4.63 และ 2.60 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.34 และ 7.61 ตามลำดับ วิธี S มีค่าเท่ากับ 1.48 และ 32.99 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.04 และ 0.84 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 0.62 และ 13.78 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 0.99 และ 22.22 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.08 และ 1.77 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากทุกวิธีมีความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D , S , d_0 , D_0 , RD_0 และ HLM มีความเบ้เท่ากับ 0.32, 0.09, 0.21, 1.30, 0.48 และ 0.21 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D , S , d_0 และ HLM มีความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีความโด่งเท่ากับ -0.65, -0.52, -0.37 และ -1.17 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีความโด่งเท่ากับ 1.81 และ 3.46 ตามลำดับ

2.2.5 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 30 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 7.17, 7.00 และ 4.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 39.13, 36.00 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -0.09 และ 0.21 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 8.13, 1.03 และ -0.77 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 11.66, 1.43 และ 68.61 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 3.59, 3.39 และ 1.96 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.27 และ 6.14 ตามลำดับ วิธี S มีค่าเท่ากับ 1.54 และ 34.53 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.04 และ 0.83 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 0.64 และ 14.38 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 1.03 และ 23.05 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.06 และ 1.37 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากทุกวิธีมีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D, S, d_0 , D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.32, 0.07, 0.30, 1.46, 0.39 และ 0.25 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D, S, d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.60, -0.56, -0.22 และ -1.15 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 1.59 และ 4.63 ตามลำดับ

2.2.6 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 20 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 4.83, 4.50 และ 2.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 40.66, 36.93 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -0.10 และ -0.64 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 7.63, 2.87 และ 14.37 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 11.57, 3.95 และ 39.42 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 2.41, 2.37 และ 2.37 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.19 และ 4.35 ตามลำดับ วิธี S มีค่าเท่ากับ 1.63 และ 36.42 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.04 และ

0.87 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 0.61 และ 13.70 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 0.95 และ 21.31 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.04 และ 1.00 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากทุกวิธีมีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D , S , d_0 , D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.24, 0.06, 0.36, 1.54, 1.25 และ 0.05 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D , S , d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.46, -0.92, -0.36 และ -1.06 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 1.76 และ 1.28 ตามลำดับ

2.2.7 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 40 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 10.23, 9.00 และ 2.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 43.51, 41.90 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, 0 และ 0.19 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 8.61, 4.89 และ -2.80 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 13.44, 7.00 และ 69.98 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 5.12, 5.16 และ 3.00 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.44 และ 7.62 ตามลำดับ วิธี S มีค่าเท่ากับ 1.90 และ 32.85 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.05 และ 0.81 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 0.79 และ 13.59 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 1.29 และ 22.41 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.10 และ 1.72 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากทุกวิธีมีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D , S , d_0 , D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.26, 0.16, 0.12, 1.13, 0.97 และ 0.06 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D , S , d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.71, -1.12, -0.54 และ -1.29 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 0.86 และ 0.94 ตามลำดับ

2.2.8 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 30 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 7.48, 7.00 และ 2.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 41.63, 38.10 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -0.05 และ 0.12 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 9.08, 3.26 และ 1.07 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 13.38, 4.48 และ 75.08 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 3.74, 3.64 และ 1.94 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.35 และ 6.07 ตามลำดับ วิธี S มีค่าเท่ากับ 1.97 และ 34.04 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.05 และ 0.81 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 0.87 และ 15.10 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 1.39 และ 24.13 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.08 และ 1.35 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากทุกวิธีมีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D, S, d_0 , D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.21, 0.14, 0.22, 1.36, 0.75 และ 0.16 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D, S, d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.84, -1.10, -0.67 และ -1.31 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 1.17 และ 2.21 ตามลำดับ

2.2.9 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 20 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 5.14, 5.00 และ 6.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 43.37, 42.86 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, 0.06 และ -0.76 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 7.44, 3.29 และ 19.35 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 11.29, 4.60 และ 45.83 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 2.57, 2.55 และ 3.83 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.24 และ 4.23 ตามลำดับ วิธี S มีค่าเท่ากับ 2.07 และ 35.92 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.05 และ

0.84 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 0.79 และ 13.73 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 1.25 และ 21.65 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.05 และ 0.85 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี ส่วนใหญ่มีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D , d_0 , D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.07, 0.25, 1.35, 0.97 และ 0.04 ตามลำดับ ส่วนวิธี S มีค่าความเบ้เป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ซ้าย โดยมีค่าความเบ้เท่ากับ -0.08 สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D , S , d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.49, -0.75, -0.55 และ -1.21 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 1.61 และ 1.41 ตามลำดับ

2.2.10 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 40 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 9.83, 9.50 และ 5.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 38.55, 40.00 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -0.14 และ -2.37 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 8.18, 3.93 และ -1.44 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 11.95, 5.42 และ -40.61 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 4.92, 4.84 และ 3.15 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.69 และ 6.92 ตามลำดับ วิธี S มีค่าเท่ากับ 2.81 และ 28.10 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0 และ 0.92 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 1.37 และ 13.66 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 2.14 และ 21.41 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.14 และ 1.45 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากทุกวิธีมีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D , S , d_0 , D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.37, 0.30, 0.14, 1.36, 1.17 และ 0.41 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D , S , d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.44, -0.43, -0.19 และ -0.78 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมี

ลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 1.77 และ 1.72 ตามลำดับ

2.2.11 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 30 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 7.39, 7.00 และ 11.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 37.53, 37.17 และ 100.00 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -0.17 และ -2.97 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 9.01, 2.69 และ -19.73 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 12.75, 3.64 และ -47.49 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 3.70, 3.62 และ 2.35 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.57 และ 5.71 ตามลำดับ วิธี S มีค่าเท่ากับ 3.04 และ 30.44 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.10 และ 0.97 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 1.56 และ 15.60 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 2.35 และ 23.54 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.12 และ 1.15 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี ส่วนใหญ่มีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D, S, D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.31, 0.18, 1.42, 1.26 และ 0.48 ตามลำดับ ส่วนวิธี d_0 มีค่าความเบ้เป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ซ้าย โดยมีค่าความเบ้เท่ากับ -0.02 สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D, S, d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.41, -0.12, -0.22 และ -0.79 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 1.64 และ 1.67 ตามลำดับ

2.2.12 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 20 ข้อ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี มีค่ากลาง คือ คะแนนเฉลี่ย ค่ามัธยฐาน และฐานนิยม ดังนี้ วิธี D มีค่ากลางเท่ากับ 5.11, 5.50 และ 7.00 ตามลำดับ วิธี S มีค่ากลางเท่ากับ 39.74, 43.30 และ 0 ตามลำดับ วิธี d_0 มีค่ากลางเท่ากับ 0, -0.17 และ -2.33 ตามลำดับ วิธี D_0 มีค่ากลางเท่ากับ 7.87, 1.68 และ -10.19 ตามลำดับ วิธี RD_0 มีค่ากลางเท่ากับ 11.54, 2.33 และ -20.41 ตามลำดับ และวิธี HLM มีค่ากลางเท่ากับ 2.56, 2.63 และ 1.83 ตามลำดับ

เมื่อพิจารณาการกระจายของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีมีค่าความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน ดังนี้ วิธี D มีค่าเท่ากับ 0.43 และ 4.27 ตามลำดับ วิธี S มีค่าเท่ากับ 3.38 และ 33.81 ตามลำดับ วิธี d_0 มีค่าเท่ากับ 0.10 และ 1.01 ตามลำดับ วิธี D_0 มีค่าเท่ากับ 1.35 และ 13.47 ตามลำดับ วิธี RD_0 มีค่าเท่ากับ 2.00 และ 20.03 ตามลำดับ และวิธี HLM มีค่าเท่ากับ 0.08 และ 0.83 ตามลำดับ

เมื่อพิจารณาลักษณะของข้อมูล พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีส่วนใหญ่มีค่าความเบ้เป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ขวา ดังนี้ วิธี D, D_0 , RD_0 และ HLM มีค่าความเบ้เท่ากับ 0.01, 1.51, 1.52 และ 0.20 ตามลำดับ ส่วนวิธี S และ d_0 มีค่าความเบ้เป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงแบบเบ้ซ้าย โดยมีค่าความเบ้เท่ากับ -0.12 และ -0.04 ตามลำดับ สำหรับความโด่งของข้อมูลนั้น พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D, S, d_0 และ HLM มีค่าความโด่งเป็นลบ นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ -0.84, -0.73, -0.49 และ -0.99 ตามลำดับ ส่วนวิธี D_0 และ RD_0 ที่มีค่าความโด่งเป็นบวก นั่นคือ ข้อมูลมีลักษณะการแจกแจงที่มียอดสูงหรือโด่งมากกว่าโค้งปกติ โดยมีค่าความโด่งเท่ากับ 1.58 และ 1.62 ตามลำดับ

ผู้วิจัยได้ตรวจสอบลักษณะการแจกแจงของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกันและขนาดกลุ่มตัวอย่างไม่เท่ากัน ว่ามีลักษณะการแจกแจงเป็นแบบโค้งปกติหรือไม่ โดยการเปรียบเทียบความน่าจะเป็นสะสมของคะแนนการเปลี่ยนแปลงกับการแจกแจงแบบโค้งปกติ ด้วยการนำข้อมูลดังกล่าวมาทดสอบโดยใช้สถิติ Kolmogorov–Smirnov Test ปรากฏผลการวิเคราะห์ดังตาราง 42

ตาราง 42 ค่าสถิติ Kolmogorov-Smirnov ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี

N	วิธีวัด	40 ชั่วโมง			30 ชั่วโมง			20 ชั่วโมง		
		MED	K-S	p	MED	K-S	p	MED	K-S	p
1,085 คน	D	0.09	2.87**	0.00	0.09	2.90**	0.00	0.08	2.54**	0.00
	S	0.09	3.09**	0.00	0.09	2.96**	0.00	0.06	2.06**	0.00
	d_{θ}	0.04	1.19	0.12	0.05	1.58	0.02	0.05	1.75**	0.00
	D_{θ}	0.17	5.48**	0.00	0.22	7.27**	0.00	0.23	7.42**	0.00
	RD_{θ}	0.19	6.09**	0.00	0.22	7.30**	0.00	0.22	7.31**	0.00
	HLM	0.11	3.49**	0.00	0.11	3.65**	0.00	0.08	2.60**	0.00
500 คน	D	0.09	1.92**	0.00	0.09	1.92**	0.00	0.08	1.76**	0.00
	S	0.09	2.04**	0.00	0.09	1.94**	0.00	0.07	1.47	0.03
	d_{θ}	0.04	0.99	0.29	0.06	1.37	0.05	0.07	1.45	0.03
	D_{θ}	0.16	3.63**	0.00	0.20	4.42**	0.00	0.22	4.83**	0.00
	RD_{θ}	0.18	4.11**	0.00	0.19	4.30**	0.00	0.23	5.02**	0.00
	HLM	0.99	2.19**	0.00	0.11	2.35**	0.00	0.07	1.62	0.01
300 คน	D	0.08	1.36	0.05	0.09	1.51	0.02	0.07	1.23	0.10
	S	0.09	1.52	0.02	0.09	1.58	0.01	0.06	1.00	0.28
	d_{θ}	0.05	0.83	0.50	0.06	0.95	0.33	0.05	0.92	0.37
	D_{θ}	0.17	2.90**	0.00	0.20	3.42**	0.00	0.19	3.27**	0.00
	RD_{θ}	0.20	3.54**	0.00	0.20	3.53**	0.00	0.21	3.59**	0.00
	HLM	0.10	1.70**	0.01	0.12	2.05**	0.00	0.08	1.41**	0.04
100 คน	D	0.11	1.10	0.18	0.08	0.84	0.49	0.10	1.01	0.26
	S	0.11	1.07	0.20	0.08	0.78	0.57	0.08	0.79	0.56
	d_{θ}	0.10	0.96	0.32	0.09	0.86	0.46	0.10	1.00	0.27
	D_{θ}	0.17	1.71**	0.02	0.25	2.47**	0.00	0.21	2.01**	0.00
	RD_{θ}	0.19	1.94**	0.00	0.25	2.51**	0.00	0.21	2.09**	0.00
	HLM	0.11	1.13	0.15	0.13	1.30	0.07	0.12	1.20	0.12

**p<.01

ผลการวิเคราะห์ตามตาราง 42 พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D และ S แนวโน้มมีลักษณะการแจกแจงคล้ายคลึงกันทุกเงื่อนไขของความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง โดยทุกเงื่อนไขความยาวของแบบทดสอบของกลุ่มตัวอย่างขนาด 1,085 คน และ 500 คน แนวโน้มมีลักษณะการแจกแจงไม่เป็นโค้งปกติ นั่นคือ มีลักษณะการแจกแจงแบบเบ้ขวา ส่วนในทุกเงื่อนไขความยาว

ของแบบทดสอบของกลุ่มตัวอย่างขนาด 300 คน และ 100 คน มีลักษณะการแจกแจงเป็นโค้งปกติ ส่วนคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี d_0 มีลักษณะการแจกแจงเป็นโค้งปกติทุกเงื่อนไข ความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง สำหรับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D_0 และ RD_0 แนวโน้มมีลักษณะการแจกแจงคล้ายคลึงกันทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง โดยมีลักษณะการแจกแจงไม่เป็นโค้งปกติ นั่นคือ มีลักษณะการแจกแจงแบบเบ้ขวา สำหรับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี HLM มีลักษณะการแจกแจงไม่เป็นโค้งปกติ นั่นคือ มีลักษณะการแจกแจงแบบเบ้ขวา ทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ยกเว้น กลุ่มตัวอย่างขนาด 100 คน ในทุกความยาวของแบบทดสอบ พบว่า มีลักษณะการแจกแจงเป็นโค้งปกติ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

สำหรับลักษณะการแจกแจงคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีที่สนใจศึกษาและวิธีโมเดลพหุระดับ ที่มีลักษณะการแจกแจงแบบเดียวกัน เป็นตัวบ่งชี้เพื่อยืนยันว่าการวัดการเปลี่ยนแปลงนั้นสามารถประมาณค่าคะแนนการเปลี่ยนแปลงได้สอดคล้องกันในภาพรวม

เมื่อเปรียบเทียบลักษณะการแจกแจงคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีที่สนใจศึกษา พบว่า วิธี D และ S มีลักษณะการแจกแจงแบบเดียวกันทุกเงื่อนไข คือ ในทุกเงื่อนไขความยาวของแบบทดสอบ ของขนาดกลุ่มตัวอย่าง 1,085 คน และ 500 คน มีลักษณะการแจกแจงแบบเบ้ขวา ขนาดกลุ่มตัวอย่าง 300 คน และ 100 คน มีลักษณะการแจกแจงแบบโค้งปกติ และวิธี D_0 และ RD_0 มีลักษณะการแจกแจงแบบเดียวกันทุกเงื่อนไข คือ มีลักษณะการแจกแจงแบบเบ้ขวา

เมื่อพิจารณาตามเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง พบว่า ในเงื่อนไขทุกขนาดความยาวของแบบทดสอบ ของขนาดกลุ่มตัวอย่าง 1,085 คน และ 500 คน นั้น วิธี D, S, D_0 , RD_0 และ HLM มีลักษณะการแจกแจงเป็นแบบเบ้ขวา สำหรับในเงื่อนไขทุกขนาดความยาวของแบบทดสอบ ของขนาดกลุ่มตัวอย่าง 300 คน นั้น วิธี D_0 , RD_0 และ HLM มีลักษณะการแจกแจงแบบเบ้ขวา และ วิธี D, S และ d_0 มีลักษณะการแจกแจงแบบโค้งปกติ ส่วนในเงื่อนไขทุกความยาวของแบบทดสอบ ของขนาดกลุ่มตัวอย่าง 100 คน วิธี D_0 และ RD_0 มีลักษณะการแจกแจงแบบเบ้ขวา วิธี D, S, d_0 และ HLM มีลักษณะการแจกแจงแบบโค้งปกติ

ตอนที่ 3 การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง

ผู้วิจัยได้แบ่งการวิเคราะห์ในตอนนี้ออกเป็น 2 ขั้นตอน ตามวัตถุประสงค์ของการวิจัยข้อ 1 เพื่อศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง คือ 1) การศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง และ 2) การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีกับวิธีโมเดลพหุระดับ โดยมีรายละเอียดดังนี้

3.1 การศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง

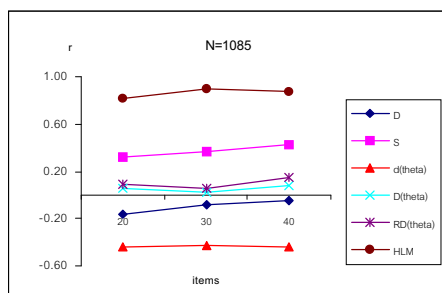
ผู้วิจัยศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธี คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ และวิธีโมเดลพหุระดับ แบบสองระดับ ที่เป็นโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดกับค่าเฉลี่ยของตัวแปรต้น (Uncentered model) ซึ่งเป็นวิธีการประมาณค่าคะแนนการเปลี่ยนแปลงที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน เพื่อทดสอบสมมติฐานของการวิจัยข้อ 1 ที่ตั้งไว้ว่า คะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีความสัมพันธ์แบบเพียร์สันไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่าวิธีอื่น ปรากฏผลการวิเคราะห์ดังตาราง 43

ตาราง 43 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดแต่ละวิธี ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน

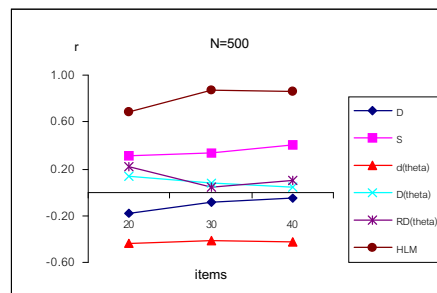
N	K	D	S	d_θ	D_θ	RD_θ	HLM
1,085	40 ข้อ	-0.05*	0.42**	-0.44**	0.09**	0.15**	0.87**
คน	30 ข้อ	-0.08**	0.37**	-0.43**	0.02	0.06*	0.90**
	20 ข้อ	-0.16**	0.32**	-0.43**	0.06*	0.09**	0.82**
500	40 ข้อ	-0.05	0.41**	-0.42**	0.04	0.11**	0.86**
คน	30 ข้อ	-0.09*	0.34**	-0.42**	0.08*	0.04	0.87**
	20 ข้อ	-0.18**	0.31**	-0.44**	0.14**	0.22**	0.68**
300	40 ข้อ	-0.06	0.49**	-0.41**	0.17*	0.29**	0.83**
คน	30 ข้อ	-0.08	0.44**	-0.41**	0.11*	0.14**	0.86**
	20 ข้อ	-0.16**	0.36**	-0.42**	0.12*	0.20**	0.84**
100	40 ข้อ	-0.15	0.27**	-0.46**	0.02	0.10	0.95**
คน	30 ข้อ	-0.19*	0.19*	-0.48**	0.03	0.08	0.96**
	20 ข้อ	-0.29**	0.12	-0.50**	0.14	0.24**	0.89**

* P<.05, **P<.01

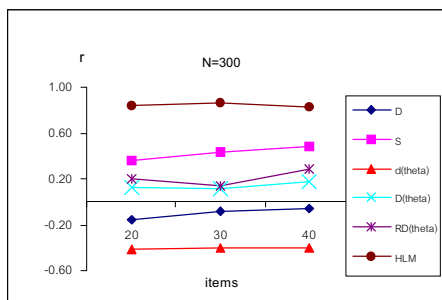
และนำผลการวิเคราะห์ข้อมูลตามตาราง 43 มาเขียนเป็นแผนภูมิได้ดังภาพประกอบ 6 (ก-ค) และ 7 (ก-ง)



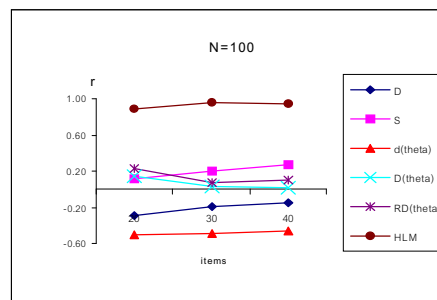
6 ก



6 ข

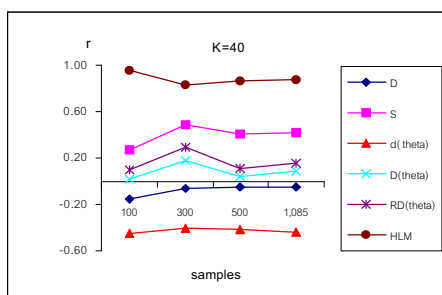


6 ค

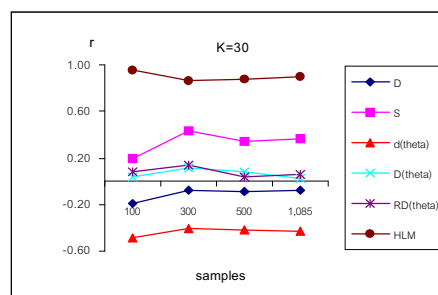


6 ง

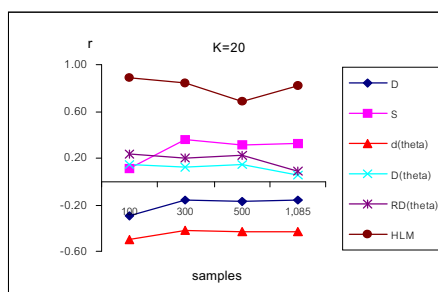
ภาพประกอบ 6 (ก-ง) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี ของกลุ่มตัวอย่างคงที่ 1,085 คน (6 ก), 500 คน (6 ข), 300 คน (6 ค) และ 100 คน (6 ง) เมื่อความยาวของแบบทดสอบไม่เท่ากัน



7 ก



7 ข



7 ค

ภาพประกอบ 7 (ก-ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีของกลุ่มตัวอย่างไม่เท่ากัน เมื่อความยาวของแบบทดสอบคงที่ 40 ข้อ (7 ก), 30 ข้อ (7 ข) และ 20 ข้อ (7 ค)

ผลการวิเคราะห์ตามตาราง 43 และภาพประกอบ 6 (ก-ง) และ 7 (ก-ค) พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี d_0 และ HLM มีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า ทุกเงื่อนไขความยาวของแบบทดสอบ และขนาดกลุ่มตัวอย่าง โดยค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ไม่เป็นลบและเข้าใกล้ศูนย์มากที่สุดในทุกเงื่อนไข รองลงมา คือ วิธี RD_0 ยกเว้น ในกลุ่มตัวอย่าง 500 คน ที่สอด้วยแบบทดสอบที่มีจำนวนข้อสอบ 30 ข้อ ที่พบว่าวิธี RD_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ไม่เป็นลบและเข้าใกล้ศูนย์มากที่สุด รองลงมา คือ วิธี D_0 ส่วนวิธี HLM มีค่าสัมประสิทธิ์สหสัมพันธ์เป็นบวกที่มีค่าสูงสุดทุกเงื่อนไข รองลงมา คือวิธี S สำหรับวิธี D และ d_0 มีค่าสัมประสิทธิ์สหสัมพันธ์เป็นลบทุกเงื่อนไข

เมื่อพิจารณาตามความยาวของแบบทดสอบที่ต่างกันในขณะที่ขนาดกลุ่มตัวอย่างคงที่ พบว่าเมื่อจำนวนข้อสอบเพิ่มมากขึ้นในทุกเงื่อนไขขนาดกลุ่มตัวอย่าง แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D มีค่าที่เป็นลบลดน้อยลง โดยมีค่าเข้าใกล้ศูนย์มากขึ้น วิธี d_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ใกล้เคียงกัน วิธี S มีค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มมากขึ้น วิธี D_0 และ RD_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ไม่เป็นระบบ ส่วนวิธี HLM มีค่าสัมประสิทธิ์สหสัมพันธ์ใกล้เคียงกัน

เมื่อพิจารณาตามขนาดกลุ่มตัวอย่างที่ไม่เท่ากันในขณะที่ความยาวของแบบทดสอบคงที่ พบว่าเมื่อจำนวนผู้สอบเพิ่มมากขึ้นในทุกเงื่อนไขความยาวของแบบทดสอบ แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D และ d_0 มีค่าที่เป็นลบลดน้อยลง โดยมีค่าเข้าใกล้ศูนย์มากขึ้นในทุกความยาวแบบทดสอบ ส่วนวิธี S มีค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มมากขึ้น จากขนาดกลุ่มตัวอย่าง 100 คน เป็น 300 คน ในขณะที่ขนาดกลุ่มตัวอย่าง 300 คน 500 คน และ 1,085 คน ในทุกความยาวแบบทดสอบ มีแนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ใกล้เคียงกัน ส่วนวิธี D_0 และ RD_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ลดลงโดยมีค่าเข้าใกล้ศูนย์มากขึ้น ส่วนวิธี HLM มีค่าสัมประสิทธิ์สหสัมพันธ์ไม่เป็นระบบ

3.2 การเปรียบเทียบความสัมพันธ์ของคะแนนการเปลี่ยนแปลง

ผู้วิจัยเปรียบเทียบความสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษา โดยมีการวัดการเปลี่ยนแปลงด้วยวิธีโมเดลพหุระดับ แบบสองระดับ ที่เป็นโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดกับค่าเฉลี่ยของตัวแปรต้น (Uncentered model) ซึ่งเป็นวิธีการประมาณค่าคะแนนการเปลี่ยนแปลงที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ ผู้วิจัยศึกษาระดับความสอดคล้องของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีกับวิธีโมเดลพหุระดับ ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างที่ไม่เท่ากัน เพื่อทดสอบสมมติฐานของการ

วิจัยข้อ 2 ที่ตั้งไว้ว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีความสัมพันธ์แบบเพียร์สันทางบวกกับวิธีโมเดลพหุระดับสูงกว่าวิธีอื่น คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีความสัมพันธ์แบบเพียร์สันทางบวกกับวิธีโมเดลพหุระดับสูงกว่าวิธีอื่น โดยมีรายละเอียดดังนี้

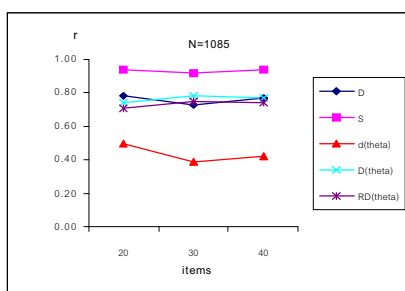
3.2.1 การวิเคราะห์ค่าความสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ พิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ปรากฏผลการวิเคราะห์ดังตาราง 44

ตาราง 44 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ได้มาจากแต่ละวิธีกับวิธีโมเดลพหุระดับ ของกลุ่มตัวอย่างที่มีขนาดไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน

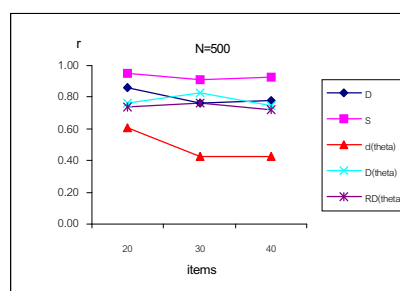
N	K	D	S	d_θ	D_θ	RD_θ
1,085 คน	40 ข้อ	0.77**	0.94**	0.42**	0.77**	0.74**
	30 ข้อ	0.73**	0.92**	0.39**	0.78**	0.75**
	20 ข้อ	0.78**	0.94**	0.50**	0.74**	0.71**
500 คน	40 ข้อ	0.78**	0.93**	0.43**	0.75**	0.72**
	30 ข้อ	0.76**	0.91**	0.43**	0.83**	0.76**
	20 ข้อ	0.86**	0.95**	0.61**	0.76**	0.74**
300 คน	40 ข้อ	0.80**	0.97**	0.49**	0.78**	0.75**
	30 ข้อ	0.77**	0.96**	0.45**	0.79**	0.74**
	20 ข้อ	0.76**	0.94**	0.49**	0.75**	0.73**
100 คน	40 ข้อ	0.65**	0.86**	0.36**	0.71**	0.71**
	30 ข้อ	0.63**	0.83**	0.34**	0.71**	0.70**
	20 ข้อ	0.71**	0.88**	0.49**	0.78**	0.78**

**P<.01

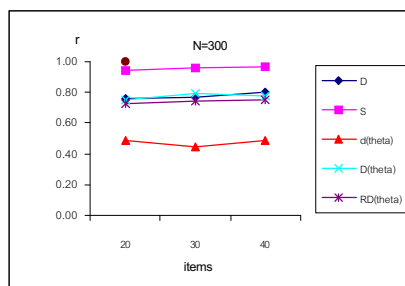
และนำผลการวิเคราะห์ข้อมูลตามตาราง 44 มาเขียนเป็นแผนภูมิ ได้ดังภาพประกอบ 8 (ก-ง) และ 9 (ก-ค)



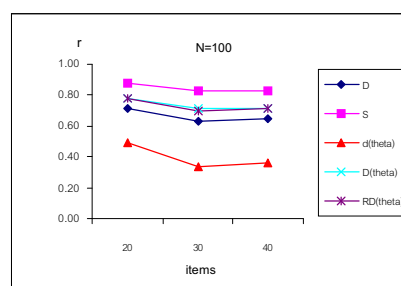
8 ก



8 ข

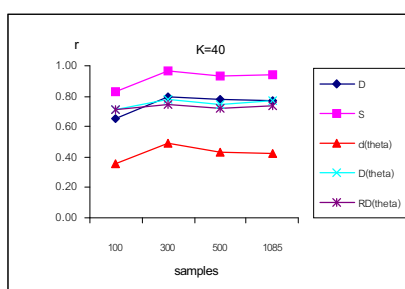


8 ค

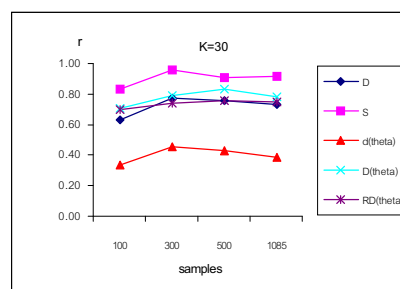


8 ง

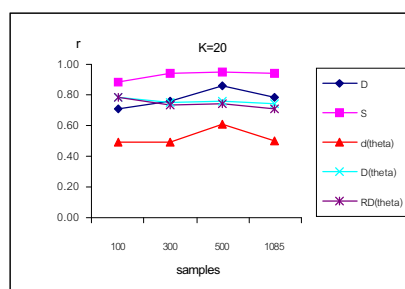
ภาพประกอบ 8 (ก-ง) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลง ของกลุ่มตัวอย่างคงที่ 1,085 คน (8 ก), 500 คน (8 ข), 300 คน (8 ค) และ 100 คน (8 ง) เมื่อความยาวของแบบทดสอบที่ต่างกัน



9 ก



9 ข



9 ค

ภาพประกอบ 9 (ก-ค) แสดงค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลง ของ กลุ่มตัวอย่าง ไม่เท่ากัน เมื่อความยาวของแบบทดสอบคงที่ 40 ข้อ (9 ก), 30 ข้อ (9 ข) และ 20 ข้อ (9 ค)

ผลการวิเคราะห์ตามตาราง 44 และภาพประกอบ 8 (ก-ง) และ 9 (ก-ค) พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนการเปลี่ยนแปลงแต่ละวิธีที่สนใจกับวิธีโมเดลพหุระดับ มีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่าในทุกเงื่อนไข (r มีค่าตั้งแต่ 0.34 ถึง 0.97) โดยคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี S กับ วิธี HLM มีความสัมพันธ์กันสูงสุดในทุกเงื่อนไข ในขณะที่ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี d_0 กับวิธี HLM มีค่าต่ำสุดในทุกเงื่อนไข ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D กับวิธี HLM, วิธี D_0 กับวิธี HLM และวิธี RD_0 กับวิธี HLM มีค่าใกล้เคียงกันในทุกเงื่อนไข

เมื่อพิจารณาตามความยาวของแบบทดสอบที่ต่างกันในขณะที่ขนาดกลุ่มตัวอย่างคงที่ พบว่า เมื่อจำนวนข้อสอบเพิ่มมากขึ้นในทุกเงื่อนไขขนาดกลุ่มตัวอย่าง แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี S กับวิธี HLM มีค่าสูงขึ้นเล็กน้อย ส่วนวิธี D_0 กับวิธี HLM มีค่าใกล้เคียงกัน เช่นเดียวกับ วิธี RD_0 กับวิธี HLM และ วิธี D กับวิธี HLM มีค่าใกล้เคียงกัน ในขณะที่แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี d_0 กับวิธี HLM มีค่าลดลงจากจำนวนข้อสอบ 20 ข้อ ไปเป็นจำนวน 30 ข้อ และมีค่าใกล้เคียงกันจากจำนวนข้อสอบ 30 ข้อไปเป็นจำนวน 40 ข้อ ในทุกเงื่อนไขขนาดกลุ่มตัวอย่าง

เมื่อพิจารณาตามขนาดกลุ่มตัวอย่างที่ไม่เท่ากันในขณะที่ความยาวของแบบทดสอบคงที่ พบว่า เมื่อจำนวนผู้สอบเพิ่มมากขึ้นในทุกเงื่อนไขความยาวของแบบทดสอบ แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี S กับวิธี HLM มีค่าใกล้เคียงกัน เช่นเดียวกับ วิธี D กับวิธี HLM, วิธี D_0 กับวิธี HLM และ วิธี RD_0 กับวิธี HLM ที่มีค่าใกล้เคียงกัน ในขณะที่แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี d_0 กับวิธี HLM มีค่าไม่เป็นระบบ

3.2.2 การทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน โดยนำค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันมาแปลงเป็นคะแนนมาตรฐาน (Z) ของฟิชเชอร์ แล้วนำมาทดสอบ โดยใช้สถิติไค-สแควร์ (χ^2) ผลปรากฏดังตาราง 45 ในภาคผนวก ก พบว่า

ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีกับวิธีโมเดลพหุระดับในแต่ละเงื่อนไขความยาวของแบบทดสอบและกลุ่มตัวอย่างของทุกเงื่อนไข มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ในแต่ละเงื่อนไขแตกต่างกันอย่างน้อย 1 คู่

เมื่อพบค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันที่นำมาทดสอบมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 จึงนำค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันมาทดสอบความแตกต่างรายคู่ต่อไป

3.2.3 การทดสอบความแตกต่างรายคู่ของค่าสัมประสิทธิ์สหสัมพันธ์ โดยใช้สถิติ Z เพื่อดูว่าค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลงคู่ใดในแต่ละเงื่อนไขความยาวของแบบทดสอบที่ต่างกันและกลุ่มตัวอย่างที่มีขนาดไม่เท่ากัน ปราบกฎผลการวิเคราะห์ดังตาราง 46 ในภาคผนวก ก พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ยกเว้น ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D-HLM กับวิธี D_0 -HLM ในเงื่อนไขขนาดกลุ่มตัวอย่าง 1,085 และ 500 คน ที่สอบด้วยแบบทดสอบจำนวน 40 ข้อ และในทุกเงื่อนไขความยาวของแบบทดสอบของขนาดกลุ่มตัวอย่าง 300 คน และ 100 คน วิธี D-HLM กับวิธี RD_0 -HLM ในเงื่อนไขขนาดกลุ่มตัวอย่าง 1,085 คน ที่สอบด้วยแบบทดสอบ 40 และ 30 ข้อ ขนาดกลุ่มตัวอย่าง 500 คน ที่สอบด้วยแบบทดสอบ 30 ข้อ และในทุกเงื่อนไขความยาวของแบบทดสอบ ของขนาดกลุ่มตัวอย่าง 300 คน และ 100 คน และค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D_0 -HLM กับวิธี RD_0 -HLM ในทุกเงื่อนไขความยาวของแบบทดสอบของขนาดกลุ่มตัวอย่าง 100 คน ที่มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .01

ตอนที่ 4 การศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง

ในการศึกษาตอนนี้ ผู้วิจัยได้แปลงคะแนนการเปลี่ยนแปลงที่ประมาณค่าจากการวัดการเปลี่ยนแปลงแต่ละวิธีให้อยู่ในมาตรเดียวกัน เพื่อสามารถเปรียบเทียบกันได้ โดยแปลงเป็นคะแนนมาตรฐาน (Z-score) แล้วมาจัดกระทำ โดยแบ่งการวิเคราะห์ออกเป็น 2 ขั้นตอน ตามวัตถุประสงค์ของการวิจัย ข้อ 2 เพื่อศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง คือ 1) การศึกษาความเชื่อมั่นของการวัดการเปลี่ยนแปลง และ 2) การศึกษาความเที่ยงตรงของการวัดการเปลี่ยนแปลง ซึ่งวิธีการประมาณค่าที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ คือ วิธีโมเดลพหุระดับ โดยมีรายละเอียดดังนี้

4.1 การศึกษาความเชื่อมั่นของการวัดการเปลี่ยนแปลง

ผู้วิจัยศึกษาความเชื่อมั่นของการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษา ด้วยการวิเคราะห์ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard error of estimate) ซึ่งผู้วิจัยใช้ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root mean squared of error; RMSE) เพื่อแสดงความคงเส้นคงวา (Consistency) ของการวัดการเปลี่ยนแปลงแต่ละวิธี เมื่อมีคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธีโมเดลพหุระดับเป็นวิธีที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ เพื่อทดสอบสมมติฐานของการวิจัยข้อ 3 ที่ตั้งไว้ว่า การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพันธ์ มีความเชื่อมั่นในการ

ประมาณค่าคะแนนการเปลี่ยนแปลงสูงกว่าวิธีอื่น ปรากฏผลการวิเคราะห์ดังตาราง 47 และภาพประกอบ 10 (ก-ง) และ 11 (ก-ค)

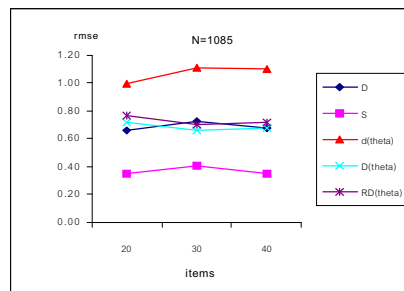
4.2 การศึกษาความเที่ยงตรงของการวัดการเปลี่ยนแปลง

ผู้วิจัยศึกษาความเที่ยงตรงของการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษา ด้วยการวิเคราะห์ค่าความลำเอียง (BIAS) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีที่สนใจศึกษา เมื่อมีคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธีโมเดลพหุระดับเป็นวิธีที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ ซึ่งค่าความลำเอียงเป็นค่าความลำเอียงเฉลี่ยที่คิดเครื่องหมาย (Average signed bias) เป็นการบอกทิศทางของความลำเอียง หรือความเที่ยงตรงของการวัดการเปลี่ยนแปลงแต่ละวิธี เพื่อทดสอบสมมติฐานของการวิจัยข้อ 4 ที่ตั้งไว้ว่า การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีความเที่ยงตรงในการประมาณค่าคะแนนการเปลี่ยนแปลงสูงกว่าวิธีอื่น ปรากฏผลการวิเคราะห์ดังตาราง 47

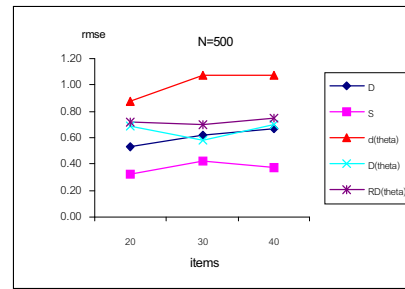
ตาราง 47 ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยและค่าความลำเอียงของคะแนนการเปลี่ยนแปลงที่ได้จากแต่ละวิธี ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน

N	K	D		S		d_θ		D_θ		RD_θ	
		RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS	RMSE	BIAS
1,085	40 ข้อ	0.68	0.00	0.35	0.00	1.10	0.00	0.68	0.00	0.72	0.00
คน	30 ข้อ	0.73	0.00	0.41	0.00	1.11	0.00	0.66	0.00	0.70	0.00
	20 ข้อ	0.66	0.00	0.35	0.00	1.00	0.00	0.72	0.00	0.77	0.00
500	40 ข้อ	0.67	0.00	0.37	0.00	1.07	0.00	0.70	0.00	0.75	0.00
คน	30 ข้อ	0.62	0.00	0.42	0.00	1.07	0.00	0.58	0.00	0.70	0.00
	20 ข้อ	0.53	0.00	0.32	0.00	0.88	0.00	0.69	0.00	0.72	0.00
300	40 ข้อ	0.64	0.00	0.25	0.00	1.00	0.00	0.66	0.00	0.70	0.00
คน	30 ข้อ	0.68	0.00	0.28	0.00	1.05	0.00	0.65	0.00	0.72	0.00
	20 ข้อ	0.69	0.00	0.34	0.00	1.01	0.00	0.71	0.00	0.74	0.00
100	40 ข้อ	0.82	0.00	0.52	0.00	1.13	0.00	0.76	0.00	0.76	0.00
คน	30 ข้อ	0.86	0.00	0.59	0.00	1.15	0.00	0.76	0.00	0.78	0.00
	20 ข้อ	0.76	0.00	0.49	0.00	1.01	0.00	0.66	0.00	0.66	0.00

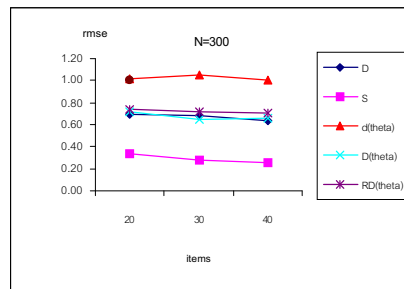
และนำผลการวิเคราะห์ตามตาราง 47 มาเขียนเป็นแผนภูมิได้ดังภาพประกอบ 10 (ก-ง) และ 11 (ก-ค) ดังนี้



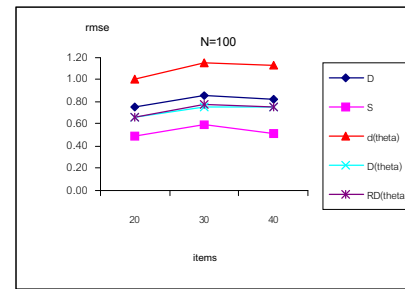
10 ก



10 ข

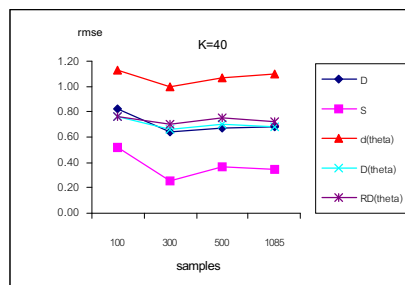


10 ค

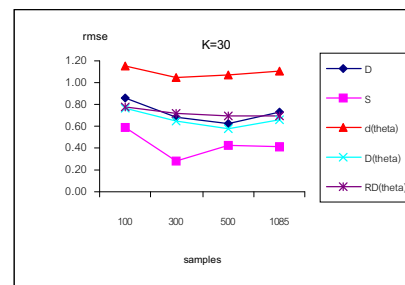


10 ง

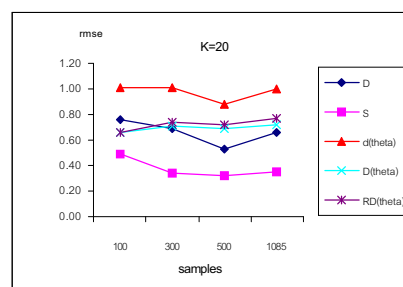
ภาพประกอบ 10 (ก-ง) แสดงค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของคะแนนการเปลี่ยนแปลงที่ได้จากแต่ละวิธี ของกลุ่มตัวอย่างคงที่ 1,085 คน (10 ก), 500 คน (10 ข), 300 คน (10 ค) และ 100 คน (10 ง) เมื่อความยาวของแบบทดสอบ 40 ข้อ 30 ข้อ และ 20 ข้อ



11 ก



11 ข



11 ค

ภาพประกอบ 11 (ก-ค) แสดงค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ยของคะแนนการเปลี่ยนแปลงที่ได้จากแต่ละวิธี ของกลุ่มตัวอย่างที่ไม่เท่ากัน เมื่อความยาวของแบบทดสอบคงที่ 40 ข้อ (11 ก), 30 ข้อ (11 ข) และ 20 ข้อ (11 ค)

ผลการวิเคราะห์ตามตาราง 47 และภาพประกอบ 10 (ก-ง) และ 11 (ก-ค) พบว่า ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี S มีค่าต่ำสุด ทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ส่วนวิธี d_0 มีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) สูงสุดทุกเงื่อนไข

เมื่อพิจารณาตามความยาวของแบบทดสอบที่ต่างกันในขณะที่ขนาดกลุ่มตัวอย่างคงที่ พบว่าเมื่อจำนวนข้อสอบเพิ่มขึ้นในทุกเงื่อนไขขนาดกลุ่มตัวอย่าง แนวโน้มค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากภายในแต่ละวิธีมีค่าใกล้เคียงกัน ยกเว้น วิธี d_0 มีแนวโน้มค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เพิ่มขึ้นเล็กน้อย

เมื่อพิจารณาตามขนาดกลุ่มตัวอย่างไม่เท่ากันในขณะที่ความยาวของแบบทดสอบคงที่ พบว่าเมื่อจำนวนผู้สอบเพิ่มขึ้นในทุกเงื่อนไขความยาวของแบบทดสอบ แนวโน้มค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากภายในแต่ละวิธีมีค่าใกล้เคียงกัน

สำหรับค่าความลำเอียงของคะแนนการเปลี่ยนแปลง พบว่า ค่าความลำเอียงของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษามีค่าเป็นศูนย์ทุกเงื่อนไข

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการศึกษาผลของการวัดการเปลี่ยนแปลงน้ำหนักตัวซ้ำสองครั้ง คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีการวัดการเปลี่ยนแปลงด้วยวิธีโมเดลพหุระดับ เป็นวิธีประมาณค่าที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ โดยพิจารณาจากค่าความสัมพันธ์ของคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง ค่าความสัมพันธ์ของคะแนนการเปลี่ยนแปลง ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าคะแนนการเปลี่ยนแปลง และค่าความลำเอียงของคะแนนการเปลี่ยนแปลง ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน ซึ่งมีความมุ่งหมายและสมมติฐานของการวิจัย ดังนี้

ความมุ่งหมายของการวิจัย

1. ศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง โดยศึกษาสองประเด็น ดังนี้

1.1 ศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษาและวิธีโมเดลพหุระดับ ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

1.2 เปรียบเทียบความสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

2. ศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง โดยศึกษาสองประเด็น ดังนี้

2.1 ศึกษาความเชื่อมั่นของการวัดการเปลี่ยนแปลง โดยพิจารณาจากความคลาดเคลื่อนมาตรฐานในการประมาณค่าคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

2.2 ศึกษาความเที่ยงตรงของการวัดการเปลี่ยนแปลง โดยพิจารณาจากค่าความลำเอียงของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ภายใต้เงื่อนไข ความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

สมมติฐานของการวิจัย

1. คะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมผัส มีความสัมพันธ์แบบเพียร์สันไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่าวิธีอื่น โดย

1.1 แบบทดสอบที่มีจำนวนข้อสอบมากจะมีค่าความสัมพันธ์ไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่าแบบทดสอบที่มีจำนวนข้อน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง

1.2 กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีค่าความสัมพันธ์ไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่ากลุ่มตัวอย่างที่มีจำนวนผู้สอบน้อยกว่า ในทุกความยาวของแบบทดสอบ

2. คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมผัส มีความสัมพันธ์แบบเพียร์สันทางบวกกับวิธีโมเดลพระดั่งสูงกว่าวิธีอื่น โดย

2.1 แบบทดสอบที่มีจำนวนข้อสอบมากจะมีค่าความสัมพันธ์สูงกว่าแบบทดสอบที่มีจำนวนข้อน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง

2.2 กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีค่าความสัมพันธ์สูงกว่ากลุ่มตัวอย่างที่มีจำนวนผู้สอบน้อยกว่า ในทุกความยาวของแบบทดสอบ

3. การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมผัส มีความเชื่อมั่นในการประมาณค่าคะแนนการเปลี่ยนแปลงสูงกว่าวิธีอื่น โดย

3.1 แบบทดสอบที่มีจำนวนข้อสอบมากจะมีความเชื่อมั่นในการประมาณค่าสูงกว่าแบบทดสอบที่มีจำนวนข้อน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง

3.2 กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีความเชื่อมั่นในการประมาณค่าสูงกว่ากลุ่มตัวอย่างที่มีจำนวนผู้สอบที่น้อยกว่า ในทุกความยาวของแบบทดสอบ

4. การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมผัส มีความเที่ยงตรงในการประมาณค่าคะแนนการเปลี่ยนแปลงสูงกว่าวิธีอื่น โดย

4.1 แบบทดสอบที่มีจำนวนข้อสอบมากจะมีความเที่ยงตรงในการประมาณค่าสูงกว่าแบบทดสอบที่มีจำนวนข้อน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง

4.2 กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีความเที่ยงตรงในการประมาณค่าสูงกว่ากลุ่มตัวอย่างที่มีจำนวนผู้สอบที่น้อยกว่า ในทุกความยาวของแบบทดสอบ

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ คือ นักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาระดับพื้นฐาน พุทธศักราช 2544 ระดับมัธยมศึกษาตอนต้น ของศูนย์บริการการศึกษานอกโรงเรียนเขต 7 แห่ง

ซึ่งเป็นสถานศึกษาในสังกัดศูนย์การศึกษานอกโรงเรียนกรุงเทพมหานคร 1 ในภาคเรียนที่ 2 ปีการศึกษา 2549 มีจำนวน 1,085 คน โดยใช้วิธีการสุ่มแบบหลายขั้นตอน (Multistage random sampling)

เครื่องมือที่ใช้ในการศึกษา

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ ได้แก่ แบบทดสอบวิชาคณิตศาสตร์ เนื้อหาเรื่อง สมการเชิงเส้น ตัวแปรเดียว ระดับมัธยมศึกษาตอนต้น หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 เป็นแบบทดสอบเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างข้อสอบขึ้นตามลักษณะเฉพาะของข้อสอบและพัฒนาโดยทดลองใช้สองครั้งเพื่อหาคุณลักษณะทางสถิติรายข้อ แบบทดสอบ และความเหมาะสมที่จะนำไปใช้กับกลุ่มตัวอย่าง ด้วยการสร้างข้อสอบจำนวน 80 ข้อ แบ่งข้อสอบออกเป็น 2 ฉบับที่มีความเป็นคู่ขนานกัน แล้วนำไปทดลองใช้ครั้งแรกกับนักศึกษาการศึกษานอกโรงเรียน ฉบับที่ 1 จำนวน 207 คน และ ฉบับที่ 2 จำนวน 197 คน เพื่อวิเคราะห์คุณลักษณะทางสถิติรายข้อและคัดเลือกข้อสอบจำนวน 50 ข้อ ต่อจากนั้นนำไปทดลองใช้ครั้งที่สองกับนักศึกษาจำนวน 1,034 คน เพื่อตรวจสอบความเป็นมิติเดียว (Unidimensional) วิเคราะห์คุณลักษณะทางสถิติรายข้อตามทฤษฎีการตอบข้อสอบด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ คัดเลือกข้อสอบที่มีคุณภาพไว้จำนวน 40 ข้อ ตามเกณฑ์ของวอร์ม (Warm, 1978: 52-53) ที่กำหนดค่าความยาก (b) ให้อยู่ระหว่าง -2.50 ถึง +2.50 ค่าอำนาจจำแนก (a) มากกว่า 0.80 และค่าการเดา (c) น้อยกว่า 0.30 โดยการสุ่มแบบแบ่งชั้น (Stratified random sampling) ที่มีมวลความรู้อยู่เป็นชั้นและข้อสอบเป็นหน่วยในการสุ่ม แล้ววิเคราะห์คุณลักษณะทางสถิติของแบบทดสอบตามทฤษฎีการตอบข้อสอบด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์

วิธีการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในการวิจัยครั้งนี้ เป็นการสอบซ้ำสามครั้งด้วยแบบทดสอบคณิตศาสตร์ ฉบับเดิมที่เป็นแบบ 4 ตัวเลือก จำนวน 40 ข้อ กับนักศึกษาการศึกษานอกโรงเรียน หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2544 ระดับมัธยมศึกษาตอนต้น ของศูนย์บริการการศึกษานอกโรงเรียนเขต 7 แห่ง ซึ่งเป็นสถานศึกษาในสังกัดศูนย์การศึกษานอกโรงเรียนกรุงเทพมหานคร 1 จำนวนนักศึกษา 1,085 คน โดยการสอบแต่ละครั้งมีระยะห่างกัน 3 สัปดาห์

การวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลในการวิจัยครั้งนี้ เป็นการเก็บรวบรวมข้อมูลกับนักศึกษา 1,085 คน ด้วยแบบทดสอบคณิตศาสตร์ที่มีความยาว 40 ข้อ หลังจากนั้นผู้วิจัยได้ทำการสุ่มตัวอย่างด้วยการสุ่มอย่างง่าย เพื่อสุ่มตัวอย่างผลการตอบของนักศึกษาเป็น 3 ขนาด คือ กลุ่มตัวอย่างที่มีจำนวน 500 คน, 300 คน และ 100 คน และสุ่มข้อสอบจำนวน 40 ข้อ เป็น 2 ขนาด คือ แบบทดสอบที่มีความยาว 30 และ 20

ข้อ โดยการสุ่มแบบแบ่งชั้น ที่มีมวลความรู้อยู่เป็นชั้น และข้อสอบเป็นหน่วยในการสุ่ม ซึ่งจะทำให้ได้ผลการตอบข้อสอบของนักศึกษาที่มีจำนวนไม่เท่ากันด้วยแบบทดสอบที่มีความยาวต่างกัน จำนวน 12 แบบแผน (4x3) เพื่อประมาณค่าคะแนนการเปลี่ยนแปลงด้วยการวัดการเปลี่ยนแปลงหาวิธีแบบวัดซ้ำสองครั้งและวิธีโมเดลพหุระดับ โดยผู้วิจัยได้แบ่งเป็น 4 ตอน คือ 1) วิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ 2) วิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลง 3) ศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง และ 4) ศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง

สรุปผลการวิจัย

การวิจัยในการศึกษาค้นคว้าครั้งนี้ สรุปผลตามประเด็นสำคัญได้ดังนี้

1. การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ
2. การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลง
3. การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง
4. การศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง

1. การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ

การศึกษาค่าสถิติพื้นฐานของคะแนนสอบแต่ละครั้งจากการสอบซ้ำสามครั้ง พบว่า คะแนนสอบเฉลี่ยทั้งสามครั้ง มีแนวโน้มเพิ่มขึ้นตามลำดับในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ค่ามัธยฐานและฐานนิยมของคะแนนจากการสอบครั้งที่ 1 และ 2 มีค่าใกล้เคียงกัน ซึ่งมีค่าต่ำกว่าค่ามัธยฐานและฐานนิยมของคะแนนที่ได้จากการสอบครั้งที่ 3 ทุกเงื่อนไข ส่วนเบี่ยงเบนมาตรฐานของคะแนนสอบมีแนวโน้มเพิ่มขึ้นตามลำดับทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง คะแนนรวมสูงสุดที่ได้จากการสอบครั้งที่ 1 มีแนวโน้มสูงขึ้นเมื่อขนาดกลุ่มตัวอย่างเพิ่มมากขึ้นในแต่ละความยาวของแบบทดสอบ คะแนนรวมสูงสุดที่ได้จากการสอบครั้งที่ 2 และ 3 มีค่าเท่ากัน คือ คะแนนเต็มในทุกเงื่อนไข ส่วนคะแนนรวมต่ำสุดที่ได้จากการสอบทั้งสามครั้งในแต่ละความยาวของแบบทดสอบของขนาดกลุ่มตัวอย่างคงที่มีค่าใกล้เคียงกัน พิสัยของคะแนนกว้างที่สุดคือ คะแนนจากการสอบครั้งที่ 2 ทุกเงื่อนไข

ลักษณะการแจกแจงของคะแนนสอบแต่ละครั้งจากค่าความเบ้ ความโด่ง และการทดสอบลักษณะการแจกแจงด้วยสถิติ Kolmogorov-Smirnov พบว่า คะแนนจากการสอบทั้งสามครั้งส่วนใหญ่มีลักษณะการแจกแจงที่คล้ายคลึงกัน คือ เบ้ขวา โดยมีความเบ้ขวาลดลงตามลำดับทุกเงื่อนไข ยกเว้นคะแนนสอบของกลุ่มตัวอย่าง 100 คน ที่สอบครั้งที่ 2 ในความยาวของแบบทดสอบ 40 ข้อ และ 20 ข้อ และ สอบครั้งที่ 3 ในทุกความยาวของแบบทดสอบ มีลักษณะการแจกแจงแบบโค้งปกติ

สำหรับผลการทดสอบการถดถอยเป็นเส้นตรงทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง พบว่า ข้อมูลทุกเงื่อนไขมีแบบแผนการเปลี่ยนแปลงเป็นเส้นตรง ซึ่งแสดงว่าสมการโพลิโนเมียลมีกำลังเป็นหนึ่ง ดังนั้นผู้วิจัยจึงเลือกที่จะทำการวิเคราะห์โมเดลพระดัดเชิงเส้นตรง และเมื่อทดสอบความแตกต่างระหว่างช่วงเวลาการสอบคณิตศาสตร์ พบว่า ช่วงเวลาในการวัดแต่ละครั้งมีผลทำให้คะแนนสอบคณิตศาสตร์ของนักศึกษาแตกต่างกัน ทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง จึงได้ทดสอบเปรียบเทียบค่าเฉลี่ยรายคู่ พบว่าค่าเฉลี่ยของคะแนนคณิตศาสตร์จากการสอบครั้งที่ 1 มีค่าต่ำกว่าค่าเฉลี่ยจากการสอบครั้งที่ 2 และครั้งที่ 3 และค่าเฉลี่ยจากการสอบครั้งที่ 2 มีค่าต่ำกว่าค่าเฉลี่ยจากการสอบครั้งที่ 3 ทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบแต่ละครั้งมีความสัมพันธ์ทางบวก มีค่าตั้งแต่ 0.43 ถึง 0.77 โดยค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบครั้งที่ 2 กับครั้งที่ 3 มีค่าสูงสุด รองลงมาคือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบครั้งที่ 1 กับครั้งที่ 2 และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบครั้งที่ 1 กับครั้งที่ 3 มีค่าต่ำสุด ทุกเงื่อนไข โดยเมื่อจำนวนข้อสอบเพิ่มมากขึ้น ในทุกขนาดกลุ่มตัวอย่างจะมีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบแต่ละครั้งเพิ่มมากขึ้น และเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นในความยาวแบบทดสอบจำนวน 20 ข้อ พบว่ามีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบแต่ละครั้งเพิ่มมากขึ้นด้วยเช่นกัน สำหรับขนาดกลุ่มตัวอย่างที่เพิ่มขึ้นในความยาวของแบบทดสอบจำนวน 40 และ 30 ข้อ มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบแต่ละครั้งเพิ่มขึ้นเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นจากจำนวน 100 คน เป็น 300 คน ส่วนขนาดกลุ่มตัวอย่าง 500 คน และ 1,085 คน มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการสอบแต่ละครั้งค่อนข้างใกล้เคียงกัน

2. การวิเคราะห์ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลง

ผู้วิจัยตรวจสอบโมเดลพระดัดตามประเภทของตัวแปรตาม ซึ่งในที่นี้ได้แก่ ตัวแปรเวลา จำนวน 3 แบบ คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) พบว่าโมเดลทั้งสามในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่างมีประสิทธิภาพของโมเดลเท่ากัน แสดงว่าโมเดลทั้งสามสามารถอธิบายการเปลี่ยนแปลงของคะแนนคณิตศาสตร์ของนักศึกษาการศึกษานอกโรงเรียน ระดับมัธยมศึกษาตอนต้นได้เหมาะสมเท่ากัน โดยในส่วนที่เป็นอิทธิพล คงที่ พบว่า คะแนนเริ่มต้นเฉลี่ยในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าต่ำกว่าอีกสองโมเดล คือ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยของตัวแปรต้น (Group centered model) และ โมเดลตัวแปรตามเป็นฟังก์ชันที่ยึดค่าเฉลี่ยรวมของตัวแปรต้น (Grand centered model) ส่วนอัตราพัฒนาการเฉลี่ยของทั้งสามโมเดลมีค่าเท่ากัน ส่วนที่เป็นอิทธิพลสุ่มพบว่า ความแปรปรวนของ

คะแนนเริ่มต้นในโมเดลตัวแปรตามเป็นฟังก์ชันที่ไม่ยึดค่าเฉลี่ยของตัวแปรต้น (Uncentered model) มีค่าน้อยกว่าอีกสองโมเดล ความแปรปรวนของอัตราพัฒนาการและความแปรปรวนของความคลาดเคลื่อนของทั้งสามโมเดลมีค่าเท่ากัน ทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกันและขนาดกลุ่มตัวอย่างที่ไม่เท่ากัน เมื่อพิจารณาลักษณะของข้อมูล พบว่าคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D, S มีลักษณะการแจกแจงคล้ายคลึงกันทุกเงื่อนไข ความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง โดยเงื่อนไขขนาดกลุ่มตัวอย่าง 1,085 คน และ 500 คน ในทุกเงื่อนไขความยาวของแบบทดสอบ มีลักษณะการแจกแจงแบบเบ้ขวา ส่วนขนาดกลุ่มตัวอย่าง 300 คน และ 100 คน มีลักษณะการแจกแจงแบบโค้งปกติ วิธี d_0 ในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง มีลักษณะการแจกแจงแบบเบ้ขวา วิธี D_0 และ RD_0 มีลักษณะการแจกแจงคล้ายคลึงกันทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง คือ มีลักษณะการแจกแจงแบบเบ้ขวา ในทุกเงื่อนไข และเช่นเดียวกับวิธี HLM ที่ส่วนใหญ่มีลักษณะการแจกแจงแบบเบ้ซ้าย ยกเว้น ขนาดกลุ่มตัวอย่าง 100 คน ในทุกความยาวของแบบทดสอบ ที่มีลักษณะการแจกแจงแบบโค้งปกติ

3. การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง

ผลการวิจัยสามารถสรุปเป็น 2 ประเด็น ตามสมมติฐานของการวิจัยได้ดังนี้

3.1 การศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงตามสมมติฐานของการวิจัยข้อ 1 : คะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีความสัมพันธ์แบบเพียร์สันไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่าวิธีอื่น

ผลการวิจัยสรุปได้ดังนี้ ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D_0 แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ไม่เป็นลบ และเข้าใกล้ศูนย์มากที่สุด รองลงมา คือ วิธี RD_0 ในทุกเงื่อนไข

เมื่อพิจารณาตามขนาดกลุ่มตัวอย่างที่ไม่เท่ากันในขณะที่ความยาวของแบบทดสอบคงที่ พบว่าเมื่อจำนวนผู้สอบเพิ่มมากขึ้นในทุกความยาวของแบบทดสอบ แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D และ d_0 มีค่าเป็นลบลดน้อยลง โดยมีค่าเข้าใกล้ศูนย์มากขึ้นในทุกความยาวแบบทดสอบ ส่วนวิธี D_0 และ RD_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ลดลงโดยมีค่าเข้าใกล้ศูนย์มากขึ้น สำหรับวิธี S และ HLM มีค่าสัมประสิทธิ์สหสัมพันธ์ไม่เป็นระบบ

เมื่อพิจารณาตามความยาวของแบบทดสอบที่ต่างกันในขณะที่ขนาดกลุ่มตัวอย่างคงที่ พบว่า เมื่อจำนวนข้อสอบเพิ่มมากขึ้นในทุกขนาดกลุ่มตัวอย่าง แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D มีค่าเป็นลบลดน้อยลง โดยมีค่าเข้าใกล้ศูนย์มากขึ้น วิธี d_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ใกล้เคียงกัน วิธี S มีค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มมากขึ้น วิธี D_0 และ RD_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ไม่เป็นระบบ ส่วนวิธี HLM มีค่าสัมประสิทธิ์สหสัมพันธ์ใกล้เคียงกัน

3.2 การเปรียบเทียบความสัมพันธ์ของคะแนนการเปลี่ยนแปลง ตามสมมติฐานของ

การวิจัย ข้อ 2 : คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีความสัมพันธ์แบบเพียร์สันทางบวกกับวิธีโมเดลพหุระดับสูงกว่าวิธีอื่น

ผลการวิจัยสรุปได้ดังนี้ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนการเปลี่ยนแปลงแต่ละวิธีที่สนใจกับวิธีโมเดลพหุระดับ มีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่าในทุกเงื่อนไข มีค่าตั้งแต่ 0.34 ถึง 0.97 โดยคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี S กับ วิธี HLM มีความสัมพันธ์กันสูงสุดในทุกเงื่อนไข ในขณะที่ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่าจากวิธี d_0 กับวิธี HLM มีค่าต่ำสุดในทุกเงื่อนไข และเมื่อทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์รายคู่ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีที่สนใจศึกษา กับวิธีโมเดลพหุระดับ พบว่า แนวโน้มมีค่าแตกต่างกันทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ยกเว้น ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D-HLM กับวิธี D_0 -HLM และ วิธี D-HLM กับวิธี RD_0 -HLM มีแนวโน้มไม่แตกต่างกันในเงื่อนไขขนาดกลุ่มตัวอย่างที่มีจำนวนผู้สอบมาก คือตั้งแต่ขนาดกลุ่มตัวอย่าง 300 คนขึ้นไปในทุกความยาวของแบบทดสอบ และค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D_0 -HLM กับวิธี RD_0 -HLM ในทุกเงื่อนไขความยาวของแบบทดสอบของขนาดกลุ่มตัวอย่าง 100 คน มีแนวโน้มไม่แตกต่างกันเช่นกัน

เมื่อพิจารณาตามขนาดกลุ่มตัวอย่างที่ไม่เท่ากันในขณะที่ความยาวของแบบทดสอบคงที่ พบว่า เมื่อจำนวนผู้สอบเพิ่มมากขึ้นในทุกเงื่อนไขความยาวของแบบทดสอบ แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีกับวิธี HLM มีค่าใกล้เคียงกัน ยกเว้นวิธี d_0 กับวิธี HLM ที่มีค่าไม่เป็นระบบ

เมื่อพิจารณาตามความยาวของแบบทดสอบที่ต่างกันในขณะที่ขนาดกลุ่มตัวอย่างคงที่ พบว่า เมื่อจำนวนข้อสอบเพิ่มมากขึ้นในทุกเงื่อนไขขนาดกลุ่มตัวอย่าง แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี S กับวิธี HLM มีค่าสูงขึ้นเล็กน้อย ส่วนวิธี D_0 กับวิธี HLM วิธี RD_0 กับวิธี HLM และ วิธี D กับวิธี HLM มีค่าใกล้เคียงกัน ในขณะที่วิธี d_0 กับวิธี HLM มีค่า

ลดลงจากจำนวนข้อสอบ 20 ข้อ ไปเป็นจำนวน 30 ข้อ และมีค่าใกล้เคียงกันจากจำนวนข้อสอบ 30 ข้อ ไปเป็นจำนวน 40 ข้อ ในทุกเงื่อนไขขนาดกลุ่มตัวอย่าง

4. การศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง

ผลการวิจัยสามารถสรุปเป็น 2 ประเด็น ตามสมมติฐานของการวิจัยได้ดังนี้

4.1 การศึกษาความเชื่อมั่นของการวัดการเปลี่ยนแปลง ตามสมมติฐานของการวิจัย

ข้อ 3 : คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำกว่าวิธีอื่นและเข้าใกล้ศูนย์มากที่สุด

ผลการวิจัยสรุปได้ดังนี้ ค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี S มีค่าต่ำที่สุด ทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ส่วนวิธี d_0 มีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) สูงที่สุดทุกเงื่อนไข แสดงว่า วิธี S เป็นวิธีประมาณค่าคะแนนการเปลี่ยนแปลงที่มีความเชื่อมั่นสูงสุด เมื่อมีวิธีโมเดลพหุระดับ เป็นวิธีประมาณค่าคะแนนการเปลี่ยนแปลงที่ใช้เป็นเกณฑ์ในครั้งนี้

เมื่อพิจารณาตามเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง พบว่า แนวโน้มค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากภายในแต่ละวิธีมีค่าใกล้เคียงกัน แสดงว่าทุกวิธีการวัดการเปลี่ยนแปลงมีความเชื่อมั่นในการประมาณค่าคะแนนการเปลี่ยนแปลงอยู่ในระดับเดียวกัน ในทุกเงื่อนไขความยาวของแบบทดสอบและทุกเงื่อนไขขนาดกลุ่มตัวอย่าง

4.2 การศึกษาความเที่ยงตรงของการวัดการเปลี่ยนแปลง ตามสมมติฐานของการวิจัย

ข้อ 4 : คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีค่าความลำเอียงต่ำกว่าวิธีอื่นและเข้าใกล้ศูนย์มากที่สุด

ผลการวิจัยสรุปได้ดังนี้ ค่าความลำเอียง (BIAS) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีมีค่าเป็นศูนย์ทุกเงื่อนไข แสดงว่าทุกวิธีการวัดการเปลี่ยนแปลงมีความเที่ยงตรงในการประมาณค่าคะแนนการเปลี่ยนแปลงสูงอยู่ในระดับเดียวกัน ในทุกเงื่อนไขความยาวของแบบทดสอบและทุกเงื่อนไขขนาดกลุ่มตัวอย่าง

สรุปผลโดยรวม ผลการศึกษาการวัดการเปลี่ยนแปลงห้าวิธีแบบวัดซ้ำสองครั้ง คือ การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝง การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัด

คะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ โดยมีการวัดการเปลี่ยนแปลงด้วยวิธีโมเดลพหุระดับ เป็นวิธีประมาณค่าที่ใช้เป็นเกณฑ์ในการศึกษาค้างนี้ พบว่า การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ (D_0) และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ (RD_0) แนวโน้มมีความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงไม่เป็นลบ เข้าใกล้ศูนย์มากกว่าวิธีอื่น ในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง สำหรับความสัมพันธ์ของคะแนนการเปลี่ยนแปลงระหว่างการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ (S) กับวิธีโมเดลพหุระดับ มีความสัมพันธ์กันมากกว่าวิธีอื่น ในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง และการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ (S) ยังมีความเชื่อมั่นในการประมาณค่าคะแนนการเปลี่ยนแปลงได้สอดคล้องกับวิธีโมเดลพหุระดับมากกว่าวิธีอื่น ในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง แต่พบว่าทุกวิธีการวัดคะแนนการเปลี่ยนแปลงมีความเที่ยงตรงในการประมาณค่าคะแนนการเปลี่ยนแปลงอยู่ในระดับเดียวกัน สามารถสรุปผลโดยรวมตามสมมติฐานของการวิจัยได้ดังตาราง 48

ตาราง 48 สรุปผลการศึกษการวัดการเปลี่ยนแปลงแต่ละวิธี ภายใต้เงื่อนไขความยาวของแบบทดสอบที่ต่างกัน และขนาดกลุ่มตัวอย่างไม่เท่ากัน

N	K	ความสัมพันธ์ของคะแนนการเปลี่ยนแปลง		ประสิทธิภาพของการวัดการเปลี่ยนแปลง	
		Cor (INT,SLP) มีค่าไม่เป็นลบ เข้าใกล้ศูนย์	Cor (com – cri) มีค่าสูงสุด	ความเชื่อมั่นสูงสุด (RMSE มีค่าต่ำสุด)	ความเที่ยงตรงสูงสุด (BIAS มีค่าต่ำสุด)
1085 คน	40 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
	30 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
	20 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
500 คน	40 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
	30 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
	20 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
300 คน	40 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
	30 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
	20 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
100 คน	40 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
	30 ข้อ	D_0, RD_0	S	S	ทุกวิธีวัด
	20 ข้อ	S, D_0	S	S	ทุกวิธีวัด

อภิปรายผล

การอภิปรายผลการวิจัยที่นำเสนอในตอนนี้เป็น 2 ประเด็น คือ 1) การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง และ 2) การศึกษาประสิทธิภาพของการวัดคะแนนการเปลี่ยนแปลง ดังรายละเอียดต่อไปนี้

1. การศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลง

ผู้วิจัยแบ่งการศึกษาความสัมพันธ์ของคะแนนการเปลี่ยนแปลงออกเป็น 2 ประเด็น คือ

1) การศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง และ 2) การเปรียบเทียบความสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ ดังนี้

1.1 การศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลง

การการศึกษาความสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงจากค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี D_0 มีแนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ไม่เป็นลบและเข้าใกล้ศูนย์มากที่สุด ในทุกเงื่อนไข รองลงมา คือ วิธี RD_0 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อ 1 ที่ตั้งไว้ว่า คะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ และการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพันธ์ มีความสัมพันธ์แบบเพียร์สันไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่าวิธีอื่น โดยสอดคล้องกับผลการวิจัยของไฮเดนเบิร์ก (Heidenberg, 2000) และ คำกล่าวของ บอค ทิสเซน และ ซิมมอสกี (Bock; Thissen; & Zimowsky, 1997) ที่กล่าวไว้ว่า คะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบสามารถลดปัญหาความสัมพันธ์ของคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงได้ และเมื่อพิจารณาภายใต้เงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ปรากฏผลดังนี้

1.1.1 เมื่อความยาวของแบบทดสอบเพิ่มขึ้นในทุกขนาดกลุ่มตัวอย่าง และเมื่อขนาดกลุ่มตัวอย่างเพิ่มมากขึ้นในทุกเงื่อนไขความยาวของแบบทดสอบ พบว่า วิธี D_0 และ RD_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงไม่เป็นลบ ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัยข้อ 1.1 ที่ตั้งไว้ว่า แบบทดสอบที่มีจำนวนข้อสอบมากจะมีค่าความสัมพันธ์ไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่าแบบทดสอบที่มีจำนวนข้อสอบน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง อาจเนื่องมาจากกลุ่มตัวอย่างที่ใช้ในการทดลองครั้งนี้ ส่วนใหญ่มีความสามารถอยู่ระดับอ่อนถึงปานกลาง ส่วนที่มีความสามารถค่อนข้างสูงมีน้อย ในขณะที่ข้อสอบส่วนใหญ่มีความยากอยู่ในระดับค่อนข้างยากถึงยาก ($b > +1.00$) ร้อยละ 77.50 ข้อสอบที่มีความยากอยู่ในระดับปานกลาง ($b = -1.00$ ถึง $+1.00$) ร้อยละ 22.50 และไม่มีข้อสอบง่าย ($b < -1.00$) ซึ่งคะแนนมวลความรู้ตามทฤษฎีการ

ตอบข้อสอบนั้นขึ้นอยู่กับค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ จึงเป็นการให้คะแนนตามค่าความยาก อำนาจจำแนก และการเดาของข้อสอบ ดังนั้นการสุ่มข้อสอบ จำนวน 30 ข้อ และ 20 ข้อ จากข้อสอบ 40 ข้อ โดยการสุ่มแบบแบ่งชั้นที่มีมวลความรู้ย่อยเป็นชั้นและข้อสอบเป็นหน่วยในการสุ่ม ซึ่งเป็นการสุ่มตามสัดส่วนมวลความรู้ย่อยและสุ่มอย่างง่าย ทำให้การประมาณค่าคะแนนการเปลี่ยนแปลงด้วยคะแนนมวลความรู้มีความคลาดเคลื่อนค่อนข้างสูง ดังนั้นผลการวิจัยจึงไม่สามารถลงสรุปได้

1.1.2 เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นในทุกความยาวของแบบทดสอบ พบว่า วิธี D_0 และ RD_0 มีค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนสอบก่อนเรียนกับคะแนนการเปลี่ยนแปลงไม่เป็นระบบ ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัยข้อ 1.2 ที่ตั้งไว้ว่า กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีค่าความสัมพันธ์ไม่เป็นลบ มีค่าเข้าใกล้ศูนย์มากกว่ากลุ่มตัวอย่างที่มีจำนวนผู้สอบน้อยกว่า ในทุกความยาวของแบบทดสอบ ทั้งนี้อาจเนื่องมาจากตัวอย่างที่สุ่มได้ในแต่ละขนาดมีทั้งที่เป็นตัวแทนที่ดีของประชากรและเป็นตัวแทนที่ไม่ดีของประชากร คือ ผลจากการทดสอบค่าเฉลี่ยคะแนนสอบก่อนเรียนของกลุ่มตัวอย่าง 500 คน 300 คน และ 100 คน กับค่าเฉลี่ยคะแนนสอบก่อนเรียนของประชากรเทียม คือ 1,085 คน พบว่าขนาดกลุ่มตัวอย่าง 500 คน และ 300 คน มีค่าเฉลี่ยไม่แตกต่างจากค่าเฉลี่ยของประชากรเทียม แต่ขนาดกลุ่มตัวอย่าง 100 คน มีค่าเฉลี่ยแตกต่างจากค่าเฉลี่ยของประชากรเทียม และเมื่อพิจารณาลักษณะการแจกแจงคะแนนสอบก่อนเรียนและหลังเรียนของผู้สอบขนาดกลุ่มตัวอย่าง 1,085 คน 500 คน และ 300 คน ในทุกเงื่อนไขความยาวของแบบทดสอบ พบว่ามีลักษณะการแจกแจงแบบเบ้ขวา ทั้งในการสอบก่อนเรียนและหลังเรียน ส่วนขนาดกลุ่มตัวอย่าง 100 คน มีลักษณะการแจกแจงคะแนนสอบก่อนเรียนเป็นแบบเบ้ขวา และคะแนนสอบหลังเรียนเป็นแบบโค้งปกติ ซึ่งแตกต่างจากลักษณะการแจกแจงคะแนนสอบก่อนเรียนและหลังเรียนของประชากรเทียม จึงอาจทำให้เกิดความคลาดเคลื่อนในการประมาณค่าคะแนนการเปลี่ยนแปลงเมื่อขนาดกลุ่มตัวอย่างแปรเปลี่ยน ในทุกเงื่อนไขความยาวของแบบทดสอบ แต่อย่างไรก็ตามยังไม่สามารถสรุปได้อย่างชัดเจนถึงสาเหตุของผลที่ได้จากการศึกษาในครั้งนี้ ควรได้มีการศึกษาวิธีการสุ่มกลุ่มตัวอย่างและลักษณะกลุ่มตัวอย่าง เพื่อตรวจสอบต่อไป

1.2 การเปรียบเทียบความสัมพันธ์ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ

จากการศึกษาค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนการเปลี่ยนแปลงแต่ละวิธีที่สนใจกับวิธีโมเดลพหุระดับ พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี S กับ วิธี HLM มีความสัมพันธ์กันสูงสุดในทุกเงื่อนไข ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัยข้อ 2 ที่ตั้งไว้ว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีความสัมพันธ์

แบบเพียร์สันทางบวกกับวิธีโมเดลพหุระดับสูงกว่าวิธีอื่น อาจเนื่องมาจาก 2 ประการ คือ ประการแรก ลักษณะแบบแผนการเปลี่ยนแปลง พบว่าแนวโน้มคะแนนการสอบสามครั้งมีแบบแผนการเปลี่ยนแปลงเป็นเส้นตรง ซึ่งเป็นไปตามคำกล่าวของ เราเดนบิวช และไบรย์ (Raudenbush; & Bry. 2002: 163) ที่ว่าการวัดจำนวนน้อยครั้ง เช่น 3 หรือ 4 ครั้ง จะมีโมเดลพัฒนาการเป็นเส้นตรง ประกอบกับเครื่องมือที่ใช้ในการศึกษาครั้งนี้ พบว่าอัตราส่วนของน้ำหนักคะแนนระหว่างความรู้พื้นฐาน : ทักษะการคำนวณ : การแก้โจทย์ปัญหา มีค่าเท่ากับ 22.5 : 62.5 : 15 ซึ่งแสดงว่าเครื่องมือที่ใช้ในการศึกษาครั้งนี้มีแนวโน้มวัดทักษะการคำนวณมากกว่าส่วนอื่น ซึ่งสอดคล้องกับผลการวิจัยของสมถวิล วิจิตรวรรณ (2543) ที่พบว่าผลการวัดการเปลี่ยนแปลงระยะยาว ชนิดตัวแปรเดียว คุณลักษณะการคิดคำนวณ มีแบบแผนการเปลี่ยนแปลงระยะยาวเป็นแบบเส้นตรง ซึ่งลักษณะแบบแผนการเปลี่ยนแปลงที่เป็นเส้นตรงนี้สอดคล้องกับข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบมาตรฐานเดิม ที่มีวิธี D และ S ในการวัดคะแนนการเปลี่ยนแปลงในการศึกษาครั้งนี้ ในขณะที่คะแนนผลความรู้ตามทฤษฎีการตอบข้อสอบเป็นคะแนนแปลง ซึ่งความสามารถ θ ไปเป็นคะแนนจริงที่ไม่เป็นเส้นตรง (Nonlinear)

ประการที่สอง การประมาณค่าพารามิเตอร์ด้วยวิธีโมเดลพหุระดับโดยใช้โปรแกรม HLM ซึ่งเป็นวิธีเกณฑ์ในการศึกษาครั้งนี้ มีการประมาณค่าด้วยวิธีความเป็นไปได้สูงสุด (Maximum likelihood) ร่วมกับการใช้กระบวนการแบบย้อนซ้ำวิธีฟิชเชอร์ สกอริง (Fisher's scoring method) (HLM for Windows. Version 6.02a) หรือ เรียกว่า Gauss-Newton ใช้ค่าคาดหวัง (Expected) ของฟังก์ชันความเป็นไปได้ (Likelihood function) มาคำนวณ ซึ่งเป็นวิธีที่ลู่ออกหาค่าพารามิเตอร์ได้อย่างรวดเร็ว (Andrew, Schwoerer.; & Peter Hovey. 2004: Online.) ดังนั้นในกรณีที่สมการเป็นเส้นตรง การประมาณค่าพารามิเตอร์ของโมเดลจะไม่มีกระบวนการย้อนซ้ำ กล่าวคือ กระบวนการย้อนซ้ำจะมีเพียงรอบเดียว เท่านั้น ค่าประมาณที่ได้จึงเป็นค่าเดียวกันกับค่าที่ได้จากวิธีความเป็นไปได้สูงสุดโดยตรงหรืออาจใช้วิธีกำลังสองน้อยที่สุดแบบถ่วงน้ำหนัก (Raudenbush, S.W.; & Bryk, A.S. 2002: 448. กัลยา วาณิชย์บัญชา. 2546: 205) จากผลการประมาณค่าพารามิเตอร์ด้วยโมเดลพหุระดับโดยใช้โปรแกรม HLM ของข้อมูลที่ได้จากผลการตอบของนักศึกษา ภายใต้เงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่างจำนวน 12 ชุด พบว่าข้อมูลทั้งหมดทุกชุดมีการประมาณค่าที่ใช้กระบวนการย้อนซ้ำเพียงรอบเดียว ด้วยการเปลี่ยนแปลงของค่าฟังก์ชันความเป็นไปได้ (Likelihood function) มีค่าน้อยมาก แสดงว่าข้อมูลเชิงประจักษ์มีความสอดคล้องกลมกลืนกับโมเดลเชิงเส้นตรงเป็นอย่างมาก

วิธีประมาณค่าด้วยวิธีความเป็นไปได้สูงสุด เป็นวิธีที่พิจารณาการแจกแจงความน่าจะเป็นร่วมของกลุ่มตัวอย่างสุ่มที่ได้เป็นเกณฑ์ในการหาตัวประมาณค่า และพบข้อบกพร่องในการประมาณค่าวิธีนี้ 2 ประการ คือ ประการแรก ไม่สามารถประมาณค่าพารามิเตอร์ได้ในกรณีที่ผู้สอบตอบข้อสอบถูกต้องหรือผิดหมด ประการที่สอง คือ บางครั้งพบว่ามีค่าความเป็นไปได้สูงสุดมากกว่าหนึ่งค่า ซึ่งสอดคล้องกับผลการวิจัยของ ฮามเบิลตัน และ สวามินาธาน (Hambleton; & Swaminathan. 1985: 87; citing Semijima. 1973) จึงไม่

สามารถประมาณค่าพารามิเตอร์ได้ ส่วนวิธี d_0 , D_0 และ RD_0 ประมาณค่าด้วยวิธีเบส แบบ EAP ซึ่งเป็นวิธีที่พิจารณาการแจกแจงความน่าจะเป็นของพารามิเตอร์ โดยการทำให้ฟังก์ชันการแจกแจงภายหลังมีค่าสูงสุด จึงมีข้อเด่นในการประมาณค่าพารามิเตอร์ได้ในกรณีที่ผู้สอบตอบข้อสอบถูกหมดหรือผิดหมด ด้วยเหตุผลดังกล่าวข้างต้นจึงทำให้คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี RD_0 มีความสัมพันธ์กับวิธี HLM น้อยกว่าวิธี S

สำหรับค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่าจากวิธี d_0 กับวิธี HLM มีค่าต่ำสุดในทุกเงื่อนไข สอดคล้องกับผลการวิจัยของ เมย์ และไนส์วันเดอร์ (May; & Nicewander. 1998: 882-897) ที่พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธีความแตกต่างระหว่างความสามารถแฝงนั้นจะบิดเบือนไปในกลุ่มผู้สอบที่มีความสามารถอ่อนที่สอบด้วยแบบทดสอบที่ยาก โดยจะประมาณค่าคะแนนการเปลี่ยนแปลงที่ต่ำกว่าที่ควรจะเป็นมากหรืออาจเนื่องมาจากคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี d_0 นี้ มีช่วงของความแตกต่างน้อยกว่าวิธีอื่นๆ จึงส่งผลให้มีค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันต่ำ ทั้งนี้อาจมีการศึกษาเพิ่มเติมเกี่ยวกับการศึกษาค่าสหสัมพันธ์อันดับของสเปียร์แมน (Spearman) เนื่องจากการแปลงคะแนนผู้สอบแต่ละคนให้เป็นอันดับ

1.2.1 เมื่อความยาวของแบบทดสอบเพิ่มขึ้นในทุกขนาดกลุ่มตัวอย่าง แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี RD_0 กับวิธี HLM มีค่าใกล้เคียงกัน ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัยข้อ 2.1 ที่ตั้งไว้ว่า แบบทดสอบที่มีจำนวนข้อสอบมากจะมีค่าความสัมพันธ์สูงกว่าแบบทดสอบที่มีจำนวนข้อน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง ทั้งนี้อาจเนื่องมาจากแบบทดสอบจำนวน 40 ข้อ ที่ใช้ในการวิจัยในครั้งนี้ ผู้วิจัยได้สร้างมาจากลักษณะเฉพาะของข้อสอบที่ได้ระบุไว้อย่างชัดเจน ทำให้ได้แบบทดสอบที่เป็นตัวแทนที่ดีของมวลความรู้ ในขณะที่แบบทดสอบที่มีความยาว 30 และ 20 ข้อ นั้น ผู้วิจัยได้ใช้วิธีการสุ่มแบบแบ่งชั้น โดยใช้มวลความรู้ย่อยเป็นชั้น และข้อสอบเป็นหน่วยในการสุ่ม จึงทำให้เป็นแบบทดสอบของมวลความรู้และเป็นตัวแทนที่ดีของมวลความรู้เช่นเดียวกันกับแบบทดสอบที่มีจำนวน 40 ข้อ สอดคล้องกับคำกล่าวของ ฮาลาดินาและรอยด์ (Haladyna; & Roid. 1983: 280) ที่กล่าวไว้ว่า แบบทดสอบที่มีจำนวน 20-30 ข้อ มีความเหมาะสมในการประมาณค่าคะแนนมวลความรู้มากที่สุดและให้ค่าการประมาณความสามารถที่แม่นยำ แต่ถ้าเพิ่มจำนวนข้อสอบมากกว่า 30 ข้อ จะทำให้สิ้นเปลืองค่าใช้จ่ายสูงเกินไป และสอดคล้องกับผลการวิจัยของ ฮัทเทน (Hutten. 1981: 4799-A) ที่พบว่า ค่าความสามารถของผู้สอบจำนวน 1,000 คน ที่ประมาณค่าด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์ ระหว่างแบบทดสอบที่มีความยาว 40 ข้อ กับ 20 ข้อ มีความสัมพันธ์กันสูง ($r=0.87$)

1.2.2 เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น แนวโน้มค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี RD_0 กับวิธี HLM มีค่าใกล้เคียงกัน ซึ่งไม่สอดคล้องกับสมมติฐาน

ของการวิจัยข้อ 2.2 ที่ตั้งไว้ว่า กลุ่มตัวอย่างที่มีจำนวนผู้ตอบมากจะมีค่าความสัมพันธ์สูงกว่ากลุ่มตัวอย่างที่มีจำนวนผู้ตอบน้อยกว่า ในทุกความยาวของแบบทดสอบ ผลการวิจัยนี้สามารถอธิบายได้ด้วยเหตุผลเช่นเดียวกับข้อ 1.1.2

2. ผลการศึกษาประสิทธิภาพของการวัดการเปลี่ยนแปลง

ผู้วิจัยแบ่งการศึกษาศักยภาพประสิทธิภาพของการวัดการเปลี่ยนแปลงแต่ละวิธีออกเป็น 2 ประเด็น คือ 1) การศึกษาความเชื่อมั่นของการวัดการเปลี่ยนแปลง และ 2) การศึกษาความเที่ยงตรงของการวัดการเปลี่ยนแปลง ดังนี้

2.1 ความเชื่อมั่นของการวัดการเปลี่ยนแปลง

การศึกษาศักยภาพความเชื่อมั่นของการวัดการเปลี่ยนแปลงแต่ละวิธีในครั้งนี้ พิจารณาจากค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ พบว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธี S มีค่าต่ำสุดและมีค่าเท่ากับศูนย์ ทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ส่วนวิธี d_0 มีค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) สูงสุดทุกเงื่อนไข ซึ่งไม่สอดคล้องกับกับสมมติฐานของการวิจัยข้อ 3 ที่ตั้งไว้ว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำกว่าวิธีอื่นและเข้าใกล้ศูนย์มากที่สุด อาจเนื่องมาจากกระบวนการประมาณค่าความสามารถของแต่ละวิธีการวัดคะแนนการเปลี่ยนแปลง ดังที่กล่าวมาแล้วในข้อ 1

2.1.1 เมื่อความยาวของแบบทดสอบเพิ่มขึ้นในทุกขนาดกลุ่มตัวอย่าง แนวโน้มค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากภายในแต่ละวิธีมีค่าใกล้เคียงกัน ยกเว้น วิธี d_0 มีแนวโน้มค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เพิ่มขึ้นเล็กน้อย ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัยข้อ 3.1 ที่ตั้งไว้ว่า แบบทดสอบที่มีจำนวนข้อสอบมากจะมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำกว่าแบบทดสอบที่มีจำนวนข้อสอบน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง ผลการวิจัยนี้สามารถอธิบายได้ด้วยเหตุผลเช่นเดียวกับข้อ 1.2.1

2.1.2 เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้นในทุกความยาวของแบบทดสอบ แนวโน้มค่ารากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) ของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากภายในแต่ละวิธีมีค่าใกล้เคียงกัน ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัยข้อ 3.2 ที่ตั้งไว้ว่า กลุ่มตัวอย่างที่มีจำนวนผู้ตอบมากจะมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำกว่ากลุ่มตัวอย่างที่มีจำนวนผู้ตอบที่น้อยกว่า ในทุกความยาวของแบบทดสอบ ผลการวิจัยนี้สามารถอธิบายได้ด้วยเหตุผลเช่นเดียวกับข้อ 1.1.2

2.2 ความเที่ยงตรงของการวัดการเปลี่ยนแปลง

การศึกษาความเที่ยงตรงของการวัดการเปลี่ยนแปลงแต่ละวิธีในครั้งนี้ พิจารณาจากค่าความลำเอียงของคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดการเปลี่ยนแปลงแต่ละวิธีที่สนใจศึกษากับวิธีโมเดลพหุระดับ พบว่า ทุกวิธีมีค่าเป็นศูนย์ทุกเงื่อนไข ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัยข้อ 4 ที่ตั้งไว้ว่า คะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ มีค่าความลำเอียงต่ำกว่าวิธีอื่นและเข้าใกล้ศูนย์มากที่สุด อาจเนื่องมาจากกระบวนการประมาณค่าความสามารถของแต่ละวิธีการวัด ดังที่กล่าวมาแล้วในข้อ 1 โดยผลการวิจัยไม่สอดคล้องกับสมมติฐานของการวิจัยข้อ 4.1 ที่ตั้งไว้ว่า แบบทดสอบที่มีจำนวนข้อสอบมากจะมีค่าความลำเอียงต่ำกว่าแบบทดสอบที่มีจำนวนข้อสอบน้อยกว่า ในทุกขนาดกลุ่มตัวอย่าง และ สมมติฐานของการวิจัยข้อ 4.2 ที่ตั้งไว้ว่า กลุ่มตัวอย่างที่มีจำนวนผู้สอบมากจะมีค่าความลำเอียงต่ำกว่ากลุ่มตัวอย่างที่มีจำนวนผู้สอบที่น้อยกว่า ในทุกความยาวของแบบทดสอบ ผลการวิจัยนี้สามารถอธิบายได้ด้วยเหตุผลเช่นเดียวกับข้อ 1.1.2 และ 1.2.1

ข้อเสนอแนะ

จากการศึกษาผลการวัดการเปลี่ยนแปลงหาวิธีแบบวัดซ้ำสองครั้ง ภายใต้เงื่อนไข ความยาวของแบบทดสอบต่างกันและขนาดกลุ่มตัวอย่างไม่เท่ากัน ผู้วิจัยมีข้อเสนอแนะในการนำผลการวิจัยไปใช้และสำหรับการวิจัยครั้งต่อไป ดังนี้

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

ผลการวิจัยครั้งนี้ชี้ให้เห็นว่าการสอบสองครั้ง ซึ่งมีแบบแผนการเปลี่ยนแปลงเป็นเส้นตรงเสมอ หรือ การสอบมากกว่าสองครั้งที่แบบแผนการเปลี่ยนแปลงมีแนวโน้มเป็นเส้นตรง ดังเช่น การสอบเกี่ยวกับทักษะการคิดคำนวณ เป็นต้น การวัดการเปลี่ยนแปลงด้วยวิธีการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ (Relative change scores; S) เป็นวิธีที่มีความเหมาะสมที่สุด เพราะสามารถประมาณค่าคะแนนการเปลี่ยนแปลงที่มีประสิทธิภาพดีในทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง เมื่อเทียบกับการวัดการเปลี่ยนแปลงด้วยวิธีโมเดลพหุระดับ แบบสองระดับ ที่มีการสอบสามครั้ง กล่าวคือ วิธีการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์เป็นวิธีที่มีความเที่ยงตรงสูงและความเชื่อมั่นสูงในการประมาณค่าคะแนนการเปลี่ยนแปลง เนื่องจากมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำกว่าวิธีอื่น มีค่าความลำเอียงเป็นศูนย์ ทุกเงื่อนไข และมีระดับความสอดคล้องในการประมาณค่าสูง คือให้ค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันของคะแนนการเปลี่ยนแปลงระหว่างคะแนนที่ประมาณค่ามาจากการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์กับวิธีโมเดลพหุระดับอยู่ในระดับสูง ทุกเงื่อนไข และเนื่องจากการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์มีจุดเด่นที่ใช้คะแนนดิบในการคำนวณ จึงคำนวณง่าย ไม่ซับซ้อน ง่ายต่อ

การเข้าใจความหมายของคะแนนการเปลี่ยนแปลง ตลอดจนเป็นวิธีที่สามารถลดอิทธิพลคะแนนสูงสุด (Ceiling effect) และอิทธิพลพื้นความรู้เดิม (Floor effect) ได้ จึงควรนำวิธีการวัดการเปลี่ยนแปลงดังกล่าวมาใช้วัดพัฒนาการหรือการเปลี่ยนแปลงรายบุคคลเมื่อได้รับการเรียนรู้ในชั้นเรียน หรือการฝึกอบรมแล้ว

การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ใช้ได้ในทุกความยาวของแบบทดสอบคือ ทั้งแบบทดสอบที่มีจำนวนน้อยข้อและมากข้อ ดังเหตุผลที่กล่าวข้างต้น ซึ่งในการกำหนดจำนวนข้อสอบว่าควรมีจำนวนกี่ข้อนั้นควรพิจารณาความสำคัญของมวลความรู้ในเนื้อหาวิชา แต่ทั้งนี้ผู้สนใจต้องคำนึงถึงความจำเป็นในทางปฏิบัติที่สำคัญ คือ เวลาที่ใช้สร้างข้อสอบ เวลาที่ใช้ดำเนินการสอบ และลักษณะของผู้สอบ เช่น ความสามารถของผู้สอบ เป็นต้น อย่างไรก็ตามแบบทดสอบที่มีจำนวนข้อน้อยมากอาจส่งผลให้เกิดอิทธิพลคะแนนสูงสุด ดังนั้นจึงควรใช้แบบทดสอบที่มีจำนวนข้อพอประมาณ โดยแบบทดสอบที่ใช้สอบต้องเป็นตัวแทนที่ดีของประชากรคุณสมบัติ (Attributes) ของพฤติกรรมผู้เรียน เช่น ความรู้ คือ เป็นแบบทดสอบที่เป็นตัวแทนที่ดีของประชากรความรู้ สามารถวัดเนื้อหาได้ครอบคลุมตามที่ต้องการ เช่น แบบทดสอบอิงมวลความรู้ ซึ่งเป็นแบบทดสอบที่สร้างมาจากลักษณะเฉพาะของมวลความรู้ที่สุ่มมาจากประชากรของข้อสอบ เป็นต้น

การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ยังสามารถนำไปใช้ได้เหมาะสมกับทุกขนาดกลุ่มตัวอย่าง ทั้งขนาดเล็ก ปานกลาง และใหญ่ นั้นแสดงว่า การวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์สามารถนำไปใช้ในการวัดคะแนนการเปลี่ยนแปลงหรือคะแนนพัฒนาการได้ทั้งเป็นรายบุคคล ห้องเรียน ชั้นเรียน โรงเรียน/สถานศึกษา หรือกลุ่มโรงเรียน เป็นต้น เพื่อเป็นสารสนเทศต่อครูผู้สอน ผู้บริหารสถานศึกษา ในการพัฒนา ปรับปรุง การจัดกิจกรรมการเรียนรู้ให้ผู้เรียนเกิดการเรียนรู้ได้อย่างเต็มศักยภาพและการบริหารงานอันนำไปสู่การประกันคุณภาพการศึกษาของสถานศึกษา

อย่างไรก็ตาม การนำการคะแนนการเปลี่ยนแปลงสัมพัทธ์ไปใช้ต้องคำนึงถึงปฏิสัมพันธ์ที่เกิดจากคุณลักษณะทางสถิติรายข้อกับความสามารถของผู้สอบ ผู้สนใจจึงควรใช้แบบทดสอบที่มีความยากระดับที่เหมาะสมกับความสามารถของกลุ่มตัวอย่าง หากใช้แบบทดสอบที่ง่ายเกินไปกับความสามารถของผู้สอบแล้วอาจทำให้ผู้สอบได้คะแนนสอบก่อนเรียนเต็ม ซึ่งไม่สามารถคำนวณคะแนนการเปลี่ยนแปลงด้วยวิธีการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์ได้ แต่ถ้าหากต้องการวัดคะแนนการเปลี่ยนแปลงรายบุคคลในกรณีคะแนนสอบก่อนเรียนของผู้สอบคนใดคนหนึ่งได้คะแนนเต็มแล้ว ผู้สนใจควรเลือกใช้การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ (Difference score; D) หรือ การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ (Difference domain score; D_0) หรือ การวัดคะแนนความแตกต่างระหว่างมวลความรู้สัมพัทธ์ (Relative difference score; RD_0) เนื่องจากเป็นวิธีที่มีความเที่ยงตรงสูงและความเชื่อมั่นสูงรองลงมา ซึ่งทั้งสามวิธีนี้ให้ผลการประมาณค่าคะแนนการเปลี่ยนแปลงที่ใกล้เคียงกัน แต่ทั้งนี้การวัดคะแนนความแตกต่างระหว่างคะแนนดิบ จะเป็นวิธีที่ใช้คะแนนดิบรวม

คำนวณคะแนนการเปลี่ยนแปลง ง่ายในการคำนวณและง่ายต่อการเข้าใจความหมายคะแนนการเปลี่ยนแปลง แต่ก็มีจุดอ่อนคือ มีทั้งอิทธิพลคะแนนสูงสุดและอิทธิพลพื้นความรู้เดิมเช่นเดียวกับการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ แต่วิธีการวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ต้องใช้ความรู้ความสามารถเฉพาะทางด้านโปรแกรมสำเร็จรูปในการคำนวณ ในขณะที่การวัดคะแนนความแตกต่างระหว่างมวลความรู้สัมพัทธ์ต้องใช้ความรู้ความสามารถเฉพาะทางด้านโปรแกรมสำเร็จรูปในการคำนวณเช่นกัน แต่เป็นวิธีที่สามารถแก้ไขปัญหาดังกล่าวได้ ดังนั้นถ้าผู้สนใจต้องการวัดคะแนนการเปลี่ยนแปลงในกรณีที่มีการสอบครั้งนั้นมีผู้สอบได้คะแนนก่อนเรียนเต็ม คำนึงถึงอิทธิพลคะแนนสูงสุดและอิทธิพลพื้นความรู้เดิมแล้ว การวัดคะแนนความแตกต่างระหว่างมวลความรู้สัมพัทธ์ เป็นวิธีที่เหมาะสมในการวัดคะแนนการเปลี่ยนแปลงมากที่สุด

2. ข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป

ผลการวิจัยครั้งนี้ชี้ให้เห็นถึงการวัดการเปลี่ยนแปลงสามประเด็นใหญ่ๆ คือ วิธีการวัดคะแนนการเปลี่ยนแปลง ความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง ซึ่งผู้วิจัยขอเสนอแนะสำหรับผู้สนใจที่จะศึกษาเกี่ยวกับเรื่องการวัดการเปลี่ยนแปลงต่อไปนี้

2.1 วิธีการวัดคะแนนการเปลี่ยนแปลง มีแนวโน้มว่าการวัดคะแนนการเปลี่ยนแปลงสัมพัทธ์เป็นวิธีการวัดการเปลี่ยนแปลงที่มีประสิทธิภาพทุกเงื่อนไขความยาวของแบบทดสอบและขนาดกลุ่มตัวอย่าง เนื่องด้วยระดับความสอดคล้องของคะแนนการเปลี่ยนแปลงกับวิธีโมเดลพหุระดับมีค่าสูงสุดคือ มีค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สันของทุกเงื่อนไขอยู่ระหว่าง 0.83-0.97 มีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าเข้าใกล้ศูนย์มากที่สุด คือมีค่าอยู่ระหว่าง 0.25-0.59 รองลงมาเป็นคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธีการวัดคะแนนความแตกต่างระหว่างคะแนนดิบ การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้ การวัดคะแนนความแตกต่างระหว่างคะแนนมวลความรู้สัมพัทธ์ ซึ่งแนวโน้มมีระดับความสอดคล้องของคะแนนการเปลี่ยนแปลงกับวิธีโมเดลพหุระดับใกล้เคียงกันทุกเงื่อนไข ในขณะที่การวัดคะแนนการเปลี่ยนแปลงความสามารถแฝงมีระดับความสอดคล้องของคะแนนการเปลี่ยนแปลงกับวิธีโมเดลพหุระดับต่ำสุด และมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าสูงสุด ทุกเงื่อนไข

ประเด็นสำคัญอยู่ที่วิธีประมาณค่าคะแนนมวลความรู้ที่ใช้ในการศึกษาครั้งนี้เป็นวิธีของแบบ EAP (Expected a posteriori) แต่เนื่องจากวิธีประมาณค่าคะแนนการเปลี่ยนแปลงโมเดลพหุระดับด้วยโปรแกรม HLM เป็นวิธีประมาณค่าคะแนนการเปลี่ยนแปลงที่ใช้เป็นเกณฑ์ในการศึกษาครั้งนี้ ซึ่งใช้วิธี ML (Maximum likelihood) ดังนั้นควรปรับเปลี่ยนวิธีประมาณค่าคะแนนมวลความรู้เป็น ML หรือปรับเปลี่ยนวิธีการวัดคะแนนการเปลี่ยนแปลงเป็นวิธีอื่นที่มีการประมาณค่าแบบเบส หรืออาจเปรียบเทียบคะแนนการเปลี่ยนแปลงที่ประมาณค่ามาจากวิธีการที่มีแนวคิดการเปลี่ยนแปลงค่าพารามิเตอร์ความ

สามารถกับค่าพารามิเตอร์ความยากของข้อสอบ และตัวแปรที่น่าสนใจอีกตัวหนึ่ง คือ แบบแผนการเปลี่ยนแปลง ซึ่งอาจไม่เป็นสมการเส้นตรงเช่นเดียวกับการศึกษาในครั้งนี้ เช่น สมการควอดราติก (Quadratic) สมการคิวบิก (Cubic) เป็นต้น

2.2 ความยาวของแบบทดสอบ เมื่อความยาวของแบบทดสอบเพิ่มมากขึ้น มีแนวโน้มว่า ระดับความสอดคล้องของคะแนนการเปลี่ยนแปลงไม่แตกต่างกัน ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าใกล้เคียงกันและค่าความลำเอียงมีค่าเป็นศูนย์ทุกเงื่อนไข ประเด็นสำคัญอยู่ที่ความยากของข้อสอบในแบบทดสอบ ควรมีลักษณะการแจกแจงเป็นเช่นไร เช่น เป็นแบบปกติ (Normal) แบบสม่ำเสมอ (Uniform) เป็นต้น และตัวแปรที่น่าสนใจอีกตัวหนึ่ง คือ ลักษณะข้อสอบ ซึ่งในการศึกษาครั้งนี้ เป็นแบบทดสอบอิงมวลความรู้ที่สร้างข้อสอบวัดมิติเดียว (Unidimensional) และให้คะแนนแบบทวิภาค (Dichotomous) หากใช้ข้อสอบที่สร้างแบบทดสอบอิงกลุ่ม หรือเป็นข้อสอบที่วัดหลายมิติ หรือให้คะแนนบางส่วน (Partial credit model) เป็นต้น

2.3 ขนาดกลุ่มตัวอย่าง เมื่อขนาดกลุ่มตัวอย่างที่เพิ่มมากขึ้น มีแนวโน้มว่า ระดับความสอดคล้องของคะแนนการเปลี่ยนแปลงไม่แตกต่างกัน ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าใกล้เคียงกันและค่าความลำเอียงมีค่าเป็นศูนย์ทุกเงื่อนไข ประเด็นสำคัญอยู่ที่ความสามารถเริ่มต้นของผู้สอบ ควรมีลักษณะการแจกแจงเป็นเช่นไร เช่น เป็นแบบปกติ แบบเบ้ซ้าย หรือ แบบเบ้ขวา หรือลักษณะของกลุ่มตัวอย่างควรแบ่งเป็นกลุ่มที่มีความสามารถต่างกัน เช่น เก่ง ปานกลาง และอ่อน ดังนั้นจึงควรจะได้มีการการวิจัยในเรื่องนี้ต่อการจำลองข้อมูลและข้อมูลเชิงประจักษ์

บรรณานุกรม

บรรณานุกรม

- กัลยา วาณิชยปัญญา. (2546). *การวิเคราะห์สถิติขั้นสูงด้วย SPSS for Windows*. พิมพ์ครั้งที่ 3. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- (2547). *หลักสถิติ*. พิมพ์ครั้งที่ 7. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- นางลักษณ์ วิรัชชัย. (2542). *โมเดลลิสเรล: สถิติวิเคราะห์สำหรับการวิจัย*. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- บุญเชิด ภิญโญนนตพงษ์. (2526). *การทดสอบแบบอิงเกณฑ์: แนวคิดและวิธีการ*. เอกสารประกอบคำบรรยายลำดับที่ 42. กรุงเทพฯ: ภาควิชาพื้นฐานการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- (2535, มกราคม-เมษายน). *การวิเคราะห์ข้อสอบอิงจุดประสงค์ชนิดครูสร้าง: วิธีการใหม่. การวัดผลการศึกษา*. 13(39): 18-19.
- ประสิทธิ์ ไชยกาล. (2539). *การเปรียบเทียบประสิทธิภาพระหว่างโมเดลลิสเรล 3 แบบที่ใช้ในการศึกษาตัวแปรที่สัมพันธ์กับการเปลี่ยนแปลงในระยะยาวของผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์*. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ผจงจิต อินทสุวรรณ. (2534). *การวิเคราะห์แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนด้วยโมเดลโลจิสติก*. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- มนต์ทิศา ไชยแก้ว. (2542). *การเปรียบเทียบผลการเปลี่ยนแปลงระยะยาวโดยใช้โมเดลประยุกต์โค้งพัฒนาการที่มีตัวแปรแฝงเมื่ออัตราการขาดหายของข้อมูล ช่วงเวลาการวัดและจำนวนครั้งที่วัดแตกต่างกัน*. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ลอร์ด เฟรดเดอริก เอ็ม. (2545). *ทฤษฎีการตอบข้อคำถาม*. แปลโดย ผจงจิต อินทสุวรรณ. กรุงเทพฯ: สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- วินิจ เทือกทอง. (2537). *การศึกษาเปรียบเทียบประสิทธิภาพของการคำนวณคะแนนเพิ่มวิธีต่างๆ ด้วยระเบียบวิธีกรมอนติคาร์โล*. วิทยานิพนธ์ กศ.ด. (วิจัยและพัฒนาหลักสูตร). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- วีระศักดิ์ คำล้าน. (2540). *การวิเคราะห์การเปลี่ยนแปลงในระยะยาวของผลสัมฤทธิ์ทางการเรียนคำศัพท์ภาษาอังกฤษ: การประยุกต์ใช้โมเดลเชิงเส้นพหุระดับ*. วิทยานิพนธ์ ค.ม. (วิจัยการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

- ศิริชัย กาญจนวาสี. (2538). การวัดการเปลี่ยนแปลง. การประชุมเชิงปฏิบัติการครั้งที่ 3 เรื่อง “หลักและวิธีวิจัยขั้นสูงเฉพาะการวิจัยและพัฒนากระบวนการไทยด้านต่างๆ” .
สำนักงานคณะกรรมการวิจัยแห่งชาติร่วมกับคณะกรรมการแห่งชาติเพื่อการวิจัยและพัฒนากระบวนการไทย. 144-149. กรุงเทพฯ: เดอะเบสท์ กราฟิค ดีไซน์.
- . (2548). การวิเคราะห์พระดุษฎี. พิมพ์ครั้งที่ 3. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี; ทวีวัฒน์ ปิตยานนท์; และ ดิเรก ศรีสุขโข. (2544). การเลือกใช้สถิติที่เหมาะสมสำหรับการวิจัย. พิมพ์ครั้งที่ 3. กรุงเทพฯ: บุญศิริ.
- สมถวิล วิจิตรวรรณ. (2543). การเปรียบเทียบประสิทธิภาพของโมเดลโค้งพัฒนาการที่มีตัวแปรแฝงโมเดลพระดุษฎีและโมเดลกึ่งซิมเพิลทิคซ์ ในการวัดการเปลี่ยนแปลงระยะยาว ชนิดตัวแปรเดียวและตัวแปรพหุ. วิทยานิพนธ์ ค.ด. . (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- สุชาติ กิระนันท์. (2542). ทฤษฎีและวิธีการสำรวจตัวอย่าง. พิมพ์ครั้งที่ 2. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- . (2545). การอนุมานเชิงสถิติ: ทฤษฎีขั้นต้น. พิมพ์ครั้งที่ 3. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.
- สุภารัตน์ เรือจันทิก. (2542). การวิเคราะห์การเปลี่ยนแปลงระยะยาวของผลสัมฤทธิ์การเรียนรู้สิ่งแวดล้อม: การเปรียบเทียบผลการวิเคราะห์ด้วยโมเดลเอชแอลเอ็มและโมเดลลิสเรล. วิทยานิพนธ์ ค.ม. (วิจัยการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- อรุณี อ่อนสวัสดิ์. (2537). การพัฒนาวิธีวัดการเปลี่ยนแปลงการเรียนรู้. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- . (2538, มกราคม – เมษายน). พัฒนาการการวัดผลการเปลี่ยนแปลงการเรียนรู้. การวัดผลการศึกษา. 16: 1-15.
- อวยพร เรืองตระกูล. (2544). การพัฒนาและวิเคราะห์ของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- อิทธิพงษ์ ตั้งสกุลเรืองไธ. (2541). การเปรียบเทียบประสิทธิภาพของโมเดลโค้งพัฒนาการที่มีตัวแปรแฝง 4 รูปแบบ ในการศึกษาการเปลี่ยนแปลงระยะยาวของการพัฒนาทางกายและผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนประถมศึกษา. วิทยานิพนธ์ ค.ม. (วิจัยการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

เอื้อมพร หลินเจริญ. (2539). *การพัฒนาโมเดลลิสเรลในการศึกษาการเปลี่ยนแปลงคุณลักษณะทางวิทยาศาสตร์ของนักเรียน*. วิทยานิพนธ์ ค.ม. (วิจัยการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

Allen, Mary J.; & Yen, Wendy M. (1979). *Introduction to Measurement Theory*. California: Brooks/Cole.

Analysis of Pretest and Posttest Scores with Gain Scores and Repeated Measures. Retrieved September 30, 2005, from <http://www.uccs.edu/lbecker/Psy590/gainscore.htm>.

Anderson, John O. (1998, October). Does Complex Analysis (IRT) Pay and Dividends in Achievement Testing?. Retrieved August 10, 2005, from <http://www.educ.uvic.ca/faculty/anderson/Paper2.doc>.

Andrew, Schworer.; & Peter Hovey. (2004). *Newton-Raphson Versus Fisher Scoring Algorithms in Calculating Maximum Likelihood Estimates*. Retrieved May 22, 2007, from [://www.udayton.edu/~mathdept/EPUMD/EPMUD.htm](http://www.udayton.edu/~mathdept/EPUMD/EPMUD.htm).

Archwamety, Teara.; Tangdhanakanond, Kamonwan.; & Pitiyanuwat, Somwung. (2005, September-December). Learning and Development Curves: Their Implications on Measuring Gains. *Journal of Research Methodology*. 18(3): 213-227.

Arvedson, Paula. (2001, June). Questions, Answers, and Comments. Retrieved November 30, 2005, from [://www.WebQnA.htm](http://www.WebQnA.htm).

Berk, Ronald A. (1980). Estimation of Test Length for Domain-Referenced Reading Comprehension Tests. *Journal of Experimental Education*. 48: 188-193.

Bock, R. Darrell.; Thissen, David.; & Zimowski, Michele F. (1997, Fall). IRT Estimation of Domain Scores. *Journal of educational measurement*. 34(3): 197-211.

Bryk, Anthony S.; & Raudenbush, Stephen W. (1987). Application of hierarchical linear models to assessing change. *Psychological Bulletin*. 101: 147-158.

----- (2002). *Hierarchical Linear Models : Applications and Data Analysis Methods*. Second Edition. California: Sage Publications. Inc.

Burchinal, Margaret.;& Appelbaum, Mark I. (1991). Estimating individual developmental functions: Methods and their assumptions. *Child Development*. 62: 23-43.

- Chou C.P.; Bentler P.M.; & Pentz M.A. (1998). Comparisons of two statistical approaches to study growth curves: The multilevel and the latent curve analysis. *Structural Equation Modeling*. 5(3): 247-266.
- Collins, Linda M. (1996, September). Is Reliability Obsolete? A Commentary on "Are Simple Gain Scores Obsolete?". *Applied Psychological Measurement*. 20(3): 289-292.
- Collins L.M.; & Horn J.L. (1995). *Best Methods for the Analysis of Change*. Washington DC: American Psychological Association Books.
- Cronbach. L.J.; & Furby. L. (1970). How should we measure "change"---or should we?. *Psychological Bulletin*. 74: 68-80.
- Embretson, Susan E. (1991, September). Multidimensional latent trait Model For Measuring Learning and Change. *Psychometrika*. 56(3): 495-515.
- (1995). A measurement model for linking individual learning to processes and knowledge: Application to mathematical research. *Journal of Educational Measurement*. 32 : 277-294.
- (1996). Multicomponent Response Models. In *Handbook of Modern Item Response Theory*. pp. 305-322. New York: Springer.
- Fischer, Gerhard H. (2001). Gain scores revisited under an IRT perspective. *Essays on item response theory*. pp.43-68. New York: Springer-Verlag.
- (1995, December). Some Neglected Problems in IRT. *Psychometrika*. 60(4): 459-487.
- (1996). Unidimensional Linear Logistic Rasch Models. In *Handbook of Modern Item Response Theory*. pp. 225-244. New York: Springer
- (2003, January). The Precision of Gain Scores Under an Item Response Theory Perspective : A Comparison of Asymptotic and Exact Conditional Inference About Change. *Applied Psychological Measurement*, 27(1): 3-26.
- Fischer,G.H.; & Parzer,P. (1991). An extension of the rating scale model with an application to measurement of change. *Psychometrika*. 56: 637-651.
- Fischer,G.H.; & Ponocny,I. (1994,June). An extension of the partial credit model with and application to the measurement of change. *Psychometrika*. 59(2): 177-192.
- Fischer,G.H.; & Prieler,J.A. (2000). An IRT methodology for the assessment of change. *Standard Progressive matrices*. pp.113-121. Oxford Psychologists Press.

- Fischer, G.H.; & Seliger, Elisabeth. (1996). Multidimensional Linear Logistic Models for Change. In *Handbook of Modern Item Response Theory*. pp. 323-346. New York: Springer.
- Glück, Judith.; & Spiel, Christiane. (1997). Item response models for repeated measures designs: Application and limitations of four different approaches. Retrieved May 19, 2005, from <http://www.mpr-online.de/issue2/art6/article.html>
- Gottman, J.M. (1996, September). The Analysis of Change. *Applied Psychological Measurement*. 20(3): 299-302.
- Gulliksen, Harold. (1950). *Theory of Mental Tests*. New York: John Wiley & Sons.
- Haladyna, Thomas M. (1983, Fall). A Comparison of Two Approaches to Criterion-Referenced Test Construction. *Journal of Educational Measurement*. 20(3): 271-282.
- Hambleton, Ronald K.; & Gruijter, Dato N. M. De. (1983, Winter). Application of Item Response Models to Criterion-Referenced Test Item Selection. *Journal of Educational Measurement*. 20(4): 355-367.
- Hambleton, Ronald K.; et al. (1978, Winter). Criterion-Referenced Testing and Measurement: A review of Technical Issues and Developments. *Review of Educational Research*. 48(1): 1-47.
- Hambleton, Ronald K.; & Linda L. Cook. (1977, Summer). Latent Trait Models and their Use in the Analysis of Educational Test Data. *Journal of Educational Measurement*. 14(2): 75-96.
- Hambleton, Ronald K.; & Swaminathan, Hariharan. (1985). *Item Response Theory : Principles and Applications*. Volume 2. Boston: Kluwer-Nijhoff.
- Hambleton, Ronald K.; Swaminathan, H.; & Roger, H. Jane. (1991). *Functionals of Item Response Theory*. London: Sage.
- Hansen, Julie. (2005). Project 4. Retrieved November 30, 2005, from <http://www.growth%20reference\Project%204>.
- Heidenberg, Alex J. (2000). *Reducing Pretest and Gain Score Correlation Using Item Response Theory Domain Sources*. Dissertation, Ph.D. (Research, Measurement, and Statistics). Georgia: Georgia State University. Photocopied.
- Henning, G. (1987). G. Retrieved November 30, 2005, from <http://www.G.htm>.

- Hulin, Charles L.; Fritz Drasgow.; & Charles K. Parsons. (1983?) *Item Response Theory: Application to Psychological Measurement*. Homewood, Illinois: Dow Jones.
- Hutten, Leah. (1981, May). The fit of empirical data to two Latent Trait Model. *Dissertation Abstracts International*. 42: 4799-A.
- Kamata, Akihito. (2001, Spring). Item Analysis by the Hierarchical Generalized Linear Model. *Journal of educational measurement*. 38(1): 79-93.
- Kolen, Michael J.; Zeng. Lingjia.; & Hanson, Bradley A. (1996, Summer). Conditional Standard Error of Measurement for Scale Scores Using IRT. *Journal of Educational Measurement*. 33(2): 129-140.
- Kolen, Michael J.; & Nicewander, W.Alan. (1999). A Rationale for Defining Achievement Levels Using IRT-Estimated Domain Scores. *Applied Psychological Measurement*. 23(4): 347-362.
- Liedka, Raymond V. (1996). Re: problems using difference scores. Retrieved November 30, 2005, from <http://www.math.yorku.ca/Who/Faculty/Monette/Ed-stat/0481.html>.
- Lin, Miao-Hsiang.; & Hsiung, Chao A. (1994, September). Empirical Bays Estimates of Domain Scores Under Binomial and Hypergeometric Distributions for Test Scores. *Psychometrika*. 59(3): 331-359.
- Linn, Robert L.; & Slinde, Jeffrey A. (1977, Winter). The Determination of the Significance of Change Between Pre-and Posttesting Periods. *Review of Educational Research*. 47(1): 121-150.
- Little, Todd D.; Schnabel, Kai U.; & Baumert, Jürgen. (2000). *Modeling Longitudinal and Multilevel Data: practical issues, applied approaches, and specific examples*. Mahwal. Lawrence Erlbaum Associates. Inc.
- Lord, F.M.; & Melvin R. Novick. (1968). *Statistical Theories of Mental test Scores*. Reading, Massachusetts: Addison-Wesley.
- Lord, Frederic M. (1958). Further Problems in the Measurement of Growth. *Educational and Psychological Measurement*. XVIII(3): 437-451.
- (1980). *Applications of Item Response Theory to Practical Testing Problems*. New Jersey: Lawrence Erlbaum Associates.

- Luecht, Richard M.; & Hirsch, Thomas M. (1992, March). Item Selection Using an Average Growth Approximation of Target Information Functions. *Applied Psychological Measurement*. 16(1): 41-51.
- Ma, Xin; & Ma, Lingling. (2004, April). Modeling stability of growth between mathematics and science achievement during middle and high school. *Education Review*. 28(2):104-122.
- May, K.; & Nicewander, W.Alan. (1994, Winter). Reliability and Information Functions for Percentile Ranks. *Journal of Educational Measurement*. 31(4): 313-325.
- (1998). Measuring change conventionally and adaptively. *Educational and Psychological Measurement*. 58: 529-554.
- McCoach, D. Betsy.; et al. (2006). Growing Readers: A Hierarchical Linear Model of Children's Reading Growth During the First 2 Years of School. *Journal of Educational Psychology*. 98(1): 14-28.
- McNemar, Quinn. (1958). On Growth Measurement. *Educational and Psychological Measurement*. XVIII (1): 47-55.
- Mellenbergh, Gideon J. (1999, March). A Note on Simple Gain Score Precision. *Applied Psychological Measurement*. 23(1): 87-89.
- Meng, Xiao-Li.; Rosenthal, Robert.; & Rubin, Donald B. (1992). Comparing Correlated Correlation Coefficients. *Psychological Bulletin*. 111(1): 172-175.
- Nering, Michael L. (1995, June). The Distribution of Person Fit Using True and Estimated Person Parameters. *Applied Psychological Measurement*. 19(2): 121-129.
- Noonan, Brian W.; Boss, Marvin W.; & Gessaroli, Marc E. (1992, December). The Effect of Test Length and IRT Model on the Distribution and Stability of Three Appropriateness Indexes. *Applied Psychological Measurement*. 16(4): 345-352.
- O'Grince, Tammy. (2005). Project 4. Retrieved November 30, 2005, from <http://www.growth%20reference\Project%204>.
- Paula. Arvedson. (2001, May-June). Questions, Answers, and Comments. Module: Weeks 1 and 2. Course: EDEL 419, section 2. Retrieved November 30, 2005, from <http://www.Web.QnA.htm>.
- Pommerich, M.; Nicewander, W.A.; & Hanson, B.A. (1999, Fall). Estimating average domain scores. *Journal of educational measurement*. 36(3): 199-216.

- Prieler, Jörg A.; & Raven, John. (2002, 20 October). The Measurement of Change in Groups and Individuals, with Particular Reference to the Value of Gain Scores: A New IRT-based Methodology for the Assessment of Treatment Effects and Utilizing Gain Scores. WebPsychEmpiricist. Retrieved June 22, 2005, from http://www.wpe.info/papers_table.html.
- Raykov, Tenko. (1999, June). Are Simple Change Scores Obsolete? An Approach to Studying Correlates and Predictors of Change. *Applied Psychological Measurement*. 23(2): 120-126.
- Russell, M. (2000). Using Expected Growth Size Estimates to Summarize Test Score Changes. Practical Assessment, Research & Evaluation. Retrieved June 6, 2005, from <http://pareonline.net/getvn.asp?v=7&n=6>.
- (2001). Summarizing Change in Test Scores: Shortcomings of Three Common Methods. ERIC Digest Retrieved June 7, 2005, from <http://www.ericdigests.org/2001-2/test.html>.
- Samejima, Fumiko. (1994, September). Some Critical Observations of the Test Information Function as a Measure of Local Accuracy in Ability Estimation. *Psychometrika*. 59 (3): 307-329.
- Schulz, E.Matthew.; Kolen, Micheal, J.; & Nicewander, W,Alan. (1999, December). A Rationale for Defining Achievement Levels Using IRT-Estimated Domain Scores. *Applied Psychological Measurement*. 23(4): 347-362.
- Schulz, E.Matthew.; Lee, Won-Chan.; & Mullen, Ken. (2005, Spring). A Domain-level Approach to Describing Growth in Achievement.. *Journal of Educational Measurement*. 42(1): 1-26.
- Shim, M.K. (1995). A longitudinal model for the study of equity issues in mathematics education. *Dissertation Abstracts*. AAC 9543721 ProQuest.
- Spray, Judith A. (1990). One – Parameter Item Response Theory Models for Psychomotor Tests Involving Repeated, Independent Attempts. *Research Quarterly for Exercise and Sport*. 61(2): 162-168.
- Spiel, Christiane.; & Glück, Judith. (1998). Item Response Models for Assessing Change in Dichotomous Items. *International Journal of Behavioral Development*, 22(3): 517-536.
- Traub, Ross E. (1994). *Reliability for the Social Sciences : theory and applications*. Volume 3, pp. 126-137. California: Sage Publications, Inc.

- Ulrich, Rich. (1997). Difference scores?. Retrieved November 11, 2005, from <http://www.pitt.edu/~wpilib/statfaq/97anocha.html>.
- Urry, Vern W. (1977, Summer). Tailored Testing: A Successful Application of Latent Trait Theory. *Journal of Educational Measurement*. 14(2): 181-196.
- Van der Linden, Wim J.; & Hambleton, Ronald K. Models for Multiple Abilities or Cognitive Components. (1996). In *Handbook of Modern Item Response Theory*. pp. 221-223. New York: Springer.
- Warm, Thomas A. (1978, December). A Primer of Item Response Theory. *Technical Report 940279*. Oklahoma: U.S. Coast Guard Institute: 52-53
- Wang, Wen-Chung.; & Chyi-In, Wu. (2004). Gain Score in Item Response Theory as an Effect Size Measure. *Educational and Psychological Measurement*. 64(5): 758-780.
- Week 2-3 Introduction to Psycho-Educational Measurement. (2005). Retrieved November 30, 2005, from [http:// www.growth%20reference\Introduction](http://www.growth%20reference\Introduction).
- Willett, John B. (1989). Some Results on Reliability for the Longitudinal Measurement of Change: Implications for the Design of Studies of Individual Growth. *Educational and Psychological Measurement*. 49: 587-602.
- Willett, John B.; & Sayer, Aline G. (1994). Quantitative Methods in Psychology: Using Covariance Structure Analysis to Detect Correlates and Predictors of Individual Change Over Time. *Psychological Bulletin*. 116(2): 363-381.
- Williams, Richard H.; & Zimmerman, Donald W. (1996). Are Simple Gain Scores Obsolete?. *Applied Psychological Measurement*. 20(1): 59-69.
- Wolfe, E.W.; & Chiu C, W.T. (1997). Measuring pretest-posttest change with a Rasch Scale model. *Journal Outcome Meas*. 3(2): 134-161.

ภาคผนวก

ภาคผนวก ก
ผลการวิเคราะห์ข้อมูล

- ค่าพารามิเตอร์ของข้อสอบเป็นรายชื่อของแบบทดสอบ ตามทฤษฎีการตอบข้อสอบ
- ค่าสถิติพื้นฐานของคะแนนสอบซ้ำสามครั้ง
- การวิเคราะห์แบบแผนการเปลี่ยนแปลง
- การวิเคราะห์โมเดลพหุระดับ 3 แบบ
- ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงแต่ละวิธี
- การทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน
- การทดสอบความแตกต่างรายคู่ของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

ตาราง 6 ค่าพารามิเตอร์ของข้อสอบเป็นรายข้อของแบบทดสอบ ตามทฤษฎีการตอบข้อสอบ

K	b	S_b	a	S_a	c	S_c	$I_i(\theta)$	$S_{I_i(\theta)}$
1	0.09	0.20	1.27	0.19	0.26	0.06	0.24	0.05
2	0.00	0.16	1.35	0.17	0.20	0.05	0.31	0.05
3	-0.29	0.12	2.06	0.23	0.20	0.05	0.72	0.12
4	-0.15	0.12	1.98	0.25	0.21	0.05	0.66	0.12
5	1.51	0.20	1.04	0.27	0.26	0.05	0.16	0.06
6	0.75	0.11	1.70	0.30	0.23	0.04	0.46	0.12
7	2.09	0.23	1.60	0.46	0.17	0.02	0.46	0.23
8	1.45	0.144	1.33	0.31	0.23	0.04	0.29	0.10
9	0.97	0.09	1.91	0.34	0.24	0.03	0.58	0.16
10	1.20	0.09	2.11	0.45	0.25	0.03	0.68	0.23
11	1.82	0.19	1.40	0.39	0.22	0.03	0.33	0.15
12	2.07	0.22	2.05	0.71	0.20	0.02	0.72	0.43
13	1.63	0.13	1.91	0.49	0.23	0.02	0.59	0.25
14	1.52	0.15	1.22	0.28	0.19	0.04	0.26	0.09
15	1.91	0.18	2.18	0.69	0.23	0.02	0.76	0.41
16	1.26	0.07	3.29	0.79	0.29	0.02	1.55	0.61
17	1.47	0.08	3.47	0.68	0.25	0.02	1.84	0.61
18	1.38	0.06	6.62	1.60	0.19	0.01	7.57	3.33
19	1.98	0.24	1.03	0.27	0.17	0.04	0.19	0.08
20	2.03	0.25	0.95	0.23	0.14	0.03	0.17	0.07
21	1.21	0.11	1.37	0.23	0.13	0.03	0.37	0.10
22	1.04	0.15	1.04	0.17	0.15	0.04	0.20	0.05
23	1.09	0.09	1.97	0.41	0.27	0.03	0.58	0.18
24	1.21	0.09	2.26	0.47	0.28	0.03	0.75	0.24
25	0.99	0.10	1.80	0.35	0.25	0.03	0.50	0.15
26	1.76	0.15	2.06	0.57	0.22	0.02	0.70	0.33
27	1.45	0.09	2.50	0.55	0.18	0.01	1.11	0.42
28	1.13	0.08	1.89	0.32	0.17	0.03	0.65	0.18
29	2.24	0.31	1.21	0.39	0.22	0.03	0.24	0.13
30	1.13	0.08	1.95	0.30	0.14	0.02	0.72	0.19

ตาราง 6 (ต่อ)

K	b	S_b	a	S_a	c	S_c	$I_i(\theta)$	$S_{I_i(\theta)}$
31	1.46	0.11	1.82	0.39	0.18	0.02	0.58	0.21
32	1.40	0.11	1.81	0.39	0.20	0.03	0.56	0.20
33	0.68	0.11	1.83	0.35	0.28	0.04	0.49	0.13
34	1.11	0.09	2.05	0.44	0.23	0.03	0.68	0.23
35	1.01	0.11	1.64	0.34	0.23	0.04	0.44	0.14
36	1.67	0.11	2.75	0.77	0.24	0.02	1.18	0.56
37	0.58	0.10	1.76	0.27	0.19	0.04	0.54	0.12
38	1.03	0.10	1.53	0.22	0.16	0.03	0.43	0.10
39	1.35	0.13	1.51	0.35	0.28	0.03	0.33	0.12
40	1.30	0.17	1.10	0.25	0.23	0.05	0.20	0.07

ตาราง 8 (ต่อ)

N	ค่า สถิติ	40 ข้อ			30 ข้อ			20 ข้อ		
		T1	T2	T3	T1	T2	T3	T1	T2	T3
300 คน	$\bar{X}$	13.87	19.36	24.10	10.09	14.21	17.57	7.06	9.95	12.20
	S.E.*	0.41	0.57	0.58	0.32	0.45	0.45	0.22	0.30	0.30
	S	7.11	9.90	10.08	5.46	7.78	7.83	3.74	5.12	5.17
	Max	38.00	40.00	40.00	28.00	30.00	30.00	19.00	20.00	20.00
	Min	2.00	2.00	5.00	1.00	1.00	5.00	1.00	1.00	2.00
	R	36.00	38.00	35.00	27.00	29.00	25.00	18.00	19.00	18.00
	Mdn	12.00	16.00	23.50	9.00	12.00	17.50	6.00	8.00	12.00
	Mo	9.00	13.00	40.00	6.00	9.00	30.00	6.00	8.00	20.00
	Sk	1.23	0.75	0.10	1.22	0.75	0.15	1.01	0.61	0.03
	S.E.*	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
	Ku	0.87	-0.68	-1.35	0.96	-0.74	-1.40	0.42	-0.79	-1.27
	S.E.*	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
100 คน	$\bar{X}$	13.33	19.44	23.16	9.57	14.40	16.96	6.83	9.99	11.93
	S.E.*	0.61	0.84	0.85	0.48	0.69	0.67	0.33	0.45	0.46
	S	6.06	8.38	8.49	4.75	6.95	6.70	3.32	4.54	4.59
	Max	34.00	40.00	40.00	26.00	30.00	30.00	18.00	20.00	20.00
	Min	5.00	7.00	9.00	3.00	4.00	7.00	2.00	3.00	4.00
	R	29.00	33.00	31.00	23.00	26.00	23.00	16.00	17.00	16.00
	Mdn	11.00	17.00	24.00	8.00	12.00	18.00	6.00	9.00	13.00
	Mo	10.00	12.00	28.00	7.00	9.00	21.00	7.00	6.00	14.00
	Sk	1.49	0.91	0.24	1.39	0.89	0.27	1.15	0.73	0.03
	S.E.*	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24	0.24
	Ku	1.94	-0.08	-1.00	1.73	-0.30	-0.98	0.97	-0.42	-1.16
	S.E.*	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48	0.48

* ความคลาดเคลื่อนมาตรฐาน ในช่วงความเชื่อมั่นระดับ 95%

ตาราง 11 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 40 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	68060.00	2	34030.00	464.15	0.00	68020.31	1	68020.31	927.89	0.00
Residual	238425.51	3252	73.32			238465.19	3253	73.31		
R – square = 0.22 Standard Error = 8.56						R – square = 0.22 Standard Error = 8.56				
Variables	Coefficients	SE	T	Sig		Coefficients	SE	t	Sig	
Constant	10.65	0.45	23.64	0.00		10.90	0.30	36.30	0.00	
Time	0.01	0.00	8.33	0.00		0.01	0.00	30.46	0.00	
Time**2	0.00	0.00	-0.05	0.46						

ตาราง 12 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 30 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	37777.55	2	18888.78	413.28	0.00	37768.52	1	37768.52	826.56	0.00
Residual	148632.67	3252	45.71			148641.71	3253	45.69		
R – square = 0.20 Standard Error = 6.76						R – square = 0.20 Standard Error = 6.76				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	7.69	0.36	21.62	0.00		7.81	0.24	32.94	0.00	
Time	0.00	0.00	7.62	0.00		0.00	0.00	28.75	0.00	
Time**2	0.00	0.00	-0.45	0.66						

ตาราง 13 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบ 20 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	17467.86	2	8733.93	432.90	0.00	17452.37	1	17452.37	865.09	0.00
Residual	65610.62	3252	20.18			65626.11	3253	20.17		
R – square = 0.21 Standard Error = 4.49						R – square = 0.21 Standard Error = 4.49				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	5.44	0.24	23.03	0.00		5.60	0.16	35.55	0.00	
Time	0.00	0.00	8.20	0.00		0.00	0.00	29.41	0.00	
Time**2	0.00	0.00	-0.88	0.38						

ตาราง 14 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 40 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	32690.49	1	32690.49	456.02	0.00	32690.49	1	32690.49	456.02	0.00
Residual	107385.99	1498	71.69			107385.99	1498	71.69		
R-square = 0.23 Standard Error = 8.47						R-square = 0.23 Standard Error = 8.47				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	10.83	0.44	24.77	0.00		10.83	0.44	24.77	0.00	
Time	0.01	0.00	21.36	0.00		0.01	0.00	21.36	0.00	
Time**2	0.00	0.00	-0.01	0.99						

ตาราง 15 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 30 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	18307.25	2	9153.63	203.44	0.00	18307.24	1	18307.24	407.16	0.00
Residual	67354.94	1497	44.99			67354.96	1498	44.96		
R-square = 0.21 Standard Error = 6.71						R-square = 0.21 Standard Error = 6.71				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	7.75	0.52	14.91	0.00		7.75	0.35	22.36	0.00	
Time	0.01	0.00	5.02	0.00		0.01	0.00	20.18	0.00	
Time**2	0.00	0.00	0.02	0.99						

ตาราง 16 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบ 20 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	8262.56	2	4131.28	208.17	0.00	8262.54	1	8262.54	416.62	0.00
Residual	29709.02	1497	19.85			29709.04	1498	19.83		
R – square = 0.22 Standard Error = 4.46						R – square = 0.22 Standard Error = 4.45				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	5.61	0.35	16.25	0.00		5.61	0.23	24.37	0.00	
Time	0.01	0.00	5.07	0.00		0.01	0.00	20.41		
Time**2	0.00	0.00	0.03	0.98						

ตาราง 17 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 40 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	21692.46	2	10846.23	141.28	0.00	21659.52	1	21659.52	282.32	0.00
Residual	68862.08	897	76.77			68895.02	898	76.72		
R – square = 0.24 Standard Error = 8.76						R – square = 0.24 Standard Error = 8.76				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	10.17	0.88	11.55	0.00		10.60	0.58	18.14	0.00	
Time	0.02	0.01	4.83	0.00		0.02	0.00	16.80	0.00	
Time**2	0.00	0.00	-0.66	0.51						

ตาราง 18 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 30 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	11826.56	2	5913.28	126.45	0.00	11799.67	1	11799.67	252.44	0.00
Residual	41948.58	897	46.77			41975.47	898	46.74		
R – square = 0.22 Standard Error = 6.84						R – square = 0.22 Standard Error = 6.84				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	7.29	0.69	10.63	0.00		7.68	0.46	16.83	0.00	
Time	0.02	0.00	4.70	0.00		0.01	0.00	15.89	0.00	
Time**2	0.00	0.00	-0.76	0.45						

ตาราง 19 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 300 คน ด้วยแบบทดสอบ 20 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	5611.35	2	2805.68	136.98	0.00	5596.63	1	5596.63	273.32	0.00
Residual	18373.24	897	20.48			18387.97	898	20.48		
R-square = 0.23 Standard Error = 4.53						R-square = 0.23 Standard Error = 4.53				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	5.13	0.45	11.30	0.00		5.41	0.30	17.93	0.00	
Time	0.01	0.00	4.95	0.00		0.01	0.00	16.53	0.00	
Time**2	0.00	0.00	-0.85	0.40						

ตาราง 20 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 40 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	5449.91	2	2724.96	47.05	0.00	5405.89	1	5405.89	93.41	0.00
Residual	17202.92	297	57.92			17246.95	298	57.88		
R – square = 0.24 Standard Error = 7.61						R – square = 0.24 Standard Error = 7.61				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	10.40	1.33	7.84	0.00		11.27	0.88	12.79	0.00	
Time	0.07	0.02	3.25	0.00		0.05	0.01	9.67	0.00	
Time**2	0.00	0.00	-0.87	0.38						

ตาราง 21 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 30 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	3090.25	2	1545.13	41.05	0.00	3064.21	1	3064.21	81.50	0.00
Residual	11178.58	297	37.64			11204.63	298	37.60		
R – square = 0.22 Standard Error = 6.14						R – square = 0.22 Standard Error = 6.13				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	7.42	1.07	6.94	0.00		8.09	0.71	11.40	0.00	
Time	0.05	0.02	3.05	0.00		0.04	0.00	9.03	0.00	
Time**2	0.00	0.00	-0.83	0.41						

ตาราง 22 ผลการวิเคราะห์แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression) และเชิงเส้นตรง (linear regression) ของกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบ 20 ข้อ

แบบแผนการเปลี่ยนแปลงเชิงเส้นโค้ง (nonlinear regression)						แบบแผนการเปลี่ยนแปลงเชิงเส้นตรง (linear regression)				
	SS	df	MS	F	Sig	SS	df	MS	F	Sig
Regression	1508.01	2	754.00	44.50	0.00	1498.46	1	1498.46	88.56	0.00
Residual	5032.91	297	16.95			5042.46	298	16.92		
R-square = 0.23 Standard Error = 4.12						R-square = 0.23 Standard Error = 4.12				
Variables	Coefficients	SE	t	Sig		Coefficients	SE	t	Sig	
Constant	5.30	0.72	7.38	0.00		5.70	0.48	11.97	0.00	
Time	0.03	0.01	3.07	0.00		0.03	0.00	9.41	0.00	
Time**2	0.00	0.00	-0.75	0.45						

ตาราง 29 ผลการวิเคราะห์หิโมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 1,085 คน ด้วยแบบทดสอบจำนวน 30 ข้อ

Fixed effect					Random Level 1 Reliability	Random effect			
Parameter	Coefficient	Standard error	t-ratio	Parameter		Variance Component	Total Observed variance	χ^2 (df = 1084)	R^2
1. uncentered model						1. uncentered model			0.56
Initial status B ₀₀	10.12**	0.17	59.30	0.66		Var (INT) R ₀	20.95**	33.74	3215.07
Linear change B ₁₀	3.59**	0.09	38.16	0.33		Var (SLP) R ₁	3.20**	15.99	1626.54
						Var (Error) E	12.79		
2. group centered model						2. group centered model			0.56
Initial status B ₀₀	13.71**	0.19	73.19	0.89		Var (INT) R ₀	33.81**	46.60	9683.68
Linear change B ₁₀	3.59**	0.09	38.16	0.33		Var (SLP) R ₁	3.20**	15.99	1626.54
						Var (Error) E	12.79		
3. grand centered model						3. grand centered model			0.56
Initial status B ₀₀	13.71**	0.19	73.19	0.89		Var (INT) R ₀	33.81**	46.60	9683.68
Linear change B ₁₀	3.59	0.09	38.16	0.33		Var (SLP) R ₁	3.20**	15.99	1626.54
						Var (Error) E	12.79		
Var (R _{null model}) = 28.85									

ตาราง 32 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบจำนวน 30 ข้อ

Fixed effect					Random effect				
Parameter	Coefficient	Standard error	t-ratio	Random Level 1 Reliability	Parameter	Variance Component	Total Observed variance	χ^2 (df = 499)	R^2
1. uncentered model					1. uncentered model				0.56
Initial status B_{00}	10.22**	0.26	40.15	0.68	Var (INT) R_0	21.98**	34.45	1554.07	
Linear change B_{10}	3.59**	0.14	26.10	0.34	Var (SLP) R_1	3.20**	15.67	754.93	
					Var (Error) E	12.47			
2. group centered model					2. group centered model				0.56
Initial status B_{00}	13.80**	0.28	49.95	0.89	Var (INT) R_0	34.02**	46.49	4581.85	
Linear change B_{10}	3.59**	0.14	26.10	0.34	Var (SLP) R_1	3.20**	15.67	754.93	
					Var (Error) E	12.47			
3. grand centered model					3. grand centered model				0.56
Initial status B_{00}	13.80**	0.28	49.95	0.89	Var (INT) R_0	34.02**	46.49	4581.85	
Linear change B_{10}	3.59**	0.14	26.10	0.34	Var (SLP) R_1	3.20**	15.67	754.93	
					Var (Error) E	12.47			
Var ($R_{\text{null model}}$) = 28.51									

ตาราง 33 ผลการวิเคราะห์โมเดลพหุระดับเชิงเส้นตรง จากการสอบซ้ำสามครั้ง ของกลุ่มตัวอย่าง 500 คน ด้วยแบบทดสอบจำนวน 20 ข้อ

Fixed effect					Random effect					
Parameter	Coefficient	Standard error	t-ratio	Random Level 1 Reliability	Parameter	Variance Component	Total Observed variance	χ^2 (df = 499)	R^2	
1. uncentered model					1. uncentered model					0.57
Initial status B ₀₀	7.27**	0.17	42.28	0.68	Var (INT) R ₀	9.99**	15.73	1541.94		
Linear change B ₁₀	2.41**	0.10	24.79	0.39	Var (SLP) R ₁	1.87**	7.61	823.94		
					Var (Error) E	5.74				
2. group centered model					2. group centered model					0.57
Initial status B ₀₀	9.68**	0.18	53.46	0.88	Var (INT) R ₀	14.48**	20.22	4277.92		
Linear change B ₁₀	2.41**	0.10	24.79	0.39	Var (SLP) R ₁	1.87**	7.61	823.94		
					Var (Error) E	5.74				
3. grand centered model					3. grand centered model					0.57
Initial status B ₀₀	9.68**	0.18	53.46	0.88	Var (INT) R ₀	14.48**	20.22	4277.92		
Linear change B ₁₀	2.41**	0.10	24.79	0.39	Var (SLP) R ₁	1.87**	7.61	823.94		
					Var (Error) E	5.74				
Var (R _{null model}) = 13.42										

ตาราง 41 ค่าสถิติพื้นฐานของคะแนนการเปลี่ยนแปลงแต่ละวิธี ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วย
แบบทดสอบที่มีความยาวต่างกัน

N	K	วิธีวัด	$\bar{X}$	S.E.	S	Max	Min	Mdn	Mo	Sk	S.E.	Ku	S.E.
1,085 คน	40 ข้อ	D	9.71	0.23	7.71	33.00	-9.00	9.00	5.00	0.33	0.70	-0.62	0.15
		S	40.58	1.00	33.08	100.00	-100.00	36.00	100	0.17	0.70	-0.82	0.15
		d_θ	0	0	0.87	2.54	-2.26	-1.00	0.38	0.13	0.70	-0.31	0.15
		D_θ	9.05	0.43	14.05	56.00	-27.98	5.40	34.27	1.21	0.70	1.07	0.15
		RD_θ	13.56	0.68	22.51	78.18	-126.35	7.74	68.60	0.72	0.70	2.20	0.15
		HLM	4.86	0.06	1.82	8.64	1.35	4.65	3.17	0.24	0.70	-1.20	0.15
	30 ข้อ	D	7.18	0.19	6.19	26.00	-7.00	7.00	4.00	0.30	0.70	-0.64	0.15
		S	39.13	1.05	34.57	100.00	-100.00	34.78	100	0.13	0.70	-0.82	0.15
		d_θ	0	0.03	0.85	2.93	-2.48	0.01	0.18	0.21	0.70	-0.23	0.15
		D_θ	9.70	0.48	15.95	84.51	-23.23	3.06	45.43	1.47	0.70	1.75	0.15
		RD_θ	13.89	0.73	24.15	84.68	-129.75	4.20	74.83	0.78	0.70	2.49	0.15
		HLM	3.59	0.04	1.41	6.49	1.09	3.38	6.15	0.32	0.70	-1.17	0.15
	20 ข้อ	D	4.88	0.13	4.32	17.00	-7.00	5.00	6.00	0.20	0.70	-0.48	0.15
		S	40.89	1.10	36.13	100.00	-70.00	37.50	100	0.04	0.70	-0.92	0.15
		d_θ	0	0.03	0.87	2.77	-3.04	-0.08	-0.60	0.26	0.70	-0.18	0.15
		D_θ	7.68	0.42	13.94	50.33	-44.33	3.15	6.97	1.47	0.70	1.88	0.15
		RD_θ	11.38	0.66	21.88	69.64	-153.34	4.42	24.11	0.69	0.70	4.32	0.15
		HLM	2.44	0.03	0.92	4.35	0.58	2.34	1.77	0.12	0.70	-1.18	0.15
500 คน	40 ข้อ	D	9.65	0.34	7.61	30.00	-9.00	8.50	5.00	0.32	0.11	-0.65	0.22
		S	40.53	1.48	32.99	100.00	-100.00	36.36	0.00	0.09	0.11	-0.52	0.22
		d_θ	0	0.04	0.84	2.32	-2.01	-0.03	0.38	0.21	0.11	-0.37	0.22
		D_θ	8.50	0.62	13.78	74.13	-26.49	5.21	-2.91	1.30	0.11	1.81	0.22
		RD_θ	12.67	0.99	22.22	78.10	-124.71	7.25	68.85	0.48	0.11	3.46	0.22
		HLM	4.83	0.08	1.77	8.58	1.31	4.63	2.60	0.21	0.11	-1.17	0.22
	30 ข้อ	D	7.17	0.27	6.14	26.00	-7.00	7.00	4.00	0.32	0.11	-0.60	0.22
		S	39.13	1.54	34.53	100.00	-100.00	36.00	100	0.07	0.11	-0.56	0.22
		d_θ	0	0.04	0.83	2.934	-2.38	-0.09	0.21	0.30	0.11	-0.22	0.22
		D_θ	8.13	0.64	14.38	54.56	-23.07	1.03	-0.77	1.46	0.11	1.59	0.22
		RD_θ	11.66	1.03	23.05	73.67	-127.98	1.43	68.61	0.39	0.11	4.63	0.22
		HLM	3.59	0.06	1.37	6.50	1.01	3.39	1.96	0.25	0.11	-1.15	0.22
	20 ข้อ	D	4.83	0.19	4.35	17.00	-5.00	4.50	2.00	0.24	0.11	-0.46	0.22
		S	40.66	1.63	36.42	100.00	-55.56	36.93	100	0.06	0.11	-0.92	0.22
		d_θ	0.00	0.04	0.87	2.62	-1.88	-0.10	-0.64	0.36	0.11	-0.36	0.22
		D_θ	7.63	0.61	13.70	49.88	-21.84	2.87	14.37	1.54	0.11	1.76	0.22
		RD_θ	11.57	0.95	21.31	69.32	-59.92	3.95	39.42	1.25	0.11	1.28	0.22
		HLM	2.41	0.04	1.00	4.71	0.29	2.37	2.37	0.05	0.11	-1.06	0.22

ตาราง 41 (ต่อ)

N	K	วิธีวัด	$\bar{X}$	S.E.	S	Max	Min	Mdn	Mo	Sk	S.E.	Ku	S.E.
300 คน	40 ข้อ	D	10.23	0.44	7.62	29.00	-6.00	9.00	2.00	0.26	0.14	-0.71	0.28
		S	43.51	1.90	32.85	100.00	-24.00	41.90	100	0.16	0.14	-1.12	0.28
		d_θ	0.00	0.05	0.81	2.00	-1.93	0.00	0.19	0.12	0.14	-0.54	0.28
		D_θ	8.61	0.79	13.59	55.10	-24.70	4.89	-2.80	1.13	0.14	0.86	0.28
		RD_θ	13.44	1.29	22.41	77.70	-56.55	7.00	69.98	0.97	0.14	0.94	0.28
		HLM	5.12	0.10	1.72	8.61	1.76	5.16	3.00	0.06	0.14	-1.29	0.28
	30 ข้อ	D	7.48	0.35	6.07	21.00	-5.00	7.00	2.00	0.21	0.14	-0.84	0.28
		S	41.63	1.97	34.04	100.00	-27.78	38.10	100	0.14	0.14	-1.10	0.28
		d_θ	0.00	0.05	0.81	1.98	-1.78	-0.05	0.12	0.22	0.14	-0.67	0.28
		D_θ	9.08	0.87	15.10	57.36	-20.10	3.26	1.07	1.36	0.14	1.17	0.28
		RD_θ	13.38	1.39	24.13	78.73	-94.19	4.48	75.08	0.75	0.14	2.21	0.28
		HLM	3.74	0.08	1.35	6.44	1.52	3.64	1.94	0.16	0.14	-1.31	0.28
	20	D	5.14	0.24	4.23	15.00	-7.00	5.00	6.00	0.07	0.14	-0.49	0.28
		S	43.37	2.07	35.92	100.00	-70.00	42.86	100	-0.08	0.14	-0.75	0.28
		d_θ	0.00	0.05	0.84	2.04	-2.01	0.06	-0.76	0.25	0.14	-0.55	0.28
		D_θ	7.44	0.79	13.73	49.05	-29.98	3.29	19.35	1.35	0.14	1.61	0.28
		RD_θ	11.29	1.25	21.65	68.20	-70.49	4.60	45.83	0.97	0.14	1.41	0.28
		HLM	2.57	0.05	0.85	4.22	0.85	2.55	3.83	0.04	0.14	-1.21	0.28
100 คน	40 ข้อ	D	9.83	0.69	6.92	29.00	-3.00	9.50	5.00	0.37	0.24	-0.44	0.48
		S	38.55	2.81	28.10	100.00	-27.00	40.00	100	0.30	0.24	-0.43	0.48
		d_θ	0.00	0.00	0.92	2.49	-2.37	-0.14	-2.37	0.14	0.24	-0.19	0.48
		D_θ	8.18	1.37	13.66	56.76	-16.75	3.93	-1.44	1.36	0.24	1.77	0.48
		RD_θ	11.95	2.14	21.41	78.89	-40.61	5.42	-40.61	1.17	0.24	1.72	0.48
		HLM	4.92	0.14	1.45	8.04	2.48	4.84	3.15	0.41	0.24	-0.78	0.48
	30 ข้อ	D	7.39	0.57	5.71	23.00	-4.00	7.00	11.00	0.31	0.24	-0.41	0.48
		S	37.53	3.04	30.44	100.00	-50.00	37.17	100.00	0.18	0.24	-0.12	0.48
		d_θ	0.00	0.10	0.97	2.66	-3.00	-0.17	-2.97	-0.02	0.24	0.22	0.48
		D_θ	9.01	1.56	15.60	58.78	-19.73	2.69	-19.73	1.42	0.24	1.64	0.48
		RD_θ	12.75	2.35	23.54	79.67	-47.49	3.64	-47.49	1.26	0.24	1.67	0.48
		HLM	3.70	0.12	1.15	6.19	1.88	3.62	2.35	0.48	0.24	-0.79	0.48
	20 ข้อ	D	5.11	0.43	4.27	14.00	-4.00	5.50	7.00	0.01	0.24	-0.84	0.48
		S	39.74	3.38	33.81	100.00	-44.44	43.30	0.00	-0.12	0.24	-0.73	0.48
		d_θ	0.00	0.10	1.01	2.22	-2.34	-0.17	-2.33	-0.04	0.24	-0.49	0.48
		D_θ	7.87	1.35	13.47	49.90	-10.19	1.68	-10.19	1.51	0.24	1.58	0.48
		RD_θ	11.54	2.00	20.03	69.83	-20.41	2.33	-20.41	1.52	0.24	1.62	0.48
		HLM	2.56	0.08	0.83	4.25	1.16	2.63	1.83	0.20	0.24	-0.99	0.48

ตาราง 45 ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลง
ระหว่างแต่ละวิธีที่สนใจกับวิธีโมเดลพหุระดับ ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มี
ความยาวต่างกัน

N	K	วิธีวัด	r	Z	χ^2 (df=4)
1,085 คน	40 ข้อ	D-HLM	0.77	1.02	2008.11**
		S-HLM	0.94	1.73	
		d_θ -HLM	0.42	0.45	
		D_θ -HLM	0.77	1.02	
		RD_θ -HLM	0.74	0.95	
	30 ข้อ	D-HLM	0.73	0.94	1574.22**
		S-HLM	0.92	1.56	
		d_θ -HLM	0.39	0.41	
		D_θ -HLM	0.78	1.06	
		RD_θ -HLM	0.75	0.98	
	20 ข้อ	D-HLM	0.78	1.05	1687.08**
		S-HLM	0.94	1.74	
		d_θ -HLM	0.50	0.55	
		D_θ -HLM	0.74	0.95	
		RD_θ -HLM	0.71	0.88	
500 คน	40 ข้อ	D-HLM	0.78	1.04	816.25**
		S-HLM	0.93	1.68	
		d_θ -HLM	0.43	0.46	
		D_θ -HLM	0.75	0.98	
		RD_θ -HLM	0.72	0.90	
	30 ข้อ	D-HLM	0.76	1.00	646.00**
		S-HLM	0.91	1.53	
		d_θ -HLM	0.43	0.46	
		D_θ -HLM	0.83	1.20	
		RD_θ -HLM	0.76	0.98	
	20 ข้อ	D-HLM	0.86	1.29	722.85**
		S-HLM	0.95	1.83	
		d_θ -HLM	0.61	0.71	
		D_θ -HLM	0.76	1.01	
		RD_θ -HLM	0.74	0.95	

ตาราง 45 (ต่อ)

N	K	วิธีวัด	R	Z	χ^2 (df=4)
300 คน	40 ข้อ	D-HLM	0.80	1.08	805.12**
		S-HLM	0.97	2.06	
		d_θ -HLM	0.49	0.54	
		D_θ -HLM	0.78	1.05	
		RD_θ -HLM	0.75	0.98	
	30 ข้อ	D-HLM	0.77	1.02	720.73**
		S-HLM	0.96	1.95	
		d_θ -HLM	0.45	0.49	
		D_θ -HLM	0.79	1.07	
		RD_θ -HLM	0.74	0.95	
	20 ข้อ	D-HLM	0.76	0.99	494.81**
		S-HLM	0.94	1.76	
		d_θ -HLM	0.49	0.53	
		D_θ -HLM	0.75	0.96	
		RD_θ -HLM	0.73	0.92	
100 คน	40 ข้อ	D-HLM	0.65	0.78	90.14**
		S-HLM	0.86	1.30	
		d_θ -HLM	0.36	0.37	
		D_θ -HLM	0.71	0.89	
		RD_θ -HLM	0.71	0.89	
	30 ข้อ	D-HLM	0.63	0.74	72.96**
		S-HLM	0.83	1.18	
		d_θ -HLM	0.34	0.35	
		D_θ -HLM	0.71	0.89	
		RD_θ -HLM	0.70	0.86	
	20 ข้อ	D-HLM	0.71	0.89	87.34**
		S-HLM	0.88	1.38	
		d_θ -HLM	0.49	0.53	
		D_θ -HLM	0.78	1.04	
		RD_θ -HLM	0.78	1.05	

** p<.01

ตาราง 46 ผลการทดสอบความแตกต่างรายคู่ของค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนการเปลี่ยนแปลง
ที่ได้จากแต่ละวิธีกับวิธีโมเดลพหุระดับ ของกลุ่มตัวอย่างที่ไม่เท่ากัน ด้วยแบบทดสอบที่มีความยาวต่างกัน

N	K	วิธีวัด	D	S	d_θ	D_θ	RD_θ
1,085 คน	40 ข้อ	Z	1.02	1.73	0.45	1.02	0.96
		D	1.02	-			
		S	1.73	-24.14**	-		
		d_θ	0.45	22.27**	34.56**	-	
		D_θ	1.02	0	22.33**	-15.21**	-
		RD_θ	0.95	1.73	25.15**	-13.28**	5.09**
	30 ข้อ	Z	0.94	1.56	0.41	1.06	0.98
		D	0.94	-			
		S	1.56	-22.83**	-		
		d_θ	0.41	25.65**	32.65**	-	
		D_θ	1.06	-3.61**	16.12**	-17.36**	-
		RD_θ	0.98	-1.16	19.62**	-15.00**	6.27**
	20 ข้อ	Z	1.05	1.74	0.55	0.95	0.88
		D	1.05	-			
		S	1.74	-23.32**	-		
		d_θ	0.55	24.66**	33.57**	-	
		D_θ	0.95	2.85**	23.97**	-11.10**	-
		RD_θ	0.88	4.68**	26.11**	-8.89**	6.90**
500 คน	40 ข้อ	Z	1.04	1.68	0.46	0.98	0.90
		D	1.04	-			
		S	1.68	-14.66**	-		
		d_θ	0.46	15.63**	22.46**	-	
		D_θ	0.98	1.21	14.81**	-9.47**	-
		RD_θ	0.90	2.71**	17.53**	-7.87**	4.09**
	30 ข้อ	Z	1.00	1.53	0.46	1.20	0.98
		D	1.00	-			
		S	1.53	-12.79**	-		
		d_θ	0.46	17.71**	20.69**	-	
		D_θ	1.20	-4.18**	7.78**	-13.85**	-
		RD_θ	0.98	0.40	13.96**	-9.66**	10.01**

ตาราง 46 (ต่อ)

N	K	วิธีวัด	D	S	d_θ	D_θ	RD_θ
500 คน	20 ข้อ	Z	1.29	1.83	0.71	1.01	0.95
		D	1.29	-			
		S	1.83	-11.22**	-		
		d_θ	0.71	18.22**	20.95**	-	
		D_θ	1.01	5.28**	17.04**	-5.50**	-
		RD_θ	0.95	6.24**	18.84**	-4.24**	4.33**
300 คน	40 ข้อ	Z	1.08	2.06	0.54	1.05	0.98
		D	1.08	-			
		S	2.06	-15.73**	-		
		d_θ	0.54	13.21**	21.25**	-	
		D_θ	1.05	0.45	16.36**	-7.13**	-
		RD_θ	0.98	1.43	17.78**	-5.88**	3.19**
	30 ข้อ	Z	1.02	1.95	0.49	1.07	0.95
		D	1.02	-			
		S	1.95	-15.89**	-		
		d_θ	0.49	13.82**	20.75**	-	
		D_θ	1.07	-0.76	14.22**	-8.11**	-
		RD_θ	0.95	1.02	16.08**	-6.25**	4.89**
	20 ข้อ	Z	0.99	1.76	0.53	0.96	0.92
		D	0.99	-			
		S	1.76	-13.33**	-		
		d_θ	0.53	12.21**	17.86**	-	
		D_θ	0.96	0.44	13.02**	-6.34**	-
		RD_θ	0.92	1.00	13.81**	-5.49**	2.41**
100 คน	40 ข้อ	Z	0.78	1.30	0.37	0.89	0.88
		D	0.78	-			
		S	1.30	-6.77**	-		
		d_θ	0.37	6.82**	8.83**	-	
		D_θ	0.89	-1.17	4.72**	-5.34**	-
		RD_θ	0.88	-0.98	5.08**	-4.86**	0.33

ตาราง 46 (ต่อ)

N	K	วิธีวัด	D	S	d_θ	D_θ	RD_θ
100 คน	30 ข้อ	Z	0.74	1.18	0.35	0.89	0.86
		D	0.74	-			
		S	1.18	-6.08**	-		
		d_θ	0.35	6.55**	8.37**	-	
		D_θ	0.89	-1.48	3.33**	-5.22**	-
		RD_θ	0.86	-1.12	3.78**	-4.70**	1.18
	20 ข้อ	Z	0.89	1.38	0.53	1.04	1.05
		D	0.89	-			
		S	1.38	-6.15**	-		
		d_θ	0.53	6.33**	8.32**	-	
		D_θ	1.04	-1.29	3.30**	-4.38**	-
		RD_θ	1.05	-1.33	3.23**	-4.26**	-0.46

**P<.01

ภาคผนวก ข
การสร้างและพัฒนาเครื่องมือการวิจัย

- รายชื่อผู้เชี่ยวชาญ
- ค่าดัชนีความสอดคล้อง (IOC)
- การวิเคราะห์องค์ประกอบเชิงยืนยันด้วยโปรแกรมลิสรел
- คำสั่งและผลการประมาณค่าพารามิเตอร์ข้อสอบด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์

รายชื่อผู้เชี่ยวชาญตรวจคุณภาพเครื่องมือ

ที่	ชื่อ-สกุล	สถานที่ทำงาน	การพิจารณา	
			ข้อ 2 และ 3	ข้อ 4 และ 6
1	นางนงนารถ ศิลากะษ	ศูนย์การศึกษานอกโรงเรียนภาค- ตะวันออกเฉียงเหนือ	✓	✓
2	นางพรทิพย์ กล้ารบ	กลุ่มงานพัฒนาการศึกษานอก- โรงเรียน สำนักบริหารงานการ- ศึกษานอกโรงเรียน	✓	
3	นางสาวสุนทรี ไกรกาบแก้ว	ศูนย์บริการการศึกษานอก- โรงเรียนอำเภอไทรน้อย		✓
4	นางสาวบุศรา แพทย์รัตน์	ศูนย์บริการการศึกษานอก- โรงเรียนเขตพญาไท		✓
5	นางสาวรุ่งทิพา สอนด้วง	ศูนย์บริการการศึกษานอก- โรงเรียนเขตพญาไท	✓	✓
6	นางพีเอ โคตรมุงคุณ	ศูนย์บริการการศึกษานอก- โรงเรียนอำเภอเมืองชลบุรี	✓	
7	นางกนกทิพย์ นาคประสพสุข	โรงเรียนชะอำคุณหญิงเนื่องบุรี	✓	✓
8	นายทวีศักดิ์ จงประดับเกียรติ	โรงเรียนมัธยมสาธิตมหาวิทยาลัย ราชภัฏบ้านสมเด็จเจ้าพระยา	✓	✓
9	นางสาวเพ็ญพร ทองคำสุก	โรงเรียนมัธยมสาธิตมหาวิทยาลัย ราชภัฏบ้านสมเด็จเจ้าพระยา	✓	✓
10	นางสาวสุวิมล มาลีลัย	โรงเรียนร่มเกล้า		✓
11	นายทินกร เนติเมธี	โรงเรียนบางช้างวิทยา		✓
12	นางสาวยุพาวดี รัชดาพรรณากุล	โรงเรียนอัสสัมชัญบางรัก	✓	
13	นายอุเทน เสวตบุญชร	โรงเรียนอัสสัมชัญบางรัก	✓	
14	มีสพุนธิวา วาปะ	โรงเรียนอัสสัมชัญบางรัก	✓	
15	นางสาวมุสดี อัครชัยสุวิกรม	วิทยาลัยการอาชีพสระบุรี		✓

ค่าดัชนีความสอดคล้อง (IOC) และจำนวนความถี่ของผลการประเมิน ของผู้เชี่ยวชาญที่ประเมิน
ความสอดคล้องระหว่างมวลความรู้ มวลความรู้ย่อยกับข้อสอบ

ที่	มวลความรู้/มวลความรู้ย่อย	ข้อที่ (ฉบับที่)	จำนวนความถี่			รวม	ดัชนีความ สอดคล้อง
			+1	0	-1		
1	ความรู้พื้นฐานเกี่ยวกับสมการ						
	1.1 นักศึกษาสามารถระบุตัวแปรได้	1(1,2)	10			10	1.00
		2(1,2)	10			10	1.00
	1.2 นักศึกษาสามารถระบุประโยคภาษาหรือ ประโยคสัญลักษณ์หรือสมการได้	3(1,2)	10			10	1.00
		4(1,2)	10			10	1.00
2	ความสามารถในการคิดคำนวณเพื่อหาคำตอบ ของสมการ โดยใช้การแทนค่าตัวแปรใน						
	2.1 สมการนักศึกษาสามารถระบุจำนวนที่กำหนด ให้ได้ว่าเป็นคำตอบของสมการ	5(1,2)	10			10	1.00
		6(1,2)	10			10	1.00
		7(1,2)	10			10	1.00
		8(1,2)	10			10	1.00
	2.2 นักศึกษาสามารถระบุจำนวนที่กำหนดให้ได้ ว่าเป็นคำตอบของสมการ	9(1,2)	10			10	1.00
		10(1,2)	10			10	1.00
3	ความสามารถในการคิดคำนวณหาคำตอบของ สมการ โดยใช้สมบัติของการเท่ากัน						
	3.1 นักศึกษาสามารถหาคำตอบของสมการโดยใช้ สมบัติการเท่ากันของการบวกได้	11(1,2)	8	2		8	0.8
		12(1,2)	8	2		8	0.8
		13(1,2)	8	2		8	0.8
		14(1,2)	8	2		8	0.8
	3.2 นักศึกษาสามารถหาคำตอบของสมการโดยใช้ สมบัติการเท่ากันของการคูณได้	15(1,2)	8	2		8	0.8
		16(1,2)	8	2		8	0.8
		17(1,2)	8	2		8	0.8
		18(1,2)	8	2		8	0.8
	3.3 นักศึกษาสามารถหาคำตอบของสมการโดยใช้ สมบัติการเท่ากันทั้งสองสมบัติได้	19(1,2)	8	2		8	0.8
		20(1,2)	8	2		8	0.8
		21(1,2)	8	2		8	0.8
		22(1,2)	8	2		8	0.8

ที่	มวลความรู้/มวลความรู้ย่อย	ข้อที่ (ฉบับที่)	จำนวนความถี่			รวม	ดัชนีความ สอดคล้อง
			+1	0	-1		
4	ความสามารถในการแก้สมการเชิงเส้นตัวแปรเดียว จากประโยคสัญลักษณ์						
4.1	นักศึกษาสามารถแก้สมการเชิงเส้นตัวแปร เดียวได้	23(1,2)	10			10	1.00
	นักศึกษาสามารถแก้สมการเชิงเส้น ตัวแปรเดียวได้	24(1,2)	10			10	1.00
		25(1,2)	10			10	1.00
		26(1,2)	10			10	1.00
4.2	นักศึกษาสามารถตรวจสอบคำตอบของ สมการเชิงเส้นตัวแปรเดียวได้	27(1,2)	10			10	1.00
		28(1,2)	10			10	1.00
		29(1,2)	10			10	1.00
		30(1,2)	10			10	1.00
5	ความสามารถในการเปลี่ยนประโยคระหว่าง ประโยคภาษากับประโยคสัญลักษณ์						
5.1	นักศึกษามารถระบุประโยคสัญลักษณ์ที่ เขียนมาจากประโยคภาษาที่กำหนดให้ได้	31(1,2)	10			10	1.00
		32(1,2)	10			10	1.00
		33(1,2)	10			10	1.00
		34(1,2)	10			10	1.00
6	ความสามารถในการแก้โจทย์ปัญหาสมการเชิงเส้น ตัวแปรเดียว						
6.1	นักศึกษาสามารถแก้โจทย์ปัญหาสมการ เชิงเส้นตัวแปรเดียวได้	35(1,2)	10			10	1.00
		36(1,2)	10			10	1.00
		37(1)	6		4	2	0.20
		37(2)	10			10	1.00
		38(1,2)	10			10	1.00
		39(1,2)	10			10	1.00
		40(1,2)	10			10	1.00

ผลการวิเคราะห์องค์ประกอบเชิงยืนยัน

DATE: 11/ 2/2006

TIME: 20:34

LISREL 8.54

BY

Karl G. J reskog & Dag S rbom

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., 1981-2002

Use of this program is subject to the terms specified in the

UniveSal Copyright Convention.

Website: www.ssicentral.com

50 items 1034 peSons

Covariance Matrix

[illegible]

Covariance Matrix

141 142 143 144 145 146 147 148 149 150

=====

141 0.25

I42	0.05	0.24									
I43	0.05	0.07	0.23								
I44	0.04	0.05	0.05	0.24							
I45	0.03	0.01	-0.01	0.04	0.21						
I46	0.06	0.06	0.07	0.06	0.02	0.25					
I47	0.04	0.04	0.01	0.04	0.03	0.06	0.23				
I48	0.04	0.03	0.04	0.03	0.04	0.05	0.04	0.24			
I49	0.05	0.05	0.04	0.03	0.01	0.07	0.07	0.04	0.25		
I50	0.03	0.04	0.04	0.03	0.01	0.05	0.06	0.02	0.07	0.24	

50 items 1034 peSons

Number of Iterations = 6

LISREL Estimates (Maximum Likelihood)

Measurement Equations

I1 = 0.14*equation, Errorvar.= 0.22 , $R^2 = 0.078$

(0.016) (0.0098)

8.42 22.37

I2 = 0.15*equation, Errorvar.= 0.17 , $R^2 = 0.12$

(0.015) (0.0078)

10.45 22.17

I3 = 0.16*equation, Errorvar.= 0.21 , $R^2 = 0.11$

(0.016) (0.0096)

10.15 22.20

I4 = 0.21*equation, Errorvar.= 0.18 , $R^2 = 0.19$

(0.015) (0.0082)

13.75 21.69

I5 = 0.21*equation, Errorvar.= 0.20 , $R^2 = 0.18$

(0.016) (0.0090)

13.32 21.76

I6 = 0.21*equation, Errorvar.= 0.19 , $R^2 = 0.19$

(0.015) (0.0086)

13.55 21.72

I7 = 0.12*equation, Errorvar.= 0.23 , $R^2 = 0.055$

(0.016) (0.010)

7.05 22.48

I8 = 0.19*equation, Errorvar.= 0.21 , $R^2 = 0.14$

(0.016) (0.0097)

11.48 22.05

I9 = 0.15*equation, Errorvar.= 0.23 , $R^2 = 0.096$
 (0.016) (0.010)
 9.41 22.28

I10 = 0.13*equation, Errorvar.= 0.22 , $R^2 = 0.072$
 (0.016) (0.0096)
 8.08 22.40

I11 = 0.076*equation, Errorvar.= 0.17 , $R^2 = 0.033$
 (0.014) (0.0075)
 5.41 22.66

I12 = 0.18*equation, Errorvar.= 0.21 , $R^2 = 0.14$
 (0.016) (0.0095)
 11.41 22.08

I13 = 0.082*equation, Errorvar.= 0.23 , $R^2 = 0.028$
 (0.017) (0.010)
 4.93 22.63

I14 = 0.15*equation, Errorvar.= 0.21 , $R^2 = 0.095$
 (0.016) (0.0096)
 9.32 22.28

I15 = 0.097*equation, Errorvar.= 0.20 , $R^2 = 0.044$
 (0.015) (0.0091)
 6.30 22.53

I16 = 0.13*equation, Errorvar.= 0.22 , $R^2 = 0.072$
 (0.016) (0.0097)
 8.06 22.40

I17 = 0.067*equation, Errorvar.= 0.18 , $R^2 = 0.025$
 (0.014) (0.0079)
 4.65 22.62

I18 = 0.12*equation, Errorvar.= 0.20 , $R^2 = 0.070$
 (0.015) (0.0089)
 7.97 22.41

I19 = 0.16*equation, Errorvar.= 0.20 , $R^2 = 0.11$
 (0.016) (0.0090)
 10.01 22.22

I20 = 0.070*equation, Errorvar.= 0.20 , $R^2 = 0.024$
 (0.015) (0.0087)
 4.61 22.62

I21 = 0.13*equation, Errorvar.= 0.22 , $R^2 = 0.070$
 (0.016) (0.0100)
 7.96 22.41

I22 = 0.16*equation, Errorvar.= 0.21 , $R^2 = 0.10$
 (0.016) (0.0095)
 9.78 22.24

I23 = 0.11*equation, Errorvar.= 0.21 , $R^2 = 0.055$
 (0.016) (0.0091)
 7.04 22.48

I24 = 0.12*equation, Errorvar.= 0.18 , $R^2 = 0.078$
 (0.014) (0.0079)
 8.43 22.37

I25 = 0.11*equation, Errorvar.= 0.19 , $R^2 = 0.063$
 (0.015) (0.0087)
 7.52 22.45

I26 = 0.11*equation, Errorvar.= 0.19 , $R^2 = 0.057$
 (0.015) (0.0082)
 7.13 22.48

I27 = 0.19*equation, Errorvar.= 0.18 , $R^2 = 0.16$
 (0.015) (0.0083)
 12.54 21.91

I28 = 0.17*equation, Errorvar.= 0.21 , $R^2 = 0.12$
 (0.016) (0.0095)
 10.77 22.14

I29 = 0.15*equation, Errorvar.= 0.22 , $R^2 = 0.096$
 (0.016) (0.0097)
 9.39 22.29

I30 = 0.18*equation, Errorvar.= 0.21 , $R^2 = 0.13$
 (0.016) (0.0096)
 11.22 22.08

I31 = 0.17*equation, Errorvar.= 0.21 , $R^2 = 0.12$
 (0.016) (0.0096)
 10.36 22.18

I32 = 0.20*equation, Errorvar.= 0.21 , $R^2 = 0.16$
 (0.016) (0.0094)
 12.34 21.94

I33 = 0.10*equation, Errorvar.= 0.19 , $R^2 = 0.052$
 (0.015) (0.0085)
 6.81 22.50

I34 = 0.14*equation, Errorvar.= 0.18 , $R^2 = 0.096$
 (0.015) (0.0081)
 9.41 22.28

I35 = 0.20*equation, Errorvar.= 0.18 , $R^2 = 0.17$
 (0.015) (0.0084)
 12.87 21.86

I36 = 0.090*equation, Errorvar.= 0.20 , $R^2 = 0.039$
 (0.015) (0.0088)
 5.90 22.56

I37 = 0.21*equation, Errorvar.= 0.17 , $R^2 = 0.20$
 (0.015) (0.0080)
 13.81 21.71

I38 = 0.15*equation, Errorvar.= 0.19 , $R^2 = 0.11$
 (0.015) (0.0084)
 10.31 22.19

I39 = 0.16*equation, Errorvar.= 0.19 , $R^2 = 0.11$
 (0.015) (0.0087)
 10.13 22.22

I40 = 0.21*equation, Errorvar.= 0.20 , $R^2 = 0.18$
 (0.016) (0.0091)
 13.29 21.76

I41 = 0.22*equation, Errorvar.= 0.20 , $R^2 = 0.20$
 (0.016) (0.0093)
 13.85 21.71

I42 = 0.19*equation, Errorvar.= 0.20 , $R^2 = 0.15$
 (0.016) (0.0091)
 11.78 22.03

I43 = 0.17*equation, Errorvar.= 0.20 , $R^2 = 0.12$
 (0.016) (0.0092)
 10.46 22.16

I44 = 0.19*equation, Errorvar.= 0.21 , $R^2 = 0.15$
 (0.016) (0.0094)
 11.97 21.99

I45 = 0.11*equation, Errorvar.= 0.20 , $R^2 = 0.057$
 (0.015) (0.0088)
 7.16 22.47

I46 = 0.25*equation, Errorvar.= 0.19 , $R^2 = 0.26$
 (0.016) (0.0087)
 16.05 21.30

I47 = 0.21*equation, Errorvar.= 0.19 , $R^2 = 0.19$
 (0.015) (0.0087)
 13.48 21.77

I48 = 0.15*equation, Errorvar.= 0.22 , $R^2 = 0.096$
 (0.016) (0.0098)
 9.38 22.28

I49 = 0.20*equation, Errorvar.= 0.21 , $R^2 = 0.17$
 (0.016) (0.0095)
 12.69 21.88

I50 = 0.15*equation, Errorvar.= 0.22 , $R^2 = 0.096$
 (0.016) (0.0098)
 9.38 22.27

Error Covariance for I2 and I1 = 0.080
 (0.0067)
 11.96

Error Covariance for I3 and I1 = 0.13
 (0.0080)
 16.20

Error Covariance for I3 and I2 = 0.073
 (0.0066)
 11.10

Error Covariance for I5 and I4 = 0.068
 (0.0065)
 10.41

Error Covariance for I6 and I4 = 0.069
 (0.0064)
 10.77

Error Covariance for I6 and I5 = 0.079
 (0.0068)
 11.56

Error Covariance for I8 and I7 = 0.036
 (0.0071)
 5.01

Error Covariance for I11 and I10 = 0.050
 (0.0061)
 8.30

Error Covariance for I12 and I11 = 0.030

(0.0057)

5.34

Error Covariance for I13 and I11 = 0.033

(0.0060)

5.54

Error Covariance for I13 and I12 = 0.046

(0.0072)

6.44

Error Covariance for I14 and I12 = 0.040

(0.0068)

5.90

Error Covariance for I14 and I13 = 0.038

(0.0071)

5.37

Error Covariance for I16 and I15 = 0.060

(0.0069)

8.62

Error Covariance for I17 and I15 = 0.039

(0.0061)

6.41

Error Covariance for I17 and I16 = 0.041

(0.0063)

6.43

Error Covariance for I21 and I20 = 0.019

(0.0066)

2.94

Error Covariance for I22 and I20 = 0.034

(0.0065)

5.22

Error Covariance for I22 and I21 = 0.051

(0.0071)

7.24

Error Covariance for I24 and I23 = 0.041

(0.0062)

6.65

Error Covariance for I26 and I25 = 0.031

(0.0061)

5.18

Error Covariance for I34 and I33 = 0.034

(0.0060)

5.73

Error Covariance for I42 and I39 = 0.033

(0.0061)

5.47

Error Covariance for I42 and I40 = 0.036

(0.0065)

5.60

Error Covariance for I43 and I40 = 0.037

(0.0066)

5.63

Error Covariance for I43 and I42 = 0.038

(0.0065)

5.90

Error Covariance for I45 and I43 = -0.03

(0.0062)

-4.67

Error Covariance for I48 and I39 = 0.033

(0.0065)

5.04

Error Covariance for I50 and I49 = 0.037

(0.0069)

5.38

Correlation Matrix of Independent Variables

equation

1.00

SI was written to file fort.11

Goodness of Fit Statistics

Degrees of Freedom = 1146

Minimum Fit Function Chi-Square = 2544.58 (P = 0.0)

Normal Theory Weighted Least Squares Chi-Square = 2759.64 (P = 0.0)

Chi-Square Difference with 1 Degree of Freedom = 37.47 (P = 0.0)

Estimated Non-centrality Parameter (NCP) = 1613.64

90 Percent Confidence Interval for NCP = (1463.59 ; 1771.34)

Minimum Fit Function Value = 2.46

Population Discrepancy Function Value (F0) = 1.56

90 Percent Confidence Interval for F0 = (1.42 ; 1.71)

Root Mean Square Error of Approximation (RMSEA) = 0.037

90 Percent Confidence Interval for RMSEA = (0.035 ; 0.039)

P-Value for Test of Close Fit (RMSEA < 0.05) = 1.00

Expected Cross-Validation Index (ECVI) = 2.92

90 Percent Confidence Interval for ECVI = (2.78 ; 3.07)

ECVI for Saturated Model = 2.47

ECVI for Independence Model = 20.37

Chi-Square for Independence Model with 1225 Degrees of Freedom = 20945.03

Independence AIC = 21045.03

Model AIC = 3017.64

Saturated AIC = 2550.00

Independence CAIC = 21342.09

Model CAIC = 3784.06
Saturated CAIC = 10125.02

Normed Fit Index (NFI) = 0.88
Non-Normed Fit Index (NNFI) = 0.92
PaSimony Normed Fit Index (PNFI) = 0.82
Comparative Fit Index (CFI) = 0.93
Incremental Fit Index (IFI) = 0.93
Relative Fit Index (RFI) = 0.87

Critical N (CN) = 512.64

Root Mean Square Residual (RMR) = 0.010
Standardized RMR = 0.045
Goodness of Fit Index (GFI) = 0.90
Adjusted Goodness of Fit Index (AGFI) = 0.89
PaSimony Goodness of Fit Index (PGFI) = 0.81

Factor Scores Regressions

	I1	I2	I3	I4	I5	I6	I7	I8	I9	I10	
equation	0.02	0.10	0.08	0.11	0.09	0.09	0.06	0.12	0.11	0.09	

	I11	I12	I13	I14	I15	I16	I17	I18	I19	I20	
equation	0.02	0.11	0.02	0.08	0.05	0.07	0.03	0.09	0.12	0.03	

	I21	I22	I23	I24	I25	I26	I27	I28	I29	I30	
equation	0.07	0.09	0.06	0.09	0.08	0.07	0.16	0.13	0.11	0.13	

	I31	I32	I33	I34	I35	I36	I37	I38	I39	I40	
equation	0.12	0.15	0.06	0.11	0.16	0.07	0.18	0.13	0.09	0.13	

	I41	I42	I43	I44	I45	I46	I47	I48	I49	I50	
equation	0.17	0.09	0.10	0.14	0.10	0.21	0.17	0.09	0.14	0.08	

Time used: 7.372Second

คำสั่งประมาณค่าพารามิเตอร์ด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์
จำนวนข้อสอบ 50 ข้อ กลุ่มตัวอย่าง 1,034 คน

item 3PL 1034 peSon 50 item

—

>GLOBAL DFName = 'C:\WINDOWS\Desktop\math.dat',

NPArm = 3,

LOGistic,

SAVe;

>SAVE PARm = 'math.PAR',

SCOr = 'math.SCO',

TSTat = 'math.TST';

>LENGTH NITems = (50);

>INPUT NTOtal = 50,

NALt = 2,

NIDchar = 4;

>ITEMS ;

>TEST1 TNAme = 'MATH',

INUmber = (1(1)50);

(4A1,50A1)

>CALIB CYCles = 50,

NEWton = 3,

CRIt = 0.0500,

ACCel = 1.0000,

CHIsquare = (22, 5);

>SCORE IDIst = 3,

SCtype = 3,

INFo = 2;

ผลการประมาณค่าพารามิเตอร์ข้อสอบด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์

BILOG-MG V3.0

REV 19990329.1300

BILOG-MG ITEM MAINTENANCE PROGRAM: LOGISTIC ITEM RESPONSE MODEL

*** BILOG-MG ITEM MAINTENANCE PROGRAM ***

*** PHASE 2 ***

item 3PL 1034 peSon 50 item

>CALIB CYCles = 50,

NEWton = 3,

CRIt = 0.0500,

ACCEl = 1.0000,

CHIsquare = (22, 5);

CALIBRATION PARAMETES

=====

MAXIMUM NUMBER OF EM CYCLES: 50

MAXIMUM NUMBER OF NEWTON CYCLES: 3

CONVERGENCE CRITERION: 0.0500

ACCELERATION CONSTANT: 1.0000

LATENT DISTRIBUTION: NORMAL PRIOR FOR EACH GROUP

PLOT EMPIRICAL VS. FITTED ICC'S: NO

DATA HANDLING: DATA ON SCRATCH FILE

CONSTRAINT DISTRIBUTION ON ASYMPTOTES: YES

CONSTRAINT DISTRIBUTION ON SLOPES: YES

CONSTRAINT DISTRIBUTION ON THRESHOLD: NO

SOURCE OF ITEM CONSTRAINT DISTRIBUTION

MEANS AND STANDARD DEVIATIONS: PROGRAM DEFAULTS

CALIBRATION OF MAINTEST

MATH

METHOD OF SOLUTION:

EM CYCLES (MAXIMUM OF 50)

FOLLOWED BY NEWTON-RAPHSON STEPS (MAXIMUM OF 3)

QUADRATURE POINTS AND PRIOR WEIGHTS:

	1	2	3	4	5
POINT	-0.4000E+01	-0.3429E+01	-0.2857E+01	-0.2286E+01	-0.1714E+01
WEIGHT	0.7648E-04	0.6387E-03	0.3848E-02	0.1673E-01	0.5245E-01

	6	7	8	9	10
POINT	-0.1143E+01	-0.5714E+00	-0.8882E-15	0.5714E+00	0.1143E+01
WEIGHT	0.1186E+00	0.1936E+00	0.2280E+00	0.1936E+00	0.1186E+00

	11	12	13	14	15
POINT	0.1714E+01	0.2286E+01	0.2857E+01	0.3429E+01	0.4000E+01
WEIGHT	0.5245E-01	0.1673E-01	0.3848E-02	0.6387E-03	0.7648E-04

CONSTRAINT DISTRIBUTIONS ON ITEM PARAMETES

(THRESHOLD, NORMAL; SLOPES, LOG-NORMAL; GUESSING, BETA)

ITEM	THRESHOLD		SLOPES		ASYMPTOTES	
	MU	SIGMA	MU	SIGMA	ALPHA	BETA

ITEM0001	-	-	1.000	1.649	11.00	11.00
ITEM0002	-	-	1.000	1.649	11.00	11.00
ITEM0003	-	-	1.000	1.649	11.00	11.00
ITEM0004	-	-	1.000	1.649	11.00	11.00
ITEM0005	-	-	1.000	1.649	11.00	11.00
ITEM0006	-	-	1.000	1.649	11.00	11.00
ITEM0007	-	-	1.000	1.649	11.00	11.00
ITEM0008	-	-	1.000	1.649	11.00	11.00
ITEM0009	-	-	1.000	1.649	11.00	11.00
ITEM0010	-	-	1.000	1.649	11.00	11.00
.						
.						
.						
ITEM0050	-	-	1.000	1.649	11.00	11.00

[E-M CYCLES]

```

-2 LOG LIKELIHOOD =    70738.502
CYCLE   1; LARGEST CHANGE=  5.10622
-2 LOG LIKELIHOOD =    62934.098
CYCLE   2; LARGEST CHANGE=  1.40957
-2 LOG LIKELIHOOD =    62322.709
CYCLE   3; LARGEST CHANGE=  1.47218
.
.
.
-2 LOG LIKELIHOOD:    62092.3515
CYCLE  16; LARGEST CHANGE=  0.04416
INTERVAL COUNTS FOR COMPUTATION OF ITEM CHI-SQUARES
-----
5. 275. 253. 271. 230.
-----
INTERVAL AVERAGE THETAS
-----
-1.761 -1.229 -0.393 0.497 1.354
-----
SUBTEST MATH; ITEM PARAMETES AFTER CYCLE 16

```

ITEM	INTERCEPT	SLOPE	THRESHOLD	LOADING	ASYMPTOTE	CHISQ	DF
	S.E.	S.E.	S.E.	S.E.	S.E.	(PROB)	
ITEM0001	-0.113	1.266	0.089	0.785	0.262	10.4	4.0
	0.262*	0.194*	0.199*	0.120*	0.064*	(0.0344)	
ITEM0002	0.694	1.700	-0.408	0.862	0.307	12.1	4.0
	0.250*	0.249*	0.190*	0.126*	0.074*	(0.0167)	
ITEM0003	0.005	1.347	-0.004	0.803	0.201	31.5	4.0
	0.207*	0.168*	0.155*	0.100*	0.054*	(0.0000)	
ITEM0004	0.600	2.061	-0.291	0.900	0.200	26.1	4.0
	0.196*	0.232*	0.117*	0.101*	0.052*	(0.0000)	
ITEM0005	0.009	1.878	-0.005	0.883	0.202	7.4	4.0
	0.221*	0.253*	0.120*	0.119*	0.049*	(0.1181)	
ITEM0006	0.288	1.983	-0.145	0.893	0.205	23.5	4.0
	0.208*	0.250*	0.119*	0.112*	0.051*	(0.0001)	
ITEM0007	-1.564	1.035	1.511	0.719	0.258	2.7	4.0
	0.448*	0.267*	0.202*	0.185*	0.047*	(0.6158)	
ITEM0008	-1.270	1.697	0.749	0.862	0.234	2.3	4.0
	0.349*	0.303*	0.109*	0.154*	0.038*	(0.6792)	

ITEM0009 | -1.385 | 1.385 | 1.000 | 0.811 | 0.293 | 0.4 4.0
| 0.416* | 0.295* | 0.148* | 0.172* | 0.044* | (0.9817)

ITEM0010 | -1.936 | 1.332 | 1.454 | 0.800 | 0.227 | 2.0 4.0
| 0.456* | 0.307* | 0.144* | 0.185* | 0.035* | (0.7362)

ITEM0011 | -3.343 | 1.597 | 2.094 | 0.847 | 0.173 | 3.4 4.0
| 0.728* | 0.456* | 0.226* | 0.242* | 0.022* | (0.4867)

ITEM0012 | -1.855 | 1.905 | 0.974 | 0.885 | 0.238 | 2.1 4.0
| 0.410* | 0.337* | 0.093* | 0.157* | 0.030* | (0.7189)

ITEM0013 | -2.655 | 1.486 | 1.787 | 0.830 | 0.331 | 0.9 4.0
| 0.762* | 0.498* | 0.213* | 0.278* | 0.031* | (0.9268)

ITEM0014 | -2.523 | 2.107 | 1.197 | 0.903 | 0.252 | 1.4 4.0
| 0.562* | 0.453* | 0.088* | 0.194* | 0.026* | (0.8435)

ITEM0015 | -2.551 | 1.402 | 1.820 | 0.814 | 0.215 | 5.4 4.0
| 0.586* | 0.385* | 0.189* | 0.224* | 0.029* | (0.2460)

ITEM0016 | -1.996 | 1.444 | 1.382 | 0.822 | 0.235 | 7.4 4.0
| 0.473* | 0.335* | 0.132* | 0.190* | 0.034* | (0.1182)

ITEM0017 | -4.246 | 2.053 | 2.068 | 0.899 | 0.202 | 2.5 4.0
| 1.149* | 0.708* | 0.223* | 0.310* | 0.020* | (0.6496)

ITEM0018 | -3.099 | 1.907 | 1.625 | 0.886 | 0.227 | 0.4 4.0
| 0.696* | 0.491* | 0.133* | 0.228* | 0.024* | (0.9797)

ITEM0019 | -1.850 | 1.221 | 1.515 | 0.774 | 0.190 | 4.4 4.0
| 0.416* | 0.276* | 0.151* | 0.175* | 0.036* | (0.3534)

ITEM0020 | -4.150 | 2.177 | 1.907 | 0.909 | 0.234 | 0.7 4.0
| 1.094* | 0.693* | 0.176* | 0.289* | 0.020* | (0.9493)

ITEM0021 | -4.687 | 3.496 | 1.341 | 0.961 | 0.320 | 2.5 4.0
| 1.117* | 0.859* | 0.077* | 0.236* | 0.020* | (0.6392)

ITEM0022 | -4.143 | 3.287 | 1.260 | 0.957 | 0.287 | 2.8 4.0
| 0.972* | 0.792* | 0.071* | 0.231* | 0.020* | (0.5904)

ITEM0023 | -5.083 | 3.465 | 1.467 | 0.961 | 0.254 | 2.5 4.0
| 0.977* | 0.678* | 0.076* | 0.188* | 0.018* | (0.6443)

ITEM0024 | -9.119 | 6.621 | 1.377 | 0.989 | 0.193 | 5.5 4.0
| 2.035* | 1.604* | 0.061* | 0.240* | 0.014* | (0.2421)

ITEM0025 | -2.041 | 1.031 | 1.979 | 0.718 | 0.172 | 4.0 4.0
| 0.447* | 0.271* | 0.238* | 0.188* | 0.035* | (0.4056)

ITEM0026 | -1.929 | 0.948 | 2.034 | 0.688 | 0.135 | 15.4 4.0
| 0.367* | 0.226* | 0.253* | 0.164* | 0.033* | (0.0040)

ITEM0027 | -1.658 | 1.370 | 1.210 | 0.808 | 0.126 | 4.9 4.0
| 0.296* | 0.225* | 0.106* | 0.132* | 0.029* | (0.2996)

ITEM0028 | -1.078 | 1.041 | 1.035 | 0.721 | 0.154 | 18.1 4.0
| 0.254* | 0.165* | 0.148* | 0.114* | 0.040* | (0.0012)

ITEM0029 | -1.480 | 1.115 | 1.328 | 0.744 | 0.218 | 2.7 4.0
| 0.377* | 0.235* | 0.164* | 0.157* | 0.043* | (0.6143)

ITEM0030 | -2.152 | 1.973 | 1.091 | 0.892 | 0.273 | 0.5 4.0
| 0.505* | 0.410* | 0.094* | 0.185* | 0.030* | (0.9772)

ITEM0031 | -2.723 | 2.258 | 1.206 | 0.914 | 0.276 | 1.2 4.0
| 0.591* | 0.466* | 0.087* | 0.189* | 0.025* | (0.8772)

ITEM0032 | -1.788 | 1.800 | 0.993 | 0.874 | 0.251 | 5.6 4.0
| 0.427* | 0.347* | 0.099* | 0.168* | 0.033* | (0.2276)

ITEM0033 | -3.640 | 2.063 | 1.764 | 0.900 | 0.216 | 2.1 4.0
| 0.839* | 0.565* | 0.147* | 0.246* | 0.021* | (0.7087)

ITEM0034 | -3.629 | 2.503 | 1.450 | 0.929 | 0.182 | 2.8 4.0
| 0.730* | 0.549* | 0.085* | 0.204* | 0.019* | (0.5974)

ITEM0035 | -2.144 | 1.893 | 1.133 | 0.884 | 0.167 | 1.8 4.0
| 0.395* | 0.319* | 0.083* | 0.149* | 0.025* | (0.7787)

ITEM0036 | -2.710 | 1.211 | 2.238 | 0.771 | 0.219 | 5.1 4.0
| 0.659* | 0.390* | 0.311* | 0.248* | 0.030* | (0.2757)

ITEM0037 | -2.211 | 1.952 | 1.132 | 0.890 | 0.144 | 3.6 4.0
| 0.373* | 0.298* | 0.078* | 0.136* | 0.023* | (0.4696)

ITEM0038 | -2.655 | 1.817 | 1.461 | 0.876 | 0.184 | 2.2 4.0
| 0.526* | 0.389* | 0.107* | 0.187* | 0.024* | (0.6961)

ITEM0039 | -2.519 | 1.805 | 1.396 | 0.875 | 0.198 | 2.5 4.0
| 0.521* | 0.392* | 0.106* | 0.190* | 0.026* | (0.6400)

ITEM0040 | -0.711 | 1.906 | 0.373 | 0.886 | 0.322 | 10.0 4.0
| 0.347* | 0.336* | 0.133* | 0.156* | 0.046* | (0.0402)

ITEM0041 | -1.238 | 1.833 | 0.675 | 0.878 | 0.278 | 9.5 4.0
| 0.384* | 0.347* | 0.114* | 0.166* | 0.040* | (0.0498)

ITEM0042 | -2.279 | 2.050 | 1.112 | 0.899 | 0.226 | 5.2 4.0
| 0.509* | 0.438* | 0.085* | 0.192* | 0.028* | (0.2689)

ITEM0043 | -0.367 | 1.294 | 0.284 | 0.791 | 0.352 | 11.8 4.0
| 0.354* | 0.260* | 0.232* | 0.159* | 0.067* | (0.0187)

ITEM0044 | -1.659 | 1.641 | 1.011 | 0.854 | 0.226 | 5.2 4.0
| 0.411* | 0.337* | 0.106* | 0.175* | 0.036* | (0.2653)

ITEM0045 | -4.587 | 2.745 | 1.671 | 0.940 | 0.243 | 5.7 4.0
| 1.140* | 0.768* | 0.112* | 0.263* | 0.018* | (0.2227)

ITEM0046 | -1.022 | 1.761 | 0.580 | 0.870 | 0.187 | 9.9 4.0
| 0.284* | 0.265* | 0.100* | 0.131* | 0.037* | (0.0414)

ITEM0047 | -1.565 | 1.527 | 1.025 | 0.837 | 0.164 | 4.3 4.0
| 0.306* | 0.223* | 0.101* | 0.122* | 0.031* | (0.3714)

ITEM0048 | -2.025 | 1.505 | 1.346 | 0.833 | 0.276 | 3.0 4.0
| 0.503* | 0.352* | 0.132* | 0.195* | 0.034* | (0.5634)

```

ITEM0049 | -1.097 | 1.657 | 0.662 | 0.856 | 0.303 | 2.2 4.0
          | 0.361* | 0.284* | 0.134* | 0.147* | 0.043* | (0.7066)
ITEM0050 | -1.431 | 1.104 | 1.296 | 0.741 | 0.225 | 5.6 4.0
          | 0.392* | 0.250* | 0.171* | 0.168* | 0.046* | (0.2284)

```

* STANDARD ERROR

LARGEST CHANGE = 0.044161 305.0 200.0
(0.0000)

PARAMETER	MEAN	STN DEV
ASYMPTOTE	0.229	0.052
SLOPE	1.873	0.894
LOG(SLOPE)	0.555	0.358
THRESHOLD	1.145	0.640

QUADRATURE POINTS, POSTERIOR WEIGHTS, MEAN AND S.D.:

	1	2	3	4	5
POINT	-0.3885E+01	-0.3327E+01	-0.2769E+01	-0.2211E+01	-0.1653E+01
POSTERIOR	0.1032E-03	0.8814E-03	0.5343E-02	0.2245E-01	0.6487E-01

	6	7	8	9	10
POINT	-0.1096E+01	-0.5377E+00	0.2012E-01	0.5780E+00	0.1136E+01
POSTERIOR	0.1274E+00	0.1755E+00	0.1923E+00	0.1901E+00	0.1543E+00

	11	12	13	14	15
POINT	0.1694E+01	0.2252E+01	0.2809E+01	0.3367E+01	0.3925E+01
POSTERIOR	0.5489E-01	0.1081E-01	0.9706E-03	0.2187E-04	0.1560E-06

MEAN 0.00000

S.D. 1.00000

78144 BYTES OF NUMERICAL WORKSPACE USED OF 8192000 AVAILABLE IN PHASE-2

5308 BYTES OF CHARACTER WORKSPACE USED OF 2048000 AVAILABLE IN PHASE-2

10/24/2006 14:37:34

ภาคผนวก ค
การวิเคราะห์ข้อมูลการวิจัย

- ตัวอย่างคำสั่งและผลการประมาณค่าความสามารถด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์
- ตัวอย่างคำสั่งและผลการวิเคราะห์คะแนนมวลความรู้ตามทฤษฎีการตอบข้อสอบ
- ตัวอย่างคำสั่งและผลการวิเคราะห์คะแนนการเปลี่ยนแปลงด้วยโมเดลพหุระดับ

ตัวอย่างคำสั่งและผลการประมาณค่าประมาณค่าความสามารถก่อนเรียน
ด้วยโมเดลโลจิสติกแบบสามพารามิเตอร์
ขนาดกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบที่มีความยาว 20 ข้อ

คำสั่งการประมาณค่าความสามารถก่อนเรียน

item 3PL 100 peSon 20 item PRETEST



```
>GLOBAL DFName = '100-20.txt',
```

```
    NPArm = 3,
```

```
    LOGistic,
```

```
    SAVe;
```

```
>SAVE PARm = '100-20.PAR',
```

```
    SCORe = '100-20.SCO',
```

```
    TSTat = '100-20.TST';
```

```
>LENGTH NITems = (20);
```

```
>INPUT NTOtal = 20,
```

```
    NALt = 2,
```

```
    NIDchar = 5;
```

```
>ITEMS ;
```

```
>TEST1 TNAme = 'TEST0001',
```

```
    INUmber = (1(1)20);
```

```
(5A1,20A1)
```

```
>CALIB CYCles = 100,
```

```
    CRIt = 0.0010,
```

```
    ACCel = 1.0000,
```

```
    TPRior,
```

```
    GPRior;
```

```
>SCORE SCtype = 3,
```

```
    INFo = 2;
```

ผลการประมาณค่าความสามารถก่อนเรียน

BILOG-MG V3.0

BILOG-MG ITEM MAINTENANCE PROGRAM: LOGISTIC ITEM RESPONSE MODEL

*** LOGISTIC MODEL ITEM ANALYSER ***

*** PHASE 3 ***

item 3PL 100 peSon 20 item PRETEST

>SCORE SCtype = 3,

INFO = 2;

PARAMETES FOR SCORING, RESCALING, AND TEST AND ITEM INFORMATION

METHOD OF SCORING SUBJECTS: EXPECTATION A POSTERIORI

(EAP; BAYES ESTIMATION)

TYPE OF PRIOR: NORMAL

SCORES WRITTEN TO FILE 100-20.SCO

SCORES WRITTEN TO FILE 100-20.PH3

TYPE OF RESCALING: IN THE SAMPLE DISTRIBUTION

REFERENCE GROUP FOR RESCALING: GROUP: 1

ITEM AND TEST INFORMATION: TEST INFORMATION CURVES

TABLE OF ITEM INFORMATION INDICES

DOMAIN SCORE ESTIMATION: NONE REQUESTED

QUAD RESCALING CONSTANTS

TEST NAME POINTS SCALE LOCATION

1 TEST0001 10 1.000 0.000

SCORING

PRIOR DISTRIBUTION(S)

=====

EAP SUBJECT ESTIMATION, TEST: TEST0001

QUADRATURE POINTS AND PRIOR WEIGHTS, MEAN AND S.D.:

1 2 3 4 5

POINT -0.4000E+01 -0.3111E+01 -0.2222E+01 -0.1333E+01 -0.4444E+00

WEIGHT 0.1190E-03 0.2805E-02 0.3002E-01 0.1458E+00 0.3213E+00

6 7 8 9 10

POINT 0.4444E+00 0.1333E+01 0.2222E+01 0.3111E+01 0.4000E+01

WEIGHT 0.3213E+00 0.1458E+00 0.3002E-01 0.2805E-02 0.1190E-03

MEAN 0.0000

S.D. 1.0000

SUMMARY STATISTICS FOR SCORE ESTIMATES

```

=====
CORRELATIONS AMONG TEST SCORES
TEST0001
TEST0001 1.0000
MEANS, STANDARD DEVIATIONS, AND VARIANCES OF SCORE ESTIMATES
TEST: TEST0001
MEAN: -0.0853
S.D.: 0.7432
VARIANCE: 0.5523
ROOT-MEAN-SQUARE POSTERIOR STANDARD DEVIATIONS
TEST: TEST0001
RMS: 0.7736
VARIANCE: 0.5985
EMPIRICAL
RELIABILITY: 0.4799

MARGINAL LATENT DISTRIBUTION(S)
=====
MARGINAL LATENT DISTRIBUTION FOR TEST TEST0001
MEAN = -0.085
S.D. = 1.048
1 2 3 4 5
POINT -0.4000E+01 -0.3111E+01 -0.2222E+01 -0.1333E+01 -0.4444E+00
WEIGHT 0.1477E-03 0.3474E-02 0.3684E-01 0.1740E+00 0.3509E+00
6 7 8 9 10
POINT 0.4444E+00 0.1333E+01 0.2222E+01 0.3111E+01 0.4000E+01
WEIGHT 0.2710E+00 0.1073E+00 0.4857E-01 0.7591E-02 0.2492E-03
1
RESCALING WITH RESPECT TO SAMPLE DISTRIBUTION
-----
RESCALING CONSTANTS
TEST SCALE LOCATION
TEST0001 1.346 0.115
1
GROUP SUBJECT IDENTIFICATION MARGINAL
WEIGHT TEST TRIED RIGHT PERCENT ABILITY S.E. PROB
-----
1 00007 | |
1.00 TEST0001 20 8 40.00 | -0.2291 1.1274 | 0.000000
1 00008 | |
1.00 TEST0001 20 7 35.00 | 0.0880 1.1581 | 0.000003
1 00015 | |
1.00 TEST0001 20 3 15.00 | -0.7977 1.0067 | 0.000067

```


1	00430						
1.00	TEST0001	20	8	40.00	0.4967	1.1655	0.000004
1	00450						
1.00	TEST0001	20	7	35.00	-0.3682	1.0655	0.000002
1	00453						
1.00	TEST0001	20	3	15.00	-0.8789	1.0006	0.000022
1	00471						
1.00	TEST0001	20	6	30.00	-0.3903	1.0327	0.000016
1	00473						
1.00	TEST0001	20	11	55.00	0.9338	1.1636	0.000000
1	00480						
1.00	TEST0001	20	5	25.00	-0.4748	1.0370	0.000033
1	00486						
1.00	TEST0001	20	5	25.00	-0.5045	1.0477	0.000018
1	00494						
1.00	TEST0001	20	7	35.00	-0.1921	1.0533	0.000007
1	00500						
1.00	TEST0001	20	5	25.00	-0.4841	1.0455	0.000017
1	00509						
1.00	TEST0001	20	4	20.00	-0.5474	1.0499	0.000042
1	00512						
1.00	TEST0001	20	12	60.00	1.1994	1.0031	0.000001
1	00514						
1.00	TEST0001	20	7	35.00	-0.1735	1.0693	0.000004
1	00530						
1.00	TEST0001	20	10	50.00	0.1340	1.1108	0.000000
1	00536						
1.00	TEST0001	20	5	25.00	-0.5227	1.0685	0.000005
1	00567						
1.00	TEST0001	20	10	50.00	0.4394	1.1500	0.000000
1	00575						
1.00	TEST0001	20	6	30.00	-0.2973	1.1408	0.000002
1	00583						
1.00	TEST0001	20	7	35.00	-0.5292	1.0370	0.000001
1	00590						
1.00	TEST0001	20	6	30.00	-0.3293	1.0990	0.000003
1	00598						
1.00	TEST0001	20	6	30.00	-0.1972	1.1600	0.000003
1	00622						
1.00	TEST0001	20	8	40.00	0.5080	1.1703	0.000004

	1	00838						
1.00	TEST0001	20	10	50.00		0.5914	1.1608	0.000000
	1	00839						
1.00	TEST0001	20	18	90.00		3.7270	0.6916	0.000009
	1	00847						
1.00	TEST0001	20	15	75.00		2.9188	0.6487	0.000001
	1	00848						
1.00	TEST0001	20	16	80.00		3.1687	0.6267	0.000003
	1	00885						
1.00	TEST0001	20	11	55.00		1.3606	1.0186	0.000001
	1	00887						
1.00	TEST0001	20	13	65.00		2.0405	0.8424	0.000001
	1	00902						
1.00	TEST0001	20	7	35.00		-0.2041	1.1044	0.000002
	1	00945						
1.00	TEST0001	20	11	55.00		0.9378	1.0654	0.000001
	1	00964						
1.00	TEST0001	20	14	70.00		2.6071	0.7311	0.000002
	1	00966						
1.00	TEST0001	20	14	70.00		2.6604	0.7182	0.000001
	1	00989						
1.00	TEST0001	20	13	65.00		2.1149	0.8434	0.000001
	1	01008						
1.00	TEST0001	20	6	30.00		-0.6068	1.0387	0.000002
	1	01027						
1.00	TEST0001	20	5	25.00		-0.6167	1.0343	0.000012
	1	01052						
1.00	TEST0001	20	3	15.00		-0.8577	0.9993	0.000025
	1	01065						
1.00	TEST0001	20	7	35.00		0.0050	1.1281	0.000003
	1	01067						
1.00	TEST0001	20	6	30.00		-0.6086	1.0414	0.000001
	1	01083						
1.00	TEST0001	20	5	25.00		-0.4495	1.0974	0.000007

MAXIMUM INFORMATION APPROXIMATELY 0.4529D+01 AT 2.0000

44 BYTES OF NUMERICAL WORKSPACE USED OF 8192000 AVAILABLE IN PHASE-3

2732 BYTES OF CHARACTER WORKSPACE USED OF 2048000 AVAILABLE IN PHASE-3

ตัวอย่างคำสั่งและผลการวิเคราะห์คะแนนผลความรู้ก่อนเรียนตามทฤษฎีการตอบข้อสอบ
จำนวนกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบที่มีจำนวนข้อสอบ 20 ข้อ

คำสั่ง

100-20.BLM - CALIBRATION OF THE 20 MATH TEST THREE-PARAMETER MODEL

>GLOBAL DFName = '100-40.txt',

NPArm = 3,

LOGistic,

SAVe;

>SAVE SCOrE = '100-20.SCO';

>LENGTH NITems = (20);

>INPUT NTOtal = 40,

NALt = 2,

NIDchar = 5;

>ITEMS INAmes = (MA01(1)MA40);

>TEST1 TNAmE = 'MATH',

INUmber = (1, 4, 6, 8, 9, 12, 14, 16, 17(2)23, 26, 27, 31,

32, 33, 38, 39, 40);

(5A1,40A1)

>CALIB SElect = (0),

ACCel = 1.0000,

TPRior,

GPRior;

>SCORE DOMain = 20,

FILE = '20.PAR.txt';

ผลการวิเคราะห์คะแนนความรู้อ่อนเรียนตามทฤษฎีการตอบข้อสอบ

BILOG-MG V3.0

BILOG-MG ITEM MAINTENANCE PROGRAM: LOGISTIC ITEM RESPONSE MODEL

*** LOGISTIC MODEL ITEM ANALYSER ***

*** PHASE 3 ***

100-20.BLM - CALIBRATION OF THE 20 MATH TEST THREE-PARAMETER MODEL

>SCORE DOMain = 20,

FILE = '20.PAR.txt';

PARAMETES FOR SCORING, RESCALING, AND TEST AND ITEM INFORMATION

METHOD OF SCORING SUBJECTS: EXPECTATION A POSTERIORI

(EAP; BAYES ESTIMATION)

TYPE OF PRIOR: NORMAL

SCORES WRITTEN TO FILE 100-20.SCO

SCORES WRITTEN TO FILE 100-20.PH3

TYPE OF RESCALING: NONE REQUESTED

ITEM AND TEST INFORMATION: NONE REQUESTED

DOMAIN SCORE ESTIMATION: DOMITEM = 20

DOMAIN FILE = 20.PAR.TXT

QUAD

TEST NAME POINTS

1 MATH 10

1

SCORING

PRIOR DISTRIBUTION(S)

=====

EAP SUBJECT ESTIMATION, TEST: MATH

QUADRATURE POINTS AND PRIOR WEIGHTS, MEAN AND S.D.:

	1	2	3	4	5
POINT	-0.4000E+01	-0.3111E+01	-0.2222E+01	-0.1333E+01	-0.4444E+00
WEIGHT	0.1190E-03	0.2805E-02	0.3002E-01	0.1458E+00	0.3213E+00
	6	7	8	9	10

POINT 0.4444E+00 0.1333E+01 0.2222E+01 0.3111E+01 0.4000E+01
 WEIGHT 0.3213E+00 0.1458E+00 0.3002E-01 0.2805E-02 0.1190E-03
 MEAN 0.0000
 S.D. 1.0000
 1

GROUP	SUBJECT IDENTIFICATION	DOMAIN SCORE	S.E.	MARGINAL
WEIGHT	TEST	TRIED RIGHT PERCENT	ABILITY	S.E. PROB

1	00007		28.78	6.54
1.00 * MATH	20 8 40.00	-0.3540	0.7871	0.000001
1	00008		28.34	6.08
1.00 * MATH	20 7 35.00	-0.4079	0.7712	0.000001
1	00015		27.21	5.15
1.00 * MATH	20 3 15.00	-0.5617	0.7631	0.000001
1	00022		27.52	5.40
1.00 * MATH	20 4 20.00	-0.5178	0.7651	0.000000
1	00029		27.91	5.97
1.00 * MATH	20 7 35.00	-0.4635	0.7994	0.000000
1	00033		27.17	5.10
1.00 * MATH	20 3 15.00	-0.5688	0.7613	0.000000
1	00042		28.83	6.42
1.00 * MATH	20 8 40.00	-0.3480	0.7679	0.000001
1	00043		28.65	6.21
1.00 * MATH	20 7 35.00	-0.3687	0.7579	0.000001
1	00047		29.19	6.91
1.00 * MATH	20 9 45.00	-0.3052	0.7925	0.000001
1	00052		27.86	5.55
1.00 * MATH	20 4 20.00	-0.4707	0.7486	0.000001
1	00057		27.65	5.44
1.00 * MATH	20 4 20.00	-0.4995	0.7562	0.000001
1	00078		28.68	6.80
1.00 * MATH	20 9 45.00	-0.3661	0.8273	0.000001
1	00102		27.99	5.84
1.00 * MATH	20 7 35.00	-0.4528	0.7738	0.000001
1	00125		28.02	5.79
1.00 * MATH	20 6 30.00	-0.4494	0.7651	0.000001
1	00135		35.16	13.52
1.00 * MATH	20 13 65.00	0.2296	0.9743	0.000001
1	00158		27.19	5.19
1.00 * MATH	20 3 15.00	-0.5661	0.7722	0.000000

1	00162				28.23	5.95	
1.00 *	MATH	20	6	30.00	-0.4224	0.7655	0.000001
1	00178				27.32	5.24	
1.00 *	MATH	20	4	20.00	-0.5464	0.7647	0.000001
1	00181				28.35	5.98	
1.00 *	MATH	20	5	25.00	-0.4062	0.7566	0.000001
1	00182				27.48	5.42	
1.00 *	MATH	20	5	25.00	-0.5238	0.7727	0.000001
1	00207				27.68	5.42	
1.00 *	MATH	20	2	10.00	-0.4946	0.7498	0.000001
1	00214				27.91	5.60	
1.00 *	MATH	20	5	25.00	-0.4644	0.7516	0.000001
1	00243				34.27	12.81	
1.00 *	MATH	20	12	60.00	0.1634	0.9742	0.000001
1	00250				31.08	8.76	
1.00 *	MATH	20	12	60.00	-0.1080	0.8400	0.000001
1	00270				27.86	5.63	
1.00 *	MATH	20	4	20.00	-0.4702	0.7594	0.000001
1	00275				27.49	5.40	
1.00 *	MATH	20	5	25.00	-0.5218	0.7679	0.000001
1	00279				27.89	5.62	
1.00 *	MATH	20	3	15.00	-0.4664	0.7554	0.000001
1	00306				27.37	5.32	
1.00 *	MATH	20	5	25.00	-0.5395	0.7701	0.000001
1	00314				27.34	5.34	
1.00 *	MATH	20	4	20.00	-0.5432	0.7760	0.000000
1	00325				27.05	5.03	
1.00 *	MATH	20	5	25.00	-0.5859	0.7647	0.000000
1	00327				28.11	5.80	
1.00 *	MATH	20	6	30.00	-0.4373	0.7567	0.000001
1	00328				28.23	5.88	
1.00 *	MATH	20	4	20.00	-0.4223	0.7562	0.000001
1	00337				27.80	5.56	
1.00 *	MATH	20	4	20.00	-0.4793	0.7568	0.000001
1	00341				27.13	5.08	
1.00 *	MATH	20	3	15.00	-0.5750	0.7633	0.000000
1	00348				27.46	5.41	
1.00 *	MATH	20	3	15.00	-0.5266	0.7735	0.000000
1	00360				27.71	5.44	
1.00 *	MATH	20	4	20.00	-0.4911	0.7503	0.000001

1	00384				27.14	5.13	
1.00 *	MATH	20	2	10.00	-0.5723	0.7689	0.000000
1	00385				26.98	4.98	
1.00 *	MATH	20	3	15.00	-0.5970	0.7669	0.000000
1	00387				27.74	5.45	
1.00 *	MATH	20	3	15.00	-0.4871	0.7482	0.000001
1	00407				28.02	5.92	
1.00 *	MATH	20	7	35.00	-0.4493	0.7822	0.000001
1	00410				29.34	7.35	
1.00 *	MATH	20	10	50.00	-0.2878	0.8299	0.000001
1	00424				36.02	14.09	
1.00 *	MATH	20	13	65.00	0.2902	0.9666	0.000001
1	00425				27.62	5.39	
1.00 *	MATH	20	3	15.00	-0.5041	0.7533	0.000001
1	00430				28.72	6.50	
1.00 *	MATH	20	8	40.00	-0.3602	0.7865	0.000001
1	00450				28.16	6.10	
1.00 *	MATH	20	7	35.00	-0.4304	0.7901	0.000001
1	00453				27.22	5.11	
1.00 *	MATH	20	3	15.00	-0.5611	0.7573	0.000001
1	00471				28.29	5.92	
1.00 *	MATH	20	6	30.00	-0.4139	0.7549	0.000001
1	00473				30.69	9.19	
1.00 *	MATH	20	11	55.00	-0.1458	0.9110	0.000001
1	00480				28.01	5.70	
1.00 *	MATH	20	5	25.00	-0.4509	0.7536	0.000001
1	00486				28.15	5.73	
1.00 *	MATH	20	5	25.00	-0.4326	0.7438	0.000001
1	00494				28.90	6.42	
1.00 *	MATH	20	7	35.00	-0.3397	0.7617	0.000001
1	00500				28.13	5.79	
1.00 *	MATH	20	5	25.00	-0.4349	0.7536	0.000001
1	00509				27.54	5.29	
1.00 *	MATH	20	4	20.00	-0.5140	0.7462	0.000001
1	00512				31.30	9.11	
1.00 *	MATH	20	12	60.00	-0.0867	0.8572	0.000001
1	00514				28.66	6.28	
1.00 *	MATH	20	7	35.00	-0.3685	0.7657	0.000001
1	00530				29.08	7.20	
1.00 *	MATH	20	10	50.00	-0.3185	0.8364	0.000001

1	00536				27.56	5.48	
1.00 *	MATH	20	5	25.00	-0.5114	0.7712	0.000000
1	00567				29.05	6.98	
1.00 *	MATH	20	10	50.00	-0.3212	0.8135	0.000001
1	00575				27.51	5.55	
1.00 *	MATH	20	6	30.00	-0.5185	0.7859	0.000000
1	00583				28.52	6.19	
1.00 *	MATH	20	7	35.00	-0.3854	0.7678	0.000001
1	00590				27.53	5.50	
1.00 *	MATH	20	6	30.00	-0.5157	0.7773	0.000000
1	00598				27.59	5.56	
1.00 *	MATH	20	6	30.00	-0.5083	0.7793	0.000000
1	00622				28.47	6.11	
1.00 *	MATH	20	8	40.00	-0.3919	0.7620	0.000001
1	00624				27.96	5.70	
1.00 *	MATH	20	4	20.00	-0.4574	0.7594	0.000001
1	00631				28.18	5.81	
1.00 *	MATH	20	5	25.00	-0.4287	0.7512	0.000001
1	00641				28.73	6.31	
1.00 *	MATH	20	8	40.00	-0.3590	0.7623	0.000001
1	00642				28.10	6.07	
1.00 *	MATH	20	7	35.00	-0.4386	0.7932	0.000001
1	00651				29.21	6.65	
1.00 *	MATH	20	8	40.00	-0.3032	0.7613	0.000001
1	00657				27.66	5.61	
1.00 *	MATH	20	6	30.00	-0.4982	0.7783	0.000001
1	00660				28.32	6.10	
1.00 *	MATH	20	7	35.00	-0.4100	0.7753	0.000001
1	00668				28.70	6.28	
1.00 *	MATH	20	6	30.00	-0.3628	0.7614	0.000001
1	00672				29.44	6.90	
1.00 *	MATH	20	10	50.00	-0.2774	0.7709	0.000001
1	00694				28.78	6.28	
1.00 *	MATH	20	7	35.00	-0.3533	0.7547	0.000001
1	00698				27.94	5.62	
1.00 *	MATH	20	5	25.00	-0.4602	0.7502	0.000001
1	00699				28.24	5.93	
1.00 *	MATH	20	7	35.00	-0.4207	0.7608	0.000001
1	00723				28.03	5.71	
1.00 *	MATH	20	5	25.00	-0.4484	0.7536	0.000001

1 00741					28.03	5.72	
1.00 * MATH	20	4	20.00		-0.4476	0.7543	0.000001
1 00743					27.78	5.49	
1.00 * MATH	20	4	20.00		-0.4819	0.7499	0.000001
1 00745					28.57	6.26	
1.00 * MATH	20	7	35.00		-0.3786	0.7709	0.000001
1 00775					27.83	5.59	
1.00 * MATH	20	4	20.00		-0.4747	0.7580	0.000001
1 00793					27.72	5.48	
1.00 * MATH	20	4	20.00		-0.4896	0.7544	0.000001
1 00796					27.83	5.54	
1.00 * MATH	20	4	20.00		-0.4744	0.7511	0.000001
1 00832					28.54	6.21	
1.00 * MATH	20	7	35.00		-0.3827	0.7681	0.000001
1 00838					29.70	8.08	
1.00 * MATH	20	10	50.00		-0.2481	0.8789	0.000001
1 00839					68.98	17.50	
1.00 * MATH	20	18	90.00		1.6878	0.6250	0.000029
1 00847					50.08	19.21	
1.00 * MATH	20	15	75.00		1.0116	0.7719	0.000004
1 00848					48.70	18.37	
1.00 * MATH	20	16	80.00		0.9553	0.7646	0.000005
1 00885					30.24	8.45	
1.00 * MATH	20	11	55.00		-0.1911	0.8720	0.000001
1 00887					37.25	15.20	
1.00 * MATH	20	13	65.00		0.3717	0.9767	0.000002
1 00902					27.95	5.94	
1.00 * MATH	20	7	35.00		-0.4585	0.7918	0.000000
1 00945					30.18	7.94	
1.00 * MATH	20	11	55.00		-0.1968	0.8237	0.000001
1 00964					37.39	14.95	
1.00 * MATH	20	14	70.00		0.3803	0.9542	0.000002
1 00966					41.57	17.61	
1.00 * MATH	20	14	70.00		0.6228	0.9296	0.000002
1 00989					35.53	13.96	
1.00 * MATH	20	13	65.00		0.2562	0.9844	0.000001
1 01008					28.02	5.77	
1.00 * MATH	20	6	30.00		-0.4496	0.7622	0.000001
1 01027					27.77	5.62	
1.00 * MATH	20	5	25.00		-0.4834	0.7685	0.000001

1	01052				27.62	5.51	
1.00 *	MATH	20	3	15.00	-0.5037	0.7692	0.000000
1	01065				28.58	6.21	
1.00 *	MATH	20	7	35.00	-0.3782	0.7641	0.000001
1	01067				27.91	5.78	
1.00 *	MATH	20	6	30.00	-0.4641	0.7754	0.000001
1	01083				27.96	5.67	
1.00 *	MATH	20	5	25.00	-0.4565	0.7547	0.000001

NOTE: THE '**' INDICATES THAT THE TEST IS NOT CALIBRATED.

SUMMARY STATISTICS FOR SCORE ESTIMATES

=====

CORRELATIONS AMONG TEST SCORES

MATH

MATH	1.0000
------	--------

MEANS, STANDARD DEVIATIONS, AND VARIANCES OF SCORE ESTIMATES

TEST:	MATH
-------	------

MEAN:	-0.3284
-------	---------

S.D.:	0.3547
-------	--------

VARIANCE:	0.1258
-----------	--------

ROOT-MEAN-SQUARE POSTERIOR STANDARD DEVIATIONS

TEST:	MATH
-------	------

RMS:	0.7883
------	--------

VARIANCE:	0.6214
-----------	--------

EMPIRICAL

RELIABILITY:	0.1684
--------------	--------

MARGINAL LATENT DISTRIBUTION(S)

=====

MARGINAL LATENT DISTRIBUTION FOR TEST MATH

MEAN	=	-0.328
------	---	--------

S.D.	=	0.833
------	---	-------

	1	2	3	4	5
POINT	-0.4000E+01	-0.3111E+01	-0.2222E+01	-0.1333E+01	-0.4444E+00
WEIGHT	0.1665E-03	0.3903E-02	0.4132E-01	0.1959E+00	0.4029E+00

	6	7	8	9	10
POINT	0.4444E+00	0.1333E+01	0.2222E+01	0.3111E+01	0.4000E+01
WEIGHT	0.2953E+00	0.5516E-01	0.5272E-02	0.1928E-04	0.1390E-07

44 BYTES OF NUMERICAL WORKSPACE USED OF 8192000 AVAILABLE IN PHASE-3

2892 BYTES OF CHARACTER WORKSPACE USED OF 2048000 AVAILABLE IN PHASE-3

ตัวอย่างคำสั่งและผลการวิเคราะห์คะแนนการเปลี่ยนแปลงด้วยโมเดลพหุระดับ
จำนวนกลุ่มตัวอย่าง 100 คน ด้วยแบบทดสอบที่มีจำนวนข้อสอบ 20 ข้อ

คำสั่ง

```
#HLM2 MDM CREATION TEMPLATE
growthmodel:y
rawdattype:spss
l1fname:C:\WINDOWS\Desktop\research analysis\HLM analysis\100-20\100-20.sav
l2fname:C:\WINDOWS\Desktop\research analysis\HLM analysis\100-20\100.sav
l1missing:n
timeofdeletion:now
mdmname:100-20
*begin l1vaS
level2id:ID
TIME
SCORES
*end l1vaS
*begin l2vaS
level2id:ID
STATUS
*end l2vaS
```

**ผลการวิเคราะห์คะแนนการเปลี่ยนแปลงด้วยโมเดลพหุระดับเชิงเส้นตรง แบบ 2 ระดับ
ขนาดกลุ่มตัวอย่าง 100 คน สอบซ้ำสามครั้งด้วยแบบทดสอบจำนวน 20 ข้อ**

Program: HLM 6 Hierarchical Linear and Nonlinear Modeling
 AuthoS: Stephen Raudenbush, Tony Bryk, & Richard Congdon
 Publisher: Scientific Software International, Inc. (c) 2000
 techsupport@ssicentral.com
 www.ssicentral.com

Module: HLM2.EXE (6.02.25138.2)

Date: 1 March 2007, Sunday

Time: 9:22:29

SPECIFICATIONS FOR THIS HLM2 RUN

Problem Title: no title

The data source for this run = 100-20

The command file for this run = C:\WINDOWS\Desktop\research analysis\HLM analysis\100-20\100-20un1.hlm

Output file name = C:\WINDOWS\Desktop\research analysis\HLM analysis\100-20\100-20un1.txt

The maximum number of level-1 units = 300

The maximum number of level-2 units = 100

The maximum number of iterations = 100

Method of estimation: restricted maximum likelihood

Weighting Specification

Weight

Variable

Weighting? Name Normalized?

Level 1 no

Level 2 no

Precision no

The outcome variable is SCORES

The model specified for the fixed effects was:

Level-1

Level-2

Coefficients

PredictoS

INTRCPT1, P0 INTRCPT2, B00

TIME slope, P1 INTRCPT2, B10

The model specified for the covariance components was:

Sigma squared (constant across level-2 units)

Tau dimensions

INTRCPT1

TIME slope

Summary of the model specified (in equation format)

Level-1 Model

$$Y = P0 + P1*(TIME) + E$$

Level-2 Model

$$P0 = B00 + R0$$

$$P1 = B10 + R1$$

Level-1 OLS regressions

Level - 2 Unit	INTRCPT1	TIME slope

1	7.66667	4.00000
2	7.66667	4.00000
3	4.50000	6.50000
4	2.66667	5.00000
5	8.16667	3.50000
6	3.16667	4.50000
7	8.66667	3.00000
8	7.83333	3.50000
9	8.00000	2.00000
10	3.66667	6.00000
.		
.		
.		
100	4.33333	6.00000

The average OLS level-1 coefficient for INTRCPT1 = 7.02500

The average OLS level-1 coefficient for TIME = 2.55500

Least Squares Estimates

sigma_squared = 17.61569

The outcome variable is SCORES
Least-squares estimates of fixed effects

Standard					
Fixed Effect	Coefficient	Error	T-ratio	d.f.	P-value

For INTRCPT1, P0					
INTRCPT2, B00	7.025000	0.383141	18.335	298	0.000
For TIME slope, P1					
INTRCPT2, B10	2.555000	0.296780	8.609	298	0.000

The outcome variable is SCORES
Least-squares estimates of fixed effects
(with robust standard errors)

Standard					
Fixed Effect	Coefficient	Error	T-ratio	d.f.	P-value

For INTRCPT1, P0					
INTRCPT2, B00	7.025000	0.340281	20.645	298	0.000
For TIME slope, P1					
INTRCPT2, B10	2.555000	0.212355	12.032	298	0.000

The least-squares likelihood value = -855.794423

Deviance = 1711.58885

Number of estimated parameters = 1

STARTING VALUES

sigma(0)_squared = 6.53833

Tau(0)		
INTRCPT1,P0	6.24745	1.36205
TIME,P1	1.36205	1.28586

The outcome variable is SCORES

Estimation of fixed effects

(Based on starting values of covariance components)

Fixed Effect	Standard	Approx.			
	Coefficient	Error	T-ratio	d.f.	P-value

	For	INTRCPT1, P0			
INTRCPT2, B00	7.025000	0.341995	20.541	99	0.000
	For	TIME slope, P1			
INTRCPT2, B10	2.555000	0.213425	11.971	99	0.000

The value of the likelihood function at iteration 1 = -8.042072E+002

Iterations stopped due to small change in likelihood function

***** ITERATION 2 *****

Sigma_squared = 6.53833

Tau		
INTRCPT1,P0	6.24745	1.36205
TIME,P1	1.36205	1.28586

Tau (as correlations)		
INTRCPT1,P0	1.000	0.481
TIME,P1	0.481	1.000

Random level-1 coefficient Reliability estimate

INTRCPT1, P0	0.534
TIME, P1	0.282

The value of the likelihood function at iteration 2 = -8.042072E+002

The outcome variable is SCORES

Final estimation of fixed effects:

	Standard	Approx.			
Fixed Effect	Coefficient	Error	T-ratio	d.f.	P-value

	For	INTRCPT1, P0				
INTRCPT2, B00	7.025000	0.341995	20.541	99	0.000	
	For	TIME slope, P1				
INTRCPT2, B10	2.555000	0.213425	11.971	99	0.000	

The outcome variable is SCORES

Final estimation of fixed effects

(with robust standard errors)

Fixed Effect	Standard	Approx.			P-value
	Coefficient	Error	T-ratio	d.f.	

	For	INTRCPT1, P0			
INTRCPT2, B00	7.025000	0.340281	20.645	99	0.000
	For	TIME slope, P1			
INTRCPT2, B10	2.555000	0.212355	12.032	99	0.000

Final estimation of variance components:

Random Effect	Standard	Variance	df	Chi-square	P-value
	Deviation	Component			
INTRCPT1, R0	2.49949	6.24745	99	212.51466	0.000
TIME slope, R1	1.13396	1.28586	99	137.93959	0.006
level-1, E	2.55702	6.53833			

Statistics for current covariance components model

Deviance = 1608.414307

Number of estimated parameters = 4

ตัวอย่าง ผลการวิเคราะห์ Residual Files level-1 ที่วิเคราะห์ด้วยโปรแกรม HLM

ของกลุ่มตัวอย่างขนาด 100 คน สอบเข้าสามครั้งด้วยแบบทดสอบจำนวน 20 ข้อ

L2ID	L1RESID	FITVAL	SIGMA	TIME	SCORES	S1
1	-.243	8.243	2.557	.000	8.000	1.000
1	-.415	11.415	2.557	1.000	11.000	1.000
1	1.413	14.587	2.557	2.000	16.000	1.000
2	-1.243	8.243	2.557	.000	7.000	1.000
2	1.585	11.415	2.557	1.000	13.000	1.000
2	.413	14.587	2.557	2.000	15.000	1.000
3	-4.690	7.690	2.557	.000	3.000	1.000
3	2.824	11.176	2.557	1.000	14.000	1.000
3	1.337	14.663	2.557	2.000	16.000	1.000
4	-1.700	5.700	2.557	.000	4.000	1.000
4	-3.340	8.340	2.557	1.000	5.000	1.000
4	3.020	10.980	2.557	2.000	14.000	1.000
5	-1.270	8.270	2.557	.000	7.000	1.000
5	2.644	11.356	2.557	1.000	14.000	1.000
5	-.442	14.442	2.557	2.000	14.000	1.000
6	-2.728	5.728	2.557	.000	3.000	1.000
6	-.282	8.282	2.557	1.000	8.000	1.000
6	1.165	10.835	2.557	2.000	12.000	1.000
7	-.298	8.298	2.557	.000	8.000	1.000
7	1.703	11.297	2.557	1.000	13.000	1.000
7	-.297	14.297	2.557	2.000	14.000	1.000
8	-1.063	8.063	2.557	.000	7.000	1.000
8	1.910	11.090	2.557	1.000	13.000	1.000
8	-.117	14.117	2.557	2.000	14.000	1.000
9	1.683	7.317	2.557	.000	9.000	1.000
9	-1.850	9.850	2.557	1.000	8.000	1.000
9	.617	12.383	2.557	2.000	13.000	1.000
10	-2.889	6.889	2.557	.000	4.000	1.000
10	-1.054	10.054	2.557	1.000	9.000	1.000
10	2.781	13.219	2.557	2.000	16.000	1.000
.						
.						
.						
100	-2.303	7.303	2.557	.000	5.000	12.000
100	-1.586	10.586	2.557	1.000	9.000	12.000
100	3.132	13.868	2.557	2.000	17.000	12.000

ตัวอย่าง ผลการวิเคราะห์ Residual Files level-2 ที่วิเคราะห์ด้วยโปรแกรม HLM

ของกลุ่มตัวอย่างขนาด 100 คน สอบเข้าสามครั้งด้วยแบบทดสอบจำนวน 20 ข้อ

L2ID	NJ	CHIPCT	MDIST	LNTOT VAR	OLSS VAR	MDS VAR	EB INTRCP	EB TIME	OL INTRCP	OL TIME	FV INTRCP	FV TIME	EC INTRCP	EC TIME	PV00	PV10	PV11
1	3	.517	.607	1.412	.255	-E1+36	1.218	.617	.642	1.445	7.025	2.555	8.243	3.172	1.536	-.180	.564
2	3	.490	.607	1.441	.650	-E1+36	1.218	.617	.642	1.445	7.025	2.555	8.243	3.172	1.536	-.180	.564
3	3	3.460	3.522	1.951	1.337	-E1+36	.665	.931	-2.525	3.945	7.025	2.555	7.690	3.486	1.536	-.180	.564
4	3	2.507	2.356	1.714	1.228	-E1+36	-1.325	.085	-4.358	2.445	7.025	2.555	5.700	2.640	1.536	-.180	.564
5	3	.310	.417	1.412	1.108	-E1+36	1.245	.531	1.142	.945	7.025	2.555	8.270	3.086	1.536	-.180	.564
6	3	1.548	1.698	1.518	.077	-E1+36	-1.297	-.001	-3.858	1.945	7.025	2.555	5.728	2.554	1.536	-.180	.564
7	3	.190	.354	1.191	.650	-E1+36	1.273	.444	1.642	.445	7.025	2.555	8.298	2.999	1.536	-.180	.564
8	3	.150	.332	1.348	.821	-E1+36	1.038	.472	.808	.945	7.025	2.555	8.063	3.027	1.536	-.180	.564
9	3	.050	.119	1.007	.973	-E1+36	.292	-.022	.975	-.555	7.025	2.555	7.317	2.533	1.536	-.180	.564
10	3	2.884	2.972	1.803	.255	-E1+36	-.136	.610	-3.358	3.445	7.025	2.555	6.889	3.165	1.536	-.180	.564
.																	
.																	
.																	
100	3	2.802	2.776	1.817	.650	-E1+36	.278	.727	-2.692	3.445	7.025	2.555	7.303	3.282	1.536	-.180	.564

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ-ชื่อสกุล	จรรยา สิงห์ทอง
สถานที่เกิด	จังหวัดชลบุรี
สถานที่อยู่ปัจจุบัน	261 ถนนรังสิต-นครนายก 51 ต.ประชาธิปไตย อ.ัญบุรี จ.ปทุมธานี
ตำแหน่งหน้าที่การงาน	ครู ค.ศ.3
สถานที่ทำงาน	ศูนย์บริการการศึกษานอกโรงเรียนเขตจอมทอง ศูนย์การศึกษานอกโรงเรียนกรุงเทพมหานคร 1 สำนักบริหารงานการศึกษานอกโรงเรียน สำนักปลัดกระทรวงศึกษาธิการ
ประวัติการศึกษา	
พ.ศ. 2528	กศ.บ. (การวัดผลการศึกษา) มหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2532	กศ.ม. (การวัดผลการศึกษา) มหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2550	กศ.ด. (การทดสอบและวัดผลการศึกษา) มหาวิทยาลัยศรีนครินทรวิโรฒ