

371-26013

ศ 818 ก

ร.3

การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า

23 ส.ค. 2547

บทคัดย่อ

ของ

สนิสา จุ้ยม่วงศรี

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุขบัณฑิต สาขาการทดสอบและวัดผลการศึกษา

ตุลาคม 2546

๒ 237699

นางสาวสุนิสา จัยม่วงศรี. (2546). การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้
คะแนนแบบหลายค่า. ปรินซ์นิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา).
กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม:
รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์, อาจารย์ ดร.ชูศักดิ์ ชัมภลิจิต,
รองศาสตราจารย์ ดร.สมสร วงษ์อยู่น้อย.

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาผลของการเทียบคะแนนที่ข้อสอบตรวจให้
คะแนนแบบหลายค่าตามวิธีการเทียบคะแนนในทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้แก่
การเทียบคะแนนเชิงเส้นตรงและการเทียบคะแนนอีควิเปอร์เซ็นไทล์ และตามวิธีในทฤษฎีการ
ตอบข้อสอบโมเดลเกรต เรสพอนส์ ได้แก่ การเทียบคะแนนวิธีค่าเฉลี่ยและส่วนเบี่ยงเบน
มาตรฐานของพารามิเตอร์ความยาก วิธีโค้งลักษณะแบบทดสอบ และวิธีปรับค่าพารามิเตอร์
พร้อมกัน ซึ่งเทียบคะแนนทั้งคะแนนจริงและคะแนนสังเกต โดยพิจารณาคุณภาพของการเทียบ
คะแนนแต่ละวิธีจาก ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ดัชนีคะแนนความ
แตกต่าง และดัชนีความแตกต่าง (C) จากการวิเคราะห์กลุ่มสอบทานผล

เครื่องมือที่ใช้ในการวิจัย ประกอบด้วยแบบทดสอบเทียบคะแนน 2 ฉบับ เป็นแบบ-
ทดสอบเลือกตอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล ได้แก่ ฉบับ X และฉบับ Y
ซึ่งผู้วิจัยสร้างขึ้นสำหรับนักเรียนชั้นมัธยมศึกษาปีที่สอง แต่ละฉบับประกอบด้วยข้อสอบที่มี
5 ตัวเลือก จำนวน 30 ข้อ รวมข้อสอบรวมจำนวน 10 ข้อ แต่ละข้อตรวจให้คะแนน 6 ค่า คือ 0
1 2 3 4 และ 5 คะแนน แบบทดสอบฉบับ X และฉบับ Y มีค่าความเชื่อมั่นเท่ากับ 0.937 และ
0.940 ตามลำดับ

กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่สอง สังกัดกรม
สามัญศึกษา ในเขตกรุงเทพมหานคร ซึ่งเลือกมาด้วยวิธีการสุ่มแบบแบ่งชั้น จำนวน 2,901 คน
แบ่งเป็น กลุ่มตัวอย่างเทียบคะแนนเพื่อนำผลการสอบไปวิเคราะห์เทียบคะแนนแบบทดสอบ
จำนวน 2,312 คน โดยการสุ่มอย่างมีระบบให้แต่ละคนทำแบบทดสอบฉบับ X หรือ ฉบับ Y
เพียงหนึ่งฉบับ ได้ผู้สอบฉบับ X จำนวน 1,155 คน และฉบับ Y จำนวน 1,157 คน อีกกลุ่มใช้
กลุ่มตัวอย่างสอบทานผลเพื่อนำผลการสอบไปตรวจสอบคุณภาพของการเทียบคะแนนในแต่ละ
วิธี จำนวน 589 คน แต่ละคนทำแบบทดสอบทั้งสองฉบับโดยทำแบบทดสอบแต่ละฉบับสลับกัน

ผลการวิจัยสรุปได้ดังนี้

1. คะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่สูงกว่าคะแนนของแบบทดสอบ
ฉบับ X จากการเทียบคะแนนทั้ง 5 วิธี 8 รูปแบบ โดยเฉพาะการเทียบคะแนนเชิงเส้นตรงให้
ค่าสูงกว่าตลอดช่วงมาตรวัดของ X

2. การแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนจากแบบทดสอบเป้าหมาย (ฉบับ X) มีความคล้ายคลึงกัน

3. ค่าดัชนีคะแนนความแตกต่าง RMSD (Root Mean Square Difference) ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก MAD (Mean Absolute Difference) ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก และ MSD (Mean Signed Difference) ไม่ถ่วงน้ำหนัก ให้ผลที่คล้ายคลึงกันคือฟังก์ชันการเทียบคะแนนจริงวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากที่สุด โดยเฉพาะฟังก์ชันการเทียบคะแนนอีคิวเปอร์เซ็นไทล์ สำหรับผลจากดัชนี MSD ถ่วงน้ำหนัก สรุปได้ว่าฟังก์ชันการเทียบคะแนนสังเกต วิธีปรับค่าพารามิเตอร์พร้อมกัน มีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากที่สุด โดยเฉพาะฟังก์ชันการเทียบคะแนนอีคิวเปอร์เซ็นไทล์ และดัชนี RMSD MAD และ MSD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนักรวม 6 ดัชนี ให้ผลที่สอดคล้องกัน คือฟังก์ชันการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบแตกต่างจากฟังก์ชันการเทียบคะแนนเชิงเส้นตรงมากที่สุด

4. ค่าดัชนีความแตกต่างจากการเทียบคะแนนแต่ละวิธี ให้ผลอยู่ในระดับที่ยอมรับได้ทั้งสิ้น คือมีค่าต่ำกว่ากำลังสองของ 0.05 ของค่าส่วนเบี่ยงเบนมาตรฐานของผลสอบจากฉบับ X และพบว่าความคลาดเคลื่อนของการเทียบคะแนนในแต่ละวิธีจากกลุ่มสอบทานผลไม่แตกต่างกัน

A STUDY OF EQUATING METHODS ON POLYTOMOUS ITEM SCORING

AN ABSTRACT
BY
SUNISA JUMONGSRI

Presented in partial fulfillment of the requirements
for the Doctor of Education degree in Testing and Measurement
at Srinakharinwirot University
October 2003

Sunisa Jumongsri.(2003). *A Study of Equating Methods on Polytomous Item Scoring*.
Dissertation, Ed. D (Testing and Educational Measurement). Bangkok :
Graduate School, Srinakharinwirot University. Advisor Committee: Assoc.
Prof.Dr.Boonchird Pinyoanuntapong, Dr. Choosak Khampalikit, Assoc.Prof.
Dr. Somson Wongyounoi.

The research aimed to investigate the results of equating two equivalent test forms with polytomous-scoring items by using two classical test theory (CTT) equating methods such as the linear and the equipercentile, and three item response theory (IRT) equating methods under the Graded Response Model (GRM) : the mean and sigma, the test response function and the concurrent calibration. Under each IRT method, the tests were equated according to both true and observed scores. The similarity of Cumulative Score Distribution, the different score indices and the discrepancy index (C) from cross validation analysis were employed to identify the quality of equating.

Two equivalent forms of mathematics achievement test on reasoning were Form X and Form Y constructed by the researcher for Matayom Suksa II Students. Each form consisted of 30 items, including 10 anchor ones, with 5 alternatives, by using polytomous-scoring as 0, 1, 2, 3, 4, or 5 according to the criteria set by the degree of correctness of each alternative. The reliability of Form X and Form Y were 0.937 and 0.940 respectively.

The samples of 2,901 Matayom Suksa II Students were stratified randomly drawn from those studying in the secondary schools under the jurisdiction of the Department of General Education in Bangkok Metro Politan Area. They were divided into two groups of 2,312 and 589 students. The first group of students was randomly assigned to take different test forms. 1,155 students did Form X and 1,157 students did Form Y. The other group took both tests for the purpose of cross validation of equating results. Each student took each form at a time as the first student took Form X, but the second one took Form Y, and they changed the form of tests alternately.

The results were as follows :

1. Most equivalent scores of Form Y were higher than those of Form X for all equating methods. This was particularly true for the linear equating method of which the scores of Form Y was higher throughout the scale of Form X.

2. The distributions of cumulative equated score and the distribution of cumulative score of Form X were similar to all equating methods.

3. The different score indices : weighed and unweighed RMSD, weighed and unweighed MAD and unweighed MSD, yielded similar results. It was found ; however, that equating function of the mean and sigma true-score method was most similar to the function of the baseline equating methods, especially equipercentile equating method. Regarding to weighed MSD, the function of observed-score equating method based on concurrent calibration was most similar to the function of the baseline equating method, especially equipercentile equating methods. The consistent result of six indices was that, the function of observed-score equating method based on the test response function was most different from the function of linear equating method.

4. All equating methods yielded equally acceptable results with equating discrepancy indices lower than the square of 0.05 of the standard deviation of test Form X and the error of each equating method of cross validation group was not significantly different at 0.01 level.

การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า

ปริญญานิพนธ์
ของ
สุนิสา จั๋ยม่วงศรี

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุขฎีบัณฑิต สาขาการทดสอบและวัดผลการศึกษา

ตุลาคม 2546

ลิขสิทธิ์เป็นของ มหาวิทยาลัยศรีนครินทรวิโรฒ

ปริญญานิพนธ์ฉบับนี้ ได้รับทุนสนับสนุนการวิจัย จากทุนทบวงมหาวิทยาลัย
มหาวิทยาลัยศรีนครินทรวิโรฒ ประจำปีการศึกษา 2545

ปริญญานิพนธ์

เรื่อง

การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า

ของ

นางสาวสุนิสา จัยม่วงศรี

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษาดุริยบัณฑิต สาขาการทดสอบและวัดผลการศึกษา
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีบัณฑิตวิทยาลัย
(รองศาสตราจารย์ ดร.นภาพร หะวานนท์)
วันที่ 24 เดือน ตุลาคม พ.ศ. 2546

คณะกรรมการสอบปริญญานิพนธ์

..... ประธาน
(รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนัดพงษ์)

..... กรรมการ
(อาจารย์ ดร.ชูศักดิ์ ขัมภลขิต)

..... กรรมการ
(รองศาสตราจารย์ ดร.สมสรร วงษ์อยู่น้อย)

..... กรรมการที่แต่งตั้งเพิ่มเติม
(ผู้ช่วยศาสตราจารย์ ดร.นิคม ดังคะพิภพ)

..... กรรมการที่แต่งตั้งเพิ่มเติม
(อาจารย์ ดร.รังสรรค์ มณีเล็ก)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จได้ด้วยความอนุเคราะห์และความกรุณาจากคณะกรรมการควบคุมปริญญานิพนธ์ รองศาสตราจารย์ ดร.บุญเชิด ภิญโญนนท์พงษ์ อาจารย์ ดร.ชูศักดิ์ ชัมภลชีต และ รองศาสตราจารย์ ดร.สมสรร วงษ์อยู่น้อย ที่กรุณาให้คำปรึกษา ข้อคิดเห็น และแก้ไขข้อบกพร่องจนทำให้งานวิจัยเรื่องนี้สำเร็จได้เป็นอย่างดี ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอกราบขอบพระคุณกรรมการที่แต่งตั้งเพิ่มเติม ผู้ช่วยศาสตราจารย์ ดร.นิคม ดังคะพิภพ และอาจารย์ ดร.รังสรรค์ มณีเล็ก ที่ให้ความกรุณาเป็นกรรมการในการตรวจปริญญานิพนธ์และให้ข้อเสนอแนะที่เป็นประโยชน์เพิ่มเติมทำให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณผู้เชี่ยวชาญทุกท่านที่ได้ให้ความช่วยเหลือในการตรวจสอบคุณภาพของเครื่องมือ ขอขอบพระคุณผู้บริหารโรงเรียนมัธยมศึกษาที่เป็นกลุ่มตัวอย่างการวิจัยและขอขอบใจนักเรียนทุกคนที่ให้ความร่วมมือเป็นอย่างดีในการรวบรวมข้อมูล

สุดท้ายนี้ ผู้วิจัยขอขอบพระคุณทุกท่านที่ได้กล่าวนามถึง ณ ที่นี้ที่สนับสนุนให้ความช่วยเหลือในการทำปริญญานิพนธ์ฉบับนี้จนสำเร็จลุล่วงไปด้วยดี คุณค่าและประโยชน์ของปริญญานิพนธ์ฉบับนี้ ขอมอบเป็นเครื่องบูชา พระคุณของบิดา มารดา ครู-อาจารย์ และผู้มีพระคุณทุกท่านของผู้วิจัย

สุนิสา จัยม่วงศรี

สารบัญ

บทที่		หน้า
1	บทนำ.....	1
	ภูมิหลัง	1
	จุดมุ่งหมายของการวิจัย	9
	ความสำคัญของการวิจัย	9
	ขอบเขตของการวิจัย	10
	ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย	10
	ตัวแปรที่ศึกษา.....	10
	เครื่องมือที่ใช้ในการวิจัย	10
	ข้อดกลงเบื้องต้น	11
	ข้อจำกัดของการวิจัย	11
	นิยามศัพท์เฉพาะ	11
	สมมุติฐานในการวิจัย.....	15
2	เอกสารและงานวิจัยที่เกี่ยวข้อง.....	17
	การประยุกต์ใช้โมเดลทฤษฎีการตอบข้อสอบ.....	17
	แนวคิดการเทียบคะแนนระหว่างแบบทดสอบ.....	26
	แบบแผนการรวบรวมข้อมูล.....	27
	การออกแบบวิธีการเทียบคะแนน.....	31
	การเทียบคะแนนตามทฤษฎีการตอบข้อสอบที่ข้อสอบตรวจให้	
	คะแนนสองค่า	35
	การเลือกแบบแผนการรวบรวมข้อมูล.....	35
	การปรับค่าพารามิเตอร์ของแบบทดสอบต่างฉบับให้อยู่ในมาตรฐานวัด	
	เดียวกัน.....	36
	การเทียบคะแนนผู้สอบ	43
	วิธีการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า	44
	โมเดลเกรต เรสพอนส์.....	44
	โมเดลนอมินอล เรสพอนส์.....	48
	โมเดลพาเชียล เครดิต.....	49
	โมเดลเจนเอรอลไลซ์ พาเชียล เครดิต.....	49
	เกณฑ์การเทียบคะแนน	53

สารบัญ(ต่อ)

บทที่	หน้า
2(ต่อ) งานวิจัยเกี่ยวกับเทคนิคการเทียบคะแนน.....	61
งานวิจัยการเทียบคะแนนในต่างประเทศ.....	63
งานวิจัยการเทียบคะแนนในประเทศ.....	-79
3 วิธีดำเนินการวิจัย.....	95
การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง.....	95
เครื่องมือที่ใช้ในการวิจัย.....	98
การเก็บรวบรวมข้อมูล.....	117
การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล.....	117
การวิเคราะห์ข้อมูลในการทดลองเครื่องมือ.....	117
การวิเคราะห์ข้อมูลในการวิจัย.....	118
4 ผลการวิเคราะห์ข้อมูล.....	134
สัญลักษณ์ และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล.....	134
การเสนอผลการวิเคราะห์ข้อมูล.....	136
ผลการวิเคราะห์แบบทดสอบที่ใช้ในการเทียบคะแนน.....	136
ผลการวิเคราะห์คุณภาพรายข้อของแบบทดสอบเทียบคะแนน.....	137
ผลการตรวจสอบความเป็นเอกมิตีของแบบทดสอบเทียบคะแนน.....	140
ผลการวิเคราะห์ค่าสถิติพื้นฐานและทดสอบความแตกต่างของค่าเฉลี่ย และการแจกแจงคะแนนของแบบทดสอบเทียบคะแนน.....	142
ผลการวิเคราะห์คุณภาพของแบบทดสอบเทียบคะแนน.....	146
ผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบรวม.....	147
ผลการสร้างตารางเทียบคะแนน.....	148
ผลการสร้างตารางเทียบคะแนนตามวิธีในทฤษฎีการตอบข้อสอบ.....	148
ผลการสร้างตารางเทียบคะแนนตามทฤษฎีการทดสอบ แบบมาตรฐานเดิม.....	170
ผลการวิเคราะห์ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม.....	178
ผลการวิเคราะห์ความแตกต่างระหว่างคะแนนแปลง.....	181

สารบัญ(ต่อ)

บทที่	หน้า
4(ต่อ) ผลการวิเคราะห์กลุ่มสอบทานผล	201
ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนจากกลุ่มสอบทานผล	201
ผลการวิเคราะห์ความสัมพันธ์ระหว่างคะแนนแปลงกับคะแนนเกณฑ์	203
ผลการประเมินค่าดัชนีความแตกต่าง	204
ผลการตรวจสอบความแตกต่างระหว่างคะแนนแปลงกับคะแนนเกณฑ์	206
5 สรุป อภิปรายผล และข้อเสนอแนะ	207
สรุปผลการวิจัย	210
อภิปรายผลการวิจัย	216
ข้อเสนอแนะ	224
บรรณานุกรม	228
ภาคผนวก	242
ภาคผนวก ก	243
ภาคผนวก ข	248
ภาคผนวก ค	271
ภาคผนวก ง	280
ภาคผนวก จ	297
ประวัติย่อผู้วิจัย	332

บัญชีตาราง

ตาราง		หน้า
1	การเปรียบเทียบสถานการณ์ที่เหมาะสมกับวิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์ วิธีเชิงเส้นตรงและวิธี IRT	34
2	สรุปประเด็นสำคัญของการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า	74
3	รายชื่อโรงเรียนและจำนวนห้องเรียนที่เป็นกลุ่มตัวอย่าง	98
4	การวิเคราะห์เนื้อหา จุดประสงค์ ระดับพฤติกรรมการวัดและจำนวนข้อของแบบทดสอบเทียบคะแนน	100
5	ค่าดัชนีความสอดคล้องและจำนวนความถี่ของผลการประเมินที่ประเมินความสอดคล้องระหว่างเนื้อหา ระดับพฤติกรรมกับข้อสอบของผู้เชี่ยวชาญ	106
6	จำนวนข้อสอบที่ใช้จริงในแต่ละเนื้อหา จุดประสงค์ และระดับพฤติกรรม การวัดของแบบทดสอบเทียบคะแนนฉบับ X และ Y	113
7	ค่าสถิติของพารามิเตอร์อำนาจจำแนก และความยากในแบบทดสอบเทียบคะแนน เมื่อวิเคราะห์ตามทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์	137
8	ค่าสถิติของอำนาจจำแนกและความยากในแบบทดสอบเทียบคะแนนเมื่อวิเคราะห์ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม	138
9	ค่า eigen-ร้อยละของความแปรปรวน ความแปรปรวนสะสมค่า KMO และค่า Barlett จำแนกตามแบบทดสอบเทียบคะแนน	140
10	ค่าดัชนีทดสอบประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันของแบบทดสอบเทียบคะแนน	142
11	ค่าสถิติพื้นฐานของคะแนนจากแบบทดสอบเทียบคะแนน	143
12	ผลการเปรียบเทียบค่าเฉลี่ยของคะแนนจากแบบทดสอบเทียบคะแนน	145
13	ผลการวิเคราะห์ความแตกต่างของการแจกแจงคะแนนดิบจากแบบทดสอบเทียบคะแนน	145
14	ผลการวิเคราะห์ค่าเชื่อมั่นของแบบทดสอบเทียบคะแนน	146
15	ค่าสถิติพื้นฐานของคะแนนจากแบบทดสอบรวม	147
16	ค่าสัมประสิทธิ์สหสัมพันธ์รายคู่ระหว่างคะแนนของแบบทดสอบรวมกับคะแนนของแบบทดสอบเทียบคะแนนแต่ละฉบับ	148
17	ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนก และความยากของแบบทดสอบรวมจำแนกตามแบบทดสอบเทียบคะแนน	149

บัญชีตาราง(ต่อ)

ตาราง		หน้า
18	ค่าสัมประสิทธิ์การเทียบคะแนนจำแนกตามวิธีการเทียบคะแนน.....	150
19	คะแนนสมมูลของการเทียบคะแนนจริงจากแบบทดสอบฉบับ Y ไปยังฉบับ X จำแนกตามวิธีการแปลงพารามิเตอร์ความสามารถ.....	151
20	คะแนนสมมูลของการเทียบคะแนนจริงจากแบบทดสอบฉบับ Y ไปยังฉบับ X วิธีปรับค่าพารามิเตอร์พร้อมกัน.....	158
21	ค่าสถิติพื้นฐานของคะแนนดิบและคะแนนสังเกตที่สร้างโดยทฤษฎี IRT โมเดลเกรด เรสพอนส์.....	162
22	ผลการเปรียบเทียบค่าเฉลี่ยของคะแนนสังเกตที่สร้างโดยใช้ทฤษฎี IRT โมเดล เกรด เรสพอนส์ และคะแนนดิบในแบบทดสอบเทียบคะแนนฉบับเดียวกัน.....	164
23	ผลการวิเคราะห์ความแตกต่างของการแจกแจงคะแนนสังเกตที่สร้างโดยใช้ ทฤษฎี IRT โมเดลเกรด เรสพอนส์ และคะแนนดิบในแบบทดสอบเทียบ คะแนนฉบับเดียวกัน.....	164
24	คะแนนสมมูลของการเทียบคะแนนสังเกตจากแบบทดสอบฉบับ Y ไปยัง ฉบับ X จำแนกตามวิธีการเทียบคะแนน.....	166
25	คะแนนสมมูลของการเทียบคะแนนดิบจากแบบทดสอบฉบับ Y ไปยังฉบับ X ตามวิธีการเทียบคะแนนแบบอิคิวเปอร์เซ็นไทล์.....	170
26	ผลการวิเคราะห์ความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบเทียบ คะแนน.....	173
27	ค่าสถิติที่ใช้ในการคำนวณค่าความชัน (A) และจุดตัด (B) ของสมการ เส้นตรง.....	174
28	ค่าความชัน (A) และจุดตัด (B) ของสมการเทียบคะแนนเชิงเส้นตรง.....	174
29	คะแนนสมมูลของการเทียบคะแนนจากแบบทดสอบฉบับ Y ไปยังฉบับ X ด้วย วิธีการเทียบคะแนนเชิงเส้นตรง.....	175
30	ค่าสถิติพื้นฐานของคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีและคะแนนจาก แบบทดสอบฉบับ X และ Y.....	178
31	ผลการเปรียบเทียบค่าเฉลี่ยของคะแนนจากแบบทดสอบฉบับ X กับคะแนน แปลงจากการเทียบคะแนนแต่ละวิธี.....	180
32	ผลการวิเคราะห์ความแตกต่างของการแจกแจงคะแนนสะสมระหว่างคะแนน จากแบบทดสอบฉบับ X กับคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี.....	180

บัญชีตาราง(ต่อ)

ตาราง	หน้า
33 ค่าดัชนี $RMSD_W$ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์.....	185
34 ค่าดัชนี $RMSD_W$ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์.....	186
35 ค่าดัชนี $RMSD_U$ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์.....	187
36 ค่าดัชนี $RMSD_U$ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์.....	188
37 ค่าดัชนี MAD_W เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์.....	189
38 ค่าดัชนี MAD_W เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์.....	190
39 ค่าดัชนี MAD_U เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์.....	191
40 ค่าดัชนี MAD_U เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์.....	192
41 ค่าดัชนี MSD_W เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์.....	193
42 ค่าดัชนี MSD_W เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์.....	194
43 ค่าดัชนี MSD_U เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์.....	195
44 ค่าดัชนี MSD_U เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์.....	196
45 ค่าดัชนีความแตกต่าง $RMSD_W$ $RMSD_U$ MAD_W MAD_U MSD_W และ MSD_U ของคะแนนแปลงจากการเทียบคะแนนที่แตกต่างกัน 2 วิธี.....	197
46 ค่าสถิติพื้นฐานของคะแนนที่ได้จากกลุ่มสอบทานผล จำแนกตามแบบทดสอบเทียบคะแนน.....	201

บัญชีตาราง(ต่อ)

ตาราง	หน้า
47 คำสัมประสิทธิ์สหสัมพันธ์รายคู่ระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนเกณฑ์ของกลุ่มสอบทานผล.....	204
48 การประเมินค่าดัชนีความแตกต่าง (C) จำแนกตามวิธีการเทียบคะแนน.....	205
49 ผลการวิเคราะห์ความแตกต่างของคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนเกณฑ์.....	206
50 ผลการเปรียบเทียบคะแนนสมมูลฉบับ Y กับคะแนนแปลง (X').....	212
51 สรุปผลการเทียบคะแนนจากเกณฑ์ประเมินผล 3 เกณฑ์ จำแนกตามวิธีการเทียบคะแนน.....	215
52 คำอำนาจจำแนก และความยากของข้อสอบในแบบทดสอบเทียบคะแนนวิเคราะห์ตามทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์ และวิเคราะห์ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม	244
53 สารสนเทศของข้อสอบและแบบทดสอบเทียบคะแนนที่วิเคราะห์ตามทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์.....	246
54 ผลการแจกแจงคะแนนความสามารถของผู้สอบฉบับ X และ Y วิธีปรับค่าพารามิเตอร์แยกฉบับ	249
55 ผลการแจกแจงคะแนนความสามารถของผู้สอบฉบับ X และ Y วิธีปรับค่าพารามิเตอร์พร้อมกัน	250
56 ผลการวิเคราะห์ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และตำแหน่งเปอร์เซ็นต์ไทล์ในการเทียบคะแนนสังเกตวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก	251
57 ผลการวิเคราะห์ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และตำแหน่งเปอร์เซ็นต์ไทล์ในการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบ	256
58 ผลการวิเคราะห์ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และตำแหน่งเปอร์เซ็นต์ไทล์ของการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน.....	261
59 ผลการวิเคราะห์ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และตำแหน่งเปอร์เซ็นต์ไทล์ของการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์.....	266

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 แบบแผนการรวบรวมข้อมูลในการเทียบคะแนน	30
2 สารสนเทศแบบทดสอบเทียบคะแนน	139
3 จำนวนองค์ประกอบในแบบทดสอบฉบับ X	141
4 จำนวนองค์ประกอบในแบบทดสอบฉบับ Y	141
5 การแจกแจงคะแนนของแบบทดสอบเทียบคะแนน	144
6 ผลการเทียบคะแนนจริง 3 วิธี จากแบบทดสอบฉบับ Y ปรับสู่สเกลฉบับ X	157
7 การแจกแจงคะแนนดิบและคะแนนสังเกตที่สร้างตามทฤษฎีการตอบข้อสอบ ในแบบทดสอบฉบับ X	163
8 การแจกแจงคะแนนดิบและคะแนนสังเกตที่สร้างตามทฤษฎีการตอบข้อสอบ ในแบบทดสอบฉบับ Y	163
9 ผลการเทียบคะแนนสังเกต 3 วิธี จากแบบทดสอบฉบับ Y ปรับสู่สเกลฉบับ X	169
10 ผลการเทียบคะแนนอีควิเปอร์เซ็นไทล์จากแบบทดสอบฉบับ Y ปรับสู่สเกล ฉบับ X	172
11 ผลการเทียบคะแนนเชิงเส้นตรงจากแบบทดสอบฉบับ Y ปรับสู่สเกลฉบับ X	177
12 คะแนนความแตกต่างเมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงเป็น เกณฑ์	182
13 คะแนนความแตกต่างเมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์- เซ็นไทล์เป็นเกณฑ์	183
14 ค่าดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนวิธี ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากกับคะแนน แปลงจากการเทียบคะแนนพื้นฐาน	198
15 ค่าดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนวิธี โค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน	199
16 ค่าดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนวิธี ปรับค่าพารามิเตอร์พร้อมกันกับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน	200
17 การแจกแจงคะแนนของกลุ่มสอบทานผลในแบบทดสอบฉบับ X และ Y	203

บทที่ 1

บทนำ

ภูมิหลัง

สถานการณ์การทดสอบในปัจจุบันมีความจำเป็นที่ต้องใช้แบบทดสอบหลายฉบับในการวัดครั้งเดียวกัน หรือใช้แบบทดสอบหลายฉบับในการประเมินการเรียนรู้ในเวลาที่แตกต่างกัน ทำให้ต้องมีการเปรียบเทียบผลที่ได้จากแบบทดสอบต่างฉบับ ให้อยู่ในมาตรฐานวัดเดียวกัน เพื่อให้คะแนนของแบบทดสอบต่างฉบับสามารถเปรียบเทียบกันได้ การดำเนินการดังกล่าวต้องอาศัยกระบวนการที่เรียกว่าการเทียบคะแนนระหว่างแบบทดสอบ (Test Equating) ซึ่งเป็นกระบวนการที่เกี่ยวข้องกับกิจกรรม 2 ประการ คือ กระบวนการที่ทำให้แบบทดสอบสองฉบับใด ๆ มีความทัดเทียมกันหรือเท่ากันในเชิงโครงสร้าง และการใช้วิธีการทางสถิติเพื่อปรับคะแนนที่ได้จากแบบทดสอบแต่ละฉบับให้อยู่ในมาตรฐานเดียวกันและเทียบกันได้ (ชูศักดิ์ ชัมภลชิต. 2529 : 144)

การเทียบคะแนนมีบทบาทสำคัญต่อการวัดและการประเมินผลการศึกษาในปัจจุบัน มากยิ่งขึ้น ดังเห็นได้ชัดเจนจากการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา โดยจัดสอบวิชาหลักปีละสองครั้ง และวิชาเฉพาะปีละหนึ่งครั้ง โดยผู้สมัครสามารถสอบกี่ครั้ง ครั้งละกี่วิชาก็ได้ และใช้คะแนนครั้งที่ดีที่สุดสมัครเข้าศึกษาในสถาบันอุดมศึกษา คะแนนสอบสามารถเก็บไว้ใช้ได้สองปี ได้มีนักวิชาการทางด้านวัดผลการศึกษา ชี้ให้เห็นประเด็นของคะแนนสอบที่เก็บไว้ในแต่ละครั้ง ซึ่งได้จากแบบทดสอบที่มีโครงสร้างเนื้อหาและคุณภาพไม่เท่าเทียมกัน เช่น ความยากของข้อสอบ ย่อมทำให้เกิดการได้เปรียบเสียเปรียบในการสอบแต่ละครั้ง เมื่อนำคะแนนสอบไม่เท่าเทียมกันมาเปรียบเทียบกัน จึงก่อให้เกิดความไม่ยุติธรรม การเทียบคะแนนแบบทดสอบจึงเป็นกระบวนการที่จำเป็นที่ควรนำมาใช้ในสถานการณ์เช่นนี้ เพื่อปรับคะแนนสอบในแต่ละครั้งให้อยู่ในมาตรฐานวัดเดียวกัน เป็นคะแนนที่เปรียบเทียบกันได้ ทำให้การสอบคัดเลือกใช้เกณฑ์ได้ถูกต้องและมีคุณภาพ

การเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนสองค่า (ตอบถูกหรือผิด) โดยทั่วไปจำแนกตามทฤษฎีการทดสอบได้สองวิธีการใหญ่ ๆ คือ การเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม การเทียบคะแนนในทฤษฎีที่รู้จักแพร่หลาย ได้แก่ การเทียบคะแนนเชิงเส้นตรง ซึ่งพิจารณาการเทียบเท่ากันของคะแนนจากแบบทดสอบต่างฉบับที่ตรงกับคะแนนมาตรฐานเดียวกัน และการเทียบคะแนนอีควิเปอร์เซ็นไทล์ ซึ่งพิจารณาการเทียบเท่ากันของคะแนนที่ได้จากแบบทดสอบต่างฉบับ ที่ตรงกับตำแหน่งเปอร์เซ็นไทล์เดียวกัน และการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ ซึ่งมีหลักการว่าคะแนนจากแบบทดสอบต่างฉบับที่วัด

ในสิ่งเดียวกันถือว่าเทียบเท่ากันได้ ถ้าผู้สอบทำแบบทดสอบต่างฉบับแล้วได้คะแนนจริง (True score) หรือ ความสามารถแท้ (Ability) เท่ากัน (สงบ ลักษณะ 2525: 22) สำหรับวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ สามารถจำแนกตามเทคนิควิธีการปรับค่าพารามิเตอร์ได้สองแนวทาง คือการปรับค่าพารามิเตอร์พร้อมกัน (Simultaneous or Concurrent Calibration) และการปรับค่าพารามิเตอร์แยกกัน (Separate Calibration) ซึ่งข้อมูลการตอบจะถูกปรับค่าอย่างใดอย่างหนึ่ง ขึ้นกับแบบแผนการรวบรวมข้อมูล (Kolen and Brennan. 1995 : 167 – 168)

วิธีการเทียบคะแนนตามทฤษฎี IRT วิธีปรับค่าพารามิเตอร์พร้อมกัน มักใช้ในแบบแผนกลุ่มผู้ที่ใช้ข้อสอบร่วมหรือแบบแผนกลุ่มเดี่ยว (กลุ่มที่ทำแบบทดสอบเทียบคะแนนทั้งสองฉบับ) ซึ่งมีหลักการว่าค่าประมาณพารามิเตอร์ผู้สอบจากแบบทดสอบคนละฉบับจะถูกปรับให้อยู่ในมาตรวัดความสามารถเดียวกัน โดยการรวมชุดข้อมูลแบบทดสอบแล้ววิเคราะห์พร้อมกัน ทำให้มีการใช้ข้อสนเทศที่หาได้ทั้งหมด และขจัดความคลาดเคลื่อนในการเทียบคะแนนบางอย่างออกไปได้ การเทียบคะแนนวิธีนี้อาจให้ผลการเทียบคะแนนที่คงที่มากกว่าวิธีการอื่นสำหรับวิธีการปรับค่าพารามิเตอร์แยกกัน คือข้อมูลการตอบของแบบทดสอบสองฉบับหรือมากกว่าสองฉบับก็ตามถูกปรับค่าแยกกัน แล้วเชื่อมต่อมาตรวัดโดยใช้ข้อสนเทศของข้อสอบร่วม แปลงค่าพารามิเตอร์ความสามารถเชิงเส้น ด้วยสัมประสิทธิ์การเทียบคะแนน ได้แก่ สัมประสิทธิ์ความชันและจุดตัด โดยมีวิธีการประมาณค่าสัมประสิทธิ์การเทียบคะแนนแยกเป็นสองกลุ่มวิธี คือ 1) การประมาณค่าจากพารามิเตอร์บางค่าหรือใช้ทุกค่าไม่พร้อมกัน ซึ่งถ้าใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากมาประมาณค่าสัมประสิทธิ์ความชันและจุดตัด เรียกว่าวิธี mean และ sigma (Macro. 1977 : 140–141) แต่ถ้าใช้ค่าเฉลี่ยของพารามิเตอร์อำนาจจำแนกและพารามิเตอร์ความยากมาประมาณค่าสัมประสิทธิ์ความชันและจุดตัดเรียกว่าวิธี mean และ mean (Kolen and Brennan. 1995 : 168; citing Loyd and Hoover. 1980) และ 2) การประมาณค่าจากพารามิเตอร์ข้อสอบทุกค่าพร้อมกันด้วยการใช้โค้งลักษณะเป็นวิธีการที่แก้ไขจุดอ่อนของวิธีการในกลุ่มแรกที่ใช้ค่าพารามิเตอร์บางค่า หรือใช้ทุกค่าไม่พร้อมกัน การประมาณค่าสัมประสิทธิ์การเทียบคะแนนด้วยวิธีนี้มีทั้งการประมาณค่าโดยใช้โค้งลักษณะข้อสอบของเฮบารา (Kolen and Brennan. 1995 : 169 – 170; citing Haebara. 1980) และการประมาณค่าโดยใช้โค้งลักษณะแบบทดสอบของสตอกกิงและลอร์ด (Stocking and Lord. 1983 : 201 - 210) ขณะนี้ยังมีการศึกษาเพื่อค้นหาวิธีการประมาณค่าสัมประสิทธิ์การเทียบคะแนนที่แม่นยำ

จากการศึกษางานวิจัยเกี่ยวกับเทคนิคการเทียบคะแนนแบบทดสอบที่ให้คะแนนสองค่า ผลที่ได้แตกต่างกันไปตามองค์ประกอบที่เกี่ยวข้องกับการเทียบคะแนน (Bastari. 2000 : 20–22) สรุปได้ว่า แบบแผนแบบทดสอบร่วมเป็นแบบแผนการเชื่อมต่อที่มีความเป็นไปได้และเป็นที่นิยมในทางปฏิบัติ (Kolen and Brennan. 1995 : 21) โดยแบบทดสอบร่วมที่ยาวกว่าให้ผลดีกว่า เช่น กฎที่ว่าจำนวนข้อสอบร่วมควรเป็นร้อยละ 20 ของจำนวนข้อสอบทั้งหมดตามที่

(Angoff. 1984 : 107) ได้กำหนดไว้ยังใช้ได้ดี เพียงแต่ยังไม่มีข้อสรุปที่ชัดเจนเกี่ยวกับความยาวของแบบทดสอบทั้งฉบับ ซึ่งขึ้นอยู่กับบริบทที่ทำการศึกษา ในส่วนของขนาดตัวอย่างยังไม่มี ความชัดเจนว่าควรเป็นจำนวนเท่าใด แต่โดยทั่วไปตัวอย่างที่มีขนาดใหญ่กว่าจะทำให้ความแม่นยำในการประมาณค่าพารามิเตอร์ข้อสอบดีกว่า และการแจกแจงความสามารถของกลุ่มตัวอย่าง เมื่อดำเนินการทดสอบกับกลุ่มสุ่มหรือกลุ่มสมมูล ซึ่งเป็นกลุ่มที่มีความสามารถใกล้เคียงกัน ผลการเทียบคะแนนจะให้ความแม่นยำมากกว่า แต่เมื่อดำเนินการกับกลุ่มที่ไม่เท่าเทียมกัน ผลจะแม่นยำเมื่อใช้แบบแผนแบบทดสอบร่วม (Vale. 1986 : 342–343) สำหรับวิธีการเทียบคะแนน ซึ่งเป็นองค์ประกอบที่สำคัญของการเทียบคะแนน จากการศึกษาส่วนใหญ่ ปรากฏว่าผลที่ได้ขึ้นกับองค์ประกอบอื่น ๆ ที่เกี่ยวข้องกับการเทียบคะแนนโดยทั่วไปวิธีการตามทฤษฎี IRT ให้ผลที่ดีกว่าวิธีการตามทฤษฎี CTT

ผลสรุปที่ได้จากงานวิจัย เกี่ยวกับการเทียบคะแนนดังกล่าวข้างต้น มีประเด็นที่น่าสนใจ คือ เมื่อคะแนนที่นำมาเปรียบเทียบเปลี่ยนแปลงไปอยู่ในรูปคะแนนหลายค่าตามการประเมินผลการเรียนรู้ วิธีการเทียบคะแนนซึ่งเป็นองค์ประกอบสำคัญต่อการเทียบคะแนน วิธีใดให้ผลการเทียบคะแนนที่มีประสิทธิภาพมากกว่าวิธีอื่น เนื่องจากการประเมินที่เหมาะสมสำหรับประเมินการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ ตามพระราชบัญญัติการศึกษาแห่งชาติ พ.ศ.2542 คือ รูปแบบการประเมินการปฏิบัติ (Performance Assessment) และการประเมินสภาพจริง (บุญเชิด ภิญญอนันตพงษ์. 2544 : 85, 125) การประเมินการปฏิบัติหากประเมินโดยการทดสอบ มีหลายรูปแบบที่ใช้ ได้แก่ การเขียนตอบ (Constructed-Response) ความเรียง การสาธิต การนำเสนอปากเปล่า บทบาทสมมุติ นิทรรศการ พอดโฟลิโอและการสังเกตโดยตรง การประเมินการปฏิบัติส่วนใหญ่ผู้เรียนเป็นผู้กำหนดคำตอบ และสามารถให้คะแนนได้หลายระดับซึ่งตรงกันข้ามกับการให้ผู้สอบตอบถูกหรือผิด (Muraki, Hombo and Lee. 2000 : 325; citing Messick. 1996) ดังนั้น ข้อสอบเขียนตอบจึงเหมาะที่จะใช้เป็นเครื่องมือในการประเมินการปฏิบัติ (บุญเชิด ภิญญอนันตพงษ์. 2544 : 87 - 88)

แบบทดสอบที่ประกอบด้วยข้อสอบเขียนตอบส่วนใหญ่มักให้คะแนนหลายค่า และยังมีแบบทดสอบอีกประเภทหนึ่งคือแบบทดสอบเลือกตอบที่วัดความรู้บางส่วน (Partial knowledge) เป็นข้อสอบที่สามารถตรวจให้คะแนนได้หลายค่าเช่นกัน ข้อสอบประเภทนี้สามารถจำแนกระดับความรู้ของผู้เรียนได้สามระดับ (Simon, Budescu and Nevo. 1997 : 65) คือ 1) ผู้สอบมีความรู้ครบถ้วนคือ รู้คำตอบอย่างสมบูรณ์และมีความเชื่อมั่น 2) ผู้สอบมีความรู้บางส่วนคือ รู้เฉพาะส่วนของคำตอบหรือไม่แน่ใจเกี่ยวกับคำตอบ และ 3) ผู้สอบขาดความรู้ คือ ไม่มีความรู้เกี่ยวกับคำตอบ ดังนั้น แบบทดสอบเลือกตอบที่ข้อสอบตรวจให้คะแนนแบบหลายค่า จึงให้สารสนเทศมากกว่าแบบทดสอบเลือกตอบที่ข้อสอบตรวจให้คะแนนแบบสองค่า การศึกษาพัฒนาแบบทดสอบเลือกตอบให้เสนอสารสนเทศเกี่ยวกับความรู้บางส่วนของผู้ตอบนั้น แบ่งตามการศึกษาได้สามกลุ่ม วิธี (Simon, Budescu and Nevo. 1997 : 66 – 69; citing Lord and

Novick. 1968) คือ 1) วิธีการปรับโครงสร้างข้อสอบ (Changing the item structure) หรือวิธีการตอบใหม่ (Changing the response Method) 2) วิธีการให้คะแนนข้อสอบต่างกัน (Differential item scoring methods) และ 3) วิธีการให้คะแนนแบบทดสอบต่างกัน (Differential test scoring methods)

วิธีการทั้งสามวิธีข้างต้นใช้แก้ไขจุดอ่อนของข้อสอบเลือกตอบ แต่ละวิธีมีจุดเด่นจุดด้อยแตกต่างกันไป วิธีการปรับโครงสร้างข้อสอบใหม่เป็นวิธีการที่ทำให้เกิดแบบทดสอบเลือกตอบถูกหลายคำตอบและผิดหลายคำตอบ ซึ่งแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบและให้ผู้สอบตัดสินทุกคำตอบ (Multiple True False : MTF) เป็นรูปแบบหนึ่งของแบบทดสอบประเภทนี้ น่าจะเป็นวิธีที่แก้ปัญหาจุดอ่อนของข้อสอบเลือกตอบได้มากที่สุด เพราะการให้ตอบในลักษณะนี้ผู้เรียนจะต้องรู้ทั้งสิ่งที่ถูกต้องและไม่ถูกต้อง จึงจะถือว่ารู้จริงและนอกจากนี้การให้ตอบข้อสอบในลักษณะนี้เหมือนกับการวัดเนื้อหาและพฤติกรรมเดียวกันซ้ำหลาย ๆ ครั้ง ทำให้แน่ใจได้ว่าได้วัดความรู้ความสามารถของผู้เรียนแม่นยำขึ้น และสามารถระบุความแตกต่างของระดับความรู้บางส่วนของผู้เรียนได้ จึงเหมาะที่จะใช้เป็นข้อสอบตัดสินผลการเรียนและวินิจฉัยข้อบกพร่องในการเรียนได้เป็นอย่างดี ส่วนการตรวจให้คะแนนก็ทำได้สะดวก คล้ายกับการให้คะแนนข้อสอบถูกผิดธรรมดา หากเปรียบเทียบแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบและให้ตัดสินทุกคำตอบกับแบบทดสอบที่เป็นคำถามปลายเปิด พบว่าแบบทดสอบปลายเปิดให้สารสนเทศในด้านความรู้บางส่วนได้เป็นอย่างดี ถึงแม้ว่าจะกำหนดให้ผู้สอบตอบเพียงคำตอบเดียว แต่ผู้สอบก็ต้องคิดวิเคราะห์ที่จะหาคำตอบเอง และถ้าคำถามปลายเปิดนั้นมีคำตอบหลายระดับ ก็จะระบุขั้นความรู้ของผู้สอบได้อย่างชัดเจน ทำให้สามารถประเมินการเรียนรู้ตามสภาพจริงได้ดีกว่า และเป็นจุดเด่นของข้อสอบประเภทปลายเปิด อย่างไรก็ตามข้อสอบเขียนตอบเป็นข้อสอบที่ใช้เวลาในการตรวจให้คะแนนทำให้มีค่าใช้จ่ายสูง (Kennedy & Walstad. 1997 : 374 - 375) และมีการเปลี่ยนแปลงผู้ให้คะแนนในแต่ละปีการศึกษา ทำให้การตรวจให้คะแนนไม่มีความคงที่ (Tate. 1999 : 344 - 345) ดังนั้นแบบทดสอบเลือกตอบที่สามารถให้สารสนเทศสำหรับความรู้บางส่วน มีข้อดีในด้านการตรวจให้คะแนนที่เป็นปรนัย จึงน่าจะเป็นทางเลือกหนึ่งสำหรับการทดสอบที่มีผู้สอบจำนวนมาก ที่ต้องการวัดความรู้ของผู้สอบให้แม่นยำขึ้น (Simon, Budescu and Nevo. 1997 : 65 - 66)

ดังนั้น เมื่อพัฒนาการของการวัดและประเมินผลการเรียนรู้เปลี่ยนไป รูปแบบของเครื่องมือวัดส่วนใหญ่ที่ใช้ในการประเมินการเรียนรู้ได้พัฒนาอยู่ในรูปแบบของข้อสอบที่ตรวจให้คะแนนหลายค่า หรือข้อสอบผสมระหว่างข้อสอบที่ให้คะแนนสองค่าและหลายค่า ทำให้มีการพัฒนาเทคนิควิธีดำเนินการเชื่อมต่อและการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ สำหรับใช้กับแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า โดยพัฒนาจากวิธีการเทียบคะแนนสำหรับแบบทดสอบที่ข้อสอบตรวจให้คะแนนสองค่า จากเดิมกระบวนการเทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบก่อนดำเนินการเทียบคะแนนตามวิธีที่กล่าวมาแล้วข้างต้นจะต้อง

พิจารณาว่าจะเลือกโมเดลการวิเคราะห์ค่าพารามิเตอร์จากโมเดลโลจิสติก หนึ่ง สอง หรือสาม พารามิเตอร์ ซึ่งเป็นโมเดลการวิเคราะห์ค่าพารามิเตอร์สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ดังนั้นเมื่อนำข้อสอบที่ตรวจให้คะแนนหลายค่ามาเทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบจึงต้องวิเคราะห์ค่าพารามิเตอร์ในโมเดลของทฤษฎีการตอบข้อสอบที่ตรวจให้คะแนนหลายค่า (Polytomous Item Response Model) ซึ่งโมเดลนี้พัฒนามาจากโมเดล IRT ที่ใช้สำหรับข้อสอบที่ให้คะแนนสองค่า คือ โมเดลโลจิสติก 1 หรือ 2 พารามิเตอร์ โมเดล IRT สำหรับแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า มีหลายโมเดล เช่น โมเดลเกรด เรสพอนส์ (Graded Response Model : GRM) (Samejima. 1997 : 85-100) โมเดลนอมินอล เรสพอนส์ (Nominal Response Model : NRM) (Bock. 1972 : 29-51) โมเดลพาเชียล เครดิต (Partial Credit Model : PCM) (Masters and Wright. 1997 : 101-121) และโมเดลเจเนอรอลไลซ์ พาเชียล เครดิต (Generalized Partial Credit Model : GPCM) (Muraki. 1997 : 153-164) เป็นต้น

โมเดลเกรด เรสพอนส์ โมเดลพาเชียล เครดิต และโมเดลเจเนอรอลไลซ์ พาเชียล เครดิต เป็นโมเดลในกลุ่มที่มีลำดับชั้นการตอบเรียงลำดับ (Ordered Polytomous Category) แต่โมเดลนอมินอล เรสพอนส์ ใช้กับข้อสอบที่ไม่สามารถเรียงลำดับตามระดับความถูกต้องหรือระดับคุณลักษณะที่ข้อนั้นมุ่งวัดได้ โมเดลเกรด เรสพอนส์ และโมเดลเจเนอรอลไลซ์ พาเชียล เครดิต เป็นโมเดลที่รู้จักกันอย่างแพร่หลาย เนื่องจากไม่เข้มงวดเกี่ยวกับข้อตกลงเบื้องต้นในเรื่องความเป็นเอกมิติของแบบทดสอบ และพัฒนามาจากโมเดลโลจิสติก 2 พารามิเตอร์ ที่ใช้วิเคราะห์ข้อสอบที่ให้คะแนนสองค่า ทำให้มีค่าพารามิเตอร์อำนาจจำแนกในฟังก์ชัน แต่ข้อแตกต่างของสองโมเดลที่สำคัญคือ ขั้นตอนและวิธีการประมาณค่าพารามิเตอร์ในแต่ละโปรแกรม ดังนั้น การที่ผู้วิจัยจะเลือกใช้โมเดล IRT สำหรับวิเคราะห์ข้อสอบที่ให้คะแนนหลายค่าต้องคำนึงถึงองค์ประกอบหลายอย่าง ได้แก่ ประเภทของข้อมูล ความสอดคล้องระหว่างข้อมูลกับโมเดลที่จะทำให้เกิดความไม่แปรผันของพารามิเตอร์ข้อสอบและผู้สอบ ข้อตกลงเบื้องต้นของโมเดลแนวคิดในการวิเคราะห์ที่ง่ายและไม่ซับซ้อน (Dodd, Ayala and Koch. 1995 : 11)

ดังนั้นการพัฒนาวิธีการเทียบคะแนนสำหรับข้อสอบที่ตรวจให้คะแนนหลายค่าจึงพัฒนาตามโมเดล IRT ที่ข้อสอบตรวจให้คะแนนสองค่า ได้แก่ วิธีการเทียบคะแนนด้วยโค้งลักษณะแบบทดสอบ ในโมเดลนอมินอล เรสพอนส์ และโมเดลเกรด เรสพอนส์ ซึ่งเบเกอร์ (Baker. 1992 : 89-92, 1993 : 243-245) ขยายจากวิธีของสตอกกิง และลอร์ด (Stocking and Lord. 1983 : 203-205) ที่ใช้กับข้อสอบที่ตรวจให้คะแนนสองค่าซึ่งสัมพันธ์กับการเทียบคะแนนที่ได้ มีการแจกแจงแบบปกติ (Baker. 1997 : 168) นอกจากนี้ได้มีการพัฒนาวิธีการเทียบคะแนนโดยขยายวิธีไคสแควร์ต่ำสุด (Minimum χ^2) ของดิฟจิ (Divgi. 1985 : 413-414) ในโมเดลเกรด เรสพอนส์ โดยคิมและโคเฮน (Kim and Cohen. 1995 : 169-171) ซึ่งพบว่าวิธี

ไคสแควร์ต่ำสุดให้สัมประสิทธิ์ที่แตกต่างเพียงเล็กน้อยกับวิธีโค้งลักษณะแบบทดสอบ และมีการคำนวณน้อยมาก

งานวิจัยเกี่ยวกับเทคนิคการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่าหรืออยู่ในรูปแบบผสมระหว่างข้อสอบที่ตรวจให้คะแนนสองค่ากับหลายค่าก็ตามนั้น ปรากฏว่ายังมีไม่มากนัก ผลการวิจัยที่ได้ส่วนใหญ่ใช้วิธีการจำลองข้อมูล ในโมเดลเกรด เรสพอนส์ และโมเดลพาเซียล เครดิต เป็นการศึกษาผลที่ได้จากวิธีการเทียบคะแนนในองค์ประกอบต่าง ๆ ที่เกี่ยวข้องกับการเทียบคะแนน เช่น แบบแผนการเชื่อมต่อ ความยาวของแบบทดสอบ ขนาดของตัวอย่าง เป็นต้น โดยเปรียบเทียบระหว่างวิธีการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมกับวิธีการตามทฤษฎีการตอบข้อสอบ ในการประเมินการปฏิบัติเมื่อใช้ข้อสอบปลายเปิดของนักเรียนที่มีความสามารถใกล้เคียงกัน ปรากฏว่าวิธีการเทียบคะแนนอีคิวเปอร์-เซ็นไทล์ให้ผลที่สมมูลกับการเทียบคะแนนในโมเดล พาเซียล เครดิต ด้วยวิธีโค้งลักษณะแบบทดสอบ เมื่อข้อสอบมีความยากปานกลาง (Huynh and Ferrara. 1994 : 139) และให้ผลที่คล้ายคลึงกับวิธีการเทียบคะแนนในโมเดลพาเซียล เครดิต ด้วยวิธีการปรับค่าพารามิเตอร์พร้อมกัน (Hennings, Hirsch and Zhang. 1996 : 6) และได้มีการศึกษาเปรียบเทียบวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบของข้อสอบเขียนตอบในโมเดลเกรด เรสพอนส์ ได้แก่ วิธีโค้งลักษณะแบบทดสอบ (TRF) วิธีไคสแควร์ต่ำสุด (MCS) และวิธีการที่ใช้พารามิเตอร์บางค่า หรือทุกค่าไม่พร้อมกันซึ่งปรับสูตรจากข้อมูลที่ได้จากข้อสอบที่ตรวจให้คะแนนสองค่า ได้แก่ วิธีใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (วิธี mean และ sigma : MS) วิธีใช้ค่าเฉลี่ยของพารามิเตอร์ความยากและอำนาจจำแนก (วิธี mean และ mean : LH) และวิธีใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากที่กำหนดน้ำหนัก (WMS) (Cohen and Kim. 1998 : 118; citing Linn and others. 1981) ปรากฏว่าส่วนใหญ่ทั้ง 5 วิธีให้ผลที่คล้ายคลึงกันเกือบทุกองค์ประกอบที่ศึกษาจากการจำลองข้อมูล (Cohen and Kim. 1998 : 126-129) ในการศึกษาส่วนใหญ่ยังคงพิจารณาความสอดคล้องของผลที่ได้เทียบกับผลจากข้อมูลที่ข้อสอบตรวจให้คะแนนสองค่า

จากกรอบแนวคิดในการพัฒนาวิธีการเทียบคะแนนของแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า และจากการศึกษาผลงานวิจัยที่ผ่านมา เท่าที่ปรากฏมีการวิจัยเพื่อพัฒนาวิธีการเทียบคะแนนตามทฤษฎี IRT ด้วยวิธีการแปลงค่าความสามารถเชิงเส้น ซึ่งประมาณค่าสัมประสิทธิ์การเทียบคะแนน ทั้งที่ใช้ค่าพารามิเตอร์บางค่าหรือทุกค่าไม่พร้อมกัน และใช้ค่าพารามิเตอร์ทุกค่าพร้อมกันจากผู้เสนอสูตร และได้มีการเปรียบเทียบวิธีการเทียบคะแนนที่พัฒนาจากวิธีการเหล่านี้บ้างบางส่วน โดยใช้วิธีศึกษาจากข้อมูลที่จำลองขึ้น ซึ่งอาจไม่ใช่ตัวแทนที่ดีของคะแนนสอบของนักเรียน ซึ่งในสภาพการณ์จริงคะแนนมีความคลาดเคลื่อน และส่งผลต่อความคลาดเคลื่อนของการเทียบคะแนน ค่าความลำเอียงและค่าเฉลี่ยของความคลาดเคลื่อนของการเทียบคะแนน อาจจะมีขนาดใหญ่กว่าในงานวิจัยที่จำลองข้อมูล (Li, Lissitz and Yang. 1999)

: 20) และการศึกษาวิธีการเทียบคะแนนในแบบทดสอบเลือกตอบที่ข้อสอบตรวจให้คะแนนหลายค่าเท่าที่ค้นพบได้ยังไม่ปรากฏ ซึ่งผลที่ได้อาจจะแตกต่างหรือไม่แตกต่างจากผลการเทียบคะแนนที่ข้อสอบอยู่ในรูปให้ผู้สอบกำหนดคำตอบ เพราะอาจมีความคลาดเคลื่อนในการให้คะแนนเข้ามาเกี่ยวข้อง ซึ่งรูปแบบของข้อสอบอาจเป็นองค์ประกอบหนึ่งที่ส่งผลต่อประสิทธิภาพของการเทียบคะแนนหรือไม่ จึงเป็นประเด็นที่ควรพิจารณา

ดังนั้น ผู้วิจัยจึงมีความสนใจที่จะศึกษาผลที่ได้จากวิธีการเทียบคะแนนทั้งในทฤษฎีการทดสอบแบบมาตรฐานเดิมและทฤษฎีการตอบข้อสอบของข้อสอบเลือกตอบที่ตรวจให้คะแนนหลายค่า ในแบบแผนกลุ่มสุ่มที่ใช้แบบทดสอบร่วม ซึ่งเป็นการเทียบคะแนนในแนวนอนของนักเรียนระดับชั้นเดียวกัน เพราะว่าการเทียบคะแนนเป็นองค์ประกอบที่สำคัญของการเทียบคะแนนจึงสมควรที่จะศึกษาว่าวิธีการใดจะให้ผลดีต่อการเทียบคะแนนในสถานการณ์ดังกล่าว โดยการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมที่ใช้ในการศึกษาค้างนี้ ได้แก่ การเทียบคะแนนเชิงเส้นตรง ซึ่งเป็นวิธีที่ง่ายและสะดวก ทั้งยังไม่ต้องมีการปรับกระบวนการเทียบคะแนน เมื่อลักษณะการให้คะแนนของข้อสอบเปลี่ยนไปยังคงใช้กระบวนการเทียบคะแนนแบบเดิม และยังเป็นวิธีที่น่าผลไปใช้ได้สะดวก วิธีการเทียบคะแนนเชิงเส้นตรงนี้ เมื่อใช้เทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนสองค่า จะให้ผลเป็นที่น่าพอใจ เมื่ออยู่ในบริบทที่การแจกแจงคะแนนในแบบทดสอบเทียบคะแนนมีความคล้ายคลึงกันซึ่งเหมาะสมกับแบบแผนกลุ่มสุ่มและการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ เป็นวิธีการที่ใช้ในการเทียบคะแนนแบบทดสอบที่ข้อสอบเป็นคำถามปลายเปิดมาแล้ว ซึ่งให้ผลคล้ายคลึงกับวิธีการเทียบคะแนนในโมเดลพหุคูณ เครดิต ด้วยโครงสร้างลักษณะแบบทดสอบและด้วยการปรับค่าพารามิเตอร์พร้อมกันเมื่อแบบทดสอบมีความยากปานกลางและผู้สอบเป็นกลุ่มสุ่ม แต่ให้ผลแตกต่างจากการเทียบคะแนนในโมเดลเกรด เรสพอนส์ ด้วยการปรับค่าพารามิเตอร์พร้อมกัน ในบางระดับชั้น (Huynh and Ferrara. 1994 : 138–139; Hennings, Hirsch and Zhang. 1996 : 6-8) วิธีการเทียบคะแนนแบบอีคิวเปอร์เซ็นต์ไทล์ จึงเป็นวิธีการหนึ่งที่น่ามาศึกษาในการเทียบคะแนนครั้งนี้ด้วย

สำหรับวิธีการเทียบคะแนนตามทฤษฎี IRT เพื่อศึกษาให้ครอบคลุม ผู้วิจัยจึงศึกษาวิธีการเทียบคะแนนตามเทคนิคการปรับค่าพารามิเตอร์ทั้งสองแนวทาง คือ การเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกัน โดยไม่ใช้การกำหนดค่า เป็นวิธีที่เหมาะสมกับกลุ่มสุ่ม ซึ่งเป็นกลุ่มของผู้สอบในระดับชั้นเดียวกันซึ่งถือว่าผู้สอบมีความสามารถใกล้เคียงกันที่จะถูกสุ่มให้ได้รับการทดสอบจากแบบทดสอบคนละฉบับ และการเทียบคะแนนด้วยการปรับค่าพารามิเตอร์แยกกัน ซึ่งต้องแปลงค่าพารามิเตอร์ความสามารถเชิงเส้นด้วยค่าประมาณสัมประสิทธิ์การเทียบคะแนนที่ใช้โครงสร้างแบบทดสอบ ซึ่งพัฒนาโดยเบเกอร์ เป็นวิธีการเทียบคะแนนที่มีความแม่นยำ (Cohen and Kim. 1998 : 128) และที่ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากข้อสอบ (วิธี mean และ sigma) ของ มาร์โค ซึ่งได้รับการพัฒนาโดยโคเฮน และคิม

(Cohen and Kim.1998 : 129) มาใช้ ซึ่งเป็นวิธีที่ให้ผลดีใกล้เคียงกับวิธีการเทียบคะแนนโดยใช้โด่งลักษณะแบบทดสอบและยังเป็นวิธีที่ง่ายต่อการคำนวณ

ทั้งนี้การเทียบคะแนนตามทฤษฎี IRT จะเทียบคะแนนในโมเดลเกรต เรสพอนส์ ซึ่งเป็นการประยุกต์ใช้กับแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบและให้ผู้สอบตัดสินทุกคำตอบ ถึงแม้ว่าโมเดลเกรต เรสพอนส์ จะเป็นโมเดลที่ตั้งอยู่บนพื้นฐานที่ลำดับขั้นการตอบเรียงลำดับ ใช้กับข้อสอบที่มีลักษณะเป็นเครดิตบางส่วน ที่บ่งบอกถึงความสำเร็จในการแก้ปัญหาของผู้สอบ แต่ก็สามารถใช้กับแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบคะแนนในแต่ละข้อได้จากผลรวมของการตัดสินที่ถูกต้องในแต่ละคำตอบ โดยให้คะแนน 1 เมื่อระบุได้ถูกต้องตรงกับเฉลย และได้ 0 เมื่อระบุได้ไม่ถูกต้อง ซึ่งสอดคล้องกับขอบเขตที่สมบูรณ์ของกระบวนการให้เหตุผลของผู้สอบในแต่ละคำตอบว่าเป็นคำตอบที่ถูกหรือผิด ผู้ที่ได้คะแนนสูงกว่าแสดงว่ามีความสามารถมากกว่าผู้ที่ได้คะแนนต่ำกว่าและโมเดลเกรต เรสพอนส์ ได้เคยนำมาใช้เทียบคะแนนในแบบทดสอบที่ประกอบด้วยแบบทดสอบย่อย (Testlets) โดยใช้แบบทดสอบย่อยเป็นหน่วยในการวิเคราะห์ ซึ่งทำให้แบบทดสอบฉบับดังกล่าวมีคุณสมบัติเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบเกี่ยวกับความเป็นอิสระของข้อสอบและผลการเทียบคะแนนในโมเดลนี้ ให้ผลดีพอ ๆ กับโมเดลนอมินอล เรสพอนส์ และดีกว่าการเทียบคะแนนในโมเดล IRT ที่ข้อสอบตรวจให้คะแนนสองค่า ซึ่งหน่วยการวิเคราะห์เป็นข้อสอบในแต่ละแบบทดสอบฉบับย่อยทำให้ฝ่าฝืนข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระของข้อสอบ (Lee and Others. 2001 : 366-367) และโมเดลเกรต เรสพอนส์ เป็นโมเดลที่มีการใช้อย่างกว้างขวางในการเทียบคะแนนโดยมีโปรแกรมสนับสนุนในการเทียบคะแนนที่ใช้โด่งลักษณะแบบทดสอบ (Baker. 1993b : 20) จึงเป็นเหตุผลหนึ่งในการเลือกใช้โมเดลเกรต เรสพอนส์ ในการประมาณค่าของข้อมูลในการศึกษาครั้งนี้ด้วย

ด้วยเหตุผลสนับสนุนดังกล่าวข้างต้น ผู้วิจัยจึงศึกษาผลที่ได้จากการเทียบคะแนนทั้ง 5 วิธี คือการเทียบคะแนนเชิงเส้นตรง การเทียบคะแนนอิกวิเปอร์เซ็นไทล์ การเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกัน การเทียบคะแนนวิธีโด่งลักษณะแบบทดสอบและการเทียบคะแนนวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก ซึ่งการเทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบเป็นการเทียบคะแนนทั้งคะแนนจริงและคะแนนสังเกต โดยเทียบคะแนนที่ได้จากแบบทดสอบวัดผลสัมฤทธิ์ประเภทเลือกตอบที่ตรวจให้คะแนนหลายค่า ในโมเดลเกรต เรสพอนส์ ซึ่งศึกษาในสถานการณ์ทดสอบจริง ที่กลุ่มผู้สอบอยู่ในระดับชั้นเรียนเดียวกัน ด้วยแบบแผนแบบทดสอบรวม ว่าผลการเทียบคะแนนในแต่ละวิธีแตกต่างกันหรือไม่ วิธีการเทียบคะแนนวิธีใดให้ผลดีที่สุด โดยพิจารณาจากเกณฑ์ความคลาดเคลื่อนของการเทียบคะแนนที่มีค่าน้อย ซึ่งแสดงว่ามีคุณภาพดี เป็นเกณฑ์ภายนอกที่ได้จากความคลาดเคลื่อนรวมของกลุ่มสอบทานผล และพิจารณาจากเกณฑ์ภายใน คือ ความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงกับคะแนนจากแบบทดสอบฉบับเป้าหมายและความแตกต่าง

ของคะแนนแปลงวัดด้วยดัชนีคะแนนความแตกต่างที่ได้จากความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี

จุดมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาผลของการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่า ตามวิธีการเทียบคะแนนในทฤษฎีการตอบข้อสอบและทฤษฎีการทดสอบแบบมาตรฐานเดิม ผลการเทียบคะแนนพิจารณาจากความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ดัชนีคะแนนความแตกต่าง และดัชนีความแตกต่าง (C) จากกลุ่มสอบทาน โดยกำหนดวัตถุประสงค์เฉพาะ ดังนี้

1. เพื่อศึกษาความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนของแบบทดสอบฉบับเป้าหมาย (ฉบับ X)
2. เพื่อศึกษาดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนพื้นฐานกับคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ
3. เพื่อเปรียบเทียบระดับคุณภาพของดัชนีความแตกต่าง (C) จากกลุ่มสอบทานผลที่ได้จากการเทียบคะแนนแต่ละวิธี
4. เพื่อเปรียบเทียบความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนที่ผู้สอบทำได้ (ฉบับ X) ของกลุ่มสอบทานผล

ความสำคัญของการวิจัย

การศึกษานี้เป็นพื้นฐานของการวิจัยเกี่ยวกับการวัดและประเมินผลเพื่อประเมินผลของการเทียบคะแนนในแบบทดสอบชนิดเลือกตอบที่มีคำตอบถูกหลายคำตอบและให้ผู้สอบตัดสินทุกคำตอบ (MTF) โดยข้อสอบตรวจให้คะแนนหลายค่า ว่าวิธีการเทียบคะแนนวิธีใดให้ผลน่าพอใจมากที่สุด ตามแบบแผนการเทียบคะแนนแบบกลุ่มสมมูลที่ใช้แบบทดสอบร่วม ผลจากการศึกษานี้จะทำให้เกิดหลักฐานเชิงประจักษ์ เพื่อยืนยันแนวความคิดของวิธีการเทียบคะแนนสำหรับแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า และได้ข้อความรู้ที่เป็นประโยชน์ในการตรวจสอบทางทฤษฎีให้สมบูรณ์และชัดเจนยิ่งขึ้น และยังเป็นประโยชน์อย่างยิ่งต่อโครงการทดสอบขนาดใหญ่ ในการนำไปใช้เพื่อตัดสินใจเลือกวิธีการเทียบคะแนนที่ให้ผลดีและสะดวกต่อการดำเนินการเมื่อประเมินความรู้ความสามารถของนักเรียนในแต่ละช่วงชั้น ด้วยการที่ใช้แบบทดสอบมากกว่าหนึ่งฉบับในช่วงเวลาที่แตกต่างกัน ผลการศึกษานี้ยังเป็นการกระตุ้นให้ครูที่สอนวิชาเดียวกัน ระดับชั้นเดียวกัน แต่ใช้แบบทดสอบคนละฉบับได้ทำการเทียบคะแนนผู้สอบก่อนตัดสินผล เพื่อให้ความยุติธรรมต่อผู้เรียนที่สอบด้วยแบบทดสอบต่างฉบับกัน และแบบทดสอบที่นำมาใช้ในการศึกษานี้เป็นเครื่องมือประเภทหนึ่งที่ใช้ในการประเมินการเรียนรู้

ที่มีความหลากหลายมากยิ่งขึ้นในปัจจุบัน ก่อให้เกิดประโยชน์ต่อการพัฒนาแบบทดสอบให้มีประสิทธิภาพยิ่งขึ้น ส่งผลให้ระบบการวัดและประเมินผลมีคุณภาพดียิ่งขึ้นต่อไป

ขอบเขตของการวิจัย

1. ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัย

เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2545 สังกัดกรมสามัญศึกษาในเขตกรุงเทพมหานคร จำนวน 52,303 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2545 สังกัดกรมสามัญศึกษาในเขตกรุงเทพมหานคร จำนวน 2,312 คน จำแนกเป็นนักเรียนที่ทดสอบด้วยแบบทดสอบฉบับ X จำนวน 1,155 คน และฉบับ Y จำนวน 1,157 คน ได้จากการสุ่มสองขั้นตอน

2. ตัวแปรที่ศึกษา

2.1 ตัวแปรอิสระ ได้แก่ วิธีการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม สองวิธีและตามทฤษฎีการตอบข้อสอบในโมเดลเกรด เรสพอนส์ สามวิธี ดังนี้

2.1.1 วิธีการเทียบคะแนนเชิงเส้นตรง

2.1.2 วิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์

2.1.3 วิธีการเทียบคะแนนด้วยการปรับค่าพารามิเตอร์พร้อมกัน

2.1.4 วิธีการเทียบคะแนนด้วยโค้งลักษณะแบบทดสอบ

2.1.5 วิธีการเทียบคะแนนด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากข้อสอบ

2.2 ตัวแปรตาม ได้แก่ ผลของการเทียบคะแนนซึ่งพิจารณาจาก

2.2.1 ความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงกับคะแนนจากแบบทดสอบฉบับเป้าหมาย (ฉบับ X)

2.2.2 ดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนพื้นฐานกับคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ

2.2.3 ดัชนีความแตกต่าง (C) ในแต่ละวิธีการเทียบคะแนนจากกลุ่มสอบทานผล

3. เครื่องมือที่ใช้ในการวิจัย

เป็นแบบทดสอบเลือกตอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผลที่มีคำตอบถูกหลายคำตอบ และให้ผู้สอบตัดสินทุกคำตอบ (MTF) จำนวน 5 ตัวเลือก แต่ละข้อตรวจให้คะแนน 6 ค่า คือ 0 1 2 3 4 และ 5 คะแนน มีสองฉบับ คือ ฉบับ X และฉบับ Y แต่ละฉบับมีจำนวนข้อเท่ากัน และวัดในโครงสร้างเนื้อหาเดียวกัน

ข้อตกลงเบื้องต้น

การศึกษาคุณภาพของคะแนนแปลงที่ได้จากการเทียบคะแนนแต่ละวิธีด้วยดัชนีคะแนนความแตกต่าง ในการวิจัยนี้ใช้คะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมเป็นเกณฑ์ ซึ่งถือว่าเป็นคะแนนเกณฑ์ที่เหมาะสม เพราะการเทียบคะแนนตามทฤษฎีนี้สามารถใช้กับคะแนนแบบหลายค่าได้โดยตรง ไม่ต้องปรับกระบวนการใด ๆ และผลที่ได้มีความแม่นยำต่อสถานการณ์ที่ผู้สอบมีความสามารถใกล้เคียงกันและแบบทดสอบเทียบคะแนนทั้งสองฉบับวัดในเนื้อหาเดียวกัน และมีค่าสถิติใกล้เคียงกัน

ข้อจำกัดของการวิจัย

เนื่องจากการวิจัยครั้งนี้ไม่สามารถประเมินผลการเทียบคะแนนตามทฤษฎีการตอบข้อสอบด้วยรากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Square Error : RMSE) ระหว่างโค้งลักษณะข้อสอบ (ICC) ที่ได้จากพารามิเตอร์ข้อสอบ และค่าประมาณพารามิเตอร์ที่ผ่านการแปลงได้ เพราะพารามิเตอร์ข้อสอบซึ่งเป็นค่าที่แท้จริงในประชากรข้อสอบไม่สามารถหาได้ในสถานการณ์ทดสอบจริง ดังนั้น จึงใช้เกณฑ์ความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงและคะแนนจากแบบทดสอบฉบับเป้าหมาย ในการตรวจสอบประสิทธิภาพของผลการเทียบคะแนนแทน จึงอาจเป็นข้อจำกัดอย่างหนึ่งของการศึกษาการเทียบคะแนนในสถานการณ์ทดสอบจริง

นิยามศัพท์เฉพาะ

1. ผลการเทียบคะแนน หมายถึง ค่าที่พิจารณาจากความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ดัชนีคะแนนความแตกต่าง และดัชนีความแตกต่างจากการวิเคราะห์กลุ่มสอบทานผล

2. ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม หมายถึง ระดับความแตกต่างของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนของแบบทดสอบฉบับเป้าหมาย ทดสอบด้วยสถิติโครโมโกรอฟ สมิโนฟ

3. ดัชนีคะแนนความแตกต่าง (Difference scores Indices) หมายถึง ค่าเฉลี่ยความแตกต่างของคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบกับคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมซึ่งถือเป็นคะแนนพื้นฐาน (Baseline) แบ่งเป็นดัชนีคะแนนความแตกต่างถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก คือ

3.1 ดัชนีคะแนนความแตกต่างถ่วงน้ำหนัก หมายถึง ค่าเฉลี่ยของคะแนนความแตกต่างที่ถ่วงน้ำหนักด้วยความถี่ของแต่ละระดับคะแนน (score point) ในแบบทดสอบฉบับ Y คำนวณได้สามวิธี คือ

3.1.1 รากที่สองของค่าเฉลี่ยกำลังสองถ่วงน้ำหนัก (Weighted Root Mean Square Difference : $RMSD_w$) หมายถึง ค่าเฉลี่ยของความแตกต่างที่คำนวณจากรากที่สองของผลรวมความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกำลังสอง และคูณด้วยความถี่ของแต่ละระดับคะแนนแล้วจึงถ่วงน้ำหนักด้วยผลรวมของความถี่ในแต่ละระดับคะแนน

3.1.2 ค่าเฉลี่ยความแตกต่างสัมบูรณ์ถ่วงน้ำหนัก (Weighted Mean Absolute Difference : MAD_w) หมายถึง ค่าเฉลี่ยของความแตกต่างที่คำนวณจากผลรวมของค่าสัมบูรณ์ความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีและคูณด้วยความถี่ของแต่ละระดับคะแนน แล้วจึงถ่วงน้ำหนักด้วยผลรวมของความถี่ในแต่ละระดับคะแนน

3.1.3 ค่าเฉลี่ยความแตกต่างที่คำนึงถึงเครื่องหมายและถ่วงน้ำหนัก (Weighted Mean Signed Difference : MSD_w) หมายถึง ค่าเฉลี่ยของความแตกต่างที่คำนวณจากผลรวมความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีและคูณด้วยความถี่ของแต่ละระดับคะแนน แล้วจึงถ่วงน้ำหนักด้วยผลรวมของความถี่ในแต่ละระดับคะแนน

3.2 ดัชนีคะแนนความแตกต่างไม่ถ่วงน้ำหนัก หมายถึง ค่าเฉลี่ยคะแนนความแตกต่างที่เฉลี่ยด้วยจำนวนระดับคะแนนที่มีค่าตั้งแต่คะแนนต่ำสุดจนถึงคะแนนเต็มในแบบทดสอบฉบับ Y จำนวนได้สามวิธี คือ

3.2.1 รากที่สองของค่าเฉลี่ยกำลังสองไม่ถ่วงน้ำหนัก (Unweighted Root Mean Square Difference : $RMSD_u$) หมายถึง ค่าเฉลี่ยของความแตกต่างที่คำนวณจากรากที่สองของผลรวมความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกำลังสอง แล้วเฉลี่ยด้วยจำนวนระดับคะแนนตามที่กำหนด

3.2.2 ค่าเฉลี่ยความแตกต่างสัมบูรณ์ไม่ถ่วงน้ำหนัก (Unweighted Mean Absolute Difference : MAD_u) หมายถึง ค่าเฉลี่ยของความแตกต่างที่คำนวณจากค่าสัมบูรณ์ของผลรวมความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี แล้วเฉลี่ยด้วยจำนวนระดับคะแนนตามที่กำหนด

3.2.3 ค่าเฉลี่ยความแตกต่างที่คำนึงถึงเครื่องหมายและไม่ถ่วงน้ำหนัก (Unweighted Mean Signed Difference : MSD_u) หมายถึง ค่าเฉลี่ยของความแตกต่างที่คำนวณจากผลรวมของความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี แล้วเฉลี่ยด้วยจำนวนระดับคะแนนตามที่กำหนด

4. ดัชนีความแตกต่าง (Discrepancy Index : C) หมายถึง ค่าความคลาดเคลื่อนรวม (Total error) ของการเทียบคะแนนจากกลุ่มสอบทานผล คำนวณจากค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างคะแนนสมมูล จากการนำคะแนนของแบบทดสอบฉบับ Y ไปแปลงกับคะแนนเกณฑ์ และถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ จึงเป็นค่ามาตรฐานที่สามารถเปรียบเทียบกันได้โดยตรง เพื่อประเมินคุณภาพในการเทียบคะแนน

5. ความเพียงพอในการเทียบคะแนน (Adequacy of Equating) หมายถึง ระดับคุณภาพของวิธีการเทียบคะแนน ซึ่งประเมินจากค่าดัชนีความแตกต่าง โดยใช้เกณฑ์ของปีเตอร์เซนและคณะ (Petersen, Macro and Stewart. 1982: 93-97) จำแนกออกเป็น 5 ระดับของการยอมรับ

6. การเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม หมายถึง กระบวนการที่ใช้วิธีการทางสถิติเพื่อปรับคะแนนจากแบบทดสอบต่างชุดที่มีโครงสร้างเนื้อหาเดียวกันให้สามารถเปรียบเทียบกันได้ แบ่งเป็นสองวิธีคือ

6.1 วิธีการเทียบคะแนนเชิงเส้นตรง (Linear equating) หมายถึง กระบวนการที่ใช้วิธีการทางสถิติเพื่อปรับคะแนนจากแบบทดสอบต่างชุด ที่มีโครงสร้างเนื้อหาเดียวกัน ให้สามารถเปรียบเทียบกันได้ โดยยึดหลักการแปลงให้เป็นคะแนนมาตรฐาน โดยใช้รูปแบบแบบทดสอบร่วมเป็นตัวปรับความแตกต่างของกลุ่มตัวอย่าง

6.2 วิธีการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์ (equipercentile equating) หมายถึง กระบวนการที่ใช้วิธีการทางสถิติ เพื่อปรับคะแนนจากแบบทดสอบต่างชุด ที่มีโครงสร้างเนื้อหาเดียวกัน ให้สามารถเปรียบเทียบกันได้ ณ ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน โดยใช้รูปแบบแบบทดสอบร่วม เป็นตัวปรับความแตกต่างของกลุ่มตัวอย่าง

7. การเทียบคะแนนตามแนวทฤษฎีการตอบข้อสอบที่ข้อสอบตรวจให้คะแนนหลายค่า หมายถึง กระบวนการวิเคราะห์ที่นำค่าพารามิเตอร์ความสามารถ และพารามิเตอร์ข้อสอบจากการวิเคราะห์ตามทฤษฎีการตอบข้อสอบโมเดลเกรต เรสพอนส์ มาดำเนินการโดยยึดหลักว่า ถ้าพารามิเตอร์ความสามารถอยู่ในสเกลเดียวกันแล้ว สามารถนำคะแนนจากแบบทดสอบสองฉบับมาเทียบกันได้ โดยใช้รูปแบบแบบทดสอบร่วมเป็นตัวปรับความแตกต่างของกลุ่ม ตัวอย่างแบ่งเป็นสามวิธีคือ

7.1 วิธีการเทียบคะแนนด้วยโค้งลักษณะแบบทดสอบ (Test Characteristic Curve Method) หมายถึง กระบวนการเทียบคะแนนที่แปลงค่าพารามิเตอร์ความสามารถเชิงเส้นจากแบบทดสอบต่างฉบับ ให้อยู่ในมาตรวัดเดียวกัน โดยประมาณค่าสัมประสิทธิ์การเทียบคะแนนด้วยโค้งลักษณะแบบทดสอบร่วม

7.2 วิธีการเทียบคะแนนด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากข้อสอบ (วิธี mean และ sigma) หมายถึง กระบวนการเทียบคะแนนที่แปลงค่าพารามิเตอร์ความสามารถเชิงเส้นจากแบบทดสอบต่างฉบับให้อยู่ในมาตรวัดเดียวกัน โดยประมาณค่าสัมประสิทธิ์การเทียบคะแนนด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของข้อสอบรวม

7.3 วิธีการเทียบคะแนนด้วยการปรับค่าพารามิเตอร์พร้อมกัน หมายถึง กระบวนการเทียบคะแนนที่วิเคราะห์ค่าพารามิเตอร์ข้อสอบและความสามารถจากข้อสอบทั้งสองฉบับพร้อมกันทำให้ค่าพารามิเตอร์อยู่ในมาตรวัดเดียวกัน

8. การเทียบคะแนนจริง หมายถึง การเทียบคะแนนจริงจากแบบทดสอบเทียบคะแนนสองฉบับ โดยยึดหลักว่า ถ้าพารามิเตอร์ความสามารถอยู่ในสเกลเดียวกันแล้ว คะแนนจริงที่คำนวณได้จากพารามิเตอร์ของแบบทดสอบทั้งสองฉบับ ณ ระดับความสามารถเดียวกัน มีค่าเทียบเท่ากัน

9. การเทียบคะแนนสังเกต หมายถึง การเทียบคะแนนที่ได้จากการสร้างตามทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์ โดยอาศัยพารามิเตอร์ความสามารถที่อยู่ในมาตรวัดเดียวกันของแบบทดสอบเทียบคะแนนทั้งสองฉบับ และนำการแจกแจงคะแนนที่สร้างในแบบทดสอบเทียบคะแนนแต่ละฉบับมาเทียบคะแนน โดยใช้วิธีการเทียบคะแนนอิกวิเปอร์เซ็นไทล์

10. คะแนนสมมูล (equivalent scores) หมายถึง คะแนนระหว่างแบบทดสอบต่างฉบับที่เทียบคะแนนกันได้ว่าเป็นคะแนนที่เท่าเทียมกัน

11. คะแนนเกณฑ์ หมายถึง คะแนนที่ได้จากแบบทดสอบฉบับเป้าหมาย (ฉบับ X) ของกลุ่มสอบทานผล

12. แบบทดสอบเทียบคะแนน หมายถึง แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผลระดับชั้นมัธยมศึกษาปีที่สองที่ผู้วิจัยสร้างขึ้น วิเคราะห์และหาคุณภาพตามทฤษฎีการทดสอบแบบมาตรฐานเดิม และทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์ และแบ่งแบบทดสอบออกเป็น 2 ฉบับ ฉบับละ 30 ข้อ คือ ฉบับ X และฉบับ Y โดยพิจารณาจากค่าสารสนเทศข้อสอบ (Item Information) ที่ใกล้เคียงกัน ในแต่ละข้อที่วัดเนื้อหาเดียวกัน แบบทดสอบเทียบคะแนนทั้งสองฉบับมีข้อสอบรวมที่กระจายตามเนื้อหาฉบับละ 10 ข้อ

13. แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล หมายถึง แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ที่วัดความสามารถด้านความคิดรวบยอดทางคณิตศาสตร์ที่เกี่ยวข้องกับหลักการ วิธีการ และโน้มน้าในการคิดแบบนามธรรม ในเรื่องของระบบจำนวนเต็ม เศษส่วนและทศนิยม แบบทดสอบมีลักษณะไม่ได้มุ่งวัดทางด้านทักษะการคำนวณ แต่ต้องการทราบวิธีการ ขั้นตอนการคำนวณ ความเข้าใจในการแปลความ ขยายความ และคิดหาเหตุผล แบบทดสอบเป็นชนิดเลือกตอบ 5 ตัวเลือก ที่มีคำตอบถูกหลายคำตอบ และให้ผู้สอบตัดสินทุกคำตอบ โดยข้อสอบแต่ละข้อตรวจให้คะแนน 6 ค่า คือ 0 1 2 3 4 และ 5 สร้างโดยยึดระดับพฤติกรรมการเรียนรู้ที่พึงประสงค์ด้านสติปัญญาของวิลสัน 4 ขั้น คือ ความรู้ ความจำเกี่ยวกับการคิดคำนวณ (Computation) ความเข้าใจ (Comprehension) การนำไปใช้ (Application) และการวิเคราะห์ (Analysis)

14. แบบทดสอบรวมภายใน (Internal or Inclusive anchor Test) หมายถึง กลุ่มข้อสอบที่รวมอยู่ในแบบทดสอบเทียบคะแนนทั้งสองฉบับ ฉบับละ 10 ข้อ เป็นแบบทดสอบชนิดเลือกตอบ 5 ตัวเลือก รูปแบบเดียวกับแบบทดสอบเทียบคะแนน และมีโครงสร้างเนื้อหาเป็นสัดส่วนกับแบบทดสอบเทียบคะแนนทั้งสองฉบับ

15. กลุ่มตัวอย่างเทียบคะแนน (Equating Samples) หมายถึง กลุ่มตัวอย่างที่ได้รับการทดสอบจากแบบทดสอบเทียบคะแนนฉบับใดฉบับหนึ่งและแบบทดสอบรวม เพื่อนำผลการสอบไปใช้ในการแปลงคะแนนสำหรับสร้างคะแนนสมมูลในการเทียบคะแนน

16. กลุ่มสอบทานผล (Cross Validation Group) หมายถึง กลุ่มตัวอย่างที่สุ่มมาจากประชากรเดียวกันกับกลุ่มตัวอย่างเทียบคะแนน โดยไม่มีหน่วยตัวอย่างที่ซ้ำซ้อนกัน ได้รับการสอบจากแบบทดสอบเทียบคะแนนทั้ง 2 ฉบับ และแบบทดสอบรวม เพื่อนำผลการสอบไปใช้ในการตรวจสอบหาค่าดัชนีความแตกต่าง

17. ตารางเทียบคะแนน หมายถึง ตารางแสดงตัวเลขคะแนนสังเกตหรือคะแนนจริงที่สมมูลกัน (equivalent scores) ระหว่างแบบทดสอบต่างชุด

สมมุติฐานในการวิจัย

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับวิธีการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนแบบหลายค่า สรุปได้ว่า

การเทียบคะแนนตามทฤษฎีการตอบข้อสอบวิธี mean และ sigma และวิธีโค้งลักษณะแบบทดสอบให้ผลที่คล้ายคลึงกันในข้อสอบเขียนตอบ (Cohen and Kim, 1998 : 126-129) ส่วนวิธีการปรับค่าพารามิเตอร์พร้อมกันให้ผลดีกว่าวิธีโค้งลักษณะแบบทดสอบในข้อสอบรูปแบบผสมระหว่างข้อสอบเลือกตอบกับเขียนตอบ (Bastari, 2000 : 52 - 54)

สำหรับผลการวิจัยระหว่างการเทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบและตามทฤษฎีการทดสอบแบบมาตรฐานเดิม พบว่า การเทียบคะแนนอีควิเปอร์เซ็นไทล์ให้ผลที่แตกต่างกับวิธีการปรับค่าพารามิเตอร์พร้อมกัน (Hennings, Hirsch and Zhang. 1996 : 6-8) ในข้อสอบปลายเปิด ขณะเดียวกันให้ผลที่คล้ายคลึงกับการเทียบคะแนนเชิงเส้นตรง และการเทียบคะแนนด้วยโค้งลักษณะแบบทดสอบในแบบทดสอบที่ประกอบด้วยแบบทดสอบฉบับย่อย เมื่อใช้แบบทดสอบฉบับย่อยเป็นหน่วยในการวิเคราะห์ (Lee and Others. 2001 : 368)

ประกอบกับการวิจัยครั้งนี้เป็นการเทียบคะแนนตามแนวนอน ในแบบแผนกลุ่มสุ่ม ซึ่งผู้สอบมีความสามารถใกล้เคียงกันและมีข้อสอบร่วมทำหน้าที่เชื่อมและปรับความแตกต่างระหว่างแบบทดสอบต่างชุด และแบบทดสอบเทียบคะแนนทั้งสองฉบับวัดในโครงสร้างเนื้อหาเดียวกัน จากเหตุผลข้างต้นผู้วิจัยคาดว่าผลที่ได้จากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบและตามทฤษฎีการทดสอบแบบมาตรฐานเดิม น่าจะมีคุณภาพอยู่ในระดับที่ยอมรับได้ทั้งสิ้น แต่เป็นระดับคุณภาพที่ต่างกัน ผู้วิจัยจึงกำหนดสมมุติฐานการวิจัยเฉพาะที่สอดคล้องกับจุดมุ่งหมายที่ใช้สถิติพิจารณา และตามเกณฑ์การประเมินที่กำหนดไว้ดังนี้

1. การแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนของแบบทดสอบฉบับเป้าหมายไม่ต่างกัน
2. ดัชนีความแตกต่าง (C) ที่ได้จากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบซึ่งวิเคราะห์ด้วยกลุ่มสอบทานผล มีคุณภาพอยู่ในระดับที่สูงกว่าที่ได้จากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม
3. ความคลาดเคลื่อนของการเทียบคะแนนแต่ละวิธีซึ่งวิเคราะห์ด้วยกลุ่มสอบทานผลให้ค่าที่แตกต่าง

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาเกี่ยวกับผลการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่า ผู้วิจัยได้ศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้า ดังต่อไปนี้

1. การประยุกต์ใช้โมเดลทฤษฎีการตอบข้อสอบ
2. แนวคิดการเทียบคะแนนระหว่างแบบทดสอบ
3. การเทียบคะแนนตามทฤษฎีการตอบข้อสอบที่ข้อสอบตรวจให้คะแนนสองค่า
4. วิธีการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า
5. เกณฑ์การเทียบคะแนน
6. งานวิจัยเกี่ยวกับเทคนิคการเทียบคะแนน

1. การประยุกต์ใช้โมเดลทฤษฎีการตอบข้อสอบ (Item Response Theory Models : IRT)

โมเดลทฤษฎีการตอบข้อสอบ (IRT) เริ่มต้นจากการพัฒนามาเพื่อวิเคราะห์ข้อสอบที่ตรวจให้คะแนนแบบสองค่า (Bastari. 2000 : 6; citing Lord and Novick.1968) เพื่อให้เกิดความสะดวกทางคณิตศาสตร์ จึงได้นำโมเดลโลจิสติกมาใช้ โมเดลทั้งสามคือ โมเดลโลจิสติก หนึ่ง สอง และสามพารามิเตอร์ (1-PL, 2-PL, 3-PL) มีแนวคิด ข้อตกลงเบื้องต้น และรูปแบบของโมเดล ซึ่งมีจุดดีและจุดด้อยพอ ๆ กัน โมเดล 3 พารามิเตอร์ มิติเดียว (Unidimension) มีสูตร ดังนี้

$$P_j(\theta) = C_j + (1 - C_j) \frac{e^{Da_j(\theta - b_j)}}{1 + e^{Da_j(\theta - b_j)}} \quad (1)$$

$P_j(\theta)$ คือ ความน่าจะเป็นของผู้สอบที่มีระดับความสามารถ θ จะตอบข้อสอบ j ถูก

θ คือ ระดับความสามารถของผู้สอบ

D คือ องค์ประกอบมาตรฐานวัด

a_j คือ พารามิเตอร์อำนาจจำแนก (ความชัน) สำหรับข้อสอบ j

b_j คือ พารามิเตอร์ความยาก (จุดเริ่มเปลี่ยน) สำหรับข้อสอบ j

c_j คือ พารามิเตอร์การเดาสำหรับข้อสอบ j

เมื่อกำหนดพารามิเตอร์ c ในโมเดลให้เป็น 0 โมเดลลดเป็นสอง พารามิเตอร์ และเมื่อกำหนดพารามิเตอร์ a ในโมเดลให้เป็น 1 โมเดลลดเป็นหนึ่งพารามิเตอร์ หรือโมเดลราสช์ (Rasch) สำหรับโมเดล IRT ที่ข้อสอบตรวจให้คะแนนสองค่า ได้มีการศึกษาเพื่อประยุกต์ในหลายศาสตร์ (Wright and Stone.1979) และมีบทบาทสำคัญต่อการพัฒนาทางด้านวัดผล จากการประยุกต์ใช้โมเดล IRT ปรากฏว่าโมเดล IRT โลจิสติกสามพารามิเตอร์ มีความเหมาะสมต่อการใช้มากกว่าโมเดลโลจิสติกสองพารามิเตอร์ และโมเดลราสช์ (Bastari. 2000 : 7; citing Traub. 1983) โปรแกรมการประเมินที่เป็นที่รู้จัก ได้แก่ NAEP CTBS ASVAB LSAT GMAT LSAT GMAT SAT GRE และ TOFEL ได้ประยุกต์ใช้โมเดลโลจิสติกสามพารามิเตอร์สำหรับข้อสอบเลือกตอบ ซึ่งโมเดลเหมาะสม (fit) กับหน่วยตัวอย่างที่ใช้ (Bastari. 2000 : 8; citing Hambleton. 1994)

การพัฒนาโมเดล IRT สำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า (Polytomous Item Response Theory Model) ได้ดำเนินการมาพร้อมกับโมเดล IRT ที่ข้อสอบตรวจให้คะแนนสองค่า โดยขยายโมเดล IRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า โมเดลดังกล่าวสามารถใช้กับข้อมูลผลการสอบจากแบบทดสอบที่เป็นเครื่องมือการประเมินการปฏิบัติหรือการประเมินสภาพจริง (Bastari. 2000 : 8; citing Hambleton. 1995) โมเดล IRT มีมติเดียวสำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า ได้แก่ โมเดลเกรด เรสพอนส์ (Graded Response Model : GRM) (Samejima. 1997 : 85 - 100) โมเดลนอมินอล เรสพอนส์ (Nominal Response Model : NRM) (Bock.1972 : 29 - 51) โมเดลเรตติ้ง สเกล (Rating Scale) (Andersen. 1977 : 69 - 81; Andrich. 1978 : 561 - 573) โมเดลพาเชียล เครดิต (Partial Credit Model : PCM) (Masters and Wright. 1982 : 101 - 121) โมเดลเจเนอรอลไลซ์ พาเชียล เครดิต (Generalized Partial Credit Model : GPCM) (Muraki. 1997 : 153 - 164) โมเดลสเตป (Step Model) (Verhelst, Glas and De Vries. 1997 : 123 - 137) และโมเดลซีควเอนเชียล (Sequential Model) (Tutz. 1997 : 139 - 152) โมเดลเหล่านี้สามารถจัดประเภทได้ว่าเป็นโมเดลชนิดที่เป็นแบบนอมินอลหรือแบบจัดลำดับ (ordinal) ขึ้นอยู่กับข้อตกลงเบื้องต้นของโมเดลเกี่ยวกับคุณลักษณะของข้อมูล (Van der Linden & Hambleton. 1997 : 30) โมเดลเกรด เรสพอนส์ โมเดลนอมินอล เรสพอนส์ โมเดลพาเชียล เครดิต และโมเดลเจเนอรอลไลซ์ พาเชียล เครดิต เป็นโมเดลสำคัญที่ใช้ในการวิเคราะห์ข้อสอบวัดผลสัมฤทธิ์ โมเดลเกรด เรสพอนส์ โมเดลพาเชียล เครดิต และโมเดลเจเนอรอลไลซ์ พาเชียล เครดิต เป็นโมเดลในกลุ่มที่ลำดับชั้น (Category) การตอบเรียงลำดับ (Ordered Polytomous Category) โดยโมเดลเกรด เรสพอนส์ โมเดลนอมินอล เรสพอนส์ และโมเดลเจเนอรอลไลซ์ พาเชียล เครดิต พัฒนามาจากโมเดลโลจิสติกสองพารามิเตอร์ที่ข้อสอบตรวจให้คะแนนสองค่า โปรแกรมที่ใช้วิเคราะห์ข้อมูลที่รู้จักกันโดยทั่วไปในปัจจุบัน ได้แก่ โปรแกรม MULTILOG (Bastari. 2000 : 2; citing Thissen. 1991) และโปรแกรม PARSCALE (Bastari. 2000 : 2; citing Muraki and Bock. 1993)

โมเดลเกรด เรสพอนส์ เป็นโมเดลการวิเคราะห์ที่แพร่หลายมักใช้กับข้อมูลที่มีการประเมินการปฏิบัติ มาตรฐานประเมินค่าทัศนคติแบบจัดอันดับเป็นไม่เห็นด้วยอย่างยิ่ง ไม่เห็นด้วย เห็นด้วย เห็นด้วยอย่างยิ่ง หรือใช้กับเกรดที่ประเมินผลนักเรียน ได้แก่ A B C D F หรือใช้กับแบบทดสอบที่มีลักษณะเป็นเครดิตบางส่วน (partial credit) ซึ่งบ่งบอกระดับความสำเร็จในการแก้ปัญหาของผู้สอบ คือ การตอบในลำดับชั้น (category) ที่ต่ำกว่าแสดงถึงการมีคุณสมบัติที่มุ่งวัดโดยข้อสอบนั้นน้อยกว่าการตอบในลำดับชั้นที่สูงกว่า (Samejima. 1997 : 85) นิพจน์ทางคณิตศาสตร์สำหรับโมเดลมีดังนี้

$$P_{jk}(\theta) = P_{jk}^+(\theta) - P_{j,k+1}^+(\theta) \quad (2)$$

เมื่อ $P_{jk}(\theta)$ คือ ความน่าจะเป็นของผู้สอบที่มีระดับความสามารถ θ ตอบข้อสอบ j ในลำดับชั้นคะแนน k

$P_{jk}^+(\theta)$ คือ ความน่าจะเป็นของผู้สอบที่มีระดับความสามารถ θ ที่เลือกลำดับชั้นคะแนน k หรือลำดับชั้นสูงกว่าของข้อสอบ j

θ คือ ระดับความสามารถของผู้สอบ

โมเดลเกรด เรสพอนส์ เป็นลักษณะของโมเดลในกลุ่มที่เรียกว่า โมเดลความแตกต่าง (Differential Model) (Dodd, Ayala and Koch. 1995 : 6) ดังนั้นความน่าจะเป็นที่จะได้คะแนน k คือความแตกต่างระหว่างความน่าจะเป็นที่ได้จากคะแนน k กับที่ได้จากคะแนน $k + 1$ หรือคะแนนที่สูงกว่า

$P_{jk}^+(\theta)$ มีสมการดังต่อไปนี้

$$P_{jk}^+ = \frac{e^{Da_j(\theta - b_{jk})}}{1 + e^{Da_j(\theta - b_{jk})}}, \quad k = 0, 1, 2, \dots, m \quad (3)$$

$$P_{j0}^+ = 1.0, \quad P_{j,m+1}^+ = 0.0$$

เมื่อ $P_{jk}^+(\theta)$ และ θ ถูกกำหนดไว้ข้างต้นและ

D คือ องค์ประกอบของมาตราวัด

m คือ ลำดับชั้นคะแนนสูงสุด

a_j คือ พารามิเตอร์อำนาจจำแนก (ความชัน) สำหรับข้อสอบ j

b_{jk} คือ พารามิเตอร์ความยาก (จุดเริ่มเปลี่ยน) สำหรับลำดับชั้นคะแนน k ของข้อสอบ j

ตามสมการ (3) หมายความว่าความน่าจะเป็นที่ได้จากทุกลำดับชั้นคะแนน ($k \geq 0$) ต้องเท่ากับ 1 ความน่าจะเป็นของคะแนนที่เป็นไปไม่ได้ (คะแนนที่สูงกว่าค่าสูงสุดของลำดับชั้นคะแนน k) ต้องเป็น 0 และความน่าจะเป็นของคะแนน k ใด ๆ ต้องไม่ต่ำกว่าศูนย์

โมเดลนอมินอล เรสพอนส์ เป็นโมเดลที่เหมาะสมกับข้อสอบที่การตอบไม่สามารถเรียงลำดับตามระดับคำตอบถูกต้องหรือระดับคุณลักษณะที่ข้อนั้นมุ่งวัดได้นำไปใช้เพื่อแก้ปัญหาเกี่ยวกับการสอบเช่น ประยุกต์ใช้กับการทดสอบแบบปรับเหมาะ โมเดลนอมินอล เรสพอนส์ได้เพิ่มความแม่นยำในการประมาณค่าความสามารถของบุคคลโดยเฉพาะคนที่มีความสามารถต่ำ โดยใช้สารสนเทศจากคำตอบที่ไม่ถูกต้องของผู้สอบ (Dodd, Ayala and Koch. 1995 : 8) โมเดลนอมินอล เรสพอนส์ เป็นโมเดลที่ใช้ฟังก์ชันโลจิสติกแบบตัวแปรพหุพร้อมกับโมเดลของฟังก์ชันการตอบลำดับชั้น (category response function : CRF เป็นโค้งแสดงลำดับชั้นหรือเส้นที่ลากจากจุดที่ตัดกัน (trace line) ของลำดับชั้นการตอบ) สำหรับลำดับชั้นการตอบ m_j ของข้อสอบ ฟังก์ชันของโมเดลนอมินอล เรสพอนส์ คือ

$$P_{jv}(\theta_i) = \frac{\exp[c_{jv} + d_{jv}(\theta_i)]}{\sum_{k=1}^{m_j} \exp[c_{jk} + d_{jk}(\theta_i)]} \quad (4)$$

เมื่อ k คือ ดัชนีลำดับชั้นการตอบของข้อสอบมีค่าเท่ากับ $1, 2, \dots, m_j$

v คือ ดัชนีลำดับชั้นการตอบที่สนใจ

c_{jk} คือ พารามิเตอร์จุดตัด (ความยาก) สำหรับลำดับชั้นการตอบ k ของข้อที่ j

d_{jk} คือ พารามิเตอร์ความชัน (อำนาจจำแนก) ของลำดับชั้นการตอบ k ของข้อที่ j

กำหนดให้ $Z_{jk}(\theta_i) = c_{jk} + d_{jk} \theta_i$ เป็นความเบี่ยงเบนแบบโลจิสติกตัวแปรพหุ สำหรับลำดับชั้นการตอบ k ของข้อ j ณ θ_i

บอค (Bock. 1972 : 31) ได้แก้ปัญหาโดยกำหนดข้อจำกัดต่อไปนี้ในข้อสอบแต่ละข้อแยกกัน

$$\sum_{k=1}^{m_j} Z_{jk}(\theta_i) = 0 \quad (5)$$

ซึ่งหมายความว่า

$$\sum_{k=1}^{m_j} c_{jk} = 0 \quad (6)$$

และ

$$\sum_{k=1}^{m_j} d_{jk} = 0 \quad (7)$$

โมเดลพาเชียล เครดิต เป็นโมเดลที่ขยายจากโมเดลราสช์ซึ่งใช้กับข้อสอบที่ตรวจให้คะแนนแบบสองค่า (0, 1) โมเดลพาเชียล เครดิต เกิดจากการตรวจให้คะแนนความรู้บางส่วนในการแก้โจทย์ปัญหาคณิตศาสตร์ โดยพิจารณาให้ลำดับขั้นของความสำเร็จในการแก้โจทย์ปัญหาเป็นคะแนนเรียงลำดับ 0 1 2 3 (Wright and Masters. 1982 : 40 – 42) และเป็นโมเดลที่ง่ายที่สุดในกลุ่มของโมเดลที่ใช้กับข้อสอบที่ตรวจให้คะแนนหลายค่าแบบเรียงลำดับขั้นการตอบมีข้อตกลงเบื้องต้นของโมเดลคือค่าอำนาจจำแนกของแต่ละข้อเท่ากัน (Masters and Wright. 1997 : 101) โมเดลพาเชียล เครดิต คล้ายกับ โมเดลเกรด เรสพอนส์ ตรงที่ โมเดล PCM เหมาะสำหรับข้อสอบที่ตรวจให้คะแนนเป็นขั้น ๆ ตามตัวเลือก ซึ่งโอกาสที่ผู้สอบระดับความสามารถ θ จะได้รับลำดับขั้นคะแนน x ในข้อที่ i ดังนี้

$$P_{ix}(\theta_i) = \frac{\exp\left[\sum_{k=0}^x (\theta_i - b_{ik})\right]}{\sum_{h=0}^{m_i} \exp\left[\sum_{k=0}^h (\theta_i - b_{ik})\right]} \quad (8)$$

โดย b_{ik} คือ พารามิเตอร์ความยากประจำลำดับขั้นที่ k ของข้อที่ i ซึ่งมีจำนวนขั้นความยากอยู่ m_i ค่าสำหรับข้อที่ i เพื่อให้ง่ายขึ้น มาสเตอร์ (Dodd, Ayala and Koch. 1995 : 9; citing Masters. 1982) กำหนดให้ $\sum (\theta - b_{ik}) = 0$ เมื่อ $k = 0$ ลำดับขั้นคะแนนในโมเดล PCM แต่ละขั้นไม่จำเป็นต้องเรียงตามความยาก เช่นลำดับขั้นคะแนนที่สองอาจง่ายกว่าลำดับขั้นคะแนนที่หนึ่ง

โมเดลเจเนอรอลไลซ์ พาเชียล เครดิต มูรากิ (Muraki. 1992 : 101) ได้พัฒนามาจากโมเดลพาเชียล เครดิต โดยยกเลิกข้อตกลงเบื้องต้นที่ว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากัน โมเดลเจเนอรอลไลซ์ พาเชียล เครดิต เป็นกรณีเฉพาะของโมเดลนอมินอล เรสพอนส์ เมื่อลำดับขั้นการตอบเรียงลำดับ แต่ค่าความยากในแต่ละลำดับขั้นไม่จำเป็นต้องเรียงลำดับ (Dodd, Ayala and Koch. 1995 : 10) มีฟังก์ชันของโมเดลดังนี้

$$P_{jik}(\theta_j) = \frac{\exp\left[\sum_{v=1}^k Da_i(\theta_j - b_i + d_k)\right]}{\sum_{c=1}^m \exp\left[\sum_{v=1}^c Da_i(\theta_j - b_i + d_k)\right]} = \frac{\exp\left[\sum_{v=1}^k Da_i(\theta_j - b_{ik})\right]}{\sum_{c=1}^m \exp\left[\sum_{v=1}^c Da_i(\theta_j - b_{ik})\right]} \quad (9)$$

เมื่อ P_{jik} คือ ความน่าจะเป็นที่ผู้สอบ j ระดับความสามารถ θ จะตอบข้อสอบ
ในลำดับขั้นการตอบ k ของข้อสอบ i

b_i คือ พารามิเตอร์ตำแหน่ง (ความยาก) ข้อสอบ i

a_i คือ พารามิเตอร์ความชัน (อำนาจจำแนก) ข้อสอบ i

d_k คือ ความยากในแต่ละลำดับขั้น (Step or Category)

ค่าของ d_k ในข้อสอบแต่ละข้อไม่จำเป็นต้องมีลำดับอย่างต่อเนื่องกัน คือ ความยาก
ลำดับขั้น k กับลำดับขั้นอื่น ๆ ภายในข้อสอบข้อเดียวกัน อาจไม่สอดคล้องกับ ตำแหน่ง
ของข้อสอบ b_i เฉพาะพารามิเตอร์ข้อสอบลำดับขั้น $m_i - 1$ ที่สามารถจะระบุจำนวนของลำดับขั้น
การตอบคือ m_i ความยากของขั้นที่หนึ่ง (d_1) ในข้อสอบแต่ละข้อถูกกำหนดอย่างอิสระเป็นศูนย์
และกำหนดให้ $\sum_{k=2}^{m_i} d_k = 0$ เพื่อกำจัดความไม่แน่นอนออกไป (Muraki. 1992 : 169 - 170) และ
 b_{ik} เท่ากับ $b_i - d_k$

ได้มีการศึกษาโมเดลต่าง ๆ เหล่านี้รวมทั้งเปรียบเทียบความเหมาะสมกับข้อมูล โดย
เฉพาะโมเดลเกรด เรสพอนส์ และโมเดลเจเนอรอลไลซ์ พาเซียล เครดิต ซึ่งโอลิวาเรส ดราสโกว์
และมิด (Olivares, Drasgow and Mead. 1994 : 254 - 255) ได้ตัดสินว่าโมเดลทั้งสองเหมาะกับ
ข้อมูลพอ ๆ กัน และจากการศึกษาในงานวิจัยอื่น ๆ ให้ข้อสรุปเหมือนกัน (Bastari. 2000 : 10;
citing Tang Eignor. 1997) แต่โมเดลทั้งสองมีข้อแตกต่างที่สำคัญคือ ทิชเซน (ธนวัฒน์ แสนสุข.
2538 : 8; อ้างอิงจาก Thissen. 1986) ได้พัฒนาโปรแกรม MULTILOG เพื่อใช้ประมาณค่าพารามิเตอร์
ด้วยความควรจะเป็นสูงสุดขอบ (Marginal Maximum likelihood : MML) โดยมีข้อดกลง
เบื้องต้นของโมเดลว่า ในแต่ละลำดับขั้นคะแนนจะต้องมีความต่อเนื่องและเรียงตามลำดับค่าความ
ยาก ขณะที่โมเดลเจเนอรอลไลซ์ พาเซียล เครดิต ใช้วิธีประมาณค่าพารามิเตอร์ข้อสอบและผู้สอบ
แบบความควรจะเป็นสูงสุดขอบ ด้วยวิธีการทางสถิติที่เรียกว่าอีเอ็ม อัลกอริทึม (EM algorithm)
เพื่อดูความสอดคล้องของการประมาณค่า 2 ขั้นตอน วิธีประมาณค่าดังกล่าวใช้ในโปรแกรม
คอมพิวเตอร์ PARSCALE

นอกจากนี้ ได้มีการตรวจสอบ โมเดลเกรด เรสพอนส์ ในแง่มุมต่าง ๆ เช่น ในรูปของ
การประมาณค่าพารามิเตอร์ ซึ่งรีสและยู (Reise and Yu. 1990 : 142) แนะนำว่า ผู้สอบอย่างน้อย
500 คน จะทำให้การปรับข้อสอบมีความเพียงพอสำหรับข้อสอบน้อยที่สุด 25 ข้อ ถึงแม้ว่าผู้สอบ
จำนวนตั้งแต่ 250 คน จะสามารถวิเคราะห์ค่าพารามิเตอร์ด้วยโปรแกรม MULTILOG ได้ก็ตาม
แต่เอ็มเบรตสัน (Embretson and Reise. 2000 : 323) ได้แนะนำให้ใช้จำนวนผู้สอบขั้นต่ำสุด
อย่างน้อย 350 คน สำหรับสถานการณ์ทดสอบจริง ซึ่งสอดคล้องกับการศึกษาของแอนเคนแมน
และ สโตน (ธนวัฒน์ แสนสุข. 2538 : 44; อ้างอิงจาก Ankenmann and Stone. 1992) ที่ใช้ข้อ
สอบจำนวน 5 10 และ 20 ข้อ กับกลุ่มตัวอย่างขนาด 250 500 และ 1000 คน ปรากฏว่าการ
ประมาณค่าพารามิเตอร์ของผู้สอบมีความแม่นยำตั้งแต่กลุ่มตัวอย่างขนาด 500 คนขึ้นไป และพบ

ว่าหากการแจกแจงของกลุ่มประชากรมีลักษณะเบ้มากควรใช้กลุ่มตัวอย่างประมาณ 1,000 คนขึ้นไป (ธนวัฒน์ แสนสุข. 2538 : 44; อ้างอิงจาก Zwinderman and Wollenberg. 1990) นอกจากนี้ยังได้มีการศึกษาการใช้โมเดล GRM กับมาตรวัดทัศนคติ แบบทดสอบวัดผลสัมฤทธิ์ และประยุกต์ใช้กับแบบทดสอบชนิดปรับเหมาะ (Computerized adaptive test) และการทำหน้าที่ต่างกันของข้อสอบ (Dodd, Ayala and Koch. 1995 : 11 – 18; Kim & Cohen. 1998 : 345 - 355)

สำหรับรูปแบบของ “ข้อสอบที่ตรวจให้คะแนนหลายค่า” โดยทั่ว ๆ ไป กล่าวถึงข้อสอบที่ต้องการให้ผู้สอบเขียนคำตอบ (Supply type) ในหลาย ๆ รูปแบบ มากกว่าการเลือกคำตอบ (Selection type) จากตัวเลือก ส่วนใหญ่ข้อสอบเขียนตอบจัดอยู่ในประเภทการตอบสั้น การตอบที่เปิดกว้าง และการแก้ปัญหา จนถึงข้อสอบเขียนตอบที่เป็นการเขียนบทความหรือข้อความยาว ๆ แต่ข้อสอบเขียนตอบมีจุดอ่อน เกี่ยวกับผู้ตรวจให้คะแนนที่เปลี่ยนแปลงไปในแต่ละปีการศึกษา (Tate. 1999 : 344 – 345) ซึ่งอาจจะทำให้เกิดความไม่คงที่ในการให้คะแนน และยังมีความเห็นของนักเศรษฐศาสตร์ (Kennedy & Walstad. 1997 : 374 – 375) ที่มีข้อวิจารณ์ว่าข้อสอบเขียนตอบเป็นข้อสอบที่ใช้เวลาในการตรวจให้คะแนนทำให้มีค่าใช้จ่ายสูง หรือหลักฐานเกี่ยวกับความแตกต่างระหว่างเพศเมื่อทำข้อสอบเขียนตอบ (DeMars. 1998 : 279; Garner & Engelhard. 1999 : 44) แต่ข้อสอบเขียนตอบเป็นเครื่องมือที่เหมาะสมในการใช้ประเมินการปฏิบัติ เพราะทำให้การประเมินเป็นไปตามสภาพจริง จึงมีการใช้อย่างกว้างขวาง

นอกจากนี้ยังมีข้อสอบเลือกตอบที่ตรวจให้คะแนนหลายค่าซึ่งใช้วัดความรู้บางส่วนของผู้เรียนได้เช่นเดียวกัน ข้อสอบประเภทนี้มีข้อดีในด้านการตรวจให้คะแนนที่มีความเป็นปรนัยมากกว่าข้อสอบเขียนตอบ โดยนักวัดผลการศึกษาได้พัฒนาจากข้อสอบเลือกตอบธรรมดาเพื่อให้สารสนเทศเกี่ยวกับความรู้ของผู้สอบได้มากขึ้น ซึ่งลอร์ดและโนวิก (Simon, Budescu and Nevo. 1997 : 66 – 69; citing Lord and Novick. 1968) ได้เสนอไว้สามวิธีการใหญ่ ๆ คือ วิธีแรกเป็นการปรับโครงสร้างข้อสอบใหม่หรือวิธีการตอบใหม่ วิธีการปรับโครงสร้างข้อสอบใหม่เป็นการเปลี่ยนแปลงลักษณะข้อสอบ (Changing the item structure) ได้แก่ ข้อสอบที่มีคำตอบถูกหลาย ๆ คำตอบ คะแนนที่ผู้สอบได้รับเท่ากับการนำคะแนนจากการระบุตัวเลือกที่ถูกต้องมารวมกัน นอกจากนี้ การตรวจให้คะแนนในบางกรณี จะมีการหักคะแนนกรณี que ผู้สอบเลือกตัวเลือกผิดแบบทดสอบในวิธีการนี้ได้มีการพัฒนาเป็นแบบทดสอบที่มีคำตอบถูกทุกข้อ หรือผิดทุกข้อ และลักษณะของข้อสอบที่เป็นกลุ่มคำถามและกลุ่มของคำตอบที่เรียกว่าข้อสอบจับคู่ ข้อสอบประเภทนี้จะทำให้การเดาข้อสอบลดลง และสามารถประเมินความรู้บางส่วนได้ แต่วิธีการนี้มีข้อจำกัดเกี่ยวกับเวลาที่ใช้ในการทำข้อสอบ สำหรับวิธีการตอบใหม่ เป็นการเปลี่ยนแปลงวิธีการตอบข้อสอบ (Changing the response method) โดยให้ผู้สอบกำหนดน้ำหนักคะแนนเอง ซึ่งเป็นการสะท้อนความรู้ที่มีอยู่ในตัวผู้สอบ วิธีการดังกล่าวเป็นวิธีการประเมินตนเอง สามารถให้สารสนเทศเกี่ยวกับความรู้บางส่วนได้อย่างน่าเชื่อถือ นอกจากนี้ยังทำให้การเดาและความคลาดเคลื่อนของ

การวัดลดลง แต่มีข้อจำกัดคือเป็นวิธีที่มีความยุ่งยากในทางปฏิบัติเพราะต้องใช้เวลาสอบมาก วิธีที่สองเป็นวิธีการให้คะแนนข้อสอบแตกต่างกัน (Differential item scoring methods) เป็นการกำหนดน้ำหนักคะแนนตัวเลือกให้แตกต่างกัน โดยใช้วิจารณ์ฐานของผู้เชี่ยวชาญ หรือเป็นการให้น้ำหนักจากข้อมูลเชิงประจักษ์ที่ผ่านมาในอดีตหรือปัจจุบัน เช่นคะแนนมาตรฐานเฉลี่ยของผู้สอบที่เลือกแต่ละตัวเลือก ความสัมพันธ์ระหว่างตัวเลือกที่ผู้สอบเลือกกับคะแนนรวมทั้งฉบับ จากการศึกษาพบว่าวิธีนี้ให้ค่าความเที่ยงตรง และความเชื่อมั่นไม่แตกต่างจากแบบทดสอบเลือกตอบธรรมดา แต่การคำนวณคะแนนมีความซับซ้อน และยากต่อการอธิบายการให้คะแนนแก่ผู้สอบ วิธีที่สามเป็นการให้คะแนนแบบทดสอบแตกต่างกัน (Differential test scoring methods) เป็นการกำหนดน้ำหนักคะแนนเมื่อข้อสอบมีคุณภาพจะได้น้ำหนักมากกว่าข้อสอบไม่มีคุณภาพ ดังนั้นเมื่อใช้ทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory : CTT) จะทำให้น้ำหนักของคะแนนแต่ละข้อไม่คงที่ จึงควรใช้ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) วิธีการทั้ง 3 วิธีใช้แก้ไขจุดอ่อนของข้อสอบเลือกตอบ แต่ละวิธีมีจุดเด่น จุดด้อยที่แตกต่างกันไป แบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบและให้ผู้สอบตัดสินทุกคำตอบ (Multiple true false : MTF) เป็นรูปแบบหนึ่งที่มีผู้นำไปใช้อย่างกว้างขวางเช่นเดียวกัน และเหมาะในการใช้กับโปรแกรมการสอบขนาดใหญ่

ปัจจุบันได้มีการนำแบบทดสอบรูปแบบผสม (Hybrid or Mix Tests) ที่ประกอบด้วยข้อสอบ ที่ตรวจให้คะแนนสองค่าและหลายค่ารวมในฉบับเดียวกันมาใช้ ข้อสอบเหล่านี้เป็นข้อสอบเลือกตอบหลายตัวเลือก (Multiple Choice : MC) และข้อสอบเขียนตอบ (Construct Response : CR) ซึ่งแฮมเบตัน (Hambleton, 1996 : 899) ได้กล่าวไว้ว่าการประเมินทางการศึกษาในทศวรรษ 20 และในทศวรรษหน้าจะมีความแตกต่างอย่างมากจากการประเมินในสองทศวรรษที่ผ่านมา รูปแบบที่เป็นปรนัยของการประเมินดังเช่นแบบทดสอบเลือกตอบจะถูกทำให้สมดุลอย่างเป็นไปได้มากที่สุด ด้วยการวัดการประเมินการปฏิบัติจากงานเขียนโครงงานและพอดโฟลิโอ การวิเคราะห์ข้อสอบรูปแบบผสมจะต้องวิเคราะห์รวมกันหลายโมเดลโดยประยุกต์ใช้โมเดล IRT ทั้ง 2 กลุ่ม เพราะโมเดล IRT โลจิสติกหนึ่ง และสองพารามิเตอร์เป็นกรณีเฉพาะของโมเดลที่ข้อสอบตรวจให้คะแนนหลายค่า อย่างไรก็ตามโมเดลโลจิสติก สามพารามิเตอร์ไม่ได้ถูกขยายไปเป็นโมเดลที่ข้อสอบตรวจให้คะแนนหลายค่า เพราะพารามิเตอร์การเดาไม่ได้ถูกนำเสนอไว้ในโมเดล (Bastari, 2000 : 10; citing Mellenbery, 1994) ดังนั้นการผสมโมเดลโลจิสติกสามพารามิเตอร์กับโมเดลที่ข้อสอบตรวจให้คะแนนหลายค่าแบบเรียงลำดับหรือไม่เรียงลำดับก็ตาม (โมเดล GRM โมเดล NRM หรือ โมเดล GPCM) มีการใช้ในหลาย ๆ สถานการณ์ของการปฏิบัติ

ได้มีการศึกษารูปแบบการวิเคราะห์ข้อมูลในโมเดล IRT ที่แบบทดสอบมีข้อสอบผสมในการประเมินด้วยข้อสอบของ Advanced Placement (AP) วิชาเคมีและประวัติศาสตร์ที่การศึกษาค่อนข้างซับซ้อน โดยแบบทดสอบประกอบด้วยข้อสอบ MC และ CR ที่ผู้สอบสามารถเลือกข้อสอบ CR ได้ด้วย เป็นการประยุกต์ใช้ โมเดลโลจิสติกสองและสาม พารามิเตอร์กับโมเดล GRM

และ โมเดล NRM ให้เข้ากับชนิดของข้อสอบและการตอบ โดยวิเคราะห์ข้อสอบ MC และ CR พร้อมกันและแยกกัน โดยใช้โปรแกรม MULTILOG ผลปรากฏว่าโดยเฉลี่ยแล้วข้อสอบ CR ให้สารสนเทศน้อยกว่าข้อสอบ MC (Lukhele, Thissen and Wainer. 1994 : 234) แต่ผลการศึกษาของโดโนห์ (Donoghue.1994 : 299, 307) ปรากฏว่าให้ผลตรงกันข้ามคือ ข้อสอบ CR ให้สารสนเทศมากกว่าข้อสอบ MC โดยผู้วิจัยประยุกต์ใช้โมเดลสามพารามิเตอร์กับข้อสอบเลือกตอบ โมเดล สองพารามิเตอร์กับข้อสอบเขียนตอบแบบสั้นที่ตรวจให้คะแนนสองค่า และโมเดลเจเนอรอลไลซ์ พาเซียล เครดิต กับข้อสอบปลายเปิด ข้อมูลเป็นข้อสอบการอ่าน จาก NAEP วิเคราะห์ค่าโดยใช้โปรแกรม PARSCALE

นอกจากนี้ยังมีการศึกษาเปรียบเทียบโมเดล IRT ผสม โดยฟิตซ์แพทริก และคณะ (Fitzpatrick et al. 1996 : 291) ซึ่งใช้ข้อมูลจริงจากการประเมินของรัฐ เป็นการเปรียบเทียบความสมมูลของโมเดลเจเนอรอลไลซ์ พาเซียล เครดิต กับ โมเดลพาเซียล เครดิต สำหรับการวิเคราะห์ข้อสอบ CR ใช้โมเดลพาเซียล เครดิต 2 พารามิเตอร์ (2PPC) และโมเดลพาเซียลเครดิต 1 พารามิเตอร์ (1PPC) ข้อสอบ MC วิเคราะห์ด้วยโมเดลโลจิสติก 3 พารามิเตอร์ และ 1 พารามิเตอร์ โดยปรับค่าพารามิเตอร์พร้อมกัน และยังศึกษาด้วยข้อมูลจำลองที่สร้างขึ้น วิเคราะห์ข้อมูลโดยใช้โปรแกรม PARDUX ปรากฏว่าโมเดล 2PPC ไม่ว่าจะวิเคราะห์รวมหรือแยกกับโมเดล 3 พารามิเตอร์ มีความเหมาะสมกับข้อมูลมากกว่าโมเดล 1PPC ที่ไม่ว่าจะรวมหรือแยกกับโมเดลโลจิสติก 1 พารามิเตอร์ก็ตาม ซึ่งแทงและอิจเนอร์ (Bastari. 2000; citing Tang and Eignor. 1997) ได้พยายามเปรียบเทียบความแตกต่างของโมเดลผสมสำหรับชุดของข้อมูลโปรแกรมการทดสอบภาษาอังกฤษเป็นภาษาต่างชาติ (TOFEL) ซึ่งประกอบด้วยข้อสอบ MC และข้อสอบ CR โดยใช้โมเดล 3 พารามิเตอร์กับโมเดลเกรด เรสพอนส์ และโมเดล 3 พารามิเตอร์กับโมเดลเจเนอรอลไลซ์ พาเซียล เครดิต วิเคราะห์โดยใช้โปรแกรม PARSCALE ปรากฏว่าโมเดลทั้งสองแบบเหมาะสมกับข้อมูลเป็นอย่างดี

สรุปได้ว่า โมเดล IRT ที่ใช้ในการวิเคราะห์ข้อสอบที่ตรวจให้คะแนนหลายค่า โมเดลที่มี 2 พารามิเตอร์ ได้แก่ โมเดลเกรด เรสพอนส์ และ โมเดลเจเนอรอลไลซ์ พาเซียล เครดิต ส่วนใหญ่จะเหมาะสมกับข้อมูลมากกว่าโมเดลพาเซียล เครดิต ซึ่งเป็นโมเดลพารามิเตอร์เดียว โมเดลเจเนอรอลไลซ์ พาเซียล เครดิต และโมเดลเกรด เรสพอนส์ เป็นโมเดลที่รู้จักแพร่หลายให้ผลการวิเคราะห์ที่คล้ายคลึงกันเป็นส่วนใหญ่ทั้งในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่าและแบบทดสอบที่ข้อสอบอยู่ในรูปแบบผสม ได้มีการประยุกต์ใช้โมเดลเกรด เรสพอนส์ กับการวัดหลาย ๆ รูปแบบ โดยเฉพาะแบบทดสอบวัดผลสัมฤทธิ์ ซึ่งส่วนใหญ่ใช้ในข้อสอบเขียนตอบ ในการศึกษาครั้งนี้ผู้วิจัยได้ประยุกต์ใช้โมเดลเกรด เรสพอนส์ ในแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบและให้ผู้สอบตัดสินใจเลือก ซึ่งเป็นแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า และเป็นเครื่องมือที่ผู้วิจัยนำมาใช้ครั้งนี้

2. แนวคิดการเทียบคะแนนระหว่างแบบทดสอบ

คำศัพท์ที่เกี่ยวข้องกับคำว่า "การเทียบคะแนนระหว่างแบบทดสอบ" (Test Equating) มีอยู่หลายคำ มีสเรวี (Johnson and Owen. 1998 : Part 3 P 1 – 2; citing Mislevy. 1992) และ ลินน์ (Johnson and Owen. 1998 : Part 3 P 1 – 2; citing Linn. 1993) ได้อธิบายให้เห็นความแตกต่างของคำศัพท์ 4 คำ คือ Equating Calibration Projection และ Moderation ต่างเป็นรูปแบบหนึ่งของการเชื่อมต่อ (linking) กระบวนการจำเป็นที่จะทำให้การเชื่อมต่อได้รับผลสำเร็จขึ้นอยู่กับว่าการประเมินทั้งสองมีความเหมือนกันในด้านเป้าหมาย การครอบคลุมเนื้อหาและคุณสมบัติของการวัดเพียงใด มีสเรวี และ ลินน์ ได้อธิบายคำทั้ง 4 คำข้างต้นว่ามีความแตกต่างกันในข้อตกลงเบื้องต้นที่กำหนด ซึ่งลดน้อยลงตามลำดับ ข้อกำหนดของการเทียบคะแนน(Equating) จะมีข้อตกลงที่เข้มงวดกว่าและมอดเดอเรชั่น (Moderation) มีข้อตกลงที่ผ่อนปรนมากที่สุด

มีสเรวีและลินน์ ได้กล่าวถึง การเทียบคะแนน ว่าเป็นการเชื่อมต่อการประเมินสอง การประเมินด้วยแบบทดสอบที่สร้างจากตารางวิเคราะห์รายละเอียดเดียวกัน ข้อกำหนดคือความเท่าเทียมกันในโครงสร้าง เช่น ความยาก รูปแบบของคำถาม วิธีการจัดการ จำนวนข้อสอบและความแม่นยำในการวัดแต่ละระดับของคะแนน ภายใต้สภาพการณ์ที่ควบคุม ทำให้ผลการประเมินสามารถสับเปลี่ยนกันได้ จึงสร้างตารางคะแนนฉบับ X ที่สอดคล้องกับคะแนนฉบับ Y ได้ ดังนั้น ทุก ๆ ข้อคำถามจึงกำหนดคะแนนจากการประเมิน X ซึ่งกำหนดในวิธีเดียวกันกับคะแนนแปลงจากการประเมิน Y เมื่อพิจารณาเทียบกับ การปรับ (Calibration) ซึ่งเป็นการเชื่อมต่อที่เงื่อนไขน้อยกว่าการเทียบคะแนน คือ ถ้าการประเมิน Y ถูกสร้างจากตารางวิเคราะห์รายละเอียดเดียวกันกับการประเมิน X แต่มีความแม่นยำแตกต่างกันหรือระดับของความยากแตกต่างกัน ในกรณีที่ไม่สามารถเทียบคะแนนได้ แต่ผลของการประเมินทั้งสองสามารถถูกปรับให้เหมาะสมได้ และสำหรับ โปรเจกชันและมอดเดอเรชั่นเป็นการเชื่อมต่อมาตรวัดเมื่อข้อสอบวัดในโครงสร้างที่แตกต่างกัน

ในขณะเดียวกันได้มีผู้ให้ความหมายการเทียบคะแนน คือ แองกอฟฟ์ (Angoff. 1984 : 85) ได้ให้ความหมายของการเทียบคะแนนระหว่างแบบทดสอบว่าเป็นการแปลงระบบคะแนนของแบบทดสอบฉบับหนึ่งให้อยู่บนสเกลของแบบทดสอบฉบับอื่น คะแนนจากแบบทดสอบทั้งสองฉบับที่แปลงแล้วถือว่าเท่าเทียมกัน จึงสามารถเปรียบเทียบกันได้ ส่วนลอร์ด (Lord. 1980 : 195) ได้กล่าวถึงการเทียบคะแนนระหว่างแบบทดสอบไว้ว่า การเทียบคะแนนทำให้เกิดความเท่าเทียมกันระหว่างคะแนนจากแบบทดสอบสองฉบับ สำหรับในความเห็นของโคลเลน และเบรนนอน (Kolen & Brennan. 1995 : 2-3) การเทียบคะแนนระหว่างแบบทดสอบ หมายถึงกระบวนการทางสถิติที่ใช้ในการปรับคะแนนจากแบบทดสอบต่างฉบับ เพื่อให้คะแนนจากแบบทดสอบสามารถใช้แทน (interchangeably) กันได้ โดยที่แบบทดสอบแต่ละฉบับอาจจะมีความยากแตกต่างกันได้ แต่ถ้ามีเนื้อหาต่างกันจะไม่สามารถเทียบกันได้ และชูล์คส์ ชัมกลิชิต (2529: 144) ได้สรุปความหมายของการเทียบคะแนนไว้ว่า เป็นกระบวนการที่เกี่ยวข้องกับกิจกรรม 2 ประการ คือ 1) กระบวนการ

ที่ทำให้แบบทดสอบสองฉบับใด ๆ มีความทัดเทียมกันหรือเท่ากันในเชิงโครงสร้าง และ 2) การใช้วิธีการทางสถิติเพื่อปรับ (adjust) คะแนนที่ได้จากแบบทดสอบแต่ละฉบับให้อยู่ในมาตรฐานเดียวกันและเทียบกันได้

ดังนั้น สรุปได้ว่า การเทียบคะแนนระหว่างแบบทดสอบ (Test Equating) เป็นวิธีการทางสถิติที่นำมาใช้ในการปรับคะแนนของแบบทดสอบต่างฉบับที่วัดคุณลักษณะเดียวกันให้อยู่ในมาตรฐานเดียวกัน คะแนนจากแบบทดสอบทั้งสองฉบับจึงเทียบกันได้อย่างมีความหมาย การเทียบคะแนนระหว่างแบบทดสอบเป็นกระบวนการวิเคราะห์ จึงต้องมีแบบแผนในการรวบรวมข้อมูลและออกแบบวิธีการเทียบคะแนน ดังมีรายละเอียดต่อไปนี้

2.1 แบบแผนการรวบรวมข้อมูล (Data Collection Designs)

การรวบรวมข้อมูลสำหรับนำผลมาใช้ในการเทียบคะแนนสามารถจำแนกได้ 3 แบบแผนใหญ่ (Angoff. 1984 : 93 – 121; Petersen & others. 1989 : 244 – 246; Kolen & Brennan. 1995 : 13 – 21; ศิริชัย กาญจนวาสี. 2541 : 8 - 10) การเลือกแบบแผนใดจะเกี่ยวข้องกันทั้งวิธีการปฏิบัติและประเด็นของสถิติที่นำมาใช้ ดังมีรายละเอียดแบบแผนการรวบรวมข้อมูลต่อไปนี้

2.1.1 แบบแผนกลุ่มเดี่ยว (Single Group Design) แบ่งเป็น 2 รูปแบบ คือ

2.1.1.1 แบบแผนกลุ่มเดี่ยวที่ไม่ได้รับการจัดให้สมดุล (Single Group Design With Incounterbalancing)

รูปแบบนี้เป็นรูปแบบที่ผู้สอบกลุ่มเดียวทำแบบทดสอบสองฉบับโดยทำแบบทดสอบฉบับหนึ่งแล้วตามด้วยแบบทดสอบอีกฉบับหนึ่งทำให้ระดับความสามารถของผู้สอบทั้งสองฉบับเท่ากัน ช่วยลดปัญหาความยากของข้อสอบต่างฉบับที่แตกต่างกันได้ แต่จะมีผลกระทบจากการเรียนรู้ ความคุ้นเคยต่อแบบทดสอบฉบับแรกทำให้ผู้สอบได้คะแนนสูงขึ้นหรืออาจเกิดความเมื่อยล้าในการทำแบบทดสอบ ซึ่งจะทำให้เกิดความไม่ตั้งใจทำแบบทดสอบคะแนนที่ได้ก็จะลดลง

2.1.1.2 แบบแผนกลุ่มเดี่ยวที่ได้รับการจัดให้สมดุล (Single Group Design With Counterbalancing)

รูปแบบนี้ได้พัฒนาขึ้นเพื่อขจัดความลำเอียงที่เกิดในแบบแผนกลุ่มเดี่ยวที่ไม่ได้รับการจัดให้สมดุล โดยสุ่มผู้สอบออกเป็นสองกลุ่ม แต่ละกลุ่มจะได้ทำแบบทดสอบทั้งสองชุดในลักษณะของการจัดลำดับก่อนหลังต่างกัน

2.1.2 แบบแผนกลุ่มสมมูลหรือกลุ่มสุ่ม (Equivalent Group or Random Group Design)

รูปแบบนี้ใช้กลุ่มผู้สอบที่มีความสามารถใกล้เคียงกันสองกลุ่ม ซึ่งได้มาโดยการสุ่มแล้วให้ผู้สอบในแต่ละกลุ่มทำแบบทดสอบฉบับเดียว ซึ่งสามารถแก้ปัญหาการเรียนรู้หรือความเมื่อยล้าที่เกิดในแบบแผนกลุ่มเดียวได้ แต่ปัญหาอาจเกิดขึ้นได้หากกลุ่มผู้สอบแต่ละกลุ่มมีความสามารถแตกต่างกัน ซึ่งทำให้เกิดความคลาดเคลื่อนในการเทียบคะแนน การลดความแตกต่างระหว่างกลุ่มทำได้โดยใช้กลุ่มตัวอย่างขนาดใหญ่

2.1.3 แบบแผนแบบทดสอบสอบรวม (Anchor Test Design) แบ่งได้เป็น 2 รูปแบบ คือ

2.1.3.1 แบบแผนแบบทดสอบรวมภายใน (Internal Anchor Test Design)

รูปแบบนี้เป็นการใช้กลุ่มผู้สอบสองกลุ่ม โดยแต่ละกลุ่มทำแบบทดสอบคนละฉบับ (ฉบับ X หรือฉบับ Y) โดยที่แบบทดสอบแต่ละฉบับจะมีแบบทดสอบรวมรวมอยู่ด้วย

2.1.3.2 แบบแผนแบบทดสอบรวมภายนอก (External Anchor Test Design)

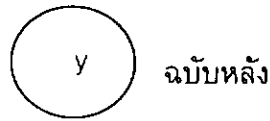
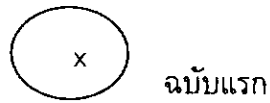
รูปแบบนี้เป็นการใช้กลุ่มผู้สอบสองกลุ่ม โดยผู้สอบแต่ละกลุ่มทำแบบทดสอบคนละฉบับและทำแบบทดสอบรวมซึ่งแยกอีกฉบับ

แบบแผนแบบทดสอบรวมนี้จะช่วยลดความลำเอียงที่เกิดจากการเรียนรู้ ความคุ้นเคย ความเมื่อยล้า และความแตกต่างระหว่างกลุ่มผู้สอบในแบบแผนกลุ่มเดียวและแบบแผนกลุ่มสมมูลได้ แบบแผนแบบทดสอบรวมจึงมีประสิทธิภาพดีกว่า แต่ประสิทธิภาพดังกล่าวขึ้นอยู่กับแบบทดสอบรวม ซึ่งต้องวัดคุณลักษณะเดียวกัน มีเนื้อหา ระดับความยาก เช่นเดียวกับแบบทดสอบที่ต้องการเทียบคะแนน และต้องมีความยาวที่เหมาะสม ซึ่งจะช่วยให้ผลการเทียบคะแนนมีความถูกต้องแม่นยำสูง ได้มีการศึกษาเกี่ยวกับความยาวของแบบทดสอบรวมภายในว่าควรมีความยาวเท่าใด แต่ยังมีได้มีข้อสรุปที่ชัดเจนนัก แต่แองกอฟฟ์ (Angoff, 1984 : 107) กล่าวว่าไว้ว่าแบบทดสอบรวมควรมีจำนวนไม่น้อยกว่า 20 ข้อ หรือน้อยกว่าร้อยละ 20 ของจำนวนข้อสอบในแบบทดสอบเทียบคะแนน แล้วแต่ว่าจำนวนไหนมากกว่าให้ใช้จำนวนนั้น ขณะที่ ไรท์ และสโตน (Wright and Stone, 1979 : 98) ได้กล่าวว่า แบบทดสอบรวมควรเป็นข้อสอบที่วัดเรื่องเดียวกับแบบทดสอบที่ใช้เทียบคะแนนทั้งสองชุดและมีจำนวนเพียง 10 ข้อ ก็เพียงพอ

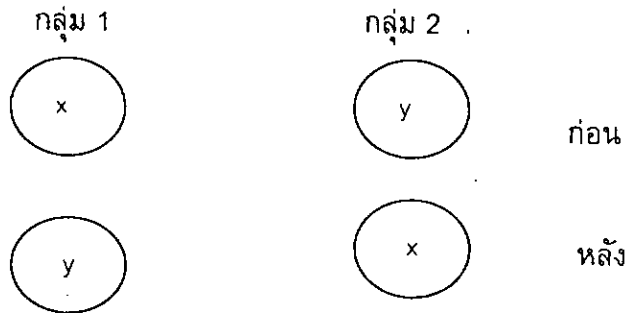
สำหรับแบบแผนแบบทดสอบรวมที่ได้กล่าวไว้ข้างต้น สามารถจำแนกตามกลุ่มผู้สอบได้ 2 ประเภท 1) แบบทดสอบอาจใช้กับกลุ่มสุ่มหรือกลุ่มเท่าเทียมกัน และ 2) แบบทดสอบอาจใช้กับกลุ่มที่ไม่เท่าเทียมกัน (Nonequivalent Groups) ซึ่งตามความหมายเมื่อนำมาใช้ในการเทียบคะแนนในแนวนอน คือ กลุ่มผู้สอบที่ไม่ได้มาจากประชากรเดียวกัน ได้แก่ ผู้สอบกลุ่มใหม่ในโปรแกรมการสอบเดียวกันหรือกลุ่มนักเรียนระดับชั้นเดียวกันคนละปีการศึกษา

โดยแบบทดสอบร่วมกันนำมาใช้อาจจะเป็นแบบทดสอบร่วมภายในหรือแบบทดสอบร่วมภายนอกก็ได้ อย่างไรก็ตาม คำว่า “กลุ่มไม่เท่าเทียมกัน” ในที่นี้เป็นคนละความหมายกับที่ใช้สำหรับการเชื่อมต่อหรือการเทียบคะแนนตามแนวตั้ง เพราะการเชื่อมต่อหรือการเทียบคะแนนในแนวตั้งนำมาใช้ในโครงสร้างของชุดแบบทดสอบผลสัมฤทธิ์ เพื่อให้มีการพัฒนามาตรวัดที่มีประสิทธิภาพตามระดับชั้นเรียน โดยกลุ่มไม่เท่าเทียมกันตามความหมายนี้คือกลุ่มผู้สอบคนละระดับชั้นเรียน (Slind & Linn. 1978 : 23 – 35; Harris & Hoover. 1987 : 151 - 159) ได้แสดงแบบแผนการรวบรวมข้อมูลในการเทียบคะแนน ตามภาพประกอบ 1

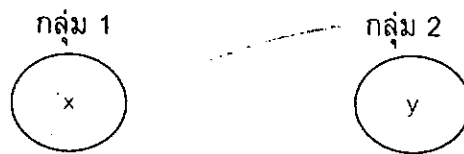
ในการศึกษาครั้งนี้ ผู้วิจัยได้ใช้แบบแผนแบบทดสอบร่วมภายในที่ใช้กับกลุ่มสุ่ม 2 กลุ่มซึ่งเป็นนักเรียนในระดับชั้นเดียวกัน แต่ละกลุ่มทำแบบทดสอบคนละฉบับที่มีความยาวฉบับละ 30 ข้อ (รวมแบบทดสอบร่วม 10 ข้อ) ซึ่งอาจจะได้รับแบบทดสอบฉบับ X (20 ข้อ) หรือฉบับ Y (20 ข้อ) และทั้งสองกลุ่มทำแบบทดสอบร่วม จำนวน 10 ข้อ ความยาวของแบบทดสอบร่วมคิดเป็นร้อยละ 33 ของแบบทดสอบเทียบคะแนนทั้งฉบับ ซึ่งยังไม่มีการศึกษาความยาวของแบบทดสอบร่วมที่ข้อสอบตรวจให้คะแนนหลายค่าไว้อย่างชัดเจนนัก แต่ผู้ที่ศึกษาเทคนิคการเทียบคะแนนส่วนใหญ่ยังใช้ข้อกำหนดของแองกอฟฟ์ คือให้แบบทดสอบร่วมมีความยาวไม่น้อยกว่าร้อยละ 20 ของแบบทดสอบเทียบคะแนน แบบแผนแบบทดสอบร่วมเป็นแบบแผนที่ผู้วิจัยส่วนใหญ่นำมาใช้กันมาก ข้อสอบร่วมจะเป็นตัวรับความแตกต่างระหว่างกลุ่มหรือระหว่างแบบทดสอบทำให้ความคลาดเคลื่อนในการเทียบคะแนนลดน้อยลง



แบบแผนกลุ่มเดียวที่ไม่ได้รับการจัดให้สมดุล



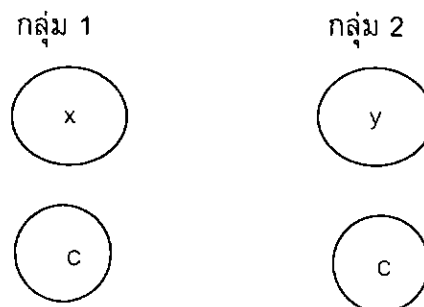
แบบแผนกลุ่มเดียวที่ได้รับการจัดให้สมดุล



แบบแผนกลุ่มสมดุลหรือกลุ่มสุ่ม



แบบแผนแบบทดสอบร่วมภายใน



แบบแผนแบบทดสอบร่วมภายนอก

ภาพประกอบ 1 แบบแผนการรวบรวมข้อมูลในการเทียบคะแนน

2.2 การออกแบบวิธีการเทียบคะแนน (Designs for Equating Methods)

วิธีการที่ใช้ในการเทียบคะแนนสามารถจำแนกตามแนวคิดพื้นฐานหรือทฤษฎีการวัดได้เป็น 2 รูปแบบ คือ การเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (Equating in Classical Test Theory) และการเทียบคะแนนตามแนวทฤษฎีการตอบข้อสอบ (Equating in Item Response Theory) (Angoff. 1984 : 93 – 121; Petersen et al. 1989 : 247 - 260; Kolen & Brennan. 1995 : 8; ศิริชัย กาญจนวาสี. 2541 : 11 - 20) มีรายละเอียดดังนี้

2.2.1 การเทียบคะแนนตามแนวทฤษฎีการทดสอบแบบมาตรฐานเดิม มีวิธีการเทียบคะแนนที่นิยมใช้กันมาก 2 วิธี คือ

2.2.1.1 วิธีการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์ (Equipercentile Equating) การเทียบคะแนนตามวิธีการนี้นิยามว่า สำหรับกลุ่มผู้สอบที่กำหนด คะแนนที่ได้จากแบบทดสอบต่างฉบับกันจะถือว่าเทียบเท่ากันถ้าตรงกับตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน กระบวนการเทียบคะแนนแบบอีควิเปอร์เซ็นต์ไทล์ประกอบด้วย 2 ขั้นตอน ได้แก่

ขั้นตอนที่ 1 การจัดทำตารางแจกแจงความถี่สะสมของแบบทดสอบ 2 ฉบับ ที่จะนำมาเทียบกัน

ขั้นตอนที่ 2 การเทียบคะแนนของแบบทดสอบ 2 ฉบับจากการแจกแจงความถี่สะสมของคะแนนที่ได้

ในทางปฏิบัติ การดำเนินการในขั้นตอนที่ 2 อาจมีปัญหาเนื่องจากไม่มีคะแนนของแบบทดสอบฉบับหนึ่งที่ตรงกับคะแนนจากแบบทดสอบอีกฉบับหนึ่ง ณ ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกันดังนั้นเพื่อให้สามารถเทียบคะแนนด้วยวิธีอีควิเปอร์เซ็นต์ไทล์ได้ จึงจำเป็นต้องปรับการแจกแจงคะแนนของแบบทดสอบให้มีลักษณะต่อเนื่อง

2.2.1.2 วิธีการเทียบคะแนนเชิงเส้นตรง (Linear Equating) การเทียบคะแนนตามวิธีการนี้นิยามว่า สำหรับกลุ่มผู้สอบที่กำหนด คะแนนที่ได้จากแบบทดสอบต่างฉบับกันถือว่าเทียบเท่ากันถ้าตรงกับคะแนนมาตรฐานค่าเดียวกัน คะแนนจากแบบทดสอบฉบับ X และแบบทดสอบฉบับ Y นั้นสามารถนำมาเทียบคะแนนเชิงเส้นตรงได้ ดังสมการต่อไปนี้

$$y = ax + b$$

โดยสัมประสิทธิ์ a และ b ประมาณค่าได้จาก (Kolen & Brennan, 1995 : 30 - 31)

$$a = \frac{\sigma(y)}{\sigma(x)}$$

$$b = \mu(y) - \frac{\sigma(y)}{\sigma(x)} \mu(x)$$

เมื่อ μ_Y และ μ_X เป็นค่าเฉลี่ยของคะแนน และ σ_Y และ σ_X เป็นส่วนเบี่ยงเบนมาตรฐานของคะแนนจากแบบทดสอบฉบับ Y และ X ตามลำดับ ดังนั้น จึงประมาณค่าคะแนนของแบบทดสอบอีกฉบับจาก

$$y = \frac{\sigma_Y}{\sigma_X} x + \left[\mu_Y - \frac{\sigma_Y}{\sigma_X} \mu_X \right] \quad (10)$$

เนื่องจากกระบวนการเทียบคะแนนเป็นการศึกษาเชิงประจักษ์ จะต้องอาศัยการออกแบบเพื่อรวบรวมข้อมูล และการออกแบบเพื่อเทียบคะแนน ซึ่งเป็นการจัดกระทำข้อมูลในทางสถิติ แองกอฟฟ์ (Angoff, 1984 : 93 – 120) ได้เสนอแบบแผนการรวบรวมข้อมูลและการจัดกระทำข้อมูลในทางสถิติของการเทียบคะแนนเชิงเส้นตรงไว้ 6 แบบแผน โดยที่แบบแผน 1 – 2 เป็นแบบแผนการเทียบคะแนนโดยไม่ใช้แบบทดสอบร่วม (คือแบบแผนที่ 2.1.1.2 และแบบแผนที่ 2.1.2 ในหน้า 28) แบบแผนที่ 3 – 6 เป็นการเทียบคะแนนโดยใช้แบบทดสอบร่วม ซึ่งแต่ละแบบแผนจะมีวิธีการประมาณค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของแบบทดสอบในสมการ (10) แตกต่างกันไป ตามเงื่อนไขของการรวบรวมข้อมูล แต่ทุกแบบแผนจะตัดสินคะแนนสมมูล จากค่าคะแนนมาตรฐานเดียวกัน คือ

$$\frac{Y - \mu_Y}{\sigma_Y} = \frac{X - \mu_X}{\sigma_X}$$

ในแต่ละแบบแผนการรวบรวมข้อมูลของแองกอฟฟ์ ได้เสนอรายละเอียดแยกระหว่างกรณีที่แบบทดสอบเทียบคะแนนมีความเชื่อมั่นเท่ากันและต่างกัน ในการวิจัยครั้งนี้ผู้วิจัยได้ใช้แบบแผนการรวบรวมข้อมูลของแองกอฟฟ์แบบแผนที่ 3 ซึ่งเป็นแบบแผนแบบทดสอบร่วมภายใน เมื่อกลุ่มที่ใช้เป็นกลุ่มสุ่ม เนื่องจากแบบทดสอบที่นำมาใช้ในการเทียบคะแนนเป็นแบบทดสอบที่สร้างขึ้นจากโครงสร้างเนื้อหาเดียวกัน และจัดแยกฉบับโดยให้มีค่าความยากใกล้เคียงกัน เมื่อนำไปใช้กับกลุ่มสุ่ม 2 กลุ่ม อาจจะทำให้ความเชื่อมั่นของแบบทดสอบ 2 ฉบับไม่แตกต่างกัน การจัดกระทำข้อมูลตามวิธีการทางสถิติ จึงใช้รูปแบบ 3A ที่แองกอฟฟ์ (Angoff, 1984 : 104 – 109) รวบรวมไว้

โดยลอร์ด (Angoff, 1984 : 105 – 106; citing Lord, 1955) ได้พัฒนาสมการสำหรับรูปแบบ เมื่อแบบทดสอบเทียบคะแนนฉบับ X และฉบับ Y มีค่าความเชื่อมั่นเท่ากัน โดยใช้การประมาณค่าความเป็นไปได้สูงสุด (Maximum likelihood estimates) ประมาณค่าเฉลี่ยและค่าความแปรปรวนของแบบทดสอบฉบับ X และฉบับ Y โดยแบบทดสอบร่วม (ฉบับ U) นั้น จะต้องมิลักษณะคู่ขนานกันในโครงสร้างกับแบบทดสอบเทียบคะแนนฉบับ X และฉบับ Y ภายใต้เงื่อนไขนี้ ถ้าแบบทดสอบร่วม U รวมเข้าเป็นส่วนหนึ่งของแบบทดสอบเทียบคะแนนฉบับ X และฉบับ Y มี

ลักษณะเป็นแบบทดสอบร่วมภายใน (Internal Anchor Test) การเทียบคะแนนจากแบบทดสอบฉบับ Y ไปยังฉบับ X จะได้ค่าสัมประสิทธิ์ความชันและจุดตัดดังนี้

$$A_{xy} = (\sigma_x^2)(\sigma_y^2) \quad (11)$$

$$B_{xy} = \mu_x - A_{xy}\mu_y \quad (12)$$

โดยที่ $\mu_x = M_{xu} + b_{xu}(\mu_u - M_{u,})$

$$\sigma_x^2 = S_{xu}^2 + b_{xu}^2(\sigma_u^2 - S_{u,}^2)$$

$$\mu_y = M_{yu} + b_{yu}(\mu_u - M_{u,})$$

$$\sigma_y^2 = S_{yu}^2 + b_{yu}^2(\sigma_u^2 - S_{u,}^2)$$

เมื่อ $\mu_u = M_{ut}$ และ $\sigma_u^2 = S_{ut}^2$ และ $t = \alpha + \beta$

ถ้าคะแนนจากแบบทดสอบฉบับ X และ Y มีลักษณะการแจกแจงที่คล้ายกัน ความสัมพันธ์ระหว่างแบบทดสอบจะมีลักษณะเป็นเส้นตรง วิธีอีควิเปอร์เซ็นไทล์จะให้ผลการเทียบคะแนนเช่นเดียวกับวิธีเชิงเส้นตรง แต่ถ้ายการแจกแจงของคะแนนจากแบบทดสอบฉบับ X และ Y แตกต่างกันความสัมพันธ์ระหว่างแบบทดสอบจะมีลักษณะเป็นเส้นโค้ง วิธีอีควิเปอร์เซ็นไทล์จะให้ผลการเทียบที่แน่นอนกว่าวิธีเชิงเส้นตรง

2.2.2 วิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ (Equating in Item Response Theory) การเทียบคะแนนตามวิธีการนี้อาศัยรูปแบบความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้สอบข้อสอบใด ๆ ว่าเป็นขึ้นอยู่กับการความสามารถของผู้สอบซึ่งประมาณค่าได้จากโค้งลักษณะข้อสอบ (Item Characteristic Curve) ของแต่ละรูปแบบที่ใช้ ซึ่งอาจจะเป็นแบบหนึ่ง สอง หรือสามพารามิเตอร์ โดยไม่ขึ้นกับการแจกแจงคะแนนของกลุ่มตัวอย่าง (สุจินดา ผ่องอักษร. 2533 : 17 – 28; อ้างอิงจาก Lord and Novick. 1968) การเทียบคะแนนโดยใช้ทฤษฎีการตอบข้อสอบมีหลักการว่าคะแนนจากแบบทดสอบต่างฉบับที่วัดในสิ่งเดียวกันจะถือว่าเทียบเท่ากันได้ ถ้าผู้สอบทำแบบทดสอบต่างฉบับแล้วได้คะแนนจริง (True Scores) หรือความสามารถแท้ (Ability) เท่ากัน (สงบ ลักษณะ. 2525 : 22) การเทียบคะแนนตามทฤษฎีนี้แบ่งเป็น 2 แบบ ตามการเสนอคะแนนสมมูลที่ได้จากการเทียบคะแนน คือ

2.2.2.1 การเทียบคะแนนโดยใช้คะแนนจริง (True Scores Equating)

2.2.2.2 การเทียบคะแนนโดยใช้คะแนนสังเกต (Observed Score Equating) รูปแบบการเทียบคะแนนทั้งสองนี้มีข้อได้เปรียบกัน คือ วิธีเทียบคะแนนโดยใช้คะแนนจริงไม่สามารถอธิบายคะแนนที่อยู่ต่ำกว่าระดับการเดาได้ โดยจะให้ความหมายของคะแนนสมมูลเฉพาะคะแนนที่อยู่เหนือค่าเฉลี่ยของการเดา ซึ่งแม้จะเป็นการเทียบคะแนน โดยใช้คะแนนจริง แต่ก็ยังเป็นค่าประมาณที่ได้จากสูตรการคำนวณ ดังนั้น จึงยังคงมีความคลาดเคลื่อนอยู่ส่วนการเทียบ

คะแนนโดยใช้คะแนนสังเกตเป็นการเทียบคะแนนโดยประมาณ แต่สามารถอธิบายคะแนนสมมูลจากแบบทดสอบทั้ง 2 ฉบับ ได้ครอบคลุมพิสัยของคะแนนที่สังเกตได้ วิธีการทั้งสองนี้มีความสอดคล้องกันมากแต่การสรุปผล เพื่อนำไปใช้อ้างอิงต้องทำอย่างพิถีพิถัน อย่างไรก็ตามการเทียบคะแนนสังเกตนี้นำไปใช้ในการวิจัยน้อยมาก ซึ่งอาจเป็นเพราะว่าวิธีนี้มีความซับซ้อนและการลงทุนสูงกว่าวิธีการเทียบคะแนนโดยใช้คะแนนจริง (Lord, 1980 : 202 – 203; Lord and Wingersky, 1984 : 453)

โคเลน และเบรนนอน (Kolen and Brennan, 1995 : 269 - 271) ได้แสดงสถานการณ์การเทียบคะแนนที่เหมาะสมกับวิธีการเทียบคะแนนแต่ละวิธีตามตาราง 1 ดังนี้

ตาราง 1 การเปรียบเทียบสถานการณ์ที่เหมาะสมกับวิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์
วิธีเชิงเส้นตรงและวิธี IRT

ประเด็น	วิธีอีควิเปอร์เซ็นไทล์	วิธีเชิงเส้นตรง	วิธี IRT
1	แบบทดสอบที่นำมาเทียบ คะแนนมีลักษณะโครงสร้างเนื้อหาเดียวกัน (same specification)	แบบทดสอบที่นำมาเทียบ คะแนนมีลักษณะโครงสร้างเนื้อหาเดียวกัน (same specification)	แบบทดสอบที่นำมาเทียบ คะแนน มีลักษณะโครงสร้างเนื้อหาเดียวกัน (same specification)
2	กลุ่มตัวอย่างควรมีขนาดใหญ่	เหมาะกับกลุ่มตัวอย่างขนาดเล็ก	กลุ่มตัวอย่างควรมีขนาดใหญ่
3	แบบทดสอบแต่ละฉบับ สามารถมีความแตกต่างกันใน ด้านความยากได้มากกว่าวิธี เชิงเส้นตรง	แบบทดสอบแต่ละฉบับ ควรมีความยากคล้ายคลึงกัน	แบบทดสอบแต่ละฉบับ สามารถมีความแตกต่างกันใน ด้านความยากได้ มากกว่าวิธีเชิงเส้นตรง
4	กระบวนการวิเคราะห์ก่อน ข้างซับซ้อน ยากที่จะทำให้ คนทั่วไปเข้าใจได้	กระบวนการวิเคราะห์ก่อน ข้างง่าย เป็นที่ยอมรับของ คนทั่วไป	กระบวนการวิเคราะห์ก่อน ข้างซับซ้อนยากที่จะทำให้ คนทั่วไปเข้าใจได้
5	ผลการเทียบคะแนนมีความ ถูกต้องทุกช่วงคะแนนแปลงที่ มีความถี่	ผลการเทียบคะแนนมี ความถูกต้องเฉพาะช่วง คะแนนที่อยู่ใกล้ค่าเฉลี่ย	ผลการเทียบคะแนนมี ความถูกต้องทุกช่วง คะแนนแปลง

สรุปได้ว่าวิธีการเทียบคะแนนแต่ละวิธีมีความเหมาะสมกับสภาพการณ์ที่แตกต่างกันไป ในการวิจัยครั้งนี้ จึงศึกษาวิธีการเทียบคะแนนทั้งในทฤษฎีการทดสอบแบบมาตรฐานเดิม ซึ่งมีกระบวนการวิเคราะห์ที่ค่อนข้างง่ายกว่า และทฤษฎีการตอบข้อสอบ ซึ่งมีวิธีประมาณค่าพารามิเตอร์ข้อสอบคงที่มากกว่า โดยหน่วยตัวอย่างที่นำมาเทียบคะแนนครั้งนี้มีขนาดไม่ต่ำกว่า 1,000 คน เพื่อให้เกิดความพอเพียงต่อการประมาณค่าพารามิเตอร์ในโปรแกรมที่ใช้ในการวิเคราะห์ข้อมูลและแบบทดสอบเทียบคะแนนแต่ละฉบับสร้างจากโครงสร้างเนื้อหาเดียวกัน โดยจัดฉบับให้มีความยากใกล้เคียงกัน

3. การเทียบคะแนนตามทฤษฎีการตอบข้อสอบที่ข้อสอบตรวจให้คะแนนสองค่า

ลอร์ด (Lord, 1980 : 199) ได้กล่าวไว้ว่าข้อกำหนดที่สำคัญ 3 ประการ สำหรับการเทียบคะแนนหรือการเชื่อมต่อแบบทดสอบมิติเดียว 2 ฉบับ ซึ่งวัดในความสามารถเดียวกัน คือ ความเสมอภาค (equity) ความไม่แปรผัน (invariance) และ ความสมมาตร (Symmetry) นอกจากนี้ แบบทดสอบเทียบคะแนนฉบับต่าง ๆ ต้องมีเนื้อหาที่ไม่แตกต่างกัน

การเทียบคะแนนตามทฤษฎีการตอบข้อสอบ ประกอบด้วย 3 ขั้นตอนคือ

3.1 การเลือกแบบแผนการรวบรวมข้อมูล แบบแผนการรวบรวมข้อมูลจากแบบทดสอบต่างฉบับที่นำมาเทียบคะแนนจะมีแบบแผนต่าง ๆ ดังที่กล่าวไว้ในหน้า 28 – 31 ซึ่งการเลือกแบบแผนใด พิจารณาได้จากธรรมชาติของการทดสอบและกลุ่มผู้ตอบ และจะส่งผลต่อรูปแบบในการวิเคราะห์ค่าพารามิเตอร์ (Kolen and Brennan, 1995 : 167; Petersen and others, 1989 : 256 - 257) ดังนี้

แบบแผนกลุ่มสุ่มหรือกลุ่มสมมูล พารามิเตอร์สำหรับแบบทดสอบฉบับ X สามารถประมาณค่าแยกจากแบบทดสอบฉบับ Y โดยกำหนดมาตรวัดความสามารถให้เหมือนกัน (เช่น ค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1) ดังนั้นค่าประมาณพารามิเตอร์สำหรับแบบทดสอบ 2 ฉบับ จึงถือว่ามาตรวัดเดียวกันโดยปราศจากการแปลง เนื่องจากถือว่ากลุ่มมีความสมมูลโดยสุ่ม และความสามารถถูกกำหนดให้มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของมาตรวัดที่เหมือนกันทั้ง 2 กลุ่ม แต่มีบางเหตุผลที่ว่า แบบแผนการกำหนดมาตรวัดของแบบทดสอบสองฉบับมีความแตกต่างกัน ดังนั้นจึงนำค่าประมาณของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานมาใช้แทนที่ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความสามารถ ดังนั้นจึงสามารถใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของมาตรวัดของพารามิเตอร์ θ ในสมการ (19c) และ (20b) หน้า 39

แบบแผนกลุ่มเดียวที่มีความสมดุลและไม่สมดุล พารามิเตอร์ของผู้สอบทุกคนที่ทำแบบทดสอบทั้ง 2 ฉบับ จะถูกประมาณค่าพร้อมกัน ดังนั้นค่าประมาณพารามิเตอร์ถือว่ามีมาตรวัดเดียวกัน ถ้าพารามิเตอร์ของแบบทดสอบทั้ง 2 ฉบับถูกประมาณค่าแยกกันโดยกำหนดมาตรวัดให้เหมือนกัน ก็ถือว่าค่าประมาณพารามิเตอร์มีมาตรวัดเดียวกันตามเหตุผลเดียวกับแบบแผนกลุ่มสุ่มหรือกลุ่มสมดุล

แบบแผนกลุ่มสุ่มที่ใช้แบบทดสอบร่วม พารามิเตอร์สำหรับแบบทดสอบฉบับ X และ Y สามารถประมาณค่าในลักษณะเดียวกับแบบแผนกลุ่มสมดุล หรือสามารถประมาณค่าพร้อมกันได้ เพราะว่าผู้สอบแต่ละคนทำแบบทดสอบร่วม ข้อมูลของข้อสอบร่วมจะเชื่อม พารามิเตอร์ผู้สอบและข้อสอบ 2 ฉบับเข้าด้วยกัน ดังนั้นค่าประมาณความสามารถและข้อสอบอยู่ในมาตรวัดเดียวกัน และสามารถประมาณค่าพารามิเตอร์ของแบบทดสอบแยกกัน แล้วใช้สารสนเทศของข้อสอบร่วมแปลงพารามิเตอร์ความสามารถให้อยู่ในมาตรวัดเดียวกัน

แบบแผนกลุ่มไม่เท่าเทียมกันที่ใช้แบบทดสอบร่วม พารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถของแบบทดสอบฉบับ X และ Y ถูกประมาณค่าแยกกันเมื่อมีการใช้แบบทดสอบแต่ละฉบับ เพราะผู้สอบในแบบทดสอบฉบับ X ไม่สมดุลกับผู้สอบในแบบทดสอบฉบับ Y ดังนั้นค่าประมาณพารามิเตอร์สองครั้งไม่ได้อยู่บนมาตรวัดเดียวกัน จึงต้องใช้ค่าประมาณพารามิเตอร์ข้อสอบร่วมในการประมาณค่าสัมประสิทธิ์การเทียบคะแนนสำหรับการแปลงมาตรวัด

3.2 การปรับค่าพารามิเตอร์ของแบบทดสอบต่างฉบับให้อยู่ในมาตรวัดเดียวกัน

การปรับค่าพารามิเตอร์ของโมเดล IRT เพื่อแปลงหรือเทียบคะแนนมาตรวัดของแบบทดสอบฉบับอื่นไปยังฉบับพื้นฐาน สามารถจัดได้ 2 ประเภท ตามเทคนิคการวิเคราะห์ข้อมูล คือ การปรับค่าพารามิเตอร์พร้อมกัน และการปรับค่าพารามิเตอร์แยกกัน ซึ่งมีรายละเอียดดังต่อไปนี้

3.2.1 การปรับค่าพารามิเตอร์พร้อมกัน (Simultaneous or Concurrent Calibration)

เมื่อปรับค่าพารามิเตอร์ของแบบทดสอบหลายฉบับพร้อม ๆ กัน โดยไม่ทราบค่าประมาณของพารามิเตอร์ข้อสอบร่วม ค่าประมาณทั้งหมดของข้อสอบทั้งที่ไม่ได้ร่วมและร่วมจะเชื่อมต่อกันโดยอัตโนมัติไปสู่มาตรวัดร่วม (Bastari, 2000 : 16; citing Mislevy and Bock, 1990) เช่น ประมาณค่าพารามิเตอร์แบบทดสอบฉบับ X และ Y พร้อมกัน โดยการ run 1 ครั้ง ในโปรแกรม LOGIST สามารถทำได้โดยใช้ข้อสอบสำหรับฉบับ X และฉบับ Y จากผู้สอบทั้งสองกลุ่ม และข้อสอบกลุ่มที่ผู้สอบกลุ่มหนึ่งไม่ได้ทำ จะใช้คำว่า "not reach" เมื่อการประมาณค่าถูกทำในลักษณะเช่นนี้ ค่าประมาณที่ได้จะอยู่ในมาตรวัดเดียวกัน

นอกจากนี้ยังมีอีกวิธีการหนึ่งเรียกว่าวิธีการกำหนดค่าก่อนปรับพารามิเตอร์ (Fixed Precalibration Item Parameter) (Bastari, 2000 : 16; citing Lord, 1980) ซึ่งเป็นวิธีการที่อยู่ในกลุ่ม การปรับค่าพารามิเตอร์พร้อมกัน สองวิธีนี้ อาจแตกต่างกันในแนวทางการประมาณค่าพารามิเตอร์ข้อสอบในการดำเนินการค่าประมาณพารามิเตอร์ข้อสอบรวมถูกกำหนดค่าระหว่างการปรับค่าแบบทดสอบฉบับอื่น ๆ ใหม่ พารามิเตอร์ที่ถูกกำหนดค่าเป็นค่าประมาณที่ถูกจัดการให้เป็นพารามิเตอร์ และค่าเหล่านี้ได้รับจากข้อสอบที่รู้ค่าแล้ว อาจจะได้จากธนาคารข้อสอบ เมื่อแบบทดสอบหนึ่งฉบับหรือมากกว่าถูกเชื่อมต่อหรือรวมค่าประมาณพารามิเตอร์ข้อสอบสำหรับข้อสอบใหม่ ถูกนำมาสู่มาตรวัดเดียวกัน วิธีการนี้สมมติว่า ค่าประมาณของพารามิเตอร์ที่กำหนดมาเหมาะสมกับข้อมูล และโมเดลเป็นอย่างดี และถูกปรับในแนวทางเดียวกันกับฉบับอื่น

วิธีการกำหนดค่าพารามิเตอร์ และวิธีการปรับค่าพารามิเตอร์พร้อมกัน สามารถรวมกันได้ เช่น เมื่อรู้ค่าประมาณพารามิเตอร์ของข้อสอบรวม ให้กำหนดค่าดังกล่าวในแบบทดสอบฉบับใหม่ ซึ่งอาจมีหนึ่งฉบับหรือมากกว่าก็ได้ ดังนั้นแบบทดสอบสองฉบับหรือหลายฉบับจะได้รับการปรับค่าพารามิเตอร์พร้อมกัน ในกรณีนี้ เป็นการประยุกต์ใช้วิธีการทั้งสองพร้อมกัน การรวมหรือแยกวิธีทั้งสองขึ้นอยู่กับค่าประมาณพารามิเตอร์ข้อสอบระหว่างขั้นตอนการปรับค่าพารามิเตอร์

3.2.2 การปรับค่าพารามิเตอร์แยกกัน (Separate Calibration)

เมื่อข้อมูลการตอบจากแบบทดสอบ 2 ฉบับ หรือมากกว่าก็ตาม ได้จากกลุ่มผู้สอบที่แตกต่างกัน จะได้รับการปรับค่าพารามิเตอร์แยกกันจึงจำเป็นต้องเชื่อมต่อมาตรวัดโดยใช้ข้อสนเทศของข้อสอบรวม เพื่อแปลงค่าพารามิเตอร์ความสามารถเชิงเส้น ซึ่งมาตรวัดความสามารถสำหรับกลุ่มพื้นฐานยังไม่มีเปลี่ยนแปลง จึงไม่มีความจำเป็นต้องสร้างมาตรวัดใหม่ ในทางตรงกันข้ามต้องแปลงมาตรวัดความสามารถของกลุ่มเทียบคะแนนไปสู่มาตรวัดที่กำหนดโดยกลุ่มพื้นฐาน ความน่าจะเป็นของการตอบถูกสำหรับข้อสอบข้อใดได้รับจากมาตรวัดค่าความสามารถใหม่ของแต่ละบุคคลสำหรับกลุ่มเทียบคะแนนยังไม่เปลี่ยนแปลง วิธีการประมาณค่าสัมประสิทธิ์การเทียบคะแนน ได้แก่ ความชัน (A) และจุดตัด (B) มีอยู่หลายวิธีกลุ่มแรกเป็นวิธีการที่อยู่บนพื้นฐานโมเมนต์และกลุ่มที่สองเป็นวิธีฟังก์ชันการตอบหรือวิธีโค้งลักษณะ (Response Function or Characteristic Curve Method) โดยพื้นฐานทุก ๆ วิธีพยายามหาสมการเชิงเส้นที่ดีที่สุดเพื่อแปลงมาตรวัดจากแบบทดสอบฉบับหนึ่งไปสู่ฉบับอื่น หรือกล่าวอีกอย่างหนึ่งว่า การประมาณค่าความชันและจุดตัด ค่าประมาณพารามิเตอร์ความสามารถ และข้อสอบได้รับการแปลงจากมาตรวัดของแบบทดสอบฉบับหนึ่งให้มีประสิทธิภาพมากที่สุดไปสู่มาตรวัดที่ใช้กับฉบับอื่น ๆ (Marco, 1977; Lord, 1980)

สมการแปลงหรือสมการเชิงเส้น (Kolen and Brennan. 1995 : 163 - 166) กำหนดให้มาตรวัด I และมาตรวัด J เป็น IRT 3 พารามิเตอร์ ดังนั้นค่า θ สำหรับ 2 มาตรวัดใดเกี่ยวข้องกัน ดังนี้

$$\theta_{ji} = A\theta_{ii} + B \quad (13)$$

เมื่อ A และ B เป็นค่าคงที่ในสมการเชิงเส้น และ θ_{ji} และ θ_{ii} คือค่าของ θ สำหรับคนที่ i บนมาตรวัด J และมาตรวัด I พารามิเตอร์ข้อสอบสำหรับ 2 มาตรวัดที่สัมพันธ์กันมีดังนี้

$$a_{ji} = \frac{a_{ij}}{A} \quad (14)$$

$$b_{ji} = Ab_{ij} + B \quad (15)$$

และ

$$C_{ji} = C_{ij} \quad (16)$$

เมื่อ a_{ji} , b_{ji} และ c_{ji} คือพารามิเตอร์ข้อสอบ j บนมาตรวัด J และ a_{ij} และ b_{ij} และ c_{ij} คือพารามิเตอร์ข้อสอบ j บนมาตรวัด I พารามิเตอร์การเดา เป็นอิสระจากการแปลงมาตรวัด ดังที่แสดงไว้ในสมการ (16)

สำหรับผู้สอบคนที่ i และ i^* หรือข้อสอบ 2 ข้อใด ๆ j และ j^* A และ B ในสมการ (13) – (16) สามารถแสดงค่าได้ในสมการ (17) และ (18)

$$A = \frac{\theta_{ji} - \theta_{ji^*}}{\theta_{ii} - \theta_{ii^*}} = \frac{b_{ji} - b_{ji^*}}{b_{ij} - b_{ij^*}} = \frac{a_{ij}}{a_{ji}} \quad (17)$$

และ

$$B = b_{ji} - Ab_{ij} = \theta_{ji} - A\theta_{ii} \quad (18)$$

ขณะเดียวกันสามารถแสดงค่าของ A และ B ในสมการ (17) – (18) ในรูปของกลุ่มข้อสอบหรือกลุ่มคนได้ดังนี้

$$A = \frac{\sigma(b_j)}{\sigma(b_i)} \quad (19a)$$

$$= \frac{\mu(a_i)}{\mu(a_j)} \quad (19b)$$

$$= \frac{\sigma(\theta_j)}{\sigma(\theta_i)} \quad (19c)$$

$$B = \mu(b_j) - A\mu(b_i) \quad (20a)$$

และ

$$= \mu(\theta_j) - A\mu(\theta_i) \quad (20b)$$

$\mu(b_j)$, $\mu(b_i)$, $\mu(a_i)$ และ $\mu(a_j)$ เป็นค่าเฉลี่ยของพารามิเตอร์ข้อสอบมากกว่า 1 ข้อทั้งในมาตรวัด I และ J

$\sigma(b_j)$ และ $\sigma(b_i)$ คือส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ข้อสอบที่มากกว่า 2 ข้อ ทั้งในมาตรวัด I และ J

$\mu(\theta_j)$ และ $\mu(\theta_i)$ เป็นค่าเฉลี่ยของกลุ่มผู้สอบมากกว่า 1 คน ทั้งในมาตรวัด I และ J

$\sigma(\theta_j)$ และ $\sigma(\theta_i)$ เป็นส่วนเบี่ยงเบนมาตรฐานของกลุ่มผู้สอบมากกว่า 2 คน ทั้งในมาตรวัด I และ J

ในการเทียบคะแนนกับกลุ่มไม่เท่าเทียมกันการประมาณค่าข้อสอบรวมจะได้จากผู้เข้าสอบทั้ง 2 กลุ่ม ค่าประมาณของพารามิเตอร์ในข้อสอบรวมสามารถใช้เพื่อกำหนดมาตรวัดให้คงที่ โดยการแทนที่พารามิเตอร์สำหรับประมาณค่าในสมการข้างต้น

วิธีการประมาณค่าสัมประสิทธิ์การเทียบคะแนน (Kolen and Brennan. 1995 : 169 – 174) มีวิธีการดังนี้

1) วิธีการประมาณค่า โดยใช้พารามิเตอร์ข้อสอบไม่พร้อมกันทุกค่า (วิธี mean และ sigma และ วิธี mean และ mean)

การแปลงมาตรวัดในแบบแผนการใช้แบบทดสอบรวมของกลุ่มที่ไม่เท่าเทียมกันคือการแทนค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ข้อสอบรวมในสมการ (19) และ (20) มาร์โค (Macro. 1977 : 140 – 141) ได้ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าประมาณพารามิเตอร์ความยาก (b) จากข้อสอบรวม เพื่อแทนค่าพารามิเตอร์ในสมการ (19a) และ (20a) เรียกว่า วิธี mean และ sigma ส่วน ลอยด์และฮูเวอร์ (Kolen and Brennan. 1995 : 168; citing Loyd and Hoover. 1980) ได้ใช้ค่าเฉลี่ยของค่าประมาณพารามิเตอร์อำนาจจำแนก (a) ของข้อสอบรวม แทนค่าพารามิเตอร์ความยากในสมการ (19b) เพื่อประมาณค่า A และค่าเฉลี่ยของค่าประมาณพารามิเตอร์ความยากของข้อสอบรวมใช้แทนค่าพารามิเตอร์ในสมการ (20a) เพื่อประมาณค่า B เรียกว่า วิธี mean และ mean นำค่า A และ B แทนในสมการ (13)-(16) เพื่อประมาณค่าพารามิเตอร์ใหม่

วิธี mean และ sigma และวิธี mean และ mean จะให้ผลที่แตกต่างกัน มีเหตุผลที่ว่า mean และ sigma อาจจะให้ค่าที่ดีกว่าวิธี mean และ mean คือ ค่าประมาณของพารามิเตอร์ความยาก (b) มีความคงที่มากกว่าค่าประมาณของพารามิเตอร์อำนาจจำแนก แต่เบเกอร์ และ อัล-คานี (Baker and Al-Karni. 1991 : 149 - 150) ได้ชี้ให้เห็นว่าวิธี mean และ mean อาจจะดีกว่า เพราะค่าเฉลี่ยมีความคงที่มากกว่าส่วนเบี่ยงเบนมาตรฐานและวิธี mean และ mean ถูกใช้เฉพาะค่าเฉลี่ยเท่านั้น ซึ่งงานวิจัยทางด้านนี้ยังขาดอยู่มาก (Kolen and Brennan. 1995 : 168) มิสเรวีและบอค (Kolen and Brennan. 1995 : 169; Citing Mislevy and Bock. 1990) ได้แนะนำให้ใช้ค่าเฉลี่ยของพารามิเตอร์ความยากและค่าเฉลี่ยเรขาคณิตของพารามิเตอร์อำนาจจำแนกและสตอกกิงและลอร์ด (Stocking and Lord. 1983) ได้เสนอกระบวนการสำหรับใช้ค่าประมาณของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากที่เข้มงวด เรียกว่าวิธี robust mean และ sigma ซึ่งให้ผลไม่น่าพอใจเมื่อเทียบกับวิธีโค้งลักษณะแบบทดสอบ

2) วิธีการแปลงด้วยโค้งลักษณะ หรือฟังก์ชันการตอบ (Characteristic Curve or Response Function Transformation Methods)

ค่าประมาณพารามิเตอร์ความยาก อำนาจจำแนกและการเดา เมื่อประมาณค่ารวมกันให้โค้งลักษณะข้อสอบที่เหมือนกันระหว่างช่วงความสามารถ ณ คะแนนสูงสุดของผู้สอบ เช่น การประมาณค่าข้อสอบด้วยค่าประมาณพารามิเตอร์ความยากที่แตกต่างกัน อาจให้โค้งลักษณะข้อสอบที่เหมือนกัน ในกรณีเช่นนี้ ถ้าใช้วิธี mean และ sigma อาจจะได้รับอิทธิพลอย่างเด่นชัดจากความแตกต่างระหว่างค่าประมาณพารามิเตอร์ความยาก แม้ว่าโค้งลักษณะข้อสอบสำหรับ 2 การประมาณค่าของข้อสอบใด ๆ จะเหมือนกัน ปัญหาที่เกิดขึ้นเพราะวิธีการแปลงมาตรฐานข้างต้นไม่ได้พิจารณาค่าประมาณพารามิเตอร์ทุกค่าอย่างพร้อม ๆ กัน เฮบารา (Kolen and Brennan. 1995 : 169; citing Haebara. 1980) เสนอวิธีการที่พิจารณาค่าพารามิเตอร์ข้อสอบทุกค่าพร้อมกัน และสตอกกิงและลอร์ด (Stocking and Lord. 1983 : 201 - 210) ได้เสนอวิธีที่คล้ายกับเฮบารา และเรียกวิธีการทั้งคู่นี้ว่าโค้งลักษณะหรือฟังก์ชันการตอบ การพัฒนาวิธีการดังกล่าวอาศัยแนวคิดการแปลงในสมการ (13) – (16) ในการแปลงมาตรฐานความสามารถ I และ J จะได้

$$p_{ij}(\theta_{ji}, a_{ji}, b_{ji}, c_{ji}) = p_{ij}\left(A\theta_{ji} + B; \frac{a_{ij}}{A}, Ab_{ij} + B, C_{ij}\right) \quad (21)$$

สำหรับผู้สอบ i และข้อสอบ j สมการ (21) กำหนดความน่าจะเป็นที่ผู้สอบมีความสามารถต่าง ๆ จะตอบข้อสอบอย่างถูกต้อง โดยไม่คำนึงว่าเป็นมาตรวัดเดียวกันที่ใช้ในการรายงานคะแนน ถ้าใช้ค่าประมาณที่แทนพารามิเตอร์ในสมการ (21) จะทำให้ไม่สามารถยืนยันความเท่าเทียมกันของข้อสอบทุก ๆ ข้อ และผู้สอบทุก ๆ คนได้ ดังนั้น A และ B อาจขาดความเท่าเทียมกันซึ่งแก้ไขโดยใช้วิธีโค้งลักษณะ

2.1) แนวทางของเฮบารา เป็นการแสดงให้เห็นฟังก์ชันความแตกต่างระหว่างโค้งลักษณะข้อสอบเป็นผลรวมของความแตกต่างกำลังสองระหว่างโค้งลักษณะข้อสอบแต่ละข้อของผู้สอบที่มีความสามารถเฉพาะ θ_i ผลรวมของความแตกต่างกำลังสองคือ

$$Hdiff(\theta_i) = \sum_{j:v} \left[p_{ij}(\theta_{ji}, a_{ji}, b_{ji}, c_{ji}) - p_{ij}\left(\theta_{ji}, \frac{a_{ij}}{A}, Ab_{ij} + B, c_{ij}\right) \right]^2 \quad (22)$$

ผลรวมของข้อสอบรวม $(j:v)$ ในสมการนี้เป็นการคำนวณ ความแตกต่างระหว่างโค้งลักษณะข้อสอบแต่ละข้อในมาตรวัดทั้งสองแล้วจึงยกกำลังก่อนจึงหาผลรวม $Hdiff$ ถูกสะสมระหว่างผู้สอบ การประมาณค่า A และ B กระทำโดยการทำให้ฟังก์ชันนี้มีค่าต่ำสุดตามเกณฑ์ดังนี้

$$Hcrit = \sum_i Hdiff(\theta_i) \quad (23)$$

ผลรวมในสมการ (23) เป็นการหาผลรวมจากผู้สอบ

2.2) แนวทางสตอกกิงและลอว์รี ตรงกันข้ามกับแนวทางของเฮบารา สตอกกิงและลอว์รี ใช้ผลรวมระหว่างข้อของความแตกต่างกำลังสองคือ

$$SLdiff(\theta_i) = \left[\sum_{j:v} p_{ij}(\theta_{ji}, a_{ji}, b_{ji}, c_{ji}) - \sum_{j:v} p_{ij}\left(\theta_{ji}, \frac{a_{ij}}{A}, Ab_{ij} + B, c_{ij}\right) \right]^2 \quad (24)$$

แนวทางของสตอกกิงและลอว์รี ผลรวมได้จากการรวมข้อสอบทุก ๆ ข้อสำหรับค่าประมาณพารามิเตอร์ในแบบทดสอบแต่ละฉบับก่อนที่จะยกกำลังในฟังก์ชัน IRT

$$\tau(\theta_i) = \sum_j p_{ij}(\theta_i) \quad (25)$$

คือโค้งลักษณะของแบบทดสอบ (test characteristic curve) ดังนั้น $SLdiff(\theta_i)$ คือความแตกต่างกำลังสองระหว่างโค้งลักษณะแบบทดสอบ สำหรับ θ_i ในทางตรงกันข้าม $Hdiff(\theta_i)$ เป็นผลรวมของความแตกต่างกำลังสองระหว่างโค้งลักษณะข้อสอบสำหรับ θ_i $SLdiff$ ถูกสะสมระหว่างผู้สอบ การประมาณค่า A และ B กระทำโดยทำให้ฟังก์ชันนี้มีค่าต่ำสุดตามเกณฑ์ดังนี้

$$SLcrit = \sum_i SLdiff(\theta_i) \quad (26)$$

ผลรวมในสมการ (26) เป็นผลรวมของผู้สอบ การหาค่า A และ B ในสมการ (23) และ (26) คำนวณโดยการทำซ้ำ

ได้มีการศึกษาเพื่อเปรียบเทียบวิธีการของเฮบาราและสตอกกิงและลอร์ด พบว่า วิธีการทั้งสองให้ผลเหมือนกัน (Kolen Brennan. 1995 : 173; citing Way and Tang. 1991) แต่ในทางทฤษฎีวิธี Hcrit อาจจะขัดแย้งกับวิธี SLcrit ด้วยเหตุผลที่ว่า 1) Hdif (θ_i) มีค่าเป็น 0 เฉพาะเมื่อโค้งลักษณะข้อสอบที่ได้เหมือนกัน ณ ระดับ θ_i ขณะที่ SLdif (θ_i) ให้ค่าเป็น 0 แม้ว่าโค้งลักษณะข้อสอบแตกต่างกัน แสดงให้เห็นว่า Hdif (θ_i) อาจจะมีค่าเข้มงวด (stringent) มากกว่า SLdif (θ_i) 2) ดังนั้น วิธีการ SLcrit อาจจะมีเหตุผลทางทฤษฎีมากกว่า เพราะมีจุดเน้นที่ความแตกต่างระหว่างโค้งลักษณะของแบบทดสอบ

ข้อจำกัดของวิธีโค้งลักษณะคือ ไม่สามารถอธิบายความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ข้อสอบได้อย่างชัดเจน (Divgi. 1985) การอธิบายความ คลาดเคลื่อนที่ชัดเจนไม่ได้อาจจะไม่ใช่ปัญหามากนักถ้ามีการใช้กลุ่มตัวอย่างขนาดใหญ่ และโค้งลักษณะข้อสอบถูกประมาณค่าอย่างดี แต่ถ้ามีการใช้กลุ่มตัวอย่างขนาดใหญ่สำหรับประมาณค่าพารามิเตอร์ในแบบทดสอบฉบับหนึ่งมากกว่าอีกฉบับหนึ่ง นอกจากจะทำให้เกิดปัญหาเกี่ยวกับความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์แล้ว ยังจะนำไปสู่ปัญหาการประมาณค่าสัมประสิทธิ์การเทียบคะแนน คือความชัน (A) และจุดตัด (B) และการประมาณค่าความสัมพันธ์ของการเทียบคะแนนด้วย ดังนั้นในการวิจัยเชิงประจักษ์จะต้องคำนึงถึงประเด็นดังกล่าวด้วย

สรุปได้ว่าวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ จำแนกตามการวิเคราะห์ได้เป็นสองแนวทางใหญ่ ๆ การจะเลือกใช้การเทียบคะแนนวิธีใดขึ้นอยู่กับแบบแผนในการรวบรวมข้อมูล หากใช้การเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกัน จะทำให้พารามิเตอร์ของแบบทดสอบทั้งสองฉบับอยู่ในมาตรวัดเดียวกัน แต่ถ้าปรับค่าพารามิเตอร์แยกกันต้องอาศัยการแปลงเชิงเส้นเพื่อให้พารามิเตอร์ความสามารถอยู่ในมาตรวัดเดียวกัน ซึ่งการประมาณค่าสัมประสิทธิ์การเทียบคะแนนมีหลากหลายวิธีขึ้นอยู่กับว่าจะนำค่าพารามิเตอร์ทุกค่ามาใช้พร้อมกันหรือใช้ค่าพารามิเตอร์บางค่า

ได้มีงานวิจัยที่ศึกษาเปรียบเทียบวิธีการเทียบคะแนนในกลุ่มวิธีการแปลงเชิงเส้นแบบที่ใช้พารามิเตอร์ไม่พร้อมกันคือวิธี mean และ sigma หรือ วิธี mean และ mean กับวิธีโค้งลักษณะ (Baker and Al-Karni. 1991 : 161; Kolen and Brennan. 1995 : 174; citing Way and Tang. 1991) ผลสรุปโดยทั่ว ๆ ไปคือวิธีโค้งคุณลักษณะให้ผลที่แม่นยำกว่า อย่างไรก็ตาม เบเกอร์ และ อัล-คานี ได้พบว่า วิธีของสตอกกิงและลอร์ด และวิธี mean และ mean ให้ผลที่เหมือนกันมากเกือบทุกกรณี แต่เมื่อการประมาณค่าพารามิเตอร์ของโมเดล IRT มีปัญหากระบวนการของสตอกกิงและลอร์ดจะให้ผลที่แม่นยำกว่า

นอกจากนี้ยังมีการศึกษาเปรียบเทียบวิธีการเทียบคะแนนด้วยการแปลงเชิงเส้นกับวิธีการเทียบคะแนนที่เกิดจากการปรับค่าพารามิเตอร์พร้อมกัน ส่วนใหญ่พบว่าโดยภาพรวมวิธีการปรับค่าพารามิเตอร์พร้อมกันเป็นวิธีหนึ่งที่ดีที่สุด แต่เมื่อจำนวนของข้อสอบรวมมีขนาดเล็ก วิธีโค้งลักษณะแบบทดสอบจะให้ผลที่ดีกว่าวิธีการปรับค่าพารามิเตอร์พร้อมกัน และให้ผลที่คล้ายคลึงกันเมื่อข้อสอบรวมมีจำนวนมากขึ้น แต่วิธีการปรับค่าพารามิเตอร์พร้อมกันอาจให้ผลที่ไม่ดีนัก เมื่อผู้สอบมีความแตกต่างกันมาก (Kim and Cohen. 1998a : 141 - 142) และวิธีการเทียบคะแนนโดยใช้โค้งลักษณะจะให้ผลที่คล้ายคลึงกันกับวิธี FPIP เมื่อข้อสอบมีความยากปานกลาง และวิธี FPIP ให้ผลดีในข้อสอบที่ง่าย (Li, Lissitz and Yang. 1999 : 5; citing Li, Griffith and Tam. 1997) ซึ่งปัจจุบันก็ยังคงมีการศึกษาวิธีการเทียบคะแนนทั้งสองกลุ่มในบริบทต่าง ๆ อยู่

3.3 การเทียบคะแนนผู้สอบ

ต้องพิจารณาว่าจะคะแนนจากแบบทดสอบจะรายงานในรูปคะแนนความสามารถหรือคะแนนจริงหรือคะแนนสังเกต การเทียบคะแนนมีวิธีการแตกต่างกัน ดังนี้

3.3.1 ในกรณีที่การรายงานผลคะแนนจากแบบทดสอบ ในรูปคะแนนความสามารถการเทียบคะแนนสามารถทำได้ โดยใช้ค่าประมาณความสามารถที่ได้จากแบบทดสอบต่างฉบับที่ผ่านการปรับแล้ว

3.3.2 ในกรณีที่การรายงานผลทำในรูปคะแนนจริง จะต้องมีการประมาณคะแนนจริงจากแบบทดสอบแต่ละฉบับโดยใช้ค่าประมาณพารามิเตอร์ที่ได้จากการแปลงในขั้นตอนที่ 2 คะแนนจริงจะถูกประมาณค่าหลาย ๆ ค่า เพื่อนำไปใช้ สร้างกราฟหรือตาราง เพื่อเทียบคะแนนจริงที่ตรงกันจากแบบทดสอบแต่ละฉบับ

3.3.3 ในกรณีที่การรายงานคะแนนทำในรูปคะแนนสังเกตจะต้องทำการแจกแจงคะแนนสังเกตอย่างมีเงื่อนไขตามทฤษฎี สำหรับความสามารถที่ตรงกันของกลุ่มผู้สอบเทียบคะแนนที่สร้างขึ้นด้วยวิธีเทียบคะแนนอิกวิเปอร์เซ็นไทล์ แล้วทำการเทียบคะแนนสังเกตที่แท้จริงจากตารางหรือกราฟของการเปรียบเทียบ

ในการศึกษารั้งนี้ ผู้วิจัยเลือกแบบแผนการใช้แบบทดสอบร่วมภายในกับกลุ่มผู้สอบ ซึ่งเป็นนักเรียนในระดับชั้นเดียวกัน เพื่อศึกษาการเทียบคะแนนในข้อสอบที่ตรวจให้คะแนนหลายค่าอย่างครอบคลุม จึงทำการเทียบคะแนนในวิธีการทั้งสองแนวทาง คือ การปรับค่าพารามิเตอร์พร้อมกัน และการปรับค่าพารามิเตอร์แยกกัน โดยในการปรับค่าพารามิเตอร์แยกกัน ได้เลือกวิธีใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (วิธี mean และ sigma) และวิธีโค้งลักษณะแบบทดสอบในการประมาณค่าสัมประสิทธิ์การเทียบคะแนน

4. วิธีการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า

จากการศึกษางานวิจัยเกี่ยวกับการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า (Baker. 1992 : 87 – 96; 1993 : 239 – 251; 1997: 157 – 172; Huynh and Ferrara. 1994 : 125 – 141; Kim and Cohen. 1995 : 167 – 176; Hennings, Hirsch and Zhang. 1996 : 1 – 13; Cohen and Kim. 1998 : 116 – 130; Li, Lissitz and Yang. 1999 : 1 – 31; Bastari 2000 : 1 – 78; Lee and others. 2001 : 357 – 372) สรุปวิธีการเทียบคะแนนในแต่ละโมเดลที่ข้อสอบตรวจให้คะแนนหลายค่าดังต่อไปนี้

4.1 โมเดลเกรด เรสพอนส์ (Grade Response Model : GRM)

มีการเทียบคะแนนในโมเดลนี้หลายวิธีคือ

4.1.1 วิธีที่ใช้พารามิเตอร์ไม่พร้อมกันทุกค่าในการประมาณค่าสัมประสิทธิ์การเทียบคะแนน

โคเฮน และคิม (Cohen and Kim. 1998 : 117 – 121) ได้ศึกษาการเทียบคะแนนโมเดลเกรด เรสพอนส์ ในกลุ่มวิธีการนี้ 3 วิธี คือ

4.1.1.1 วิธีการใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (วิธี mean และ sigma : MS)

มาร์โค (Macro. 1977 : 140 – 141) ได้เสนอวิธีการนี้โดยสมมุติให้ค่าประมาณพารามิเตอร์ตำแหน่ง (location) จากกลุ่ม 1 คือ b_{jk1} กลุ่ม 2 คือ b_{jk2} โดยที่ $j = 1, \dots, n_c$; $k = 1, \dots, m_j - 1$ และ n_c เป็นข้อสอบร่วมในการเชื่อมต่อ ความสัมพันธ์เชิงเส้น ของพารามิเตอร์ข้อสอบมีดังนี้

$$b_{jk1} = Ab_{jk2} + B \quad (27)$$

และ

$$a_{j1} = a_{j2}/A \quad (28)$$

สัมประสิทธิ์การเชื่อมต่อของวิธี MS คือ A และ B ได้จากความสัมพันธ์ดังนี้

$$\bar{b}_1 = A\bar{b}_2 + B \quad (29)$$

เมื่อ $\bar{b}_1$ และ $\bar{b}_2$ คือ ค่าเฉลี่ยของ b_{jk1} และ b_{jk2} ตามลำดับ

และ

$$S(b_1) = AS(b_2) \quad (30)$$

เมื่อ $S(b_1)$ และ $S(b_2)$ คือ ส่วนเบี่ยงเบนมาตรฐานของ b_{jk1} และ b_{jk2} ตามลำดับ

ดังนั้น A และ B ได้จาก

$$A = S(b_1)/S(b_2) \quad (31)$$

และ

$$B = \bar{b}_1 - A\bar{b}_2 \quad (32)$$

4.1.1.2 วิธีการใช้ค่าเฉลี่ยของพารามิเตอร์ความยากและอำนาจจำแนก (วิธี mean และ mean : LH)

ลอยด์และฮูเวอร์ (Cohen and Kim. 1998 : 118 – 119; citing Loyd & Hoover. 1980) ได้เสนอให้ประมาณค่า A โดยใช้อัตราส่วนของค่าเฉลี่ยอำนาจจำแนกที่ได้จากการประมาณค่าแบบทดสอบ 2 ฉบับ วิธีการนี้ใช้กับโมเดลราสช์มาแล้ว ต่อมาเบเกอร์ และ อัล-คานี (Baker & Al-Karni 1991 : 161) ได้ใช้วิธี LH เพื่อเชื่อมต่อการวัดในโมเดล IRT โลจิสติก 3 พารามิเตอร์ ความสัมพันธ์เชิงเส้นของพารามิเตอร์อำนาจจำแนก คือ

$$a_{1j} = a_{2j} / A$$

A สำหรับวิธี LH ในโมเดล GRM จะได้จาก

$$A = \bar{a}_2 / \bar{a}_1 \quad (33)$$

และ

$$B = \bar{b}_1 - A\bar{b}_2 \quad (34)$$

เมื่อ $\bar{a}_1$ และ $\bar{a}_2$ คือ ค่าเฉลี่ยของ a_{1j} และ a_{2j} ตามลำดับ

ข้อจำกัดของวิธี MS และ LH คือถ้าประมาณค่าความยากของข้อสอบได้ไม่ดี อาจทำให้ค่า A และ B มีค่าเป็นลบ

4.1.1.3 วิธีการใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากที่กำหนดน้ำหนัก (วิธี Weighted mean และ sigma : WMS)

ลินน์และคณะ (Cohen and Kim. 1998 : 119 – 120; citing Linn, et al. 1980, 1981) ได้แก้ไขปัญหที่พบในวิธี MS และ LH โดยปรับปรุงกระบวนการ MS เพื่อรวมน้ำหนักของค่าประมาณความยากของข้อสอบ โดยอินเวอร์สของความคลาดเคลื่อนมาตรฐานกำลังสองมีขนาดใหญ่กว่า ต่อมาสตอกกิงและลอร์ด (Stocking and Lord. 1983 : 203, 208 - 209) กำหนดน้ำหนักโดยใช้ผลรวมของน้ำหนักระหว่างข้อสอบทุกข้อ สำหรับโมเดล GRM กำหนดน้ำหนักสำหรับค่าประมาณพารามิเตอร์ตำแหน่งของข้อสอบ j ลำดับชั้น k คือ w_{jk} ดังนี้

$$W_{jk} = \frac{\{\max[SE(b_{jk1}), SE(b_{jk2})]\}^{-2}}{\sum_{j=1}^{n_c} \sum_{k=1}^{m_j-1} \{\max[SE(b_{jk1}), SE(b_{jk2})]\}^{-2}} \quad (35)$$

ค่าประมาณพารามิเตอร์ตำแหน่งที่กำหนดน้ำหนัก คือ

$$b_{jk1}^w = W_{jk} b_{jk1} \quad (36)$$

และ

$$b_{jk2}^w = W_{jk} b_{jk2} \quad (37)$$

แล้ว

$$\bar{b}_1^w = A\bar{b}_2^w + B \quad (38)$$

และ

$$S(b_1^w) = AS(b_2^w) \quad (39)$$

เมื่อ $\bar{b}_1^w$ และ $S(b_1^w)$ คือค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่กำหนดน้ำหนักค่า ดังนี้

$$\bar{b}_1^w = \frac{\sum_{j=1}^{n_c} \sum_{k=1}^{m_j-1} w_{jk} b_{jk1}}{\sum_{j=1}^{n_c} \sum_{k=1}^{m_j-1} w_{jk}} \quad (40)$$

และ

$$S(b_1^w) = \sqrt{\frac{\sum_{j=1}^{n_c} \sum_{k=1}^{m_j-1} (w_{jk} b_{jk1} - \bar{b}_1^w)^2}{\sum_{j=1}^{n_c} \sum_{k=1}^{m_j-1} w_{jk}}} \quad (41)$$

$\bar{b}_2^w$ และ $S(b_2^w)$ กำหนดไว้เช่นเดียวกัน สำหรับค่า A และ B คำนวณจาก

$$A = S(b_1^w)/S(b_2^w) \quad (42)$$

และ

$$B = \bar{b}_1^w - A\bar{b}_2^w \quad (43)$$

นอกเหนือจากนี้ยังมีวิธีการอื่นในกลุ่มนี้ที่โคเฮนและคิมไม่ได้ศึกษาไว้คือ วิธี robust mean และ sigma ซึ่งอธิบายโดยบีจาและวิงเกอสกี (Cohen and Kim. 1998 : 120; citing Bejar & Wingersky. 1981) และวิธีการทำซ้ำของ mean และ sigma โดยสตอกกิงและลอร์ด (Stocking & Lord. 1983 : 203,209)

4.1.2 วิธีฟังก์ชันการตอบหรือโค้งลักษณะ (Response Function Method or Characteristic Curve Method : TRF)

4.1.2.1 เบเกอร์ (Baker. 1992 : 89 -92) ได้ขยายวิธีฟังก์ชันการตอบแบบทดสอบ (TRF) ของ สตอกกิงและลอร์ด (Stocking and Lord. 1983 : 203 - 205) ไปยังโมเดล GRM เทคนิคของเบเกอร์ ในการคำนวณสัมประสิทธิ์เชิงเส้น 2 ค่าโดยการทำให้ฟังก์ชันรอส (Loss function) กำลังสองมีค่าต่ำสุด

$$F = \frac{1}{N} \sum_{i=1}^N (T_{i1} - T_{i2}^*)^2 \quad (44)$$

เมื่อ N คือ จำนวนของจุดในการวัด θ_i และ

T_{i1} และ T_{i2}^* คือค่าคาดหวังของจำนวนคะแนนที่ตอบถูกสำหรับกลุ่ม 1 และกลุ่ม 2 กำหนดไว้ดังนี้

$$T_{i1} = \sum_{j=1}^{n_i} \sum_{k=1}^{m_j} u_{jk} p_{jk1}(\theta_i) \quad (45)$$

$$T_{i2}^* = \sum_{j=1}^{n_i} \sum_{k=1}^{m_j} u_{jk} p_{jk2}^*(\theta_i) \quad (46)$$

เมื่อ u_{jk} คือ น้ำหนักที่กระจายไปตามลำดับชั้นการตอบ k (Response category) ในข้อสอบ j เป็นดัชนีที่เป็นจำนวนเต็มของลำดับชั้นการตอบ (category)

4.1.2.2 วิธีไคสแควร์ต่ำสุด (Minimum χ^2 Method : MCS)

คิมและโคเฮน (Kim and Cohen. 1995 : 169 - 171) ขยายวิธีไคสแควร์ต่ำสุด (MCS) ของดีฟิ (Divgi. 1985 : 413 - 414) ไปยังโมเดล GRM วิธีการ คือ การทำให้ฟังก์ชันกำลังสองมีค่าต่ำสุด ดังสมการต่อไปนี้

$$\chi^2 = \sum_{j=1}^{n_i} \chi_{jm_j}^2 = \sum_{j=1}^{n_i} \xi'_{jm_j} \Sigma_{jm_j}^{-1} \xi_{jm_j} \quad (47)$$

เมื่อ

$$\xi_{jm_j} = \xi_{jm_j,1} - \xi_{jm_j,2}^* \quad (48)$$

$$\xi_{jm_j,1} = (a_{j1}, b_{j11}, \dots, b_{jk1}, \dots, b_{j(m_j-1)1})^T \quad (49)$$

$$\xi_{jm_j,2}^* = (a_{j2}^*, b_{j12}^*, \dots, b_{jk2}^*, \dots, b_{j(m_j-1)2}^*)^T \quad (50)$$

และ

$$\Sigma_{jm_j} = \Sigma_{jm_j,1} + \Sigma_{jm_j,2}^* \quad (51)$$

เมื่อ $\Sigma_{jm_j,1}$ คือค่าประมาณเมตริกซ์ ความแปรปรวน - ความแปรปรวนร่วม (variance - covariance) ของ $\xi_{jm_j,1}$

และ $\Sigma_{jm_j,2}^*$ คือค่าประมาณของเมตริกซ์ความแปรปรวน - ความแปรปรวนร่วม (variance - covariance) ที่แปลงแล้วของ $\xi_{jm_j,2}^*$

A และ B คำนวณจากการทำให้ χ^2 มีค่าต่ำสุด โดยหาอนุพันธ์เทียบกับ A และ B

4.1.3 วิธีการปรับค่าพารามิเตอร์พร้อมกัน

เฮนนิ่ง เอิร์ช และแซง (Hennings, Hirsch and Zhang. 1996 : 4 - 5) ได้ใช้วิธีปรับค่าพารามิเตอร์พร้อมกันชนิดกำหนดค่าในโมเดลเกรด เรสพอนส์ สำหรับข้อสอบที่เป็นคำถามปลายเปิด ต่อมา บาสตารี (Bastari. 2000 : 26) ได้ใช้วิธีการนี้ในแบบทดสอบข้อสอบผสม ซึ่งข้อสอบที่ตรวจให้คะแนนหลายค่าเป็นข้อสอบเขียนตอบถูกวิเคราะห์ค่ารวมกับข้อสอบที่ตรวจให้คะแนนสองค่าที่เป็นข้อสอบเลือกตอบในโมเดลเกรด เรสพอนส์ และโลจิสติก 3 พารามิเตอร์ โดยแบบทดสอบเทียบคะแนนทั้งสองฉบับปรับค่าพารามิเตอร์พร้อมกัน

4.2 โมเดลนอมินอล เรสพอนส์ (Nominal Response Model : NRM)

เบเกอร์ (Baker. 1993 : 243 - 245) ได้ขยายฟังก์ชันการตอบแบบทดสอบ (TRF) ของสตอกกิงและลอร์ด (Stocking and Lord. 1983 : 203 - 205) ไปยังโมเดลนอมินอล เทคนิคของเบเกอร์ สำหรับการหาสัมประสิทธิ์เชิงเส้น คือ การทำให้ค่าฟังก์ชันรอส กำลังสอง มีค่าต่ำสุด โดย

$$F = \frac{1}{NS} \sum_{i=1}^N \sum_{j=1}^n \sum_{k=1}^{m_j} [p_{jk}(\theta_i) - p_{jk}^*(\theta_i)]^2 \quad (52)$$

เมื่อ

$$S = \sum_{j=1}^n m_j \quad (53)$$

เมื่อ $k = 1, \dots, m_j$ เป็นลำดับชั้นการตอบข้อสอบ
 $j = 1, \dots, n$ เป็นจำนวนข้อสอบ
 $i = 1, \dots, N$ เป็นจำนวนผู้สอบ

เบเกอร์ได้สร้างโปรแกรม EQUATE 2.0 (Baker. 1993b : 20) ไว้สำหรับหาค่าสัมประสิทธิ์การเทียบคะแนนที่ใช้คงลักษณะแบบทดสอบในโมเดลเกรด เรสพอนส์ และโมเดลนอมินอล เรสพอนส์

4.3 โมเดลพาเชียล เครดิต (Partial Credit Model : PCM)

ฮันและเฟอรรารา (Huynh and Ferrara. 1994 : 134 - 135) ได้เทียบคะแนนในโมเดลพาเชียล เครดิต ด้วยโค้งลักษณะแบบทดสอบ โดยกำหนดค่าเฉลี่ยความยากของข้อสอบเท่ากับศูนย์ทั้งสองฉบับ แปลงค่าพารามิเตอร์ความสามารถเชิงเส้น แล้วเทียบคะแนนจริง นอกจากนี้ เฮนนิ่ง เอิร์ช และแซง (Hennings, Hirsch and Zhang. 1996 : 5 - 6) ได้เทียบคะแนนในโมเดลพาเชียลเครดิตวิธีปรับค่าพารามิเตอร์พร้อมกัน ด้วยการกำหนดค่าพารามิเตอร์

4.4 โมเดลเจเนอราไลซ์ พาเชียล เครดิต (Generalize Partial Credit Model : GPCM)

ไล ลิซซี และ แยง (Li, Lissitz and Yang. 1991 : 7 - 8) ได้กล่าวถึงการแปลงค่าประมาณพารามิเตอร์ข้อสอบของโมเดล GPCM คือ

$$b'_B = Ab_E + B \quad (54)$$

$$d'_{kB} = Ad_{kE} \quad (55)$$

$$b'_{kB} = Ab_{kE} + B, \text{ โดยที่ } (b_k = b - d_k) \quad (56)$$

$$a'_B = -a_E/A \quad (57)$$

เมื่อ * แทนค่ามาตรวัดที่แปลงไปสู่มาตรวัดพื้นฐาน

A คือ ความชัน และ B คือ จุดตัด

สัญลักษณ์ B,E แทน กลุ่มพื้นฐานและกลุ่มเทียบคะแนน

หลังจากการแปลงข้างต้น ค่าประมาณพารามิเตอร์ข้อสอบซึ่งขึ้นอยู่กับมาตรวัดความสามารถของกลุ่มเทียบคะแนนเป็นการอ้างอิงใหม่ไปยังมาตรวัดความสามารถของกลุ่มพื้นฐาน

$$\theta'_B = A\theta_E + B \quad (58)$$

สำหรับการเทียบคะแนนในโมเดล GPCM จากการศึกษางานวิจัย ปรากฏว่ามีการเทียบคะแนนโดยใช้วิธีโค้งลักษณะ (Characteristic Curve Method : CCM) ซึ่งเป็นการศึกษาในแบบทดสอบที่มีรูปแบบผสม (Li, Lissitz and Yang. 1999 : 8 - 11) สัมประสิทธิ์การเทียบคะแนนของ A และ B ได้จากการวิเคราะห์ค่าเฉลี่ยความแตกต่างกำลังสองระหว่างคะแนนจริงที่ถูกคาดหวัง 2 ค่า จาก 2 ชุดของค่าประมาณพารามิเตอร์ข้อสอบรวมสำหรับ N ระดับความสามารถที่กำหนดให้ค่า N ต่างกันน้อยที่สุดเท่าที่เป็นไปได้ ค่าประมาณพารามิเตอร์การเทียบคะแนนค่อนข้างจะคงที่เมื่อ N มีค่ามากกว่า 100 ฟังก์ชันต่อไปนี้ ถูกทำให้มีค่าต่ำที่สุด (Hambleton, Swaminathan & Rogers. 1991) คือ

$$f(A, B) = \frac{1}{N} \sum_{j=1}^N [t(\theta_{jB}) - t^*(\theta_{jE})]^2 \quad (59)$$

เมื่อ

$$t(\theta_{jB}) = \sum_{i=1}^L \sum_{k=1}^m x_{ik} P(X_i = x_{ik} | \theta) \quad (60)$$

และ

$$t^*(\theta_{jE}) = \sum_{i=1}^L \sum_{k=1}^m x_{ik} P(X_i = x_{ik} | \theta) \quad (61)$$

โดยที่ X_{ik} เป็นคะแนนสังเกตของลำดับชั้น k ในข้อสอบ i อยู่ในช่วง 1 – 4

(หรือ 0 ถึง 3) สำหรับข้อสอบที่มีการให้คะแนน 4 ลำดับชั้น

L เป็นจำนวนของข้อสอบรวม

M เป็นจำนวนของลำดับชั้นคะแนน

$P(\)$ เป็นความน่าจะเป็นของลำดับชั้นการตอบ k สามารถคำนวณได้จากโมเดล GPCM คะแนนจริงที่คาดหวังของผู้สอบที่มีความสามารถ θ_j บนกลุ่มของข้อสอบรวม ในแบบทดสอบพื้นฐาน (B) และแบบทดสอบเทียบคะแนน (E) คำนวณในสมการ (60) และ (61) ตามลำดับ

สัมประสิทธิ์ A และ B คำนวณจากการทำให้ฟังก์ชัน (สมการ 59) มีค่าต่ำสุดด้วยการหาอนุพันธ์ของฟังก์ชันเทียบกับ A และ B โดยกำหนดสมการของอนุพันธ์ย่อย (partial derivative) ทั้งสองสมการให้มีค่าเท่ากับ 0 โดยการกำหนดค่า B เป็นค่าคงที่ B_0 และ พิจารณาฟังก์ชัน $f(A, B_0)$ เป็นฟังก์ชันที่มีคำตอบเดียวในกรณีของฟังก์ชัน $f(A, B)$ ที่มี 2 คำตอบ (สมการ 59) การประมาณค่าความแตกต่างสำหรับการหาอนุพันธ์ย่อยเทียบกับ A ที่ A_0 สามารถเขียนสมการ ดังนี้ (Li, Lissitz and Yang. 1999 : 9; citing Nakamura. 1996)

$$\frac{\partial f(A, B)}{\partial A} \approx \frac{f((A_0 + \Delta A), B_0) - f(A_0, B_0)}{\Delta A} = 0 \quad (62)$$

เมื่อ ΔA เป็นช่วงระหว่างจุดที่ต่อเนื่องกัน 2 จุดบนเส้นจำนวนในการประมาณค่าของ A การกำหนดโดยทั่วไปจะกำหนดค่าที่เล็กมาก ๆ เช่น 0.001 ด้วยหลักการเดียวกันประยุกต์ใช้การหาอนุพันธ์ย่อยเทียบกับ B ณ ค่า B_0 จะได้

$$\frac{\partial f(A, B)}{\partial B} \approx \frac{f(A_0, (B_0 + \Delta B)) - f(A_0, B_0)}{\Delta B} = 0 \quad (63)$$

เมื่อ ΔB เป็นช่วงระหว่างจุดที่ต่อเนื่องกัน 2 จุดบนเส้นจำนวนในการประมาณค่า B สมการทั้ง 2 สมการ (62 และ 63) ไม่ใช่สมการเชิงเส้น สามารถแก้สมการเพื่อหาค่า A และ B ได้ โดยใช้กระบวนการ Newton-Raphson ด้วยการทำซ้ำ t ครั้ง (Li, Lissitz and Yang. 1999 : 10; citing Baker. 1992) ค่าประมาณพารามิเตอร์ของข้อสอบรวมในแบบทดสอบเทียบคะแนนนำมาสร้างมาตรวัดใหม่ โดยการหาค่า A และ B ที่ทำซ้ำ t ครั้ง ขณะที่คำนวณค่าที่คาดหวังของคะแนนจริง (สมการ 61) สมการที่ทำซ้ำ คือ

$$\begin{bmatrix} A \\ B \end{bmatrix}_{t+1} = \begin{bmatrix} A \\ B \end{bmatrix}_t - \begin{bmatrix} \frac{\partial^2 f}{\partial A^2} & \frac{\partial^2 f}{\partial A \partial B} \\ \frac{\partial^2 f}{\partial B \partial A} & \frac{\partial^2 f}{\partial B^2} \end{bmatrix}^{-1} \begin{bmatrix} \frac{\partial f}{\partial A} \\ \frac{\partial f}{\partial B} \end{bmatrix}_t \quad (64)$$

การประมาณค่าความแตกต่างสำหรับอนุพันธ์ลำดับสองของฟังก์ชัน $f(A, B)$ ที่ (A_0, B_0) อธิบายได้ดังนี้ (Li, Lissitz and Yang. 1999 : 10; citing Nakamura. 1996)

$$\frac{\partial^2 f}{\partial A^2} \approx \frac{f((A_0 + \Delta A), B_0) - 2f(A_0, B_0) + f((A_0 - \Delta A), B_0)}{\Delta A^2} \quad (65)$$

$$\frac{\partial^2 f(A, B)}{\partial B^2} \approx \frac{f(A_0, (B_0 + \Delta B)) - 2f(A_0, B_0) + f(A_0, (B_0 - \Delta B))}{\Delta B^2} \quad (66)$$

$$\begin{aligned} \frac{\partial^2 f(A, B)}{\partial A \partial B} &\approx \frac{f((A_0 + \Delta A), (B_0 + \Delta B)) - f((A_0 - \Delta A), (B_0 + \Delta B))}{\Delta A \Delta B} \\ &\quad - \frac{f((A_0 + \Delta A), (B_0 - \Delta B)) + f((A_0 - \Delta A), (B_0 - \Delta B))}{\Delta A \Delta B} \end{aligned} \quad (67)$$

$$\frac{\partial^2 f(B, A)}{\partial B \partial A} = \frac{\partial^2 f(A, B)}{\partial A \partial B} \quad (68)$$

ค่าเริ่มต้นของ A และ B เป็นสิ่งสำคัญสำหรับการแก้สมการทำซ้ำ (64) ค่าเริ่มต้นสำหรับ A และ B คำนวณโดยสมการข้างล่างนี้ จะต้องเป็นค่าที่ทำให้มีการลู่เข้าเป็นส่วนใหญ่

$$A_{\text{starValue}} = \frac{\sum_{i=1}^L a_{iE}}{\sum_{i=1}^L a_{iB}} \quad (69)$$

$$B_{\text{starValue}} = \left[\frac{\sum_{i=1}^L b_{iB}}{L} \right] - A_{\text{starValue}} \left[\frac{\sum_{i=1}^L b_{iE}}{L} \right] \quad (70)$$

จากนั้นจะได้พารามิเตอร์การเทียบคะแนน ค่าประมาณความสามารถและข้อสอบจากแบบทดสอบต่างฉบับซึ่งแปลงไปสู่มาตรฐานวัดเดียวกัน

จากการศึกษาวิธีการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า ผู้วิจัยคิดว่าควรจะได้มีการศึกษาวิธีการเทียบคะแนนทั้งในทฤษฎีการทดสอบแบบมาตรฐานเดิม และในทฤษฎีการตอบข้อสอบ โดยวิธีการเทียบคะแนนในทฤษฎีการทดสอบแบบมาตรฐานเดิม ผู้วิจัยได้ศึกษาวิธีการเทียบคะแนนเชิงเส้นตรง ซึ่งเป็นวิธีง่ายและสะดวก ไม่ต้องปรับกระบวนการเทียบคะแนน ถึงแม้ว่าข้อสอบตรวจให้คะแนนหลายค่าก็ตาม และยังเลือกใช้วิธีอิกวิเปอร์เซ็นไทล์ ซึ่งเป็นวิธีการที่ให้ผลใกล้เคียงกับวิธีการเทียบคะแนนในทฤษฎีการตอบข้อสอบ สำหรับวิธีการเทียบคะแนนในทฤษฎีการตอบข้อสอบ ผู้วิจัยได้ศึกษาการเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกัน เนื่องจากแบบแผนในการเทียบคะแนนครั้งนี้เป็นการศึกษากับกลุ่มสุ่มหรือกลุ่มสมมูลในนักเรียนระดับชั้นเดียวกัน และใช้แบบทดสอบรวมภายในเป็นตัวปรับความแตกต่างของกลุ่ม และความแตกต่างของข้อสอบ จึงสามารถวิเคราะห์แบบทดสอบเทียบคะแนนทั้งสองฉบับได้พร้อมกัน นอกจากนี้ยังได้ศึกษาวิธีการเทียบคะแนนที่แปลงค่าพารามิเตอร์ความสามารถเชิงเส้นอีก 2 วิธี คือ วิธีการเทียบคะแนนโดยใช้ค่าพารามิเตอร์ข้อสอบบางค่า และที่ใช้พารามิเตอร์ข้อสอบทุกค่าพร้อมกัน ได้แก่ วิธีการเทียบคะแนนที่ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (วิธี mean และ sigma) และ วิธีการเทียบคะแนนด้วยโค้งลักษณะแบบทดสอบ ซึ่งเป็นการศึกษาการเทียบคะแนนโดยใช้ข้อมูลจริง การเทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบเป็นการเทียบทั้งคะแนนจริงและคะแนนสังเกตและโมเดล IRT ที่ใช้ในการศึกษาครั้งนี้เลือกใช้โมเดล GRM ที่สามารถใช้วิเคราะห์ข้อมูลดังกล่าวได้ รวมทั้งมีโปรแกรมการเทียบคะแนนด้วยโค้งลักษณะแบบทดสอบในโมเดลนี้ด้วย จึงเป็นสิ่งสนับสนุนการวิจัยครั้งนี้ เพื่อให้เกิดสารสนเทศ และข้อความรู้ในการนำไปใช้ต่อไป

5. เกณฑ์การเทียบคะแนน (Equating Criteria)

กระบวนการเทียบคะแนน นอกเหนือจากการเลือกแบบแผนการรวบรวมข้อมูล การกำหนดนิยามและวิธีการสำหรับประมาณความสัมพันธ์ของการเทียบคะแนนแล้ว หลังจากดำเนินการเทียบคะแนนแล้ว ควรประเมินผลการเทียบคะแนนที่ได้ โดย American Psychological Association (Harris and Crouse. 1993 : 195) ได้กำหนดไว้ในมาตรฐานที่ 4.6 ว่า "ให้ระบุการตรวจสอบความเพียงพอของการเทียบคะแนน (Adequacy of equating) ไว้ในรายงาน" การประเมินผลการเทียบคะแนนจำเป็นต้องใช้เกณฑ์ในการเทียบคะแนน แฮร์ริส และ ครูส (Harris and Crouse. 1993 : 195 – 240) ได้รวบรวมเกณฑ์การเทียบคะแนนที่ใช้ในงานวิจัยต่าง ๆ รวมทั้งโคเลนและเบรนนอน (Kolen and Brennan. 1995:23-24) ได้กล่าวถึงคุณสมบัติต่าง ๆ ของการเทียบคะแนนที่สามารถใช้เป็นเกณฑ์การเทียบคะแนนได้ มีรายละเอียดดังต่อไปนี้

1. Weak Equity or Equivalent Expected Score

ดีฟิจิ (Harris and Crouse. 1993 : 201; citing Divgi. 1981) และเยน (Harris and Crouse 1993 : 201; citing Yen. 1983) ได้เสนอกรณีพิเศษของลอร์ด อีควิตี้ (Lord's Equity) ที่เรียกว่า Weak Equity จากการที่ลอร์ด (Lord. 1980 : 195) ได้กล่าวถึงคุณสมบัติอีควิตี้ไว้ว่า "การแจกแจงคะแนนจริงแบบมีเงื่อนไขของแบบทดสอบแต่ละฉบับ หลังจากเทียบคะแนนแล้ว ต้องเท่ากัน" แต่ Weak Equity พิจารณา "ค่าคาดหวังของการแจกแจงแบบมีเงื่อนไขของแบบทดสอบแต่ละฉบับหลังจากการเทียบคะแนนแล้วต้องเท่ากัน" ซึ่งเป็น First order equity ความหมายคือ ค่าคาดหวังแบบมีเงื่อนไขของคะแนนจริงที่เทียบเท่ากับ X มีค่าเท่ากับค่าคาดหวังแบบมีเงื่อนไขของคะแนนจริงฉบับ X ดังสูตร

$$E = [e(Y) | \gamma(\tau)] = E(X | \tau) = \tau$$

เมื่อ τ คือ คะแนนจริงของผู้สอบจากแบบทดสอบฉบับเก่า (X)
 $\gamma(\tau)$ คือ ฟังก์ชันของ τ จากแบบทดสอบฉบับใหม่ (Y)
 X คือ คะแนนฉบับเก่า
 $e(Y)$ คือ คะแนนของแบบทดสอบฉบับใหม่ที่เทียบเท่ากับฉบับเก่า

การเทียบคะแนนจริงจึงมีความสัมพันธ์กับคุณสมบัติของอีควิตี้

2. ดัชนี (Indices)

ใช้เพื่อเปรียบเทียบคะแนนแปลงที่ได้จากการเทียบคะแนน 2 วิธี ดัชนีนี้ใช้สรุปภาพรวมของความแตกต่าง ดัชนีเหล่านี้เดิมใช้ประเมินความคลาดเคลื่อนทางสถิติ แต่นำมาใช้กับการเทียบคะแนนอาจจะไม่ชัดเจนนักแต่สามารถประยุกต์ใช้ดัชนีเหล่านี้ได้โดยให้ค่าที่แท้จริง (true value) เป็นเกณฑ์ ดัชนีกลุ่มนี้จำแนกเป็นดัชนีถ่วงน้ำหนักและดัชนีไม่ถ่วงน้ำหนัก

2.1 ดัชนีถ่วงน้ำหนัก ด้วยความถี่ของคะแนนดิบของแบบทดสอบฉบับเป้าหมาย ประกอบด้วย

$$2.1.1 \quad \text{RMSD} = \left(\frac{\sum_i f_i (A_i - B_i)^2}{\sum_i f_i} \right)^{1/2}$$

$$2.1.2 \quad \text{MAD} = \left(\frac{\sum_i f_i |A_i - B_i|}{\sum_i f_i} \right)$$

$$2.1.3 \quad \text{MSD} = \frac{\sum_i f_i (A_i - B_i)}{\sum_i f_i}$$

โดยที่ A_i คือ คะแนนสมมูล i จากแบบทดสอบฉบับใหม่ที่ยกคะแนนไปยังสเกลของฉบับเก่าด้วยการเทียบคะแนนวิธีที่หนึ่ง
 B_i คือ คะแนนเกณฑ์ ซึ่งเป็นคะแนนที่เป็นจริง หรือคะแนนสมมูล i จากแบบทดสอบฉบับใหม่ที่ยกคะแนนไปยังสเกลของฉบับเก่าด้วยการเทียบคะแนนวิธีอื่น

2.2 ดัชนีไม่ถ่วงน้ำหนัก เป็นการพิจารณาความแตกต่างที่เกิดขึ้นตลอดช่วงคะแนนของมาตราวัด โดยไม่กำหนดน้ำหนักในแต่ละระดับคะแนน ประกอบด้วย

$$2.2.1 \quad \text{RMSD} = \left(\frac{\sum_i (A_i - B_i)^2}{k} \right)^{1/2}$$

$$2.2.2 \quad \text{MAD}_u = \left(\frac{\sum_i |A_i - B_i|}{k} \right)$$

$$2.2.3 \quad MSD_u = \frac{\sum (A_i - B_i)}{k}$$

โดยที่ k คือ จำนวนของระดับคะแนน (Score Point)

นอกจากนี้ ดัชนี RMSD ยังสามารถนำมาแยกเป็นความแปรปรวนของความแตกต่างและความลำเอียงยกกำลังสอง คือ

$$\sum f_i(d_i)^2/n = \sum f_i(d_i - \bar{d})^2/n + \bar{d}^2$$

โดยที่ d_i คือ $(A_i - B_i)$
 $\bar{d}$ คือ ความแตกต่างเฉลี่ย

ความลำเอียง (bias) ในการเทียบคะแนน คือ ความคลาดเคลื่อนเชิงระบบ (systematic error)

การใช้ดัชนีเหล่านี้เป็นเกณฑ์ในการเทียบคะแนนสามารถแปลความหมายตามขนาดดัชนี ซึ่งเป็นค่าที่เรียงลำดับได้ว่าดัชนีที่มีค่าน้อยแสดงว่าความแตกต่างนั้นมีน้อย แต่ดัชนีขนาดเท่าใดที่ทำให้ผลการเทียบคะแนนเป็นที่น่าพอใจยังไม่ได้มีการระบุไว้

ปีเตอร์เซนและคณะ (Petersen and others, 1982 :71- 135) ได้เสนอดัชนีความแตกต่าง (Discrepancy Index) มีสูตรการคำนวณคือ

$$\text{total error} = \sum f_j d_j^2 / n S_i^2$$

เมื่อ $d_j = (t_j - t'_j)$
 t_j = คะแนนเกณฑ์ ซึ่งเป็นคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ
 t'_j = คะแนนแปลงจากการเทียบคะแนนวิธีอื่น
 S_i^2 = ความแปรปรวนของคะแนน t_j

ค่าดัชนีที่ได้มีลักษณะเป็นค่ามาตรฐาน เพราะถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ สามารถนำมาเปรียบเทียบกันได้โดยตรง ถึงแม้ในสถานการณ์ที่ได้ข้อมูลต่างกันก็ตาม นอกจากนี้ปีเตอร์เซนและคณะ ยังได้กำหนดหลักเกณฑ์การประเมินความเพียงพอในการเทียบคะแนน ซึ่งเป็นการประเมินประสิทธิภาพของการเทียบคะแนนตามระดับการยอมรับ 5 ระดับ โดยใช้ร้อยละของส่วนเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ยกกำลังสองเป็นเกณฑ์ในการเปรียบเทียบแต่ละระดับ

นอกจากนี้เจเกอร์ (Jaeger, 1981 : 26 – 28) ยังได้เสนอดัชนีที่ใช้ตรวจสอบความพอเพียงของกระบวนการเทียบคะแนนเชิงเส้นตรง เป็นการตรวจสอบความคล้ายคลึงกันของการแจกแจงคะแนนสะสม (Similarity of Cumulative Score Distributions) ระหว่างคะแนนแปลงกับคะแนนของแบบทดสอบฉบับเป้าหมายด้วยสถิติทดสอบโคโมโกรอฟ-สมีนอฟ (Kolmogorov – Sminov two sample test)

3. ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน (Standard error of equating)

ความคลาดเคลื่อนโดยทั่วไปจำแนกเป็นความคลาดเคลื่อนเชิงสุ่มและความคลาดเคลื่อนเชิงระบบ ความคลาดเคลื่อนเชิงสุ่มเป็นความคลาดเคลื่อนที่เกิดจากการใช้หน่วยตัวอย่างจากประชากรมาประมาณค่าความสัมพันธ์ของการเทียบคะแนน ซึ่งอาจจะเกิดความแตกต่างจากความสัมพันธ์ของการเทียบคะแนนในประชากร สำหรับคลาดเคลื่อนเชิงระบบได้แก่ วิธีการประมาณค่าทำให้เกิดความลำเอียงในการประมาณค่าความสัมพันธ์ของการเทียบคะแนน เช่น เทคนิคการปรับเรียบ มีประโยชน์คือลดความคลาดเคลื่อนเชิงสุ่ม-แต่อาจทำให้เกิดความคลาดเคลื่อนเชิงระบบ คือ ขนาดของความคลาดเคลื่อนเชิงสุ่มจะลดลงเมื่อหน่วยตัวอย่างมีขนาดเพิ่มขึ้น แต่ไม่ทำให้ความคลาดเคลื่อนเชิงระบบลดลง ดังนั้นความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนก็คือ ค่าประมาณของความคลาดเคลื่อนอันเนื่องจากการสุ่มตัวอย่าง คือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนแปลงที่เกิดจากการทำซ้ำในกระบวนการเทียบคะแนน

ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนเป็นเกณฑ์ที่สำคัญในการเทียบคะแนน แต่ในการเทียบคะแนนบางวิธี เช่น การเทียบคะแนนสังเกตตามทฤษฎีการตอบข้อสอบ ยังไม่มีสูตรเบื้องต้นในการคำนวณ ทำให้การเทียบคะแนนบางวิธีไม่สามารถคำนวณความคลาดเคลื่อนมาตรฐานได้ การใช้เกณฑ์ความคลาดเคลื่อนมาตรฐานยังคงใช้ขนาดความคลาดเคลื่อนที่เล็กกว่า แสดงถึงประสิทธิภาพที่ดีกว่า แต่ยังไม่สามารถระบุได้ว่าขนาดความคลาดเคลื่อนเท่าใดที่ยอมรับได้

4. การจำลองข้อมูล (Generated Data)

เป็นเทคนิคที่มักใช้กันบ่อย ๆ ในการศึกษาการเทียบคะแนน โดยนิยามการเทียบคะแนนที่แท้จริง (true equating) ด้วยโมเดลทางจิตวิทยาและสร้างข้อมูลให้เหมาะสมกับโมเดล การเทียบคะแนนที่แท้จริงจึงใช้เป็นเกณฑ์ของการเทียบคะแนน โมเดลเฉพาะที่สร้างขึ้นอาจมีแนวโน้มที่สนับสนุนวิธีการบางวิธี แต่แนวทางนี้จะเป็นประโยชน์ถ้าจำลองได้ใกล้เคียงกับข้อมูลที่ได้จากการทดสอบจริง

5. การเทียบคะแนนกลับสู่แบบทดสอบฉบับเดิมหรือการเทียบคะแนนลูกโซ่ (Test Equated to itself / Circular Equating)

การเทียบคะแนนกลับสู่แบบทดสอบฉบับเดิม ไม่ว่าจะเทียบโดยตรงหรือเทียบผ่านแบบทดสอบอื่น แล้วปรับสู่แบบทดสอบเดิมทำให้มีการเทียบคะแนนมากกว่า 1 ครั้ง เช่นในสถานการณ์ที่มีแบบทดสอบ 3 ฉบับ คะแนนจากฉบับ X เทียบไปสู่มาตรวัดฉบับ Y คะแนนจากแบบทดสอบฉบับ Y เทียบไปสู่มาตรวัดฉบับ Z และคะแนนจากแบบทดสอบฉบับ Z เทียบกันไปสู่ฉบับ X จะเห็นว่าการเทียบคะแนนกลับสู่ฉบับเดิม การเทียบคะแนนนี้มีความพอเพียงเมื่อคะแนนดิบของแบบทดสอบฉบับ X มีค่าเท่ากับ 1 แปลงไปสู่คะแนน 1 หรือ 2 ไปสู่ 2 แนวคิดนี้สามารถใช้กับการเทียบคะแนนแบบทดสอบ 3 ฉบับ คือ X Y และ Z ในแบบแผนกลุ่มสุ่ม หรือแบบแผนกลุ่มที่ไม่เท่าเทียมกันเมื่อใช้ข้อสอบร่วมระหว่าง X กับ Y Y กับ Z และ Z กับ X

เบรนนอนและโคเลน (Brennan and Kolen, 1987 : 283-284) ได้ชี้ประเด็นที่น่าสนใจว่า 1) การเทียบคะแนน Identity จะให้ผลดีเมื่อเทียบคะแนนแบบลูกโซ่ และวิธีการเทียบคะแนนที่ประมาณค่าพารามิเตอร์น้อยกว่า เช่น การเทียบคะแนนเชิงเส้นตรง มีความโน้มเอียงว่าจะเป็นวิธีที่ดีกว่าวิธีที่มีการประมาณค่าพารามิเตอร์มากกว่า เช่นการเทียบคะแนนอีควิเปอร์-เซ็นไทล์ ภายใต้แบบแผนกลุ่มไม่ตัดเทียมกันที่ใช้ข้อสอบร่วม 2) ผลของการเปรียบเทียบขึ้นอยู่กับแบบทดสอบที่ใช้เริ่มต้น คือ เมื่อแบบทดสอบฉบับ X เทียบกลับสู่ฉบับเดิม โดยผ่าน Y และ Z อาจแตกต่างจากผลที่ได้เมื่อ Z เทียบไปสู่ฉบับเดิม โดยผ่าน X และ Y

6. หน่วยตัวอย่างขนาดใหญ่ (Large Sample Criterion)

ข้อมูลจากตัวอย่างผู้สอบที่มีขนาดใหญ่ใช้เป็นตัวแทนของประชากร ตัวอย่างขนาดเล็กจะถูกสุ่มออกมา และผลที่ได้จากหน่วยตัวอย่างขนาดเล็ก จะนำมาเปรียบเทียบผลการเทียบคะแนนจากหน่วยตัวอย่างขนาดใหญ่ ซึ่งถือว่าเป็นความสัมพันธ์ของการเทียบคะแนนที่แท้จริง เกณฑ์นี้มีข้อจำกัดเพราะว่ากลุ่มผู้สอบขนาดใหญ่ไม่สามารถหาได้ ผลที่ได้มีความหมายเฉพาะกลุ่มผู้สอบที่มีลักษณะเดียวกัน

7. การเปรียบเทียบความคงเส้นคงวาของผลการเทียบคะแนน (Comparisons of the Consistency of Equating Results)

ความคงเส้นคงวาของการเทียบคะแนน ได้จากการนำวิธีการเทียบคะแนนหลายวิธี ไปใช้ในการเทียบคะแนนในสถานการณ์อื่น ๆ แล้วพิจารณาผลที่ได้จากการเทียบคะแนนว่า แตกต่างหรือเหมือนกันกับผลที่ได้จากสถานการณ์เดิม ถ้าให้ผลเหมือนกันแสดงว่ามีความคงเส้นคงวาในการเทียบคะแนน แต่ไม่สามารถระบุได้ว่ามีความแม่นยำในการเทียบคะแนน การใช้ความคงเส้นคงวาเป็นเกณฑ์ในการศึกษาการเทียบคะแนนต้องระมัดระวังว่าอาจเกิดความไม่คงเส้นคงวาในสถานการณ์ที่ไม่เหมาะสมหรือไม่เพียงพอกับวิธีการดังกล่าว

8. การสอบทานผล (Replication Samples / Cross – Validation Studies / Stability)

เป็นการนำผลที่ได้รับในกลุ่มตัวอย่างหนึ่งไปใช้กับกลุ่มตัวอย่างอื่น เป็นการตรวจสอบความคงที่ของการเทียบคะแนน ซึ่งเป็นเกณฑ์ภายนอกที่ใช้เพื่อเปรียบเทียบความแตกต่างของการเทียบคะแนนด้วยวิธีที่แตกต่างกัน คุณภาพของการวัดด้วยการสอบทานผลแสดงความคงที่ของการเทียบคะแนน กลุ่มตัวอย่างที่ใช้ตรวจสอบผล เรียกว่า กลุ่มสอบทานผลมีลักษณะเช่นเดียวกับกลุ่มตัวอย่างเทียบคะแนน กลุ่มสอบทานผลจะทำแบบทดสอบทั้งสองฉบับ แล้วนำคะแนนจากแบบทดสอบฉบับหนึ่งไปแปลงจากตารางเทียบคะแนนที่สร้างไว้ แล้วพิจารณาผลต่างระหว่างคะแนนที่ทำได้จริงกับคะแนนแปลงเป็นการตรวจสอบความคลาดเคลื่อนจากการเทียบคะแนน

โคเลนและวิทนี (Kolen and Whitney, 1982 : 284-285) ได้ใช้เกณฑ์สอบทานผลในการเปรียบเทียบความแตกต่างระหว่างวิธีการเทียบคะแนน 4 วิธี โดยนำคะแนนของกลุ่มสอบทานผล ที่ทำแบบทดสอบทั้งสองฉบับ ไปตรวจสอบด้วยดัชนีความแตกต่างระหว่างการแจกแจงคะแนนแปลงกับคะแนนที่ทำได้จริงเรียกว่าดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ เป็นดัชนีความแตกต่างกำลังสองเฉลี่ยระหว่างคะแนนที่ได้กับคะแนนแปลง มีสูตรในการคำนวณดังนี้

$$C = \sum_i (X_i - Y_i')^2 / nk$$

เมื่อ n คือ จำนวนคะแนนดิบของกลุ่มสอบทานผล
 k คือ จำนวนข้อสอบในแบบทดสอบร่วมที่ใช้

ถ้า C มีค่าน้อย หมายความว่าวิธีการเทียบคะแนนที่ใช้สร้างคะแนนแปลงนั้น มีความเหมาะสมและเพียงพอที่ผลให้การแปลงคะแนนคงเส้นคงวา ดังนั้นนี้เหมาะที่จะนำไปใช้กับกลุ่มสมมูลแบบสุ่ม 2 กลุ่ม ที่ทำแบบทดสอบคนละฉบับ หรือกลุ่มหนึ่งกลุ่มที่ทำแบบทดสอบทั้งสองฉบับ ซึ่งโคเลนและวิทนีย์ ได้กล่าวถึงเกณฑ์สอบทานผลว่าเป็นเกณฑ์ภายนอกที่เป็นองค์ประกอบสำคัญของการเปรียบเทียบความแตกต่างระหว่างการเทียบคะแนนด้วยวิธีการที่แตกต่างกัน

9. เกณฑ์อื่น ๆ (Other Criteria)

เกณฑ์อื่น ๆ ที่เคยใช้ประเมินการเทียบคะแนน เช่น บูดิส (Harris and Crouse. 1993 : 228; citing Budescu. 1985) ใช้ฟังก์ชันโมโนโทนิกของความสัมพันธ์ระหว่างข้อสอบร่วมกับไม่ได้รวมเป็นตัววัดประสิทธิภาพของการเทียบคะแนน เพเทียน (Harris and Crouse. 1993: 228; citing Patience.1981) เคยใช้ความสัมพันธ์ของคะแนนผู้สอบจากมาตรวัดของคะแนนแปลงกับเกณฑ์ซึ่งเป็นมาตรวัดอ้างอิง เพื่อศึกษาประสิทธิภาพสัมพัทธ์ของการเทียบคะแนนแนวดิ่ง โคป (Harris and Crouse. 1993 : 229; citing Cope. 1992) ได้แนะนำให้ใช้เกณฑ์ภายนอกเพื่อประเมินการเทียบคะแนน เช่น การเพิ่มความเที่ยงตรงเชิงพยากรณ์หรือการเปรียบเทียบอัตราสอบผ่านในการเทียบคะแนนวิธีต่าง ๆ

นอกจากนี้โคเลนและเบรนนอน (Kolen and Brennan. 1995 : 24) ได้แนะนำให้ใช้คุณสมบัติของการเทียบคะแนน เป็นเกณฑ์ในการประเมินผลจากการเทียบคะแนน คุณสมบัติเหล่านั้น คือ

คุณสมบัติการสมมาตร (Symmetry Property) คือ ผลของการเทียบคะแนนจากแบบทดสอบฉบับ X ไปยังฉบับ Y หรือการเทียบคะแนนจากแบบทดสอบฉบับ Y ไปยังฉบับ X จะต้องให้ผลเหมือนกัน

คุณสมบัติโครงสร้างเนื้อหาที่คล้ายคลึงกัน (Same Specifications Property) คือ แบบทดสอบสองฉบับที่นำมาเทียบคะแนนต้องมีความคล้ายคลึงกันของเนื้อหา ควรสร้างจากตารางวิเคราะห์หลักสูตรเดียวกัน เพื่อให้คะแนนของแบบทดสอบทั้งสองฉบับสามารถแทนที่กันได้

คุณสมบัติการเทียบคะแนนสังเกต (Observed Score Equating Properties) คือ ในการเทียบคะแนนสังเกตลักษณะของการแจกแจงคะแนนถูกกำหนดให้เท่ากันในประชากรของผู้สอบ (Kolen and Brennan. 1995 ; citing Angoff. 1984. 11) เช่น คุณสมบัติของการเทียบคะแนนอีควิเปอร์เซ็นไทล์ คะแนนแปลงของแบบทดสอบฉบับ X มีการแจกแจงคะแนนเหมือนกับแบบทดสอบฉบับ Y มีฟังก์ชันการเทียบคะแนนคือ ;

$$G^*[e_y(x)] = G(y)$$

ภายใต้คุณสมบัติของการเทียบคะแนนเชิงเส้นตรง ซึ่งคะแนนแปลงของแบบทดสอบทั้งสองฉบับมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเท่ากัน เมื่อการเทียบคะแนนอีควิเปอร์เซ็นไทล์มีคุณสมบัติดังกล่าว การเทียบคะแนนเชิงเส้นตรงก็มีคุณสมบัติของการเทียบคะแนนสังเกตด้วย

คุณสมบัติการไม่แปรเปลี่ยนกลุ่ม (Group Invariance Property) คือ คะแนนที่เทียบแล้วจะไม่แปรผัน ไม่ว่าจะเทียบคะแนนจากกลุ่มใด สามารถนำไปใช้กับกลุ่มอื่นได้

สำหรับเกณฑ์ที่ใช้ประเมินผลการเทียบคะแนนที่ได้จากการประเมินการปฏิบัติ การประเมินตามสภาพจริงที่ข้อสอบตรวจให้คะแนนหลายค่า ส่วนใหญ่ไม่ได้พัฒนาเกณฑ์ขึ้นมาใหม่ ยังคงใช้เกณฑ์ที่ใช้กับการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนสองค่า ซึ่งแฮร์ริสและครุส (Harris and Crouse, 1993 : 229) ได้กล่าวว่า อาจเป็นเพราะวิทยาการของการเทียบคะแนนในแบบทดสอบประเภทนี้ยังได้รับการพัฒนาไม่มากนัก แต่การทดสอบด้วยเครื่องมือที่ใช้คำถามปลายเปิด และการประเมินการปฏิบัติได้มีการใช้สูงมากขึ้นเรื่อย ๆ ในการวัดและประเมินผลในปัจจุบัน การประเมินผลการเทียบคะแนนประเภทนี้ควรต้องมีการพัฒนาเกณฑ์ตามไปด้วย

การนำเกณฑ์ดังกล่าวไปใช้ในการประเมินผลเทียบคะแนน มิได้นำเกณฑ์เดียวไปใช้ได้เลย เช่น การใช้เกณฑ์กลุ่มสอบทานผลควบคู่กับดัชนีที่บอกความแตกต่าง หรือการจำลองข้อมูลแล้วเทียบคะแนนสู่แบบทดสอบฉบับเดิมและใช้ดัชนีความแตกต่างเป็นตัวตรวจสอบความคลาดเคลื่อน หรือกรณีของปีเตอร์เซนและคณะที่ใช้ดัชนีความแตกต่างเป็นเกณฑ์การเทียบคะแนน และยังสร้างเกณฑ์ประเมินค่าดัชนีความแตกต่างด้วย ดังนั้น ไม่ควรมีการโต้เถียงว่าเกณฑ์ใดดีที่สุด แต่ควรพิจารณาว่าเกณฑ์นั้นมีประโยชน์หรือจำเป็นในสถานการณ์นั้น ๆ เกณฑ์การเทียบคะแนนยังมีความจำเป็นต่อการประเมินผล (Harris and Crouse, 1993 : 232) และการเปลี่ยนแปลงเกณฑ์จะทำให้มีการเปลี่ยนแปลงผลสรุปที่ได้ (Harris and Crouse, 1993 : 199; citing Skaggs, 1990)

สำหรับการศึกษารั้วนี้ ผู้วิจัยได้ประเมินผลการเทียบคะแนน โดยใช้เกณฑ์ภายในคือ

- 1) ดัชนีคะแนนความแตกต่างทั้งถ่วงน้ำหนักและไม่ได้ถ่วงน้ำหนัก 6 ดัชนีคือ ดัชนี RMSD MAD และ MSD เป็นการพิจารณาความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ และจากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ซึ่งเป็นคะแนนเกณฑ์ว่า ฟังก์ชันการเทียบคะแนนมีความคล้ายคลึงกันมากน้อยเพียงใด การใช้คะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมเป็นเกณฑ์ เนื่องจากวิธีการเทียบคะแนนตามทฤษฎีนี้ใช้คะแนนดิบที่เป็นคะแนนรวมของผู้สอบแต่ละคนมาดำเนินการเทียบคะแนน ไม่ต้องเปลี่ยนแปลงวิธีการหรือสูตรที่ใช้ในการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนสองค่า แต่ในการเทียบคะแนนตามทฤษฎีการตอบข้อสอบมีการเปลี่ยนแปลงสูตรต่าง ๆ ที่ใช้ค่อนข้างมาก รวมทั้งโมเดลการวิเคราะห์ที่แตกต่างไปจากการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนสองค่า
- 2) เกณฑ์ความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนและคะแนนจากแบบทดสอบฉบับเป้าหมาย ซึ่งเจเกอร์

(Jaeger. 1981 : 26-27) เสนอไว้สำหรับเป็นดัชนีในการตรวจสอบความเพียงพอในการเทียบคะแนนเชิงเส้นตรงหรือเชิงเส้นโค้ง ในครั้งนี้นำมาประยุกต์ใช้กับคะแนนแปลงที่ได้จากการเทียบคะแนนตามทฤษฎีต่าง ๆ ซึ่งเทียบทั้งคะแนนสังเกตและคะแนนจริง โดยตรวจสอบความแตกต่างด้วยสถิติโคโมโกรอฟ-สมิโนฟ สำหรับตัวอย่างสองกลุ่ม เนื่องจากการตรวจสอบความคล้ายคลึงกันของการแจกแจงคะแนนสะสมเป็นคุณสมบัติของการเทียบคะแนนสังเกตและเป็นคุณสมบัติของ Weak Equity ของการเทียบคะแนนจริงด้วย นอกจากนี้ยังใช้เกณฑ์ภายนอก คือ 3) การสอบทานผลการเทียบคะแนนด้วยกลุ่มสอบทานผลที่ทำแบบทดสอบสองฉบับเปรียบเทียบความแตกต่างของคะแนนฉบับเป้าหมายซึ่งเป็นคะแนนเกณฑ์กับคะแนนแปลงด้วยดัชนีความแตกต่างของปีเตอร์ เซนและคณะ (Petersen and others. 1982: 91) และประเมินค่าดัชนีความแตกต่างด้วยเกณฑ์ 5 ระดับการยอมรับของของปีเตอร์ เซนและคณะ (Petersen and others. 1982 : 93-97) เพื่อแสดงคุณภาพของผลการเทียบคะแนนแต่ละวิธีว่าอยู่ในระดับใด เกณฑ์สอบทานผลเป็นเกณฑ์ตรวจสอบความเพียงพอของการเทียบคะแนน เกณฑ์ทั้งสามเกณฑ์ที่ใช้ในการประเมินผลการเทียบคะแนนครั้งนี้เป็นเกณฑ์ที่ใช้ประเมินผลการเทียบคะแนนที่ข้อสอบตรวจสอบให้คะแนนสองค่ามาแล้วทั้งสิ้น

6. งานวิจัยเกี่ยวกับเทคนิคการเทียบคะแนน

ปัจจุบันได้มีการศึกษาเทคนิคการเทียบคะแนนกันอย่างต่อเนื่อง และกว้างขวาง ฟลูเออร์ (Bastari. 2000 : 19- 20; citing Feuer, et al. 1999) ได้กล่าวว่าเป็นสิ่งที่ท้าทายอย่างมากต่อการอ้างอิงข้อค้นพบเกี่ยวกับการเทียบคะแนน ทั้งนี้เพราะมีความแตกต่างในความครอบคลุมเนื้อหา ประเภทของข้อสอบ การดำเนินการสอบ และการใช้แบบทดสอบในสถานการณ์ที่เป็นไปไม่ได้ ตลอดจนการพัฒนาการเชื่อมต่อมาตรฐานวัด เพื่อเปรียบเทียบการประเมินที่แตกต่างกัน วารสาร Applied Psychological Measurement (Brennan. 1987) และ วารสาร Applied Measurement in Education (Bastari. 2000 : 20; citing Dorans. 1990) ซึ่งเป็นวารสารที่เป็นผู้นำทางการวัดผลทางการศึกษาและจิตวิทยา ได้เผยแพร่ประเด็นสำคัญของหัวข้อการศึกษาการเทียบคะแนน การศึกษาส่วนใหญ่เป็นการเปรียบเทียบความแตกต่างของวิธีการแปลง นอกเหนือจากนี้ สแควและริสสิซ (Skagg and Lissitz. 1986) ได้ทบทวนวรรณคดีที่เกี่ยวข้องกับการเทียบคะแนนภายใต้ทฤษฎี IRT อย่างกว้างขวางและยังมีหนังสืออีก 2 เล่มที่เกี่ยวข้องเป็นพิเศษกับการเทียบคะแนน (Holland & Rubin. 1982 ; Kolen & Brennan. 1995) จึงเห็นได้ว่า การศึกษาการเทียบคะแนนได้มุ่งประเด็นไปยังข้อสอบที่ตรวจให้คะแนนสองค่า ถึงแม้ว่าเทคนิคที่ได้จะไม่ใช่วิธีการที่ดีที่สุด เนื่องจากผลที่ได้จากการศึกษาวิธีการเทียบคะแนนในแต่ละบริบทยังคงแตกต่างกันขึ้นอยู่กับวัตถุประสงค์และสถานการณ์เฉพาะ โดยทั่วไปวิธีการภายใต้ทฤษฎี IRT ให้ผลที่ดีกว่า เมื่อเปรียบเทียบกับวิธีการเทียบคะแนนภายใต้ทฤษฎี CTT แต่ก็ไม่ได้เป็นเช่นนั้นทุกสถานการณ์ (Petersen, Cook and Stocking. 1983 : 155) ข้อค้นพบ

ที่ได้จากการศึกษาการเทียบคะแนนในแบบทดสอบที่ตรวจให้คะแนนสองค่า สามารถสรุปได้ตามองค์ประกอบที่เกี่ยวข้องกับการเทียบคะแนนดังนี้

6.1 แบบแผนการเชื่อมต่อ (Linking design) แบบแผนแบบทดสอบร่วมเป็นแบบแผนที่มีความเป็นไปได้และเป็นที่ยอมรับในทางปฏิบัติ การศึกษาส่วนมากกำหนดและอภิปรายในแบบแผนนี้ (Angoff. 1984 : 105 – 107; Petersen, Kolen and Hoover 1989; Kolen and Brennan. 1995 : 21) ซึ่งไม่มีความซับซ้อนในการจัดการกับแบบทดสอบหลาย ๆ ฉบับ

6.2 ความยาวแบบทดสอบ (Test length) แต่ละการศึกษาใช้แบบทดสอบที่มีความยาวแตกต่างกัน เช่นงานวิจัยของเบเกอร์ ใช้ 40 ข้อ (Baker. 1996 : 40) ขณะที่การศึกษาอื่น ๆ ประยุกต์ใช้กับข้อสอบจำนวน 50 ข้อ (Kim & Cohen. 1998 : 133) จึงยังไม่มีข้อสรุปที่ชัดเจนเกี่ยวกับความยาวของแบบทดสอบ

6.3 ความยาวของข้อสอบร่วม (Anchor length) เช่นเดียวกัน มีความแตกต่างในแต่ละการศึกษา แต่บางการศึกษาใช้ข้อสอบทุกข้อเป็นข้อสอบร่วม (Kim & Cohen. 1998 : 133) บางการศึกษาใช้ข้อสอบร่วมจำนวน 2 ข้อ และให้ผลดี (Wingersky & Lord. 1984, Vale 1986 : 33) ซึ่งผลที่ได้อาจจะต้องตีความด้วยความรอบคอบ (Bastari. 2000 : 21; citing Hambleton. 1986) อย่างไรก็ตามจำนวนข้อสอบร่วมที่ยาวกว่าย่อมให้ผลที่ดีกว่า

6.4 ขนาดของหน่วยตัวอย่าง (Sample size) ไม่มีการตกลงเกี่ยวกับประเด็นนี้ แต่ละการศึกษาประยุกต์ใช้หน่วยตัวอย่างขนาดแตกต่างกัน ซึ่งเบรนนอนและโคเลน (Brennan and Kolen. 1987 : 285) ได้กล่าวไว้ว่ากลุ่มตัวอย่างที่ใช้ในการเทียบคะแนนแต่ละกลุ่มไม่ควรมีขนาดต่ำกว่า 400 คน แต่การศึกษาหนึ่งใช้หน่วยตัวอย่างจริงจำนวน 3,000 คน (Han, Kolen and Pohlmann. 1997 : 107) ขณะที่การศึกษาอื่น ๆ ใช้ผู้เข้าสอบ 1,000 คน (Baker. 1996 : 49) และ 500 คน (Kim & Cohen. 1998 : 133) โดยทั่วไปแล้วหน่วยตัวอย่างขนาดใหญ่จะทำให้เกิดความแม่นยำสำหรับการประมาณค่าพารามิเตอร์ข้อสอบ

6.5 การแจกแจงความสามารถ (Ability distribution) เมื่อนำแบบทดสอบหลายฉบับดำเนินการต่อกลุ่มผู้สอบหรือกลุ่มที่สมมูลกัน ผลที่ได้จะให้ความแม่นยำมากกว่า เมื่อแบบทดสอบดำเนินการในกลุ่มที่ไม่เท่าเทียมกัน แต่ถ้าต้องการเทียบคะแนนในกลุ่มที่ไม่เท่าเทียมกัน ควรใช้แบบแผนแบบทดสอบร่วมจะทำได้ดีกว่า (Vale. 1986 : 33)

6.6 วิธีการเทียบคะแนนมาตรฐาน (Method of Equating the scales) สตอกกิงและลอร์ด (Stocking and Lord. 1983 : 207) ได้เปรียบเทียบวิธีโค้งคุณลักษณะแบบทดสอบกับวิธี robust mean และ sigma และพบว่าวิธีโค้งคุณลักษณะแบบทดสอบให้ผลที่ดีกว่า และเมื่อเปรียบเทียบวิธีโค้งคุณลักษณะแบบทดสอบกับวิธี mean และ mean เมื่อข้อมูลมีความซับซ้อนต่อการปรับ วิธีโค้งคุณลักษณะแบบทดสอบให้ผลที่แม่นยำกว่า (Baker and Al-karni. 1991 : 161) ปีเตอร์เซ็น คูก และสตอกกิง (Bastari. 2000 : 22; citing Petersen, Cook and Stocking. 1983) ได้เปรียบเทียบวิธีการปรับค่าพารามิเตอร์พร้อมกัน ด้วยวิธีกำหนดค่า b

(fixed b) กับวิธีโค้งลักษณะแบบทดสอบ ผลที่ได้ปรากฏว่าโดยภาพรวมวิธีการปรับค่าพารามิเตอร์พร้อมกันเป็นวิธีที่ดีที่สุด นอกเหนือจากนี้คิมและโคเฮน (Kim and Cohen. 1998 : 141 - 142) ได้พบว่า เมื่อจำนวนของข้อสอบรวมมีขนาดเล็ก วิธีโค้งลักษณะแบบทดสอบให้ผลที่ดีกว่าวิธีการปรับค่าพารามิเตอร์พร้อมกัน

ดังได้กล่าวมาแล้วข้างต้น โดยภาพรวมข้อค้นพบที่ได้จากการศึกษาการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนสองค่า อาจจะสามารถอ้างอิงหรือไม่สามารถอ้างอิงไปยังสถานการณ์และเงื่อนไขอื่นได้ เพราะว่าแต่ละการศึกษาจะเกี่ยวข้องกับบริบทเฉพาะ ผลสรุปอาจจะเหมือนหรือไม่เหมือนกับการศึกษาอื่น ๆ นอกเหนือจากนี้บางการศึกษาใช้ข้อมูลจริง และบางการศึกษาอาจจำลองข้อมูล ทำให้ผลของการศึกษาที่ได้มีความแตกต่างกัน แต่ก็ยังให้ภาพรวมว่าเงื่อนไขใดทำให้การเทียบคะแนนในแต่ละวิธีให้ผลที่แม่นยำที่สุด

งานวิจัยการเทียบคะแนนต่างประเทศ

สำหรับการศึกษาเกี่ยวกับการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า ได้มีการศึกษาไม่มากนัก ส่วนใหญ่เป็นการศึกษาในแบบทดสอบเขียนตอบ ในแบบทดสอบเลือกตอบยังไม่ปรากฏหลักฐาน โดยวิธีการที่ศึกษาเป็นวิธีการเทียบคะแนนทั้งตามทฤษฎีการตอบข้อสอบและตามทฤษฎีการทดสอบแบบมาตรฐานเดิม สำหรับวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบได้แก่ วิธีโค้งลักษณะแบบทดสอบและวิธีไคสแควร์ต่ำสุด ซึ่งขยายมาจากวิธีการที่ใช้กับข้อมูลที่ข้อสอบตรวจให้คะแนนสองค่า นอกจากนี้ยังมีวิธีการแปลงเชิงเส้นโดยใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (วิธี mean และ sigma) วิธีที่ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากที่กำหนดน้ำหนัก (วิธี Weighted mean และ sigma) และ วิธีใช้ค่าเฉลี่ยของพารามิเตอร์ความยากและอำนาจจำแนก (วิธี mean และ mean) ซึ่งปรับสูตรจากการเชื่อมต่อในข้อมูลที่ข้อสอบตรวจให้คะแนนสองค่า และวิธีการปรับค่าพารามิเตอร์พร้อมกัน โดยการศึกษาตามทฤษฎีการตอบข้อสอบ มีการศึกษาทั้งในโมเดลเกรด เรสพอนส์ โมเดลนอมินอล เรสพอนส์ โมเดลพาสซีล เครดิต และโมเดลเจเนอรัลไลซ์ พาสซีล เครดิต สำหรับวิธีการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้มีการนำวิธีการเทียบคะแนนเชิงเส้นตรง วิธีการเทียบคะแนนแบบค่าเฉลี่ย และวิธีอีควิเปอร์เซ็นไทล์มาศึกษา

การศึกษาวิธีการเทียบคะแนนในโมเดลเกรด เรสพอนส์ เริ่มจากเบเกอร์ (Baker. 1996 : 87 - 96) ได้พัฒนาวิธีการเทียบคะแนนด้วยโค้งลักษณะแบบทดสอบและวิธีไคสแควร์ต่ำสุด โดยขยายจากวิธีการที่ใช้ในโมเดล IRT ที่ข้อสอบตรวจให้คะแนนสองค่าในการศึกษาใช้การจำลองข้อมูล ปรากฏว่า สมประสิทธิ์การแปลงคะแนนที่ได้จากวิธีการนี้มีค่าเหมือนกันกับค่าที่ใช้ในการสร้างข้อมูล ดังนั้นวิธีการนี้สามารถดำเนินการกับข้อมูลที่ข้อสอบไม่ได้ให้คะแนนแบบสอง

ค่าได้ และเบเกอร์ (Baker. 1997 : 157 - 172) ได้ศึกษาเพื่อประเมินวิธีการเทียบคะแนนด้วย โค้งคุณลักษณะแบบทดสอบในสถานการณ์เทียบคะแนนที่แตกต่างกัน ได้แก่การพิจารณาค่า พารามิเตอร์ข้อสอบที่ได้ใหม่ การเทียบคะแนนไปสู่วิธีการเทียบคะแนนพื้นฐาน (baseline) และการเทียบคะแนนรายคู่ (pairwise equating) โดยนำแบบแผนแบบทดสอบร่วม และการแจกแจง โดยการจับคู่ (Matched distribution) มาประยุกต์ใช้ นอกเหนือจากนี้ข้อมูลถูกทำซ้ำโดยใช้ หน่วยตัวอย่างที่มีขนาดพอเพียง พบว่าทุกเงื่อนไขที่ต้องตรวจสอบ การแจกแจงการสุ่มเชิง ประจักษ์ของสัมประสิทธิ์การเทียบคะแนนที่สร้างจากวิธีนี้มีสมบัติที่ดีพอประมาณ

นอกจากนี้ คิมและโคเฮน (Kim & Cohen. 1995 : 167 - 176) ได้ขยายวิธี ไคสแควร์ต่ำสุด (Minimum χ^2) ซึ่งเป็นวิธีการที่ต่างจากวิธีโค้งลักษณะแบบทดสอบตรงที่ ฟังก์ชันรอส ซึ่งได้จากคะแนนจริงของข้อสอบร่วมที่บรรจุอยู่ในแบบทดสอบ 2 ฉบับ ได้รวม ความคลาดเคลื่อนเชิงสุ่มเข้าในเมตริกซ์ความแปรปรวนร่วม วิธีไคสแควร์ต่ำสุดเป็นวิธีการที่ทำให้ฟังก์ชันรอสมีค่าต่ำสุด นอกจากนี้ คณะผู้วิจัยได้เปรียบเทียบวิธีการนี้ กับวิธีโค้งลักษณะ แบบทดสอบพบว่า วิธีไคสแควร์ต่ำสุดมีการคำนวณน้อยมากและให้สัมประสิทธิ์การเทียบ คะแนนที่แตกต่างเพียงเล็กน้อยจากผลที่ได้จากวิธีโค้งลักษณะแบบทดสอบ

ต่อมาโคเฮนและคิม (Cohen & Kim. 1998 : 116 - 130) ได้ศึกษาวิธีการเชื่อมต่อในทฤษฎีการตอบข้อสอบที่เป็นการแปลงเชิงเส้น 5 วิธี ได้แก่ วิธีโค้งลักษณะแบบทดสอบ (TRF) วิธีไคสแควร์ต่ำสุด (MCS) และวิธีใช้ค่าพารามิเตอร์บางค่าหรือใช้ทุกค่าไม่พร้อมกันในการประมาณค่าสัมประสิทธิ์การเทียบคะแนนได้แก่วิธีที่ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ของค่าความยากข้อสอบ (วิธี mean และ sigma : MS) ของ มาร์โค (Marco. 1977 : 140 - 141) วิธีใช้ค่าเฉลี่ยของพารามิเตอร์อำนาจจำแนกและค่าความยาก (วิธี mean และ mean : LH) ของ ลอยด์และฮูเวอร์ (Cohen and Kim. 1998 : 118 - 119; citing Loyd & Hoover. 1980) และวิธี ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากที่กำหนดน้ำหนัก (วิธี Weighted mean และ sigma : WMS) ของลินน์และคณะ (Cohen and Kim. 1998 : 119-120; citing Linn, et al. 1980, 1981) ข้อมูลที่ใช้ในการศึกษาได้จำลองขึ้นมาโดยโปรแกรม GENIRV จากพารามิเตอร์ข้อสอบ 36 ข้อ โดยใช้แบบแผนแบบทดสอบร่วมที่จำลองด้วยจำนวน ข้อสอบร่วมแตกต่างกัน 4 ขนาด คือ 5 10 20 และ 30 ขนาดของหน่วยตัวอย่าง 2 ขนาด คือ 300 และ 1000 สร้างด้วยการแจกแจง 0 สองแบบคือ แบบปกติ (0,1) และแบบปกติ (1,1) ดังนั้นจึงมีเงื่อนไขของการศึกษา 16 เงื่อนไขที่ทำซ้ำในแต่ละเงื่อนไข 50 ครั้ง ชุดของข้อมูลที่ สร้างขึ้น 800 ชุด ประมาณค่าพารามิเตอร์ข้อสอบโดยใช้โปรแกรม MULTLOG ดำเนินการตาม แบบแผนการเทียบคะแนนแนวนอน (จับคู่ 0) และตามแนวตั้ง (ไม่จับคู่ 0) โดยเกณฑ์ที่ใช้ ประเมินคุณภาพ วิธีที่ 1 ประเมินจากการตรวจสอบความแตกต่างระหว่างสัมประสิทธิ์การเชื่อมต่อ กับค่าที่คาดหวัง วิธีที่ 2 พิจารณาจากรากที่สองของความแตกต่างกำลังสองเฉลี่ย (Root

mean square difference : RMSD) ระหว่างค่าประมาณพารามิเตอร์และพารามิเตอร์ที่สร้างใหม่ โดยคำนวณแยกในแต่ละพารามิเตอร์ ($RMSD_a$, $RMSD_b$) และดัชนี Mean Euclidian distance (MED) เป็นค่าเฉลี่ยของรากที่สองของผลรวมของความแตกต่างกำลังสองระหว่างพารามิเตอร์อำนาจจำแนก (α_j) และความยาก (β_{jk}) ที่ประมาณจากค่าพารามิเตอร์ที่สร้างใหม่

ผลการศึกษา ปรากฏว่าโดยภาพรวมทุกวิธีการเชื่อมต่อให้สัมประสิทธิ์การแปลงคล้ายคลึงกัน แตกต่างกันเพียงเล็กน้อย แต่ไม่ปรากฏผลต่อการใช้ในทุกเงื่อนไข โดยขนาดของ RMSD และ MED ในแต่ละวิธีการเชื่อมต้อมีขนาดเล็กมาก ดังนั้นจึงถือว่าในทางทฤษฎีวิธีการเชื่อมต่อทั้ง 5 วิธีให้ผลที่เหมือนกันโดยส่วนใหญ่ ไม่มีเหตุผลที่จะต้องเลือกใช้ แต่ในทางปฏิบัติวิธีที่ง่ายที่สุดในการดำเนินการคือ LH รองลงมาคือ MS และ MCS ซึ่งง่ายต่อการดำเนินการกว่าวิธี TRF แต่วิธีการทั้งคู่มิมีการคำนวณที่ค่อนข้างยุ่งยากกว่าวิธี LH และ MS แต่โดยทั่วไปวิธี WMS TRF และ MCS ให้ผลที่ดีกว่าแม้ว่าการประมาณค่าจะเบี่ยงเบนไป แต่ก็ไม่ทำให้เกิดผลทางลบต่อวิธีการทั้งสามวิธี แต่จะส่งผลต่อคุณภาพของวิธี LH และ MS เพราะค่าเฉลี่ยและความแปรปรวนจะมีความไวต่อ Outlier เหตุที่วิธีการเชื่อมต่อทั้ง 5 วิธีให้ผลที่เหมือนกันเป็นเพราะใช้วิธีการประมาณค่าด้วยความควรจะเป็นสูงสุดขอบ (Marginal Maximum-likelihood Estimation : MMLE) ซึ่งให้ค่าประมาณใกล้เคียงค่าพื้นฐานมากกว่าค่าภายใต้การประมาณแบบ ความควรจะเป็นสูงสุดร่วม (Joint Maximum likelihood) ถึงแม้ว่าหน่วยตัวอย่างจะมีขนาดเล็กก็ตาม ซึ่งผลการศึกษาครั้งนี้สอดคล้องกับงานวิจัยของเบเกอร์ และอัล-คานี (Baker and Al-Karni, 1991) ที่ใช้โมเดล IRT แบบที่ข้อสอบตรวจให้คะแนนสองค่า ซึ่งไม่พบความแตกต่างระหว่างวิธี LH และ TRF ในการศึกษาครั้งนี้เมื่อประมาณค่าพารามิเตอร์ภายใต้เงื่อนไขที่ดีที่สุด จึงทำให้คุณภาพของข้อค้นพบในแบบทดสอบรวมที่มีขนาดสั้นกว่า หรือกลุ่มตัวอย่างที่มีขนาดเล็กกว่ายังคงให้ผลดีอยู่ แต่การใช้กลุ่มตัวอย่างขนาดใหญ่คือ 1,000 คน ยังให้ผลที่ดีกว่า 300 คน ขณะเดียวกันจำนวนข้อสอบรวมที่ยาวกว่าให้ผลที่ดีกว่า แม้ว่าอยู่ในเงื่อนไขการไม่จับคู่ 0 ก็ตาม

นอกจากนี้ ในการศึกษาได้แก้ไขการคำนวณในวิธีสแควร์ต่ำสุด (MCS) ซึ่งเป็นวิธีที่ต้องใช้พจน์ของ off-diagonal covariance สำหรับแต่ละข้อ แต่โปรแกรมคอมพิวเตอร์ในปัจจุบันที่ใช้สำหรับการประมาณค่าพารามิเตอร์ข้อสอบ ไม่สามารถให้ค่าดังกล่าวได้ ดังนั้นไคสแควร์ (ฟังก์ชันกำลังสอง) ให้ค่าต่ำสุดเฉพาะพจน์ diagonal ของเมตริกซ์ ความแปรปรวน – ความแปรปรวนร่วม (variance – covariance) เท่านั้น ผลที่นำมาใช้ในครั้งนี้จึงเป็นสัมประสิทธิ์เพียร์สันที่มีความสัมพันธ์กันสูงกับ Mahalanobis D^2 เพราะมีการคำนวณที่ง่าย คิม และโคเฮน (Cohen and Kim, 1998 : 128; citing Kim and Cohen, 1995) ได้แสดงให้เห็นว่าไม่มีความแตกต่างระหว่างการประมาณค่าที่ใช้กับค่าประมาณที่ใช้เมตริกซ์ ความแปรปรวน – ความแปรปรวนร่วม และการศึกษาครั้งนี้เป็นการศึกษาในสถานการณ์ที่ใช้เฉพาะพารามิเตอร์ของข้อสอบรวมในการแปลงคะแนนเท่านั้น ไม่รวมสถานการณ์ที่ข้อสอบรวมเป็นส่วนหนึ่งของความยาวของแบบทดสอบที่จำลองขึ้น (คือเป็นแบบทดสอบฉบับที่มีทั้งข้อสอบรวม (Common) และข้อสอบที่

ไม่ได้ร่วม (Noncommon)) โดยโมเดลเกรด เรสพอนส์ มีความเหมาะสมกับรูปแบบของข้อสอบเขียนตอบ ที่ข้อสอบตรวจให้คะแนน 5 ลำดับขั้น ผลที่ได้สอดคล้องกับงานวิจัยในโมเดลที่ข้อสอบตรวจให้คะแนนสองค่า แสดงว่าผลที่ได้จากการศึกษาวิธีการเชื่อมต่อไม่ขึ้นกับจำนวนลำดับขั้น ไม่ว่าจะมียี่ลำดับขั้นก็ตาม จะให้ผลเหมือนกันหรือไม่ยังไม่มียงานวิจัยเชิงประจักษ์สนับสนุน ผลการศึกษาวิธีการเชื่อมต่อในข้อสอบที่ตรวจให้คะแนนหลายค่าอาจจะมี หรือไม่มี การอ้างอิงไปสู่สถานการณ์อื่น ๆ เพราะว่าทุก ๆ การศึกษาได้จำลองข้อมูลตามเงื่อนไข ได้แก่ การเลือกความยาวของแบบทดสอบ จำนวนของข้อสอบรวม และขนาดของหน่วยตัวอย่างซึ่งอาจมีความแตกต่างในบริบทอื่น ๆ นอกจากนี้คณะผู้วิจัยได้แนะนำว่าควรจะได้มีการศึกษาวิธีการเชื่อมต่อที่นำค่าประมาณสำหรับข้อสอบที่ไม่ได้ร่วมมาประเมินด้วย โดยใช้วิธีการเทียบคะแนนด้วยการปรับค่าพารามิเตอร์พร้อมกันหรือการกำหนดค่าพารามิเตอร์

บาสทารี (Bastari, 2000 : 1 - 78) ได้ศึกษาผลขององค์ประกอบที่มีต่อความแม่นยำของการเชื่อมต่อค่าประมาณพารามิเตอร์ข้อสอบผสมระหว่างข้อสอบเลือกตอบ (MC) และข้อสอบเขียนตอบ (CR) โดยใช้โมเดลโลจิสติก 3 พารามิเตอร์ร่วมกับโมเดลเกรด เรสพอนส์ การศึกษาครั้งนี้ใช้แบบแผนแบบทดสอบร่วมที่มีทั้งข้อสอบเลือกตอบ และข้อสอบเขียนตอบ องค์ประกอบที่ใช้ในการศึกษามี 6 องค์ประกอบคือ 1) ความยาวแบบทดสอบ มี 2 ชนิด คือ 30 ข้อ และ 50 ข้อ 2) สัดส่วนของข้อสอบแต่ละชนิดในแบบทดสอบ มี 2 ระดับคือ ข้อสอบ MC 80% - CR 20% และข้อสอบ MC 90% - CR 10% 3) ความยาวข้อสอบรวม โดยใช้กฎร้อยละ 10 และร้อยละ 20 ของความยาวทั้งฉบับ 4) ขนาดของหน่วยตัวอย่าง 3 ขนาดคือ 400 1000 และ 3000 คน 5) การแจกแจงความสามารถโดยเลือกการแจกแจงที่เป็นการแจกแจงปกติ และจัดให้มีระดับความแตกต่างในค่าเฉลี่ย 2 ระดับ คือมีความแตกต่างเป็น 0 และ 1 หน่วยเบี่ยงเบนมาตรฐาน 6) วิธีการเชื่อมต่อที่นำมาใช้ในการศึกษามี 2 วิธีคือ ใช้วิธีการปรับค่าพารามิเตอร์พร้อมกัน (Simultaneous Calibration) และอีกวิธีเป็นวิธีโค้งลักษณะแบบทดสอบ การศึกษาครั้งนี้เป็นการศึกษาโดยการจำลองข้อมูลจากค่าประมาณพารามิเตอร์ข้อสอบของแบบทดสอบ MCAS ปี 1998 สำหรับวิชาวิทยาศาสตร์ ในนักเรียนเกรด 10 สุ่มข้อมูลการตอบข้อสอบทั้งข้อสอบ MC และ CR เพื่อให้ได้แบบทดสอบฉบับพื้นฐานสองฉบับ แต่ละฉบับมีข้อสอบ MC 40 ข้อ และข้อสอบ CR 9 ข้อ ตามลำดับ แบบทดสอบร่วมประกอบด้วยข้อสอบ MC 9 ข้อ และข้อสอบ CR 2 ข้อ ตามลำดับ แบบทดสอบทั้งสองฉบับถูกสร้างจนเกือบคู่ขนานกัน โดยเฉพาะในค่าเฉลี่ยความยากของข้อสอบ MC และดำเนินการตามเงื่อนไขในองค์ประกอบได้ทั้งหมด 96 สถานการณ์ มีการทำซ้ำข้อมูล 50 ครั้ง โปรแกรมที่ใช้ในการวิเคราะห์คือ MULTLOG โดยการปรับค่าพารามิเตอร์พร้อมและแยกกัน ค่าประมาณพารามิเตอร์เมื่อปรับข้อมูลพร้อมกันจะนำมาใช้ได้โดยตรงสำหรับค่าประมาณที่ปรับค่าพารามิเตอร์แยกกันถูกนำมาเชื่อมต่อตามวิธีโค้งลักษณะแบบทดสอบ เกณฑ์ที่ใช้ในการประเมิน มี 2 เกณฑ์ เกณฑ์ที่หนึ่ง คือความแตกต่างกำลังสองเฉลี่ย (MSD) ประกอบด้วยความลำเอียง และความแปรปรวนสำหรับแต่ละค่าประมาณ

พารามิเตอร์ ได้แก่ พารามิเตอร์ความยากและพารามิเตอร์อำนาจจำแนกและเกณฑ์ที่สองคือ รากของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) เป็นการพิจารณาความแตกต่างระหว่างโค้ง ลักษณะข้อสอบ (ICC) จากพารามิเตอร์ข้อสอบกับค่าประมาณ เป็นการประเมินโดยใช้ค่าพารามิเตอร์ข้อสอบพร้อม ๆ กันทั้งสองเกณฑ์จะถูกเปรียบเทียบจากข้อสอบที่ไม่ได้รวมในแต่ละฉบับเท่านั้น เกณฑ์ทั้งสองใช้หลักการที่ว่าค่า RMSD และ RMSE ที่มีค่าน้อยกว่าจะมีความแม่นยำมากกว่า นอกจากนี้ยังมีการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อสำรวจข้อค้นพบทั่ว ๆ ไป ก่อนที่จะมีการศึกษาในรายละเอียดในขั้นต่อไป

ผลการวิจัยปรากฏว่า รากที่สองของความคลาดเคลื่อนกำลังสองเฉลี่ย (RMSE) และรากของความแตกต่างกำลังสองเฉลี่ย (RMSD) สำหรับพารามิเตอร์ความยากให้ผลการประเมินที่สอดคล้องกัน เมื่อพิจารณาในภาพรวมแล้วพบว่าเงื่อนไขในองค์ประกอบต่าง ๆ ที่ทำให้ได้ผลมีประสิทธิภาพมากกว่าคือ แบบทดสอบที่ยาวกว่า สัดส่วนขนาดใหญ่ของข้อสอบเลือกตอบในแบบทดสอบ แบบทดสอบรวมที่ยาวกว่า หน่วยตัวอย่างขนาดใหญ่กว่า ความไม่แตกต่างระหว่างความสามารถของกลุ่มที่ใช้ในการเชื่อมต่อแบบทดสอบ และวิธีการปรับค่าพารามิเตอร์พร้อมกัน โดยขนาดของตัวอย่างและวิธีการแปลงคะแนน จะมีอิทธิพลต่อผลที่ได้ หน่วยตัวอย่างขนาดใหญ่ก่อให้เกิดค่า RMSE RMSD ความลำเอียง และความแปรปรวนน้อยที่สุด สำหรับแต่ละชนิดของข้อสอบ ในสถานการณ์เดียวกันนี้เกิดจากวิธีการปรับค่าพารามิเตอร์พร้อมกัน องค์ประกอบอื่น ๆ ได้แก่ความยาวแบบทดสอบ สัดส่วนของข้อสอบ ความยาวของข้อสอบรวม และความแตกต่างของความสามารถของหน่วยตัวอย่าง มีอิทธิพลต่อผลที่ได้ไม่ชัดเจนนัก และปรากฏในสถานการณ์เฉพาะคือในองค์ประกอบความยาวของแบบทดสอบที่เป็นข้อสอบ MC แบบทดสอบที่ยาวกว่าจะให้ค่า RMSE RMSD ความลำเอียงและความแปรปรวนน้อยกว่าแบบทดสอบที่สั้นกว่า แต่ให้ผลตรงกันข้ามในข้อสอบ CR ซึ่งผู้วิจัยได้อภิปรายว่าน่าจะมาจากพารามิเตอร์ข้อสอบที่ใช้ในการสร้างข้อมูลพารามิเตอร์ความยากของข้อสอบ CR ที่ไม่ได้รวมมีขนาดใหญ่กว่าข้อสอบ MC มาก (ค่าเฉลี่ยพารามิเตอร์ความยากสำหรับข้อสอบ MC มีค่า 0.097 ขณะที่ข้อสอบ CR มีค่า 1.546) แสดงว่าแบบทดสอบ CR มีความยากกว่า จึงส่งผลให้เกิดความคลาดเคลื่อนขนาดใหญ่ในการเชื่อมต่อ สำหรับองค์ประกอบความแตกต่างในความสามารถ ผลที่ผิดปกติดูเห็นได้ชัดเจนคือความคลาดเคลื่อนในการเชื่อมต่อต่ำกว่าเมื่อความแตกต่างของความสามารถในกลุ่มเพิ่มขึ้น ที่เป็นเช่นนี้เนื่องจากความยากของข้อสอบ CR มีค่าเฉลี่ยถึง 1.546 เมื่อใช้กับผู้สอบที่มีความสามารถเฉลี่ยเท่ากับ 0 ข้อสอบ CR จึงยากเกินไป เมื่อเป็นเช่นนี้ การนำมาใช้กับผู้สอบที่มีความสามารถเฉลี่ยเป็น 1 พารามิเตอร์ CR จึงประมาณค่าอย่างแม่นยำมากกว่า ทำให้เกิดความคลาดเคลื่อนของการเชื่อมต่อที่เล็กกว่า สำหรับข้อสอบ MC ที่มีค่าเฉลี่ยความยากเป็น 0.096 จะให้ค่าประมาณที่แม่นยำเมื่อการแจกแจงความสามารถของผู้สอบสองกลุ่มที่เท่าเทียมกัน

ลีและคณะ (Lee and others. 2001 : 357 - 372) ได้เปรียบเทียบผลการเทียบคะแนน เมื่อใช้ทฤษฎีการตอบข้อสอบที่ตรวจให้คะแนนสองค่า คือ โมเดลโลจิสติก 3 พารามิเตอร์ และโมเดล IRT ที่ข้อสอบตรวจให้คะแนนหลายค่า ได้แก่ โมเดลเกรด เรสพอนส์ และโมเดลนอมินอล เรสพอนส์ ในบริบทของการเทียบคะแนนที่แบบทดสอบประกอบด้วยแบบทดสอบฉบับย่อย (Testlets) ข้อมูลที่ใช้ได้จากแบบทดสอบ IOWA Tests of Basic Skills (ITBS) ปี 1995 ฉบับ M และฉบับ K หน่วยตัวอย่างเป็นนักเรียนในเกรด 8 แบบทดสอบ 3 ฉบับประกอบด้วยแบบทดสอบ ด้านการอ่าน แผนที่และไดอะแกรม การแก้ปัญหาและการตีความทางคณิตศาสตร์ (Reading Maps และ Math) ในแบบทดสอบฉบับ K ข้อสอบที่วัดเกี่ยวกับการอ่านมีจำนวน 49 ข้อ 8 ตอน แต่ละตอนประกอบด้วย ข้อสอบจำนวน 9 4 7 5 5 6 3 10 ข้อ ตามลำดับ ข้อสอบที่วัดด้านแผนที่และไดอะแกรมมีจำนวน 33 ข้อ 5 ตอน แต่ละตอนประกอบด้วยข้อสอบจำนวน 7 7 6 6 7 ข้อตามลำดับ ข้อสอบที่วัดทางด้านคณิตศาสตร์มีจำนวน 36 ข้อ 8 ตอน แต่ละตอนประกอบด้วยข้อสอบจำนวน 8 4 4 4 4 4 4 ข้อตามลำดับ ในแบบทดสอบฉบับ M ข้อสอบที่วัดเกี่ยวกับการอ่านมีจำนวน 49 ข้อ 7 ตอน แต่ละตอนประกอบด้วยข้อสอบจำนวน 8 4 9 4 5 8 11 ข้อ ตามลำดับ ข้อสอบที่วัดด้านแผนที่และไดอะแกรมมีจำนวน 33 ข้อ 5 ตอน แต่ละตอนประกอบด้วยข้อสอบจำนวน 7 7 6 6 7 ข้อ ตามลำดับ ข้อสอบที่วัดทางด้านคณิตศาสตร์มีจำนวน 36 ข้อ 7 ตอน แต่ละตอนประกอบด้วยข้อสอบจำนวน 8 6 6 4 4 4 4 ข้อ ตามลำดับ ข้อสอบแต่ละข้อในแต่ละตอนให้คะแนนสองค่า จำนวนหน่วยตัวอย่างในแบบทดสอบฉบับ K จำแนกตามองค์ประกอบการวัด จำนวน 663 632 และ 537 คน ตามลำดับ และจำนวนหน่วยตัวอย่างในแบบทดสอบฉบับ M จำแนกตามองค์ประกอบของการวัด มีจำนวน 680 653 และ 561 คน ตามลำดับ

แบบแผนการเทียบคะแนนเป็นแบบแผนกลุ่มสุ่มโดยใช้วิธีการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมเป็นเกณฑ์ ได้แก่ วิธีการเทียบคะแนนแบบค่าเฉลี่ยเชิงเส้นตรง และอีควิเปอร์เซ็นไทล์ เหตุที่ใช้วิธีการเทียบคะแนนในทฤษฎีการทดสอบแบบมาตรฐานเดิมเป็นเกณฑ์เนื่องจากวิธีการในทฤษฎีนี้ คะแนนรวมไม่ได้รับอิทธิพลจากการฝ่าฝืนข้อตกลงเบื้องต้นของความเป็นอิสระของข้อสอบ ตามแนวทฤษฎีการตอบข้อสอบและวิธีการเทียบคะแนนในโปรแกรมการทดสอบของ IOWA ที่ปฏิบัติกันอยู่นั้นใช้การเทียบคะแนนวิธีอีควิเปอร์เซ็นไทล์อยู่แล้ว การวิเคราะห์ข้อมูลที่น่าสนใจในการศึกษานี้ใช้โปรแกรม RAGE ในการเทียบคะแนนโดยวิธีค่าเฉลี่ย เชิงเส้นตรงและอีควิเปอร์เซ็นไทล์ สำหรับการเทียบคะแนน IRT ที่ข้อสอบตรวจให้คะแนนสองค่าใช้โปรแกรม BILOG ในการประมาณค่าพารามิเตอร์ แล้วใช้โปรแกรมคอมพิวเตอร์เทียบคะแนนจริงและคะแนนสังเกต สำหรับการเทียบคะแนนตามทฤษฎี IRT ที่ข้อสอบตรวจให้คะแนนหลายค่า ทั้ง 2 โมเดลใช้โปรแกรม MULTILOG ในการประมาณค่าพารามิเตอร์ แล้วจึงวิเคราะห์ไค์ลักษณะแบบทดสอบในทั้งสองฉบับ และเทียบคะแนนจริงและคะแนนสังเกต โดยใช้โปรแกรมฟอร์แทนที่เขียนขึ้น โดยเปรียบเทียบ

โมเมนต์ของการแจกแจงคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับโมเมนต์ของการแจกแจงจากแบบทดสอบฉบับเป้าหมายและเปรียบเทียบความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนน IRT แต่ละวิธีกับคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมแต่ละวิธี โดยใช้ดัชนีรากของค่าเฉลี่ยกำลังสองถ่วงน้ำหนัก (RMS) และไม่ได้ถ่วงน้ำหนัก (URMS)

ผลที่ได้จากการศึกษา ปรากฏว่า ข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระและความเป็นมิติเดียวสำหรับโมเดล IRT ถูกฝ่าฝืนสำหรับแบบทดสอบที่ประกอบด้วยแบบทดสอบฉบับย่อย เมื่อข้อสอบแต่ละข้อถูกใช้เป็นหน่วยของการวิเคราะห์ โดยแบบทดสอบทางด้านคณิตศาสตร์มีการฝ่าฝืนน้อยที่สุด แต่เมื่อใช้แบบทดสอบย่อยเป็นหน่วยของการวิเคราะห์ กลับพบว่าเป็นไปตามข้อตกลงเบื้องต้น ดังนั้นจึงเป็นเหตุผลสนับสนุนว่าการเทียบคะแนนแบบทดสอบที่ประกอบด้วยแบบทดสอบฉบับย่อยควรใช้โมเดลการตอบที่ข้อสอบตรวจให้คะแนนหลายค่ามิติเดียว แทนที่โมเดล IRT ที่ข้อสอบตรวจให้คะแนนสองค่า โดยการเทียบคะแนนจริงและคะแนนสังเกตในโมเดล IRT ที่ข้อสอบตรวจให้คะแนนหลายค่าให้ผลการเทียบคะแนนเหมือนกันกับการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม มากกว่าการเทียบคะแนนจริงหรือคะแนนสังเกตของโมเดล IRT ที่ข้อสอบตรวจให้คะแนนสองค่า โดยเฉพาะในแบบทดสอบการอ่าน และแผนที่ แต่ให้ผลไม่ชัดเจนนักสำหรับแบบทดสอบคณิตศาสตร์ อาจจะเนื่องด้วยมีการฝ่าฝืนข้อตกลงเบื้องต้นของโมเดล IRT น้อยกว่าแบบทดสอบฉบับอื่น ๆ (ในการศึกษานี้ใช้วิธีการตรวจสอบความเที่ยงตรงของวิธีการเทียบคะแนนโมเดล IRT ด้วยการเปรียบเทียบฟังก์ชันการเทียบคะแนนกับฟังก์ชันการเทียบคะแนนในทฤษฎีการทดสอบแบบมาตรฐานเดิม) ในโมเดล IRT ที่ตรวจให้คะแนนหลายค่าแบบทดสอบเป็นไปตามข้อตกลงเบื้องต้นของโมเดล ดังนั้น ความสัมพันธ์ของการเทียบคะแนนในโมเดล IRT ที่ข้อสอบตรวจให้คะแนนหลายค่าจึงดีกว่าการเทียบคะแนนในโมเดล IRT ที่ข้อสอบตรวจให้คะแนนสองค่า โดยเฉพาะโมเดลนอมินอล เรสพอนส์ และโมเดลเกรด เรสพอนส์ จะเป็นโมเดลที่เป็นทางเลือกที่ดีที่สุดสำหรับการเทียบคะแนนแบบทดสอบที่ประกอบด้วยแบบทดสอบฉบับย่อย

นอกจากนี้ยังมีการศึกษาการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่าในโมเดลพหุเชิงเส้น เครดิต และโมเดลเจเนอรัลไลซ์ พหุเชิงเส้น เครดิต ดังรายละเอียดต่อไปนี้

ฮันและเฟอร์รา (Huyuh and Ferrara, 1994 : 125 - 141) ได้เปรียบเทียบวิธีการเทียบคะแนน 2 วิธี คืออีคิวเปอร์เซ็นไทล์ (EP) และการเทียบคะแนนจริงในโมเดลพหุเชิงเส้น เครดิต (PC) โดยใช้การเชื่อมต่อดัชนีลักษณะแบบทดสอบ เป็นการวิจัยเชิงประจักษ์เกี่ยวกับการเทียบคะแนนในข้อสอบที่มีลักษณะการตอบเป็นอิสระหรือคำถามปลายเปิด (free response หรือ open-ended) เป็นแบบทดสอบทางด้านภาษาและคณิตศาสตร์ แบบทดสอบภาษาเป็นการ

ประเมินพื้นฐานการอ่าน การเขียน การใช้ภาษาคือ การให้นักเรียนตอบสั้นจนถึงการตอบยาว สำหรับแบบทดสอบคณิตศาสตร์เป็นการทดสอบการแก้ปัญหาคณิตศาสตร์หลายชั้น เพื่อให้ นักเรียนได้ใช้ความสามารถในการประยุกต์ความรู้ ทักษะ กระบวนการคิด การให้คะแนนสำหรับการทดสอบทางด้านคณิตศาสตร์แบ่งเป็น ตอบแบบสั้นตอบถูกได้ 1 คะแนน ตอบผิดไม่ได้ คะแนน กฎการตอบยาวถ้าตอบถูกต้องหรือพอเพียงได้ 2 คะแนน ตอบถูกเป็นบางส่วนได้ 1 คะแนน ตอบไม่ถูกต้องหรือไม่พอเพียงไม่ได้คะแนน สำหรับด้านภาษาจะมี รูบิก (rubrics) การให้คะแนนเกี่ยวกับการเขียนความเรียง เป็น 0 1 2 3 คะแนน ในแต่ละหัวเรื่อง แล้วจึงรวม คะแนนที่ได้ ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลที่ได้จาก MSPAP (Maryland School Performance Assessment Program) ในปี 1991 โดยผู้สอบเป็นนักเรียนในเกรด 3 5 และ 8 ในการศึกษาได้สุ่มแบบทดสอบจากฉบับเดิมออกเป็นแบบทดสอบ 2 ฉบับ ที่มีจำนวนข้อสอบ ใกล้เคียงกัน จำแนกออกเป็น 5 คู่ คือ คู่ A เป็นแบบทดสอบการอ่านในเกรด 3 คือ R35 และ R36 แต่ละฉบับมี 30 ข้อ คู่ B เป็นแบบทดสอบการอ่านในเกรด 5 คือ R51 มี 31 ข้อ และ R52 มี 30 ข้อ คู่ C เป็นแบบทดสอบทางคณิตศาสตร์ในเกรด 5 คือ M61 มี 37 ข้อ และ M65 มี 38 ข้อ คู่ D เป็นการทดสอบการอ่านเกรด 8 คือ R81 มี 27 ข้อ และ R82 มี 30 ข้อ คู่ E เป็นแบบทดสอบคณิตศาสตร์ในเกรด 8 คือ M91 มี 49 ข้อ และ M93 มี 47 ข้อ แบบทดสอบแต่ละ ฉบับได้รับการสุ่มตัวอย่างจำนวนประมาณ 1,500 คน การประมาณค่าพารามิเตอร์ในโมเดล PCM ใช้โปรแกรมฟอร์แทน HCREDTI (Huynh, 1992) ประเมินผลที่ได้โดยใช้ขนาดความแตกต่างของคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี (ADES) แสดงด้วยกราฟและตาราง ในการวิจัยนี้ได้มีการตรวจสอบข้อตกลงเบื้องต้นตามทฤษฎี IRT เกี่ยวกับความเป็นมิติเดียวและการประเมินความอิสระของข้อสอบ

ผลการวิจัยปรากฏว่า เมื่อเทียบคะแนน ณ ระดับ 0 ที่เท่ากัน วิธีการเทียบคะแนนทั้งสองวิธีให้ผลที่เหมือนกันเมื่อแบบทดสอบมีความยากง่ายปานกลางและกลุ่มผู้สอบมีความสามารถใกล้เคียงกัน (กลุ่มสุ่ม) ในกรณีเช่นนี้ คะแนนดิบมีแนวโน้มที่จะสมมาตรหรือเบ้เพียงเล็กน้อย อย่างไรก็ตามมีความแตกต่างระหว่างคะแนนที่เทียบได้ในช่วงปลายของคะแนนดิบที่มีความเบ้ คะแนนแปลงด้วยวิธี PC และ EP มีความเบ้ช่วงปลาย จึงทำให้ไม่สามารถสรุปได้ว่าทั้ง 2 วิธีนั้นวิธีใดดีที่สุด ผลการวิจัยครั้งนี้เหมือนกับผลที่ได้จากการศึกษาของโฮม (Huynh and Ferrara, 1994 : 130; citing Holms, 1980) ในการเปรียบเทียบการเทียบคะแนนในโมเดลราสช์ และวิธีอีควิเปอร์เซ็นไทล์สำหรับแบบทดสอบมาตรฐาน หลาย ๆ ฉบับที่ข้อสอบตรวจให้คะแนน 2 ค่า คณะผู้วิจัยได้ตัดข้อสอบที่ไม่เหมาะกับโมเดล (fit) ออกไป เพื่อลดความแตกต่างระหว่างวิธีการเทียบคะแนน 2 วิธี ผลที่ได้คือรูปแบบ PC อาจจะเป็นประโยชน์สำหรับการเทียบคะแนนในแบบทดสอบที่ข้อสอบตอบอย่างอิสระ ประกอบด้วยข้อสอบที่มีความยากปานกลาง อธิบายความแปรปรวนรวมในองค์ประกอบสำคัญองค์ประกอบแรกไม่น้อยกว่าร้อยละ 25 และมีความเป็นอิสระภายในแต่ละกลุ่มข้อสอบ

เฮนนิ่ง เอิร์ชและแซง (Hennings, Hirsch and Zhang. 1996 : 1 - 13) ได้ศึกษาวิธีการเทียบคะแนนที่เหมาะสมสำหรับแบบทดสอบปลายเปิด ได้แก่ วิธีอีควิเปอร์เซ็นไทล์ วิธีการเทียบคะแนนในโมเดลพหุคูณ เครดิต และเกรด เรสพอนส์ ด้วยวิธีการปรับค่าพารามิเตอร์ร่วมกัน โดยใช้โปรแกรมคอมพิวเตอร์ BIGSTEPS และ MULTILOG การเทียบคะแนนทุกวิธีประยุกต์ใช้กับข้อมูลชุดเดียวกัน เป็นการประเมินการปฏิบัติในคณิตศาสตร์และการอ่าน ปี 1993 และ 1994 กับนักเรียนในเกรด 3 5 8 และ 10 แบบทดสอบเบตเตอร์ IOWA Tests of Basic Skills (ITBS) ฉบับ k ใช้กับนักเรียนเกรด 3 5 และ 8 และแบบทดสอบ The Tests of Achievement and Proficiency (TAP) ฉบับ K ใช้กับนักเรียนเกรด 10 จากข้อมูลดังกล่าวได้สร้างจุดตัด 2 จุดตัด เป็นการจับหมวดหมู่การปฏิบัติของนักเรียนออกเป็น 3 กลุ่ม คือ 1) อยู่เหนือกว่ามาตรฐาน 2) อยู่ในมาตรฐาน 3) ต่ำกว่ามาตรฐาน เพื่อที่จะรักษามาตรฐานให้คงที่ติดต่อกันหลายปี เป็นการเทียบคะแนนดิบจากแบบทดสอบฉบับปี 1993 ไปยังคะแนนดิบของแบบทดสอบปี 1994 ในแต่ละระดับเกรดด้วยการใช้แบบแผนแบบทดสอบร่วม นักเรียนปี 1993 จะได้รับการทดสอบด้วยแบบทดสอบฉบับ A เป็นการประเมินการปฏิบัติในแบบทดสอบคณิตศาสตร์ และการอ่าน (เช่นเดียวกับ ITBS/TAP) และนักเรียนปี 1994 ได้รับการทดสอบด้วยแบบทดสอบฉบับ B ซึ่งเป็นการประเมินการปฏิบัติ เช่นเดียวกับ ITBS/TAP และแบบทดสอบ ITBS/TAP เป็นแบบทดสอบร่วม ประเมินผลการเทียบคะแนนจากคะแนนสมมูลจากการเทียบคะแนนแต่ละวิธีที่มีต่อคะแนนจุดตัดซึ่งสรุปในรูปร้อยละของนักเรียนในแต่ละมาตรฐาน

ผลที่ได้จากการศึกษาปรากฏว่า ทั้ง 3 วิธีการให้ผลที่ค่อนข้างคงที่ วิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์และโมเดลพหุคูณ เครดิต ให้ผลที่คล้ายคลึงกัน สำหรับการเทียบคะแนนในโมเดลเกรด เรสพอนส์ จะมีบางส่วนให้ความไม่คงที่มากกว่า 2 วิธีการแรก อาจจะเนื่องจากจำนวนของนักเรียนน้อยกว่าจำนวนที่เคยใช้ในการปรับค่าในโปรแกรม ปรากฏรายละเอียด คือ การเทียบคะแนนในเกรด 8 และ 10 ทั้ง 3 วิธีให้ผลที่คล้ายคลึงกัน สำหรับเกรด 3 วิธีการเทียบคะแนนในโมเดลเกรด เรสพอนส์ ให้ผลไม่คงที่มากที่สุด และสำหรับเกรด 5 วิธีการเทียบคะแนนในโมเดลพหุคูณ เครดิตให้ผลไม่คงที่มากที่สุด และการเทียบคะแนนทั้ง 3 วิธีให้ผลที่คล้ายคลึงกัน สำหรับเกรด 3 และ 10 ในช่วงคะแนนดิบที่อยู่รอบ ๆ จุดตัด ในเกรด 5 และ 8 การเทียบคะแนนในโมเดลเกรด เรสพอนส์ มีความโน้มเอียงไปสู่การแปรผันมากกว่า 2 วิธีอื่น ๆ ผลการเทียบคะแนนในครั้งนี้ไม่ได้ประยุกต์ใช้ เนื่องจากเป็นวิธีการที่เฉพาะอาจจะไม่แม่นยำเมื่อต้องเกี่ยวข้องกับบริบทอื่น ในการศึกษาครั้งนี้ได้มีการตรวจสอบผลของการเทียบคะแนนในรูปร้อยละของนักเรียนที่ถูกจัดกลุ่มในมาตรฐาน 3 ระดับ ซึ่งถือว่าเป็นเกณฑ์สำหรับเปรียบเทียบ-เทียบวิธีการเทียบคะแนนทั้ง 3 วิธีด้วย วิธีอีควิเปอร์เซ็นไทล์ให้ผลที่คงที่มากที่สุด วิธีการเทียบคะแนนในโมเดลพหุคูณ เครดิต และโมเดลเกรด เรสพอนส์ เบี่ยงเบนจากอัตราการจัดระดับมาตรฐาน โดยโมเดลเกรด เรสพอนส์ ให้ผลที่ไม่คงที่มากที่สุด ซึ่งสอดคล้องกับผลที่ได้จากการตรวจสอบในตอนต้น

ไล ลิซซิงและแยง (Li, Lissitz and Yang 1999 : 1 – 31) ได้ศึกษาวิธีการเทียบคะแนนโดยใช้โค้งลักษณะแบบทดสอบกับแบบทดสอบที่มีข้อสอบผสม คือ ข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่า ในโมเดลเจเนอรัลไลซ์ พาเซียล เดรดิต การศึกษาจำลองข้อมูลจากพารามิเตอร์ข้อสอบของแบบทดสอบการอ่านและการเขียนที่ใช้กับนักเรียนเกรด 4 ในโรงเรียนระดับตำบล องค์ประกอบต่าง ๆ ที่ใช้ในการศึกษา คือ 1) สัดส่วนของข้อสอบรวมที่ให้คะแนนสองค่ากับหลายค่า ข้อสอบที่ให้คะแนนหลายค่ามีจำนวน 5 ข้อ แต่ละข้อมี 4 ลำดับชั้นถูกรวมอยู่ในข้อสอบรวมที่ให้คะแนนสองค่า 3 เงื่อนไข คือ 10 15 และ 20 ข้อ ซึ่งเงื่อนไขดังกล่าวใช้ผลจากการศึกษาของโดโนฮิว (Donoghue, 1994) 2) ขนาดของหน่วยตัวอย่างมี 3 ขนาดคือ 1000 2000 3000 คน 3) สถานการณ์ในการเทียบคะแนน มี 2 สถานการณ์ คือ สถานการณ์ที่ศึกษาค่าของพารามิเตอร์ใหม่ (parameter recovery) เมื่อใช้เฉพาะความคลาดเคลื่อนของค่าประมาณพารามิเตอร์ในแบบทดสอบเทียบคะแนน สำหรับสถานการณ์ที่สอง เป็นสถานการณ์เทียบคะแนนในกลุ่มข้อมูลจริง เมื่อความคลาดเคลื่อนของค่าประมาณพารามิเตอร์ที่มีอยู่ทั้งแบบทดสอบเทียบคะแนน และแบบทดสอบพื้นฐาน 4) ประเภทของการเทียบคะแนน มี 2 ประเภทคือการเทียบคะแนนตามแนวนอนและแนวตั้ง จะได้เงื่อนไขต่าง ๆ ที่ใช้ในการศึกษา 36 เงื่อนไข (3 สัดส่วนของข้อสอบสองชนิด $\times$ 3 ขนาดหน่วยตัวอย่าง $\times$ 2 สถานการณ์การศึกษา $\times$ 2 สถานการณ์การเทียบคะแนน) การประเมินความแม่นยำของสัมประสิทธิ์การเทียบคะแนนโดยใช้ค่าความลำเอียง (BIAS คือ ค่าเฉลี่ยความแตกต่างพารามิเตอร์จริงและค่าประมาณที่สอดคล้องกัน) และ ค่ารากที่สองของความคลาดเคลื่อนเฉลี่ย (RMSE) ใช้วิธีการวิเคราะห์การถดถอยพหุคูณในการสรุปองค์ประกอบต่าง ๆ ที่ใช้ในการจำลองข้อมูลว่ามีผลกระทบต่อความแม่นยำของสัมประสิทธิ์การเทียบคะแนนหรือไม่ โดยเมื่อพิจารณาองค์ประกอบใดก็ได้ให้องค์ประกอบอื่น ๆ เป็นค่าคงที่

ผลจากการศึกษาปรากฏว่า วิธีการเทียบคะแนนโดยใช้โค้งลักษณะแบบทดสอบ ในโมเดล GPCM สามารถประยุกต์ใช้กับแบบทดสอบที่มีข้อสอบผสม ภายใต้เงื่อนไขต่าง ๆ ตามที่กำหนดได้ ความลำเอียงของสัมประสิทธิ์การเทียบคะแนนมีค่าใกล้ศูนย์และสอดคล้องกับค่า RMSE ที่มีค่าน้อย ซึ่งผู้วิจัยได้อภิปรายไว้ว่าค่าสองค่านี้ อาจจะมีค่ามากกว่านี้ถ้าประมาณค่าจากการทดสอบจริง นอกจากนี้รูปร่างการแจกแจงการสุมเชิงประจักษ์ของสัมประสิทธิ์การเทียบคะแนนที่ได้รับการประมาณค่าตามเงื่อนไขต่าง ๆ มีรูปร่างเป็นโค้งระฆังคว่ำที่สมมาตร ด้วยความแปรปรวนน้อยและไม่มี outlier สิ่งหนึ่งที่ปรากฏอยู่ในผลการวิจัยครั้งนี้ก็คือภายใต้สถานการณ์ที่ข้อสอบตรวจให้คะแนนหลายค่ารวมอยู่ในข้อสอบรวมที่ให้คะแนนสองค่าที่มีจำนวนข้อมากกว่า ไม่ได้ให้ค่าความลำเอียงและค่า RMSE ของสัมประสิทธิ์การเทียบคะแนนที่ต่ำกว่า ซึ่งคณะผู้วิจัยคาดว่าเกิดจากผลกระทบของการกำหนดน้ำหนักสัดส่วนของรูปแบบข้อสอบที่แตกต่างกัน จึงอาจเกิดผลต่อความแม่นยำของสัมประสิทธิ์การเทียบคะแนนที่ได้ จึงควรจะได้มีการศึกษาเงื่อนไขดังกล่าวให้ชัดเจนต่อไป โดยใช้แบบแผนที่มีเงื่อนไขที่ดีที่สุดของแบบทดสอบ

ที่มีรูปแบบผสม (ซึ่งไม่ใช่หัวข้อของการศึกษาในครั้งนี้) ผลที่ได้ข้างต้นอาจจะเป็นการเตือนนักพัฒนาแบบทดสอบ ว่าลักษณะพารามิเตอร์ข้อสอบของกลุ่มข้อสอบรวมมีความสำคัญต่อความแม่นยำของสัมประสิทธิ์การเทียบคะแนนมากกว่าองค์ประกอบของจำนวนข้อสอบรวมและสำหรับการวิเคราะห์ผลการถดถอยของ $\log [RMSE]$ ในค่าประมาณ A หรือ B จากองค์ประกอบต่าง ๆ ที่จำลองขึ้นมา พบว่ามีความไวต่อการผันแปรและความแม่นยำของการประมาณค่าในแต่ละการประมาณค่าพารามิเตอร์ของการแปลง แสดงว่าองค์ประกอบต่าง ๆ ที่จำลองขึ้นมาส่งผลต่อการประมาณค่าสัมประสิทธิ์การเทียบคะแนน จุดหนึ่งที่น่าสนใจจากข้อค้นพบนี้คือ คณะผู้วิจัยได้ตั้งคำถามว่า ข้อค้นพบนี้สามารถที่จะอ้างอิงไปยังการเทียบคะแนนที่ข้อมูลได้รับการทดสอบจากแบบทดสอบ 2 ฉบับใด ๆ ในสถานการณ์จริง แล้วให้ผลที่ตรงกันหรือไม่ในทุก ๆ เงื่อนไขที่เหมือนกัน

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการเทียบคะแนนในข้อสอบที่ตรวจให้คะแนนหลายค่า สรุปได้ว่า มีการศึกษาในแบบทดสอบที่ใช้ในการประเมินการปฏิบัติเป็นส่วนใหญ่ เป็นข้อสอบลักษณะคำถามปลายเปิด หรือข้อสอบเขียนตอบทั่ว ๆ ไป และยังมีการศึกษาในแบบทดสอบรูปแบบผสมที่ประกอบด้วยข้อสอบเลือกตอบ และข้อสอบเขียนตอบ ส่วนใหญ่เป็นการศึกษาโดยการจำลองข้อมูล ด้วยโปรแกรมคอมพิวเตอร์ตามค่าพารามิเตอร์ที่เป็นเงื่อนไข หรือจำลองจากข้อมูลที่เป็นผลการสอบระดับตำบล ระดับประเทศโดยจัดเงื่อนไขตามองค์ประกอบที่ศึกษา (Partial simulation) ดังสรุปประเด็นสำคัญจากงานวิจัยดังกล่าวข้างต้นไว้ในตาราง 2

ตาราง 2 สรุปประเด็นสำคัญของการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนน
หลายค่า

ผู้วิจัย/ ปีการศึกษา	วิธีการศึกษา	แบบทดสอบ /กลุ่มตัวอย่าง	ตัวแปรที่ศึกษา	วิธีการประเมิน
อันและ เฟอรารา (1994)	วิจัยเชิงประจักษ์ โดยใช้ผลการสอบ ของ MSPAP ปี 1991 จำลอง ข้อมูลตามลักษณะ แบบทดสอบโดยใช้ แบบแผนกลุ่มสุ่ม	แบบทดสอบทาง ด้านภาษา และ คณิตศาสตร์ ที่ เป็นคำถาม ปลายเปิด มี ระบบการให้ คะแนนที่ชัดเจน จัดแบบทดสอบ ออกเป็น 5 คู่ เพื่อใช้ในการ เทียบคะแนนใน เกรด 3 5 และ 8	วิธีการเทียบคะแนน 1. อีคิวเปอร์- เซ็นไทล์ 2. โมเดล PCM ด้วยการใช้โค้ง ลักษณะแบบ ทดสอบ	พิจารณาจากขนาด ความแตกต่างของ คะแนนแปลงจาก การเทียบคะแนนแต่ ละวิธี (ADES)
เฮนนิ่ง เอิร์ช และแซง (1996)	เป็นการเทียบ คะแนนในสถาน การณ์ทดสอบจริง โดยใช้แบบแผนแบบ ทดสอบร่วมและ สร้างจุดตัดเพื่อจัด นักเรียนเข้าสู่ มาตรฐาน 3 กลุ่ม คือ กลุ่มอยู่เหนือ ใน และต่ำกว่ามาตรฐาน	แบบทดสอบวัด การปฏิบัติใน คณิตศาสตร์และ การอ่านเป็นคำ ถามปลายเปิด นักเรียนปี 1993 เกรด 3 5 และ 8 ได้รับแบบ ทดสอบ ITBS ฉบับ K และ แบบทดสอบ A นักเรียนปี 1994 เกรด 3 5 และ 8 ได้รับแบบ ทดสอบ ITBS ฉบับ K และ	วิธีการเทียบคะแนน 1. อีคิวเปอร์- เซ็นไทล์ 2. โมเดล GRM วิธีการปรับค่า พร้อมกัน 3. โมเดล PCM วิธีการปรับค่า พร้อมกัน	1. พิจารณาจาก ขนาดความ แตกต่างของ คะแนนแปลงจาก การเทียบคะแนน แต่ละวิธี 2. ร้อยละของ นักเรียนในแต่ละ มาตรฐานหลัง จากการเทียบ คะแนนในแต่ละ วิธี

ตาราง 2 (ต่อ)

ผู้วิจัย/ ปีการศึกษา	วิธีการศึกษา	แบบทดสอบ /กลุ่มตัวอย่าง	ตัวแปรที่ศึกษา	วิธีการประเมิน
เฮนนิ่ง เอิร์ช และแซง (1996) (ต่อ)		แบบทดสอบ B สำหรับนักเรียน เกรด 10 ปี 1993 ได้รับแบบ ทดสอบ TAP ฉบับ k และแบบ ทดสอบ A ปี 1994 ได้รับแบบ ทดสอบ TAP ฉบับ K และแบบ ทดสอบฉบับ B		
โคเฮนและ คิม (1998)	จำลองข้อมูลจาก โปรแกรม GENIRV ใน เงื่อนไขที่ต้องการ ศึกษาใช้แบบแผน แบบทดสอบรวม	ข้อสอบที่ตรวจ ให้คะแนนหลาย ค่ามี 5 ลำดับชั้น	1. วิธีการเทียบ คะแนน - โค้งลักษณะแบบ ทดสอบ - ไคสแควร์ต่ำสุด - mean และ sigma - กำหนดน้ำหนัก mean และ sigma - Mean และ Mean 2. ความยาว ข้อสอบรวม 5 10 20 และ 30 ข้อ 3. ขนาดหน่วย ตัวอย่าง 300 และ 1000 คน 4. การแจกแจงพารา- มิเตอร์แบบ $N(0,1)$ และ $N(1,1)$	1. ใช้ดัชนี RMSD ในแต่ละพารา- มิเตอร์และดัชนี Mean Euclidian Distance (MED) 2. ตรวจสอบความ แตกต่างระหว่าง สัมประสิทธิ์การ เทียบคะแนน ที่ได้กับค่าที่ คาดหวัง

ตาราง 2 (ต่อ)

ผู้วิจัย/ ปีการศึกษา	วิธีการศึกษา	แบบทดสอบ /กลุ่มตัวอย่าง	ตัวแปรที่ศึกษา	วิธีการประเมิน
ไล ลิซซี และ แยง (1999)	จำลองข้อมูลจาก พารามิเตอร์การ อ่านและการเขียน ในโรงเรียนระดับ ตำบล	แบบทดสอบที่มี ข้อสอบผสมคือ ข้อสอบที่ตรวจ ให้คะแนนสองค่า และหลายค่า ใช้กับนักเรียน เกรด 4	1. วิธีการเทียบ คะแนนในโมเดล GPCM ด้วยโค้ด ลักษณะแบบ ทดสอบ 2. สัดส่วนของข้อ สอบที่ตรวจให้ คะแนน หลายค่ามี 5 ข้อ แต่ละข้อมี 4 ลำดับขั้นรวมกับ ข้อสอบรวมที่ให้ คะแนนสองค่ามี ความยาว 10 15 และ 20 ข้อ 3. ขนาดหน่วยตัว อย่าง 3 ขนาด คือ 1000 2000 และ 3000 คน 4. สถานการณ์ใน การเทียบคะแนนที่ ใช้เฉพาะค่าความ คลาดเคลื่อนที่ได้ จากแบบทดสอบ เทียบคะแนน และใน กลุ่มข้อมูลจริงที่ใช้ ความคลาดเคลื่อน จากค่าประมาณ พารามิเตอร์ของแบบ	1. ค่าความ ลำเอียง 2. RMSE

ตาราง 2 (ต่อ)

ผู้วิจัย/ ปีการศึกษา	วิธีการศึกษา	แบบทดสอบ /กลุ่มตัวอย่าง	ตัวแปรที่ศึกษา	วิธีการประเมิน
โล ลิซซี และ แยง (1999) (ต่อ)			ทดสอบทั้งสองฉบับ 5. การเทียบ คะแนนในแนวนอน และแนวตั้ง	
บาทหาว (2000)	จำลองข้อมูลจาก ค่าประมาณพารามิเตอร์ ข้อสอบของ แบบทดสอบ MCAS ปี 1998 จะได้แบบ ทดสอบ 2 ฉบับ ที่ เกือบคู่ขนานกัน แต่ละฉบับมีข้อสอบ MC 40 ข้อ และ CR 9 ข้อ และ แบบทดสอบรวม MC 9 ข้อ และ CR 2 ข้อ	แบบทดสอบ รูปแบบผสม ระหว่างข้อสอบ เลือกตอบและ เขียนตอบ วิชา วิทยาศาสตร์ใช้ กับนักเรียนเกรด 10	1. วิธีการเทียบ คะแนนในโมเดล GRM - ปรับค่าพารามิเตอร์ พร้อมกัน - โค้งลักษณะแบบ ทดสอบ 2. สัดส่วนของ ข้อสอบ - ข้อสอบ MC 80% และ CR 20% - ข้อสอบ MC 90% และ CR 10% 3. ความยาว ข้อสอบรวม - ร้อยละ 10 - ร้อยละ 20 ของ ความยาวทั้งฉบับ 4. ขนาดหน่วย ตัวอย่าง 400 1000 และ 3000 คน	1. RMSD แยก ตามพารามิเตอร์ ข้อสอบ 2. RMSE

ตาราง 2 (ต่อ)

ผู้วิจัย/ ปีการศึกษา	วิธีการศึกษา	แบบทดสอบ /กลุ่มตัวอย่าง	ตัวแปรที่ศึกษา	วิธีการประเมิน
บาสทาร์ (2000) (ต่อ)			5. การแจกแจงความ สามารถแบบ N (0,1) และ N (1,1) 6. ความยาวของ แบบทดสอบเทียบ คะแนน จำนวน 30 ข้อ และ 50 ข้อ	
ลีและคณะ (2001)	จากสถานการณ์การ ทดสอบจริง แบบ แผนกลุ่มสมมูล	แบบทดสอบ ITBS ฉบับ M และ K ประกอบ ด้วยแบบทดสอบ 3 ฉบับ คือ การอ่าน แผนที่ และการ แก้ปัญหา คณิตศาสตร์ เกรด 8 แต่ละ ฉบับมีแบบ ทดสอบฉบับ ย่อย (Testlets)	วิธีการเทียบคะแนน - โมเดลเกรด เรสพอนส์ โดยใช้ โค้งลักษณะแบบ ทดสอบ - โมเดล นอมินอล เรสพอนส์โดยใช้ โค้งลักษณะแบบ ทดสอบ - โมเดล โลจิสติก 3 พารามิเตอร์ เทียบคะแนนด้วย คะแนนจริงและ คะแนนสังเกต - เชิงเส้นตรง - อีควิเปอร์เซ็นไทล์ - ค่าเฉลี่ย	1. ดัชนีคะแนน ความแตกต่าง ชนิดถ่วง น้ำหนัก (RMS) 2. ดัชนีคะแนน ความแตกต่าง ชนิดไม่ถ่วง น้ำหนัก (URMS)

จะเห็นได้ว่าการศึกษาข้อความรู้ทางด้านนี้ยังมีไม่มากนัก งานวิจัยส่วนใหญ่เป็นการศึกษาความแม่นยำของวิธีการเทียบคะแนนด้วยเกณฑ์ในการประเมินที่แตกต่างกันไป ผลการศึกษาส่วนใหญ่มักจะสอดคล้องกับผลที่เคยศึกษาในข้อมูลที่ข้อสอบตรวจให้คะแนนสองค่า แต่ยังไม่สามารถอ้างอิงไปยังสถานการณ์อื่น ๆ ได้ เนื่องจากเป็นการศึกษาในบริบทเฉพาะ ซึ่งเกิดจากการจำลองข้อมูลเป็นส่วนใหญ่

งานวิจัยการเทียบคะแนนในประเทศ

สำหรับงานวิจัยในประเทศไทย เท่าที่สำรวจได้เป็นงานวิจัยในข้อมูลที่ข้อสอบตรวจให้คะแนนสองค่า จำแนกได้เป็นงานวิจัยเกี่ยวกับการเทียบคะแนนที่ใช้ในการตัดสินผลการเรียนรู้ ซึ่ง ยาวาดี รังชัยกุล (Yavadee Rangchaikul, 1975 : 87-94) เป็นบุคคลแรกของไทยที่เห็นความสำคัญของการเทียบมาตรา ได้ทำการเทียบมาตราคะแนนผลการสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 5 ซึ่งใช้แบบทดสอบต่างชุดกันว่ามีความเท่าเทียมกันเพียงใดที่จะนำมาใช้ตัดสินผลการสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 5 โดยใช้ข้อมูลผลการสอบซึ่งกระทรวงศึกษาธิการได้จัดดำเนินการสอบให้กับนักเรียนในระบบโรงเรียนเมื่อเดือนมีนาคม ปี 2516 และปี 2517 และจัดให้กับบุคคลทั่วไปที่อยู่นอกระบบโรงเรียนเมื่อเดือนสิงหาคม พ.ศ.2516 และปี 2517 เทียบคะแนนของแบบทดสอบทั้ง 4 ชุด กับมาตราคะแนนร่วม (Common Scale Score) ที่มีช่วงคะแนน 1-175 ค่าเฉลี่ย 100 ส่วนเบี่ยงเบนมาตรฐาน 25 กลุ่มตัวอย่างที่ใช้เป็นนักเรียนที่สอบแต่ละครั้งครั้งละ 500 คน เทียบคะแนนในวิชาภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์และวิทยาศาสตร์ โดยใช้เทคนิคการเทียบคะแนนเชิงเส้นตรงรูปแบบที่ 1 ซึ่งเสนอโดยแองกอฟฟ์ (Angoff, 1971)

ผลการศึกษาปรากฏว่าคะแนนที่เทียบกับมาตราคะแนนร่วมทั้ง 4 ฉบับ มีคะแนนไม่เท่าเทียมกันทุกวิชา กล่าวคือ นักเรียนคนหนึ่งสอบได้คะแนนรวมของวิชาภาษาไทยฉบับที่หนึ่ง (มีนาคม 2516) สูงกว่าคะแนนรวมของวิชาภาษาไทยฉบับที่สอง (มีนาคม 2517) และพบว่าคะแนนจุดตัดของแบบทดสอบแต่ละชุดไม่สามารถเปรียบเทียบกันได้ เช่น พบว่าคะแนนจุดตัดสำหรับการสอบผ่านเมื่อเดือนมีนาคม พ.ศ.2516 ในมาตราคะแนนร่วมเท่ากับ 88 แต่การสอบเมื่อเดือนมีนาคม 2517 มีคะแนนจุดตัดสำหรับการสอบผ่านในมาตราคะแนนร่วมเท่ากับ 78 ภายหลังการเทียบคะแนน ได้ตรวจสอบสัดส่วนของผู้สอบที่ได้รับการตัดสินใจได้และตกด้วยเกณฑ์ร้อยละ 50 พบว่าสัดส่วนนักเรียนที่ควรสอบได้ แต่ได้รับการตัดสินใจให้ตกมีจำนวนน่าสนใจ ในส่วนการศึกษาในรายวิชาทั้ง 5 วิชา พบว่าแต่ละวิชามีความแปรเปลี่ยนของคะแนนแตกต่างกัน วิชาภาษาไทยมีความแปรปรวนเปลี่ยนน้อยที่สุด วิชาคณิตศาสตร์มีความแปรเปลี่ยนของคะแนนมากที่สุด ผู้วิจัยได้เสนอแนะวิธีการ ซึ่งควรใช้ในการเทียบคะแนนใน

สถานการณ์การสอบนี้ว่าควรใช้รูปแบบที่ 4 ตามแองกอฟฟิเสนอไว้ ซึ่งเป็นรูปแบบการใช้แบบสอบร่วมกับกลุ่มตัวอย่างที่ไม่ได้สุ่มจะมีความเหมาะสม

สุจินดา ผ่องอักษร (2533 : 143 – 157) ได้ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะ (วิชาคณิตศาสตร์และภาษาไทย) ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบมาตรฐานรูปแบบราสส์ กลุ่มตัวอย่างทั้งสิ้นจำนวน 1,454 คน จำนวน 4 อำเภอ จาก 7 อำเภอ ในจังหวัดปทุมธานี และกลุ่มตัวอย่างเพื่อศึกษาค่าคงที่ในการเทียบมาตรฐาน สุ่มอย่างง่ายจากนักเรียนในอำเภอธัญบุรี จำนวน 581 คน ผู้วิจัยได้ออกแบบเพื่อทำการเทียบมาตรฐาน โดยหาค่าคงที่เพื่อใช้ในสมการเชิงเส้นตรง สำหรับปรับคะแนนที่ได้จากแบบสอบต่างชุดในแต่ละกลุ่มวิชาให้อยู่ในสเกลเดียวกัน โดยใช้สเกลคะแนนผลการสอบในปีการศึกษา 2529 เป็นหลัก และเทียบมาตรฐานคะแนนในปี 2530 และ 2531 ไปสู่สเกลเดียวกับสเกลคะแนนในปี 2529 หลังจากนั้นจึงนำคะแนนที่สมมูลกันระหว่างแบบทดสอบต่างชุดมาเปรียบเทียบกันโดยตรง

ผลการศึกษาปรากฏว่า หลังจากเทียบมาตรฐานแล้วจุดตัดคะแนนในแต่ละปีการศึกษามีค่าแตกต่างกันในกลุ่มวิชาทักษะภาษาไทย ตั้งแต่ปีการศึกษา 2529 ถึง 2531 มีค่าเท่ากับ 15, 18 และ 14 ตามลำดับ กลุ่มทักษะวิชาคณิตศาสตร์มีค่าเท่ากับ 19, 24 และ 18 ตามลำดับ สำหรับผลการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียน พบว่าในกลุ่มทักษะวิชาภาษาไทยนักเรียนมีผลสัมฤทธิ์ทางการเรียนสูงกว่าระดับที่น่าพอใจ และมีผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิม ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์ พบว่ามีผลสัมฤทธิ์ทางการเรียนต่ำกว่าระดับที่น่าพอใจ และมีผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิมเฉพาะในปีการศึกษา 2530 แต่ในปีการศึกษา 2531 พบว่ามีผลสัมฤทธิ์ลดลงกว่าเดิม แต่ยังสูงกว่าผลสัมฤทธิ์ทางการเรียนเมื่อปีการศึกษา 2529 และพบความแตกต่างที่เพิ่มขึ้นและลดลงนี้ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ พบว่ามีความแตกต่างกัน คือ ในปี 2530 พบว่ากลุ่มทักษะวิชาภาษาไทยมีนักเรียนร้อยละ 10 และในกลุ่มทักษะวิชาคณิตศาสตร์มีนักเรียนร้อยละ 20 ที่ควรเป็นผู้สอบผ่าน แต่ได้รับการตัดสินใจให้เป็นผู้สอบไม่ผ่าน ส่วนในปี 2531 พบว่ากลุ่มทักษะวิชาภาษาไทยมีนักเรียนร้อยละ 7 และกลุ่มทักษะวิชาคณิตศาสตร์พบว่ามีนักเรียนร้อยละ 5 ที่ควรเป็นผู้สอบไม่ผ่านแต่ได้รับการตัดสินใจให้ผ่าน

การศึกษาการเทียบคะแนนในประเทศไทยส่วนใหญ่เป็นการศึกษาการเทียบคะแนนในแนวนอนซึ่งมีการศึกษาไว้อย่างกว้างขวาง ทั้งการเปรียบเทียบวิธีการเทียบคะแนน ศึกษาคุณภาพของข้อสอบรวม ขนาดของหน่วยตัวอย่าง การพัฒนาโปรแกรมที่ใช้ในการเทียบคะแนนเชิงเส้นตรง การศึกษาความคลาดเคลื่อนมาตรฐานและความลำเอียงของการเทียบคะแนนและการพัฒนาเกณฑ์ในการเทียบคะแนน โดยข้อมูลการทดสอบเป็นข้อมูลในสถานการณ์จริงเป็นส่วนใหญ่และมีการจำลองข้อมูลบ้างเป็นบางการศึกษา ซึ่ง ชูชีพ พงษ์สมบูรณ์ (2528 : 112-118) ได้เปรียบเทียบประสิทธิภาพและความคงที่ของการเทียบมาตรฐานระหว่างรูปแบบที่ใช้แบบสอบรวมกับรูปแบบที่ใช้ผู้สอบรวม ในการเทียบมาตรา 3 วิธี คือ การเทียบมาตราเชิงเส้นตรง การเทียบโดยใช้อิควิเปอร์เซ็นไทล์ และการเทียบโดยใช้โค้งคุณลักษณะข้อสอบ กลุ่มตัวอย่างที่ใช้เป็นผลการสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศที่เข้าสอบตามโครงการตรวจสอบคุณภาพการศึกษาของกรมวิชาการ กระทรวงศึกษาธิการ ซึ่งสุ่มมาจำนวน 3,721 คน แบ่งศึกษาเป็นกลุ่มย่อย 6 กลุ่ม ๆ ละ 620 คน แต่ละกลุ่มมีความสามารถใกล้เคียงกัน แบบสอบที่ใช้เป็นแบบทดสอบคณิตศาสตร์มีข้อสอบทั้งหมด 60 ข้อ แบ่งออกเป็น 2 ฉบับ ๆ ละ 38 ข้อ มีแบบสอบรวมภายใน 20 ข้อ แบบสอบทั้งสองชุดมีความยากใกล้เคียงกัน การเปรียบเทียบประสิทธิภาพของการเทียบมาตรา พิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา ส่วนการเปรียบเทียบความคงที่ของการเทียบมาตรานั้นพิจารณาจากความไม่แปรผันไปตามกลุ่มตัวอย่าง ความสมมาตร และความเสมอภาค โดยทำการทดสอบนัยสำคัญของความแตกต่างโดยใช้สถิติทดสอบโครโมโกรอฟ-สมิรโนฟ (Kolmogorov-Smirnow two sample test)

ผลการวิจัยปรากฏว่าประสิทธิภาพของการเทียบมาตรฐานระหว่างรูปแบบที่ใช้แบบทดสอบรวมกับรูปแบบที่ใช้ผู้สอบรวมในแต่ละวิธีของการเทียบมาตราไม่แตกต่างกัน ส่วนผลการเปรียบเทียบความคงที่ของการเทียบมาตรารูปแบบที่ใช้ผู้สอบรวม พบว่าการเทียบมาตราโดยใช้อิควิเปอร์เซ็นไทล์มีความคงที่มากกว่าการเทียบเชิงเส้นตรง และมีความคงที่พอ ๆ กับการเทียบโดยใช้โค้งลักษณะข้อสอบ ส่วนการเปรียบเทียบความคงที่ของการเทียบมาตราโดยใช้รูปแบบแบบทดสอบรวมนี้ให้ผลเช่นเดียวกับรูปแบบที่ใช้ผู้สอบรวม

ภาวิณี ศรีสุขวัฒนานนท์ (2529 : 155 - 170) ได้เปรียบเทียบผลของการใช้รูปแบบการเทียบมาตราที่ต่างกัน 3 รูปแบบ คือ รูปแบบอิควิเปอร์เซ็นไทล์ รูปแบบเชิงเส้นตรงและรูปแบบอิงทฤษฎีการตอบข้อสอบแบบสามพารามิเตอร์ โดยใช้แบบสอบรวมภายในที่ต่างกัน 3 ขนาด คือ ขนาดร้อยละ 20 (7 ข้อ), ขนาดร้อยละ 40 (14 ข้อ) และขนาดร้อยละ 60 (21 ข้อ) โดยมีข้อสอบทั้งหมด 100 ข้อ จัดแบ่งเป็น 2 ชุด ชุดละ 35 ข้อ ซึ่งทั้งสองชุดมีความยากง่ายใกล้เคียงกันที่เหลือนจัดเป็นแบบสอบรวมขนาดต่าง ๆ ดังกล่าว สำหรับกลุ่มตัวอย่างแยกเป็นกรณี 2 กรณี คือ กรณีแบบทดสอบคัดเลือก 2 กลุ่ม กลุ่มละ 1,500 คน และกลุ่มสอบทานผลอีก 1 กลุ่ม

จำนวน 1,500 คน อีกกรณีหนึ่งคือกรณีแบบวัดผลสัมฤทธิ์ ซึ่งมีจำนวนกลุ่มตัวอย่างเช่นเดียวกับกรณีแรก ในการประเมินผลการเทียบมาตรฐานใช้เปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน (SEE) ตามวิธีการเทียบมาตรฐานที่ต่างกันของแบบทดสอบแต่ละกรณี ในแต่ละความยาวของแบบทดสอบรวม และตรวจสอบความเพียงพอของวิธีการเทียบมาตรฐานแต่ละวิธีในแต่ละความยาวของแบบทดสอบรวม โดยใช้ค่าดัชนีเปรียบเทียบความแตกต่าง (index C) ที่ได้จากกลุ่มสอบทานผล

ผลการวิจัยปรากฏว่าการใช้แบบสอบรวมที่ยาวกว่าให้ค่าความคลาดเคลื่อนมาตรฐานที่น้อยกว่า และให้ผลในระดับที่น่าพอใจมากกว่าทั้งสองกรณีของแบบสอบ ส่วนการเปรียบเทียบรูปแบบของการเทียบมาตรฐาน พบว่าในกรณีของแบบสอบคัดเลือกนั้น รูปแบบที่ให้ผลที่มีความเพียงพอมากที่สุด คือ รูปแบบอิกวิเปอร์เซ็นไทล์ รองลงมาคือรูปแบบทฤษฎีการตอบข้อสอบ และรูปแบบเชิงเส้นตรงตามลำดับ ซึ่งให้ผลที่แตกต่างกับในกรณีของแบบสอบวัดผลสัมฤทธิ์ที่พบว่า รูปแบบที่ให้ผลที่มีความเพียงพอมากที่สุด คือ รูปแบบเชิงเส้นตรง รองลงมาคือรูปแบบอิกวิเปอร์เซ็นไทล์ และรูปแบบทฤษฎีการตอบข้อสอบตามลำดับ และเมื่อทดสอบความแตกต่างของค่าดัชนี C เป็นรายคู่พบว่ากรณีแบบทดสอบวัดผลสัมฤทธิ์วิธีการที่ใช้รูปแบบเชิงเส้นตรงมีความเพียงพอมากกว่า วิธีที่ใช้รูปแบบทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ .05 ส่วนการเปรียบเทียบรูปแบบอื่นไม่สามารถสรุปความแตกต่างอย่างมีนัยสำคัญ

เรวดี อินทสระ (2530 : 64 – 69) ได้เปรียบเทียบความคลาดเคลื่อนและความเพียงพอในการเทียบมาตรฐาน 2 รูปแบบ คือ การเทียบมาตรฐานรูปแบบอิงทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2529 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดพัทลุง โดยแบบทดสอบที่ใช้เทียบมาตรฐานเป็นแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ เรื่องอัตราส่วน ร้อยละซึ่งผู้วิจัยสร้างขึ้นแล้วนำข้อสอบที่วิเคราะห์หาคุณภาพแล้วมาจัดแบ่งเป็นแบบทดสอบ 2 ชุด ชุดละ 45 ข้อ และแต่ละชุดมีแบบสอบรวมภายใน 15 ข้อ นำไปดำเนินการสอบกลุ่มตัวอย่างอื่น จำนวน 1,557 คน โดยทุกคนต้องทำแบบทดสอบทั้ง 2 ชุด หลังจากนั้นจัดสุ่มใช้คะแนนผลการสอบจากแบบทดสอบชุดที่ 1 จำนวน 779 คน และอีก 778 คน ใช้คะแนนจากแบบทดสอบชุดที่ 2 ส่วนคะแนนที่เหลือใช้เป็นคะแนนตรวจทานผลของการเทียบมาตรฐานแต่ละรูปแบบ

ผลการวิจัยปรากฏว่าการเทียบมาตรฐานรูปแบบอิงทฤษฎีการตอบข้อสอบให้ความเพียงพอ ในระดับที่น่าพอใจ ส่วนรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบให้ความเพียงพอในระดับปานกลาง สำหรับการเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตรฐานรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบมีความคลาดเคลื่อนสูงกว่ารูปแบบอิงทฤษฎีการตอบข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (t - test) แต่เมื่อพิจารณาในแต่ละช่วงคะแนนพบว่ารูปแบบทั้งสองให้คะแนนสมมูลใกล้เคียงกันในช่วงตรงกลางของการแจกแจง

คะแนน (คะแนน 19 – 23) ส่วนช่วงคะแนนต่ำและคะแนนสูง พบว่าวิธีทั้งสองให้ผลแตกต่างกันทางตรงกันข้าม กล่าวคือในช่วงคะแนนต่ำ (คะแนน 1 – 18) นั้น รูปแบบอิงทฤษฎีการตอบข้อสอบมีคะแนนสมมูลสูงกว่ารูปแบบการใช้เทคนิคการวิเคราะห์ห้องค์ประกอบ ส่วนช่วงคะแนนสูง (คะแนน 24 – 35) รูปแบบการใช้เทคนิคการวิเคราะห์ห้องค์ประกอบ มีคะแนนสมมูลสูงกว่ารูปแบบการอิงทฤษฎีการตอบข้อสอบ

สุรัตน์ ขวัญบุญจันทร์ (2531 : 77 – 82) ได้ศึกษาการสร้างแบบทดสอบและตารางเทียบมาตรฐานคะแนนตามแนวนอนในวิชาคณิตศาสตร์ เรื่องพลังงานและการเปลี่ยนแปลงสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 โดยใช้วิธีการเทียบมาตรฐานรูปแบบราศี ผู้วิจัยนำแบบทดสอบที่สร้างขึ้นไปทดลองกับกลุ่มตัวอย่างแล้วคัดเลือกข้อสอบที่มีคุณภาพจำนวน 110 ข้อ แบ่งเป็น 2 ฉบับ ๆ ละ 60 ข้อ ซึ่งมีแบบสอบรวมภายในอยู่ฉบับละ 10 ข้อ นำไปสอบกับกลุ่มตัวอย่างเทียบมาตรฐานจำนวน 819 คน สอบแบบทดสอบคนละฉบับตามการสุ่มได้ฉบับที่ 1 408 คน ฉบับที่สอง 411 คน และนำไปสอบกับกลุ่มทนายผลจำนวน 166 คน ให้ทำแบบทดสอบทั้งสองฉบับ สร้างตารางเทียบมาตรฐานและวิเคราะห์ค่าดัชนีความเพียงพอของการเทียบมาตรฐานเท่ากับ .1019 ซึ่งอยู่ในระดับที่น่าพอใจมาก

สุนิสา จั๋ยม่วงศรี (2536 : บทคัดย่อ) ได้ตรวจสอบคุณภาพของการเทียบมาตรฐานเชิงเส้นตรง จากการใช้แบบทดสอบรวมภายในขนาดความยาวต่างกัน 4 ขนาด คือ 10 15 20 และ 25 ข้อ โดยการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน (SEE) และดัชนีความแตกต่าง (C) จากการวิเคราะห์กลุ่มสอบทานผล เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ปลายภาค ในวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่สอง ภาคต้น ของกลุ่มโรงเรียนมัธยมศึกษาตอนต้น จังหวัดขอนแก่น กลุ่มที่สอง จำนวน 2 ฉบับ ฉบับละ 60 ข้อ สร้างในปีการศึกษา 2532 (ฉบับ X) และ 2533 (ฉบับ Y) และแบบทดสอบรวมวัดผลสัมฤทธิ์ที่ผู้วิจัยสร้างขึ้น เพื่อใช้ในการเทียบมาตรฐาน จำนวน 25 ข้อ ได้แบบทดสอบที่ใช้ในการเทียบมาตรฐาน 2 ฉบับ คือ ฉบับ X และฉบับ Y กลุ่มตัวอย่างที่ใช้ในการศึกษามีสองกลุ่ม เป็นกลุ่มตัวอย่างเทียบมาตรฐาน เพื่อนำผลการสอบไปวิเคราะห์เทียบมาตรฐานคะแนนของแบบสอบ จำนวน 810 คน สุ่มอย่างมีระบบให้แต่ละคนสอบแบบสอบฉบับ X และฉบับ Y คนละฉบับสลับกัน ได้ผู้สอบฉบับละ 405 คน และเป็นกลุ่มสอบทานผล เพื่อนำผลการสอบไปตรวจสอบคุณภาพของวิธีการเทียบมาตรฐานเชิงเส้นตรง จำนวน 117 คน โดยให้สอบแบบทดสอบทั้งสองฉบับด้วยการสุ่มอย่างมีระบบ แต่ละคนสอบแบบทดสอบฉบับ X และฉบับ Y สลับกัน

ผลการวิจัยปรากฏว่าคะแนนสมมูลของแบบสอบฉบับ Y น้อยกว่าคะแนนของแบบฉบับ X จากการเทียบมาตรฐาน โดยใช้แบบสอบรวมทั้ง 4 ขนาด ส่วนค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานเชิงเส้นตรง ในแบบสอบรวมที่ยาวกว่าให้ค่าน้อยกว่า โดยพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของการใช้แบบทดสอบรวม 4 ขนาด ณ ระดับ

คะแนนมาตรฐาน (Z_v) เดียวกัน 7 ระดับ ทำให้ค่าประสิทธิภาพสัมพัทธ์ ที่ได้จากอัตราส่วน ร้อยละของความคลาดเคลื่อน มาตรฐาน เมื่อใช้แบบสอบรวมขนาด 10 15 และ 20 ข้อ เทียบ กับแบบสอบรวมขนาด 25 ข้อ ให้ค่า 86.26 89.08 และ 93.30 ตามลำดับ และค่าดัชนี ความแตกต่างของการเทียบมาตรา จากการใช้แบบทดสอบรวม 4 ขนาด ให้ผลอยู่ในระดับที่ ยอมรับได้ทั้งสิ้น คือ .4319 .3886 .3630 และ .3354 ตามลำดับ ซึ่งแตกต่างอย่างไม่มี นัยสำคัญทางสถิติที่ระดับ .05

วรเนตร พิชิตเกริกพล (2536 : บทคัดย่อ) ได้เปรียบเทียบคุณภาพของวิธีการ เทียบมาตราระหว่างแบบอิกวิเปอร์เซ็นไทล์กับแบบเชิงเส้นตรง โดยศึกษากรณีของลักษณะ แบบทดสอบที่มีความยากใกล้เคียงกันและแตกต่างกัน สอบกับกลุ่มตัวอย่างที่มีระดับความ สามารถใกล้เคียงกันและแตกต่างกัน การประเมินผลการเทียบมาตรา กระทำโดยวิเคราะห์กลุ่ม ตรวจสอบผล ซึ่งเป็นเกณฑ์ภายนอกเพื่อตรวจสอบความคลาดเคลื่อนของคะแนนที่ได้จากการ เทียบมาตราโดยตรงกับผลสุดท้ายของการเทียบมาตรา ประชากรที่ศึกษาเป็นนักเรียนชั้นมัธยม- ศึกษาปีที่ 2 ปีการศึกษา 2534 จังหวัดพระนครศรีอยุธยา จำนวน 6,752 คน โดยเป็นกลุ่ม ตัวอย่างสำหรับเทียบมาตรา จำนวน 900 คน และกลุ่มตัวอย่างตรวจสอบผล จำนวน 240 คน เครื่องมือที่ใช้ในการเทียบมาตราเป็นแบบสอบวิชาคณิตศาสตร์ จำนวน 3 ฉบับ ๆ ละ 40 ข้อ โดยมีข้อกระทงรวมฉบับละ 20 ข้อ การเทียบมาตราในแต่ละกรณีใช้วิธีเทียบมาตรา 2 วิธี คือ วิธีอิกวิเปอร์เซ็นไทล์และวิธีเชิงเส้นตรง แล้วสร้างตารางแปลงคะแนนสมมูลเพื่อวิเคราะห์ดัชนี ความแตกต่างมาตรฐาน (C) และความคลาดเคลื่อนของคะแนนที่ได้จากการเทียบมาตราจาก กลุ่มตัวอย่างตรวจสอบผล

ผลการวิจัยปรากฏว่า ในกรณีแบบทดสอบมีความยากใกล้เคียงกันและกลุ่มตัวอย่าง ที่มีระดับความสามารถใกล้เคียงกัน วิธีเชิงเส้นตรง ให้คุณภาพดีกว่า วิธีอิกวิเปอร์เซ็นไทล์ใน กรณีแบบทดสอบมีความยากใกล้เคียงกัน และกลุ่มตัวอย่างที่มีระดับความสามารถแตกต่างกัน วิธีอิกวิเปอร์เซ็นไทล์ให้คุณภาพดีกว่า วิธีเชิงเส้นตรง ในกรณีแบบทดสอบมีความยากแตกต่างกัน ในกลุ่มตัวอย่างที่มีระดับความสามารถใกล้เคียงกัน วิธีอิกวิเปอร์เซ็นไทล์ให้คุณภาพดีกว่า วิธีเชิงเส้นตรง และในกรณีแบบทดสอบมีความยากแตกต่างกัน ในกลุ่มตัวอย่างที่มีระดับความ สามารถแตกต่างกัน การเทียบทั้งสองวิธีมีความเพียงพออยู่ในระดับไม่น่าพอใจ

สมพงศ์ แก้วแจ่ม (2537 : บทคัดย่อ) ได้พัฒนาโปรแกรมคอมพิวเตอร์สำหรับการ เทียบคะแนนเชิงเส้นตรงตามแนวนอน โดยใช้แบบสอบรวม ตามโมเดลของ แองกออฟฟ์ ใน รูปแบบที่ 3 โดยใช้สูตรคำนวณการประมาณค่าเฉลี่ยและความแปรปรวน ตามวิธีของลอร์ด โดยมีขั้นตอนในการวิจัยและพัฒนาโปรแกรมคือ กำหนดปัญหา วิเคราะห์ปัญหา ออกแบบโปรแกรม เลือกภาษาและเขียนโปรแกรม ทดลองใช้โปรแกรมทดสอบและประเมินประสิทธิภาพ และจัดทำ คู่มือการใช้โปรแกรม ผลการวิจัยและพัฒนาครั้งนี้ ทำให้ได้โปรแกรมคอมพิวเตอร์สำเร็จรูปเพื่อ

ใช้สำหรับการเทียบคะแนนเชิงเส้นตรงตามแนวนอน โดยใช้แบบทดสอบร่วม (TELEQUAT) และคู่มือการใช้โปรแกรม โดยโปรแกรมที่พัฒนาในครั้งนี้ เขียนด้วยภาษาบอร์แลนด์ปาสคาล เวอร์ชัน 7.0 และคอมไพล์ให้เป็นไฟล์ .EXE เพื่อสามารถเรียกโปรแกรมทำงานได้เร็วขึ้น โดยโปรแกรมนี้สามารถประมวลผลค่าสถิติพื้นฐาน, คะแนนแปลงจากแบบสอบเทียบมาตรฐานฉบับหนึ่ง (ฟอร์ม Y) ไปยังแบบสอบเทียบมาตรฐานอีกฉบับหนึ่ง (ฟอร์ม X) ความแปรปรวนของความคลาดเคลื่อนในการเทียบมาตรฐาน (SE^2_Y) ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนในการเทียบมาตรฐาน (SE_Y), คะแนนมาตรฐานที (T-Score) และคะแนนมาตรฐานซี (Z-Score) การรายงานผลสามารถรายงานได้บนทั้งจอภาพและเครื่องพิมพ์ โดยรายงานผลในรูปตาราง จากการประเมินผลการใช้โปรแกรมพบว่า โปรแกรมมีความถูกต้องในการสั่งงาน สะดวกและง่ายในการนำไปใช้ มีความเร็วในการประมวลผลข้อมูลและมีประสิทธิภาพในการทำงาน และผลการประมวลมีความถูกต้องแม่นยำ

พิชัย ละแมนชัย (2538 : บทคัดย่อ) ได้ศึกษาเพื่อหาขนาดกลุ่มตัวอย่างขั้นต่ำที่ทำการปรับเทียบคะแนนในแนวระดับ ตามแนวทฤษฎีการตอบสนองรายข้อ โมเดล หนึ่ง และสาม พารามิเตอร์ ในแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วม ที่มีความยาวแบบสอบต่าง ๆ กัน คือ 30, 50, 70, 90, 110, 130 และ 150 ข้อ เพื่อให้ผลการปรับเทียบคะแนนในแนวระดับเกิดประสิทธิภาพ ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นการจำลองจากโปรแกรม IRTDATA เกณฑ์ที่ใช้ในการตัดสินใจคือการหาค่าดัชนีความแตกต่างที่อยู่ในระดับน่าพอใจอย่างมาก ผลการวิจัยปรากฏว่า

1. ในแบบแผนกลุ่มสมมูลขนาดกลุ่มตัวอย่างขั้นต่ำตามโมเดล 1 พารามิเตอร์ ที่มีความยาวแบบทดสอบเป็น 30, 50, 70, 90, 110, 130 และ 150 ข้อ คือ 1000, 850, 700, 600, 550, 450 และ 350 ตามลำดับ และขนาดกลุ่มตัวอย่างขั้นต่ำตามโมเดล 3 พารามิเตอร์ ที่มีความยาวแบบทดสอบเป็น 30, 50, 70, 90, 110, 130 และ 150 ข้อ คือ 2550, 2350, 2150, 1750, 1550, 1400 และ 1150 ตามลำดับ

2. ในแบบแผนข้อสอบร่วมขนาดกลุ่มตัวอย่างขั้นต่ำตามโมเดล 1 พารามิเตอร์ ที่มีความยาวแบบทดสอบเป็น 30, 50, 70, 90, 110, 130 และ 150 ข้อ คือ 750, 650, 500, 400, 350, 300 และ 200 ตามลำดับ และขนาดกลุ่มตัวอย่างขั้นต่ำตามโมเดล 3 พารามิเตอร์ ที่มีความยาวแบบทดสอบเป็น 30, 50, 70, 90, 110, 130 และ 150 ข้อ คือ 2000, 1800, 1600, 1350, 1200, 1000 และ 800 ตามลำดับ

3. ทั้งแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วมในโมเดลการตอบสนองรายข้อ 1 และ 3 พารามิเตอร์ ขนาดกลุ่มตัวอย่างขั้นต่ำในแต่ละความยาวแบบทดสอบแตกต่างกัน โดยโมเดล 1 พารามิเตอร์ใช้ขนาดกลุ่มตัวอย่างขั้นต่ำที่น้อยกว่า และขนาดกลุ่มตัวอย่างขั้นต่ำในแต่ละความยาวแบบทดสอบแตกต่างกัน โดยแบบแผนข้อสอบร่วมใช้ขนาดกลุ่มตัวอย่างขั้นต่ำที่น้อยกว่าแบบแผนกลุ่มสมมูล

นิภาพร ยอดเมือง (2539 : บทคัดย่อ) ได้ศึกษาความแตกต่างของความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน ความลำเอียงทางสถิติในการเทียบคะแนน และคะแนนที่ถูกเทียบโดยเฉลี่ย เมื่อขนาดของกลุ่มตัวอย่างเป็น 30 60 120 และ 250 คน ประชากรที่ใช้ในการศึกษาเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2538 สังกัดกรุงเทพมหานคร เขตบางเขน เขตดอนเมือง เขตจตุจักร จำนวนนักเรียนทั้งสิ้น 1,826 คน เครื่องมือเป็นแบบทดสอบวัดการรับรู้ 3 ฉบับ โดยจำแนกเป็นแบบทดสอบเทียบคะแนน 2 ฉบับ เป็นแบบทดสอบเชื่อมโยง 1 ฉบับ ประชากรที่ใช้ในการศึกษาจำแนกเป็น 2 กลุ่ม กลุ่มที่ 1 ทำแบบทดสอบเทียบคะแนนฉบับที่หนึ่งกับแบบทดสอบเชื่อมโยง กลุ่มที่ 2 ทำแบบทดสอบเทียบคะแนนฉบับที่สองกับแบบทดสอบเชื่อมโยง

ผลการวิจัยปรากฏว่า คะแนนที่ถูกเทียบโดยเฉลี่ยของกลุ่มตัวอย่างขนาด 30 60 120 และ 250 คน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความลำเอียงทางสถิติในการเทียบคะแนน ในกลุ่มตัวอย่างขนาด 30 60 120 และ 250 คน แตกต่างกันอย่างน้อยมาก ค่าความลำเอียงทางสถิติในการเทียบคะแนนของกลุ่มตัวอย่างขนาด 60 และ 250 คน มีค่าเท่ากับ 0 เมื่อคะแนนสังเกตเข้าใกล้คะแนนเฉลี่ยและค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนในกลุ่มตัวอย่างขนาด 30 60 120 และ 250 คน แตกต่างกันอย่างเล็กน้อย เมื่อคะแนนสังเกตใกล้ค่าเฉลี่ย (38.8204) ดังนั้นจากผลการวิจัยผู้วิจัยได้สรุปว่า กลุ่มตัวอย่างที่สามารถนำไปใช้ในการเทียบคะแนนได้คือกลุ่มตัวอย่างทั้ง 4 ขนาด คือ ขนาด 30 60 120 และ 250 คน แต่ถ้าต้องการให้ค่าความคลาดเคลื่อนมาตรฐาน และความลำเอียงทางสถิติในการเทียบคะแนนมีค่าน้อยก็สมควรใช้กลุ่มตัวอย่างขนาด 250 คน

สมนา โสติกผลอนันต์ (2539 : บทคัดย่อ) ได้เปรียบเทียบวิธีการเทียบคะแนนแบบเส้นตรงที่ได้จากการนำคะแนนสอบมาคำนวณด้วยสูตรสี่วิธี โดยมีข้อตกลงว่า แบบทดสอบที่นำมาเทียบคะแนนนั้น ต้องเป็นไปตามแบบจำลองคะแนนจริงสัมพันธ์ และมีความเชื่อมั่นเท่าเทียมกัน กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักศึกษาพยาบาลชั้นปีที่ 1 และชั้นปีที่ 2 หลักสูตรพยาบาลศาสตรและพยาบาลศาสตรระดับต้น ปีการศึกษา 2538 ของวิทยาลัยพยาบาลสังกัดสถาบันพัฒนากำลังคนด้านสาธารณสุข สำนักงานปลัดกระทรวงสาธารณสุข กระทรวงสาธารณสุข ในเขตภาคกลาง จำนวน 864 คน ได้มาจากวิธีการสุ่มอย่างง่าย มาจัดเป็นกลุ่มเทียบคะแนน 12 กลุ่ม ที่มีความสามารถทางสมองเท่าเทียมกัน ทำการทดสอบด้วยแบบทดสอบวัดการคิดเอกลักษณะสัมพันธ์ที่มีฟอร์มต่างกันสามฟอร์ม คือ ฟอร์ม A ฟอร์ม B และฟอร์ม C โดยใช้ฟอร์ม C จัดทำเป็นแบบทดสอบหลักที่มีลักษณะการใช้ของแบบทดสอบหลัก ที่เป็นแบบทดสอบเชื่อมโยงและเป็นข้อสอบรวม และมีขนาดความยาวสามขนาด คือคิดเป็นสัดส่วน .2, .3 และ .4 ของแบบทดสอบเทียบคะแนนฟอร์ม A แล้วนำคะแนนสอบมาวิเคราะห์ทางสถิติ โดยการแปลงคะแนนดิบให้เป็นคะแนนแปลง ด้วยวิธีการคำนวณจากสูตรสี่วิธี แล้วจึงเปรียบเทียบ

คะแนนแปลงและความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน ด้วยวิธีการหาความแปรปรวนแบบการวัดซ้ำ

ผลการศึกษาปรากฏว่า คะแนนแปลงที่ได้จากวิธีการคำนวณด้วยสูตรส่ววิธี ให้ผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เมื่อเปรียบเทียบผลของการใช้แบบทดสอบเชื่อมโยงกับข้อสอบร่วมพบว่ามีผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 รวมทั้งผลของขนาดความยาวของแบบทดสอบหลักก็พบว่ามีผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทำนองเดียวกันผลของการเปรียบเทียบความคลาดเคลื่อนมาตรฐานที่เกิดจากการใช้วิธีการคำนวณด้วยสูตรส่ววิธีก็พบว่ามีผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ด้วยเช่นกัน

พรโสภา พินิจผล (2539 : บทคัดย่อ) ได้เปรียบเทียบผลการเทียบคะแนนในแบบทดสอบวิชาการพยาบาลสูติศาสตร์ตามทฤษฎีการตอบข้อสอบที่ใช้วิธีการวิเคราะห์ 2 วิธี คือวิธีการเทียบคะแนนจริงโดยการใช้แบบทดสอบร่วม และวิธีการเทียบคะแนนดิบโดยการใช้แบบทดสอบร่วม กลุ่มตัวอย่างเป็นนักศึกษาพยาบาลศาสตร์ (ระดับต้น) ชั้นปีที่ 2 ปีการศึกษา 2538 ของวิทยาลัยพยาบาลในสังกัดสถาบันพัฒนากำลังคนด้านสาธารณสุข กระทรวงสาธารณสุข ในเขตภาคเหนือ ภาคกลางและภาคตะวันออก จำนวน 980 คน ซึ่งได้มาโดยวิธีการสุ่มแบบง่าย เครื่องมือที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาการพยาบาลสูติศาสตร์ 1 ที่ผู้วิจัยสร้างขึ้นจากการประมาณค่าพารามิเตอร์ของข้อสอบ ตามทฤษฎีการตอบข้อสอบ โดยใช้โปรแกรม BILOG เวอร์ชัน 3.04 แล้วแบ่งออกเป็น 2 ฉบับ ๆ ละ 10 ข้อ

ผลการวิจัยสรุปได้ว่า คะแนนที่ได้จากการเทียบคะแนนตามวิธีการเทียบคะแนนจริงกับวิธีการเทียบคะแนนดิบแตกต่างกัน อย่างไม่มีนัยสำคัญทางสถิติ โดยคะแนนความคลาดเคลื่อนที่ได้จากการเทียบคะแนนตามวิธีการเทียบคะแนนจริงกับวิธีการเทียบคะแนนดิบแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และความเพี่ยงพ้อของการเทียบคะแนนตามวิธีการเทียบคะแนนจริงกับวิธีการเทียบคะแนนดิบ แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ นอกจากนี้ผู้วิจัยยังได้เสนอแนะวิธีการเทียบคะแนนโดยวิธีการเทียบคะแนนจริง เพราะให้คะแนนความคลาดเคลื่อนและความเพี่ยงพ้อของการเทียบคะแนนไม่แตกต่างกับวิธีการเทียบคะแนนดิบ ซึ่งเป็นวิธีการที่มีการคำนวณที่ซับซ้อนและลงทุนมากกว่า

ดรุณี พรประเสริฐ (2541 : บทคัดย่อ) ได้สร้างตารางเทียบคะแนนผลสัมฤทธิ์ทางการเรียนตามทฤษฎีคลาสสิกอล โดยใช้รูปแบบอิกคิวเปอร์เซ็นไทล์และรูปแบบเชิงเส้นตรง และตามทฤษฎีการตอบสนองข้อสอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2540 สังกัดสำนักงานการประถมศึกษาจังหวัดนนทบุรี ที่เข้าสอบตามโครงการประเมินประสิทธิภาพและคุณภาพการศึกษาของสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ จำนวน 2 กลุ่ม ๆ ละ 380 คน โดยการสุ่มอย่างง่าย เครื่องมือที่ใช้คือแบบทดสอบคณิตศาสตร์

(การคิดคำนวณ) ระดับประเทศ และแบบทดสอบคณิตศาสตร์ (การคิดคำนวณ) ระดับจังหวัด โดยทั้งสองฉบับมีข้อสอบรวม 7 ข้อ ข้อมูลในการวิจัยนี้ใช้ข้อมูลจากสถานการณ์การสอบจริง พร้อมกันทั้งประเทศ

ผลการวิจัยปรากฏว่า การเทียบคะแนนตามทฤษฎีคลาสสิกคอลโดยใช้การเทียบตาม อิกวิปอร์เซ้นไทล์และการเทียบเชิงเส้นตรง ปรากฏว่าจะแนนสังเกตของแบบทดสอบทั้งสอง ฉบับสามารถเทียบเท่ากันได้ในบางค่า สำหรับการเทียบคะแนนตามทฤษฎีการตอบสนอง ข้อสอบพบว่า ค่าพารามิเตอร์ความยากของแบบทดสอบระดับประเทศมีค่าตั้งแต่ -0.271 กับ 1.962 โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานเป็น 0.537 และ $.692$ ตามลำดับ ส่วนค่า พารามิเตอร์ความยากของแบบทดสอบระดับจังหวัดมีค่าตั้งแต่ -0.024 ถึง 2.389 โดยมีค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานเป็น $.972$ และ $.562$ ตามลำดับ การปรับค่าพารามิเตอร์ความยาก ของแบบทดสอบระดับจังหวัดให้เป็นสเกลเดียวกับค่าพารามิเตอร์ความยากของแบบทดสอบ ระดับประเทศ ปรากฏว่าค่าพารามิเตอร์ความยากของข้อสอบรวมทั้ง 7 ข้อ มีค่าตั้งแต่ 0.540 ถึง 1.471 ส่วนค่าพารามิเตอร์ความยากของข้อสอบที่ไม่ใช่ข้อสอบรวมของแบบทดสอบระดับ จังหวัดมีค่าตั้งแต่ -1.178 ถึง 1.725 และในการเทียบค่าความสามารถจริงของผลการสอบจาก แบบทดสอบทั้ง 2 ฉบับ ปรากฏว่าค่าความสามารถจริงของผลการสอบแบบทดสอบระดับ ประเทศมีค่ามากที่สุด 24.230851 และน้อยที่สุด 1.654381 ส่วนค่าความสามารถจริงของ ผลการสอบแบบทดสอบระดับจังหวัดมีค่ามากที่สุดและน้อยที่สุด 23.96726 และ 2.254844 ตาม ลำดับ

ทิวาภรณ์ ทุงฤทธิ์ (2544 : บทคัดย่อ) ได้เปรียบเทียบคะแนนที่เทียบได้ ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน และความเพียงพอของการเทียบคะแนนใน แนวนอน โดยใช้วิธีเชิงเส้นตรง เมื่อใช้การเทียบคะแนนที่ใช้วิธีการแปลงคะแนนที่แตกต่างกัน 3 วิธี คือ 1) การแปลงคะแนนจากการสุ่มกลุ่มนักเรียนให้สอบ โดยใช้แบบทดสอบกลุ่มละฉบับ 2) โดยใช้แบบทดสอบทั้งสองฉบับที่เรียงอันดับการสอบต่างกัน 3) โดยใช้แบบทดสอบกลุ่มละ ฉบับและมีแบบทดสอบรวมที่ต้องสอบทั้งสองกลุ่ม เครื่องมือที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ ทางการเรียนรู้วิชาคณิตศาสตร์ เรื่อง สมการและการแก้สมการ จำนวน 3 ฉบับ ประกอบด้วย ฉบับ X ฉบับ Y และฉบับรวม มีกลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2542 ของโรงเรียนในสังกัดสำนักงานการประถมศึกษาจังหวัดปราจีนบุรี จำนวน 820 คน ซึ่งได้ มาจากการสุ่มแบบแบ่งชั้น ผลการวิจัยสรุปได้ดังนี้ คะแนนที่ได้จากการเทียบคะแนน 3 วิธี และ ความเพียงพอของการเทียบคะแนน 3 วิธี แตกต่างกันอย่างมีนัยสำคัญทางสถิติ สำหรับ ความคลาดเคลื่อนมาตรฐานแตกต่างอย่างไม่มีนัยสำคัญทางสถิติ

อุทัยวรรณ พงศ์อร่าม (2545 : บทคัดย่อ) ได้ศึกษาขนาดของกลุ่มตัวอย่างที่ เหมาะสมสำหรับการเปรียบเทียบคะแนนด้วยวิธีอิกวิปอร์เซ้นไทล์และวิธีเชิงเส้นตรงตามแบบ

จำลองคะแนนจริงสัมพันธ์ในแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบรวมที่มีความยาวของแบบทดสอบต่างกัน คือ 60 50 40 และ 30 ข้อ กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ของโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา จังหวัดชุมพร ภาคเรียนที่ 1 ปีการศึกษา 2544 จำนวน 3,552 คน ซึ่งได้มาโดยการสุ่มสองขั้นตอน แบ่งเป็นสองกลุ่ม คือ กลุ่มตัวอย่างเปรียบเทียบคะแนนและกลุ่มตัวอย่างสอบทานผล เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ ค 101 ชั้นมัธยมศึกษาปีที่ 1 จำนวนสองฉบับ การวิเคราะห์ข้อมูลใช้วิธีการเปรียบเทียบคะแนนด้วยวิธีอีคิวเปอร์เซ็นต์ไทล์และวิธีเชิงเส้นตรงตามแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบรวม เพื่อสร้างตารางแปลงคะแนนสมมูลและวิเคราะห์ดัชนีความแตกต่าง (C) หาขนาดกลุ่มตัวอย่างที่เหมาะสมโดยใช้เกณฑ์ตัดสินจากค่าดัชนีความแตกต่าง (C) ที่อยู่ในระดับน่าพอใจและน่าพอใจมาก ผลการวิจัยพบว่า

1. ขนาดของกลุ่มตัวอย่างที่เหมาะสมในการเปรียบเทียบคะแนนวิธีเชิงเส้นตรง ในแบบแผนกลุ่มสมมูล ที่มีความยาวของแบบทดสอบเป็น 60 50 และ 40 ข้อ เป็น 100 คน ส่วน 30 ข้อ เป็น 200 คน สำหรับการเปรียบเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ในแบบแผนกลุ่มสมมูลที่มีความยาวของแบบทดสอบเป็น 60 ข้อ มีขนาดกลุ่มตัวอย่างที่เหมาะสมเป็น 200 คน และ 50 กับ 40 ข้อเป็น 300 คน ส่วน 30 ข้อ เป็น 400 คน

2. ขนาดของกลุ่มตัวอย่างที่เหมาะสมในการเปรียบเทียบคะแนนวิธีเชิงเส้นตรง ในแบบแผนข้อสอบรวม ที่มีความยาวของแบบทดสอบเป็น 60 50 และ 40 ข้อ เป็น 100 คน ส่วน 30 ข้อ เป็น 300 คน สำหรับการเปรียบเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ในแบบแผนข้อสอบรวม ที่มีความยาวของแบบทดสอบเป็น 60 และ 50 ข้อ เป็น 200 คน ส่วน 40 และ 30 เป็น 500 คน

3. การเปรียบเทียบคะแนนด้วยวิธีเชิงเส้นตรงและวิธีอีคิวเปอร์เซ็นต์ไทล์ทั้งแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบรวม ขนาดของกลุ่มตัวอย่างที่เหมาะสมแตกต่างกันในแต่ละความยาวของแบบทดสอบ โดยการเปรียบเทียบคะแนนด้วยวิธีเชิงเส้นตรงใช้กลุ่มตัวอย่างที่น้อยกว่า

4. ตามแบบแผนการเปรียบเทียบคะแนนแบบกลุ่มสมมูลและแบบแผนข้อสอบรวมทั้งวิธีเชิงเส้นตรงและวิธีอีคิวเปอร์เซ็นต์ไทล์ ขนาดของกลุ่มตัวอย่างที่เหมาะสมในแต่ละความยาวของแบบทดสอบไม่แตกต่างกัน

นอกจากนี้ อติสร ศรีบุญวงศ์ (2545 : บทคัดย่อ) ได้พัฒนาเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบและตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้นด้วยการหาความสอดคล้องของผลการตัดสินคุณภาพการเปรียบเทียบคะแนนด้วยเกณฑ์ที่พัฒนาขึ้นกับเกณฑ์ความเสมอภาคของลอว์ด และเกณฑ์ของปีเตอร์เซ็นและคณะ โดยใช้ข้อมูลจำลองจากโปรแกรม IRTDATA ที่ตรวจสอบความถูกต้องโดยการวิเคราะห์ซ้ำด้วยโปรแกรม BILOG การจำลองข้อมูลตามเงื่อนไขการเปรียบเทียบคะแนนด้วยโมเดลตอบสนองข้อสอบแบบ 1 และ 3 พารามิเตอร์ วิเคราะห์คุณภาพของการเปรียบเทียบคะแนนด้วยการปรับเทียบกลับสู่แบบ-

สอบเดิมและการใช้กลุ่มสอบทานผล แบบแผนการเก็บรวบรวมโดยใช้กลุ่มสมมูลและการใช้ข้อสอบร่วม วิธีปรับเทียบคะแนน 4 วิธี คือ วิธี The Same Scaling Convention วิธี Mean and Mean วิธี Mean and Sigma และวิธี Characteristic Curve จำนวนผู้สอบตั้งแต่ 100, 200,....., 3000 คน และเงื่อนไขจำนวนข้อสอบ 10, 20,....., 200 ข้อ จัดกลุ่มดัชนีความแตกต่าง AMD MAD และ RMS โดยใช้จุดตัด ซึ่งเป็นค่าดัชนีความแตกต่างเมื่อทดสอบความแตกต่างระหว่างค่าเฉลี่ยของคะแนนจริงที่ปรับเทียบแล้วกับคะแนนจริงที่ยังไม่ได้ปรับเทียบ สำหรับผู้สอบกลุ่มเดียวกันด้วยสถิติ t-test แล้วพบค่าวิกฤตที่ระดับนัยสำคัญทางสถิติที่ระดับ .001 .01 และ .02 คำนวณค่าดัชนีความแตกต่าง แล้วนำกลุ่มค่าดัชนีความแตกต่างแต่ละกลุ่มมาระบุระดับคุณภาพการปรับเทียบเพื่อกำหนดเป็นเกณฑ์

ผลการวิจัยสรุปได้ว่า เกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบที่พัฒนาขึ้นมีทั้งหมด 7 เกณฑ์ คือ เกณฑ์รวมทุกเงื่อนไข เกณฑ์สำหรับโมเดล 1 พารามิเตอร์ เกณฑ์สำหรับโมเดล 3 พารามิเตอร์ เกณฑ์สำหรับการใช้กลุ่มสมมูล เกณฑ์สำหรับการใช้แบบสอบร่วม เกณฑ์สำหรับการปรับเทียบกลุ่มสู่แบบสอบเดิม และเกณฑ์สำหรับการใช้กลุ่มสอบทานผล แต่ละเกณฑ์เป็นกลุ่มค่าดัชนี AMD MAD และ RMS ที่บอกระดับคุณภาพการปรับเทียบคะแนน 4 ระดับ คือ การปรับเทียบน่าพอใจอย่างยิ่ง การปรับเทียบน่าพอใจ การปรับเทียบไม่น่าพอใจ และการปรับเทียบไม่น่าพอใจอย่างยิ่ง ผลการตัดสินคุณภาพการปรับเทียบคะแนนเมื่อใช้เกณฑ์ที่พัฒนาขึ้นปรับเทียบกับผลการใช้ตามเกณฑ์ของปีเตอร์เซนและคณะ พบว่า ไม่สอดคล้องกัน แต่เมื่อเทียบกับเกณฑ์ความเสมอภาคของลอว์ด ผลการตัดสินคุณภาพการปรับเทียบมีความสอดคล้องกัน

นอกจากนี้ยังมีการศึกษาการเทียบคะแนนตามแนวทางตั้ง ในกลุ่มผู้สอบที่มีความสามารถแตกต่างกัน ซึ่ง อากรณ์ กาญจนกิจโสภณ (2531 : 66 – 67) ได้ศึกษาการสร้างแบบทดสอบและตารางเทียบมาตรฐานคะแนนตามแนวทางตั้งในวิชาคณิตศาสตร์ เรื่องสมการและสมการ สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้นโดยใช้วิธีการเทียบมาตรฐานแบบราสส์ กลุ่มตัวอย่างที่ใช้เทียบมาตรฐานเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 2 และ 3 จำนวนระดับชั้นละ 490 482 และ 415 คน ตามลำดับ ผู้วิจัยได้นำแบบสอบที่สร้างขึ้นระดับชั้นละ 90 ข้อ ไปทดลองกับกลุ่มตัวอย่างอื่นแล้ว คัดเลือกข้อสอบที่มีคุณภาพเรียงตามค่าความยากที่ต้องการ ระดับชั้นละ 40 30 และ 30 ข้อ โดยแต่ละชั้นจะมีข้อสอบรวมระหว่างชั้นอยู่ 10 ข้อ หลังจากนั้นนำไปทดสอบกับกลุ่มตัวอย่างที่ใช้เทียบมาตรฐาน และสร้างตารางเทียบมาตรฐานเพื่อใช้ศึกษาพัฒนาการของนักเรียนกลุ่มเดิมที่ผ่านขึ้นไปเรียนในชั้นสูงขึ้นไปว่า นักเรียนมีความก้าวหน้าเป็นไปตามปกติหรือเบี่ยงเบนไปจากปกติ

พรพิมล นาคเวช (2537 : 99 – 101) ได้ศึกษาคุณภาพของการเทียบคะแนนใน แนวตั้งโดยใช้ทฤษฎีการตอบสนองข้อสอบ สำหรับผู้มีความสามารถต่างกัน 3 ระดับ ซึ่งได้แก่นักเรียนชั้น ม.1 ม.2 และ ม.3 โดยวิเคราะห์ความคลาดเคลื่อนมาตรฐานและความเพียงพอของการเทียบคะแนน กลุ่มตัวอย่างเป็นนักเรียนชั้น ม.1 ม.2 และ ม.3 สังกัดกองการมัธยมศึกษา กรมสามัญศึกษา จำนวน 6,429 คน ที่ได้รับการสุ่มกลุ่มตัวอย่างแบบแบ่งชั้นหลายชั้นตอน แบ่งเป็นกลุ่มตัวอย่างเทียบคะแนนเพื่อนำผลไปเทียบคะแนนระหว่างแบบทดสอบ จำนวน 3,279 คน และกลุ่มตัวอย่างสอบทานผลเพื่อนำผลไปตรวจสอบคุณภาพของการเทียบคะแนนจำนวน 3,105 คน เครื่องมือที่ใช้เป็นแบบทดสอบคณิตศาสตร์ เรื่องสมการและอสมการ สำหรับระดับชั้น ม.1 ม.2 และ ม.3 รวม 3 ฉบับ ๆ ละ 25 ข้อ โดยมีข้อสอบร่วมระหว่างระดับชั้นที่ติดกันระดับชั้นละ 5 ข้อ ผลการศึกษาพบว่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนมีค่าอยู่ระหว่าง .031 - .051 ซึ่งเป็นความคลาดเคลื่อนมาตรฐานที่ค่อนข้างต่ำ และผลการวิเคราะห์ความเพียงพอของการเทียบคะแนนตามแนวตั้งพบว่า ค่าดัชนีความแตกต่างของการเทียบคะแนนอยู่ในระดับที่น่าพอใจ

วรรณดี แสงประทีปทอง (2538 : บทคัดย่อ) ได้ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.1 ม.2 และ ม.3 โดยใช้วิธีการเทียบคะแนนตามแนวตั้ง ตามทฤษฎีการตอบสนองข้อสอบ แบบจำลองโลจิสติก สามพารามิเตอร์ การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนศึกษาทั้งในกลุ่มรวม และในกลุ่มย่อยตามระดับความสามารถของนักเรียน กลุ่มตัวอย่างการวิจัยเป็นนักเรียนชั้น ม.1 ม.2 และ ม.3 ของโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษาในจังหวัดสิงห์บุรี ภาคเรียนที่ 1 ปีการศึกษา 2537 กลุ่มตัวอย่างแบ่งเป็น 2 กลุ่มคือกลุ่มตัวอย่างเทียบคะแนน เป็นนักเรียนชั้น ม.1 ม.2 และ ม.3 จำนวน 1,245 1,280 และ 1,171 คน ตามลำดับ และกลุ่มตัวอย่างสอบทานผลเป็นนักเรียนชั้น ม.1 ม.2 และ ม.3 จำนวน 373 384 และ 352 คนตามลำดับ เครื่องมือที่ใช้ในการวิจัยประกอบด้วยแบบทดสอบเทียบคะแนน 3 ฉบับ ซึ่งเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทย ท 101 (ม.1) ท 203 (ม.2) และ ท 305 (ม.3) และแบบทดสอบรวม ซึ่งเป็นแบบทดสอบความสามารถทั่วไปทางภาษาไทย 1 ฉบับ ข้อมูลจากการวิจัยรวบรวมจากสถานการณ์การสอบจริง ของการสอบปลายภาคเรียนที่ 1 ปีการศึกษา 2537 พร้อมกันทุกระดับชั้น การวิเคราะห์ข้อมูลใช้วิธีการเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ แบบจำลองโลจิสติก สามพารามิเตอร์ เพื่อเทียบคะแนนจากการสอบของนักเรียนชั้น ม.1 ม.2 และ ม.3 จากแบบทดสอบต่างฉบับให้อยู่ในสเกลเดียวกัน แล้วเปรียบเทียบผลสัมฤทธิ์ทางการเรียนของนักเรียนแต่ละชั้นภายหลังจากการเทียบคะแนนนักเรียนแต่ละชั้นให้อยู่ในสเกลเดียวกันแล้วเพื่อศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียน ผลการวิจัยปรากฏว่า

1. ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.3 สูงกว่า ม.2 แต่ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.2 และ ม.1 ไม่แตกต่างกัน หรือกล่าวได้ว่า ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.2 และ ม.3 ก้าวหน้ามากกว่าปกติ แต่ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.1 และ ม.2 ก้าวหน้าตามปกติ

2. ในกลุ่มนักเรียนที่มีความสามารถสูง ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.3 และ ม.2 ไม่แตกต่างกัน และผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.2 และ ม.1 ไม่แตกต่างกัน หรือกล่าวได้ว่าผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม.2 และ ม.3 ก้าวหน้าตามปกติ และผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.1 ไป ม.2 ก้าวหน้าตามปกติ

3. ในกลุ่มนักเรียนที่มีความสามารถปานกลาง ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.3 สูงกว่า ม.2 และผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.2 สูงกว่า ม.1 หรือกล่าวได้ว่าผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.2 และ ม.3 ก้าวหน้ามากกว่าปกติ และผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.1 ไป ม.2 ก้าวหน้ามากกว่าปกติ

4. ในกลุ่มนักเรียนที่มีความสามารถต่ำ ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.3 สูงกว่า ม.2 และผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.2 และ ม.1 ไม่แตกต่างกัน หรือกล่าวได้ว่าผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.2 ไป ม.3 ก้าวหน้ามากกว่าปกติ แต่ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.1 ไป ม.2 ก้าวหน้าตามปกติ

ภัทราพร เกษสังข์ (2546 : ปทคัดย่อ) ได้ศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนและประสิทธิภาพของการเทียบคะแนนในแนวตั้งตามทฤษฎีการตอบข้อสอบแบบไลจีสมติสองและสามพารามิเตอร์ ที่ใช้แบบทดสอบร่วมภายในและแบบทดสอบเชื่อมโยงภายนอกที่หลักระดับความยาก (ยากและปานกลาง) และความยาว (15 10 และ 5 ข้อ) ของแบบทดสอบร่วม จากกลุ่มผู้สอบที่มีความสามารถแตกต่างกันสามระดับชั้น คือมัธยมศึกษาปีที่ 1 2 และ 3 โดยการวิเคราะห์ความคลาดเคลื่อนมาตรฐานและความเพียงพอของการเทียบคะแนนแต่ละรูปแบบที่ศึกษา แล้วนำรูปแบบการเทียบคะแนนที่เหมาะสมมาใช้ในการศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียน กลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนระดับมัธยมศึกษาปีที่ 1 2 และ 3 สังกัดกองการมัธยมศึกษา กรมสามัญศึกษา ในจังหวัดชัยนาท จำนวน 4,041 คน ที่ได้จากการสุ่มกลุ่มตัวอย่างแบบแบ่งชั้น แบ่งกลุ่มตัวอย่างเป็นสองกลุ่ม กลุ่มแรก จำนวน 3,081 คน ใช้สำหรับศึกษาการเทียบคะแนน เพื่อนำผลไปเทียบคะแนนระหว่างแบบทดสอบ และศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ กลุ่มที่สอง จำนวน 960 คน ใช้สำหรับสอบทานผล เพื่อนำผลไปตรวจสอบประสิทธิภาพของการเทียบคะแนน เครื่องมือวัดที่ใช้เป็นแบบทดสอบคณิตศาสตร์ เรื่องสมการและอสมการ ที่มีความเชื่อมั่น

ของระดับชั้นมัธยมศึกษาปีที่ 1 2 และ 3 เท่ากับ 0.745 0.945 และ 0.951 ตามลำดับ และแบบทดสอบวัดความถนัดด้านตัวเลขที่มีความเชื่อมั่นของระดับชั้นมัธยมศึกษาปีที่ 1 2 และ 3 เท่ากับ 0.792 0.847 และ 0.708 ตามลำดับ

ผลการวิจัยสรุปได้ว่า ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนตามรูปแบบที่ศึกษาสำหรับผู้ที่มีความสามารถต่างกันสามระดับ พบว่า มีค่าระหว่าง 0.1011 ถึง 0.1830 เป็นค่าระดับต่ำ ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนต่ำที่สุด เป็นการเทียบคะแนนในแนวตั้ง ตามทฤษฎีการตอบข้อสอบรูปแบบโลจิสติกสองพารามิเตอร์ ที่แบบทดสอบรวมภายใน 15 ข้อ มีความยากของข้อสอบรวมเฉลี่ยอยู่ในระดับยากมาก ($\bar{x} = 2.2$) สำหรับความเพียงพอของการเทียบคะแนน พบว่า ค่าดัชนีความแตกต่างของการเทียบคะแนน (C) อยู่ในระดับน่าพอใจมากและระดับน่าพอใจ โดยการเทียบคะแนนในแนวตั้งตามทฤษฎีการตอบข้อสอบรูปแบบโลจิสติกสองพารามิเตอร์ที่มีแบบทดสอบรวมภายใน 15 ข้อ และมีลักษณะความยากของข้อสอบรวมเฉลี่ยระดับยากมาก ให้ความเพียงพอของการเทียบคะแนนดีที่สุด และพัฒนาความสามารถทางคณิตศาสตร์จากรูปแบบการเทียบคะแนนที่มีประสิทธิภาพ พบว่าความสามารถทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ไปชั้นมัธยมศึกษาปีที่ 3 มีพัฒนาการสูงขึ้น และนักเรียนชั้นมัธยมศึกษาปีที่ 1 ไปชั้นมัธยมศึกษาปีที่ 3 มีพัฒนาการสูงขึ้นมาก

สรุปได้ว่าผลการวิจัยเกี่ยวกับการเทียบคะแนนทั้งต่างประเทศและในประเทศที่ผ่านมา ไม่ว่าจะเป็นการศึกษาในแบบทดสอบที่ให้คะแนนสองค่าหรือหลายค่าก็ตาม ส่วนใหญ่เป็นการศึกษาวิธีการเทียบคะแนน เพื่อให้ได้ผลการเทียบคะแนนที่แม่นยำ ผลจากการวิจัยยังไม่สามารถสรุปได้ว่าวิธีการเทียบคะแนนวิธีใดให้ผลดีในทุกสถานการณ์ วิธีการเทียบคะแนนให้ผลดีในบางบริบท ขึ้นอยู่กับองค์ประกอบที่เกี่ยวข้องกับการเทียบคะแนนที่ส่งผลต่อประสิทธิภาพของการเทียบคะแนน ได้แก่ขนาดของหน่วยตัวอย่าง ความยาวข้อสอบรวม ความแตกต่างของความสามารถกลุ่มตัวอย่าง แต่ถ้าในสถานการณ์ทดสอบสามารถจัดเงื่อนไขได้ดีสมดุลแล้ว วิธีการเทียบคะแนนในโมเดล IRT ที่ปรับค่าพารามิเตอร์พร้อมกันจะให้ผลดีที่สุด ผลที่ได้ในการศึกษาเกี่ยวกับการเทียบคะแนน ยังไม่สามารถนำไปอ้างอิงในสถานการณ์อื่นได้ เนื่องจากการศึกษาส่วนใหญ่เป็นการศึกษาในบริบทเฉพาะ สำหรับการศึกษาการเทียบคะแนนในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่า ยังมีไม่มากนักในงานวิจัยต่างประเทศ และสำหรับประเทศไทยยังไม่ปรากฏหลักฐานการเทียบคะแนนในแบบทดสอบประเภทนี้ ผู้วิจัยเห็นว่าปัจจุบันแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่าได้มีการนำมาใช้มากขึ้น ควรจะได้มีการศึกษาเพื่อพัฒนาวิธีการเทียบคะแนนในข้อสอบที่ตรวจให้คะแนนหลายค่าตามทฤษฎีการตอบข้อสอบ เพื่อจะได้ดำเนินการเทียบคะแนนในแบบทดสอบลักษณะดังกล่าวได้ มิเช่นนั้นจะพบว่าเมื่อต้องเทียบคะแนนในข้อสอบรูปแบบผสมก็มักัดัดสารสนเทศของข้อสอบเขียนตอบออกไป

เทียบคะแนนเฉพาะข้อสอบเลือกตอบที่ให้คะแนนสองค่าหรือเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ซึ่งไม่ต้องปรับกระบวนการเทียบคะแนนใด ๆ ประกอบกับข้อสอบที่ตรวจให้คะแนนหลายค่าที่นำมาศึกษาในการเทียบคะแนนส่วนใหญ่อยู่ในรูปแบบของข้อสอบเขียนตอบ ซึ่งมีจุดอ่อนในด้านการให้คะแนนที่เกิดจากการเปลี่ยนแปลงผู้ตรวจ ทำให้คะแนนที่ได้ไม่คงที่ อาจก่อให้เกิดความคลาดเคลื่อนในการวัดสูงขึ้น และส่งผลให้การเทียบคะแนนขาดความแม่นยำได้ ผู้วิจัยจึงนำแบบทดสอบเลือกตอบที่ข้อสอบตรวจให้คะแนนหลายค่า ซึ่งมีวิธีการตรวจให้คะแนนที่เป็นปรนัยมากกว่ามาศึกษาเพื่อตรวจสอบว่าวิธีการเทียบคะแนนสำหรับแบบทดสอบรูปแบบนี้ ในสถานการณ์ทดสอบจริงวิธีการเทียบคะแนนใดให้ผลการเทียบคะแนนที่คงที่ ซึ่งอาจจะสอดคล้องหรือไม่สอดคล้อง กับผลงานวิจัยที่มีการศึกษาด้วยการจำลองข้อมูล ในแบบทดสอบที่ข้อสอบตรวจให้คะแนนหลายค่าชนิดเขียนตอบก็ได้ เพื่อจะได้ข้อค้นพบที่ครบถ้วนสมบูรณ์ว่า รูปแบบของข้อสอบที่ให้คะแนนหลายค่ามิใช่ตัวแปรหนึ่งของการเทียบคะแนน ดังมีวิธีการศึกษาในบทถัดไป

บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้เป็นการศึกษาเชิงประจักษ์ (Empirical Study) มีจุดมุ่งหมายเพื่อศึกษาผลการเทียบคะแนน 5 วิธี ที่ได้จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ด้านเหตุผลแบบเลือกตอบ 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบ และให้ผู้สอบตัดสินทุกคำตอบ ซึ่งข้อสอบตรวจให้คะแนนหลายค่า ผู้วิจัยได้ดำเนินการวิจัยตามขั้นตอน ดังนี้

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

การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 1 ปีการศึกษา 2545 ของโรงเรียนระดับมัธยมศึกษา สังกัดกรมสามัญศึกษา จังหวัดกรุงเทพมหานคร ซึ่งมีทั้งหมด 47 เขต 116 โรงเรียน จำนวน 52,303 คน 1,207 ห้องเรียน (ฝ่ายสถิติ กองแผนงาน กรมสามัญศึกษา)

กลุ่มตัวอย่าง

กลุ่มตัวอย่างในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 1 ปีการศึกษา 2545 ของโรงเรียนระดับมัธยมศึกษา สังกัดกรมสามัญศึกษา จังหวัดกรุงเทพมหานคร ประมาณขนาดตัวอย่างรวมสำหรับประมาณค่าเฉลี่ยของประชากรแบบ Stratified Random Sampling เมื่อกำหนดขนาดตัวอย่างของชั้นภูมิตามขนาดชั้นภูมิ (Proportional allocation) ได้กลุ่มตัวอย่างจำนวน 665 คน มีสูตรการคำนวณ (สุชาติ กระนั้นนท์, 2538 : 130) ดังนี้

$$n_0 = \frac{1}{V} \sum W_h S_h^2$$

โดยที่ n_0 เท่ากับ ขนาดตัวอย่างรวม (ไม่พิจารณาค่า fpc)
 S_h^2 เท่ากับ ความแปรปรวนในแต่ละชั้นภูมิ (กำหนดเท่ากันทุกชั้นภูมิมีค่าเท่ากับ $(30)^2$)

W_h เท่ากับ $\frac{N_h}{N}$

N_h	เท่ากับ ขนาดประชากรในแต่ละชั้นภูมิ (ขนาดใหญ่พิเศษ 25,916 คน ขนาดใหญ่ 21,266 คน และขนาดกลางและขนาดเล็ก 5,121 คน)
N	เท่ากับ ขนาดประชากรรวม (52,303 คน)
v	เท่ากับ ความแปรปรวนของค่าเฉลี่ย (เท่ากับ $(d/z)^2$)
d	เท่ากับ ความคลาดเคลื่อนของการประมาณค่าเฉลี่ย ไม่เกิน 3 คะแนน
z	เท่ากับ z ณ ระดับความเชื่อมั่น 99% (มีค่าเท่ากับ 2.58)

แต่การออกแบบการเทียบคะแนนจะต้องใช้กลุ่มตัวอย่างเทียบคะแนน 2 กลุ่ม และในการวิเคราะห์ข้อมูลตามทฤษฎีการตอบข้อสอบ ซึ่งประมาณค่าพารามิเตอร์ด้วยวิธีความควรจะเป็นสูงสุดขอบ (Marginal Maximum likelihood : MML) ควรใช้หน่วยตัวอย่างขนาด 500 คนขึ้นไป สำหรับข้อสอบจำนวน 5 10 และ 20 ข้อ (ชนวัฒน์ แสนสุข. 2538 : 44 อ้างอิงจาก Ankenmann and Stone. 1992) และหากข้อมูลมีการแจกแจงที่มีลักษณะเบ้มาก ควรใช้กลุ่มตัวอย่างประมาณ 1,000 คน (ชนวัฒน์ แสนสุข. 2538 : 44 อ้างอิงจาก Zwinderman and Wollenberg. 1990) ในการศึกษาครั้งนี้จึงใช้หน่วยตัวอย่างขนาด 1,000 คน ดังนั้นกลุ่มตัวอย่างที่ใช้ในการเทียบคะแนนรวมทั้งสิ้น 2,312 คน และกลุ่มสอบทานผลซึ่งใช้ประเมินคุณภาพในการเทียบคะแนน ควรมีขนาดไม่น้อยกว่า 10% ของกลุ่มเทียบคะแนน (Kolen and Whitney. 1982: 287 - 291) ในศึกษานี้ใช้ จำนวน 589 คน สำหรับหน่วยตัวอย่างที่ใช้ในการทดลองเครื่องมือ ใช้ขนาดประมาณ 250 คน ซึ่งเป็นจำนวนขั้นต่ำสุดที่สามารถวิเคราะห์พารามิเตอร์จากโมเดลเกรด เรสพอนส์ ในโปรแกรม MULTILOG ได้ (Reise and Yu. 1990 : 143) โดยสุ่มตัวอย่างแบบสองขั้นตอน มีขั้นตอนในการสุ่มดังนี้

1. สุ่มโรงเรียน

จากการสำรวจข้อมูลโรงเรียนระดับมัธยมศึกษาในกรุงเทพมหานคร มีจำนวน 116 โรงเรียน สามารถจำแนกตามเกณฑ์การแบ่งขนาดโรงเรียน ของกรมสามัญศึกษา ออกเป็น 4 ขนาดได้จำนวนโรงเรียนในแต่ละขนาดดังนี้

- 1.1 โรงเรียนขนาดใหญ่พิเศษ (พิจารณาจากจำนวนนักเรียนทั้งหมดมากกว่า 2,499 คน) จำนวน 43 โรงเรียน
- 1.2 โรงเรียนขนาดใหญ่ (พิจารณาจากจำนวนนักเรียนทั้งหมดตั้งแต่ 1,500 ถึง 2,499 คน) จำนวน 49 โรงเรียน
- 1.3 โรงเรียนขนาดกลาง (พิจารณาจากจำนวนนักเรียนทั้งหมดตั้งแต่ 500 ถึง 1,499 คน) จำนวน 23 โรงเรียน

1.4 โรงเรียนขนาดเล็ก (พิจารณาจากจำนวนนักเรียนทั้งหมดน้อยกว่า 500 คน)
จำนวน 1 โรงเรียน

ผู้วิจัยดำเนินการสุ่มแบบแบ่งชั้นภูมิ (Stratified Random Sampling) โดยมีขนาดโรงเรียนเป็นชั้นภูมิ ในการศึกษาครั้งนี้รวมโรงเรียนขนาดเล็กและขนาดกลางไว้ในชั้นภูมิเดียวกันและโรงเรียนเป็นหน่วยในการสุ่ม สุ่มตามสัดส่วนจำนวนโรงเรียนในแต่ละขนาด (จำนวนร้อยละ 25 ของแต่ละขนาดโรงเรียน) จะได้โรงเรียนขนาดใหญ่พิเศษ จำนวน 6 โรงเรียน โรงเรียนขนาดใหญ่จำนวน 7 โรงเรียน และโรงเรียนขนาดกลางและเล็ก จำนวน 3 โรงเรียน รวม 16 โรงเรียน

2. สุ่มห้องเรียน

หลังจากได้โรงเรียนซึ่งเป็นหน่วยตัวอย่างในแต่ละขนาดแล้ว จึงทำการสุ่มห้องเรียนในแต่ละโรงเรียน มีห้องเรียนเป็นหน่วยในการสุ่มด้วยวิธีการสุ่มอย่างง่าย (Simple Random Sampling) ตามสัดส่วนจำนวนห้องเรียนในแต่ละโรงเรียน โดยกลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ ผู้วิจัยแบ่งออกเป็น 2 กลุ่มคือ

กลุ่มที่ 1 เป็นกลุ่มที่ใช้ในการวิจัย เพื่อนำผลการสอบไปวิเคราะห์เทียบคะแนนแบบทดสอบ ด้วยวิธีการเทียบคะแนน 5 วิธี แต่เนื่องจากในการวิจัยครั้งนี้ต้องทำการเก็บข้อมูลโดยใช้แบบทดสอบ 2 ฉบับ ซึ่งนักเรียนทำแบบทดสอบคนละ 1 ฉบับ เพื่อให้ได้นักเรียนที่สอบแบบทดสอบแต่ละฉบับมีความสามารถทางด้านการเรียนใกล้เคียงกัน จึงสุ่มนักเรียนให้ได้รับแจกแบบทดสอบสลับกันไป ได้จำนวนนักเรียนที่เป็นหน่วยตัวอย่างเทียบคะแนน จำนวนรวมทั้งหมด 2,312 คน ดังแสดงไว้ในตาราง 3

กลุ่มที่ 2 เป็นกลุ่มที่ใช้ในการวิจัย เพื่อนำผลการสอบไปวิเคราะห์เพื่อตรวจสอบคุณภาพของวิธีการเทียบคะแนนในแต่ละวิธี โดยกลุ่มสอบทานผลแต่ละคนจะต้องทำแบบทดสอบทั้ง 2 ฉบับ จำนวน 589 คน ดังแสดงไว้ในตาราง 3

ตาราง 3 รายชื่อโรงเรียนและจำนวนห้องเรียนที่เป็นกลุ่มตัวอย่าง

ขนาด	โรงเรียน	จำนวน ห้องเรียน (คน)	จำนวน ห้องที่สุ่ม	จำนวน ตัวอย่าง เทียบคะแนน	จำนวน ตัวอย่าง สอบทานผล
ใหญ่ พิเศษ	ร.ร.สวนกุหลาบวิทยาลัย	12 (674)	5	234	-
	ร.ร.ศึกษานารี	12 (658)	3	-	164
	ร.ร.บดินทรเดชา (สิง สิงหนเสนีย์)	12 (697)	5	272	-
	ร.ร.นนทรีวิทยา	14 (567)	6	266	-
	ร.ร.ทวิธาภิเศก	10 (588)	3	-	162
	ร.ร.พรตพิทยพยัต	12 (557)	5	250	-
ใหญ่	ร.ร.เบญจมราชาลัย	8 (420)	4	217	-
	ร.ร.ชิโนรสวิทยาลัย	8 (356)	4	180	-
	ร.ร.ศิลาจารพิพัฒน์	10 (436)	4	177	-
	ร.ร.สตรีวัดอัมพรสวรรค์	12 (533)	5	200	-
	ร.ร.มัธยมวัดธาตุทอง	10 (414)	3	-	112
	ร.ร.ราชวินิตมัธยม	9 (418)	4	172	-
	ร.ร.วัดราชโอรส	8(382)	3	-	151
กลาง	ร.ร.มัธยมวัดดาวคอง	6 (315)	3	124	-
	ร.ร.มัธยมวัดมกุฏกษัตริย์	6 (235)	3	110	-
	ร.ร.วัดสังเวช	7 (254)	3	110	-

เครื่องมือที่ใช้ในการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ออกแบบเครื่องมือเป็นแบบทดสอบเลือกตอบที่มีคำตอบ ถูกหลายคำตอบ และให้ตัดสินทุกคำตอบ (Multiple True-false answer) วัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล ซึ่งวัดความสามารถด้านความคิดรวบยอดทางคณิตศาสตร์ที่เกี่ยวข้องกับหลักการ วิธีการและมโนภาพในการคิดแบบนามธรรม ในเรื่องจำนวนเต็ม เศษส่วน และทศนิยม ระดับชั้นมัธยมศึกษาตอนต้น จำนวน 2 ฉบับ แต่ละฉบับประกอบด้วยข้อสอบชนิด 5 ตัวเลือก จำนวน 30 ข้อ เป็นข้อสอบรวม จำนวน 10 ข้อ (รวมอยู่ใน 30 ข้อ) ที่ใช้ในการเทียบคะแนนในแบบทดสอบทั้ง 2 ฉบับ มีรายละเอียดดังต่อไปนี้

ขั้นตอนในการสร้างแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบ วัดผลสัมฤทธิ์ทางการเรียน คณิตศาสตร์ด้านเหตุผล ของนักเรียนชั้นมัธยมศึกษาปีที่สอง เพื่อนำคะแนนมาใช้ในการศึกษา วิธีการเทียบคะแนนหลายค่า
2. ศึกษาทฤษฎีเกี่ยวกับการสร้างแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบ และการวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์จากเอกสารและงานวิจัยที่เกี่ยวข้องเพื่อเป็นแนวทางในการสร้างข้อสอบ
3. ศึกษาหลักสูตร คู่มือครู และแบบเรียนวิชาคณิตศาสตร์ชั้นมัธยมศึกษาตอนต้น (ค.101 ค.102 และ ค.203) เพื่อวิเคราะห์เนื้อหา จุดประสงค์การเรียนรู้และมโนคติเรื่องจำนวนเต็ม เศษส่วน และทศนิยม
4. สร้างตารางวิเคราะห์เนื้อหา โดยกำหนดเนื้อหา จุดประสงค์เชิงพฤติกรรมและมโนคติให้สอดคล้องกับพฤติกรรมที่ต้องการวัดของวิลสัน (Wilson 1971 : 645-696) แล้วให้ผู้เชี่ยวชาญทางด้านการสอนคณิตศาสตร์จำนวน 4 ท่าน ซึ่งเป็นอาจารย์ที่สอนวิชาคณิตศาสตร์ในระดับมัธยมศึกษาไม่ต่ำกว่า 5 ปี และจบการศึกษาขั้นต้นระดับปริญญาโท (ดังมีรายชื่อในภาคผนวก) เป็นผู้พิจารณาว่าการกำหนดเนื้อหา และพฤติกรรมการวัดมีความเหมาะสมเพียงใด รวมทั้งพิจารณาการกำหนดน้ำหนักในแต่ละพฤติกรรม แล้วนำผลที่ได้มาพิจารณารวมกับเอกสารหลักสูตร เช่นเวลาที่ใช้ในการสอนแต่ละเนื้อหา เพื่อกำหนดจำนวนข้อสอบในแต่ละเนื้อหาตามระดับพฤติกรรมที่ต้องการวัด ดังมีรายละเอียดของจำนวนข้อสอบในแต่ละระดับพฤติกรรมแต่ละเนื้อหาดังตาราง 4

ตาราง 4 การวิเคราะห์เนื้อหา จุดประสงค์ ระดับพฤติกรรมการวัดและจำนวนข้อของ
แบบทดสอบเทียบคะแนน

เนื้อหา/จุดประสงค์ เชิงพฤติกรรม	ระดับพฤติกรรม (จำนวนข้อ)				รวม จำนวนข้อ ที่สร้าง (ใช้จริง)
	ความรู้ ความจำ	ความ เข้าใจ	การนำ ไปใช้	การ วิเคราะห์	
ระบบจำนวนเต็ม					
1. จำนวนเต็มและการเปรียบเทียบ					
1.1 ระบุโครงสร้างของระบบจำนวน เต็มได้	1	1			2
1.2 ระบุได้ว่าจำนวนที่กำหนดให้เป็น จำนวนเต็มหรือจำนวนเต็ม บวก เต็มลบ และศูนย์		2			2
1.3 เปรียบเทียบจำนวนเต็มได้		2			2
รวม					6(3)
2. การบวกและลบจำนวนเต็ม					
2.1 อธิบายหลักการบวก ลบจำนวน เต็มได้		2			2
2.2 เปลี่ยนรูปการบวก ลบจำนวน เต็มให้อยู่ในรูปที่เท่ากันได้		2			2
2.3 หาผลบวกและผลลบของจำนวน เต็มได้	3	2	1	2	8
รวม					12(6)
3. การคูณและการหารจำนวนเต็ม					
3.1 อธิบายหลักการคูณและหาร จำนวนเต็มได้	1	2		1	4
3.2 หาผลคูณและผลหารจำนวนเต็มได้	1	5	1	1	8
3.3 นำหลักการบวก ลบ คูณ และ หารจำนวนเต็มไปแก้ปัญหาได้		6		2	8
รวม					20(10)

ตาราง 4 (ต่อ)

เนื้อหา/จุดประสงค์ เชิงพฤติกรรม	ระดับพฤติกรรม (จำนวนข้อ)				รวม จำนวนข้อ ที่สร้าง (ใช้จริง)
	ความรู้ ความจำ	ความ เข้าใจ	การนำ ไปใช้	การ วิเคราะห์	
4. สมบัติของหนึ่งและศูนย์					
4.1 ระบบการใช้สมบัติการบวก ลบ คูณ และหารจำนวนใด ๆ ด้วยหนึ่งและศูนย์ได้	1	1			2
4.2 นำสมบัติของหนึ่งและศูนย์ ไปแก้ ปัญหาการบวก ลบ คูณและหาร กับจำนวนเต็มได้		3	1	2	6
รวม					8(4)
5. สมบัติของจำนวนเต็ม					
5.1 ระบบสมบัติการเท่ากัน การสลับที่ เปลี่ยนกลุ่ม และการแจกแจง ที่เกี่ยวกับการบวกและการคูณ จำนวนเต็มได้	2				2
5.2 นำความรู้เกี่ยวกับสมบัติการ เท่ากัน การสลับที่ เปลี่ยนกลุ่ม และแจกแจงของการบวกและการ คูณจำนวนเต็มไปแก้ปัญหา เกี่ยวกับการบวกลบ คูณและหาร จำนวนเต็มได้		2	3	2	7
รวม					9(4)
เศษส่วน และทศนิยม					
6. เศษส่วนและการเปรียบเทียบ					
6.1 อธิบายความหมายของเศษส่วนได้		2			2
6.2 เปรียบเทียบเศษส่วนที่กำหนด ให้ว่าเท่ากัน มากกว่าหรือ น้อยกว่ากันได้		3		1	4
รวม					6(3)

ตาราง 4 (ต่อ)

เนื้อหา/จุดประสงค์ เชิงพฤติกรรม	ระดับพฤติกรรม (จำนวนข้อ)				รวม จำนวนข้อ ที่สร้าง (ใช้จริง)
	ความรู้ ความจำ	ความ เข้าใจ	การนำ ไปใช้	การ วิเคราะห์	
7. การบวก ลบ คูณ และหารเศษส่วน					
7.1 อธิบายหลักการบวก ลบ คูณ หารเศษส่วนได้	1	3			4
7.2 หาผลบวก ลบ คูณ และหาร เศษส่วนได้			2	4	6
7.3 บอกความสัมพันธ์ของข้อมูลใน ปัญหาที่กำหนดให้ได้		3			3
7.4 นำหลักการบวก ลบ คูณ และหาร เศษส่วนไปแก้ปัญหาได้		5			5
รวม					18(9)
8. ทศนิยมและการเปรียบเทียบ					
8.1 ระบุได้ว่าจำนวนทศนิยมที่กำหนด ให้จำนวนใด มากกว่า น้อยกว่า หรือเท่ากันได้		4			4
8.2 เปรียบเทียบจำนวนที่อยู่ในรูป จำนวนเต็ม เศษส่วน หรือทศนิยม ได้ว่าจำนวนใดเท่ากัน มากกว่า หรือน้อยกว่ากันได้		2			2
รวม					6(3)

ตาราง 4 (ต่อ)

เนื้อหา/จุดประสงค์ เชิงพฤติกรรม	ระดับพฤติกรรม (จำนวนข้อ)				รวม จำนวนข้อ ที่สร้าง (ใช้จริง)
	ความรู้ ความจำ	ความ เข้าใจ	การนำ ไปใช้	การ วิเคราะห์	
9. การบวก ลบ คูณ และหารทศนิยม					
9.1 อธิบายหลักการ บวก ลบ คูณ และหารทศนิยมได้		3			3
9.2 หาผลบวก ลบ คูณและหารทศนิยมได้		3	1	3	7
9.3 บอกความสัมพันธ์ของข้อมูลในปัญหาที่กำหนดให้ได้		2		1	3
9.4 นำหลักการ บวก ลบ คูณ และหารทศนิยมไปแก้ปัญหาได้		2			2
รวม					15(8)
รวมทั้งฉบับ	10	62	9	19	100(50)

5. สร้างข้อสอบตามเนื้อหาและระดับพฤติกรรมการวัดที่กำหนดไว้ในตารางวิเคราะห์เนื้อหา ได้แบบทดสอบเลือกตอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล 5 ตัวเลือก จำนวน 100 ข้อ เพื่อใช้จริง จำนวน 50 ข้อ (2 ฉบับ) แต่ละข้อมีคำตอบถูกหลายคำตอบ

6. กำหนดเกณฑ์การตรวจให้คะแนน โดยให้ผู้สอบตัดสินทุกตัวเลือกว่าถูกหรือผิด ถ้าตอบได้ตรงกับเฉลยให้ 1 คะแนน ถ้าไม่ตรงเฉลยหรือไม่ตอบให้ 0 คะแนน คะแนนรวมแต่ละข้อได้จากการรวมคะแนนในแต่ละตัวเลือก

โดยผู้สอบที่ตอบไม่ถูกทุกตัวเลือกหรือไม่ตอบได้ 0 คะแนน ให้คะแนนอยู่ในลำดับชั้น (Category) ที่ 1

ผู้สอบที่ตอบถูกเพียงตัวเดียวได้ 1 คะแนน ให้คะแนนอยู่ในลำดับชั้นที่ 2

ผู้สอบที่ตอบถูกสองตัวเลือกได้ 2 คะแนน ให้คะแนนอยู่ในลำดับชั้นที่ 3

ผู้สอบที่ตอบถูกสามตัวเลือกได้ 3 คะแนน ให้คะแนนอยู่ในลำดับชั้นที่ 4

ผู้สอบที่ตอบถูกสี่ตัวเลือกได้ 4 คะแนน ให้คะแนนอยู่ในลำดับชั้นที่ 5

ผู้สอบที่ตอบถูกทุกคำตอบ ได้ 5 คะแนน ให้คะแนนอยู่ในลำดับชั้นที่ 6

เป็นความสามารถจากระดับต่ำไปสูง ผู้ที่ได้คะแนนสูงกว่าแสดงว่ามีความสามารถมากกว่าผู้ที่ได้คะแนนต่ำกว่า

ตัวอย่างข้อสอบ

(0) ให้ a และ b เป็นจำนวนเต็ม ค่าของ a และ b ในข้อใด ทำให้ผลลัพธ์ของ $a - b$ เป็นจำนวนเต็มลบ

- ก. a เป็นจำนวนเต็มลบและ b เป็นศูนย์
- ข. a และ b เป็นจำนวนเต็มลบ โดยที่ a อยู่ห่างจากศูนย์มากกว่า b
- ค. a และ b เป็นจำนวนเต็มบวก โดยที่ a อยู่ห่างจากศูนย์มากกว่า b
- ง. a เป็นจำนวนเต็มลบ และ b เป็นจำนวนเต็มบวก
- จ. a เป็นศูนย์และ b เป็นจำนวนเต็มบวก

(คำตอบถูก คือ ก,ข,ง,จ)

(00) ถ้า $P + Q = 10$ และ $R + S = -12$ แล้ว $(P + Q) - (R + S)$ มีค่าตรงกับข้อใด

- ก. $10 - 12$
- ข. $10 - (-12)$
- ค. $10 + 12$
- ง. $(-12) + 10$
- จ. $(-10) - 12$

(คำตอบถูก คือ ข,ค)

ตัวอย่างการตอบ

ข้อ	ก	ข	ค	ง	จ
(0)	x	x	✓	✓	✓
(00)	x	✓	✓	x	x

ข้อ (0) ได้ 2 คะแนน ข้อ (00) ได้ 5 คะแนน

คุณภาพของแบบทดสอบและการคัดเลือกข้อสอบ

1. ตรวจสอบความเที่ยงตรงเชิงเนื้อหา

นำข้อสอบที่สร้างขึ้นประกอบด้วยข้อสอบเลือกตอบ 5 ตัวเลือก พร้อมเฉลยคำตอบไปให้ผู้เชี่ยวชาญด้านการสอนคณิตศาสตร์ จำนวน 6 ท่าน (ดังมีรายชื่ออยู่ในภาคผนวก) ตรวจสอบความสอดคล้องระหว่างข้อสอบกับจุดประสงค์เชิงพฤติกรรมและพฤติกรรมการวัดตามวิธีของโรวินลลี และแฮมเบิลตัน (บุญเชิด ภิญโญนนตพงษ์ 2527:67-70; อ้างอิงจาก Rovinelli and Hambleton 1977) มีสูตรดังนี้

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

เมื่อ IOC คือ ดัชนีความสอดคล้องระหว่างข้อสอบกับพฤติกรรม
การวัด

R คือ คะแนนความคิดเห็นของผู้เชี่ยวชาญ

N คือ จำนวนผู้เชี่ยวชาญ

โดยพิจารณาคะแนนการตัดสินเฉลี่ยมีค่าตั้งแต่ 0.5 ถือว่าข้อสอบนั้นวัดได้สอดคล้องกับพฤติกรรมการวัดนั้นจริง โดยให้นำน้ำหนักคะแนน ดังนี้

+1 หมายถึง แน่ใจว่าข้อสอบวัดได้ตรงตามเนื้อหาและพฤติกรรมการวัด

0 หมายถึง ไม่แน่ใจว่าข้อสอบวัดได้ตรงเนื้อหาและพฤติกรรมการวัด

-1 หมายถึง แน่ใจว่า ข้อสอบวัดได้ไม่ตรงตามเนื้อหาและ

พฤติกรรมการวัด

พร้อมทั้งตรวจสอบความถูกต้องของภาษาที่ใช้ และเฉลยการตอบข้อสอบ

จากการนำข้อสอบไปให้ผู้เชี่ยวชาญตรวจสอบ แล้วนำคะแนนความคิดเห็นมา
คำนวณค่า IOC จะได้ค่าคะแนนเฉลี่ยจากผลการตัดสินของผู้เชี่ยวชาญอยู่ในช่วง .17 ถึง 1.00
(ดังผลการตรวจสอบที่ปรากฏในตาราง 5) ข้อสอบที่มีค่าเฉลี่ยไม่ถึง .5 จำนวน 12 ข้อ ได้มีการ
ปรับปรุงทั้งข้อคำถามและตัวเลือก แล้วนำไปให้ผู้เชี่ยวชาญพิจารณาอีกครั้งหนึ่ง ก่อนนำไป
ทดลองใช้

ตาราง 5 ค่าดัชนีความสอดคล้องและจำนวนความถี่ของผลการประเมินที่ประเมินความ
สอดคล้องระหว่างเนื้อหา ระดับพฤติกรรมกับข้อสอบของผู้เชี่ยวชาญ

ลำดับ ที่	เนื้อหา/จุดประสงค์เชิงพฤติกรรม	ข้อที่ (ฉบับ)	จำนวนความถี่ของ ผลการประเมิน			รวม	ดัชนีความ สอดคล้อง	
			-1	0	+1			
1.	ระบบจำนวนเต็ม							
	จำนวนเต็มและการเปรียบเทียบ							
	1.1 ระบบโครงสร้างของระบบจำนวนเต็มได้	1(X) 2(Y)	 1	 1	 4	 3	 100 0.5	
	1.2 ระบบได้ว่าจำนวนที่กำหนดให้เป็นจำนวนเต็มหรือจำนวนเต็มบวก เต็มลบ และศูนย์	3(X) 4(Y)	 	 1	 5	 5	 0.83 0.83	
	1.3 เปรียบเทียบจำนวนเต็มได้	5(X) 6(Y)	 2	 1	 5	 5	 0.83 0.33	
	2.	การบวกและลบจำนวนเต็ม						
		2.1 อธิบายหลักการบวก ลบจำนวนเต็มได้	7(Y) 8(X)	 1	 2	 3	 2	 0.33 1.00
		2.2 เปลี่ยนรูปการบวก ลบ จำนวนเต็มให้อยู่ในรูปที่เท่ากันได้	9(X) 10(Y)	 	 1	 5	 5	 0.83 1.00
		2.3 หาผลบวกและผลลบของจำนวนเต็มได้	11(Y) 12(X) 13(Y) 14(X) 15(Y) 16(X) 17(Y) 18(X)	 	 1 1 2	 6 5 5 6 6 5 4 6	 6 5 5 6 6 5 4 6	 1.00 0.83 0.83 1.00 1.00 0.83 0.67 1.00
		3.	การคูณและการหารจำนวนเต็ม					
3.1 อธิบายหลักการคูณและหารจำนวนเต็มได้			19(Y) 20(X) 21(Y) 22(X)	 1 1 1	 1 2	 5 5 4 4	 4 4 3 4	 0.67 0.67 0.5 0.67

ตาราง 5 (ต่อ)

ลำดับ ที่	เนื้อหา/จุดประสงค์เชิงพฤติกรรม	ข้อที่ (ฉบับ)	จำนวนความถี่ของ ผลการประเมิน			รวม	ดัชนีความ สอดคล้อง
			-1	0	+1		
	3.2 หาผลคูณและผลหารจำนวน เต็มได้	23(Y)		1	5	5	0.83
		24(Y)		2	4	4	0.67
		25(X)	1	1	4	3	0.50
		26(X)		2	4	4	0.67
		27(Y)		2	4	4	0.67
		28(X)		2	4	4	0.67
		29(X)			6	6	1.00
		30(Y)			6	6	1.00
	3.3 นำหลักการ บวก ลบ คูณ และ หารจำนวนเต็มไปแก้ปัญหาคิด	31(Y)		1	5	5	0.83
		32(Y)		1	5	5	0.83
		33(Y)		1	5	5	0.83
		34(X)	2	1	3	1	0.17
		35(X)			6	6	1.00
		36(Y)			6	6	1.00
		37(X)	1		5	4	0.67
		38(X)	1		5	4	0.67
4.	สมบัติของหนึ่งและศูนย์						
	4.1 ระบุการใช้สมบัติการบวก ลบ คูณ และหารจำนวนเต็มใด ๆ ด้วยหนึ่งและศูนย์ได้	39(X)	1	1	4	3	0.5
		40(Y)			6	6	1.00
	4.2 นำสมบัติของหนึ่งและศูนย์ไป แก้ปัญหาคการบวก ลบ คูณ และ หารกับจำนวนเต็มได้	41(Y)			6	6	1.00
		42(X)	2		4	2	0.33
		43(Y)			6	6	1.00
		44(X)		1	5	5	0.83
		45(Y)	1	1	4	3	0.5
		46(X)	1	2	3	2	0.33

ตาราง 5 (ต่อ)

ลำดับ ที่	เนื้อหา/จุดประสงค์เชิงพฤติกรรม	ข้อที่ (ฉบับ)	จำนวนความถี่ของ ผลการประเมิน			รวม	ดัชนีความ สอดคล้อง
			-1	0	+1		
5.	สมบัติของจำนวนเต็ม						
	5.1 ระบุสมบัติการเท่ากัน การสลับที่ เปลี่ยนกลุ่ม และการแจกแจงเกี่ยวกับการบวกและการคูณจำนวนเต็มได้	47(X) 48(Y)		1	6 5	6 5	1.00 0.83
	5.2 นำความรู้เกี่ยวกับสมบัติการเท่ากัน การสลับที่ เปลี่ยนกลุ่ม และการแจกแจงของการบวกและการคูณของจำนวนเต็มไปแก้ปัญหาเกี่ยวกับการบวก ลบ คูณและหารจำนวนเต็มได้	49(X) 50(Y) 51(X) 52(Y) 53(Y) 54(X) 55(Y)	2		6 6 6 6 6 4 6	6 6 6 6 6 2 6	1.00 1.00 1.00 1.00 1.00 0.33 1.00
6.	เศษส่วนและทศนิยม						
	6.1 อธิบายความหมายของเศษส่วนได้	56(Y) 57(X)			6 6	6 6	1.00 1.00
	6.2 เปรียบเทียบเศษส่วนที่กำหนดให้ว่าเท่ากัน มากกว่า หรือน้อยกว่ากันได้	58(X) 59(Y) 60(X) 61(Y)		1	6 5 6 6	6 5 6 6	1.00 0.83 1.00 1.00
7.	การบวก ลบ คูณและหารเศษส่วน						
	7.1 อธิบายหลักการบวก ลบ คูณและหารเศษส่วนได้	62(Y) 63(X) 64(Y) 65(X)	1	1 1 1	6 4 5 5	6 3 5 5	1.00 0.5 0.83 0.83

ตาราง 5 (ต่อ)

ลำดับ ที่	เนื้อหา/จุดประสงค์เชิงพฤติกรรม	ข้อที่ ฉบับ	จำนวนความถี่ของ ผลการประเมิน			รวม	ดัชนีความ สอดคล้อง
			-1	0	+1		
	7.2 ทาผลบวก ลบ คูณและหาร เศษส่วนได้	66(Y)			6	6	1.00
		67(X)		1	5	5	0.83
		68(X)		1	5	5	0.83
		69(Y)		1	5	5	0.83
		70(Y)			6	6	1.00
		71(X)			6	6	1.00
	7.3 บอกความสัมพันธ์ของข้อมูล ในปัญหาที่กำหนดให้ได้	72(Y)			6	6	1.00
		73(X)			6	6	1.00
		74(Y)			6	6	1.00
	7.4 นำหลักการ บวก ลบ คูณ และ หารเศษส่วนไปแก้ปัญหาคได้	75(X)	2		4	2	0.33
		76(Y)	1	1	4	3	0.5
		77(X)	1		5	4	0.67
		78(Y)	2	1	3	1	0.17
		79(Y)		1	5	5	0.83
8.	ทศนิยมและการเปรียบเทียบ						
	8.1 ระบุได้ว่าจำนวนทศนิยมที่ กำหนดให้จำนวนใดมากกว่า น้อยกว่าหรือเท่ากันได้	80(Y)	1	2	3	2	0.33
		81(X)			6	6	1.00
		82(Y)			6	6	1.00
		83(X)			6	6	1.00
	8.2 เปรียบเทียบจำนวนที่อยู่ในรูป ของจำนวนเต็ม เศษส่วนหรือ ทศนิยมได้ว่า จำนวนใดเท่ากัน มากกว่าหรือน้อยกว่ากันได้	84(Y)			6	6	1.00
		85(X)		1	5	5	0.83
9.	การบวก ลบ คูณและหารทศนิยม						
	9.1 อธิบายหลักการบวก ลบ คูณ และหารทศนิยมได้	86(X)		1	5	5	0.83
		87(X)			6	6	1.00
		88(Y)			6	6	1.00

ตาราง 5 (ต่อ)

ลำดับ ที่	เนื้อหา/จุดประสงค์เชิงพฤติกรรม	ข้อที่ (ฉบับ)	จำนวนความถี่ของ ผลการประเมิน			รวม	ดัชนีความ สอดคล้อง
			-1	0	+1		
	9.2 ทาผลบวก ลบ คูณ และหาร ทศนิยมได้	89(Y)	1	2	3	2	0.33
		90(Y)		1	5	5	0.83
		91(X)		2	4	4	0.67
		92(X)	1		5	4	0.67
		93(X)			6	6	1.00
		94(Y)			6	6	1.00
		95(X)			6	6	1.00
	9.3 บอกความสัมพันธ์ของข้อมูล ในปัญหาที่กำหนดให้ได้	96(X)			6	6	1.00
		97(Y)			6	6	1.00
		98(X)			6	6	1.00
	9.4 นำหลักการ บวก ลบ คูณ และ หารทศนิยมไปแก้ปัญหาได้	99(Y)	2	1	3	1	0.17
		100(X)	2	1	3	1	0.17

2. ทดลองใช้แบบทดสอบ

2.1 ทดลองใช้ครั้งที่ 1

2.1.1 นำแบบทดสอบที่ปรับปรุงแก้ไขแล้วไปทดลองใช้ (try out) เพื่อหาคุณภาพรายข้อและความเหมาะสมที่จะนำไปใช้กับกลุ่มตัวอย่างจริง โดยนำไปสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2545 จำนวน 179 คน เป็นนักเรียนโรงเรียนสตรีศรีสุริโยทัย จำนวน 105 คน และโรงเรียนวัดน้อยใน จำนวน 74 คน โดยแบ่งข้อสอบออกเป็น 2 ฉบับ ๆ ละ 50 ข้อ แต่ละฉบับคู่ขนานในด้านเนื้อหา (รายละเอียดดังตาราง 5) ทดสอบใช้เวลาฉบับละ 2 ชั่วโมง โดยกำหนดให้นักเรียนแต่ละคนทำแบบทดสอบคนละฉบับ โดยสุ่มแจกแบบทดสอบสลับกันในแต่ละห้องสอบ เพื่อให้แบบทดสอบแต่ละฉบับได้รับการทดสอบจากหน่วยตัวอย่างที่มีความสามารถใกล้เคียงกันแบบทดสอบแต่ละฉบับ มีผู้สอบ จำนวน 92 และ 87 คน

2.1.2 นำแบบทดสอบทั้งสองฉบับมาตรวจให้คะแนนตามเกณฑ์ พร้อมทั้งวิเคราะห์คุณภาพของข้อสอบ ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม เพื่อหาค่าความยาก (p) และอำนาจจำแนก (r) โดยใช้โปรแกรม SPSS Version 7.5 ได้ค่าความยากของข้อสอบในช่วง 0.45 ถึง 0.874 และช่วง 0.478 ถึง 0.805 ในแบบทดสอบฉบับ X และ Y ตามลำดับ ค่าอำนาจจำแนกในช่วง 0.0050 ถึง 0.6251 และช่วง -0.1474 ถึง 0.6177 ในแบบทดสอบฉบับ X และ Y ตามลำดับ มีข้อสอบที่มีคุณภาพไม่อยู่ในเกณฑ์ คือ ค่าอำนาจจำแนกต่ำกว่า 0.2 จำนวนรวม 18 ข้อ และค่าความยากมากกว่า 0.8 จำนวนรวม 2 ข้อ จึงปรับปรุงข้อสอบ โดยปรับค่าในตัวเลือกให้ใกล้เคียงกับตัวเลือกอื่น ๆ และปรับเงื่อนไขของโจทย์ในข้อสอบ เพื่อให้ข้อสอบง่ายขึ้น และนำไปทดลองใช้ครั้งที่ 2

2.2 ทดลองใช้ครั้งที่ 2

2.2.1 นำแบบทดสอบที่ปรับปรุงแก้ไขจากการทดลองใช้ครั้งที่ 1 นำไปสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 1 ปีการศึกษา 2545 จำนวน 439 คน เป็นนักเรียนโรงเรียนสุรศักดิ์มนตรี จำนวน 157 คน โรงเรียนวัดราชพิพิธ จำนวน 170 คน และโรงเรียนเตรียมอุดมศึกษาน้อมเกล้า จำนวน 112 คน โดยแบ่งข้อสอบออกเป็น 2 ฉบับ ๆ ละ 50 ข้อ ดำเนินการเช่นเดียวกับการทดลองใช้ครั้งที่ 1 แบบทดสอบแต่ละฉบับมีผู้สอบ จำนวน 223 คน และ 216 คน ตามลำดับ

2.2.2 นำแบบทดสอบทั้งฉบับมาตรวจให้คะแนนตามเกณฑ์ที่กำหนด พร้อมทั้งวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้แก่ ค่าความยาก (p) และค่าอำนาจจำแนก (r) ด้วยโปรแกรม SPSS เวอร์ชัน 7.5 ได้ค่าสัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบทั้งฉบับ โดยใช้สูตรของแอลฟา ครอนบัค เท่ากับ .936 และ .918 เมื่อวิเคราะห์คุณภาพของข้อสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิม, แบบทดสอบทั้งสองฉบับมีค่าความยากอยู่ในช่วง 0.4646 ถึง 0.9363 และ 0.4713 ถึง 0.9269 ตามลำดับ และมีค่าอำนาจจำแนกอยู่ในช่วง 0.1021 ถึง 0.6580 และ 0.1109 ถึง 0.6315 มีข้อสอบที่มีคุณภาพไม่อยู่ในเกณฑ์ คือ มีค่าความยากเกิน 0.85 จำนวนรวม 8 ข้อ และค่าอำนาจจำแนกไม่ถึง 0.2 จำนวนรวม 4 ข้อ และวิเคราะห์ข้อสอบตามทฤษฎีการตอบข้อสอบในโมเดลเกรต เรสพอนส์ โดยใช้โปรแกรม MULTILOG เวอร์ชัน 7.03 for Windows ได้แก่ ค่าความยาก (b_{jk}) ค่าอำนาจจำแนก (a_j) และค่าสารสนเทศของข้อสอบแต่ละข้อ แบบทดสอบทั้งสองฉบับมีค่าอำนาจจำแนกไม่ติดลบ ซึ่งถือว่าอยู่ในเกณฑ์ที่มีคุณภาพ คือ มีอำนาจจำแนกอยู่ในช่วง 0.38 ถึง 2.18 และอยู่ในช่วง 0.47 ถึง 2.52 ตามลำดับ และส่วนใหญ่สอดคล้องกับค่าอำนาจจำแนกที่วิเคราะห์ได้จากทฤษฎีการทดสอบแบบมาตรฐานเดิม คือ ถ้าข้อสอบข้อใดผลวิเคราะห์จากทฤษฎีการทดสอบแบบมาตรฐานเดิมมีค่าสูง ผลที่ได้จากการวิเคราะห์ตามทฤษฎีการตอบข้อสอบส่วนใหญ่มีค่าสูงด้วย

สำหรับค่าความยากของข้อสอบทั้งสองฉบับเรียงตามลำดับชั้นความยากจากน้อยไปหามาก แต่พบว่าข้อสอบข้อที่ 8 ในแบบทดสอบฉบับที่ 2 ค่าความยากไม่เรียงตามลำดับชั้น จึงถือว่ามีความคุณภาพที่ใช้ไม่ได้

2.2.3 คัดเลือกข้อสอบที่มีคุณภาพอยู่ในเกณฑ์ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม คือ ข้อสอบที่มีค่าความยากของข้อสอบระหว่าง 0.2 – 0.85 และค่าอำนาจจำแนกตั้งแต่ .2 ขึ้นไป และเกณฑ์ตามทฤษฎีการตอบข้อสอบ คือ ข้อสอบที่มีค่าอำนาจจำแนกไม่ติดลบ และค่าความยากในแต่ละลำดับชั้นเรียงจากน้อยไปหามาก ได้จำนวนข้อสอบที่อยู่ในเกณฑ์ทั้งค่าความยากและอำนาจจำแนก จำนวน 88 ข้อ จำแนกตามเนื้อหาย่อยได้ดังนี้

เนื้อหาที่ 1	จำนวนข้อสอบที่อยู่ในเกณฑ์	5 ข้อ	จำนวนข้อสอบที่ต้องการใช้	3
เนื้อหาที่ 2	จำนวนข้อสอบที่อยู่ในเกณฑ์	8 ข้อ	จำนวนข้อสอบที่ต้องการใช้	6
เนื้อหาที่ 3	จำนวนข้อสอบที่อยู่ในเกณฑ์	15 ข้อ	จำนวนข้อสอบที่ต้องการใช้	10
เนื้อหาที่ 4	จำนวนข้อสอบที่อยู่ในเกณฑ์	8 ข้อ	จำนวนข้อสอบที่ต้องการใช้	4
เนื้อหาที่ 5	จำนวนข้อสอบที่อยู่ในเกณฑ์	9 ข้อ	จำนวนข้อสอบที่ต้องการใช้	4
เนื้อหาที่ 6	จำนวนข้อสอบที่อยู่ในเกณฑ์	6 ข้อ	จำนวนข้อสอบที่ต้องการใช้	3
เนื้อหาที่ 7	จำนวนข้อสอบที่อยู่ในเกณฑ์	18 ข้อ	จำนวนข้อสอบที่ต้องการใช้	9
เนื้อหาที่ 8	จำนวนข้อสอบที่อยู่ในเกณฑ์	5 ข้อ	จำนวนข้อสอบที่ต้องการใช้	3
เนื้อหาที่ 9	จำนวนข้อสอบที่อยู่ในเกณฑ์	14 ข้อ	จำนวนข้อสอบที่ต้องการใช้	8

การคัดเลือกและจัดเตรียมแบบทดสอบเทียบคะแนน

1. นำข้อสอบที่มีคุณภาพอยู่ในเกณฑ์มาจัดเตรียมเป็นแบบทดสอบเทียบคะแนน โดยคัดเลือกให้ได้น้ำหนักตามตารางโครงสร้างเนื้อหา และระดับพฤติกรรมการวัด และครอบคลุมจุดประสงค์ในแต่ละเนื้อหาให้ได้มากที่สุด โดยพิจารณาจำนวนข้อสอบที่มีคุณภาพในแต่ละจุดประสงค์ แล้วกำหนดสัดส่วนจำนวนข้อสอบในแต่ละจุดประสงค์ ในกรณีที่จุดประสงค์ใดมีข้อสอบ จำนวน 1 ข้อ จะพิจารณาให้เป็นข้อสอบร่วม และกรณีที่ มีจำนวนข้อสอบเกินจากจำนวนที่ต้องการในแต่ละจุดประสงค์จะพิจารณาข้อสอบที่มีค่าอำนาจจำแนกอยู่ในเกณฑ์ดี และมีความสารสนเทศของข้อสอบใกล้เคียงกันจับคู่กันไว้ (โดยพิจารณาระดับพฤติกรรมการวัดไปพร้อมกัน)

2. เลือกข้อสอบรวมให้กระจายไปตามเนื้อหา จำนวน 10 ข้อ (เป็นไปตามเงื่อนไขของข้อสอบที่มีอยู่ในแต่ละจุดประสงค์) แล้วจึงจำแนกข้อสอบที่ได้จับคู่ไว้ออกเป็น 2 ฉบับ ที่มีเนื้อหา ระดับพฤติกรรมการวัด และค่าสถิติของข้อสอบใกล้เคียงกัน (ดังรายละเอียดจำนวนข้อสอบในแต่ละจุดประสงค์แต่ละระดับพฤติกรรมการวัดในตาราง 6) โดยข้อสอบรวมกระจายอยู่ในเนื้อหาย่อยที่ 1 2 4 5 6 7 8 และ 9 เป็นข้อสอบข้อที่ 2 3 4 12 16 17 22 25 29 และ 30 ในแบบทดสอบเทียบคะแนนทั้งสองฉบับคือ ฉบับ X และ Y

3. แบบทดสอบทั้งฉบับ X และ Y มีค่าคะแนนเฉลี่ยจากผลการตัดสินของผู้เชี่ยวชาญอยู่ในช่วง .05 ถึง 1 โดยค่า IOC เฉลี่ยทั้งฉบับของแบบทดสอบฉบับ X และ Y มีค่า 0.905 และ 0.900 ตามลำดับ

ตาราง 6 จำนวนข้อสอบที่ใช้จริงในแต่ละเนื้อหา จุดประสงค์ และระดับพฤติกรรมการวัดของแบบทดสอบเทียบคะแนนฉบับ X และ Y

เนื้อหา/จุดประสงค์ เชิงพฤติกรรม	ระดับพฤติกรรมกาารวัด (จำนวนข้อ และ (ฉบับ))				รวม จำนวน ข้อสอบ ฉบับ		รวมจำนวน ข้อสอบที่ใช้ จริงทั้ง สองฉบับ
	ความรู้ ความจำ	ความ เข้าใจ	การนำ ไปใช้	การ วิเคราะห์	X	Y	
ระบบจำนวนเต็ม							
1. จำนวนเต็มและการเปรียบเทียบ							
1.1 ระบบโครงสร้างของระบบ จำนวนเต็มได้	1(Y)	1(X)			1	1	2
1.2 ระบบได้ว่าจำนวนที่ กำหนดให้เป็นจำนวน เต็มหรือจำนวนเต็มบวก เต็มลบและศูนย์		1*(X,Y)			1	1	1*
รวม					2	2	3
2. การบวกและลบจำนวนเต็ม							
2.1 อธิบายหลักการบวก ลบ จำนวนเต็มได้		1*(X,Y)			1	1	1*
2.2 เปลี่ยนรูปการบวกและ ลบจำนวนเต็มให้อยู่ในรูป ที่เท่ากันได้		1*(X,Y)			1	1	1*
2.3 หาผลบวกและผลลบของ จำนวนเต็มได้	1(X)	1(Y)	1(X)	1(Y)	2	2	4
รวม					4	4	6
3. การคูณและการหารจำนวนเต็ม							
3.1 อธิบายหลักการคูณและ หารจำนวนเต็มได้	1(Y)	1(X)			1	1	2
3.2 หาผลคูณและผลหาร จำนวนเต็มได้		2(X,Y)	1(X)	1(Y)	2	2	4

ตาราง 6 (ต่อ)

เนื้อหา/จุดประสงค์ เชิงพฤติกรรม	ระดับพฤติกรรมการวัด (จำนวนข้อ และ (ฉบับ))				รวม จำนวน ข้อสอบ ฉบับ		รวมจำนวน ข้อสอบที่ใช้ จริงทั้ง สองฉบับ
	ความรู้ ความจำ	ความ เข้าใจ	การนำ ไปใช้	การ วิเคราะห์	X	Y	
3.3 นำหลักการบวก ลบ คูณ และหารจำนวนเต็มไปใช้ แก้ปัญหาได้		1(Y) 2(X)		1(Y)	2	2	4
รวม					5	5	10
4. สมบัติของหนึ่งและศูนย์							
4.1 ระบุการใช้สมบัติการบวก ลบ คูณ และหารจำนวน ใด ๆ ด้วยหนึ่งและศูนย์ ได้	1*(X,Y)				1	1	1*
4.2 นำสมบัติของหนึ่งและ ศูนย์ไปแก้ปัญหการ บวก ลบ คูณ และหารกับ จำนวนเต็มได้		1(X)	1(Y)	1(X)	2	1	3
รวม					3	2	4
5. สมบัติของจำนวนเต็ม							
5.1 ระบุสมบัติการเท่ากัน การสลับที่ เปลี่ยนกลุ่ม และการแจกแจงเกี่ยว กับการบวกและการคูณ จำนวนเต็มได้	1*(X,Y)				1	1	1*
5.2 นำความรู้เกี่ยวกับสมบัติ การเท่ากัน การสลับที่ เปลี่ยนกลุ่ม และแจกแจง ของการบวกและการคูณ จำนวนเต็มไปแก้ปัญหา เกี่ยวกับการบวก ลบ คูณ และหารจำนวนเต็มได้		1(Y)	1(Y)	1(X)	1	2	3
รวม					2	3	4

ตาราง 6 (ต่อ)

เนื้อหา/จุดประสงค์ เชิงพฤติกรรม	ระดับพฤติกรรมการวัด (จำนวนข้อ และ (ฉบับ))				รวม จำนวน ข้อสอบ ฉบับ		รวมจำนวน ข้อสอบที่ใช้ จริงทั้ง สองฉบับ
	ความรู้ ความจำ	ความ เข้าใจ	การนำ ไปใช้	การ วิเคราะห์	X	Y	
เศษส่วน และทศนิยม							
6. เศษส่วนและการเปรียบเทียบ							
6.1 อธิบายความหมายของ เศษส่วนได้		1*(X,Y)			1	1	1*
6.2 เปรียบเทียบเศษส่วนที่ กำหนดให้ว่าเท่ากันมาก กว่าหรือน้อยกว่ากันได้		1(X)		1(Y)	1	1	2
รวม					2	2	3
7. การบวก ลบ คูณ และหาร เศษส่วน							
7.1 อธิบายหลักการบวก ลบ คูณ หารเศษส่วนได้	1(X)	1(Y)			1	1	2
7.2 หาผลบวก ลบ คูณ และ หารเศษส่วนได้			1(Y)	1(Y) 2(X)	2	2	4
7.3 บอกความสัมพันธ์ของ ข้อมูลในปัญหาที่กำหนด ให้ได้		1*(X,Y)			1	1	1*
7.4 นำหลักการบวก ลบ คูณ และหารเศษส่วนไปแก้ ปัญหาได้		2(X,Y)			1	1	2
รวม					5	5	9

ตาราง 6 (ต่อ)

เนื้อหา/จุดประสงค์ เชิงพฤติกรรม	ระดับพฤติกรรมการวัด (จำนวนข้อ และ (ฉบับ))				รวม จำนวน ข้อสอบ ฉบับ		รวมจำนวน ข้อสอบที่ใช้ จริงทั้ง สองฉบับ
	ความรู้ ความ จำ	ความ เข้าใจ	การนำ ไปใช้	การ วิเคราะห์	X	Y	
8. ทศนิยมและการเปรียบเทียบ							
8.1 ระบุได้ว่าจำนวน ทศนิยมที่กำหนดให้จำนวน ใดมากกว่า น้อยกว่าหรือ เท่ากันได้		2(X,Y)			1	1	2
8.2 เปรียบเทียบจำนวนที่อยู่ใน รูปของจำนวนเต็มเศษส่วน หรือทศนิยมได้ว่าจำนวนใด เท่ากัน มากกว่า หรือ น้อยกว่ากันได้		1*(X,Y)			1	1	1*
รวม					2	2	3
9. การบวก ลบ คูณ และหาร ทศนิยม							
9.1 อธิบายหลักการ บวกลบ คูณ และหารทศนิยมได้		2(X,Y)			1	1	2
9.2 หาผลบวก ลบ คูณและหาร ทศนิยมได้		2(X,Y)		2(X,Y)	2	2	4
9.3 บอกความสัมพันธ์ของข้อ- มูลในปัญหาที่กำหนดให้ได้		1*(X,Y)			1	1	1*
9.4 นำหลักการบวก ลบ คูณ และหารทศนิยมไปแก้ ปัญหาได้		1*(X,Y)			1	1	1*
รวม					5	5	8
รวมทั้งหมด	6	28	5	11	30	30	50
จำแนกตามฉบับ	2*(X,Y) 4 (2X,2Y)	8*(X,Y) (11X,9Y)	(2X,3Y)	(5X,6Y)			

* หมายถึง ข้อสอบรวมในแบบทดสอบเทียบคะแนนทั้งสองฉบับ

การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลตั้งแต่วันที่ 19 สิงหาคม 2545 - 28 ตุลาคม 2545 ตามขั้นตอนต่อไปนี้

1. ทำหนังสือและติดต่อขอความร่วมมือจากผู้อำนวยการโรงเรียนมัธยมศึกษาที่เป็นกลุ่มตัวอย่างเทียบคะแนนและกลุ่มสอบทานผล เพื่อกำหนดวันและเวลาสอบ
2. ดำเนินการสอบโดยนำแบบทดสอบที่ได้จัดเตรียมไว้ทั้ง 2 ฉบับคือฉบับ X และฉบับ Y แต่ละฉบับมีความยาว จำนวน 30 ข้อ ไปทดสอบกับกลุ่มตัวอย่างเทียบคะแนน โดยนักเรียนแต่ละคนในแต่ละห้อง ของแต่ละโรงเรียนถูกสุ่มอย่างมีระบบให้สอบแบบทดสอบ เพียงคนละ 1 ฉบับ เท่านั้น เพื่อให้การสอบแบบทดสอบทั้ง 2 ฉบับ มีลักษณะเหมือนสุ่ม และนักเรียนมีความสามารถด้านการเรียนใกล้เคียงกัน พร้อมทั้งแยกแบบทดสอบเทียบคะแนนฉบับ X และฉบับ Y ของผู้สอบแต่ละคน ได้ผู้สอบแบบทดสอบฉบับที่ 1 (ฉบับ X) จำนวน 1,155 คน และฉบับที่ 2 (ฉบับ Y) จำนวน 1,157 คน ก่อนการสอบได้ชี้แจงให้นักเรียนกลุ่มตัวอย่างทราบถึงวัตถุประสงค์ของการสอบและวิธีการทำข้อสอบโดยโน้มน้าวให้นักเรียนตั้งใจทำข้อสอบ เพื่อให้ได้ผลการสอบที่เชื่อถือได้
3. ดำเนินการสอบกับกลุ่มสอบทานผลโดยก่อนสอบได้ชี้แจงรายละเอียดเช่นเดียวกับกลุ่มตัวอย่างเทียบคะแนน นักเรียนแต่ละคน แต่ละห้องเรียน แต่ละโรงเรียน ได้สอบแบบทดสอบทั้ง 2 ฉบับ โดยสุ่มอย่างมีระบบให้แต่ละคนสอบแบบทดสอบฉบับ X ก่อนแล้วตามด้วยฉบับ Y หรือสอบฉบับ Y ก่อนแล้วตามด้วยฉบับ X สลับกันโดยสอบวันละ 1 ฉบับ เว้นอีก 1-7 วัน สอบอีก 1 ฉบับ แล้วแยกแบบทดสอบเทียบคะแนนฉบับ X และฉบับ Y ของผู้สอบแต่ละคน โดยตัดนักเรียนที่เข้าสอบไม่ครบทั้งสองครั้งออกไปได้กลุ่มสอบทานผลที่สอบแบบทดสอบทั้งสองฉบับ จำนวน 589 คน
4. ตรวจสอบให้คะแนนตามเกณฑ์ที่กำหนดไว้ในแบบทดสอบเทียบคะแนนแต่ละฉบับ ทั้งในกลุ่มตัวอย่างเทียบคะแนน และกลุ่มสอบทานผล

การจัดกระทำข้อมูลและการวิเคราะห์ข้อมูล

มีขั้นตอนในการจัดกระทำและวิเคราะห์ข้อมูล ดังต่อไปนี้

1. การวิเคราะห์ข้อมูลในการทดลองเครื่องมือ
 - 1.1 ตรวจสอบให้คะแนนตามเกณฑ์กำหนด
 - 1.2 วิเคราะห์ค่าความยากและอำนาจจำแนกตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ใช้ผู้สอบทุกคน โดยไม่จำแนกออกเป็นกลุ่มสูงกลุ่มต่ำ มีรายละเอียด ดังนี้

1.2.1 ดัชนีความยาก ใช้ค่าเฉลี่ยคะแนนรายข้อ (item mean) ที่ได้จากการวิเคราะห์ด้วยโปรแกรม SPSS Version 7.5 แล้วหารด้วยน้ำหนักคะแนนสูงสุด (ในการศึกษาครั้งนี้เท่ากับ 5) ดัชนีความยาก คือ

$$P = \frac{\sum X}{5N} \quad (71)$$

เมื่อ N แทน จำนวนผู้สอบ
 X แทน คะแนนของข้อคำถาม

1.2.2 ดัชนีค่าอำนาจจำแนก ใช้สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนในแต่ละข้อกับคะแนนรวมจากข้ออื่น ๆ ที่เหลือทั้งหมด วิเคราะห์ด้วยโปรแกรม SPSS Version 7.5 มีสูตรดังนี้ (บุญเชิด ภิญโญนนตพงษ์. 2545: 84)

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}} \quad (72)$$

N แทน จำนวนผู้สอบ
X แทน คะแนนของข้อคำถาม
Y แทน คะแนนผลรวมของข้ออื่น ๆ ที่เหลือทุกข้อ

1.3 วิเคราะห์ค่าความยาก อำนาจจำแนก และสารสนเทศของข้อสอบตามทฤษฎีการตอบข้อสอบในโมเดลเกรต เรสพอนส์ ด้วยโปรแกรม MULTILOG Version 7.03 for Windows

2. การวิเคราะห์ข้อมูลในการวิจัย

ข้อมูลที่ได้รับจากการสอบของนักเรียนนำมาตรวจให้คะแนนตามเกณฑ์ที่กำหนด แล้วนำคะแนนที่ได้ในแบบทดสอบเทียบคะแนนในแต่ละฉบับ มาวิเคราะห์ข้อมูลดังต่อไปนี้

2.1 การวิเคราะห์ข้อมูลแบบทดสอบในกลุ่มเทียบคะแนน

2.2 การสร้างตารางเทียบคะแนนแบบทดสอบ

2.3 การวิเคราะห์ผลการเทียบคะแนน

ดังมีรายละเอียดต่อไปนี้

2.1 การวิเคราะห์ข้อมูลแบบทดสอบในกลุ่มเทียบคะแนน

2.1.1 หาค่าสถิติพื้นฐานของกลุ่มตัวอย่าง ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน คะแนนสูงสุด คะแนนต่ำสุด มัธยฐาน ฐานนิยม ความเบ้ ความโด่ง ค่าความคลาดเคลื่อนมาตรฐาน โดยใช้โปรแกรมสำเร็จรูป SPSS Version 7.50

2.1.2 วิเคราะห์คุณภาพของแบบทดสอบ เพื่อหาความเชื่อมั่นของแบบทดสอบโดยใช้สัมประสิทธิ์ แอลฟาของครอนบัค (Cronbach) สัมประสิทธิ์ของราซ และสัมประสิทธิ์เบต้าเค (β_k) (บุญเชิด ภิญโญนนตพงษ์. 2545: 126,131-134, 137) โดยใช้โปรแกรมสำเร็จรูป SPSS Version 7.50 และความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) ด้วยโปรแกรม LISREL Version 8.30

2.1.3 ตรวจสอบการแจกแจงคะแนนของแบบทดสอบฉบับ X ฉบับ Y โดยใช้สถิติทดสอบนันทพาราเมตริกซ์ Mann-Whitney U เมื่อตัวอย่าง 2 กลุ่มสุ่มอย่างเป็นอิสระกัน (กัลยา วาณิชย์บัญชา. 2542 : 300-302) และตรวจสอบการแจกแจงคะแนนระหว่างคะแนนดิบและคะแนนสังเกตจากแบบทดสอบเทียบคะแนนฉบับเดียวกันโดยใช้สถิติทดสอบนันทพาราเมตริกซ์ Wilcoxon Signed-Ranks สำหรับตัวอย่าง 2 กลุ่มที่สัมพันธ์กัน (กัลยา วาณิชย์บัญชา. 2542 : 306-307) โดยใช้โปรแกรมสำเร็จรูป SPSS Version 7.5

2.1.4 ตรวจสอบความแตกต่างของค่าเฉลี่ยระหว่างคะแนนดิบของแบบทดสอบเทียบคะแนนฉบับ X และ Y จากกลุ่มตัวอย่างเทียบคะแนนและกลุ่มสอบทานผล โดยใช้สถิติทดสอบ t-independent เมื่อตัวอย่างจากประชากรเป็นอิสระจากกัน (กัลยา วาณิชย์บัญชา. 2542 : 167-173) และตรวจสอบความแตกต่างของค่าเฉลี่ยระหว่างคะแนนดิบและคะแนนสังเกตจากแบบทดสอบเทียบคะแนนฉบับเดียวกัน โดยใช้สถิติทดสอบ t-dependent เมื่อตัวอย่างจากประชากรสัมพันธ์กัน (กัลยา วาณิชย์บัญชา. 2542 : 173-178) ด้วยโปรแกรมสำเร็จรูป SPSS Version 7.50

2.1.5 ตรวจสอบความสัมพันธ์ระหว่างคะแนนของแบบทดสอบเทียบคะแนนกับคะแนนของแบบทดสอบร่วมในฉบับเดียวกัน และความสัมพันธ์ระหว่างคะแนนแปลงกับคะแนนเกณฑ์จากการวิเคราะห์กลุ่มสอบทานผล ด้วยสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน (กัลยา วาณิชย์บัญชา. 2542 : 220) ด้วยโปรแกรมสำเร็จรูป SPSS Version 7.50

2.1.6 ตรวจสอบคุณสมบัติการวัดมิติเดียวของแบบทดสอบแต่ละฉบับด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ โดยสกัดองค์ประกอบด้วยวิธีวิเคราะห์องค์ประกอบหลัก (Principle Component : PC) โดยใช้โปรแกรมสำเร็จรูป SPSS Version 7.50 และยืนยันองค์ประกอบ (Confirmatory Analysis) ด้วยโปรแกรม LISREL Version 8.30

2.1.7 วิเคราะห์ข้อสอบรายข้อ ตามทฤษฎีการตอบข้อสอบ (IRT) ในโมเดลเกรต เรสพอนส์ เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยาก ค่าอำนาจจำแนก โดยใช้โปรแกรม MULTILOG Version 7.03 for Windows ประมาณค่าพารามิเตอร์ 2 ลักษณะคือ

2.1.7.1 ประมาณค่าพารามิเตอร์จากแบบทดสอบเทียบคะแนนทั้งสองฉบับพร้อมกัน โดยข้อสอบไม่ได้ร่วม (Unique item) ใส่รหัส "9" เป็น Missing ดังนั้น จะได้มาตรวจวัดร่วม ซึ่งเกิดจากสารสนเทศของข้อสอบร่วมจากแบบทดสอบทั้งสองฉบับ

2.1.7.2 ประมวลค่าพารามิเตอร์จากแบบทดสอบเทียบคะแนนแต่ละฉบับแยกกัน โดยครั้งที่หนึ่งจากแบบทดสอบฉบับ X และครั้งที่สองจากแบบทดสอบฉบับ Y

2.1.8 คำนวณค่าความน่าจะเป็นของการตอบข้อสอบแต่ละข้อในแบบทดสอบเทียบคะแนนที่วิเคราะห์ด้วยโมเดลเกรต เรสพอนส์ มีสูตรดังนี้ (Bastari, 2000 : 9)

$$P_{jk}(\theta) = P_{jk}^+(\theta) - P_{j,k+1}^+(\theta) \quad (73)$$

เมื่อ $P_{jk}(\theta)$ คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถระดับ θ จะตอบข้อสอบข้อที่ j ได้ลำดับชั้นคะแนน k :

P_{jk}^+ คือ ค่าความน่าจะเป็นของผู้สอบที่มีความสามารถระดับ θ จะตอบข้อสอบข้อที่ j ได้คะแนนในลำดับชั้น k หรือสูงกว่า

θ คือ ระดับความสามารถของผู้สอบ

$$\text{และ } P_{jk}^+(\theta) = \frac{e^{[Da_j(\theta - b_{jk})]}}{1 + e^{[Da_j(\theta - b_{jk})]}} \quad k = 0, 1, \dots, m \quad (74)$$

$$P_{j0}^+ = 1.0, \quad P_{j,m+1}^+ = 0.0$$

โดยที่ D คือ ค่าคงที่ของสเกล (scaling constant) เมื่อปรับโค้งฟังก์ชันของโมเดลโลจิสติก จากโมเดลนอร์มอล ออโงไฟฟ มีค่าเท่ากับ 1.70

m คือ ลำดับชั้นคะแนนสูงสุด

a_j คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ j

b_{jk} คือ ค่าความยากของข้อสอบข้อที่ j ในลำดับชั้นที่ k

e คือ ค่าคงที่ มีค่าเท่ากับ 2.7183

2.1.9 คำนวณค่าฟังก์ชันสารสนเทศของแบบทดสอบเทียบคะแนนแต่ละฉบับ (TIF) โดยนำค่าพารามิเตอร์ของข้อสอบจากแบบทดสอบแต่ละฉบับมาคำนวณ ณ ระดับความสามารถต่าง ๆ มีสูตรการคำนวณในโมเดลเกรต เรสพอนส์ (Dodd, Ayala and Koch, 1955 : 10-11) ดังต่อไปนี้

$$I(\theta) = \sum_{j=1}^J I_j(\theta) \quad (75)$$

เมื่อ $I_j(\theta)$ คือ ฟังก์ชันสารสนเทศของข้อสอบ (IIF) ตั้งแต่ข้อ 1 ถึงข้อที่ j

$$I_j(\theta) = \sum_{k=1}^{m_j} \frac{[p'_{jk}(\theta)]^2}{P_{jk}(\theta)} \quad (76)$$

เมื่อ $P'_{jk}(\theta)$ คืออนุพันธ์อันดับหนึ่งของ $P_{jk}(\theta)$

และ $P_{jk}(\theta)$ มีความหมายตาม 2.1.8

วิเคราะห์ค่าโดยใช้โปรแกรม MULTLOG version 7.03 for windows โดยโปรแกรมกำหนดค่า θ อยู่ในช่วง -3 ถึง $+3$ มีค่าแตกต่างกันช่วงละ .20

2.2 การสร้างตารางเทียบคะแนนแบบทดสอบ

เนื่องจากการวิจัยครั้งนี้ ศึกษาวิธีการเทียบคะแนน 5 วิธี ซึ่งเป็นวิธีตามทฤษฎีการทดสอบแบบมาตรฐานเดิม 2 วิธี คือ การเทียบคะแนนเชิงเส้นตรง และการเทียบคะแนนอิกวิเปอร์เซ็นไทล์ และ เป็นวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ 3 วิธี คือ วิธีการเทียบคะแนนด้วยการปรับค่าพารามิเตอร์พร้อมกัน วิธีการเทียบคะแนนโดยใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (วิธี mean และ sigma) ซึ่งเป็นรูปแบบของ มาร์โค (Marco. 1977 : 140 - 141) และพัฒนาให้สอดคล้องกับโมเดลเกรด เรสพอนส์ (Coken and Kim. 1998 : 118) และวิธีการเทียบคะแนนด้วยโค้งลักษณะแบบทดสอบซึ่งพัฒนาโดย เบเกอร์ (Baker. 1992: 89 - 92) รายละเอียดการสร้างตารางเทียบคะแนนในแต่ละวิธี ดังต่อไปนี้

2.2.1 การเทียบคะแนนเชิงเส้นตรง

ตามสูตรซึ่งใช้ในแบบแผนการรวบรวมข้อมูลของแองกอฟฟ์ รูปแบบที่ 3 (Angoff. 1984:104-109) ซึ่งเป็นแบบแผนกลุ่มสุ่มที่ใช้แบบทดสอบร่วม มีขั้นตอนในการดำเนินการ ดังนี้

2.2.1.1 ทดสอบความแตกต่างของค่าความเชื่อมั่นระหว่างแบบทดสอบเทียบคะแนน

เนื่องจากแบบแผนการรวบรวมข้อมูลของแองกอฟฟ์ รูปแบบที่ 3 ได้คำนวณค่าความชัน และจุดตัดแกนของสมการเชิงเส้นตรง แยกตามระดับความเชื่อมั่นของแบบทดสอบเทียบคะแนนว่ามีความแตกต่างกันหรือไม่ ดังนั้นในการศึกษานี้ จึงทดสอบความแตกต่างของค่าความเชื่อมั่น ที่ประมาณค่าโดยใช้สัมประสิทธิ์แอลฟาของครอนบัค เมื่อแบบทดสอบทั้งสองฉบับใช้กลุ่มตัวอย่างที่เป็นอิสระต่อกัน สถิติทดสอบที่ใช้ (Feldt. 1969 : 365) คือ

$$W = \frac{1 - r_2}{1 - r_1} \quad (77)$$

ค่าวิกฤติ คือ Central F ที่มี df เท่ากับ $N_1 - 1$ และ $N_2 - 1$ (Feldt. 1969 : 369)

โดยที่ N_1 = จำนวนผู้สอบในแบบทดสอบฉบับ X
 N_2 = จำนวนผู้สอบในแบบทดสอบฉบับ Y
 r_1 = ความเชื่อมั่นของแบบทดสอบฉบับ X
 r_2 = ความเชื่อมั่นของแบบทดสอบฉบับ Y

2.2.1.2 คำนวณค่าความชัน (Slope) และจุดตัดแกน (Intercept) ของสมการเส้นตรง เพื่อแปลงคะแนนของสเกล Y ให้อยู่ในสเกล X จากสูตร 3A ของแองกอฟฟ์ คือ

$$A_{xy} = (\sigma_x^2) / (\sigma_y^2) \quad (78)$$

$$B_{xy} = \mu_x - A_{xy}\mu_y \quad (79)$$

เมื่อ

$$\begin{aligned} \mu_x &= M_{x\alpha} + b_{xu\alpha}(\mu_U - M_{U\alpha}) \\ \sigma_x^2 &= S_{x\alpha}^2 + b_{xu\alpha}^2(\sigma_U^2 - S_{U\alpha}^2) \\ \mu_y &= M_{y\beta} + b_{yu\beta}(\mu_U - M_{U\beta}) \\ \sigma_y^2 &= S_{y\beta}^2 + b_{yu\beta}^2(\sigma_U^2 - S_{U\beta}^2) \end{aligned}$$

โดยที่ $\mu_U = M_{Ut}$; $\sigma_U^2 = S_{Ut}^2$; $t = \alpha + \beta$

μ_x, σ_x^2 คือ ค่าประมาณคะแนนเฉลี่ย และความแปรปรวนของกลุ่มตัวอย่างที่ทำแบบทดสอบฉบับ X ที่ปรับแล้ว

μ_y, σ_y^2 คือ ค่าประมาณคะแนนเฉลี่ย และความแปรปรวนของกลุ่มตัวอย่างที่ทำแบบทดสอบฉบับ Y ที่ปรับแล้ว

$M_{x\alpha}, S_{x\alpha}^2$ คือ ค่าเฉลี่ย และความแปรปรวนของคะแนนฉบับ X

$M_{y\beta}, S_{y\beta}^2$ คือ ค่าเฉลี่ย และความแปรปรวนของคะแนนฉบับ Y

$M_{U\alpha}, S_{U\alpha}^2$ คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนแบบทดสอบร่วม U จากกลุ่ม α

$M_{U\beta}, S_{U\beta}^2$ คือ ค่าเฉลี่ย และความแปรปรวนของคะแนนแบบทดสอบร่วม U จากกลุ่ม β

M_{Ut}, S_{Ut}^2 คือ ค่าเฉลี่ย และความแปรปรวนของคะแนนแบบทดสอบร่วม U จากกลุ่มรวม t

2.2.1.2 นำค่าพารามิเตอร์มาสร้างสมการเส้นตรงเพื่อแปลงคะแนน Y ให้อยู่ในมาตราของ X

$$X^* = e(y) = AY + B \quad (80)$$

2.2.1.3 สร้างตารางเทียบคะแนนจากแบบทดสอบฉบับ Y ไปยังคะแนนตามมาตราคะแนนของแบบทดสอบฉบับ X

การดำเนินการในขั้นตอนที่ 2.2.1.1 – 2.2.1.3 ใช้โปรแกรมสำเร็จรูป SPSS

Version 7.50

2.2.2 การเทียบคะแนนอีควิเปอร์เซ็นไทล์

ขั้นตอนการเทียบคะแนนอีควิเปอร์เซ็นไทล์ ในแบบแผนกลุ่มสุ่มที่ใช้แบบทดสอบรวม โดยกลุ่ม α สอบแบบทดสอบฉบับ X และแบบทดสอบรวม U กลุ่ม β สอบแบบทดสอบฉบับ Y และแบบทดสอบรวม U คือการประมาณค่าความถี่ในแบบทดสอบฉบับ X และ Y สำหรับกลุ่มรวม $t(\alpha + \beta)$ (Equipercentile Analog – Frequency Estimation) (Angoff, 1982 : 62 – 63) มีขั้นตอน ดังนี้

2.2.2.1 รวมการแจกแจงคะแนนของแบบทดสอบรวม U ที่สอบโดยกลุ่ม α, β เป็นการแจกแจงคะแนนของกลุ่ม $t(\alpha + \beta)$

2.2.2.2 หาอัตราส่วนความถี่ $f_{it}/f_{i\alpha}$ และ $f_{it}/f_{i\beta}$ ที่ทุกช่วงคะแนน i

2.2.2.3 แจกแจงความถี่ของแบบทดสอบรวม U ที่ทุกช่วงคะแนนของแบบทดสอบฉบับ X และ Y โดยใช้ตารางแจกแจงความถี่สองทาง ;

2.2.2.4 คูณความถี่ของแบบทดสอบรวม U ในแต่ละช่วงคะแนนของแบบทดสอบฉบับ X ด้วยอัตราส่วน $f_{it}/f_{i\alpha}$

2.2.2.5 คูณความถี่ของแบบทดสอบรวม U ในแต่ละช่วงคะแนนของแบบทดสอบฉบับ Y ด้วยอัตราส่วน $f_{it}/f_{i\beta}$

2.2.2.6 รวมความถี่แต่ละค่าคะแนนในฉบับ X และ Y จะได้รับการแจกแจงของ X และ Y สำหรับกลุ่ม t

2.2.2.7 คำนวณสัดส่วนความถี่และสัดส่วนความถี่สะสมของแบบทดสอบเทียบคะแนนทั้งสองฉบับที่ปรับการแจกแจงแล้ว

2.2.2.8 เทียบคะแนนโดยใช้กระบวนการวิเคราะห์ (Analytic Procedures) หาดำแหน่งเปอร์เซ็นไทล์ของแบบทดสอบฉบับ Y ด้วยสูตร (Kolen and Brennan, 1995 : 39)

$$Q(y) = 100[G(y - 1) + (g(y)/2)] \quad (81)$$

เมื่อ $G(y)$ คือ สัดส่วนของผู้สอบที่ได้คะแนนเท่ากับหรือน้อยกว่าคะแนน y

$g(y)$ คือ สัดส่วนของผู้สอบที่ได้คะแนน y

y คือ คะแนนของผู้สอบที่เป็นจำนวนเต็ม

มีค่า = 0, 1, 2, ..., K_y

K_y คือ คะแนนเต็มสูงสุดเท่ากับ 150

2.2.2.9 หาฟังก์ชันผกผันของตำแหน่งเปอร์เซ็นต์ไทล์ เพื่อคำนวณคะแนนของแบบทดสอบฉบับ X ที่มีตำแหน่งเปอร์เซ็นต์ไทล์เดียวกับคะแนนของแบบทดสอบฉบับ Y โดยใช้สูตร (Kolen and Brennan. 1995 : 45)

$$\begin{aligned} e_x(y) &= x = P^{-1}[Q(y)] \\ &= \frac{Q(y)/100 - F(x_u^* - 1)}{F(x_u^*) - F(x_u^* - 1)} + (x_u^* - .5), \quad 0 \leq Q(y) < 100 \\ &= K_y + .5, \quad Q(y) = 100 \end{aligned} \quad (82)$$

โดยที่ $Q(y)$ คือ ตำแหน่งเปอร์เซ็นต์ไทล์ของคะแนน y

x_u^* คือ คะแนนที่เป็นจำนวนเต็มที่มีค่าน้อยสุด ซึ่ง $[100F(x_u^*)]$ มีขนาดใหญ่กว่า $Q(y)$

$F(x_u^*)$ คือ สัดส่วนของผู้สอบที่ได้คะแนนน้อยกว่าหรือเท่ากับ คะแนน x_u^*

2.2.2.10 เนื่องจากการแจกแจงคะแนนดิบของแบบทดสอบฉบับ X มีช่วงคะแนน 60 - 150 และฉบับ Y มีช่วงคะแนน 59 - 150 ทำให้ช่วงคะแนน 0 - 58 ในแบบทดสอบทั้งสองฉบับมีความถี่เป็นศูนย์ และในช่วงคะแนน 60-150 ของแบบทดสอบฉบับ X และช่วงคะแนน 59-150 ของแบบทดสอบฉบับ Y อาจมีบางระดับคะแนนที่มีความถี่เป็นศูนย์ เพื่อแก้ปัญหาดังกล่าวจึงใช้การปรับความถี่แต่ละระดับคะแนนด้วยค่า $adj = 10^{-6}$ ซึ่งไม่ทำให้เกิดความลำเอียงของการเทียบคะแนน (Kolen and Brennan. 1995 : 47) และผลรวมของความถี่สัมพัทธ์ที่ปรับแล้วจะต้องมีค่าเท่ากับ 1 ความถี่สัมพัทธ์ที่ปรับแล้วของแบบทดสอบฉบับ Y ดังสูตร (Kolen and Brennan. 1995 : 47)

$$g_{adj}(y) = \frac{g(y) + adj}{1 + (K_y + 1)adj} \quad (83)$$

โดยที่ $g(y)$ คือ ความถี่สัมพัทธ์ที่สังเกตได้ของแบบทดสอบฉบับ Y

adj คือ ค่าที่ใช้ปรับเท่ากับ 10^{-6}

และใช้กระบวนการเดียวกันปรับความถี่สัมพัทธ์ของแบบทดสอบฉบับ X โดยขั้นตอนนี้จะดำเนินการต่อจาก 2.2.2.7 ก่อนที่จะดำเนินการเทียบคะแนน

2.2.2.11 สร้างตารางเทียบคะแนนดิบจากแบบทดสอบฉบับ Y ปรับสเกลสู่แบบทดสอบฉบับ X

การดำเนินการในขั้นตอนที่ 2.2.2.1–2.2.2.11 ใช้โปรแกรม

Excel

2.2.3 การเทียบคะแนนตามทฤษฎีการตอบข้อสอบวิธีปรับค่าพารามิเตอร์พร้อมกัน

การประมาณค่าพารามิเตอร์ข้อสอบจากแบบทดสอบเทียบคะแนนฉบับ X และฉบับ Y ด้วยการวิเคราะห์ค่าพร้อม ๆ กันในโมเดลเกรด เรสพอนส์จะทำให้มาตรวัดของพารามิเตอร์ความสามารถ (θ) ของแบบทดสอบฉบับ X และ Y อยู่บนมาตรวัดเดียวกัน ดังนั้นไม่จำเป็นต้องแปลงคะแนน θ ให้อยู่ในมาตรวัดเดียวกัน แต่การรายงานผลการเทียบคะแนนความสามารถบนมาตรวัดของคะแนนความสามารถ ซึ่งมีค่าที่เป็นไปได้ทั้งบวกและลบ ทำให้มีปัญหาในการแปลความหมาย จึงมีการเทียบคะแนนจริงและคะแนนสังเกต เพื่อให้เข้าใจได้ง่ายขึ้นเมื่อนำไปใช้ โดยมีขั้นตอนของการเทียบคะแนนดังนี้

2.2.3.1 การเทียบคะแนนจริง

เนื่องจากพารามิเตอร์ความสามารถของผู้สอบแต่ละคนที่สอบแบบทดสอบฉบับ X และ Y อยู่บนมาตรวัดเดียวกัน ดังนั้น ณ ระดับ θ เดียวกัน คะแนนจริงที่ได้จากแบบทดสอบฉบับ Y จึงเทียบเท่าคะแนนจริงฉบับ X มีขั้นตอนในการเทียบคะแนนจริงดังนี้

1) คำนวณคะแนนจริงของผู้สอบฉบับ Y ที่มีระดับความสามารถ θ โดยใช้สูตร (Lee and others. 2001 : 358)

$$T(\theta) = \sum_{j=1}^J \sum_{k=1}^K u_{jk} P_{jk}(\theta) \quad (84)$$

โดยที่ u_{jk} คือ คะแนนที่ได้ในลำดับขั้นคะแนน k ของข้อที่ j

$P_{jk}(\theta)$ คือ ความน่าจะเป็นที่จะตอบข้อสอบได้คะแนนในลำดับขั้น k ข้อที่ j ณ ระดับความสามารถ θ ในโมเดลเกรด เรสพอนส์

$$\text{และ } P_{jk}(\theta) = P_{jk}^+(\theta) - P_{j,k+1}^-(\theta)$$

2) หาฟังก์ชันผกผันของคะแนนจริงในฉบับ Y จะทำให้ได้คะแนนความสามารถ (θ_i) ที่สมนัยกับคะแนนจริงของฉบับ Y

3) จาก θ_i ที่หาได้ในแต่ละระดับคะแนนจริงของฉบับ Y นำมาคำนวณคะแนนจริงของผู้สอบในแบบทดสอบฉบับ X จะได้คะแนนจริงของแบบทดสอบฉบับ X ที่สมนัยกับฉบับ Y

การดำเนินการในขั้นตอนที่ 1) – 3) ใช้โปรแกรมสำเร็จรูป SPSS Version 7.50 ด้วยการเขียน Syntax

2.2.3.2 การเทียบคะแนนสังเกต

1) ประมาณการแจกแจงคะแนนสังเกตเชิงทฤษฎีในแบบทดสอบฉบับ X และ Y ในแต่ละระดับ θ โดยใช้ค่าพารามิเตอร์ข้อสอบของแบบทดสอบฉบับ X และ Y และพารามิเตอร์ความสามารถของผู้สอบฉบับ X และ Y (โดยพารามิเตอร์ความสามารถที่ได้จากแบบทดสอบฉบับ Y อยู่บนมาตรฐานวัดเดียวกับฉบับ X)

การแจกแจงของคะแนนสังเกตในแต่ละระดับความสามารถของผู้สอบสร้างโดยใช้ การแจกแจง Multinomial สำหรับคะแนนที่มีหลายค่า โดย แวง และคณะ (Wang, Kolen & Harris. 2000:146) ได้ขยายจากขั้นตอนของลอร์ดและวิงเกอร์สกี (Lord and wingersky. 1984 : 454) ซึ่งใช้การแจกแจง Compound binomial กับการเทียบคะแนนทวิภาค ดังมีสูตร (Lee and others. 2001 : 360) คือ

สำหรับข้อสอบ 1 ข้อ

$$P_1(X = x|\theta) = P(U_1 = x|\theta) \quad (85)$$

เมื่อ $x = 0, 1, 2, 3, 4, 5$

สำหรับข้อสอบ 2, 3, ..., 30 ข้อ

$$P_k(X = x|\theta) = \sum_{u=0}^n P_{k-1}(X = x - u)P(U_k = u|\theta) \quad (86)$$

เมื่อ $x = 0, 1, \dots, \sum_{k=1}^K n_k$

โดยที่ U_k คือ ตัวแปรสุ่มสำหรับคะแนนที่เป็นไปได้ของข้อสอบ k ข้อ ซึ่งมีค่าเท่ากับ $0, 1, 2, \dots, n_k$

$P(U_i = x|\theta)$ คือ $P_{jk}(\theta)$ ในสมการ 73

การแจกแจงคะแนนสังเกตของแบบทดสอบแต่ละฉบับต้องคำนวณค่าความน่าจะเป็นจนถึง 30 ข้อ โดยความน่าจะเป็นที่ได้คะแนนที่อยู่นอกเหนือ $X = x - u$ จะถูกกำหนดให้มีค่าเป็น 0 โดยเงื่อนไขดังกล่าว $P_k(Y = y|\theta)$ ในแบบทดสอบฉบับ Y ก็ถูกดำเนินการเช่นเดียวกัน

2) เมื่อได้การแจกแจงคะแนนสังเกตของแบบทดสอบเทียบคะแนนฉบับ X และ Y ในแต่ละระดับความสามารถ และแต่ละระดับคะแนน ($0 - 150$ คะแนน) ($f(x|\theta)$) แล้ว จึงรวมการแจกแจงคะแนนสังเกตในทุกระดับความสามารถ ($f(x)$) โดยใช้สูตร (Lee and others. 2001 : 360)

$$f(x) = \sum_{\theta} f(x|\theta)\psi(\theta) \quad (87)$$

เมื่อ $\psi(\theta)$ คือ การแจกแจงของ θ ที่ได้จากการวิเคราะห์ด้วยการแจกแจงที่ไม่ต่อเนื่องโดยกำหนดระยะเท่า ๆ กัน

$f(x|\theta)$ คือ การแจกแจงคะแนน ณ ระดับความสามารถ θ ที่ได้จากการสมการ 86 เช่นเดียวกับการแจกแจงคะแนนของแบบทดสอบฉบับ Y ($g(y)$) ก็สามารถหาได้โดยใช้สมการข้างต้น

การดำเนินการในขั้นตอนที่ 1) และ 2) วิเคราะห์ด้วยโปรแกรม SPSS Version 7.50 โดยการเขียน Syntax

3) เทียบคะแนนสังเกตจากแบบทดสอบฉบับ Y ไปยังฉบับ X โดยใช้วิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์ โดยเริ่มดำเนินการคำนวณค่าความน่าจะเป็นสะสมในแต่ละระดับคะแนนในแบบทดสอบเทียบคะแนนทั้งสองฉบับ แล้วจึงดำเนินการตามข้อ 2.2.2.8 – 2.2.2.11 โดยใช้โปรแกรม Excel

2.2.4 การเทียบคะแนนตามทฤษฎีการตอบข้อสอบ วิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (วิธี mean และ sigma)

การเทียบคะแนนตามวิธีการนี้เกิดจากการประมาณค่าพารามิเตอร์แบบทดสอบเทียบคะแนนแต่ละฉบับแยกกัน จึงต้องสร้างสมการแปลงค่า θ เชิงเส้น โดยประมาณค่าสัมประสิทธิ์การเทียบคะแนน (A , B) ด้วยค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของข้อสอบรวม ดังมีรายละเอียดต่อไปนี้

2.2.4.1 ประมาณค่าพารามิเตอร์ข้อสอบในแบบทดสอบเทียบคะแนนแต่ละฉบับและประมาณค่าพารามิเตอร์ความสามารถของผู้สอบแต่ละคน โดยใช้โปรแกรม MULTILOG Version 7.03 for Windows

2.2.4.2 แปลงค่าพารามิเตอร์ความสามารถที่ได้จากแบบทดสอบเทียบคะแนนแต่ละฉบับ ให้อยู่บนสเกลเดียวกันโดยใช้พารามิเตอร์ข้อสอบรวมในฉบับ X และ Y มีสูตร (Cohen and Kim. 1998 : 118) คือ

$$\bar{b}_1 = A\bar{b}_2 + B \quad (88)$$

$$S(b_1) = AS(b_2) \quad (89)$$

โดยที่ $\bar{b}_1$ และ $\bar{b}_2$ คือ ค่าเฉลี่ยของค่าความยากของข้อสอบรวมในแบบทดสอบฉบับ X ฉบับ Y ตามลำดับ

$S(b_1)$ และ $S(b_2)$ คือ ส่วนเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบรวมในแบบทดสอบฉบับ X และฉบับ Y ตามลำดับ

ดังนั้น A และ B หาได้จาก

$$A = S(b_1)/S(b_2) \quad \text{และ} \quad (90)$$

$$B = \bar{b}_1 - A\bar{b}_2 \quad (91)$$

ดังนั้นสามารถแปลงพารามิเตอร์ความสามารถให้อยู่ในมาตรฐานเดียวกันได้โดยใช้สูตร

$$\theta'_x = A\theta_y + B \quad (92)$$

2.2.4.3 การเทียบคะแนนจริง ตามขั้นตอนดังนี้

1) กำหนดคะแนนจริงในมาตรฐานของแบบทดสอบฉบับ Y ทุกระดับคะแนน (1-150)

2) หาฟังก์ชันผกผันของคะแนนจริงในแบบทดสอบฉบับ Y จะได้คะแนนความสามารถ (θ_i) ที่สมนัยกับคะแนนจริงของแบบทดสอบฉบับ Y

3) นำ θ_i ใน 2) ไปแปลงในสมการที่ 92 ได้ θ ในมาตรฐานเดียวกันกับฉบับ X แล้วคำนวณคะแนนจริงของแบบทดสอบฉบับ X ที่สมนัยกับฉบับ Y ด้วยพารามิเตอร์ข้อสอบฉบับ X และพารามิเตอร์ความสามารถที่แปลงค่าแล้ว

2.2.4.4 การเทียบคะแนนสังเกต ตามขั้นตอน ดังนี้

1) แจกแจงพารามิเตอร์ความสามารถที่ได้จากผู้สอบในฉบับ X และ Y โดยการจัดชั้นคะแนนความสามารถ และใช้จุดกลางของชั้นเป็นตัวแทนของพารามิเตอร์ความสามารถ จะได้การแจกแจงความน่าจะเป็นของพารามิเตอร์ความสามารถ ($\psi(\theta)$)

2) ประมาณการแจกแจงคะแนนสังเกตเชิงทฤษฎีในแบบทดสอบเทียบคะแนนฉบับ X ในแต่ละระดับความสามารถและแต่ละระดับคะแนน ($f(x|\theta)$) ใช้สูตร 85 และ 86 การคำนวณใช้ θ ที่เป็นตัวแทนระดับความสามารถ คือ จุดกลางของชั้นคะแนน θ ในแบบทดสอบฉบับ X และพารามิเตอร์ข้อสอบของฉบับ X ในการคำนวณตามสูตรข้างต้น

3) แปลง θ ที่เป็นจุดกลางของชั้นคะแนนในแบบทดสอบฉบับ Y ด้วยสมการแปลงตามสูตร 92 ได้ θ ที่อยู่ในมาตรฐานเดียวกันกับฉบับ X

4) ประมาณการแจกแจงคะแนนสังเกตเชิงทฤษฎีในแบบทดสอบเทียบคะแนนฉบับ Y ในแต่ละระดับความสามารถและแต่ละระดับคะแนน ($g(y|\theta)$) ใช้สูตร 85 และ 86 การคำนวณใช้ θ ที่เป็นระดับความสามารถที่แปลงแล้วใน 3) และพารามิเตอร์ข้อสอบของฉบับ Y ในการคำนวณตามสูตรข้างต้น

5) คำนวณค่าความน่าจะเป็นในแต่ละระดับคะแนน (0 - 150 คะแนน) ของแบบทดสอบเทียบคะแนนฉบับ X และ Y ($f(x), g(y)$) โดยใช้สูตร 87 ซึ่งเป็นผลรวมของความน่าจะเป็นในแต่ละระดับคะแนนสังเกตในทุกะดับความสามารถ

การดำเนินการในขั้นตอนที่ 1) – 5) วิเคราะห์ด้วยโปรแกรม SPSS Version 7.50 โดยการเขียน Syntax

6) เทียบคะแนนจากแบบทดสอบฉบับ Y ไปยังฉบับ X โดยใช้วิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์ โดยเริ่มดำเนินการคำนวณค่าความน่าจะเป็นสะสมในแต่ละระดับคะแนนในแบบทดสอบเทียบคะแนนทั้งสองฉบับ แล้วจึงดำเนินการตามข้อ 2.2.2.8 – 2.2.2.11 โดยใช้โปรแกรม Excel

2.2.5 การเทียบคะแนนตามทฤษฎีการตอบข้อสอบ วิธีฟังก์ชันการตอบแบบทดสอบ

การเทียบคะแนนตามวิธีการนี้ เกิดจากการประมาณค่าพารามิเตอร์ของแบบทดสอบเทียบคะแนนแต่ละฉบับแยกกัน จึงต้องสร้างสมการแปลงพารามิเตอร์ θ เชิงเส้น โดยประมาณค่าสัมประสิทธิ์การเทียบคะแนน (A,K) จากฟังก์ชันการตอบแบบทดสอบ ซึ่งเบเกอร์ (Baker. 1992 : 89 - 92) ได้พัฒนาสูตร และสร้างโปรแกรม Equate 2.1 (Baker. 1993b : 20) เพื่อประมาณค่าสัมประสิทธิ์การเทียบคะแนน (A,K) มีวิธีดำเนินการดังต่อไปนี้

2.2.5.1 ดำเนินการตามขั้นตอนที่ 2.2.4.1

2.2.5.2 ประมาณค่าสัมประสิทธิ์การเทียบคะแนนด้วยสารสนเทศของข้อสอบรวม โดยการทำให้ค่าฟังก์ชัน loss กำลังสองมีค่าต่ำสุด ดังสูตรต่อไปนี้ (Cohen and Kim. 1998 : 120)

$$F = \frac{1}{N} \sum_{i=1}^N (T_{i1} - T_{i2}^*)^2 \quad (93)$$

เมื่อ N คือ จำนวนจุด θ_i ที่กำหนดบนมาตราวัด

T_{i1} และ T_{i2}^* คือ ค่าคาดหวังของคะแนนสำหรับกลุ่ม 1 และ 2 ตามลำดับ กำหนดไว้ดังนี้

$$T_{i1} = \sum_{j=1}^{n_c} \sum_{k=1}^{m_j} u_{jk} P_{jk1}(\theta_i)$$

$$T_{i2} = \sum_{j=1}^{n_c} \sum_{k=1}^{m_j} u_{jk} P_{jk2}^*(\theta_i)$$

เมื่อ u_{jk} คือ น้ำหนักที่กำหนดต่อลำดับชั้นการตอบ k สำหรับข้อสอบข้อที่ j

น้ำหนักคือดัชนีที่เป็นจำนวนเต็มของลำดับชั้นการตอบ

(category)

ค่าประมาณสัมประสิทธิ์การเทียบคะแนนที่ได้จากการคำนวณด้วยโปรแกรม EQUATE 2.1 ใช้ได้กับค่าประมาณพารามิเตอร์ข้อสอบและความสามารถจากโปรแกรม MULTILOG เท่านั้น

2.2.5.3 ดำเนินการเทียบคะแนนจริงตามขั้นตอนใน 2.2.4.3 โดยในขั้นตอน 3) ใช้พารามิเตอร์ θ ที่แปลงจากสัมประสิทธิ์การเทียบคะแนนใน 2.2.5.2

2.2.5.4 ดำเนินการเทียบคะแนนสังเกตตามขั้นตอนใน 2.2.4.4 โดยในขั้นตอน 3) ใช้พารามิเตอร์ θ ที่แปลงจากสัมประสิทธิ์การเทียบคะแนนใน 2.2.5.2

2.3 การวิเคราะห์ผลการเทียบคะแนน

การศึกษาผลการเทียบคะแนนวิธีต่าง ๆ ศึกษาโดยใช้เกณฑ์ ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ดัชนีคะแนนความแตกต่าง (difference scores indices) และดัชนีความแตกต่าง (C) จากกลุ่มสอบทานผล มีรายละเอียดดังต่อไปนี้

2.3.1 ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม

การตรวจสอบความคล้ายคลึงกันของการแจกแจงคะแนนสะสมเป็นเกณฑ์ที่ใช้ตรวจสอบคุณสมบัติของการเทียบคะแนนระหว่าง การแจกแจงคะแนนสะสมของคะแนนแปลงจากการเทียบคะแนนวิธีต่าง ๆ กับการแจกแจงคะแนนสะสมจากแบบทดสอบฉบับเป้าหมาย (ฉบับ X) หากมีความคล้ายคลึงกัน แสดงว่า วิธีการเทียบคะแนนนั้นมีคุณสมบัติของการเทียบคะแนนซึ่งแสดงถึงควมมีประสิทธิภาพของการเทียบคะแนน โดยตรวจสอบด้วยสถิติทดสอบนันทราเมตริกซ์ Kolmogorov Smirnov two-Sample (Jaeger, 1981 : 26-27) ซึ่งทดสอบโดยเปรียบเทียบลักษณะของประชากรทั้งตำแหน่ง (location) และรูปร่าง (shape) ว่าแตกต่างกันหรือไม่ (กัลยา วาณิชยบัญชา.2542 : 300)

2.3.2 ดัชนีคะแนนความแตกต่าง เป็นดัชนีแสดงผลรวมของการเปรียบเทียบคะแนนที่แปลงจากการเทียบคะแนนพื้นฐาน (baseline) ในการศึกษาครั้งนี้ได้แก่วิธีการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้แก่ วิธีการเทียบคะแนนเชิงเส้นตรง และวิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์ กับคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบแต่ละวิธี ได้แก่ การเทียบคะแนนจริงและคะแนนสังเกตด้วยวิธีปรับค่าพารามิเตอร์พร้อมกัน การเทียบคะแนนจริงและคะแนนสังเกตด้วยวิธีโค้งลักษณะแบบทดสอบ และการเทียบคะแนนจริงและคะแนนสังเกตด้วยวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก มี 6 ดัชนี จำแนกเป็นสองกลุ่ม คือ

2.3.2.1 ดัชนีคะแนนความแตกต่างถ่วงน้ำหนัก

ดัชนีคะแนนความแตกต่างถ่วงน้ำหนักด้วยความถี่ในแต่ละระดับคะแนนของแบบทดสอบฉบับ Y เป็นการกำหนดน้ำหนักความแตกต่างในแต่ละระดับคะแนนด้วยความถี่แตกต่างกัน ประกอบด้วยดัชนี 3 ดัชนีคือ

1) รากของค่าเฉลี่ยความแตกต่างกำลังสองถ่วงน้ำหนัก (RMSD) มีสูตร คือ (Harris and Crouse, 1993 : 203)

$$RMSD_w = \left(\sum_i f_i (A_i - B_i)^2 / \sum_i f_i \right)^{1/2} \quad (94)$$

เมื่อ A_i คือ คะแนนแปลง i จากวิธีการเทียบคะแนนตามทฤษฎีการตอบ
ข้อสอบ

B_i คือ คะแนนแปลง i จากวิธีการเทียบคะแนนพื้นฐาน

i คือ ค่าของคะแนนดิบในช่วงที่เป็นไปได้

f_i คือ ความถี่ในแต่ละระดับคะแนนของแบบทดสอบฉบับ Y

2) ค่าเฉลี่ยสัมบูรณ์ของความแตกต่างถ่วงน้ำหนัก (MAD)

มีสูตร คือ (Harris and Crouse. 1993 : 203)

$$MAD_w = \frac{\sum_i f_i |A_i - B_i|}{\sum_i f_i} \quad (95)$$

(ค่าต่าง ๆ มีความหมายตาม 1))

3) ค่าเฉลี่ยความแตกต่างถ่วงน้ำหนัก (MSD) มีสูตร คือ

(Harris and Crouse. 1993 : 203)

$$MSD_w = \frac{\sum_i f_i (A_i - B_i)}{\sum_i f_i} \quad (96)$$

(ค่าต่าง ๆ มีความหมายตาม 1))

2.3.2.2 ดัชนีคะแนนความแตกต่างไม่ถ่วงน้ำหนัก

ความแตกต่างของคะแนนที่อยู่ต้นหรือปลายช่วง ซึ่งไม่มี
ผู้สอบทำคะแนนได้ จะไม่ถูกคำนวณความแตกต่างถ้าใช้ดัชนีถ่วงน้ำหนัก ดัชนีไม่ถ่วงน้ำหนักจึง
เป็นดัชนีที่ตรวจสอบความแตกต่างที่เกิดขึ้นตลอดช่วงคะแนนของมาตรวัดที่เป็นไปได้ ประกอบ
ด้วยดัชนี 3 ดัชนี คือ

4) รากของค่าเฉลี่ยความแตกต่างกำลังสองไม่ถ่วงน้ำหนัก

(RMSD) มีสูตรคือ (Harris and Crouse. 1993 : 203)

$$RMSD_u = \left(\sum_i (A_i - B_i)^2 / k \right)^{1/2} \quad (97)$$

เมื่อ k คือ จำนวนระดับคะแนน (score point) ของมาตรวัด

(สำหรับค่าอื่นมีความหมายเช่นเดียวกับ 1))

5) ค่าเฉลี่ยสัมบูรณ์ของความแตกต่างไม่ถ่วงน้ำหนัก (MAD)

มีสูตรคือ (Harris and Crouse. 1993 : 204)

$$MAD_u = \frac{\sum_i |A_i - B_i|}{k} \quad (98)$$

(สำหรับค่าอื่นมีความหมายเดียวกับ 4))

6) ค่าเฉลี่ยความแตกต่างไม่ถ่วงน้ำหนัก (MSD) มีสูตร คือ (Harris and Crouse. 1993 : 204)

$$MSD_U = \frac{\sum_i (A_i - B_i)}{k} \quad (99)$$

(สำหรับค่าอื่นมีความหมายเดียวกับ 4)

2.3.3 การวิเคราะห์กลุ่มสอบทานผล

วิธีวิเคราะห์กลุ่มสอบทานผล เป็นวิธีประเมินความคงที่ (Stability) ในการเทียบคะแนนโดยโคเลนและวิทนี (Kolen and Whitney. 1982: 284) เป็นผู้แนะนำให้ใช้ข้อมูลจากคะแนนของผู้สอบเองเป็นเกณฑ์ในการหาความแตกต่าง ข้อมูลเหล่านี้ได้มาจากการออกแบบใช้กลุ่มสอบทานผล ซึ่งสุ่มมาจากประชากรเดียวกันกับกลุ่มตัวอย่างเทียบคะแนน (ที่ใช้สร้างตารางเทียบคะแนน) และไม่มีหน่วยตัวอย่างที่ซ้ำกัน ได้รับการทดสอบด้วยแบบทดสอบเทียบคะแนนทั้งสองฉบับ โดยนำผลสอบจากแบบทดสอบเทียบคะแนนฉบับ Y มาแปลงคะแนนจากตารางเทียบคะแนนแต่ละวิธีให้เป็นคะแนนในแบบทดสอบฉบับ X (X^*) แล้ววัดความแตกต่างระหว่างคะแนนแปลงกับคะแนนสอบจากแบบทดสอบฉบับ X ซึ่งเป็นคะแนนเกณฑ์ ถ้าความแตกต่างโดยเฉลี่ยทั้งกลุ่มมีค่าน้อย แสดงว่าวิธีการเทียบคะแนนที่นำมาสร้างตารางคะแนนแปลงนั้นมีความเหมาะสมและเพียงพอที่จะให้ผลการแปลงคะแนนอย่างคงเส้นคงวา วิธีนี้สามารถใช้ตรวจสอบคุณภาพได้ทุกวิธีการเทียบคะแนน มีขั้นตอนดังนี้

2.3.3.1 คำนวณค่าดัชนีความแตกต่าง (Discrepancy Index: C) จากสูตรของปีเตอร์เซนและคณะ (Petersen and others. 1982 : 91)

$$C = \sum_i (X_i - X_i^*)^2 / nS_x^2 \quad (100)$$

เมื่อ C คือ ดัชนีความแตกต่าง

X_i คือ คะแนนจากแบบทดสอบฉบับ X ของผู้สอบคนที่ i

X_i^* คือ คะแนนจากแบบทดสอบฉบับ X ที่ได้จากการแปลงคะแนนของผู้สอบคนที่ i

n คือ จำนวนคนของกลุ่มสอบทานผล

S_x^2 คือ ค่าความแปรปรวนของการแจกแจงคะแนนจากแบบทดสอบฉบับ X

2.3.3.2 ประเมินค่าดัชนีความแตกต่าง (C) โดยอาศัยเกณฑ์การประเมินของปีเตอร์เซนและคณะ (Petersen and others. 1982 : 93-97) ซึ่งกำหนดคุณภาพในการเทียบคะแนนตามระดับการยอมรับ ดังนี้

ระดับน่าพอใจอย่างมาก เมื่อ $C \leq (.05 SD_x)^2$

ระดับน่าพอใจ เมื่อ $(.05 SD_x)^2 < C \leq (.10 SD_x)^2$

ระดับปานกลาง เมื่อ $(.10 SD_x)^2 < C \leq (.15 SD_x)^2$

ระดับไม่น่าพอใจ เมื่อ $(.15 SD_x)^2 < C \leq (.20 SD_x)^2$

ระดับไม่น่าพอใจอย่างยิ่ง เมื่อ $(.20 SD_x)^2 \leq C$

2.3.3.3 การทดสอบความแตกต่างของคะแนนแปลงกับคะแนนเกณฑ์

การทดสอบความแตกต่างของคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี กับคะแนนเกณฑ์ของกลุ่มสอบทานผล โดยใช้สถิติทดสอบ Friedman ซึ่งเป็นสถิติทดสอบนัยพาราเมตริกซ์ที่ไม่ต้องมีข้อตกลงเบื้องต้นของการแจกแจงความแตกต่างของคะแนน ซึ่งใช้ทดสอบความแตกต่างของข้อมูลที่ตัวแปรมากกว่าหรือเท่ากับ 2 ที่มีความสัมพันธ์กัน สถิติทดสอบมีการแจกแจงโดยประมาณแบบไคสแควร์ (กัลยา วาณิชย์บัญชา. 2542 : 309)

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาผลของการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่าตามวิธีการเทียบคะแนนในทฤษฎีการตอบข้อสอบ ในโมเดลเกรต เรสพอนส์ คือ (1) การเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกัน (2) การเทียบคะแนนวิธีโค้งลักษณะแบบทดสอบ (3) การเทียบคะแนนวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก แต่ละวิธีเทียบทั้งคะแนนจริงและคะแนนสังเกต และตามวิธีการในทฤษฎีการทดสอบแบบมาตรฐานเดิม คือ (4) การเทียบคะแนนเชิงเส้นตรง และ (5) การเทียบคะแนนอีควิเปอร์เซ็นไทล์ ผลการเทียบคะแนนพิจารณาจาก ความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงกับคะแนนจากแบบทดสอบเป้าหมาย ดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมกับการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ และดัชนีความแตกต่าง (C) ในแต่ละวิธีการเทียบคะแนนจากกลุ่มสอบ-ทานผล การนำเสนอผลการวิเคราะห์ข้อมูลแยกเป็น 5 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์แบบทดสอบที่ใช้ในการเทียบคะแนน

ตอนที่ 2 ผลการสร้างตารางเทียบคะแนนตามทฤษฎีการตอบข้อสอบและทฤษฎีการทดสอบแบบมาตรฐานเดิม

ตอนที่ 3 ผลการวิเคราะห์ความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนจากแบบทดสอบฉบับเป้าหมาย

ตอนที่ 4 ผลการวิเคราะห์ความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบและจากทฤษฎีการทดสอบแบบมาตรฐานเดิม

ตอนที่ 5 ผลการวิเคราะห์คุณภาพของการเทียบคะแนนจากกลุ่มสอบทานผล ในการนำเสนอผลการวิเคราะห์ข้อมูล เพื่อให้เกิดความเข้าใจตรงกันผู้วิจัยขอ กำหนดสัญลักษณ์และอักษรย่อที่ใช้แทนค่าสถิติและตัวแปรต่าง ๆ ดังนี้

Sk	แทน	ค่าความเบ้ของการแจกแจงความถี่ของคะแนน
Ku	แทน	ค่าความโด่งของการแจกแจงความถี่ของคะแนน
OPM	แทน	ความยากของแบบทดสอบคำนวณจากอัตราส่วนของค่าเฉลี่ยต่อคะแนนรวมสูงสุดของแบบทดสอบ
THY	แทน	ระดับความสามารถที่สมนัยกับระดับคะแนนจริงในแบบ-ทดสอบฉบับ Y

THYB	แทน	ระดับความสามารถที่แปลงด้วยสัมประสิทธิ์การเทียบ คะแนนวิธี MS
THYC	แทน	ระดับความสามารถที่แปลงด้วยสัมประสิทธิ์การเทียบ คะแนนวิธี TRF
THYS	แทน	ระดับความสามารถที่สมนัยกับระดับคะแนนจริงในแบบ- ทดสอบฉบับ Y ซึ่งวิเคราะห์พารามิเตอร์พร้อมฉบับ X
LIN	แทน	วิธีการเทียบคะแนนเชิงเส้นตรง
EQU	แทน	วิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์
TMS	แทน	การเทียบคะแนนจริงวิธีค่าเฉลี่ยและส่วนเบี่ยงเบน มาตรฐานของพารามิเตอร์ความความยาก
TTRF	แทน	การเทียบคะแนนจริงวิธีโค้งลักษณะแบบทดสอบ
TS	แทน	การเทียบคะแนนจริงวิธีปรับค่าพารามิเตอร์พร้อมกัน
OMS	แทน	การเทียบคะแนนสังเกตวิธีค่าเฉลี่ยและส่วนเบี่ยงเบน มาตรฐานของพารามิเตอร์ความยาก
OTRF	แทน	การเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบ
OS	แทน	การเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน
SY	แทน	คะแนน (สังเกต) สมมูลของแบบทดสอบฉบับ Y
STY	แทน	คะแนน (จริง) สมมูลของแบบทดสอบฉบับ Y
XBE	แทน	คะแนน (สังเกต) แปลงของแบบทดสอบฉบับ X จากการ เทียบคะแนนสังเกตวิธีค่าเฉลี่ยและส่วนเบี่ยงเบน มาตรฐานของพารามิเตอร์ความยาก
XCE	แทน	คะแนน (สังเกต) แปลงของแบบทดสอบฉบับ X จากการ เทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบ
XSE	แทน	คะแนน (สังเกต) แปลงของแบบทดสอบฉบับ X จากการ เทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน
TRUYB	แทน	คะแนน (จริง) แปลงของแบบทดสอบฉบับ X จากการ เทียบคะแนนจริงวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน ของพารามิเตอร์ความยาก
TRUYC	แทน	คะแนน (จริง) แปลงของแบบทดสอบฉบับ X จากการ เทียบคะแนนจริงวิธีโค้งลักษณะแบบทดสอบ
TRUXS	แทน	คะแนน (จริง) แปลงของแบบทดสอบฉบับ X จากการ เทียบคะแนนจริงวิธีปรับค่าพารามิเตอร์พร้อมกัน

XEQ	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนอีคิวเปอร์เซ็นไทล์
XEE	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนเชิงเส้นตรง
XVEE	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนเชิงเส้นตรงของกลุ่มสอบทานผล
XVEQ	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนอีคิวเปอร์เซ็นไทล์ของกลุ่มสอบทานผล
XVETMS	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนจริงวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของกลุ่มสอบทานผล
XVETTRF	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนจริงวิธีโค้งลักษณะแบบทดสอบของกลุ่มสอบทานผล
XVETS	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนจริงวิธีปรับค่าพารามิเตอร์พร้อมกันของกลุ่มสอบทานผล
XVEOMS	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนสังเกตวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของกลุ่มสอบทานผล
XVEOTRF	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบของกลุ่มสอบทานผล
XVEOS	แทน	คะแนนแปลงของแบบทดสอบฉบับ X จากการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกันของกลุ่มสอบทานผล

ตอนที่ 1 ผลการวิเคราะห์แบบทดสอบที่ใช้ในการเทียบคะแนน

การวิเคราะห์แบบทดสอบที่ใช้ในการเทียบคะแนน จำแนกเป็น การวิเคราะห์คุณภาพรายข้อของแบบทดสอบเทียบคะแนน การตรวจสอบความเป็นเอกมิติของแบบทดสอบเทียบคะแนน การวิเคราะห์ค่าสถิติพื้นฐานและทดสอบความแตกต่างของค่าเฉลี่ยและการแจกแจงคะแนนของแบบทดสอบเทียบคะแนน การวิเคราะห์คุณภาพของแบบทดสอบเทียบคะแนนทั้งฉบับ และการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบรวม มีรายละเอียดดังนี้

1.1 ผลการวิเคราะห์คุณภาพรายข้อของแบบทดสอบเทียบคะแนน

การวิเคราะห์คุณภาพรายข้อของแบบทดสอบเทียบคะแนนทั้งสองฉบับ คือฉบับ X และฉบับ Y ซึ่งแต่ละฉบับมีข้อสอบ จำนวน 30 ข้อ แต่ละข้อตรวจให้คะแนน 6 ค่า คือ 0 1 2 3 4 และ 5 คะแนน คุณภาพของข้อสอบแต่ละข้อวิเคราะห์ตามทฤษฎี การตอบข้อสอบด้วยโมเดลเกรด เรสพอนส์ ในแต่ละฉบับและวิเคราะห์ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ดังผลการวิเคราะห์ต่อไปนี้

1.1.1 ผลการวิเคราะห์พารามิเตอร์ข้อสอบ

จากข้อมูลของผู้สอบจำนวน 1,155 คน และ 1,157 คน ในแบบทดสอบฉบับ X และ Y ตามลำดับ เมื่อวิเคราะห์ตามทฤษฎีการตอบข้อสอบ โมเดลเกรด เรสพอนส์ ด้วยโปรแกรม MULTILOG Version 7.03 for Windows ได้พารามิเตอร์ความยากและอำนาจจำแนกดังผลปรากฏตามตาราง 7 และผลการวิเคราะห์คุณภาพรายข้อตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ปรากฏผลตามตาราง 8

ตาราง 7 ค่าสถิติของพารามิเตอร์อำนาจจำแนก และความยาก ในแบบทดสอบเทียบคะแนน เมื่อวิเคราะห์ตามทฤษฎีการตอบข้อสอบ โมเดลเกรด เรสพอนส์

ค่าสถิติ	ค่าต่ำสุด – ค่าสูงสุด		ค่าเฉลี่ย		ส่วนเบี่ยงเบนมาตรฐาน	
	ฉบับ X	ฉบับ Y	ฉบับ X	ฉบับ Y	ฉบับ X	ฉบับ Y
A	1.055 ถึง 2.414	1.197 ถึง 2.404	1.605	1.673	0.350	0.323
B ₁	-5.817 ถึง -2.050	-4.872 ถึง -1.499	-3.162	-3.045	0.824	0.804
B ₂	-2.994 ถึง -0.963	-2.484 ถึง -0.921	-1.727	-1.625	0.480	0.380
B ₃	-1.883 ถึง -0.031	-1.118 ถึง 0.019	-0.727	-0.681	0.389	0.280
B ₄	-0.743 ถึง 0.792	-0.714 ถึง 0.895	0.182	0.141	0.402	0.398
B ₅	-0.104 ถึง 2.444	-0.082 ถึง 2.714	1.225	1.153	0.597	0.659
B ₆	-5.817 ถึง 2.444	-4.872 ถึง 2.714	-0.841	-0.811	1.620	1.545

ผลการวิเคราะห์ตามตาราง 7 พบว่า ค่าอำนาจจำแนกของแบบทดสอบฉบับ X มีค่าอยู่ระหว่าง 1.055 ถึง 2.414 ฉบับ Y มีค่าอยู่ระหว่าง 1.197 ถึง 2.404 สำหรับค่าความยากในแต่ละลำดับขั้น คือ ลำดับขั้นคะแนนที่หนึ่ง (B₁) ของแบบทดสอบฉบับ X อยู่ระหว่าง -5.817 ถึง -2.050 ฉบับ Y อยู่ระหว่าง -4.872 ถึง -1.499 ลำดับขั้นคะแนนที่สอง (B₂) ของแบบทดสอบฉบับ X อยู่ระหว่าง -2.994 ถึง -0.963 ฉบับ Y อยู่ระหว่าง -2.484 ถึง -0.921 ลำดับขั้นคะแนนที่สาม (B₃) ของแบบทดสอบฉบับ X อยู่ระหว่าง -1.883 ถึง -0.031 ฉบับ Y อยู่ระหว่าง -1.118 ถึง +0.019 ลำดับขั้นคะแนนที่สี่ (B₄) ของแบบทดสอบฉบับ X อยู่ระหว่าง -0.743 ถึง

+0.792 ฉบับ Y อยู่ระหว่าง -0.714 ถึง +0.895 และลำดับชั้นคะแนนที่ห้า (B_5) ของแบบทดสอบฉบับ X มีค่าอยู่ระหว่าง -0.104 ถึง +2.444 ฉบับ Y มีค่าอยู่ระหว่าง -0.082 ถึง +2.714

สรุปได้ว่าผลวิเคราะห์คุณภาพรายข้อของแบบทดสอบเทียบคะแนนทั้งสองฉบับ เมื่อพิจารณาจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์อำนาจจำแนกและความยาก มีความใกล้เคียงกันและมีคุณภาพอยู่ในระดับดีคือค่าอำนาจจำแนกของแบบทดสอบทั้งสองฉบับไม่มีค่าในข้อใดติดลบ และมีค่าเป็นบวกตั้งแต่ 1.00 ขึ้นไปทุกข้อ ส่วนค่าความยากนั้นเรียงตามลำดับชั้นคะแนนจากง่ายไปหายากเป็นไปตามรูปแบบของโมเดลเกรด เรสพอนส์

ตาราง 8 ค่าสถิติของอำนาจจำแนกและความยากในแบบทดสอบเทียบคะแนน เมื่อวิเคราะห์ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม

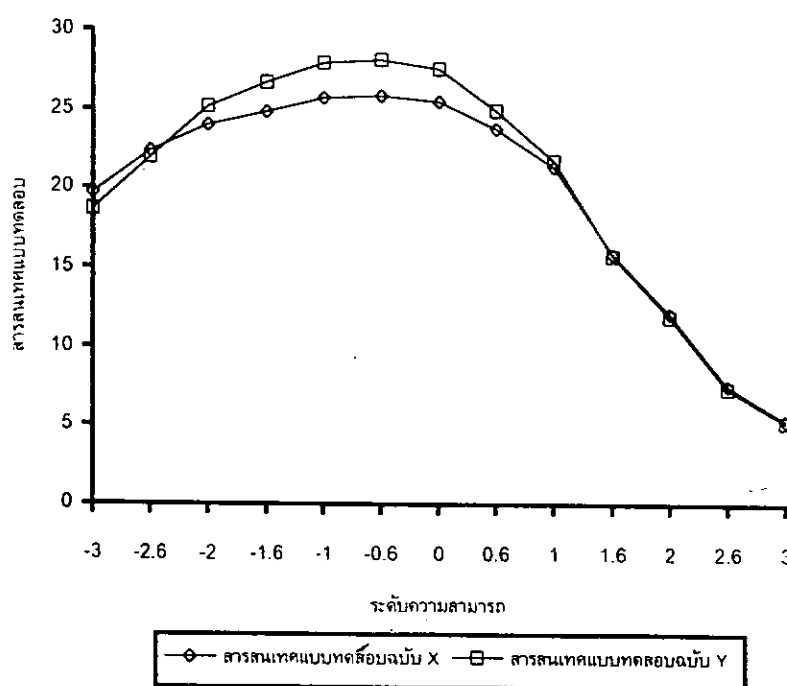
ค่าสถิติ	ค่าต่ำสุด - ค่าสูงสุด		ค่าเฉลี่ย		ส่วนเบี่ยงเบนมาตรฐาน	
	ฉบับ X	ฉบับ Y	ฉบับ X	ฉบับ Y	ฉบับ X	ฉบับ Y
อำนาจจำแนก	0.460 ถึง 0.676	0.479 ถึง 0.668	0.560	0.568	0.054	0.047
ความยาก	0.581 ถึง 0.839	0.570 ถึง 0.844	0.689	0.702	0.065	0.064

ผลการวิเคราะห์ตามตาราง 8 พบว่า ค่าความยากของแบบทดสอบฉบับ X มีค่าอยู่ระหว่าง 0.581 ถึง 0.839 ฉบับ Y มีค่าอยู่ระหว่าง 0.570 ถึง 0.844 ซึ่งค่าความยากของข้อสอบในฉบับ X และฉบับ Y ฉบับละ 2 ข้อ มีค่าเกิน 0.800 แต่ค่าอำนาจจำแนกของแบบทดสอบทั้งสองฉบับถือว่ามีความอยู่ในระดับดี คือ ฉบับ X มีค่าอยู่ระหว่าง 0.460 ถึง 0.676 ฉบับ Y มีค่าอยู่ระหว่าง 0.479 ถึง 0.668

1.1.2 ผลการวิเคราะห์ฟังก์ชันสารสนเทศของข้อสอบและแบบทดสอบ

ผลการวิเคราะห์ฟังก์ชันสารสนเทศของข้อสอบ ซึ่งเป็นดัชนีที่ใช้บ่งชี้คุณภาพของข้อสอบ ฟังก์ชันสารสนเทศของข้อสอบเป็นดัชนีที่ประกอบด้วยค่าพารามิเตอร์ความยากและค่าอำนาจจำแนก ผลปรากฏตามตาราง 53 (ภาคผนวก) พบว่า ค่าสารสนเทศของข้อสอบที่มีค่าอำนาจจำแนกสูง จะให้ค่าสารสนเทศของข้อสอบมีค่าสูงด้วย เช่น ในฉบับ X ข้อ 7 ให้ค่าอำนาจจำแนกสูงสุดขณะเดียวกันให้ค่าสารสนเทศของข้อสอบ ณ $\theta = 0$ เท่ากับ 1.807 สำหรับฉบับ Y ข้อ 14 ให้ค่าอำนาจจำแนกสูงสุด ขณะเดียวกันให้ค่าสารสนเทศของข้อสอบ ณ $\theta = -0.6$ เท่ากับ 1.839 และเมื่อพิจารณาข้อสอบในฉบับ X และ Y ที่วัดในจุดประสงค์เดียวกัน พบว่า สารสนเทศของข้อสอบทั้งสองฉบับ ณ ระดับ θ เดียวกัน มีค่าใกล้เคียงกัน

เพื่อให้เห็นสารสนเทศของแบบทดสอบเทียบคะแนน ซึ่งเป็นภาพรวมที่ใช้ในการพิจารณาความใกล้เคียงกันของแบบทดสอบ ผู้วิจัยจึงได้นำสารสนเทศของแบบทดสอบเทียบคะแนนฉบับ X และ Y ในแต่ละระดับความสามารถมาเขียนเป็นกราฟ ดังภาพประกอบ 2



ภาพประกอบ 2 สารสนเทศแบบทดสอบเทียบคะแนน

ผลการวิเคราะห์ตามภาพประกอบ 2 พบว่า ค่าสารสนเทศของแบบทดสอบฉบับ X มีค่าสูงสุดที่ $\theta = -0.6$ เท่ากับ 25.871 ฉบับ Y มีค่าสูงสุด ณ $\theta = -0.6$ เท่ากับ 28.127 แสดงว่าแบบทดสอบทั้งสองฉบับเหมาะที่จะนำมาใช้เพื่อประมาณค่าความสามารถของผู้สอบในช่วง θ มีค่า ระหว่าง -1 ถึง 0 ซึ่งจะให้ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าต่ำคือ ฉบับ X ประมาณ 0.20 และ ฉบับ Y ประมาณ 0.19 และเมื่อพิจารณาประสิทธิภาพสัมพัทธ์ ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบทดสอบฉบับ Y ต่อฉบับ X ณ ตำแหน่ง θ เดียวกันมีค่าอยู่ในช่วง 0.949 ถึง 1.087 และเมื่อพิจารณาประสิทธิภาพสัมพัทธ์เฉลี่ย (RAI) ซึ่งเป็นอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศเฉลี่ยของแบบทดสอบฉบับ Y ต่อฉบับ X ณ ทุกตำแหน่ง θ มีค่าเท่ากับ 1.036 แสดงว่า แบบทดสอบทั้งสองฉบับมีความแม่นยำในการประมาณค่า θ ได้ใกล้เคียงกัน

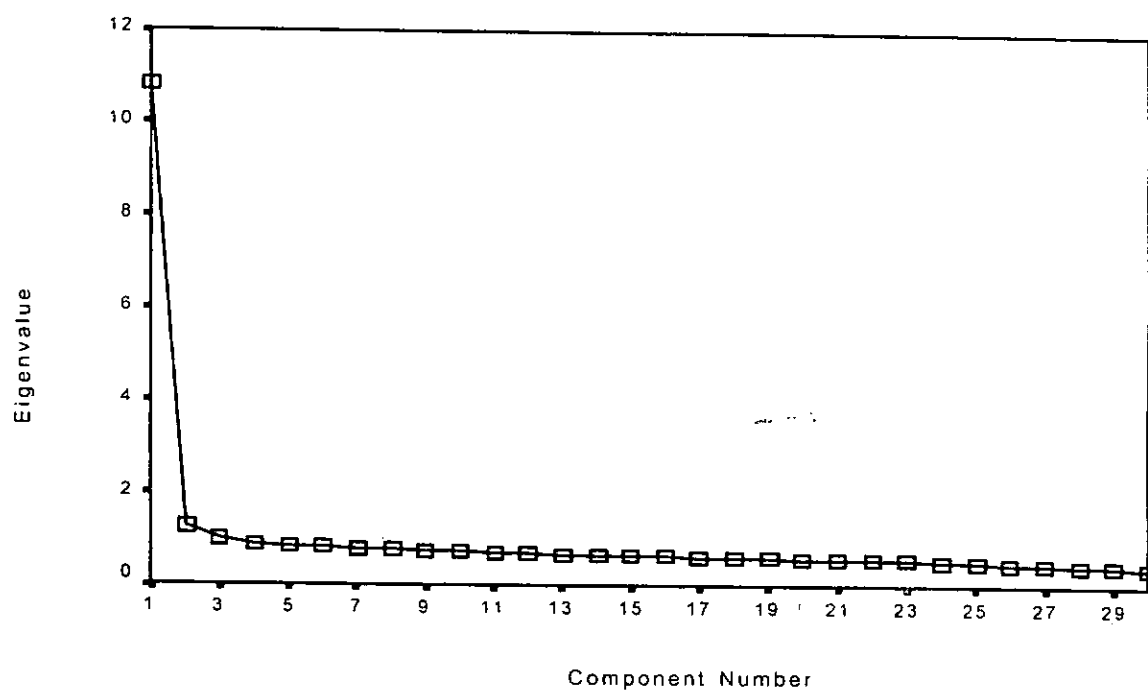
1.2 ผลการตรวจสอบความเป็นเอกมิติ (Unidimensional) ของแบบทดสอบ เทียบคะแนน

สำหรับแบบทดสอบเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่า ซึ่งผู้วิจัยสร้างขึ้น เพื่อให้เป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ ผู้วิจัยได้ตรวจสอบจำนวนองค์ประกอบในแบบทดสอบทั้งสองฉบับด้วยการวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory factor analysis) แบบ principle component โดยใช้เกณฑ์ค่า eigen ตั้งแต่ 1 ขึ้นไป นับเป็น 1 องค์ประกอบ วิเคราะห์ค่า KMO (Kaiser – Meyer -Olkin Measure of Sampling Adequacy) และสถิติทดสอบ Bartlett (Bartlett's Test of Sphericity) รวมทั้งผลที่ได้จาก Scree plot และการวิเคราะห์เพื่อยืนยันองค์ประกอบ (Confirmatory factor analysis) ดังผลปรากฏตามตาราง 9 – 10 และภาพประกอบ 3 – 4 ดังนี้

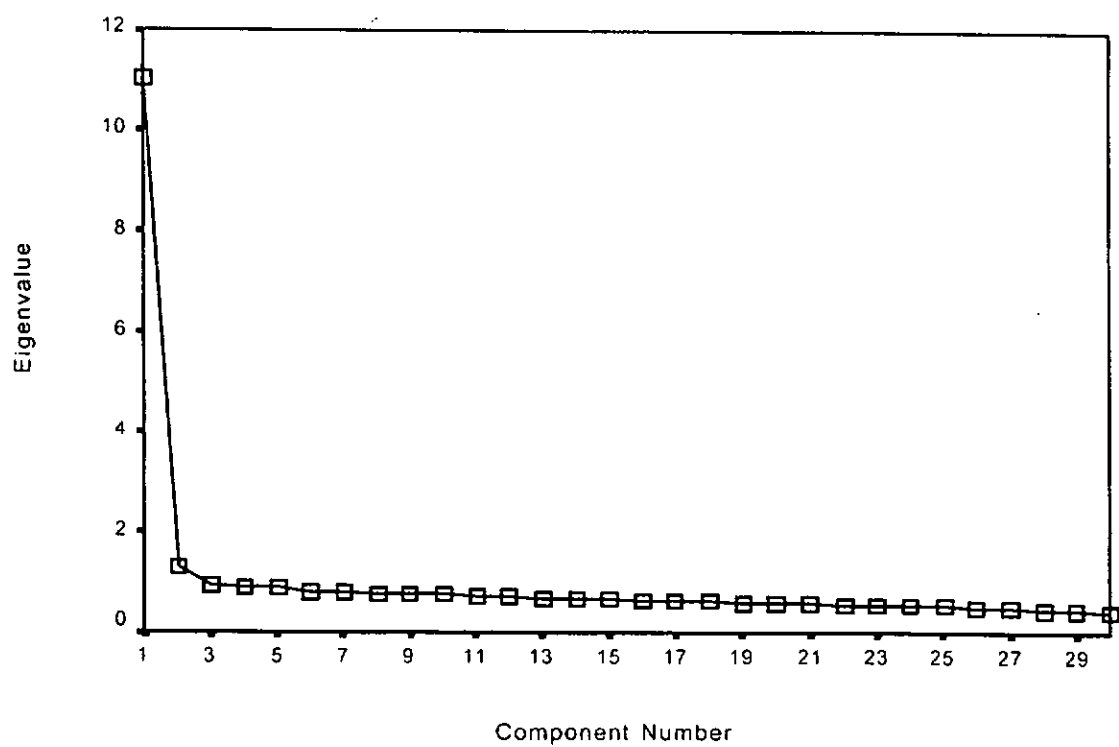
ตาราง 9 ค่า eigen ร้อยละของความแปรปรวน ความแปรปรวนสะสม ค่า KMO และค่า Bartlett จำแนกตามแบบทดสอบเทียบคะแนน

แบบทดสอบ	ค่า eigen	ร้อยละของความแปรปรวน	ความแปรปรวนสะสม	ค่า KMO	ค่า Bartlett
ฉบับ X	10.803	36.009	36.009	0.976	0.000
	1.266	4.220	40.229		
	1.002	3.340	43.569		
ฉบับ Y	11.038	36.792	36.792	0.978	0.000
	1.282	4.272	41.064		

ผลการวิเคราะห์ตามตาราง 9 พบว่า ค่า KMO มีค่าเท่ากับ 0.976 และ 0.978 ในแบบทดสอบฉบับ X และ Y ตามลำดับ ซึ่งมากกว่า 0.5 แสดงว่าตัวแปรต่าง ๆ มีความสัมพันธ์กัน ข้อมูลในแบบทดสอบฉบับ X และ Y จึงมีความเหมาะสมที่จะวิเคราะห์องค์ประกอบ โดยที่แบบทดสอบฉบับ X มีจำนวนองค์ประกอบที่มีค่า eigen ตั้งแต่ 1 ขึ้นไป 3 องค์ประกอบมีค่าเท่ากับ 10.803 1.266 และ 1.002 ตามลำดับ สำหรับฉบับ Y มีจำนวนองค์ประกอบที่มีค่า eigen ตั้งแต่ 1 ขึ้นไป 2 องค์ประกอบ มีค่าเท่ากับ 11.038 และ 1.282 ตามลำดับ



ภาพประกอบ 3 จำนวนองค์ประกอบในแบบทดสอบฉบับ X



ภาพประกอบ 4 จำนวนองค์ประกอบในแบบทดสอบฉบับ Y

ผลการวิเคราะห์ตามภาพประกอบ 3 – 4 พบว่า ค่า eigen จากองค์ประกอบที่ 1 ในแบบทดสอบทั้งสองฉบับลดลงอย่างรวดเร็ว และองค์ประกอบที่ 1 ของแบบทดสอบทั้งสองฉบับยังอธิบายความแปรปรวนได้มากกว่าร้อยละ 20 (Raju and others. 1993 : 304; citing Reckase.1979) และหากพิจารณาอัตราส่วนระหว่างองค์ประกอบที่ 1 และองค์ประกอบที่ 2 ในแบบทดสอบฉบับ X และ Y มีค่าเท่ากับ 8.533 และ 8.610 ตามลำดับ ซึ่งสูงกว่าเกณฑ์ที่วอร์นุช แหยมแสง (2537) ตั้งไว้ (5.739)

ดังนั้น เพื่อเป็นการยืนยันว่าแบบทดสอบแต่ละฉบับมี 1 องค์ประกอบจริง จึงได้ทำการวิเคราะห์องค์ประกอบเชิงยืนยัน โดยแปลงตัวแปรคะแนนให้เป็นตัวแปรสุ่มแบบปกติ ปรากฏผลการวิเคราะห์ตามตาราง 10

ตาราง 10 ค่าดัชนีทดสอบประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันของแบบทดสอบเทียบคะแนน

แบบทดสอบ	χ^2	p	df	NNFI	RMSEA	GFI	AGFI	CFI
ฉบับ X	520.232	0.000	405	0.917	0.013	0.973	0.969	0.923
ฉบับ Y	435.687	0.141	405	0.977	0.004	0.977	0.973	0.978

ผลการวิเคราะห์ตามตาราง 10 พบว่า แบบทดสอบฉบับ X และ Y มีค่าไคสแควร์เท่ากับ 520.232 และ 435.687 ตามลำดับ ค่า p เท่ากับ 0.000 และ 0.141 ตามลำดับ โดยมีชั้นแห่งความเป็นอิสระเท่ากับ 405 และ 405 ตามลำดับ แสดงว่าโมเดลไม่สอดคล้องกับข้อมูลในแบบทดสอบฉบับ X แต่เมื่อวิเคราะห์โดยหารค่าไคสแควร์ด้วยชั้นแห่งความเป็นอิสระได้ค่าเท่ากับ 1.285 และค่าดัชนีอื่น ๆ ได้แก่ NNFI RMSEA GFI AGFI และ CFI มีค่าเท่ากับ 0.917 0.013 0.973 0.969 และ 0.923 ตามลำดับ ซึ่งค่าดังกล่าวอยู่ในเกณฑ์ที่ยอมรับได้ จึงสรุปได้ว่าแบบทดสอบทั้งสองฉบับ มี 1 องค์ประกอบจริง และแสดงให้เห็นว่า แบบทดสอบเทียบคะแนนแต่ละฉบับมีความเที่ยงตรงเชิงโครงสร้าง

1.3 ผลการวิเคราะห์ค่าสถิติพื้นฐานและทดสอบความแตกต่างของค่าเฉลี่ยและการแจกแจงคะแนนของแบบทดสอบเทียบคะแนน

1.3.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนน

การวิเคราะห์ข้อมูลเบื้องต้นด้วยค่าสถิติพื้นฐานเพื่ออธิบายคุณลักษณะของคะแนนดิบที่ได้จากการวัดด้วยแบบทดสอบเทียบคะแนน ได้แก่ ค่าต่ำสุด ค่าสูงสุด ค่าเฉลี่ย ค่ามัธยฐาน ส่วนเบี่ยงเบนมาตรฐาน ความแปรปรวน ค่าความเบ้ ค่าความโด่ง จากกลุ่มตัวอย่าง

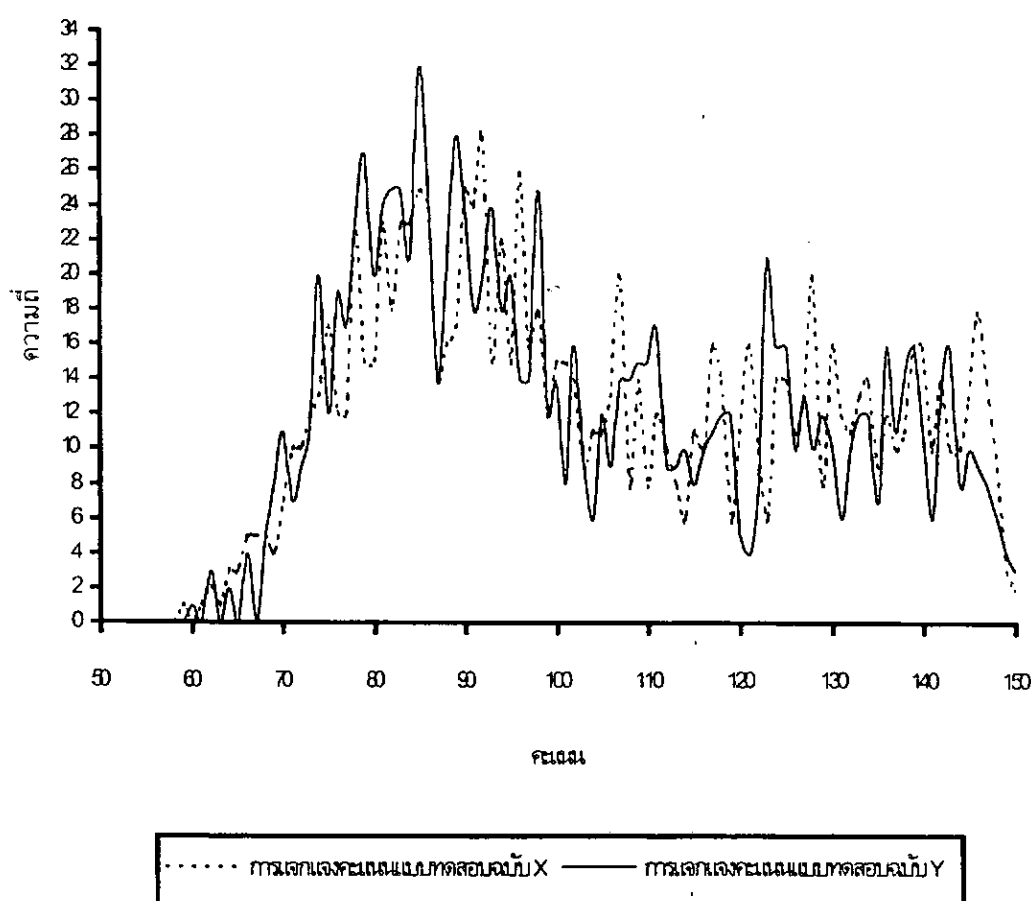
จำนวน 1155 คน และ 1157 คน ในแบบทดสอบฉบับ X และ Y ตามลำดับ ดังผลปรากฏในตาราง 11 และภาพประกอบ 5

ตาราง 11 ค่าสถิติพื้นฐานของคะแนนจากแบบทดสอบเทียบคะแนน

ค่าสถิติ	แบบทดสอบ	
	ฉบับ X	ฉบับ Y
ค่าต่ำสุด	60	59
ค่าสูงสุด	150	150
ค่าเฉลี่ย	103.335	105.224
ค่ามัธยฐาน	98.000	101.000
ส่วนเบี่ยงเบนมาตรฐาน	22.770	23.289
ความแปรปรวน	518.488	542.392
ความเบ้	0.373	0.252
ความคลาดเคลื่อนมาตรฐานของค่าความเบ้	0.072	0.072
ความโด่ง	-1.077	-1.142
ความคลาดเคลื่อนมาตรฐานของความโด่ง	0.144	0.144
OPM	0.689	0.701

ผลการวิเคราะห์ตามตาราง 11 พบว่า แบบทดสอบฉบับ X และ Y มีคะแนนเต็ม 150 คะแนน คะแนนรวมต่ำสุดคือ 60 และ 59 คะแนนตามลำดับ แสดงว่าต่ำสุด ผู้สอบสามารถตอบข้อสอบได้ถูกต้อง อย่างน้อย 2 ใน 5 คะแนน คะแนนรวมสูงสุดทั้งสองฉบับมีค่าเท่ากัน คือ 150 คะแนน สำหรับคะแนนเฉลี่ยของแบบทดสอบฉบับ X มีค่าต่ำกว่าฉบับ Y คือ ฉบับ X มีค่าเฉลี่ย 103.335 คะแนน ฉบับ Y มีค่าเฉลี่ย 105.224 คะแนน แสดงว่า แบบทดสอบฉบับ Y ค่อนข้างง่ายกว่า แบบทดสอบคะแนนฉบับ X โดยที่ร้อยละ 50 ของผู้สอบแบบทดสอบฉบับ X ได้คะแนนต่ำกว่า 98 คะแนน ฉบับ Y ได้คะแนนต่ำกว่า 101 คะแนน โดยพิสัยของคะแนนฉบับ X และฉบับ Y มีค่าใกล้เคียงกัน ($PR_{25} - PR_{75}$ อยู่ในช่วง 84 - 123 และ 86 - 126 คะแนน) คะแนนของแบบทดสอบฉบับ X และ Y มีค่ามัธยฐานแตกต่างกันเล็กน้อยคือ 98.000 และ 101.000 คะแนน

เมื่อพิจารณารูปแบบการแจกแจงคะแนนจากค่าความเบ้และความโด่ง พบว่า คะแนนจากแบบทดสอบทั้งสองฉบับมีลักษณะการแจกแจงที่คล้ายกันคือ เบ้ขวา โดยคะแนนจากแบบทดสอบฉบับ X มีการเบ้มากกว่าฉบับ Y ($S_k = 0.373, 0.252$ ตามลำดับ) ความโด่งของแบบทดสอบทั้งฉบับ X และ Y มีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ โดยคะแนนจากแบบทดสอบฉบับ Y มีความแบนมากกว่าฉบับ X ($K_u = -1.077, -1.142$) เมื่อพิจารณาความคลาดเคลื่อนของความเบ้และความโด่งของคะแนนจากแบบทดสอบทั้ง 2 ฉบับมีค่าเท่ากับ 0.072 และ 0.144 ตามลำดับ ซึ่งมีค่าไม่เกิน 2.00 หรือไม่น้อยกว่า -2.00 ซึ่งถือว่ามีลักษณะการแจกแจงเป็นโค้งปกติ (อวยพร เรื่องตระกูล, 2544, : 116; อ้างอิงจาก SPSS Base 8.0, 1998) ดังภาพประกอบ 5



ภาพประกอบ 5 การแจกแจงคะแนนของแบบทดสอบเทียบคะแนน

1.3.2 ผลการทดสอบความแตกต่างของค่าเฉลี่ยของคะแนน

เนื่องจากค่าเฉลี่ยและมัธยฐานของคะแนนจากแบบทดสอบเทียบคะแนนทั้งสองฉบับมีความแตกต่างกันไม่มากนัก และพิสัยของคะแนนฉบับ X และ Y มีค่าใกล้เคียงกัน จึงทำการทดสอบความแตกต่างของค่าเฉลี่ยในแบบทดสอบฉบับ X และ Y ด้วยสถิติทดสอบ t - independent ปรากฏผลตามตาราง 12

ตาราง 12 ผลการเปรียบเทียบค่าเฉลี่ยของคะแนนจากแบบทดสอบเทียบคะแนน

แบบทดสอบ	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	ผลต่างของค่าเฉลี่ย	สถิติทดสอบ t	df	p
ฉบับ X	103.335	22.770	-1.889	-1.972	2310	0.049
ฉบับ Y	105.224	23.289				

$p > 0.01$

ผลการวิเคราะห์ตามตาราง 12 พบว่า ค่าเฉลี่ยของคะแนนจากแบบทดสอบฉบับ X และฉบับ Y ไม่แตกต่างกัน

1.3.3 ผลการทดสอบการแจกแจงคะแนน

จากภาพประกอบ 5 แบบทดสอบเทียบคะแนนทั้งสองฉบับมีรูปร่างการแจกแจงคะแนนที่ใกล้เคียงกันและมีความคลาดเคลื่อนของความเบ้และความโด่งเท่ากัน จึงทำการทดสอบการแจกแจงคะแนนของแบบทดสอบทั้งสองฉบับ โดยใช้สถิติทดสอบนัยพารามेटริกซ์ Mann – Whitney U ดังผลปรากฏตามตาราง 13

ตาราง 13 ผลการวิเคราะห์ความแตกต่างของการแจกแจงคะแนนดิบจากแบบทดสอบเทียบคะแนน

แบบทดสอบ	ผลรวมของอันดับ	สถิติทดสอบ Z	p
ฉบับ X	1302772.5	-2.055	0.040
ฉบับ Y	1371055.5		

$p > 0.01$

ผลการวิเคราะห์ตามตาราง 13 พบว่า การแจกแจงคะแนนของแบบทดสอบเทียบคะแนนทั้งสองฉบับไม่แตกต่างกัน

1.4 ผลการวิเคราะห์คุณภาพของแบบทดสอบเทียบคะแนน

การวิเคราะห์คุณภาพของแบบทดสอบเทียบคะแนนทั้งฉบับ ได้แก่วามเชื่อมั่นและความเที่ยงตรงของแบบทดสอบ โดยแสดงหลักฐานความเที่ยงตรงเชิงเนื้อหาไว้ในบทที่ 3 และหลักฐานความเที่ยงตรงเชิงโครงสร้างที่ได้จากการยืนยันองค์ประกอบในหัวข้อที่ 1.2 ของบทที่ 4 แล้วนั้น สำหรับความเชื่อมั่นของแบบทดสอบที่วิเคราะห์ด้วยการแบ่งเครื่องมือออกเป็นสองส่วนย่อยตามเนื้อหา คือ จำนวนเต็มและเศษส่วนและทศนิยม โดยพิจารณาว่าหากแต่ละส่วน มีระดับความคู่ขนานแบบคะแนนจริงสมมูล คำนวณด้วยสัมประสิทธิ์แอลฟา ถ้าแต่ละส่วนมีระดับความคู่ขนานแบบคะแนนจริงสัมพันธ์คำนวณตามสูตรราชู และคำนวณค่าความเชื่อมั่นด้วยการแบ่งเครื่องมือเป็นส่วนย่อยหลาย ๆ ส่วนตามจำนวนข้อสอบในแต่ละฉบับ คำนวณด้วยสัมประสิทธิ์แอลฟา เมื่อพิจารณาว่าแต่ละข้อมีระดับความคู่ขนานแบบคะแนนจริงสมมูลและคำนวณด้วยสัมประสิทธิ์เบต้าเคเมื่อพิจารณาว่าแต่ละข้อมีระดับความคู่ขนานแบบคะแนนจริงสัมพันธ์ ปรากฏผลตามตาราง 14

ตาราง 14 ผลการวิเคราะห์ค่าเชื่อมั่นของแบบทดสอบเทียบคะแนน

ความเชื่อมั่น	แบบทดสอบ	
	X	Y
สัมประสิทธิ์แอลฟาแบ่งหลายส่วน	0.937	0.940
สัมประสิทธิ์เบต้าเค	0.937	0.940
สัมประสิทธิ์แอลฟาแบ่งสองส่วน	0.908	0.915
สัมประสิทธิ์ของราชู	0.912	0.919

ผลการวิเคราะห์ตามตาราง 14 พบว่า ค่าความเชื่อมั่นของแบบทดสอบ เมื่อแบ่งส่วนย่อยตามเนื้อหา สัมประสิทธิ์ของราชูให้ค่าที่สูงกว่าสัมประสิทธิ์แอลฟาในแบบทดสอบทั้งสองฉบับไม่มากนัก แต่ค่าความเชื่อมั่นของสัมประสิทธิ์แอลฟาเมื่อแปลงส่วนย่อยตามจำนวนข้อให้ค่าสูงกว่าความเชื่อมั่น เมื่อแบ่งเครื่องมือวัดออกเป็นสองส่วนในแบบทดสอบเทียบคะแนนทั้งสองฉบับ และค่าความเชื่อมั่นยังใกล้เคียงกับค่าที่วิเคราะห์ด้วยสัมประสิทธิ์เบต้าเค จะเห็นได้ว่าค่าความเชื่อมั่นที่ได้จากสัมประสิทธิ์ความเชื่อมั่นแต่ละสูตรยังคงให้ค่าที่สูง (มากกว่า 0.90) และค่าความเชื่อมั่นของแบบทดสอบ ฉบับ Y สูงกว่า ฉบับ X เล็กน้อย ดังนั้นการวิเคราะห์ผลที่เกี่ยวข้องกับค่าความเชื่อมั่นจะใช้ค่าสัมประสิทธิ์แอลฟาแบบแบ่งตามจำนวนข้อเป็นหลัก

1.5 ผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบร่วม

การเทียบคะแนนในครั้งนี้ใช้แบบแผนการรวบรวมข้อมูลด้วยแบบทดสอบร่วม ภายในกับกลุ่มสุ่ม เพื่อให้แบบทดสอบร่วมเป็นตัวปรับความแตกต่างระหว่างกลุ่มตัวอย่างเทียบคะแนน คือ กลุ่ม α เป็นกลุ่มตัวอย่างที่ทำแบบทดสอบฉบับ X และกลุ่ม β เป็นกลุ่มตัวอย่างที่ทำแบบทดสอบฉบับ Y แบบทดสอบเทียบคะแนนฉบับ X และ Y มีข้อสอบฉบับละ 30 ข้อ แต่ละฉบับได้รวมแบบทดสอบร่วมไว้ฉบับละ 10 ข้อ เมื่อนำคะแนนของแบบทดสอบร่วมที่ทำโดยกลุ่ม α β และกลุ่มรวม $t(\alpha + \beta)$ มาวิเคราะห์ค่าสถิติพื้นฐาน และความสัมพันธ์กับแบบทดสอบเทียบคะแนน ปรากฏผลการวิเคราะห์ตามตาราง 15 และ 16

ตาราง 15 ค่าสถิติพื้นฐานของคะแนนจากแบบทดสอบร่วม

ค่าสถิติ	กลุ่มตัวอย่าง		
	α	β	t
ค่าต่ำสุด	16	15	15
ค่าสูงสุด	50	50	50
ค่าเฉลี่ย	34.386	34.279	34.333
ค่ามัธยฐาน	33.000	33.000	33.000
ความแปรปรวน	53.735	58.343	56.020
ส่วนเบี่ยงเบนมาตรฐาน	7.330	7.638	7.485
ความเชื่อมั่น	0.807	0.822	0.815

ผลการวิเคราะห์ตามตาราง 15 พบว่า ค่าเฉลี่ยของแบบทดสอบร่วมในฉบับ X และ Y แตกต่างกันเล็กน้อย โดยคะแนนเฉลี่ยของแบบทดสอบร่วมในฉบับ X มีค่าสูงกว่าแบบทดสอบร่วมในฉบับ Y คือ 34.386 และ 34.279 ตามลำดับ ซึ่งสูงกว่าครึ่งหนึ่งของคะแนนเต็ม เมื่อพิจารณาค่าเบี่ยงเบนมาตรฐาน พบว่า แตกต่างกันไม่มาก โดยค่าเบี่ยงเบนมาตรฐานของแบบทดสอบร่วมในฉบับ X กลับต่ำกว่าค่าเบี่ยงเบนมาตรฐานของแบบทดสอบร่วมในฉบับ Y คือ 7.330 และ 7.638 ตามลำดับ มีผลทำให้ค่าความเชื่อมั่นของแบบทดสอบร่วมในฉบับ X ต่ำกว่าความเชื่อมั่นของแบบทดสอบฉบับ Y เล็กน้อย คือ 0.807 และ 0.822 ตามลำดับ

เมื่อพิจารณาค่าเฉลี่ยของแบบทดสอบร่วมที่ได้จากกลุ่มตัวอย่างรวม t ให้ค่าสูงกว่าค่าเฉลี่ยของแบบทดสอบร่วมที่ได้จากกลุ่ม β แต่ต่ำกว่ากลุ่ม α สำหรับความแปรปรวนมีค่าสูงกว่าความแปรปรวนของแบบทดสอบร่วมที่ได้จากกลุ่ม α แต่ต่ำกว่ากลุ่ม β

ตาราง 16 ค่าสัมประสิทธิ์สหสัมพันธ์รายคู่ระหว่างคะแนนของแบบทดสอบร่วมกับคะแนนของแบบทดสอบเทียบคะแนนแต่ละฉบับ

แบบทดสอบเทียบคะแนน	แบบทดสอบรวม
ฉบับ X	0.929**
ฉบับ Y	0.937**

**p < 0.01

ผลการวิเคราะห์ตามตาราง 16 พบว่า คะแนนของแบบทดสอบรวมในฉบับ X และ ฉบับ Y มีค่าสัมประสิทธิ์สหสัมพันธ์ กับคะแนนของแบบทดสอบเทียบคะแนนในแต่ละฉบับ อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยมีค่าสูงมาก คือ เท่ากับ 0.929 และ 0.937 ตามลำดับ แสดงว่าคะแนนจากแบบทดสอบรวมในแต่ละฉบับกับคะแนนจากแบบทดสอบเทียบคะแนนมีความสอดคล้องกันสูงมาก

ตอนที่ 2 ผลการสร้างตารางเทียบคะแนน

การศึกษาการเทียบคะแนนครั้งนี้ ใช้แบบแผนแบบทดสอบรวมในกลุ่มผู้เทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบ และตามทฤษฎีการทดสอบแบบมาตรฐานเดิม โดยวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ จำแนกตามวิธีการปรับค่าพารามิเตอร์ ซึ่งมีทั้งวิเคราะห์แยกฉบับ และวิเคราะห์พร้อมกันรวม 3 วิธี แต่ละวิธีเทียบคะแนนโดยใช้ทั้งคะแนนจริงและคะแนนสังเกต สำหรับวิธีการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมใช้วิธีการเทียบคะแนนเชิงเส้นตรงและอีคิวเปอร์เซ็นต์ไทล์ จึงขอเสนอผลการวิเคราะห์ค่าสถิติต่าง ๆ ที่ใช้ในการเทียบคะแนนแต่ละวิธี รวมทั้งตารางเทียบคะแนนที่ได้ในแต่ละวิธีดังต่อไปนี้

2.1 ผลการสร้างตารางเทียบคะแนนตามวิธีในทฤษฎีการตอบข้อสอบ

การเทียบคะแนนตามทฤษฎีการตอบข้อสอบ เมื่อวิเคราะห์ค่าพารามิเตอร์ของแบบทดสอบฉบับ X และ Y แยกกัน ทำให้ต้องแปลงค่าพารามิเตอร์ความสามารถให้อยู่ในมาตรวัดเดียวกันก่อนดำเนินการเทียบคะแนน โดยสัมประสิทธิ์ของสมการแปลงคะแนนอาศัยสารสนเทศจากข้อสอบรวม วิธีการประมาณค่าสัมประสิทธิ์การเทียบคะแนนมี 2 วิธีคือ วิธีประมาณค่าที่ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของข้อสอบรวม (MS) และวิธีประมาณค่าโดยใช้โค้งลักษณะแบบทดสอบรวม (TRF) เมื่อแปลงพารามิเตอร์ความสามารถของผู้สอบในแบบทดสอบฉบับ Y ให้อยู่บนมาตรวัดเดียวกับผู้สอบในแบบทดสอบฉบับ X แล้วจึงดำเนินการเทียบคะแนนโดยใช้คะแนนจริงและคะแนนสังเกต สำหรับอีกวิธีการหนึ่งซึ่งวิเคราะห์ค่าพารามิเตอร์ของแบบทดสอบทั้งสองฉบับพร้อมกันโดยพารามิเตอร์ของข้อสอบรวม

ทั้งสองฉบับมีค่าเดียวกัน ทำให้พารามิเตอร์ความสามารถของผู้สอบจากแบบทดสอบทั้งสองฉบับอยู่ในมาตรวัดเดียวกัน ไม่มีความจำเป็นต้องแปลงค่าพารามิเตอร์ความสามารถของผู้สอบทั้งสองฉบับ (Kolen, 1995) จึงสามารถเทียบคะแนนจริงและคะแนนสังเกตได้ทันที ดังนั้น จึงขอเสนอผลการประมาณค่าสัมประสิทธิ์การเทียบคะแนน ผลการสร้างตารางเทียบคะแนนจริงและคะแนนสังเกตในแต่ละวิธีการเทียบคะแนน ดังต่อไปนี้

2.1.1 ผลการประมาณค่าสัมประสิทธิ์การเทียบคะแนน

2.1.1.1 ค่าสถิติที่ใช้ในการคำนวณค่าสัมประสิทธิ์การเทียบคะแนน

ในการสร้างสมการแปลงคะแนนความสามารถต้องอาศัยค่าสถิติจากพารามิเตอร์ของข้อสอบรวมทั้งวิเคราะห์จากแบบทดสอบฉบับ X และ Y ฉบับละ 10 ข้อ จากกลุ่มตัวอย่างเทียบคะแนนสองกลุ่ม คือ กลุ่ม α (ทำแบบทดสอบฉบับ X) และ กลุ่ม β (ทำแบบทดสอบฉบับ Y) ค่าสถิติที่ใช้ประมาณค่าสัมประสิทธิ์การเทียบคะแนนวิธี MS ปรากฏผลตามตาราง 17 สำหรับการประมาณค่าสัมประสิทธิ์การเทียบคะแนนโดยใช้โด่งลักษณะแบบทดสอบเป็นการใช้ค่าพารามิเตอร์อำนาจจำแนก (A) และความยาก ($B_1 - B_5$) ทั้งสองค่าพร้อมกัน โดยนำค่าพารามิเตอร์ข้อสอบรวมทั้งข้อจากแบบทดสอบเทียบคะแนนทั้ง 2 ฉบับ (รายละเอียดค่าพารามิเตอร์ในตาราง 52 (ภาคผนวก)) ไปคำนวณในโปรแกรม EQUATE Version 2.1

ตาราง 17 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกและความยากของแบบทดสอบร่วมจำแนกตามแบบทดสอบเทียบคะแนน

ค่าสถิติ	A	B ₁	B ₂	B ₃	B ₄	B ₅	B _i
แบบทดสอบฉบับ X							
ค่าเฉลี่ย	1.453	-3.666	-1.945	-0.838	0.202	1.451	-0.959
ส่วนเบี่ยงเบนมาตรฐาน	0.246	0.994	0.494	0.346	0.425	0.642	1.875
แบบทดสอบฉบับ Y							
ค่าเฉลี่ย	1.502	-3.602	-1.802	-0.695	0.231	1.450	-0.884
ส่วนเบี่ยงเบนมาตรฐาน	0.230	0.904	0.465	0.352	0.428	0.641	1.837

ผลการวิเคราะห์ตามตาราง 17 พบว่า ค่าเฉลี่ยของอำนาจจำแนกจากแบบทดสอบร่วมในฉบับ X และ Y มีค่าเท่ากับ 1.453 และ 1.502 ตามลำดับ และส่วนเบี่ยงเบนมาตรฐานของอำนาจจำแนกเท่ากับ 0.246 และ 0.230 ตามลำดับ ซึ่งมีค่าใกล้เคียงกัน โดยที่ค่าเฉลี่ยของอำนาจจำแนกของแบบทดสอบร่วมในฉบับ Y สูงกว่าในฉบับ X แต่ส่วนเบี่ยงเบนมาตรฐานของอำนาจจำแนกของแบบทดสอบร่วมในฉบับ Y มีค่าต่ำกว่าฉบับ X และเมื่อพิจารณาค่าเฉลี่ยของ

ความยากทุกลำดับชั้นของแบบทดสอบรวมในฉบับ X และ Y มีค่าเท่ากับ -0.959 และ -0.884 ตามลำดับ และส่วนเบี่ยงเบนมาตรฐานมีค่าเท่ากับ 1.875 และ 1.837 ตามลำดับ จะเห็นได้ว่าความยากของแบบทดสอบรวมในฉบับ Y สูงกว่าฉบับ X เล็กน้อย แต่ส่วนเบี่ยงเบนมาตรฐานมีค่าต่ำกว่าเล็กน้อย ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความยากที่คำนวณได้จะนำไปประมาณค่าสัมประสิทธิ์การเทียบคะแนน

2.1.1.2 สมการแปลงพารามิเตอร์ความสามารถ

จากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของแบบทดสอบรวมในแบบทดสอบฉบับ X และ Y นำมาคำนวณค่าสัมประสิทธิ์การเทียบคะแนนตามสูตร 90 และ 91 ในหน้า 128 เรียกว่าวิธี MS ปรากฏผลตามตาราง 18 สำหรับวิธีที่ 2 เรียกว่าวิธี TRF เป็นการคำนวณค่าสัมประสิทธิ์การเทียบคะแนนโดยใช้โค้งลักษณะแบบทดสอบรวมจากฉบับ X และ Y ปรากฏผลตามตาราง 18

ตาราง 18 ค่าสัมประสิทธิ์การเทียบคะแนนจำแนกตามวิธีการเทียบคะแนน

วิธีการ เทียบคะแนน	สัมประสิทธิ์		สมการแปลงพารามิเตอร์ความสามารถ
	A	K	
MS	1.021	-0.057	$thyb = 1.021 (thy) - 0.057$
TRF	1.026	-0.062	$thyc = 1.025 (thy) - 0.062$

ผลการวิเคราะห์ตามตาราง 18 เมื่อพิจารณาสัมประสิทธิ์การเทียบคะแนน A ซึ่งเป็นค่าความชัน และ K ซึ่งเป็นจุดตัด พบว่า ค่า A และ K ในทั้งสองวิธีประมาณค่าสัมประสิทธิ์การเทียบคะแนนมีค่าใกล้เคียงกัน แต่วิธี MS ประมาณค่า A และ K เข้าใกล้ 1 และ 0 มากกว่าวิธี TRF

2.1.2 ผลการสร้างตารางเทียบคะแนนจริง

2.1.2.1 ผลการสร้างตารางเทียบคะแนนจริงวิธีวิเคราะห์ค่า

พารามิเตอร์แยกฉบับ

การสร้างตารางเทียบคะแนนจริง สามารถดำเนินการโดยนำข้อมูลจากแบบทดสอบฉบับ Y มาประมาณค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถ แล้วกำหนดค่าคะแนนจริงตั้งแต่ 1 – 150 และหาระดับความสามารถที่สมนัยกับระดับคะแนนจริง โดยวิเคราะห์ฟังก์ชันผกผันของคะแนนจริง ว่าคะแนนจริงที่มีค่าเท่านี้เกิดจากความสามารถระดับใดของผู้สอบ ได้ค่าคะแนนจริงและความสามารถที่สมนัยกันในช่อง 1 – 2 ของตาราง 19

เมื่อได้ค่าพารามิเตอร์ความสามารถของผู้สอบในฉบับ Y แล้ว จึงคำนวณคะแนนจริงของผู้สอบในแบบทดสอบฉบับ X ที่เทียบเท่ากับฉบับ Y โดยนำพารามิเตอร์ความสามารถ (THY) ในฉบับ Y แปลงไปสู่มาตรฐานวัดของฉบับ X ด้วยสมการแปลงในตาราง 18 จะได้พารามิเตอร์ความสามารถของฉบับ Y ที่สมมูลกับพารามิเตอร์ความสามารถฉบับ X (THYC และ THYB ในช่อง 3 และ 5 ของตาราง 19) นำค่าพารามิเตอร์ความสามารถที่แปลงจากแต่ละสมการ ไปคำนวณค่าคะแนนจริงด้วยพารามิเตอร์ข้อสอบฉบับ X ตามสูตร 84 ในหน้า 125 ปรากฏผลในช่อง 4 (TRUYC) และ 6 (TRUYB) ของตาราง 19

ตาราง 19 คะแนนสมมูลของการเทียบคะแนนจริงจากแบบทดสอบฉบับ Y ไปยังฉบับ X
จำแนกตามวิธีการแปลงพารามิเตอร์ความสามารถ

STY	THY	THYC	TRUYC	THYB	TRUYB
1	-5.1013	-5.2916	1.4024	-5.2645	1.4464
2	-4.7230	-4.9037	2.1746	-4.8784	2.2382
3	-4.4681	-4.6424	2.9363	-4.6182	3.0197
4	-4.2667	-4.4359	3.7262	-4.4126	3.8271
5	-4.0961	-4.2610	4.5458	-4.2385	4.6621
6	-3.9465	-4.1077	5.3909	-4.0857	5.5228
7	-3.8123	-3.9701	6.2631	-3.9488	6.4087
8	-3.6905	-3.8452	7.1620	-3.8244	7.3227
9	-3.5790	-3.7309	8.0888	-3.7106	8.2651
10	-3.4762	-3.6255	9.0459	-3.6057	9.2379
11	-3.3807	-3.5276	10.0354	-3.5082	10.2440
12	-3.2917	-3.4364	11.0541	-3.4173	11.2800
13	-3.2081	-3.3506	12.1050	-3.3320	12.3453
14	-3.1294	-3.2700	13.1790	-3.2517	13.4349
15	-3.0549	-3.1936	14.2774	-3.1756	14.5478
16	-2.9841	-3.1210	15.3948	-3.1034	15.6764
17	-2.9167	-3.0519	16.5245	-3.0346	16.8173
18	-2.8523	-2.9859	17.6624	-2.9688	17.9663
19	-2.7907	-2.9227	18.8035	-2.9059	19.1149

ตาราง 19 (ต่อ)

STY	THY	THYC	TRUYC	THYB	TRUYB
20	-2.7315	-2.8620	19.9438	-2.8455	20.2607
21	-2.6745	-2.8036	21.0783	-2.7873	21.4010
22	-2.6194	-2.7471	22.2072	-2.7311	22.5319
23	-2.5661	-2.6925	23.3237	-2.6777	23.6509
24	-2.5144	-2.6395	24.4281	-2.6239	24.7566
25	-2.4641	-2.5879	25.5201	-2.5725	25.8488
26	-2.4150	-2.5376	26.5980	-2.5224	26.9260
27	-2.3670	-2.4883	27.6655	-2.4734	27.9901
28	-2.3199	-2.4401	28.7187	-2.4254	29.0416
29	-2.2738	-2.3928	29.7607	-2.3783	30.0817
30	-2.2284	-2.3463	30.7930	-2.3320	31.1121
31	-2.1838	-2.3005	31.8178	-2.2864	32.1349
32	-2.1399	-2.2555	32.8329	-2.2416	33.1481
33	-2.0967	-2.2112	33.8407	-2.1975	34.1543
34	-2.0541	-2.1676	34.8418	-2.1540	35.1560
35	-2.0122	-2.1246	35.8388	-2.1113	36.1492
36	-1.9710	-2.0824	36.8274	-2.0692	37.1389
37	-1.9305	-2.0408	37.8127	-2.0279	38.1206
38	-1.8907	-2.0000	38.7903	-1.9872	39.0993
39	-1.8516	-1.9600	39.7601	-1.9473	40.0706
40	-1.8132	-1.9206	40.7273	-1.9081	41.0368
41	-1.7754	-1.8818	41.6920	-1.8695	42.0004
42	-1.7384	-1.8439	42.6465	-1.8318	42.9539
43	-1.7020	-1.8066	43.5983	-1.7946	43.9071
44	-1.6662	-1.7699	44.5471	-1.7581	44.8548
45	-1.6311	-1.7339	45.4900	-1.7222	45.7991
46	-1.5965	-1.6984	46.4319	-1.6869	46.7395
47	-1.5624	-1.6635	47.3695	-1.6521	47.6783
48	-1.5289	-1.6291	48.3051	-1.6179	48.6121
49	-1.4957	-1.5951	49.2408	-1.5840	49.5486

ตาราง 19 (ต่อ)

STY	THY	THYC	TRUYC	THYB	TRUYB
50	-1.4631	-1.5617	50.1703	-1.5508	50.4759
51	-1.4308	-1.5286	51.1014	-1.5178	51.4073
52	-1.3988	-1.4957	52.0362	-1.4851	52.3394
53	-1.3672	-1.4634	52.9628	-1.4529	53.2658
54	-1.3359	-1.4313	53.8918	-1.4209	54.1945
55	-1.3049	-1.3995	54.8199	-1.3893	55.1191
56	-1.2742	-1.3680	55.7464	-1.3579	56.0449
57	-1.2437	-1.3367	56.6737	-1.3268	56.9683
58	-1.2134	-1.3057	57.5984	-1.2959	57.8920
59	-1.1834	-1.2749	58.5231	-1.2652	58.8155
60	-1.1536	-1.2444	59.4444	-1.2348	59.7355
61	-1.1240	-1.2140	60.3680	-1.2046	60.6546
62	-1.0945	-1.1838	61.2906	-1.1745	61.5757
63	-1.0653	-1.1538	62.2120	-1.1447	62.4924
64	-1.0362	-1.1240	63.1318	-1.1150	63.4105
65	-1.0073	-1.0944	64.0499	-1.0855	64.3267
66	-0.9786	-1.0650	64.9659	-1.0562	65.2409
67	-0.9500	-1.0356	65.8859	-1.0270	66.1558
68	-0.9216	-1.0065	66.8003	-0.9980	67.0680
69	-0.8933	-0.9775	67.7150	-0.9691	67.9805
70	-0.8651	-0.9486	68.6297	-0.9403	68.8929
71	-0.8370	-0.9198	69.5442	-0.9116	69.8050
72	-0.8090	-0.8911	70.4581	-0.8831	70.7132
73	-0.7811	-0.8625	71.3711	-0.8546	71.6237
74	-0.7533	-0.8340	72.2830	-0.8262	72.5329
75	-0.7256	-0.8056	73.1934	-0.7979	73.4405
76	-0.6979	-0.7772	74.1053	-0.7697	74.3463
77	-0.6703	-0.7489	75.0151	-0.7415	75.2532
78	-0.6427	-0.7206	75.9259	-0.7133	76.1610
79	-0.6151	-0.6923	76.8375	-0.6851	77.0695

ตาราง 19 (ต่อ)

STY	THY	THYC	TRUYC	THYB	TRUYB
80	-0.5875	-0.6640	77.7497	-0.6570	77.9754
81	-0.5600	-0.6358	78.6592	-0.6289	78.8818
82	-0.5324	-0.6075	79.5723	-0.6007	79.7917
83	-0.5049	-0.5793	80.4825	-0.5726	80.6988
84	-0.4773	-0.5510	81.3963	-0.5445	81.6062
85	-0.4497	-0.5227	82.3105	-0.5163	82.5172
86	-0.4220	-0.4943	83.2283	-0.4880	83.4320
87	-0.3943	-0.4659	84.1466	-0.4597	84.3472
88	-0.3666	-0.4375	85.0655	-0.4315	85.2597
89	-0.3388	-0.4090	85.9882	-0.4031	86.1794
90	-0.3109	-0.3804	86.9149	-0.3746	87.1029
91	-0.2830	-0.3518	87.8422	-0.3461	88.0271
92	-0.2549	-0.3230	88.7767	-0.3174	88.9585
93	-0.2268	-0.2942	89.7117	-0.2888	89.8871
94	-0.1986	-0.2653	90.6505	-0.2600	90.8227
95	-0.1703	-0.2363	91.5927	-0.2311	91.7617
96	-0.1419	-0.2072	92.5382	-0.2021	92.7039
97	-0.1134	-0.1780	93.4867	-0.1730	93.6490
98	-0.0847	-0.1485	94.4442	-0.1437	94.6000
99	-0.0559	-0.1190	95.4007	-0.1143	95.5530
100	-0.0270	-0.0894	96.3590	-0.0848	96.5077
101	0.0021	-0.0595	97.3249	-0.0551	97.4669
102	0.0315	-0.0294	98.2948	-0.0251	98.4332
103	0.0610	0.0008	99.2650	0.0050	99.3997
104	0.0907	0.0313	100.2414	0.0353	100.3691
105	0.1207	0.0620	101.2201	0.0660	101.3473
106	0.1510	0.0931	102.2071	0.0969	102.3274
107	0.1816	0.1245	103.1986	0.1281	103.3119
108	0.2125	0.1562	104.1940	0.1597	104.3036
109	0.2438	0.1882	105.1928	0.1916	105.2986

ตาราง 19 (ต่อ)

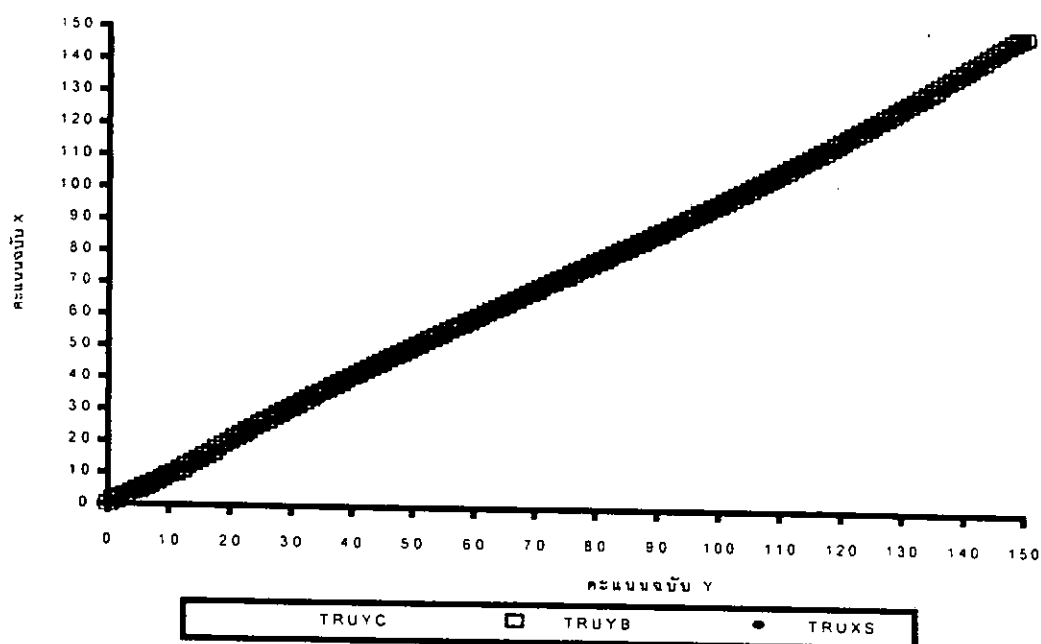
STY	THY	THYC	TRUYC	THYB	TRUYB
110	0.2754	0.2206	106.1976	0.2239	106.2995
111	0.3074	0.2534	107.2077	0.2565	107.3028
112	0.3399	0.2868	108.2286	0.2897	108.3168
113	0.3729	0.3206	109.2534	0.3234	109.3379
114	0.4063	0.3548	110.2816	0.3575	110.3624
115	0.4402	0.3896	111.3184	0.3921	111.3925
116	0.4747	0.4250	112.3629	0.4273	112.4304
117	0.5097	0.4608	113.4085	0.4630	113.4724
118	0.5453	0.4973	114.4631	0.4994	114.5234
119	0.5816	0.5346	115.5286	0.5364	115.5797
120	0.6184	0.5723	116.5927	0.5740	116.6403
121	0.6560	0.6108	117.6656	0.6124	117.7099
122	0.6942	0.6500	118.7436	0.6514	118.7818
123	0.7332	0.6900	119.8280	0.6912	119.8603
124	0.7730	0.7308	120.9175	0.7318	120.9440
125	0.8137	0.7725	122.0133	0.7734	122.0367
126	0.8553	0.8152	123.1158	0.8158	123.1312
127	0.8980	0.8589	124.2229	0.8594	124.2355
128	0.9418	0.9038	125.3369	0.9041	125.3443
129	0.9869	0.9501	126.4594	0.9502	126.4618
130	1.0335	0.9978	127.5866	0.9977	127.5842
131	1.0816	1.0472	128.7209	1.0468	128.7118
132	1.1316	1.0984	129.8596	1.0979	129.8486
133	1.1836	1.1517	131.0035	1.1509	130.9867
134	1.2379	1.2074	132.1527	1.2064	132.1325
135	1.2949	1.2658	133.3060	1.2646	133.2828
136	1.3548	1.3272	134.4616	1.3257	134.4341
137	1.4182	1.3922	135.6221	1.3904	135.5908
138	1.4856	1.4613	136.7866	1.4592	136.7522
139	1.5576	1.5352	137.9556	1.5327	137.9173

ตาราง 19 (ต่อ)

STY	THY	THYC	TRUYC	THYB	TRUYB
140	1.6352	1.6147	139.1288	1.6119	139.0889
141	1.7193	1.7009	140.3065	1.6978	140.2658
142	1.8114	1.7953	141.4892	1.7918	141.4473
143	1.9137	1.9002	142.6786	1.8962	142.6356
144	2.0292	2.0186	143.8710	2.0141	143.8285
145	2.1633	2.1561	145.0670	2.1510	145.0261
146	2.3250	2.3219	146.2585	2.3161	146.2212
147	2.5338	2.5360	147.4368	2.5292	147.4049
148	2.8434	2.8534	148.5855	2.8452	148.5631
149	3.6132	3.6426	149.7043	3.6310	149.6973
150			150.0000		150.0000

ผลการวิเคราะห์ตามตาราง 19 พบว่า คะแนนสมมูลของแบบทดสอบฉบับ Y เมื่อประมาณค่าสัมประสิทธิ์การเทียบคะแนน โดยใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (MS) ในช่วงคะแนน 1 – 3 ต่ำกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.02 ถึง 0.446 ช่วงคะแนน 4 – 18 สูงกว่า คะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.034 ถึง 0.762 ช่วงคะแนน 19 – 56 ต่ำกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.045 ถึง 1.156 ช่วงคะแนน 57 – 144 สูงกว่าคะแนนแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.032 ถึง 3.701 ช่วงคะแนน 145 – 149 ต่ำกว่าคะแนนแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.026 ถึง 0.697 และที่คะแนน 150 มีค่าเท่ากัน

สำหรับคะแนนสมมูลของแบบทดสอบฉบับ Y เมื่อประมาณค่าสัมประสิทธิ์การเทียบคะแนนโดยใช้โค้งลักษณะแบบทดสอบ (TRF) พบว่า ในช่วงคะแนน 1 – 2 ต่ำกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.175 ถึง 0.402 ช่วงคะแนน 3 – 20 สูงกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.056 ถึง 0.965 ช่วงคะแนน 21 – 52 ต่ำกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.036 ถึง 0.842 ช่วงคะแนน 53 – 144 สูงกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.037 ถึง 3.807 ช่วงคะแนน 145 – 149 ต่ำกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.067 ถึง 0.704 และที่คะแนน 150 มีค่าเท่ากัน ดังภาพประกอบ 6



ภาพประกอบ 6 ผลการเทียบคะแนนจริง 3 วิธีจากแบบทดสอบฉบับ Y ไปฉบับ X

2.1.2.2 ผลการสร้างตารางเทียบคะแนนจริงวิธีปรับค่า

พารามิเตอร์พร้อมกัน (SIM)

เมื่อวิเคราะห์ข้อมูลจากแบบทดสอบฉบับ X และฉบับ Y พร้อมกัน ทำให้ค่าประมาณพารามิเตอร์ความสามารถของผู้สอบฉบับ Y อยู่ในมาตรวัดเดียวกับฉบับ X ดังนั้นสามารถเทียบคะแนนจริงโดยไม่จำเป็นต้องแปลงพารามิเตอร์ความสามารถจากผู้สอบฉบับ Y ไปสู่ฉบับ X แต่สามารถเปรียบเทียบค่าคะแนนจริงได้โดยตรง ดังนั้น การสร้างตารางเทียบคะแนนจริง สามารถดำเนินการโดยกำหนดค่าคะแนนจริงตั้งแต่ 1-150 และหาระดับความสามารถที่สมนัยกับระดับคะแนนจริง โดยวิเคราะห์ฟังก์ชันผกผันของคะแนนจริงว่าคะแนนจริงที่มีค่าเท่านี้เกิดจากความสามารถระดับใดของผู้สอบ ปรากฏค่าในช่อง 1 - 2 ของตาราง 20 และคำนวณค่าคะแนนจริงในแบบทดสอบฉบับ X ที่ระดับความสามารถเดียวกันด้วยพารามิเตอร์ข้อสอบฉบับ X ตามสูตร 84 ในหน้า 125 ปรากฏค่าในช่อง 3 ของตาราง 20

ตาราง 20 คะแนนสมมูลของการเทียบคะแนนจริงจากแบบทดสอบฉบับ Y ไปยังฉบับ X วิธีปรับ
ค่าพารามิเตอร์พร้อมกัน

STY	THYS	TRUXS	STY	THYS	TRUXS
1	-5.2649	1.1398	29	-2.3476	29.2569
2	-4.7811	2.2697	30	-2.3017	30.2807
3	-4.4861	3.3250	31	-2.2565	31.2980
4	-4.2738	4.3021	32	-2.2120	32.3087
5	-4.1044	5.2215	33	-2.1681	33.3154
6	-3.9611	6.1033	34	-2.1248	34.3183
7	-3.8353	6.9666	35	-2.0822	35.3155
8	-3.7224	7.8250	36	-2.0402	36.3094
9	-3.6192	8.6918	37	-1.9989	37.2981
10	-3.5238	9.5750	38	-1.9583	38.2817
11	-3.4346	10.4823	39	-1.9184	39.2603
12	-3.3507	11.4161	40	-1.8792	40.2338
13	-3.2714	12.3770	41	-1.8407	41.2023
14	-3.1960	13.3657	42	-1.8029	42.1655
15	-3.1241	14.3799	43	-1.7658	43.1233
16	-3.0553	15.4172	44	-1.7293	44.0779
17	-2.9894	16.4723	45	-1.6935	45.0263
18	-2.9261	17.5417	46	-1.6582	45.9733
19	-2.8653	18.6188	47	-1.6235	46.9159
20	-2.8066	19.7028	48	-1.5894	47.8533
21	-2.7499	20.7881	49	-1.5557	48.7905
22	-2.6951	21.8695	50	-1.5224	49.7269
23	-2.6418	22.9487	51	-1.4896	50.6590
24	-2.5900	24.0204	52	-1.4572	51.5891
25	-2.5395	25.0840	53	-1.4250	52.5221
26	-2.4901	26.1400	54	-1.3932	53.4519
27	-2.4417	27.1877	55	-1.3617	54.3806
28	-2.3943	28.2249	56	-1.3305	55.3077

ตาราง 20 (ต่อ)

STY	THYS	TRUXS	STY	THYS	TRUXS
57	-1.2995	56.2358	86	-0.4631	82.9897
58	-1.2688	57.1614	87	-0.4348	83.9206
59	-1.2383	58.0871	88	-0.4065	84.8521
60	-1.2080	59.0126	89	-0.3781	85.7877
61	-1.1779	59.9376	90	-0.3497	86.7239
62	-1.1480	60.8618	91	-0.3212	87.6643
63	-1.1183	61.7850	92	-0.2926	88.6088
64	-1.0887	62.7101	93	-0.2639	89.5572
65	-1.0593	63.6337	94	-0.2352	90.5063
66	-1.0301	64.5556	95	-0.2063	91.4623
67	-1.0010	65.4786	96	-0.1773	92.4218
68	-0.9721	66.3994	97	-0.1482	93.3844
69	-0.9433	67.3208	98	-0.1190	94.3498
70	-0.9146	68.2425	99	-0.0896	95.3208
71	-0.8860	69.1643	100	-0.0601	96.2936
72	-0.8575	70.0857	101	-0.0304	97.2712
73	-0.8291	71.0066	102	-0.0005	98.2528
74	-0.8008	71.9264	103	0.0296	99.2381
75	-0.7725	72.8482	104	0.0599	100.2263
76	-0.7443	73.7684	105	0.0905	101.2204
77	-0.7161	74.6900	106	0.1213	102.2163
78	-0.6880	75.6094	107	0.1524	103.2168
79	-0.6599	76.5296	108	0.1839	104.2246
80	-0.6318	77.4506	109	0.2157	105.2358
81	-0.6037	78.3722	110	0.2478	106.2498
82	-0.5756	79.2942	111	0.2803	107.2692
83	-0.5475	80.2166	112	0.3133	108.2965
84	-0.5194	81.1394	113	0.3467	109.3278
85	-0.4913	82.0627	114	0.3805	110.3624

ตาราง 20 (ต่อ)

STY	THYS	TRUXS	STY	THYS	TRUXS
115	0.4149	111.4057	133	1.1601	131.0053
116	0.4497	112.4508	134	1.2141	132.1426
117	0.4851	113.5030	135	1.2705	133.2805
118	0.5211	114.5614	136	1.3299	134.4236
119	0.5576	115.6221	137	1.3924	135.5655
120	0.5947	116.6873	138	1.4587	136.7101
121	0.6325	117.7588	139	1.5293	137.8554
122	0.6709	118.8328	140	1.6049	139.0010
123	0.7101	119.9137	141	1.6863	140.1451
124	0.7500	120.9974	142	1.7748	141.2888
125	0.7908	122.0878	143	1.8720	142.4305
126	0.8325	123.1831	144	1.9802	143.5680
127	0.8752	124.2837	145	2.1030	144.6981
128	0.9191	125.3921	146	2.2467	145.8184
129	0.9641	126.5026	147	2.4228	146.9223
130	1.0106	127.6215	148	2.6571	147.9994
131	1.0586	128.7445	149	3.0298	149.0307
132	1.1084	129.8735	150		150.0000

ผลการวิเคราะห์ตามตาราง 20 พบว่า คะแนนสมมูลจากแบบทดสอบฉบับ Y ในช่วงคะแนน 1 – 6 ต่ำกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.103 ถึง 0.325 ช่วงคะแนน 7 – 23 สูงกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.033 ถึง 0.634 ช่วงคะแนน 24 – 45 ต่ำกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.02 ถึง 0.318 ช่วงคะแนน 46 – 147 สูงกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.027 ถึง 3.784 ที่คะแนน 148 และ 150 มีค่าเท่ากับคะแนนจริงของแบบทดสอบฉบับ X และที่คะแนน 149 ต่ำกว่าคะแนนจริงของแบบทดสอบฉบับ X มีค่าเท่ากับ .031 ดังภาพประกอบ 6

2.1.3 ผลการสร้างตารางเทียบคะแนนสังเกต

ในการสร้างตารางเทียบคะแนนสังเกต จำเป็นต้องสร้างการแจกแจงคะแนนสังเกตเชิงทฤษฎีตามวิธีการของทฤษฎีการตอบข้อสอบโดยวิเคราะห์พารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถจากข้อมูลของแบบทดสอบฉบับ X และ Y เพื่อนำมาคำนวณค่าความน่าจะเป็นที่จะเกิดคะแนนสังเกตแต่ละค่าที่เป็นไปได้ การสร้างฟังก์ชันการแจกแจงคะแนนสังเกต มีขั้นตอนดังนี้

2.1.3.1 สร้างการแจกแจงคะแนนความสามารถของผู้สอบ

โดยการคำนวณคะแนนความสามารถของผู้สอบที่ทำแบบทดสอบฉบับ X และ Y แล้วจัดชั้นคะแนนความสามารถ เพื่อหาสัดส่วนคะแนนความสามารถของผู้สอบในแต่ละชั้นคะแนนต่อจำนวนผู้สอบทั้งหมด $\psi(\theta_i)$ แล้วนำค่าจุดกลางแต่ละชั้นคะแนนเป็นค่าประมาณพารามิเตอร์ความสามารถซึ่งมีจำนวน 36 ค่า เป็นระดับความสามารถที่กำหนดขึ้นเพื่อใช้คำนวณค่าความน่าจะเป็นในแต่ละระดับคะแนน ปรากฏค่าในตาราง 54 และ 55 (ภาคผนวก) ซึ่งพบว่า ความน่าจะเป็นของแต่ละชั้นคะแนนความสามารถในแบบทดสอบฉบับ X และ Y มีค่าแตกต่างกันไม่มากนัก

2.1.3.2 สร้างการแจกแจงคะแนนสังเกตในแต่ละระดับคะแนน แต่ละระดับความสามารถของแบบทดสอบฉบับ X และ Y

โดยการคำนวณค่าความน่าจะเป็นในแต่ละระดับคะแนนแต่ละระดับความสามารถทั้ง 36 ค่า ($f(x|\theta), g(y|\theta)$) โดยคำนวณตามสูตร 85 – 86 ในหน้า 126

2.1.3.3 สร้างการแจกแจงคะแนนสังเกตในแต่ละระดับคะแนนของ แบบทดสอบฉบับ X และ Y

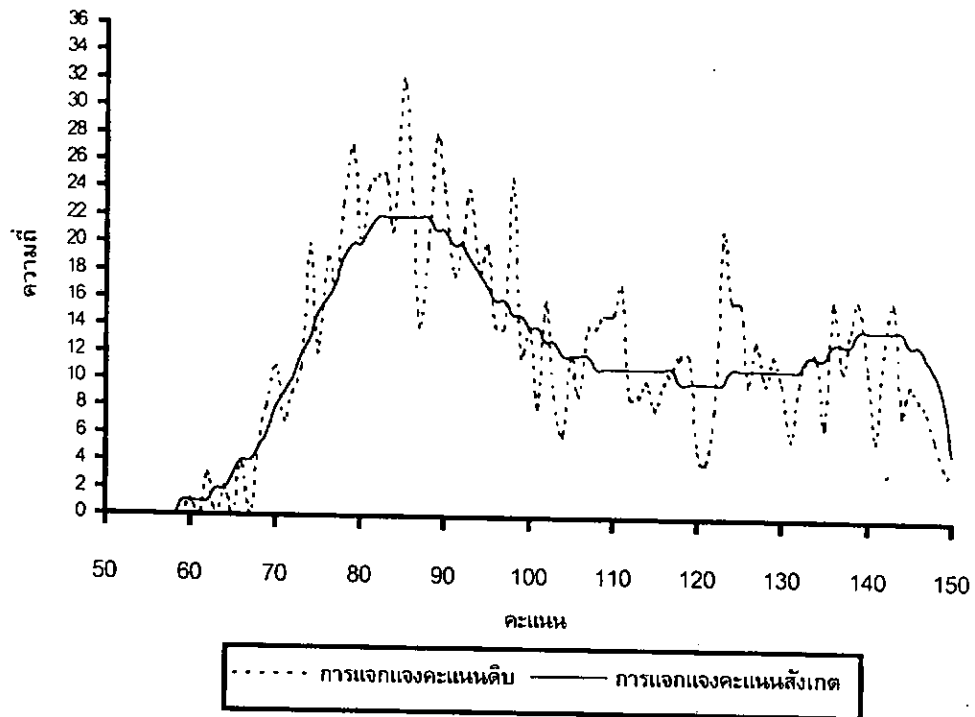
โดยการนำค่าความน่าจะเป็นที่ได้จากแต่ละระดับความสามารถทั้ง 36 ค่า (ในขั้น 2.1.3.1) คูณกับความน่าจะเป็นที่คำนวณได้ในแต่ละระดับคะแนนของแต่ละระดับความสามารถ (ในขั้น 2.1.3.2) แล้วรวมผลที่ได้ทั้ง 36 ค่าตามสูตร 87 ในหน้า 126 จะได้ฟังก์ชันการแจกแจงคะแนนสังเกตของแบบทดสอบฉบับ X และ Y ($f(x), f(y)$) ผลการวิเคราะห์ค่าสถิติพื้นฐานของการแจกแจงคะแนนดิบกับการแจกแจงคะแนนสังเกตเชิงทฤษฎีของแบบทดสอบฉบับ X ฉบับ Y ที่สร้างขึ้นตามทฤษฎีการตอบข้อสอบ ปรากฏผลตามตาราง 21

ตาราง 21 ค่าสถิติพื้นฐานของคะแนนดิบและคะแนนสังเกตที่สร้างโดยใช้ทฤษฎี IRT
โมเดลเกรด เรสพอนส์

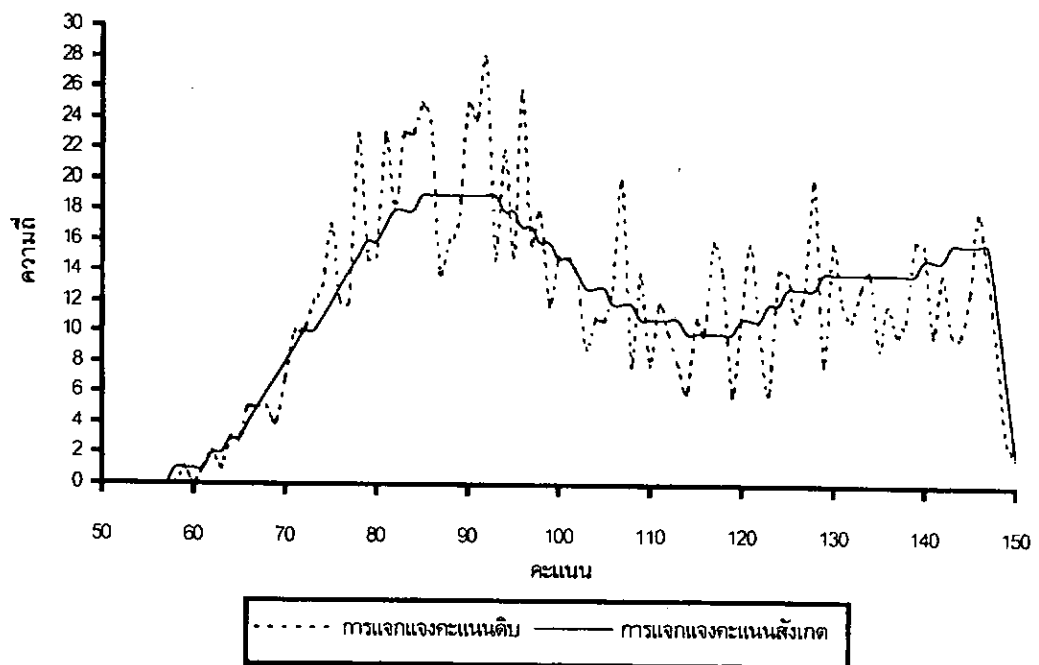
แบบ ทดสอบ	คะแนนดิบ				คะแนนสังเกต			
	$\bar{X}$	SD	Sk	Ku	$\bar{X}$	SD	Sk	Ku
ฉบับ X	103.335	22.770	0.373	-1.077	105.205	23.80	0.280	-1.164
ฉบับ Y	105.224	23.289	0.252	-1.142	107.262	23.974	0.120	-1.215

ผลการวิเคราะห์ตามตาราง 21 พบว่า ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนดิบในแบบทดสอบฉบับ X มีค่าต่ำกว่าฉบับ Y ซึ่งสอดคล้องกับคะแนนสังเกตที่สร้างขึ้น สำหรับความเบ้ของคะแนนดิบในแบบทดสอบฉบับ X มีความเบ้มากกว่าฉบับ Y เป็นการเบ้ขวา ซึ่งสอดคล้องกับความเบ้ของคะแนนสังเกต และความโด่งที่ได้จากคะแนนดิบของแบบทดสอบฉบับ Y มีความแบนกว่าฉบับ X ซึ่งสอดคล้องกับความโด่ง ที่ได้จากคะแนนสังเกตเช่นเดียวกัน

โดยภาพรวมค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ได้จากคะแนนสังเกตของแบบทดสอบทั้งสองฉบับมีค่าสูงกว่าที่ได้จากคะแนนดิบ ขณะเดียวกันความเบ้ที่ได้จากคะแนนสังเกตในแบบทดสอบฉบับ X และฉบับ Y มีค่าน้อยกว่าที่ได้จากคะแนนดิบ ทำให้ความโด่งที่ได้จากคะแนนสังเกตในแบบทดสอบฉบับ X และฉบับ Y แบนกว่าที่ได้จากคะแนนดิบ การแจกแจงของคะแนนดิบและคะแนนสังเกตในแบบทดสอบฉบับเดียวกัน ปรากฏผลดังภาพประกอบ 7 และ 8



ภาพประกอบ 7 การแจกแจงคะแนนดิบและคะแนนสังเกตที่สร้างตามทฤษฎีการตอบ
ข้อสอบในแบบทดสอบฉบับ X



ภาพประกอบ 8 การแจกแจงคะแนนดิบและคะแนนสังเกตที่สร้างตามทฤษฎีการตอบข้อสอบ
ในแบบทดสอบฉบับ y

ผลการวิเคราะห์ตามภาพประกอบ 7 และ 8 พบว่าการแจกแจงคะแนน
สังเกตที่สร้างตามทฤษฎีตอบข้อสอบในแบบทดสอบฉบับ X และ Y มีความเรียบมากกว่า
คะแนนดิบ

การทดสอบความแตกต่างระหว่างค่าเฉลี่ยของคะแนนดิบและคะแนน
สังเกตในแบบทดสอบฉบับเดียวกัน ด้วยสถิติทดสอบ t แบบ Dependent ปรากฏผลตาม
ตาราง 22

ตาราง 22 ผลการเปรียบเทียบค่าเฉลี่ยของคะแนนสังเกตที่สร้างโดยใช้ทฤษฎี IRT โมเดลเกรด
เรสพอนส์ และคะแนนดิบในแบบทดสอบเทียบคะแนนฉบับเดียวกัน

แบบทดสอบ	ผลต่างของค่าเฉลี่ย	สถิติทดสอบ t	df	p
ฉบับ X (คะแนนสังเกต – ดิบ)	1.870	1.861	1154	0.063
ฉบับ Y (คะแนนสังเกต – ดิบ)	2.038	1.965	1156	0.050

p > 0.01

ผลการวิเคราะห์ตามตาราง 22 พบว่า ค่าเฉลี่ยระหว่างคะแนนดิบกับคะแนนสังเกต
ในแบบทดสอบฉบับ X และ ฉบับ Y ไม่แตกต่างกัน

การทดสอบความแตกต่างระหว่างการแจกแจงคะแนนดิบและคะแนนสังเกตใน
แบบทดสอบฉบับเดียวกัน ด้วยสถิติทดสอบนัยพาราเมตริกซ์ Wilcoxon Signed Ranks ปรากฏ
ผลตามตาราง 23

ตาราง 23 ผลการวิเคราะห์ความแตกต่างของการแจกแจงคะแนนสังเกตที่สร้างโดยใช้ทฤษฎี
IRT โมเดลเกรด เรสพอนส์ และคะแนนดิบในแบบทดสอบเทียบคะแนนฉบับเดียวกัน

แบบทดสอบ	ผลรวมของอันดับ	สถิติทดสอบ Z	p
ฉบับ X	304488.50	-2.056	0.040
ฉบับ Y	305441.00	-2.165	0.030

p > 0.01

ผลการวิเคราะห์ตามตาราง 23 พบว่า คะแนนดิบและคะแนนสังเกตที่สร้าง
โดยใช้ทฤษฎี IRT ในแบบทดสอบฉบับ X และ ฉบับ Y มีการแจกแจงคะแนนไม่แตกต่างกัน

2.1.3.4 แปลงคะแนนความสามารถจากแบบทดสอบฉบับ Y ปรับ

สเกลสู่อันดับ X

สำหรับการเทียบคะแนนวิธีปรับค่าพารามิเตอร์แยกกัน เพื่อให้พารามิเตอร์ความสามารถอยู่ในมาตรวัดเดียวกัน นำคะแนนความสามารถทั้ง 36 ค่า (ในชั้น 2.1.3.1) มาแปลงด้วย สมการแปลงของวิธี MS และ TRF จะได้ค่าคะแนนความสามารถที่แปลงจากฉบับ Y ปรับสู่อันดับ X ดังค่าในตาราง 54 (ภาคผนวก) และนำค่าคะแนนความสามารถที่แปลงแล้วไปสร้างการแจกแจงคะแนนสังเกตในแบบทดสอบฉบับ Y ในชั้นที่ 2.1.3.2 – 2.1.3.3 สำหรับวิธีปรับค่าพารามิเตอร์พร้อมกัน ไม่จำเป็นต้องแปลงคะแนนความสามารถ สามารถใช้คะแนนความสามารถที่เป็นจุดกลางของชั้นคะแนน (ดังค่าในตาราง 55 (ภาคผนวก)) เป็นตัวแทนในการสร้างการแจกแจงคะแนนสังเกตของแบบทดสอบฉบับ Y ได้เลย

2.1.3.5 เทียบคะแนนสังเกตด้วยวิธีอีควิเปอร์เซ็นต์ไทล์

โดยการนำฟังก์ชันการแจกแจงคะแนนสังเกตของแบบทดสอบฉบับ X และ Y ปรับค่าความน่าจะเป็นให้ต่อเนื่อง (เพราะบางค่าคะแนนไม่มีความน่าจะเป็นที่จะเกิดคะแนนดังกล่าว) โดยใช้สูตร 83 ในหน้า 124 และนำฟังก์ชันการแจกแจงคะแนนสังเกตของแบบทดสอบฉบับ X และ Y ที่ปรับค่าแล้ว ดำเนินการเทียบคะแนนแบบอีควิเปอร์เซ็นต์ไทล์ โดยคำนวณค่าความน่าจะเป็นสะสมของแบบทดสอบทั้ง 2 ฉบับ ($G(X)$, $G(Y)$) แล้วคำนวณตำแหน่งเปอร์เซ็นต์ไทล์ในแต่ละระดับคะแนนของแบบทดสอบฉบับ Y จากสูตร 81 ในหน้า 123 และคำนวณฟังก์ชันผกผันของตำแหน่งเปอร์เซ็นต์ไทล์ จากสูตร 82 ในหน้า 124 ทำให้ได้คะแนนของแบบทดสอบฉบับ X ที่เทียบเท่ากับ Y ปรากฏผลการเทียบคะแนนสังเกตจากทั้ง 3 วิธี ในตาราง 24

ตาราง 24 คะแนนสมมูลของการเทียบคะแนนสังเกตจากแบบทดสอบฉบับ Y ไปยัง
ฉบับ X จำแนกตามวิธีการเทียบคะแนน

SY	XBE	XCE	XSE	SY	XBE	XCE	XSE
1	1.0000	1.0000	1.0000	29	29.0000	29.0000	29.0000
2	2.0000	2.0000	2.0000	30	30.0000	30.0000	30.0000
3	3.0000	3.0000	3.0000	31	31.0000	31.0000	31.0000
4	4.0000	4.0000	4.0000	32	32.0000	32.0000	32.0000
5	5.0000	5.0000	5.0000	33	33.0000	33.0000	33.0000
6	6.0000	6.0000	6.0000	34	34.0000	34.0000	34.0000
7	7.0000	7.0000	7.0000	35	35.0000	35.0000	35.0000
8	8.0000	8.0000	8.0000	36	36.0000	36.0000	36.0000
9	9.0000	9.0000	9.0000	37	37.0000	37.0000	37.0000
10	10.0000	10.0000	10.0000	38	38.0000	38.0000	38.0000
11	11.0000	11.0000	11.0000	39	39.0000	39.0000	39.0000
12	12.0000	12.0000	12.0000	40	40.0000	40.0000	40.0000
13	13.0000	13.0000	13.0000	41	41.0000	41.0000	41.0000
14	14.0000	14.0000	14.0000	42	42.0000	42.0000	42.0000
15	15.0000	15.0000	15.0000	43	43.0000	43.0000	43.0000
16	16.0000	16.0000	16.0000	44	44.0000	44.0000	44.0000
17	17.0000	17.0000	17.0000	45	45.5000	46.0000	45.0000
18	18.0000	18.0000	18.0000	46	48.0000	48.6000	46.5000
19	19.0000	19.0000	19.0000	47	49.5000	49.8750	48.2500
20	20.0000	20.0000	20.0000	48	50.7500	51.1250	49.6000
21	21.0000	21.0000	21.0000	49	51.8929	52.2857	50.7000
22	22.0000	22.0000	22.0000	50	53.0319	53.4255	51.6579
23	23.0000	23.0000	23.0000	51	54.1410	54.5163	52.6029
24	24.0000	24.0000	24.0000	52	55.2236	55.5699	53.5508
25	25.0000	25.0000	25.0000	53	56.2957	56.6195	54.4915
26	26.0000	26.0000	26.0000	54	57.3566	57.6636	55.3889
27	27.0000	27.0000	27.0000	55	58.4065	58.7015	56.2994
28	28.0000	28.0000	28.0000	56	59.4454	59.7315	57.2269

ตาราง 24 (ต่อ)

SY	XBE	XCE	XSE	SY	XBE	XCE	XSE
57	60.4755	60.7548	58.1676	86	86.2715	86.4569	84.3887
58	61.4979	61.7722	59.1212	87	87.1291	87.3119	85.1910
59	62.5108	62.7845	60.0888	88	87.9919	88.1722	86.0041
60	63.5201	63.7926	61.0688	89	88.8598	89.0377	86.8259
61	64.5240	64.7948	62.0571	90	89.7326	89.9084	87.6552
62	65.5218	65.7896	63.0508	91	90.6106	90.7843	88.4921
63	66.5109	66.7745	64.0466	92	91.4940	91.6656	89.3423
64	67.4864	67.7470	65.0422	93	92.3872	92.5529	90.2043
65	68.4437	68.7047	66.0361	94	93.2874	93.4488	91.0803
66	69.3836	69.6459	67.0277	95	94.1950	94.3540	91.9725
67	70.3053	70.5698	68.0174	96	95.1097	95.2662	92.8824
68	71.2092	71.4725	69.0056	97	96.0307	96.1844	93.8107
69	72.0959	72.3468	69.9922	98	96.9563	97.1065	94.7566
70	72.9673	73.2067	70.9761	99	97.8838	98.0299	95.7169
71	73.8253	74.0544	71.9545	100	98.8100	98.9512	96.6867
72	74.6723	74.8924	72.9232	101	99.7316	99.8673	97.6597
73	75.5104	75.7229	73.8767	102	100.6465	100.7764	98.6295
74	76.3282	76.5479	74.8094	103	101.5542	101.6783	99.5918
75	77.1438	77.3597	75.7161	104	102.4571	102.5745	100.5449
76	77.9596	78.1694	76.5934	105	103.3587	103.4692	100.9887
77	78.7766	78.9817	77.4353	106	104.2614	104.3675	102.4338
78	79.5955	79.7970	78.2413	107	105.1689	105.2709	103.3792
79	80.4139	80.6155	79.0246	108	106.0834	106.1816	104.3319
80	81.2347	81.4360	79.7916	109	107.0051	107.0992	105.2954
81	82.0613	82.2595	80.5494	110	107.9310	108.0205	106.2697
82	82.8934	83.0889	81.2987	111	108.8560	108.9402	107.2507
83	83.7306	83.9237	82.0538	112	109.7743	109.8527	108.2310
84	84.5725	84.7634	82.8201	113	110.6813	110.7535	109.2022
85	85.4192	85.6078	83.5987	114	111.5755	111.6415	110.1576

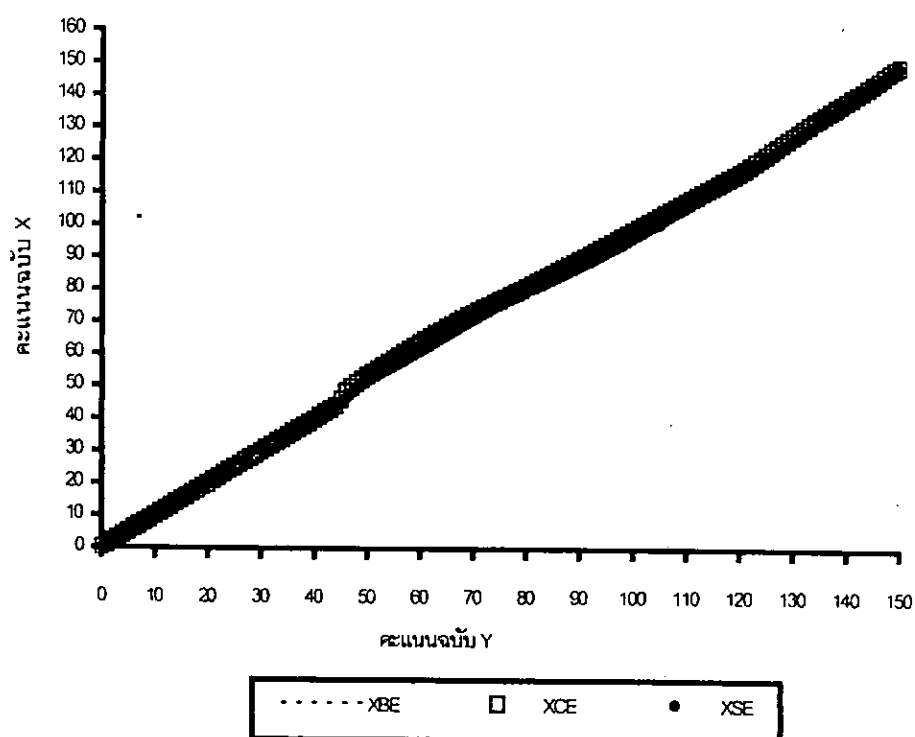
ตาราง 24 (ต่อ)

SY	XBE	XCE	XSE	SY	XBE	XCE	XSE
115	112.4600	112.5201	111.0951	133	131.9437	131.9262	130.8248
116	113.3420	113.3969	112.0180	134	133.0708	133.0498	132.0217
117	114.2323	114.2826	112.9339	135	134.1686	134.1444	133.2017
118	115.1430	115.1895	113.8539	136	135.2369	135.2098	134.3545
119	116.0859	116.1292	114.7903	137	136.2824	136.2527	135.4765
120	117.0710	117.1111	115.7556	138	137.3149	137.2828	136.5698
121	118.1048	118.1417	116.7614	139	138.3430	138.3087	137.6456
122	119.1886	119.2220	117.8161	140	139.3731	139.3370	138.7144
123	120.3171	120.3464	118.9225	141	140.4097	140.3722	139.7844
124	121.4784	121.5029	120.0759	142	141.4573	141.4187	140.8615
125	122.6536	122.6727	121.2636	143	142.5216	142.4822	141.9490
126	123.8288	123.8427	122.4690	144	143.6148	143.5744	143.0483
127	124.9971	125.0059	123.6711	145	144.7537	144.7130	144.1653
128	126.1580	126.1619	124.8652	146	145.9520	145.9127	145.3174
129	127.3152	127.3144	126.0528	147	147.1931	147.1583	146.5223
130	128.4732	128.4680	127.2384	148	148.4027	148.3773	147.7747
131	129.6341	129.6246	128.4274	149	149.4623	149.4486	148.9872
132	130.7943	130.7807	129.6235	150	150.3300	150.3200	150.0700

ผลการวิเคราะห์ตามตาราง 24 เมื่อเทียบคะแนนสังเกตจากแบบทดสอบฉบับ Y ไปยังฉบับ X โดยประมาณค่าสัมประสิทธิ์การเทียบคะแนนด้วยวิธี MS พบว่า คะแนนสมมูลจากแบบทดสอบฉบับ Y ในช่วงคะแนน 1 – 44 เท่ากับคะแนนสังเกตของแบบทดสอบฉบับ X (ซึ่งเป็นช่วงคะแนนที่ไม่มีผู้สอบทำได้ทั้งฉบับ X และ Y) ช่วงคะแนน 45 – 87 ต่ำกว่าคะแนนสังเกตของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.129 ถึง 3.524 ช่วงคะแนน 88 – 146 สูงกว่าคะแนนสังเกตของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.008 ถึง 2.929 และช่วงคะแนน 147 – 150 ต่ำกว่าคะแนนสังเกตของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.193 ถึง 0.462

สำหรับการเทียบคะแนนสังเกตจากแบบทดสอบฉบับ Y ไปยังฉบับ X เมื่อประมาณค่าสัมประสิทธิ์การเทียบคะแนนด้วยวิธี TRF พบว่า คะแนนสมมูลจากแบบทดสอบฉบับ Y ช่วงคะแนน 1 – 44 เท่ากับคะแนนสังเกตของแบบทดสอบฉบับ X (เป็นช่วงคะแนนที่ไม่มีผู้สอบทำได้ทั้งในฉบับ X และ Y) ช่วงคะแนน 45 – 89 ต่ำกว่าคะแนนสังเกตของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.038 ถึง 3.795 ช่วงคะแนน 90 – 146 สูงกว่าคะแนนสังเกตของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.087 ถึง 2.889 และช่วงคะแนน 147 – 150 ต่ำกว่าคะแนนสังเกตของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.158 ถึง 0.449

และการเทียบคะแนนสังเกตจากแบบทดสอบฉบับ Y ไปยังฉบับ X วิธีปรับค่าพารามิเตอร์พร้อมกัน พบว่า คะแนนสมมูลจากแบบทดสอบฉบับ Y ช่วงคะแนน 1 – 45 เท่ากับคะแนนสังเกตของแบบทดสอบฉบับ X (เป็นช่วงคะแนนที่ไม่มีผู้สอบทำได้ ทั้งในฉบับ X และ Y) ช่วงคะแนน 46 – 79 ต่ำกว่าคะแนนสังเกตของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.025 ถึง 1.70 ช่วงคะแนน 80 – 149 สูงกว่าคะแนนสังเกตของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.013 ถึง 4.244 และที่ระดับคะแนน 150 ต่ำกว่าคะแนนสังเกตของแบบทดสอบฉบับ X มีค่าเท่ากับ 0.07 ซึ่งคะแนนแปลงที่ได้จากการเทียบคะแนนสังเกตทั้ง 3 วิธีให้ผลที่ใกล้เคียงกัน ดังภาพประกอบ 9



ภาพประกอบ 9 ผลการเทียบคะแนนสังเกต 3 วิธี จากแบบทดสอบฉบับ Y ปรับสู่สเกล
ยังฉบับ X

2.2 ผลการสร้างตารางเทียบคะแนนตามทฤษฎีการทดสอบแบบ

มาตรฐานเดิม

การเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม เป็นการเทียบคะแนนดิบด้วยวิธีอีควิเปอร์เซ็นไทล์และวิธีเชิงเส้นตรง ในแบบแผนการรวบรวมข้อมูลด้วยการใช้แบบทดสอบรวมภายในกับกลุ่มสุ่ม มีรายละเอียดของการสร้างตารางเทียบคะแนนดังต่อไปนี้

2.2.1 ผลการสร้างตารางเทียบคะแนนวิธีอีควิเปอร์เซ็นไทล์

การเทียบคะแนนวิธีอีควิเปอร์เซ็นไทล์เป็นการเทียบคะแนนดิบจากแบบทดสอบฉบับ Y ไปยังฉบับ X วิธีการที่นำมาใช้เป็นการประมาณความถี่ในการแจกแจงคะแนนของแบบทดสอบฉบับ X และ Y โดยใช้ข้อสนเทศของกลุ่มตัวอย่างรวม ที่ทำข้อสอบรวมในแบบทดสอบเทียบคะแนนทั้ง 2 ฉบับ การแจกแจงคะแนนของแบบทดสอบฉบับ X และ Y ที่ปรับแล้ว ปรากฏว่าระดับคะแนน 0 – 59 และระดับคะแนน 0 – 60 ในแบบทดสอบฉบับ Y และ X ตามลำดับ ไม่มีผู้สอบทำคะแนนได้ ดำเนินการคำนวณค่าความถี่สัมพัทธ์ของแบบทดสอบแต่ละฉบับและปรับค่าให้ต่อเนื่อง ตามสูตร 83 หน้า 124 แล้วคำนวณค่าความถี่สัมพัทธ์สะสมในแบบทดสอบแต่ละฉบับ ต่อมาคำนวณตำแหน่งเปอร์เซ็นต์ไทล์ของคะแนนในแบบทดสอบฉบับ Y ตามสูตร 81 หน้า 123 และคำนวณฟังก์ชันผกผันของตำแหน่งเปอร์เซ็นต์ไทล์ เพื่อหาค่าคะแนนฉบับ X ที่เทียบเท่าฉบับ Y ตามสูตร 82 หน้า 124 ค่าดังกล่าวข้างต้นปรากฏในตาราง 59 (ภาคผนวก) ปรากฏผลการเทียบคะแนนดังตาราง 25

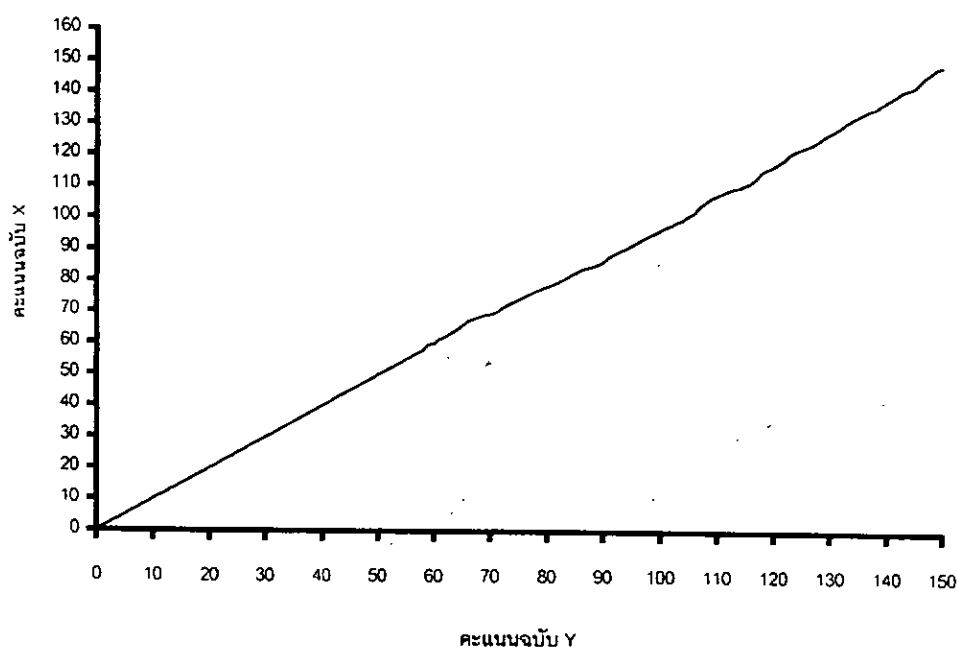
ตาราง 25 คะแนนสมมูลของการเทียบคะแนนดิบจากแบบทดสอบฉบับ Y ไปยังฉบับ X ตามวิธีการเทียบคะแนนแบบอีควิเปอร์เซ็นไทล์

SY	XEQ	SY	XEQ	SY	XEQ	SY	XEQ
1	1.0000	10	10.0000	19	19.0000	28	28.0000
2	2.0000	11	11.0000	20	20.0000	29	29.0000
3	3.0000	12	12.0000	21	21.0000	30	30.0000
4	4.0000	13	13.0000	22	22.0000	31	31.0000
5	5.0000	14	14.0000	23	23.0000	32	32.0000
6	6.0000	15	15.0000	24	24.0000	33	33.0000
7	7.0000	16	16.0000	25	25.0000	34	34.0000
8	8.0000	17	17.0000	26	26.0000	35	35.0000
9	9.0000	18	18.0000	27	27.0000	36	36.0000

ตาราง 25 (ต่อ)

SY	XEQ	SY	XEQ	SY	XEQ	SY	XEQ
37	37.0000	66	67.5302	95	91.8820	124	122.6185
38	38.0000	67	68.2352	96	92.8933	125	123.2729
39	39.0000	68	68.8531	97	93.9792	126	123.9572
40	40.0000	69	69.3671	98	94.9593	127	124.6943
41	41.0000	70	69.7575	99	95.8773	128	125.8788
42	42.0000	71	70.3492	100	96.8923	129	127.0952
43	43.0000	72	71.5204	101	97.8143	130	128.1155
44	44.0000	73	72.6238	102	98.4453	131	129.2764
45	45.0000	74	73.6593	103	99.4016	132	130.3087
46	46.0000	75	74.4117	104	100.2101	133	131.8805
47	47.0000	76	75.5216	105	101.4815	134	133.0727
48	48.0000	77	76.1586	106	102.2947	135	133.9726
49	49.0000	78	77.1530	107	104.4281	136	135.0134
50	50.0000	79	77.9946	108	105.8481	137	135.9172
51	51.0000	80	78.5461	109	107.0184	138	136.5222
52	52.0000	81	79.2235	110	107.9440	139	137.6782
53	53.0000	82	80.1214	111	108.7392	140	138.7982
54	54.0000	83	81.0317	112	109.5592	141	139.6972
55	55.0000	84	81.9809	113	110.1772	142	140.9655
56	56.0000	85	82.9320	114	110.6421	143	142.1508
57	57.0000	86	83.9606	115	111.1657	144	142.8174
58	58.0000	87	84.7274	116	112.1220	145	143.5276
59	59.8263	88	85.1985	117	113.6552	146	145.1617
60	60.1543	89	85.7683	118	115.5112	147	146.8420
61	61.7087	90	86.7037	119	116.5177	148	148.2898
62	62.2285	91	88.1744	120	117.2771	149	149.5074
63	63.6479	92	89.1909	121	118.4959	150	150.2100
64	64.3409	93	90.0255	122	119.7585		
65	65.8649	94	90.8906	123	121.5594		

ผลการวิเคราะห์ตามตาราง 25 พบว่า คะแนนสมมูลจากแบบทดสอบฉบับ Y ช่วงคะแนน 1 – 58 เท่ากับคะแนนของแบบทดสอบฉบับ X (เป็นช่วงคะแนนที่ไม่มีผู้สอบทำได้ทั้งฉบับ X และ Y) ช่วงคะแนน 59 – 69 ต่ำกว่าคะแนนของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.154 ถึง 1.53 ช่วงคะแนน 70 – 147 สูงกว่าคะแนนของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.158 ถึง 3.878 และช่วงคะแนน 148 – 150 ต่ำกว่าคะแนนของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.21 ถึง 0.507 ดังภาพประกอบ 10



ภาพประกอบ 10 ผลการเทียบคะแนนวิธีอีควิเปอร์เซ็นไทล์จากแบบทดสอบฉบับ Y ไปยังฉบับ X

2.2.2 ผลการสร้างตารางเทียบคะแนนวิธีเชิงเส้นตรง

การเทียบคะแนนวิธีเชิงเส้นตรงเมื่อใช้แบบแผนแบบทดสอบร่วมในกลุ่มสุ่มนั้น การสร้างสมการแปลงคะแนนต้องอาศัยข้อมูลจากแบบทดสอบร่วม มาปรับค่าเฉลี่ยและความแปรปรวนของแบบทดสอบเทียบคะแนนทั้ง 2 ฉบับ จึงจำเป็นต้องมีการตรวจสอบความเชื่อมั่นของแบบทดสอบเทียบคะแนนทั้งสองฉบับ เพื่อเลือกรูปแบบการคำนวณค่าความชันและจุดตัดของสมการแปลงคะแนน จึงขอเสนอผลการวิเคราะห์ข้อมูลต่าง ๆ รวมทั้งตารางเทียบคะแนนดังต่อไปนี้

2.2.2.1 ผลการเปรียบเทียบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบเทียบคะแนน

ผู้วิจัยนำค่าความเชื่อมั่นของแบบทดสอบเทียบคะแนนทั้งสองฉบับที่วิเคราะห์โดยใช้สัมประสิทธิ์แอลฟาของครอนบัคมาทดสอบว่ามีความแตกต่างกันหรือไม่ โดยใช้สถิติทดสอบ W ของ เฟลด์ (Feldt.1969 : 365) เพื่อเป็นเงื่อนไขในการวิเคราะห์ค่าความชัน (A) และ จุดตัด (B) ของสมการแปลง ปรากฏผลการทดสอบตามตาราง 26

ตาราง 26 ผลการวิเคราะห์ความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบเทียบคะแนน

ความเชื่อมั่นของแบบ ทดสอบฉบับ X	ความเชื่อมั่นของแบบ ทดสอบฉบับ Y	สถิติทดสอบ W	p
.9372	.9400	.9554	.7809

$p > 0.01$

ผลการวิเคราะห์ตามตาราง 26 พบว่า แบบทดสอบเทียบคะแนนทั้ง 2 ฉบับมีความเชื่อมั่นของแบบทดสอบไม่แตกต่างกัน จึงคำนวณค่าความชัน (A) และจุดตัด (B) จากรูปแบบ 3A ของแองกอฟฟ์ (Angoff. 1984 : 106)

2.2.2.2 ค่าสถิติที่ใช้ในการคำนวณความชัน (A) และจุดตัด (B) ของสมการเส้นตรง

การคำนวณค่าความชัน (A) และจุดตัด (B) ของสมการเส้นตรงใช้ค่าสถิติต่าง ๆ จากกลุ่มตัวอย่างเทียบคะแนน 2 กลุ่ม กลุ่ม α สอบแบบทดสอบฉบับ X กลุ่ม β สอบแบบทดสอบฉบับ Y ฉบับละ 30 ข้อ คำนวณค่าเฉลี่ย และความแปรปรวนของคะแนนจากแบบทดสอบฉบับ X และ Y ที่ได้จากกลุ่ม α และ β ตามลำดับ ค่าเฉลี่ยและความแปรปรวนของคะแนนจากแบบทดสอบรวมทั้งที่สอบโดยกลุ่ม α , β และกลุ่มรวม $t(\alpha + \beta)$ แล้วปรับค่าเฉลี่ยและความแปรปรวนของแบบทดสอบเทียบคะแนนฉบับ X และ Y ด้วยค่าสถิติในแบบทดสอบรวม ตามวิธีการประมาณค่าความเป็นไปได้สูงสุดของ ลอร์ด (Angoff. 1984 : 105 – 106; citing Lord 1955a) ปรากฏผลตามตาราง 27

ตาราง 27 ค่าสถิติที่ใช้ในการคำนวณค่าความชัน (A) และจุดตัด (B) ของสมการเส้นตรง

แบบทดสอบ	ค่าสถิติ	
	ค่าเฉลี่ย	ความแปรปรวน
ฉบับ X	103.167	537.575
ฉบับ Y	105.363	523.468

ผลการวิเคราะห์ตามตาราง 27 พบว่า ค่าเฉลี่ยที่ปรับแล้วของแบบทดสอบฉบับ Y สูงกว่าฉบับ X ส่วนความแปรปรวนที่ปรับแล้วของแบบทดสอบฉบับ X สูงกว่าฉบับ Y เมื่อพิจารณาเปรียบเทียบกับค่าเฉลี่ยของแบบทดสอบฉบับ X และ Y ที่ไม่ได้ปรับ พบว่า ค่าเฉลี่ยของแบบทดสอบฉบับ X ที่ยังไม่ได้ปรับมีค่าสูงกว่าค่าเฉลี่ยของแบบทดสอบฉบับ X ที่ปรับแล้ว และค่าเฉลี่ยของแบบทดสอบฉบับ Y ที่ยังไม่ได้ปรับมีค่าต่ำกว่าค่าเฉลี่ยของแบบทดสอบฉบับ Y ที่ปรับแล้ว สำหรับความแปรปรวนของแบบทดสอบฉบับ X ที่ยังไม่ได้ปรับมีค่าต่ำกว่าความแปรปรวนของแบบทดสอบฉบับ X ที่ปรับแล้วและความแปรปรวนของแบบทดสอบฉบับ Y ที่ยังไม่ได้ปรับมีค่าสูงกว่าความแปรปรวนของแบบทดสอบฉบับ Y ที่ปรับแล้ว

2.2.2.3 สมการเส้นตรงที่ใช้ในการเทียบคะแนน

นำค่าเฉลี่ยและความแปรปรวนของแบบทดสอบฉบับ X และ Y ที่ปรับแล้วมาคำนวณค่าความชัน (A) และจุดตัด (B) ตามสูตรของเลวิน (Angoff, 1984 : 105 – 106; citing Levine, 1955) เพื่อสร้างสมการเทียบคะแนน ปรากฏผลดังตาราง 28

ตาราง 28 ค่าความชัน (A) และจุดตัด (B) ของสมการเทียบคะแนนเชิงเส้นตรง

A	B	สมการเส้นตรงในการเทียบคะแนน
1.027	5.041	$X^* = 1.027 (Y) - 5.041$

ผลการวิเคราะห์ตามตาราง 28 พบว่า สมการเส้นตรงในการเทียบคะแนนมีค่าความชันเป็นบวกและค่าจุดตัดเป็นลบ

2.2.2.4 ผลการสร้างตารางเทียบคะแนนเชิงเส้นตรง

นำคะแนนดิบจากแบบทดสอบฉบับ Y แทนค่าในสมการเชิงเส้นได้คะแนนสมมูลปรากฏผลตามตาราง 29

ตาราง 29 คะแนนสมมูลของการเทียบคะแนนจากแบบทดสอบฉบับ Y ไปยังฉบับ X ด้วย
วิธีการเทียบคะแนนเชิงเส้นตรง

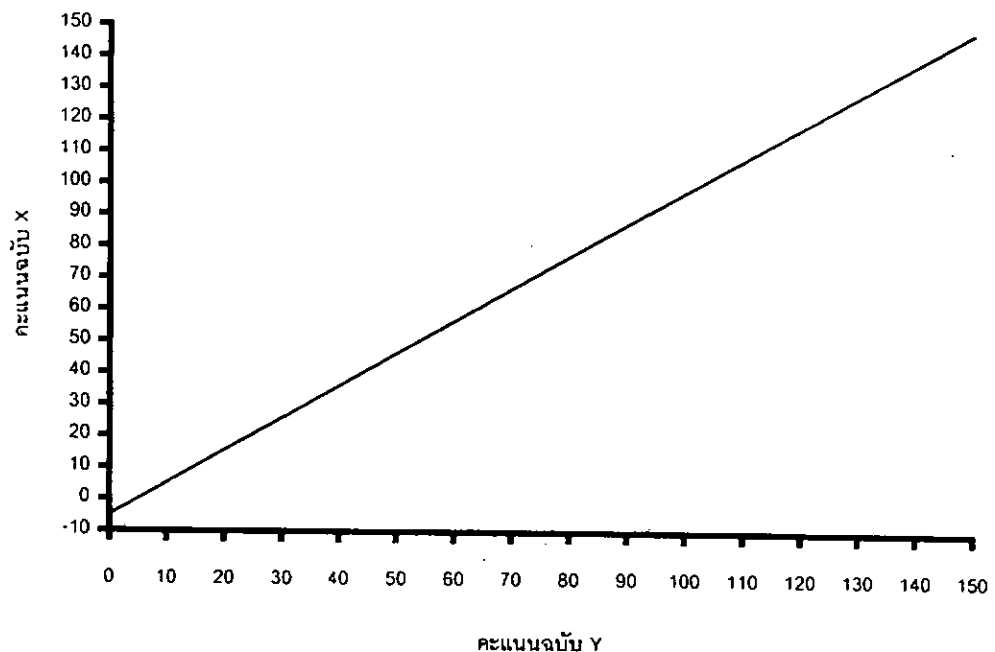
SY	XEE	SY	XEE	SY	XEE
1	-4.0085	28	23.7191	55	51.4468
2	-2.9815	29	24.7461	56	52.4737
3	-1.9546	30	25.7730	57	53.5007
4	-0.9276	31	26.8000	58	54.5276
5	0.0993	32	27.8269	59	55.5546
6	1.1263	33	28.8539	60	56.5815
7	2.1532	34	29.8808	61	57.6085
8	3.1802	35	30.9078	62	58.6354
9	4.2071	36	31.9347	63	59.6624
10	5.2341	37	32.9617	64	60.6893
11	6.2610	38	33.9886	65	61.7163
12	7.2879	39	35.0156	66	62.7432
13	8.3149	40	36.0425	67	63.7702
14	9.3418	41	37.0695	68	64.7971
15	10.3688	42	38.0964	69	65.8240
16	11.3957	43	39.1234	70	66.8510
17	12.4227	44	40.1503	71	67.8779
18	13.4496	45	41.1773	72	68.9049
19	14.4766	46	42.2042	73	69.9318
20	15.5035	47	43.2312	74	70.9588
21	16.5305	48	44.2581	75	71.9857
22	17.5574	49	45.2851	76	73.0127
23	18.5844	50	46.3120	77	74.0396
24	19.6113	51	47.3390	78	75.0666
25	20.6383	52	48.3659	79	76.0935
26	21.6652	53	49.3929	80	77.1205
27	22.6922	54	50.4198	81	78.1474

ตารางที่ 29 (ต่อ)

SY	XEE	SY	XEE	SY	XEE
82	79.1744	105	102.7942	128	126.4140
83	80.2013	106	103.8212	129	127.4410
84	81.2283	107	104.8481	130	128.4679
85	82.2552	108	105.8751	131	129.4949
86	83.2822	109	106.9020	132	130.5218
87	84.3091	110	107.9290	133	131.5488
88	85.3361	111	108.9559	134	132.5757
89	86.3630	112	109.9829	135	133.6027
90	87.3900	113	111.0098	136	134.6296
91	88.4169	114	112.0368	137	135.6566
92	89.4439	115	113.0637	138	136.6835
93	90.4708	116	114.0907	139	137.7105
94	91.4978	117	115.1176	140	138.7374
95	92.5247	118	116.1446	141	139.7644
96	93.5517	119	117.1715	142	140.7913
97	94.5786	120	118.1985	143	141.8183
98	95.6056	121	119.2254	144	142.8452
99	96.6325	122	120.2524	145	143.8722
100	97.6595	123	121.2793	146	144.8991
101	98.6864	124	122.3063	147	145.9261
102	99.7134	125	123.3332	148	146.9530
103	100.7403	126	124.3601	149	147.9800
104	101.7673	127	125.3871	150	149.0069

ผลการวิเคราะห์ตามตาราง 28 พบว่า คะแนนสมมูลของแบบทดสอบฉบับ Y ทุกระดับคะแนน (1 - 150) สูงกว่าคะแนนของแบบทดสอบฉบับ X มีค่าตั้งแต่ 0.993 ถึง 5.009 โดยในช่วงคะแนน 1 - 58 (ไม่มีผู้สอบได้คะแนนดังกล่าว) คะแนนสมมูลของแบบทดสอบฉบับ Y สูงกว่า คะแนนแบบทดสอบฉบับ X มีค่าตั้งแต่ 3.472 ถึง 5.009 ช่วงคะแนน 59 - 104 มีค่าตั้งแต่ 2.233 ถึง 3.445 ช่วงคะแนน 105 - 150 มีค่าตั้งแต่ 0.991 ถึง 2.206 แสดงว่าในระดับคะแนน

ที่เข้าใกล้คะแนนเต็ม คะแนนสมมูลของแบบทดสอบฉบับ Y มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X น้อยกว่าเมื่อคะแนนสมมูลมีระดับคะแนนต่ำ ดังภาพประกอบ 11



ภาพประกอบ 11 ผลการเทียบคะแนนเชิงเส้นตรงจากแบบทดสอบฉบับ Y ปรับสู่สเกลฉบับ X

สรุปได้ว่า การเทียบคะแนนภายใต้วิธีการเทียบคะแนนที่แตกต่างกัน 5 วิธี แต่ละวิธี ได้ผลสุดท้ายเป็นตารางแปลงคะแนนจากมาตรวัดของแบบทดสอบฉบับ Y ไปยังมาตรวัดของแบบทดสอบฉบับ X มีทั้งหมด 8 ตาราง จำแนกเป็นการเทียบคะแนนดิบ คะแนนจริง และคะแนนสังเกต (ตาราง 19 20 24 25 และ 29) เมื่อพิจารณาเปรียบเทียบกัน ปรากฏว่า มีส่วนที่สอดคล้องกันทั้ง 5 วิธีการเทียบคะแนน คือ คะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่มีค่าสูงกว่าคะแนนแบบทดสอบฉบับ X และช่วงคะแนนแปลงที่อยู่ใกล้ค่าเฉลี่ยมีความคล้ายคลึงกัน ความแตกต่างของคะแนนแปลงเกิดขึ้นเฉพาะส่วนปลายที่อยู่ใกล้คะแนนสูงสุดและต่ำสุด

ตอนที่ 3 ผลการวิเคราะห์ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม

นำคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีมาวิเคราะห์ค่าสถิติพื้นฐานและเปรียบเทียบความแตกต่างของค่าเฉลี่ยระหว่างคะแนนแปลงและคะแนนจากแบบทดสอบฉบับเป้าหมาย (ฉบับ X) พร้อมทั้งตรวจสอบความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงที่ได้จากการเทียบคะแนนแต่ละวิธีกับคะแนนจากแบบทดสอบฉบับเป้าหมาย เพื่อแสดงคุณสมบัติของการเทียบคะแนน ปรากฏผลตามตาราง 30-32 ดังนี้

3.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนแปลง

ผู้วิจัยได้วิเคราะห์ค่าสถิติพื้นฐาน คือ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานความเบ้ และความโด่ง ของคะแนนจากแบบทดสอบเทียบคะแนนฉบับ X และ Y และคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี ความแตกต่างสัมบูรณ์ของค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐานความเบ้ และความโด่ง ระหว่างคะแนนจากแบบทดสอบฉบับเป้าหมายกับคะแนนแปลงมีความแตกต่างมากน้อยต่างกันไปตามค่าสถิติแต่ละประเภท ปรากฏผลดังตาราง 30

ตาราง 30 ค่าสถิติพื้นฐานของคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีและคะแนนจากแบบทดสอบฉบับ X และ Y

ฉบับ/ วิธีการ	ค่าเฉลี่ย	ความ แตกต่าง	ส่วนเบี่ยงเบน มาตรฐาน	ความ แตกต่าง	ความ เบ้	ความ แตกต่าง	ความ โด่ง	ความแตกต่าง
ฉบับ Y	105.224		23.289		0.252		-1.142	
ฉบับ X	103.335		22.770		0.373		-1.077	
LIN	103.024	0.311	23.917	1.147	0.252	0.121	-1.142	0.065
EQU	103.233	0.102	23.351	0.581	0.360	0.013	-1.132	0.055
TMS	104.766	1.431	24.055	1.285	0.273	0.100	-1.205	0.128
TTRF	104.656	1.321	24.157	1.387	0.272	0.101	-1.207	0.130
TS	104.663	1.328	24.326	1.556	0.266	0.107	-1.215	0.138
OMS	104.740	1.405	22.363	0.407	0.391	0.018	-1.055	0.022
OTRF	104.847	1.512	22.271	0.499	0.392	0.019	-1.054	0.023
OS	103.176	0.159	22.755	0.015	0.415	0.042	-1.044	0.033

ผลการวิเคราะห์ตามตาราง 30 พบว่า ความแตกต่างของค่าเฉลี่ยระหว่างคะแนนแปลงจากการเทียบคะแนนวิธีอีคิวเปอร์เซ็นต์ไทล์ และการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์ พร้อมกันกับคะแนนจากแบบทดสอบเป้าหมายให้ค่าน้อยที่สุดเท่ากับ 0.102 และ 0.159 ตามลำดับ ขณะที่ความแตกต่างของค่าเฉลี่ยระหว่างคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงกับคะแนนจากแบบทดสอบเป้าหมายมีค่า เท่ากับ 0.311 ซึ่งควรจะมีค่าความแตกต่างน้อยที่สุดเมื่อเทียบกับคะแนนแปลงจากวิธีการเทียบคะแนนอื่น ๆ และความแตกต่างของค่าเฉลี่ยระหว่าง

คะแนนแปลง จากการเทียบคะแนนสังเกตวิธี TRF กับคะแนนจากแบบทดสอบเป้าหมายให้ค่าสูงสุดเท่ากับ 1.512

สำหรับความแตกต่างของส่วนเบี่ยงเบนมาตรฐานระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตทุกวิธี คือวิธี SIM MS และ TRF กับคะแนนจากแบบทดสอบเป้าหมายให้ค่าน้อยที่สุด เท่ากับ 0.015 0.407 และ 0.499 ตามลำดับ และความแตกต่างระหว่างส่วนเบี่ยงเบนมาตรฐานของคะแนนแปลงจากการเทียบคะแนนจริงทุกวิธี คือวิธี MS TRF และ SIM กับคะแนนจากแบบทดสอบฉบับเป้าหมายให้ค่าสูงสุดเท่ากับ 1.285 1.387 และ 1.556 ตามลำดับ

และความแตกต่างของความเบ้และความโด่งระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตทุกวิธี คือ วิธี MS TRF และ SIM กับคะแนนจากแบบทดสอบฉบับเป้าหมายให้ค่าน้อยที่สุด โดยที่ความแตกต่างของความเบ้ มีค่าเท่ากับ 0.018 0.019 และ 0.042 ตามลำดับ และความแตกต่างของความโด่งมีค่าเท่ากับ 0.022 0.023 และ 0.033 ตามลำดับ สำหรับคะแนนแปลงจากการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ให้ความแตกต่างเฉพาะความเบ้จากคะแนนของแบบทดสอบเป้าหมายมีค่าเท่ากับ 0.013 ซึ่งเป็นค่าน้อยที่สุด

จากผลดังกล่าวข้างต้นทำให้ไม่สามารถสรุปได้ว่าโมเมนต์ของคะแนนแปลงจากการเทียบคะแนนวิธีใดให้ผลที่คล้ายคลึงกับโมเมนต์ของแบบทดสอบเป้าหมายมากกว่าวิธีอื่น ๆ จึงได้ศึกษาเพื่อเปรียบเทียบค่าเฉลี่ยและการแจกแจงคะแนนสะสมของคะแนนแปลงกับคะแนนจากแบบทดสอบเป้าหมายต่อไป

3.2 ผลการเปรียบเทียบความแตกต่างของค่าเฉลี่ยระหว่างคะแนนแปลงกับคะแนนจากแบบทดสอบฉบับเป้าหมาย

การเปรียบเทียบค่าเฉลี่ยระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนจากแบบทดสอบเป้าหมาย ทดสอบด้วยสถิติทดสอบ t แบบ Independent ผลที่ได้ปรากฏตามตาราง 31.

ตาราง 31 ผลการเปรียบเทียบค่าเฉลี่ยของคะแนนจากแบบทดสอบฉบับ X กับคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี

ฉบับ - วิธีการ	ผลต่างของค่าเฉลี่ย	สถิติทดสอบ t	df	p
X - LIN	0.311	0.320	2310	0.749
X - EQU	0.102	0.106	2310	0.915
X - TMS	-1.431	-1.469	2303	0.142
X - TTRF	-1.321	-1.353	2302	0.176
X - TS	-1.328	-1.355	2300	0.176
X - OMS	-1.405	-1.497	2310	0.135
X - OTRF	-1.512	-1.614	2310	0.107
X - OS	0.159	0.168	2310	0.867

$p > 0.01$

ผลการวิเคราะห์ตามตาราง 31 พบว่า ค่าเฉลี่ยระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนจากแบบทดสอบเป้าหมายไม่แตกต่างกัน

3.3 ผลการเปรียบเทียบความแตกต่างของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงกับคะแนนจากแบบทดสอบฉบับเป้าหมาย

การวิเคราะห์ความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี กับคะแนนจากแบบทดสอบเป้าหมาย โดยใช้สถิติทดสอบนันทราเมตริกซ์ Two – Sample Kolmogorov – Samirnov ปรากฏผลดังตาราง 32

ตาราง 32 ผลการวิเคราะห์ความแตกต่างของการแจกแจงคะแนนสะสมระหว่างคะแนนจากแบบทดสอบฉบับ X กับคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี

วิธีการ เทียบคะแนน	ความแตกต่าง สูงสุดสัมบูรณ์	สถิติทดสอบ Z	p
LIN	0.050	1.204	0.110
EQU	0.034	0.823	0.507
TMS	0.047	1.134	0.153
TTRF	0.047	1.134	0.153
TS	0.048	1.155	0.139
OMS	0.063	1.505	0.022
OTRF	0.064	1.547	0.017
OS	0.033	0.793	0.555

$p > 0.01$

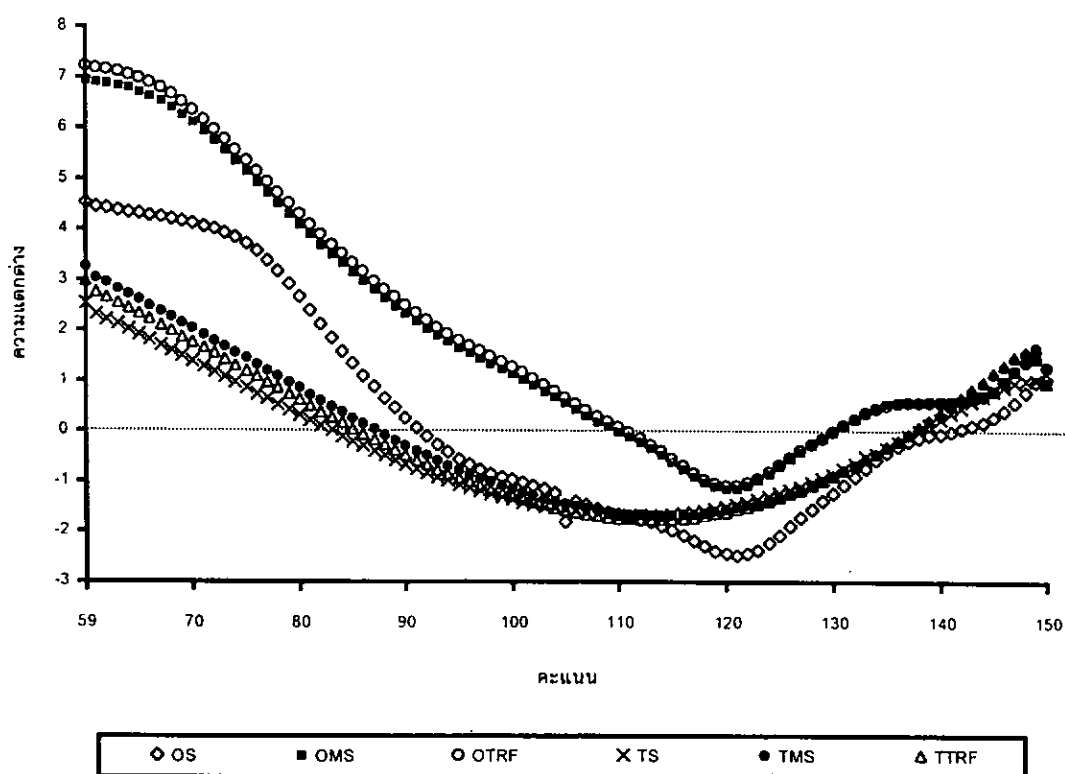
ผลการวิเคราะห์ตามตาราง 32 พบว่า การแจกแจงคะแนนสะสมของคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนจากแบบทดสอบเป้าหมาย มีความคล้ายคลึงกันแสดงว่าการเทียบคะแนนทุกวิธีมีคุณสมบัติการเทียบคะแนนสังเกตและการเทียบคะแนนจริง

ตอนที่ 4 ผลการวิเคราะห์ความแตกต่างระหว่างคะแนนแปลง

นำคะแนนแปลงจากการเทียบคะแนนที่แตกต่างกัน 2 วิธี มาตรวจสอบความแตกต่าง ซึ่งสามารถใช้คะแนนความแตกต่างที่คำนวณโดยการลบคะแนนแปลงจากการเทียบคะแนนพื้นฐาน (ในการศึกษานี้ก็คือวิธีการเทียบคะแนนเชิงเส้นตรงและอีควิเปอร์เซ็นไทล์) ออกจากคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบโมเดลเกรต เรสพอนส์ เป็นการเสนอให้เห็นความแตกต่างในแต่ละระดับคะแนน ส่วนผลรวมของความแตกต่างวิเคราะห์ด้วยดัชนี 6 ดัชนี ซึ่งจำแนกเป็น 2 กลุ่ม คือ กลุ่มดัชนีถ่วงน้ำหนักด้วยความถี่ของคะแนนดิบที่สมมูลกันกับกลุ่มดัชนีไม่ได้ถ่วงน้ำหนัก ดัชนีที่ใช้วิเคราะห์คะแนนความแตกต่างได้แก่ ดัชนีรากที่สองของค่าเฉลี่ยกำลังสองของความแตกต่าง ($RMSD_w$, $RMSD_u$) ดัชนีค่าเฉลี่ยของความแตกต่างสัมบูรณ์ (MAD_w , MAD_u) และดัชนีค่าเฉลี่ยของความแตกต่างที่คำนึงถึงเครื่องหมาย (MSD_w , MSD_u) ดัชนีทั้ง 6 ดัชนี คำนวณด้วยสูตร 94 – 99 ในหน้า 130 – 131 โดยใช้คะแนนแปลงจากการเทียบคะแนนตามวิธีการในทฤษฎีการทดสอบแบบมาตรฐานเดิมเป็นเกณฑ์ ปรากฏผลตามตาราง 33-45 และภาพประกอบ 12-16 ดังรายละเอียดต่อไปนี้

4.1 ผลการวิเคราะห์คะแนนความแตกต่าง

การวิเคราะห์คะแนนความแตกต่าง คำนวณโดยใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง หรือการเทียบคะแนนอีควิเปอร์เซ็นไทล์หักออกจากคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ 3 วิธี 6 รูปแบบ ผลวิเคราะห์ตามภาพประกอบ 12 และ 13

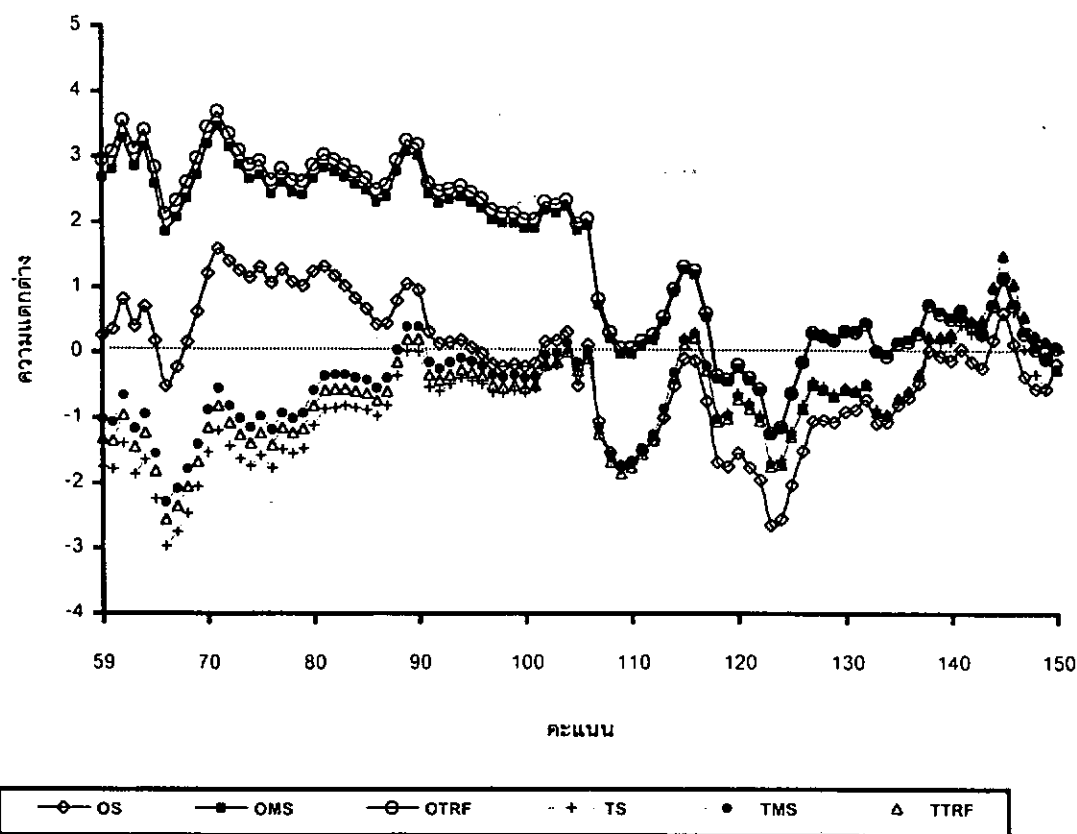


ภาพประกอบ 12 กะแนบความแตกต่างเมื่อใช้กะแนบแปลงจากการเทียบกะแนบเชิงเส้นตรงเป็นเกณฑ์

ผลการวิเคราะห์ตามภาพประกอบ 12 แสดงให้เห็นว่า กะแนบความแตกต่างระหว่างกะแนบแปลงจากการเทียบกะแนบตามทฤษฎีการตอบสนองแต่ละวิธีกับกะแนบแปลงจากการเทียบกะแนบเชิงเส้นตรง เมื่อโค้งของพล็อตใกล้กับเส้นพื้นฐาน (จุด 0 บนแกนตั้ง) หมายถึงฟังก์ชันการเทียบกะแนบวิธีนั้น คล้ายคลึงกับฟังก์ชันการเทียบกะแนบเชิงเส้นตรงในช่วงกะแนบดังกล่าวมากที่สุด ดังนั้นเมื่อพิจารณาวิธีการเทียบกะแนบสังเกต 3 วิธี พบว่า ช่วงกะแนบ 59 – 101 (เป็นช่วงกะแนบดิบที่ต่ำกว่าค่าเฉลี่ยของแบบทดสอบฉบับ Y) กะแนบแปลงจากการเทียบกะแนบสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกันมีความคล้ายคลึงกับกะแนบแปลงจากการเทียบกะแนบเชิงเส้นตรงมากที่สุด ช่วงกะแนบ 102 – 110 กะแนบแปลงจากการเทียบกะแนบสังเกตวิธี MS มีความคล้ายคลึงกับกะแนบแปลงจากการเทียบกะแนบเชิงเส้นตรงมากที่สุด ช่วงกะแนบ 111 – 134 กะแนบแปลงจากการเทียบกะแนบสังเกตวิธี TRF มีความคล้ายคลึงกับกะแนบแปลงจากการเทียบกะแนบเชิงเส้นตรงมากที่สุด ช่วงกะแนบ 135 – 150 กะแนบแปลงจากการเทียบกะแนบสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน มีความคล้ายคลึงกับกะแนบแปลงจากการเทียบกะแนบเชิงเส้นตรงมากที่สุด

เมื่อพิจารณาการเทียบคะแนนจริง 3 วิธี พบว่า ช่วงคะแนน 59 – 84 คะแนนแปลงจากการเทียบคะแนนจริงวิธีปรับค่าพารามิเตอร์พร้อมกัน มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงมากที่สุด ช่วงคะแนน 85 – 114 ส่วนใหญ่คะแนนแปลงจากการเทียบคะแนนจริงวิธี MS มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงมากที่สุด ช่วงคะแนน 115 – 149 ส่วนใหญ่คะแนนแปลงจากการเทียบคะแนนจริงวิธีปรับค่าพารามิเตอร์พร้อมกัน มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงมากที่สุดและที่คะแนน 150 คะแนนแปลงจากการเทียบคะแนนจริงทั้ง 3 วิธีมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง

โดยภาพรวม สรุปได้ว่า ฟังก์ชันการเทียบคะแนนจริงทุกวิธี ให้ความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนเชิงเส้นตรงมากกว่าฟังก์ชันการเทียบคะแนนสังเกต และฟังก์ชันการเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกัน มีความคล้ายคลึงกับ ฟังก์ชันการเทียบคะแนนเชิงเส้นตรง มากกว่าวิธี MS และ TRF



ภาพประกอบ 13 คะแนนความแตกต่างเมื่อใช้คะแนนแปลงจากการเทียบคะแนน
อีควิเปอร์เซ็นไทล์เป็นเกณฑ์

ผลการวิเคราะห์ตามภาพประกอบ 13 แสดงให้เห็นว่า คะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบแต่ละวิธีกับคะแนนแปลงจากการเทียบคะแนนวิธีอีควิเปอร์เซ็นไทล์ เมื่อโค้งของพล็อตใกล้กับเส้นพื้นฐาน (จุด 0 บนแกนตั้ง) หมายถึง ฟังก์ชันการเทียบคะแนนวิธีนั้น คล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นไทล์ในช่วงคะแนนดังกล่าวมากที่สุด ดังนั้นเมื่อพิจารณาการเทียบคะแนนสังเกต 3 วิธี พบว่า ในช่วงคะแนน 59 - 106 คะแนนแปลงจากการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกันมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด ช่วงคะแนน 107 - 113 คะแนนแปลงจากการเทียบคะแนนสังเกตวิธี MS มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด ช่วงคะแนน 114 - 116 คะแนนแปลงจากการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกันมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด และช่วงคะแนน 117 - 150 ส่วนใหญ่คะแนนแปลงจากการเทียบคะแนนสังเกตวิธี TRF มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด โดยมีคะแนนแปลงจากวิธี MS สลับบางช่วงคะแนน

และเมื่อพิจารณาการเทียบคะแนนจริง 3 วิธี พบว่า ในช่วงคะแนน 59 - 116 ส่วนใหญ่คะแนนแปลงจากการเทียบคะแนนจริงวิธี MS มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด ช่วงคะแนน 117 - 133 คะแนนแปลงจากการเทียบคะแนนจริงวิธีปรับค่าพารามิเตอร์พร้อมกันมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด ช่วงคะแนน 134 - 149 ส่วนใหญ่คะแนนแปลงจากการเทียบคะแนนจริงวิธีปรับค่าพารามิเตอร์พร้อมกันมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด โดยมีคะแนนแปลงจากวิธี MS และ TRF สลับบางช่วงคะแนน และที่ระดับคะแนน 150 คะแนนแปลงจากการเทียบคะแนนจริงทั้ง 3 วิธีมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์

โดยภาพรวม สรุปได้ว่า ในการเปรียบเทียบความคล้ายคลึงกันระหว่างคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ ให้ผลที่ไม่ชัดเจนว่าการเทียบคะแนนวิธีใดมีความคล้ายคลึงกันมากที่สุด ทั้งการเทียบคะแนนจริงและคะแนนสังเกต เพราะในแต่ละช่วงคะแนนจะสลับกันให้ผลที่คล้ายคลึงกัน ดังนั้น จึงวิเคราะห์คะแนนความแตกต่างด้วยดัชนีเพื่อสรุปผลรวมของความแตกต่างต่อไป

4.2 ผลการวิเคราะห์ความแตกต่างของคะแนนแปลงด้วยดัชนี

เพื่อให้เห็นความแตกต่างรวมทุกระดับคะแนนระหว่างคะแนนแปลงจากการเทียบคะแนนที่แตกต่างกัน 2 วิธี จึงวิเคราะห์ผลด้วยดัชนี 6 ดัชนี ดังกล่าวข้างต้น ปรากฏผลจำแนกตามดัชนีต่าง ๆ ดังต่อไปนี้

4.2.1 ผลการวิเคราะห์ความแตกต่างด้วยดัชนี RMSD

4.2.1.1 ผลการวิเคราะห์ความแตกต่างด้วยดัชนี RMSD

ถ่วงน้ำหนัก

การวิเคราะห์ความแตกต่างของคะแนนแปลงด้วยดัชนี RMSDถ่วงน้ำหนัก ($RMSD_w$) ด้วยความถี่ของคะแนนดิบในแบบทดสอบฉบับ Y ผลการคำนวณจำแนกตามคะแนนเกณฑ์คือเมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงเป็นเกณฑ์ ปรากฏผลตามตาราง 33 และเมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ เป็นเกณฑ์ ปรากฏผลตามตาราง 34

ตาราง 33 ค่าดัชนี $RMSD_w$ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum f_i(A_i - B_1)^2$	$\sum f_i$	$RMSD_w$	ลำดับที่
TMS – LIN	1535.901	1157	1.152	3
TTRF – LIN	1531.691	1157	1.151	2
TS – LIN	1301.576	1157	1.061	1
OMS – LIN	7976.312	1157	2.626	5
OTRF – LIN	8761.441	1157	2.752	6
OS – LIN	4123.717	1157	1.888	4

ผลการวิเคราะห์ตามตาราง 33 พบว่า ค่าดัชนี $RMSD_w$ ระหว่างคะแนนแปลงจากการเทียบคะแนนจริง วิธีปรับค่าพารามิเตอร์พร้อมกันกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (TS – LIN) มีค่าน้อยที่สุด เท่ากับ 1.061 และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (OTRF – LIN) มีค่ามากที่สุดเท่ากับ 2.752 และโดยภาพรวมคะแนนแปลงจากการเทียบคะแนนจริงมีความคล้อยลึงกับคะแนนแปลงจากเทียบคะแนนเชิงเส้นตรงมากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตทุกวิธี

ตาราง 34 ค่าดัชนี $RMSD_w$ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ (B_2) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum f_i (A_i - B_2)^2$	$\sum f_i$	$RMSD_w$	ลำดับที่
TMS - EQU	670.625	1157	0.761	1
TTRF - EQU	856.656	1157	0.860	2
TS - EQU	1041.654	1157	0.949	3
OMS - EQU	4253.930	1157	1.917	5
OTRF - EQU	4869.303	1157	2.051	6
OS - EQU	1085.090	1157	0.968	4

ผลการวิเคราะห์ตามตาราง 34 พบว่า ค่าดัชนี $RMSD_w$ ระหว่างคะแนนแปลงจากการเทียบคะแนนจริงวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก กับคะแนนแปลงจากการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ (TMS - EQU) มีค่าน้อยที่สุดเท่ากับ 0.761 และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ (OTRF - EQU) มีค่ามากที่สุดเท่ากับ 2.051 และโดยภาพรวม คะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึง กับคะแนนแปลงจากการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ มากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตทุกวิธี

และเมื่อพิจารณาดัชนี $RMSD_w$ ที่คะแนนเกณฑ์แตกต่างกัน พบว่า คะแนนแปลงจากการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนตามทฤษฎี IRT ทั้งการเทียบคะแนนจริงและคะแนนสังเกต มากกว่าการเทียบคะแนนเชิงเส้นตรง ถึงแม้ลำดับที่ของความแตกต่างของคะแนนแปลงระหว่างการเทียบคะแนนจริง 3 วิธี จะมีความไม่สอดคล้องกัน ขณะที่ความแตกต่างของคะแนนแปลงจากการเทียบคะแนนสังเกตมีลำดับที่ของความแตกต่างลำดับเดียวกันไม่ว่าคะแนนเกณฑ์จะเป็นคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงหรืออีคิวเปอร์เซ็นต์ไทล์ก็ตาม

4.2.1.2 ผลการวิเคราะห์ความแตกต่างด้วยดัชนี RMSD ไม่ถ่วงน้ำหนัก

การวิเคราะห์ความแตกต่างของคะแนนแปลงด้วยดัชนี RMSD ไม่ถ่วงน้ำหนัก ($RMSD_u$) เป็นการตรวจสอบความแตกต่างในแต่ละระดับคะแนน ตลอดช่วงคะแนนต่ำสุด - สูงสุด (59 - 150 คะแนน) โดยไม่คำนึงว่าระดับคะแนนใดมีผู้สอบได้คะแนนดังกล่าวก็คน ผลการคำนวณจำแนกตามคะแนนเกณฑ์ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงเป็นเกณฑ์ ปรากฏผลตามตาราง 35 และเมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์เป็นเกณฑ์ ปรากฏผลตามตาราง 36

ตาราง 35 ค่าดัชนี $RMSD_u$ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum_i (A_i - B_1)^2$	k	$RMSD_u$	ลำดับที่
TMS - LIN	195.657	92	1.458	3
TTRF - LIN	181.966	92	1.406	2
TS - LIN	143.483	92	1.249	1
OMS - LIN	947.809	92	3.210	5
OTRF - LIN	1029.674	92	3.345	6
OS - LIN	485.996	92	2.298	4

ผลการวิเคราะห์ตามตาราง 35 พบว่า ค่าดัชนี RMSD ระหว่างคะแนนแปลงจากการเทียบคะแนนจริง วิธีปรับค่าพารามิเตอร์พร้อมกัน กับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (TS - LIN) มีค่าน้อยที่สุดเท่ากับ 1.249 และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงมีค่ามากที่สุดเท่ากับ 3.345 และโดยภาพรวมคะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงมากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตทุกวิธี

ตาราง 36 ค่าดัชนี $RMSD_u$ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum (A_i - B_2)^2$	K	$RMSD_u$	ลำดับที่
TMS – EQU	69.247	92	0.868	1
TTRF – EQU	89.061	92	0.984	3
TS – EQU	113.164	92	1.109	4
OMS – EQU	333.121	92	1.903	5
OTRF – EQU	384.179	92	2.043	6
OS – EQU	88.657	92	0.982	2

ผลการวิเคราะห์ตามตาราง 36 พบว่า ค่าดัชนี $RMSD_u$ ระหว่างคะแนนแปลงจากการเทียบคะแนนจริง วิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก กับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (TMS – EQU) มีค่าน้อยที่สุดเท่ากับ 0.868 และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (OTRF – EQU) มีค่ามากที่สุดเท่ากับ 2.043 และโดยภาพรวม คะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ มากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตบางวิธี

และเมื่อพิจารณาดัชนี $RMSD_u$ ที่คะแนนเกณฑ์แตกต่างกัน พบว่า คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนตามทฤษฎี IRT ทั้งการเทียบคะแนนจริงและคะแนนสังเกตมากกว่าการเทียบคะแนนเชิงเส้นตรง ถึงแม้ลำดับที่ของความแตกต่าง ระหว่างคะแนนแปลงจากการเทียบคะแนนจริงทั้ง 3 วิธี และคะแนนสังเกต 1 วิธีจะมีความไม่สอดคล้องกัน

4.2.2 ผลการวิเคราะห์ความแตกต่างด้วยดัชนี MAD

4.2.2.1 ผลการวิเคราะห์ความแตกต่างด้วยดัชนี MAD

ถ่วงน้ำหนัก

การวิเคราะห์ความแตกต่างของคะแนนแปลงด้วยดัชนี MAD ถ่วงน้ำหนัก (MAD_w) ด้วยความถี่ของคะแนนดิบในแบบทดสอบฉบับ Y ผลการคำนวณ จำแนกตามคะแนนเกณฑ์ คือ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงเป็นเกณฑ์ ปรากฏผลตามตาราง 37 และเมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์เป็น เกณฑ์ ปรากฏผลตามตาราง 38

ตาราง 37 ค่าดัชนี MAD_w เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum f_i A_i - B_1 $	$\sum f_i$	MAD_w	ลำดับที่
TMS – LIN	1154.108	1157	0.998	2
TTRF – LIN	1157.598	1157	1.001	3
TS – LIN	1066.411	1157	0.922	1
OMS – LIN	2270.069	1157	1.962	5
OTRF – LIN	2377.999	1157	2.055	6
OS – LIN	1722.260	1157	1.489	4

ผลการวิเคราะห์ตามตาราง 37 พบว่า ค่าดัชนี MAD_w ระหว่างคะแนนแปลงจากการเทียบคะแนนจริง วิธีปรับค่าพารามิเตอร์พร้อมกัน กับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (TS – LIN) มีค่าน้อยที่สุดเท่ากับ 0.922 และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (OTRF – LIN) มีค่ามากที่สุดเท่ากับ 2.055 และโดยภาพรวมคะแนนแปลงจากการเทียบคะแนนจริง มีความคล้ายคลึงกับ คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง มากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตทุกวิธี

ตาราง 38 ค่าดัชนี MAD_w เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum_i f_i A_i - B_2 $	$\sum f_i$	MAD_w	ลำดับที่
TMS – EQU	707.536	1157	0.612	1
TTRF – EQU	830.164	1157	0.718	2
TS – EQU	903.976	1157	0.781	4
OMS – EQU	1855.362	1157	1.604	5
OTRF – EQU	1973.044	1157	1.705	6
OS – EQU	893.46	1157	0.772	3

ผลการวิเคราะห์ตามตาราง 38 พบว่า ค่าดัชนี MAD_w ระหว่างคะแนนแปลงจากการเทียบคะแนนจริงวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (TMS – EQU) มีค่าน้อยที่สุดเท่ากับ 0.612 และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (OTRF – EQU) มีค่ามากที่สุด เท่ากับ 1.705 และโดยภาพรวม คะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึง กับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตบางวิธี

เมื่อพิจารณาดัชนี MAD_w ที่คะแนนเกณฑ์แตกต่างกัน พบว่า คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนตามทฤษฎี IRT ทั้งคะแนนจริงและคะแนนสังเกต มากกว่าคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง ถึงแม้ลำดับที่ของความแตกต่างระหว่างการเทียบคะแนนจริง ทั้ง 3 วิธี และคะแนนสังเกต 1 วิธี จะมีความไม่สอดคล้องกัน

4.2.2.2 ผลการวิเคราะห์ความแตกต่างด้วยดัชนี MAD

ไม่ถ่วงน้ำหนัก

การวิเคราะห์ความแตกต่างของคะแนนแปลงด้วยดัชนี MAD ไม่ถ่วงน้ำหนัก (MAD_0) เป็นการตรวจสอบความแตกต่างในแต่ละระดับคะแนน ตลอดช่วงคะแนนต่ำสุด-สูงสุด (59-150 คะแนน) ผลการคำนวณ จำแนกตามคะแนนเกณฑ์ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงเป็นเกณฑ์ ปรากฏผลตามตาราง 39 และเมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์เป็นเกณฑ์ ปรากฏผลตามตาราง 40

ตาราง 39 ค่าดัชนี MAD_0 เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum_i A_i - B_i $	K	MAD_0	ลำดับที่
TMS – LIN	115.245	92	1.253	3
TTRF – LIN	113.057	92	1.229	2
TS – LIN	100.924	92	1.097	1
OMS – LIN	210.682	92	2.290	5
OTRF – LIN	219.232	92	2.383	6
OS – LIN	168.356	92	1.830	4

ผลการวิเคราะห์ตามตาราง 39 พบว่า ค่าดัชนี MAD_0 ระหว่างคะแนนแปลงจากการเทียบคะแนนจริง วิธีปรับค่าพารามิเตอร์พร้อมกันกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (TS – LIN) มีค่าน้อยที่สุดเท่ากับ 1.097 และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (OTRF – LIN) มีค่ามากที่สุดเท่ากับ 2.383 และโดยภาพรวมคะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงมากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตทุกวิธี

ตาราง 40 ค่าดัชนี MAD_u เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum A_i - B_2 $	k	MAD_u	ลำดับที่
TMS – EQU	64.772	92	0.704	1
TTRF – EQU	75.190	92	0.817	3
TS – EQU	83.024	92	0.902	4
OMS – EQU	142.676	92	1.551	5
OTRF – EQU	152.162	92	1.654	6
OS – EQU	71.596	92	0.778	2

ผลการวิเคราะห์ตามตาราง 40 พบว่า ค่าดัชนี MAD_u ระหว่างคะแนนแปลงจากการเทียบคะแนนจริง วิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากกับคะแนนแปลงจากวิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (TMS – EQU) มีค่าน้อยที่สุดเท่ากับ 0.704 และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบ (OTRF – EQU) มีค่ามากที่สุดเท่ากับ 1.654 และโดยภาพรวม คะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตบางวิธี

เมื่อพิจารณาดัชนี MAD_u ที่คะแนนเกณฑ์แตกต่างกัน พบว่า คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนตามทฤษฎี IRT ทั้งการเทียบคะแนนจริงและคะแนนสังเกต มากกว่าคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง ถึงแม้ลำดับที่ของความแตกต่างระหว่างการเทียบคะแนนจริง ทั้ง 3 วิธี และคะแนนสังเกต 1 วิธี จะมีความไม่สอดคล้องกัน

4.2.3 ผลการวิเคราะห์ความแตกต่างด้วยดัชนี MSD

4.2.3.1 ผลการวิเคราะห์ความแตกต่างด้วยดัชนี MSD

ถ่วงน้ำหนัก

การวิเคราะห์ความแตกต่างของคะแนนแปลงด้วยดัชนี MSD ถ่วงน้ำหนัก (MSD_w) ด้วยความถี่ของคะแนนดิบในแบบทดสอบฉบับ Y ผลการคำนวณจำแนกตามคะแนนเกณฑ์ คือ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงเป็นเกณฑ์ ปรากฏผลตามตาราง 41 และเมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์เป็นเกณฑ์ ปรากฏผลตามตาราง 42

ตาราง 41 ค่าดัชนี MSD_w เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum f_i(A_i - B_1)$	$\sum f_i$	MSD_w	ลำดับที่
TMS – LIN	-238.480	1157	-0.206	2
TTRF – LIN	-374.630	1157	-0.324	3
TS – LIN	-536.680	1157	-0.464	4
OMS – LIN	1985.260	1157	1.716	5
OTRF – LIN	2109.110	1157	1.823	6
OS – LIN	175.750	1157	0.152	1

ผลการวิเคราะห์ตามตาราง 41 พบว่า ค่าดัชนี MSD_w ระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีปรับค่าพารามิเตอร์พร้อมกัน กับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (OS – LIN) มีค่าน้อยที่สุดเท่ากับ 0.152 แสดงให้เห็นว่าคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน ส่วนใหญ่มีค่าสูงกว่าคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (OTRF – LIN) มีค่ามากที่สุดเท่ากับ 1.823 โดยภาพรวมคะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงมากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตบางวิธี

ตาราง 42 ค่าดัชนี MSD_w เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum f_i(A_i - B_2)$	$\sum f_i$	MSD_w	ลำดับที่
TMS – EQU	-480.320	1157	-0.415	2
TTRF – EQU	-616.470	1157	-0.533	3
TS – EQU	-778.520	1157	-0.673	4
OMS – EQU	1743.420	1157	1.507	5
OTRF – EQU	1867.270	1157	1.614	6
OS – EQU	-66.090	1157	-0.057	1

ผลการวิเคราะห์ตามตาราง 42 พบว่า ค่าดัชนี MSD_w ระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีปรับค่าพารามิเตอร์พร้อมกัน กับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (OS – EQU) มีค่าน้อยที่สุดเท่ากับ -0.057 แสดงให้เห็นว่าคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีปรับค่าพารามิเตอร์พร้อมกัน ส่วนใหญ่มีค่าต่ำกว่าคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบ กับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (OTRF – EQU) มีค่ามากที่สุดเท่ากับ 1.614 โดยภาพรวมคะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ มากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตบางวิธี

เมื่อพิจารณาดัชนี MSD_w ที่คะแนนเกณฑ์แตกต่างกัน พบว่า คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนจริงมากกว่าคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ และคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนสังเกตมากกว่าคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง

4.2.3.2 ผลการวิเคราะห์ความแตกต่างด้วยดัชนี MSD

ไม่ถ่วงน้ำหนัก

การวิเคราะห์ความแตกต่างของคะแนนแปลงด้วยดัชนี MSD ไม่ถ่วงน้ำหนัก (MSD_0) เป็นการตรวจสอบความแตกต่างที่ได้ในแต่ละระดับคะแนน ตลอดช่วงคะแนนต่ำสุด-สูงสุด (59 – 150 คะแนน) ผลการคำนวณ จำแนกตามคะแนนเกณฑ์ เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงเป็นเกณฑ์ ปรากฏผลตามตาราง 43 และเมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีคิวเปอร์เซ็นไทล์เป็นเกณฑ์ ปรากฏผลตามตาราง 44

ตาราง 43 ค่าดัชนี MSD_0 เมื่อใช้คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (B_1) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum_i (A_i - B_i)$	k	MSD_0	ลำดับที่
TMS – LIN	5.571	92	0.061	1
TTRF – LIN	-5.479	92	-0.060	1
TS – LIN	-19.839	92	-0.216	3
OMS – LIN	185.681	92	2.017	5
OTRF – LIN	195.711	92	-2.127	6
OS – LIN	41.171	92	0.448	4

ผลการวิเคราะห์ตามตาราง 43 พบว่า ค่าดัชนี MSD_0 ระหว่างคะแนนแปลงจากการเทียบคะแนนจริง วิธีโค้งลักษณะแบบทดสอบและวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (TTRF – LIN, TMS – LIN) มีค่าน้อยที่สุดเท่ากับ -0.060 และ 0.061 ตามลำดับ แสดงให้เห็นว่าคะแนนแปลงจากการเทียบคะแนนจริง วิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก ส่วนใหญ่มีค่าสูงกว่าหรือเกือบเท่ากับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง และคะแนนแปลงจากการเทียบคะแนนจริง วิธีโค้งลักษณะแบบทดสอบส่วนใหญ่มีค่าต่ำกว่าหรือเกือบเท่ากับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง นอกจากนี้ความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีโค้งลักษณะแบบทดสอบ กับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง (OTRF – LIN) มีค่ามากที่สุดเท่ากับ 2.127 โดยภาพรวมคะแนนแปลงที่ได้จากการเทียบคะแนนจริงมีความคล้ายคลึงกับ คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง มากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตทุกวิธี

ตาราง 44 ค่าดัชนี MSD_u เมื่อใช้คะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (B_2) เป็นเกณฑ์

วิธีการเทียบคะแนน	$\sum_i (A_i - B_2)$	k	MSD_u	ลำดับที่
TMS – EQU	-47.776	92	-0.519	2
TTRF – EQU	-58.826	92	-0.639	3
TS – EQU	-73.186	92	-0.796	4
OMS – EQU	132.334	92	1.438	5
OTRF – EQU	142.364	92	1.547	6
OS – EQU	-12.176	92	-0.132	1

ผลการวิเคราะห์ตามตาราง 44 พบว่า ค่าดัชนี MSD_u ระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีปรับค่าพารามิเตอร์พร้อมกัน กับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (OS – EQU) มีค่าน้อยที่สุดเท่ากับ -0.132 แสดงให้เห็นว่าคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีวิเคราะห์ค่าพารามิเตอร์พร้อมกัน ส่วนใหญ่มีค่าต่ำกว่าคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ และความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีโค้งลักษณะแบบทดสอบ กับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ (OTRF – EQU) มีค่ามากที่สุดเท่ากับ 1.547 โดยภาพรวมคะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากกว่าคะแนนแปลงจากการเทียบคะแนนสังเกตบางวิธี

เมื่อพิจารณาดัชนี MSD_u ที่คะแนนเกณฑ์แตกต่างกัน พบว่า คะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนจริงมากกว่าคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ และคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนสังเกตมากกว่าคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง

4.3 สรุปผลความแตกต่างของการเทียบคะแนนในแต่ละวิธี

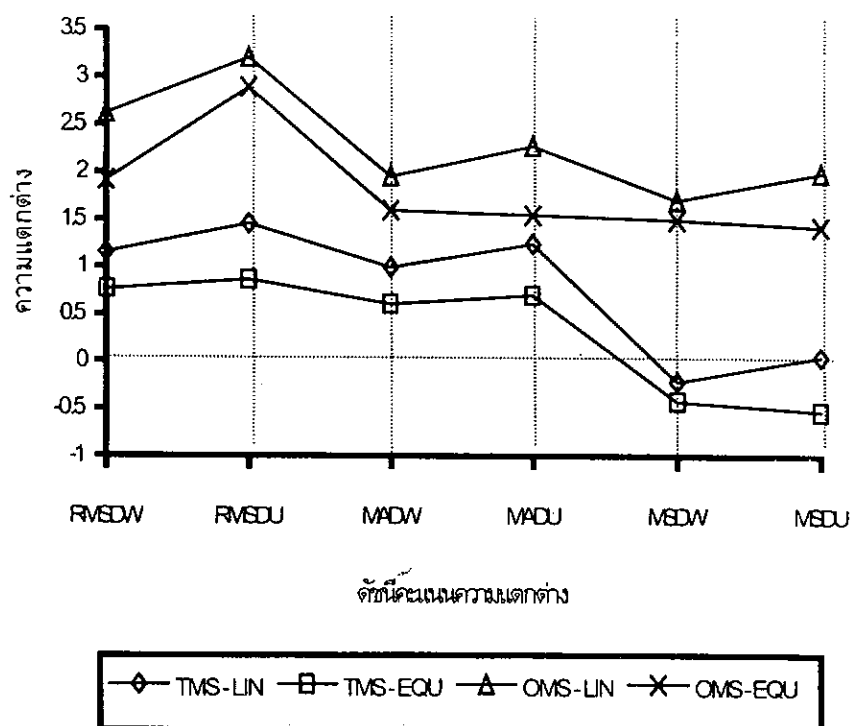
การวิเคราะห์ความแตกต่างของคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงและการเทียบคะแนนอีควิเปอร์เซ็นไทล์ด้วยดัชนีทั้ง 6 ดัชนี สามารถนำมาสรุปผลรวม ปรากฏผลตามตาราง 45

ตาราง 45 ค่าดัชนี $RMSD_w$ $RMSD_u$ MAD_w MAD_u MSD_w และ MSD_u ของคะแนนแปลงจากการเทียบคะแนนที่แตกต่างกัน 2 วิธี

วิธีการเทียบ คะแนน	ดัชนี					
	$RMSD_w$	$RMSD_u$	MAD_w	MAD_u	MSD_w	MSD_u
TMS – LIN	1.152	1.458	0.998	1.253	-0.206	0.061
TMS – EQU	0.761	0.868	0.612	0.704	-0.415	-0.519
TTRF – LIN	1.151	1.406	1.001	1.229	-0.324	-0.060
TTRF – EQU	0.860	0.984	0.718	0.817	-0.533	-0.639
TS – LIN	1.061	1.249	0.922	1.097	-0.464	-0.216
TS – EQU	0.949	1.109	0.781	0.902	-0.673	-0.796
OMS – LIN	2.626	3.210	1.962	2.290	1.716	2.017
OMS – EQU	1.917	1.903	1.604	1.551	1.507	1.438
OTRF – LIN	2.752	3.345	2.055	2.383	1.823	2.127
OTRF – EQU	2.051	2.043	1.705	1.654	1.614	1.547
OS – LIN	1.888	2.298	1.489	1.830	0.152	0.448
OS – EQU	0.968	0.982	0.772	0.778	-0.057	-0.132

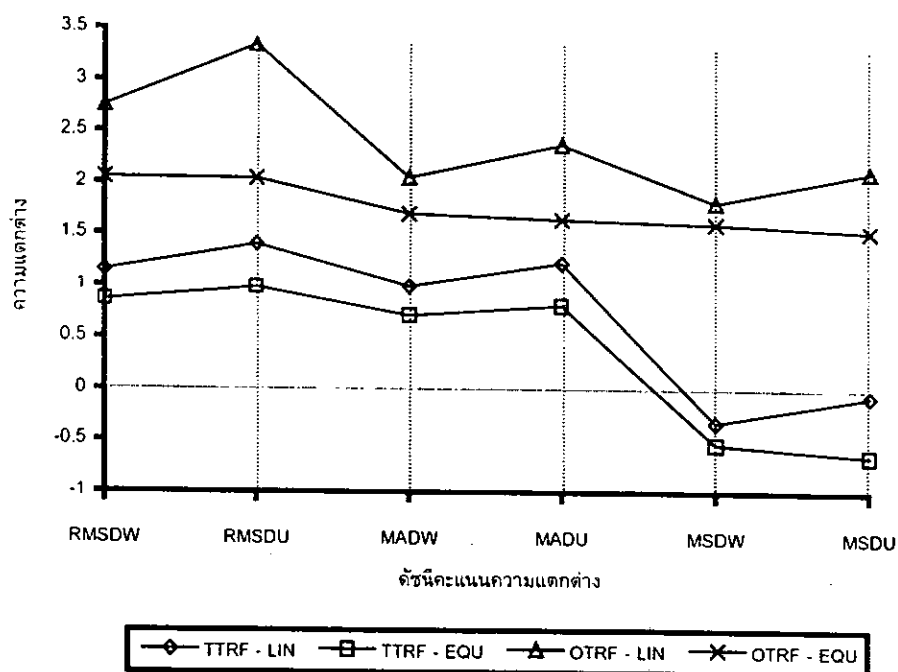
ผลการวิเคราะห์ตามตาราง 45 พบว่า ความแตกต่างของคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง มีค่าความแตกต่างสูงสุดทั้ง 6 ดัชนี ขณะที่คะแนนแปลงจากการเทียบคะแนนจริงมีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงและอีควิเปอร์เซ็นไทล์ มากกว่าการเทียบคะแนนสังเกต โดยคะแนนแปลงจากการเทียบคะแนนจริง วิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนพื้นฐานมากที่สุด โดยเฉพาะการเทียบคะแนนอีควิเปอร์เซ็นไทล์ ในดัชนี $RMSD_w$ $RMSD_u$ MAD_w MAD_u และ MSD_u และคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน มีความคล้ายคลึงกับคะแนนแปลงจากการเทียบคะแนนพื้นฐานมากที่สุด โดยเฉพาะการเทียบคะแนนอีควิเปอร์เซ็นไทล์ในดัชนี MSD_w

นอกจากนี้ ผู้วิจัยขอเสนอกราฟที่แสดงความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบแต่ละวิธีกับคะแนนแปลงจากการเทียบคะแนนพื้นฐานในแต่ละดัชนี ปรากฏผลตามภาพประกอบ 14 - 16



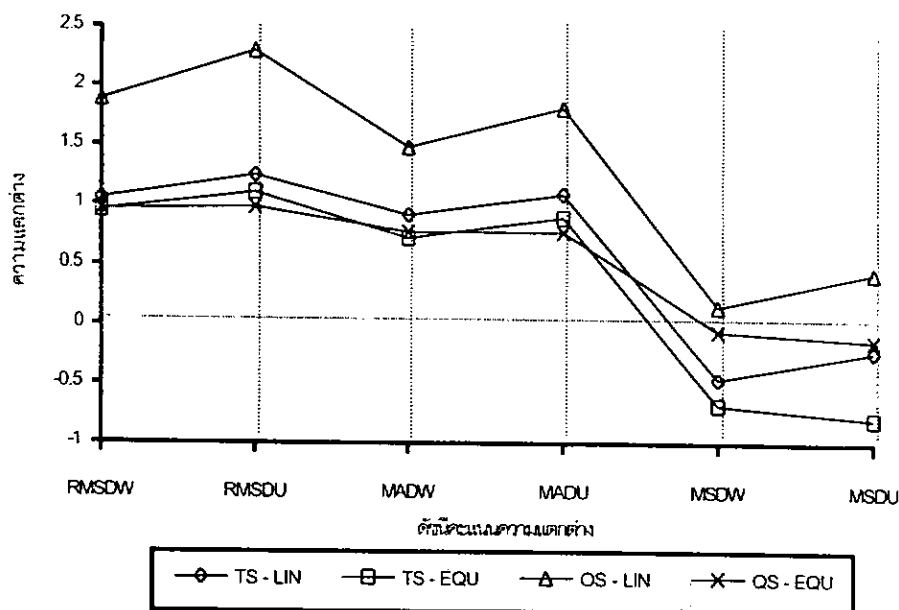
ภาพประกอบ 14 ค่าดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากกับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน

ผลการวิเคราะห์ตามภาพประกอบ 14 แสดงให้เห็น ความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากกับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน ซึ่งพบว่า ฟังก์ชันการเทียบคะแนนจริงมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด ในดัชนี $RMSD_W$, $RMSD_U$, MAD_W และ MAD_U และมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนเชิงเส้นตรงมากที่สุดในดัชนี MSD_W และ MSD_U สำหรับฟังก์ชันการเทียบคะแนนสังเกตแตกต่างจากฟังก์ชันการเทียบคะแนนพื้นฐานมากกว่าฟังก์ชันการเทียบคะแนนจริง



ภาพประกอบ 15 ค่าดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนน
วิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน

ผลการวิเคราะห์ตามภาพประกอบ 15 แสดงให้เห็น ความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนวิธีโค้งลักษณะแบบทดสอบกับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน ซึ่งพบว่า ฟังก์ชันการเทียบคะแนนจริงมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด ในดัชนี $RMSD_W$, $RMSD_U$, MAD_W และ MAD_U และมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนเชิงเส้นตรงมากที่สุด ในดัชนี MSD_W และ MSD_U สำหรับฟังก์ชันการเทียบคะแนนสังเกตแตกต่างจากฟังก์ชันการเทียบคะแนนพื้นฐานมากกว่าฟังก์ชันการเทียบคะแนนจริง



ภาพประกอบ 16 ค่าดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนน
วิธีปรับค่าพารามิเตอร์พร้อมกันกับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน

ผลการวิเคราะห์ตามภาพประกอบ 16 แสดงให้เห็น ความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนน วิธีปรับค่าพารามิเตอร์พร้อมกัน กับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน ซึ่งพบว่า ฟังก์ชันการเทียบคะแนนสังเกตมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด ในดัชนี $RMSDU$, $MADW$, $MADU$, $MSDW$ และ $MSDU$ และฟังก์ชันการเทียบคะแนนจริงมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด ในดัชนี $RMSDW$ สรุปได้ว่า ฟังก์ชันการเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกันมีความคล้ายคลึงกับ ฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นไทล์ มากกว่าฟังก์ชันการเทียบคะแนนเชิงเส้นตรง

โดยภาพรวม สรุปได้ว่า ฟังก์ชันการเทียบคะแนน วิธี MS และ วิธี TRF ให้ผลที่สอดคล้องกันทุกดัชนี คือ ฟังก์ชันการเทียบคะแนนจริงมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากกว่าฟังก์ชันการเทียบคะแนนสังเกต สำหรับฟังก์ชันการเทียบคะแนนวิธี SIM มีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นไทล์ทุกดัชนี

ตอนที่ 5 ผลการวิเคราะห์กลุ่มสอบทานผล

การวิเคราะห์กลุ่มสอบทานผลเป็นวิธีการหนึ่งที่น่ามาใช้ในการประเมินผลการเทียบคะแนน กลุ่มสอบทานผล สุ่มจากประชากรเดียวกับหน่วยตัวอย่างเทียบคะแนน จำนวน 589 คน ทำแบบทดสอบเทียบคะแนนทั้งสองฉบับ คือ ฉบับ X และ Y นำผลการสอบทั้งสองฉบับไปวิเคราะห์ ค่าสถิติพื้นฐาน ค่าความสัมพันธ์ระหว่างคะแนนแปลงกับคะแนนเกณฑ์ ประเมินค่าดัชนีความแตกต่าง (C) และเปรียบเทียบความแตกต่างของคะแนนแปลงกับคะแนนเกณฑ์ ดังรายละเอียดต่อไปนี้

5.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนจากกลุ่มสอบทานผล

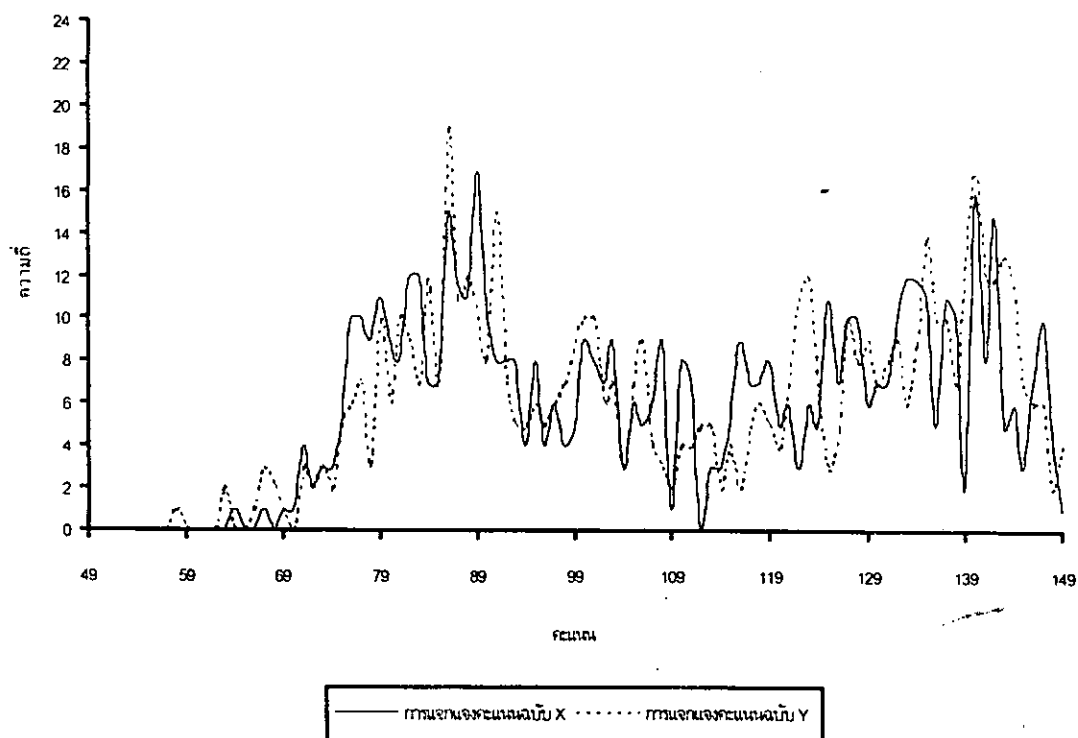
การวิเคราะห์ข้อมูลเบื้องต้นด้วยค่าสถิติพื้นฐาน เพื่ออธิบายคุณลักษณะของคะแนนดิบที่ได้จากการวัดด้วยกลุ่มสอบทานผล จำนวน 589 คน ซึ่งทำแบบทดสอบทั้งฉบับ X และฉบับ Y ซึ่งเป็นแบบทดสอบเทียบคะแนนที่มีคะแนนเต็มฉบับละ 150 คะแนน ปรากฏผลดังตาราง 46

ตาราง 46 ค่าสถิติพื้นฐานของคะแนนที่ได้จากกลุ่มสอบทานผล จำแนกตามแบบทดสอบเทียบคะแนน

ค่าสถิติ	แบบทดสอบ	
	ฉบับ X	ฉบับ Y
ค่าต่ำสุด	64	58
ค่าสูงสุด	149	149
ค่าเฉลี่ย	109.424	111.082
มัธยฐาน	108.000	111.000
ส่วนเบี่ยงเบนมาตรฐาน	23.260	23.696
ความแปรปรวน	541.013	561.493
ความเบ้	0.042	-0.066
ความคลาดเคลื่อนมาตรฐานของค่าความเบ้	0.101	0.101
ความโด่ง	-1.392	-1.374
ความคลาดเคลื่อนมาตรฐานของค่าความโด่ง	0.201	0.201
OPM	0.734	0.746

ผลการวิเคราะห์ตามตาราง 46 พบว่า คะแนนรวมต่ำสุดของแบบทดสอบฉบับ X และ Y คือ 64 และ 58 คะแนน ตามลำดับ แสดงว่าต่ำสุดผู้สอบสามารถตอบข้อสอบได้ถูกต้องอย่างน้อย 2 ใน 5 คะแนน คะแนนรวมสูงสุดของแบบทดสอบทั้งสองฉบับมีค่าเท่ากัน คือ 149 คะแนน สำหรับคะแนนเฉลี่ยและความแปรปรวนของแบบทดสอบฉบับ X ต่ำกว่าฉบับ Y คือ ฉบับ X มีค่าเฉลี่ย 109.424 คะแนน ความแปรปรวน 541.013 คะแนน และฉบับ Y มีค่าเฉลี่ย 111.082 คะแนน ความแปรปรวน 561.493 คะแนน สอดคล้องกับผลที่ได้จากกลุ่มตัวอย่างเทียบคะแนน โดยที่ร้อยละ 50 ของผู้สอบแบบทดสอบฉบับ X ได้คะแนนต่ำกว่า 108 คะแนน ฉบับ Y ได้คะแนนต่ำกว่า 111 คะแนน พิสัยของคะแนนฉบับ X และ Y ใกล้เคียงกัน ($PR_{25} - PR_{75}$ อยู่ในช่วง 88 – 132 และ 89 – 134 คะแนนตามลำดับ) คะแนนของแบบทดสอบฉบับ X และ Y มีค่ามัธยฐานแตกต่างกันเล็กน้อย คือ 108 .000 และ 111 .000 คะแนน ตามลำดับ

เมื่อพิจารณารูปแบบการแจกแจงคะแนนจากค่าความเบ้และความโด่ง พบว่า คะแนนจากแบบทดสอบทั้งสองฉบับ มีลักษณะการแจกแจงที่เกือบปกติ โดยคะแนนจากแบบทดสอบฉบับ Y มีการเบ้มากกว่า ($Sk = 0.042, -0.066$ ตามลำดับ) ความโด่งของแบบทดสอบทั้งฉบับ X และ Y มีลักษณะแบนกว่าโค้งปกติ โดยแบบทดสอบฉบับ X และ Y มีความโด่งใกล้เคียงกัน ($Ku = -1.392, -1.374$) และเมื่อพิจารณาความคลาดเคลื่อนของความเบ้และความโด่งของคะแนนจากแบบทดสอบทั้งสองฉบับมีค่าเท่ากัน คือ 0.101 และ 0.201 ตามลำดับ ซึ่งมีค่าไม่เกิน 2.00 หรือ น้อยกว่า -2.00 ซึ่งถือว่าไม่มีการแจกแจงเป็นโค้งปกติ (อวยพร เรืองตระกูล, 2544 : 116; อ้างอิงจาก SPSS BASE 8.0. 1998) ดังรูปร่างการแจกแจงคะแนนในภาพประกอบ 17



ภาพประกอบ 17 การแจกแจงคะแนนของกลุ่มสอบทานผลในแบบทดสอบฉบับ X และ Y

5.2 ผลการวิเคราะห์ความสัมพันธ์ระหว่างคะแนนแปลงกับคะแนนเกณฑ์

การวิเคราะห์ความสัมพันธ์ระหว่างคะแนนแปลง (X^*) ที่ได้จากการนำคะแนนแบบทดสอบฉบับ Y ไปแปลงให้อยู่ในมาตรวัดเดียวกันกับคะแนนของแบบทดสอบฉบับ X กับคะแนนที่ผู้สอบทำได้จากแบบทดสอบฉบับ X ปรากฏผลตามตาราง 47

ตาราง 47 ค่าสัมประสิทธิ์สหสัมพันธ์รายคู่ระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธี กับคะแนนเกณฑ์ของกลุ่มสอบทานผล

คะแนนแปลง	คะแนนเกณฑ์
XVEE	0.909**
XVEP	0.913**
XVETMS	0.910**
XVETTRF	0.910**
XVETS	0.910**
XVEOMS	0.913**
XVEOTRF	0.913**
XVEOS	0.914**

**p < 0.01

ผลการวิเคราะห์ตามตาราง 47 พบว่า คะแนนเกณฑ์ซึ่งเป็นคะแนนที่ผู้สอบทำได้ จากแบบทดสอบฉบับ X มีค่าสัมประสิทธิ์สหสัมพันธ์กับคะแนนแปลงจากการเทียบคะแนนแต่ละ วิธี อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 โดยมีค่าสูงกว่า 0.90 แสดงว่า การเทียบคะแนนใน แต่ละวิธีมีประสิทธิภาพค่อนข้างสูง

5.3 ผลการประเมินค่าดัชนีความแตกต่าง (C)

ดัชนีความแตกต่าง (C) เป็นดัชนีที่ใช้ประเมินความคงที่ของการเทียบ คะแนน เป็นการคำนวณค่าความแตกต่างระหว่างคะแนนแปลงกับคะแนนเกณฑ์ (ฉบับ X) ซึ่งเป็นคะแนนที่ผู้สอบแต่ละคนทำได้ตามสูตร 100 หน้า 132 ประเมินค่าดัชนี C โดยใช้เกณฑ์ ของปีเตอร์เซนและคณะ (Petersen and others. 1982: 93 – 97) ซึ่งกำหนดคุณภาพของผล การเทียบคะแนนตามระดับการยอมรับ 5 ระดับ (เกณฑ์ในหน้า 133) คือ

ระดับน่าพอใจมาก	เมื่อ $C \leq 1.353$
ระดับน่าพอใจ	เมื่อ $1.353 < C \leq 5.410$
ระดับปานกลาง	เมื่อ $5.410 < C \leq 12.173$
ระดับไม่น่าพอใจ	เมื่อ $12.173 < C \leq 21.641$
ระดับไม่น่าพอใจอย่างยิ่ง	เมื่อ $21.641 \leq C$

ผลการประเมินค่าดัชนีความแตกต่างปรากฏผลตามตาราง 48

ตาราง 48 การประเมินค่าดัชนีความแตกต่าง (C) จำแนกตามวิธีการเทียบคะแนน

วิธีการเทียบคะแนน	$\sum (X_i - X_i^*)^2$	S_x^2	C	ระดับการยอมรับ
LIN	61328.545	541.013	0.192	น่าพอใจอย่างมาก
EQU	57154.167	541.013	0.179	น่าพอใจอย่างมาก
TMS	61935.752	582.272	0.181	น่าพอใจอย่างมาก
TTRF	62437.017	582.272	0.182	น่าพอใจอย่างมาก
TS	63164.85	582.272	0.184	น่าพอใจอย่างมาก
OMS	55287.649	541.013	0.174	น่าพอใจอย่างมาก
OTRF	55207.709	541.013	0.173	น่าพอใจอย่างมาก
OS	55228.512	541.013	0.173	น่าพอใจอย่างมาก

ผลการวิเคราะห์ตามตาราง 48 พบว่า ค่าดัชนีความแตกต่าง (C) ของการเทียบคะแนน ทั้ง 8 รูปแบบ มีค่าน้อยกว่ากำลังสองร้อยละ 5 ของส่วนเบี่ยงเบนมาตรฐานของแบบทดสอบฉบับ X เท่ากับ 1.353 นั่นคือ ค่าดัชนีความแตกต่างจากการเทียบคะแนนทั้ง 5 วิธีการให้ความพอเพียงในการเทียบคะแนนอยู่ในระดับที่น่าพอใจอย่างมาก โดยค่าดัชนีความแตกต่างจากการเทียบคะแนนสังเกตวิธีปรับค่าพร้อมกัน (OS) และวิธีโค้งลักษณะแบบทดสอบ (OTRF) มีค่าน้อยที่สุด เท่ากับ 0.173 และค่าดัชนีความแตกต่างจากการเทียบคะแนนเชิงเส้นตรงมีค่าสูงสุดเท่ากับ 0.192

ค่าดัชนีความแตกต่างที่ได้แสดงถึงความคลาดเคลื่อนรวมในแต่ละวิธีการเทียบคะแนน เป็นการตรวจสอบจากเกณฑ์ภายนอก และสอดคล้องกับดัชนีคะแนนความแตกต่างทั้ง 6 ดัชนี ที่ตรวจสอบในเบื้องต้นว่า ดัชนีคะแนนความแตกต่างทั้ง 6 วิธี ให้ค่าสูงสุด เมื่อเปรียบเทียบคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรงกับคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีโค้งลักษณะแบบทดสอบ ขณะเดียวกันคะแนนแปลงจากการเทียบคะแนนสังเกตทุกวิธี มีค่าใกล้เคียงกับคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ และคะแนนแปลงจากการเทียบคะแนนจริงทุกวิธี มีค่าใกล้เคียงคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง แต่การเทียบคะแนนทุกวิธี ให้ความพอเพียงของการเทียบคะแนนอยู่ในระดับคุณภาพเดียวกัน คือ ระดับน่าพอใจอย่างมาก ดังนั้น จึงทดสอบความแตกต่างของคะแนนแปลงในแต่ละวิธีการเทียบคะแนนกับคะแนนเกณฑ์ ว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติหรือไม่

5.4 ผลการตรวจสอบความแตกต่างระหว่างคะแนนแปลงกับคะแนน

เกณฑ์

การทดสอบความมีนัยสำคัญทางสถิติของความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนเกณฑ์ โดยใช้สถิติทดสอบนัยพาราเมตริกซ์ Friedman Test ซึ่งใช้กับข้อมูลที่ทำกรวัดซ้ำโดยการจัดอันดับ และไม่ต้องมีข้อตกลงของความ เป็นปกติของการแจกแจง ความแตกต่างของคะแนนพิจารณาเฉพาะค่าไม่คำนึงถึงเครื่องหมาย ปรากฏผลตามตาราง 49

ตาราง 49 ผลการวิเคราะห์ความแตกต่างของคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับ คะแนนเกณฑ์

N	Chi – Square	df	p
589	17.058	7	.017

$p > 0.01$

ผลการวิเคราะห์ตามตาราง 49 พบว่า คะแนนแปลงจากการเทียบคะแนนแต่ละวิธี กับคะแนนเกณฑ์ไม่แตกต่าง แสดงว่าความคลาดเคลื่อนของการเทียบคะแนนในแต่ละวิธีไม่ แตกต่างกัน

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีจุดมุ่งหมาย เพื่อศึกษาผลของการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่า เมื่อเทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบโมเดลเกรต เรสพอนส์ คือ (1) การเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกัน (SIM) (2) การเทียบคะแนนวิธีโค้งลักษณะแบบทดสอบ (TRF) (3) การเทียบคะแนนวิธีค่าเฉลี่ยและส่วนเบี่ยงมาตรฐานของพารามิเตอร์ความยากข้อสอบ (MS) ซึ่งเทียบทั้งคะแนนจริงและคะแนนสังเกต และตามวิธีการในทฤษฎีการทดสอบแบบมาตรฐานเดิม คือ (4) การเทียบคะแนนเชิงเส้นตรง และ (5) การเทียบคะแนนอีควิเปอร์เซ็นไทล์ โดยผลการเทียบคะแนนพิจารณาจาก ความคล้ายคลึงกันของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงกับคะแนนจากแบบทดสอบฉบับเป้าหมาย ดัชนีคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมกับการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ และดัชนีความแตกต่าง (C) ในแต่ละวิธีการเทียบคะแนนที่วิเคราะห์จากกลุ่มสอบทานผล โดยมีสมมุติฐานการวิจัย 3 ข้อ คือ ข้อแรกการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนของแบบทดสอบฉบับเป้าหมายไม่ต่างกัน ข้อสองดัชนีความแตกต่าง (C) ที่ได้จากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบซึ่งวิเคราะห์จากกลุ่มสอบทานผล มีคุณภาพอยู่ในระดับที่สูงกว่าที่ได้จากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ข้อสามความคลาดเคลื่อนของการเทียบคะแนนแต่ละวิธีจากกลุ่มสอบทานผลให้ค่าที่แตกต่างกัน

กลุ่มตัวอย่างที่ใช้ในการวิจัย ได้แก่ นักเรียนชั้นมัธยมศึกษาที่สอง ภาคเรียนที่ 1 ปีการศึกษา 2545 ของโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา จังหวัดกรุงเทพมหานคร ประกอบด้วย (1) กลุ่มตัวอย่างเทียบคะแนน จำนวน 2,312 คน เป็นกลุ่มที่ใช้เพื่อนำผลการสอบไปวิเคราะห์เทียบคะแนนแบบทดสอบจำแนกเป็นนักเรียนที่ทดสอบด้วยแบบทดสอบฉบับ X จำนวน 1,155 คน และฉบับ Y จำนวน 1,175 คน และ (2) กลุ่มสอบทานผล จำนวน 589 คน เป็นกลุ่มที่ใช้เพื่อประเมินคุณภาพของวิธีการเทียบคะแนน โดยกลุ่มสอบทานผลแต่ละคนต้องทำแบบทดสอบทั้งสองฉบับ

เครื่องมือที่ใช้ในการวิจัยประกอบด้วยแบบทดสอบเทียบคะแนน 2 ฉบับ เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ด้านเหตุผล ฉบับละ 30 ข้อ โดยแต่ละฉบับได้รวมแบบทดสอบร่วมที่วัดเนื้อหาเดียวกันกับแบบทดสอบเทียบคะแนน จำนวน 10 ข้อ แบบทดสอบเทียบคะแนนเป็นแบบทดสอบเลือกตอบ 5 ตัวเลือก แต่ละข้อมีคำตอบให้ผู้สอบตัดสินทุกตัวเลือก ตรวจให้คะแนน 6 ลำดับชั้นคะแนน คือ ให้คะแนน 0 คะแนน เมื่อผู้สอบตอบไม่ถูกหรือไม่ตอบทุกตัวเลือก ให้คะแนน 1 คะแนน เมื่อผู้สอบตอบถูกหนึ่งตัวเลือก ให้คะแนน 2 คะแนน เมื่อผู้สอบตอบถูกสองตัวเลือก ให้คะแนน 3 คะแนน เมื่อผู้สอบตอบถูกสามตัวเลือก ให้คะแนน 4 คะแนน เมื่อผู้สอบตอบถูกสี่ตัวเลือก และให้คะแนน 5 คะแนน เมื่อผู้สอบตอบถูกห้าตัวเลือก แบบทดสอบทั้งสองฉบับวัดความรู้ครอบคลุมเนื้อหา เรื่องจำนวนเต็ม เศษส่วนและทศนิยม สร้างจากจุดประสงค์เชิงพฤติกรรมที่สอดคล้องกับระดับพฤติกรรมการวัด 4 ระดับชั้นของวิสัยทัศน์ จำนวน 100 ข้อ ทดลองใช้แบบทดสอบเพื่อหาคุณภาพและความเหมาะสมที่จะนำไปใช้กับกลุ่มตัวอย่างจริง ครั้งหนึ่ง จำนวน 179 คน และครั้งที่สอง จำนวน 439 คน แล้วคัดเลือกข้อสอบจากกลุ่มข้อสอบที่มีคุณภาพ ตามจำนวนข้อในแต่ละจุดประสงค์ โดยพิจารณาข้อสอบที่มีค่าจำแนกอยู่ในเกณฑ์ดีและมีสารสนเทศของข้อสอบใกล้เคียงกัน จำแนกเป็นข้อสอบที่ไม่ได้ร่วม (Unique item) ฉบับละ 20 ข้อ และข้อสอบร่วมฉบับละ 10 ข้อ

การเก็บรวบรวมข้อมูล โดยนำแบบทดสอบเทียบคะแนนทั้งสองฉบับ คือ ฉบับ X และฉบับ Y ฉบับละ 30 ข้อ ไปทดสอบกับกลุ่มตัวอย่างเทียบคะแนน จำนวน 12 โรงเรียน โดยสุ่มอย่างมีระบบให้นักเรียนในแต่ละห้อง ได้สอบแบบทดสอบฉบับ X และฉบับ Y สลับกัน แล้วนำผลการสอบของนักเรียนมาแยกตามลักษณะของแบบทดสอบเทียบคะแนน ตรวจให้คะแนนในแต่ละฉบับ และผู้วิจัยได้นำแบบทดสอบเทียบคะแนนทั้งสองฉบับไปสอบกับกลุ่มสอบทานผลเพื่อประเมินคุณภาพของวิธีการเทียบคะแนน จำนวน 4 โรงเรียน โดยให้หน่วยตัวอย่างแต่ละคนสอบแบบทดสอบเทียบคะแนนทั้งสองฉบับ โดยสุ่มอย่างมีระบบให้แต่ละคนสอบแบบทดสอบฉบับ X แล้วตามด้วยฉบับ Y หรือสอบฉบับ Y ก่อน แล้วตามด้วยฉบับ X นำคะแนนผลการสอบของกลุ่มสอบทานผลมาดำเนินการเช่นเดียวกับกลุ่มตัวอย่างเทียบคะแนน ได้คะแนนผลการสอบแต่ละคู่

การวิเคราะห์ข้อมูลจำแนกเป็น 5 ตอน คือ ตอนแรกเป็นการวิเคราะห์ข้อมูลคะแนนของแบบทดสอบในกลุ่มเทียบคะแนน เพื่อวิเคราะห์คุณภาพของข้อสอบรายข้อ ตรวจสอบความเป็นเอกมิติของแบบทดสอบเทียบคะแนน วิเคราะห์ค่าสถิติพื้นฐานของคะแนนรวมทั้งตรวจสอบค่าเฉลี่ยและความแตกต่างของการแจกแจงคะแนนของแบบทดสอบเทียบคะแนนทั้งสองฉบับ วิเคราะห์คุณภาพของแบบทดสอบเทียบคะแนน และวิเคราะห์ค่าสถิติพื้นฐานและความสัมพันธ์ของแบบทดสอบร่วมกับแบบทดสอบเทียบคะแนน

การวิเคราะห์ข้อมูลตอนที่สองเป็นการสร้างตารางเทียบคะแนน จำแนกตามวิธีการเทียบคะแนนในทฤษฎีการทดสอบออกเป็นสองกลุ่มใหญ่ ๆ คือ กลุ่มที่หนึ่ง สร้างตารางการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ซึ่งเป็นการเทียบคะแนนดิบจากฉบับ Y ปรับสู่สเกลฉบับ X มี 2 วิธี คือ 2.1.1) การสร้างตารางเทียบคะแนนเชิงเส้นตรง มีขั้นตอนคือ 2.1.1.1) คำนวณค่าเฉลี่ยและความแปรปรวนของแบบทดสอบเทียบคะแนนฉบับ X และ Y ที่ปรับค่าด้วย ค่าเฉลี่ยและความแปรปรวนของแบบทดสอบรวมจากกลุ่ม α (สอบแบบทดสอบฉบับ X) กลุ่ม β (สอบแบบทดสอบฉบับ Y) และกลุ่มรวม ($\alpha + \beta$) 2.1.1.2) คำนวณค่าความชันและจุดตัดแกนของสมการเส้นตรงที่ใช้ในการเทียบคะแนน 2.1.1.3) เทียบคะแนนจากแบบทดสอบฉบับ Y ปรับสู่สเกลฉบับ X โดยแทนค่าคะแนนดิบลงในสมการ 2.1.1.2) 2.1.2) การสร้างตารางเทียบคะแนนอีควิเปอร์เซ็นไทล์ มีขั้นตอนคือ 2.1.2.1) ปรับการแจกแจงคะแนนดิบของแบบทดสอบฉบับ X และ Y โดยใช้อัตราส่วนความถี่ของแต่ละระดับของคะแนนของแบบทดสอบรวมจากกลุ่มรวมต่อกลุ่ม α และกลุ่ม β ตามลำดับ 2.1.2.2) ปรับการแจกแจงคะแนนดิบของแบบทดสอบฉบับ X และ Y ให้มีความต่อเนื่อง 2.1.2.3) คำนวณตำแหน่งเปอร์เซ็นต์ไทล์ของคะแนนแบบทดสอบฉบับ Y 2.1.2.4) คำนวณคะแนนของแบบทดสอบฉบับ X ณ ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกับแบบทดสอบฉบับ Y จะได้ตารางเทียบคะแนนอีควิเปอร์เซ็นไทล์

กลุ่มที่สอง การสร้างตารางเทียบคะแนนตามทฤษฎีการตอบข้อสอบมี 3 วิธี ในแต่ละวิธีสร้างตารางเทียบคะแนนจริงและคะแนนสังเกตจำแนกเป็น 2.2.1) การสร้างตารางเทียบคะแนนตามทฤษฎีการตอบข้อสอบเมื่อวิเคราะห์ข้อมูลแยกกัน มีขั้นตอนคือ 2.2.1.1) วิเคราะห์ค่าพารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถในแบบทดสอบเทียบคะแนนทั้งสองฉบับแยกกัน 2.2.1.2) สร้างสัมประสิทธิ์การเทียบคะแนน 2 วิธี คือ วิธีที่หนึ่ง ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของข้อสอบรวม และวิธีที่สองใช้โค้งสารสนเทศแบบทดสอบรวม ในการประมาณค่าสัมประสิทธิ์การเทียบคะแนน 2.2.1.3) สร้างตารางเทียบคะแนนจริง มีขั้นตอนคือ 2.2.1.3.1) คำนวณระดับความสามารถผู้สอบฉบับ Y ที่สอดคล้องกับคะแนนจริงในช่วงคะแนน 1-150 2.2.1.3.2) คำนวณคะแนนจริงของผู้สอบฉบับ X ที่สมนัย ณ ระดับความสามารถเดียวกัน (คือแปลง θ ที่ได้จาก 2.2.1.3.1) ด้วยสัมประสิทธิ์การเทียบคะแนนในแต่ละวิธี) จะได้ตารางเทียบคะแนนจริง 2 วิธี 2.2.1.4) สร้างตารางเทียบคะแนนสังเกต มีขั้นตอนดังนี้ 2.2.1.4.1) สร้างการแจกแจงคะแนนความสามารถของผู้สอบฉบับ X และ Y ใช้จุดกลางของขั้นคะแนนเป็นตัวแทนระดับความสามารถในแต่ละฉบับ 2.2.1.4.2) คำนวณค่าความน่าจะเป็นของผู้สอบในแต่ละระดับความสามารถที่ทำคะแนนได้ตั้งแต่ 0-150 ในแบบทดสอบฉบับ X และ ฉบับ Y (สำหรับระดับความสามารถในฉบับ Y เป็นระดับความสามารถที่แปลงจากสมการใน 2.2.1.2) ในแต่ละวิธี) 2.2.1.4.3) คำนวณผลรวมความน่าจะเป็นที่ผู้มีความสามารถทุกระดับจะตอบข้อสอบได้คะแนน 0-150 ในแบบทดสอบ X และ Y 2.2.1.4.4) สร้างตารางเทียบคะแนน

สังเกตด้วยวิธีการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ ตามขั้นตอนใน 2.1.2.2) - 2.1.2.4) 2.2.2) การสร้างตารางเทียบคะแนนตามทฤษฎีการตอบข้อสอบเมื่อวิเคราะห์ ข้อมูลพร้อมกันมี ขั้นตอน คือ 2.2.2.1) วิเคราะห์ค่าพารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถของผู้สอบ พร้อมกันทั้งสองฉบับ 2.2.2.2) การสร้างตารางเทียบคะแนนจริงมีขั้นตอนตาม 2.2.1.3.1) - 2.2.1.3.2) โดยไม่ต้องแปลงระดับความสามารถ 2.2.2.3) การสร้างตารางเทียบคะแนนสังเกตมี ขั้นตอนตาม 2.2.1.4.1) - 2.2.1.4.4) โดยไม่ต้องแปลงระดับความสามารถในฉบับ Y

การวิเคราะห์ข้อมูลตอนที่สาม เป็นการตรวจสอบความคล้ายคลึงกันของการแจกแจง คะแนนสะสม โดยเปรียบเทียบความแตกต่างของการแจกแจงคะแนนสะสมระหว่างคะแนนแปลง ที่ได้กับคะแนนจากแบบทดสอบเป้าหมาย (ฉบับ X) ด้วยสถิติทดสอบโคโมโกรอฟ - สมิโนฟ ตอนที่สี่เป็นการวิเคราะห์ความแตกต่างของคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการ ตอบข้อสอบกับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน ด้วยการเปรียบเทียบความแตกต่าง ในแต่ละระดับคะแนนและตรวจสอบด้วยดัชนีคะแนนความแตกต่าง 6 ดัชนี ได้แก่ ดัชนี RMSD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก ดัชนี MAD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก และดัชนี MSD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก และตอนที่ห้าเป็นการวิเคราะห์คุณภาพของการเทียบ คะแนนวิธี ต่าง ๆ โดยใช้กลุ่มสอบทานผล เพื่อประเมินคุณภาพของการเทียบคะแนนแต่ละวิธี ด้วยดัชนีความแตกต่างตามเกณฑ์ของปีเตอร์เซ็น และคณะ และตรวจสอบความแตกต่าง ระหว่างคะแนนแปลงที่ได้จากการเทียบคะแนนในแต่ละวิธีกับคะแนนเกณฑ์ด้วยสถิติทดสอบ ฟรีดแมน

สรุปผลการวิจัย

ผลการวิจัยในการศึกษาค้างนี้ สรุปตามประเด็นสำคัญได้ดังนี้

1. ผลการสร้างตารางเทียบคะแนนแบบทดสอบ จากวิธีการเทียบคะแนนทั้ง 5 วิธี 8 รูปแบบ จำแนกเป็นการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้แก่วิธีการ เทียบคะแนนเชิงเส้นตรง และอีคิวเปอร์เซ็นต์ไทล์ ซึ่งเป็นการเทียบคะแนนดิบจากแบบทดสอบ ฉบับ Y ปรับสู่สเกลของแบบทดสอบฉบับ X และวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ จำแนกตามวิธีการวิเคราะห์ค่าพารามิเตอร์ ซึ่งมีทั้งวิเคราะห์แยกฉบับและวิเคราะห์พร้อมกันรวม 3 วิธี แต่ละวิธีเทียบคะแนนโดยใช้คะแนนจริงและคะแนนสังเกต จากแบบทดสอบฉบับ Y ปรับสู่ สเกลฉบับ X ปรากฏผลโดยรวมคือคะแนนสมมูลของแบบทดสอบฉบับ Y มีค่าสูงกว่าคะแนน ของแบบทดสอบฉบับ X ในช่วงคะแนนที่แตกต่างกันของแต่ละวิธีการเทียบคะแนน ดังนี้

1.1 ในการเทียบคะแนนเชิงเส้นตรง คะแนนสมมูลของแบบทดสอบฉบับ Y ทุก ระดับคะแนน คือ 1-150 คะแนน มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X ระหว่าง 0.993 ถึง 5.009 โดยในช่วงคะแนน 1-58 ซึ่งเป็นช่วงคะแนนที่ไม่มีผู้สอบได้คะแนนดังกล่าว คะแนนสมมูล ของแบบทดสอบฉบับ Y มีค่าสูงกว่าคะแนนแบบทดสอบฉบับ X ค่อนข้างมาก และลดต่ำลงใน

ช่วงคะแนนที่เข้าใกล้คะแนนเต็ม สำหรับการเทียบคะแนนอีควิเปอร์เซ็นไทล์ ปรากฏผลคือ คะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X คือ ในช่วงคะแนน 70-147 สำหรับช่วงคะแนน 59-69 และ 148-150 คะแนนสมมูลของแบบทดสอบฉบับ Y ให้ค่าต่ำกว่าคะแนนของแบบทดสอบฉบับ X โดยในช่วงคะแนน 1-58 ซึ่งเป็นช่วงที่ไม่มีผู้สอบทำคะแนนได้นั้น ไม่นำมาพิจารณาเพราะต้องมีการทำ extrapolation ซึ่งทำให้คะแนนแปลงมีความคลาดเคลื่อนสูง

1.2 การเทียบคะแนนตามทฤษฎีการตอบข้อสอบวิธีปรับค่าพารามิเตอร์พร้อมกัน ในการเทียบคะแนนจริงปรากฏผล คือคะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X คือในช่วงคะแนน 46-147 สำหรับในช่วงคะแนนที่ต่ำกว่า 46 คะแนน คะแนนสมมูลของแบบทดสอบฉบับ Y จะให้ค่าที่สูงหรือต่ำกว่าคะแนนของแบบทดสอบฉบับ X สลับกันไป และในช่วงคะแนนที่สูงกว่า 147 คะแนน คะแนนสมมูลของแบบทดสอบฉบับ Y มีค่าที่ใกล้เคียงกับคะแนนของแบบทดสอบฉบับ X และผลที่ได้จากการเทียบคะแนนสังเกตปรากฏว่าคะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X คือในช่วงคะแนน 80-149 สำหรับในช่วงคะแนน 46-79 และ 150 คะแนนสมมูลของแบบทดสอบฉบับ Y มีค่าต่ำกว่าคะแนนของแบบทดสอบฉบับ X และในช่วงคะแนน 1-45 เป็นช่วงคะแนน ที่ค่าความน่าจะเป็นที่จะได้ระดับคะแนนดังกล่าวเป็นศูนย์ทั้งฉบับ X และ Y จึงไม่นำมาพิจารณา เพราะต้องทำ extrapolation ซึ่งทำให้คะแนนแปลงที่ได้มีความคลาดเคลื่อนสูง

1.3 การเทียบคะแนนตามทฤษฎีการตอบข้อสอบ เมื่อปรับค่าพารามิเตอร์แยกกัน ซึ่งจำแนกเป็น 2 วิธี ปรากฏผลดังนี้

1.3.1 การเทียบคะแนนวิธีใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก ในการเทียบคะแนนจริง ปรากฏว่าคะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X คือในช่วงคะแนน 57-144 สำหรับในช่วงคะแนนที่ต่ำกว่า 57 คะแนน คะแนนสมมูลของแบบทดสอบฉบับ Y มีค่าสูงหรือต่ำกว่าคะแนนของแบบทดสอบฉบับ X สลับกันไป ส่วนในช่วงคะแนน 145-149 คะแนนสมมูลของแบบทดสอบฉบับ Y มีค่าต่ำกว่าคะแนนของแบบทดสอบฉบับ X และเมื่อเทียบคะแนนสังเกตปรากฏว่าคะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X คือในช่วงคะแนน 88-146 สำหรับช่วงคะแนน 45-87 มีค่าต่ำกว่าคะแนนของแบบทดสอบฉบับ X และช่วงคะแนนที่สูงกว่า 146 คะแนนสมมูลของแบบทดสอบฉบับ Y มีค่าต่ำกว่าแบบทดสอบฉบับ X ส่วนในช่วงคะแนน 1-44 เป็นช่วงคะแนนที่ค่าความน่าจะเป็นที่จะได้ระดับคะแนนดังกล่าวเป็นศูนย์ ทั้งในฉบับ X และ Y จึงไม่นำมาพิจารณาเพราะต้องทำ extrapolation ในช่วงคะแนนดังกล่าว ซึ่งทำให้คะแนนแปลงมีความคลาดเคลื่อนสูง

1.3.2 การเทียบคะแนนวิธีโค้งลักษณะแบบทดสอบในการเทียบคะแนนจริง
ปรากฏว่า คะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่มีค่าสูงกว่าคะแนนของแบบทดสอบ
ฉบับ X ในช่วงคะแนน 53-144 สำหรับในช่วงคะแนนต่ำกว่า 53 คะแนนสมมูลของแบบทดสอบ
ฉบับ Y มีค่าสูงหรือต่ำกว่าคะแนนของแบบทดสอบฉบับ X สลับกันไปและในช่วงคะแนนที่สูง
กว่า 144 คะแนน คะแนนสมมูลของแบบทดสอบฉบับ Y มีค่าต่ำกว่าคะแนนสมมูลของแบบ-
ทดสอบฉบับ X และเมื่อเทียบคะแนนสังเกตพบว่าคะแนนสมมูลของแบบทดสอบฉบับ Y ส่วน
ใหญ่มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X ในช่วงคะแนน 90-146 สำหรับในช่วงคะแนน
45-89 และ 147-150 คะแนนสมมูลของแบบทดสอบฉบับ Y ต่ำกว่าคะแนนของแบบทดสอบ
ฉบับ X และเช่นเดียวกับผลใน 1.3.1 ช่วงคะแนน 1-44 เป็นช่วงคะแนนที่ค่าความน่าจะเป็นที่จะ
ได้ระดับคะแนนดังกล่าวเป็นศูนย์ ทั้งในฉบับ X และ Y จึงไม่นำมาพิจารณา เพราะต้องทำ
extrapolation ในช่วงคะแนนดังกล่าว ซึ่งทำให้คะแนนแปลงที่ได้มีความคลาดเคลื่อนสูง ดังผล
สรุปตามตาราง 50

ตาราง 50 ผลการเปรียบเทียบคะแนนสมมูลฉบับ Y กับคะแนนแปลง (X*)

วิธีการเทียบคะแนน	ระดับคะแนนสมมูลฉบับ Y	ลักษณะของคะแนนสมมูลฉบับ Y เทียบ กับคะแนนแปลง (X*)
LIN	1-150	สูงกว่า*
EQU	70-147	สูงกว่า*
	59-69, 148-150	ต่ำกว่า
TS	46-147	สูงกว่า*
	148-150	ใกล้เคียงกัน
	1-45	ต่ำและสูงสลับกัน
OS	80-149	สูงกว่า*
	46-79, 150	ต่ำกว่า
TMS	57-144	สูงกว่า*
	1-56	สูงและต่ำสลับกัน
	145-149	ต่ำกว่า
OMS	88-146	สูงกว่า*
	45-87, 146-150	ต่ำกว่า
TTRF	53-144	สูงกว่า*
	144-149	ต่ำกว่า
	1-52	สูงและต่ำสลับกัน
OTRF	90-146	สูงกว่า*
	147-150, 45-89	ต่ำกว่า

2. ผลการวิเคราะห์ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม

ผลการตรวจสอบความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ระหว่างคะแนนแปลงจากการเทียบคะแนนในแต่ละวิธีกับคะแนนจากแบบทดสอบเป้าหมาย (ฉบับ X) ปรากฏว่าคะแนนแปลงจากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ มีค่าเฉลี่ยและความเบ้แตกต่างจากแบบทดสอบฉบับ X น้อยที่สุด ส่วนคะแนนแปลงจากการเทียบคะแนนสังเกต วิธีปรับค่าพารามิเตอร์พร้อมกันมีค่าส่วนเบี่ยงเบนมาตรฐานแตกต่างจากแบบทดสอบฉบับ X น้อยที่สุด และคะแนนแปลงจากการเทียบคะแนนสังเกตวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากมีค่าความโด่งแตกต่างจากแบบทดสอบฉบับ X น้อยที่สุด ซึ่งผลดังกล่าวทำให้ไม่สามารถสรุปได้ว่าโมเมนต์ของคะแนนแปลงจากการเทียบคะแนนวิธีใดเหมือนกับโมเมนต์ของคะแนนจากแบบทดสอบเป้าหมาย แต่เมื่อวิเคราะห์ผลเพื่อเปรียบเทียบค่าเฉลี่ยและการแจกแจงสะสมของคะแนนแปลงจากการเทียบคะแนนในแต่ละวิธี กับคะแนนจากแบบทดสอบเป้าหมาย พบว่าไม่มีความแตกต่างกัน แสดงว่าสัดส่วนความถี่สะสมของคะแนนแปลงกับคะแนนจากแบบทดสอบเป้าหมายในแต่ละระดับมีค่าเท่ากัน สรุปได้ว่าการเทียบคะแนนแต่ละวิธีมีคุณสมบัติของการเทียบคะแนน ซึ่งสอดคล้องกับสมมติฐานที่ตั้งไว้

3. ผลการวิเคราะห์ความแตกต่างของคะแนนแปลง จำแนกได้ 2 กรณี ปรากฏผลดังนี้

3.1 ผลการวิเคราะห์คะแนนความแตกต่าง... ระหว่างคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ 3 วิธี (6 รูปแบบ) กับคะแนนแปลงจากการเทียบคะแนนพื้นฐาน (การเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม) สามารถสรุปได้ว่าคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนจริง และคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน กับการเทียบคะแนนเชิงเส้นตรงมีค่าน้อย แสดงว่าฟังก์ชันการเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกัน มีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนเชิงเส้นตรง และฟังก์ชันการเทียบคะแนนจริงทุกวิธีมีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนเชิงเส้นตรงมากกว่าฟังก์ชันการเทียบคะแนนสังเกตทุกวิธี แต่ผลที่ได้ของคะแนนความแตกต่างระหว่างคะแนนแปลงจากการเทียบคะแนนตามทฤษฎีการตอบข้อสอบกับวิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์ ให้ผลที่ไม่ชัดเจนว่าวิธีใดมีฟังก์ชันการเทียบคะแนนที่คล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์เซ็นไทล์ ทั้งการเทียบคะแนนจริงและการเทียบคะแนนสังเกต เพราะในแต่ละช่วงคะแนนสลับกันให้ผลที่คล้ายคลึงกัน

3.2 ผลการวิเคราะห์ด้วยดัชนีคะแนนความแตกต่างทั้ง 6 ดัชนี คือ ดัชนีรากที่สองของค่าเฉลี่ยความแตกต่างกำลังสอง (RMSD) ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก ดัชนีค่าเฉลี่ยความแตกต่างสัมบูรณ์ (MAD) ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก และดัชนีค่าเฉลี่ยความแตกต่างที่คำนึงเครื่องหมาย (MSD) ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก ปรากฏว่าผลที่ได้จากการวิเคราะห์ดัชนี RMSD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก ดัชนี MAD ทั้งถ่วงน้ำหนักและ

ไม่ถ่วงน้ำหนัก และดัชนี MSD ไม่ถ่วงน้ำหนัก สรุปได้ว่าการเทียบคะแนนจริงวิธีค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (MS) มีฟังก์ชันการเทียบคะแนนที่ คล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากกว่าการเทียบคะแนนวิธีอื่น โดยคล้ายคลึงกับ ฟังก์ชันการเทียบคะแนนอีคิวเปอร์เซ็นไทล์มากกว่าฟังก์ชันการเทียบคะแนนเชิงเส้นตรง และผล ที่ได้จากการวิเคราะห์ MSD ที่ถ่วงน้ำหนัก สรุปได้ว่าการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์ พร้อมกันมีฟังก์ชันที่คล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐาน โดยเฉพาะมีความ คล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีคิวเปอร์เซ็นไทล์มากกว่าฟังก์ชันการเทียบคะแนนเชิง-เส้นตรง นอกจากนี้ยังพบว่าผลวิเคราะห์ที่ได้จากการวิเคราะห์ดัชนีทั้ง 6 ดัชนีนั้น ฟังก์ชันการ เทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบ มีความแตกต่างกับฟังก์ชันการเทียบคะแนน เชิงเส้นตรงมากที่สุด และโดยภาพรวมพบว่าการเทียบคะแนนจริงทุกวิธีมีฟังก์ชันการเทียบ คะแนนที่คล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากกว่าการเทียบคะแนนสังเกต ยกเว้น การเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน มีฟังก์ชันการเทียบคะแนนที่คล้ายคลึง กับการเทียบคะแนนอีคิวเปอร์เซ็นไทล์

4. ผลการประเมินคุณภาพของการเทียบคะแนน โดยใช้ค่าดัชนีความแตกต่าง (c) จากการวิเคราะห์กลุ่มสอบทานผล - ซึ่งเป็นหน่วยตัวอย่างที่สุ่มมาจากประชากรเดียวกันกับกลุ่ม เทียบคะแนน ผลการวิเคราะห์ค่าสถิติพื้นฐานของคะแนนพบว่า คะแนนต่ำสุด สูงสุดที่มีผู้สอบ ทำได้ในแบบทดสอบฉบับ X และ Y มีค่าเท่ากับ 64 และ 149 คะแนน และเท่ากับ 58 และ 149 คะแนน ตามลำดับ ซึ่งใกล้เคียงกับที่กลุ่มเทียบคะแนนทำได้ ค่าเฉลี่ยของคะแนนแบบทดสอบ ฉบับ X ต่ำกว่าฉบับ Y เช่นเดียวกับกลุ่มเทียบคะแนน แสดงว่ากลุ่มสอบทานผลซึ่งเป็นกลุ่มที่ใช้ ตรวจสอบความคลาดเคลื่อนรวมของการเทียบคะแนน มีคุณลักษณะที่ใกล้เคียงกับกลุ่มตัวอย่าง เทียบคะแนน ซึ่งเป็นกลุ่มที่ข้อมูลถูกนำมาสร้างตารางเทียบคะแนน ปรากฏผลการประเมิน คุณภาพดังนี้

4.1 ค่าดัชนีความแตกต่างจากการเทียบคะแนนแต่ละวิธี คือ การเทียบคะแนน เชิงเส้นตรง อีคิวเปอร์เซ็นไทล์ การเทียบคะแนนจริง และคะแนนสังเกตวิธีค่าเฉลี่ยและส่วน เบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก การเทียบคะแนนจริงและคะแนนสังเกตวิธี โค้งลักษณะแบบทดสอบ และการเทียบคะแนนจริง และคะแนนสังเกตวิธีปรับค่าพารามิเตอร์ พร้อมกัน มีค่าเท่ากับ 0.192 0.179 0.181 0.174 0.182 0.173 0.184 และ 0.173 ตาม ลำดับ ซึ่งวิธีการเทียบคะแนนเชิงเส้นตรงให้ความคลาดเคลื่อนรวมสูงสุด วิธีการปรับค่าพารามิเตอร์พร้อมกัน และวิธีโค้งลักษณะแบบทดสอบ เมื่อเทียบคะแนนสังเกตให้ความคลาดเคลื่อน รวมต่ำสุด และเมื่อประเมินด้วยเกณฑ์ของปีเตอร์เซ็น และคณะ พบว่าผลการเทียบคะแนนทุกวิธี เมื่อเทียบกับเกณฑ์ให้ผลอยู่ในระดับที่น่าพอใจอย่างมาก ซึ่งขัดแย้งกับสมมุติฐานที่กำหนดไว้

4.2 การเปรียบเทียบความแตกต่างสัมบูรณ์ระหว่างคะแนนแปลงจากการเทียบคะแนนกับคะแนนเกณฑ์ในกลุ่มสอบทานผล พบว่าไม่แตกต่างกัน แสดงว่าความคลาดเคลื่อนของการเทียบคะแนนในแต่ละวิธีไม่แตกต่างกัน ซึ่งขัดแย้งกับสมมุติฐานที่กำหนดไว้

โดยภาพรวมผลการเทียบคะแนนที่ได้จากเกณฑ์ทั้ง 3 เกณฑ์ คือ ดัชนีคะแนนความแตกต่าง 6 ดัชนี ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม และดัชนีความแตกต่าง มีความสอดคล้องกัน คือ ผลที่ได้จากคะแนนความแตกต่างทั้ง 5 ดัชนี คือ $RMSD_w$, $RMSD_u$, MAD_w , MAD_u , MSD_u สรุปได้ว่าการเทียบคะแนนจริงวิธี MS มีฟังก์ชันการเทียบคะแนนคล้ายคลึงกับอีควิเปอร์เซ็นไทล์มากที่สุด หากพิจารณาค่าดัชนีความแตกต่างประกอบจะพบว่าวิธีการเทียบคะแนนทั้งคู่มิค่าดัชนีความแตกต่าง ต่างกันน้อยที่สุดเมื่อเทียบกับวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบอื่น ๆ สำหรับผลที่ได้จากดัชนีคะแนนความแตกต่างที่สอดคล้องกันทั้ง 6 ดัชนี คือ การเทียบคะแนนจริงวิธี TRF มีฟังก์ชันการเทียบคะแนนแตกต่างกับการเทียบคะแนนเชิงเส้นตรงมากที่สุด หากพิจารณาค่าดัชนีความแตกต่างประกอบจะพบว่าวิธีการเทียบคะแนนทั้งคู่มิค่าดัชนีความแตกต่าง ต่างกันมากที่สุดเมื่อเทียบกับวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบอื่น ๆ เมื่อประเมินค่าดัชนีความแตกต่างด้วยเกณฑ์ของปีเตอร์เซ็นและคณะ ยังพบว่ามีคุณภาพอยู่ในระดับเดียวกัน ซึ่งสอดคล้องกับการตรวจสอบความคล้ายคลึงกันของการแจกแจงความถี่สะสมระหว่างคะแนนแปลงกับคะแนนของแบบทดสอบฉบับเป้าหมาย (X) ว่าไม่มีความแตกต่างกันในทุกวิธีการเทียบคะแนน ดังผลสรุปตามตาราง 51

ตาราง 51 สรุปผลการเทียบคะแนนจากเกณฑ์ประเมินผล 3 เกณฑ์ จำแนกตามวิธีการเทียบคะแนน

เกณฑ์การเทียบคะแนน	วิธีการเทียบคะแนน							
	LIN	EQU	OMS	TMS	OTRF	TRRF	OSIM	TSIM
ดัชนีคะแนนความแตกต่าง								
$RMSD_w$				EQU(1)	LIN(12)			
$RMSD_u$				EQU(1)	LIN(12)			
MAD_w				EQU(1)	LIN(12)			
MAD_u				EQU(1)	LIN(12)			
MSD_w					LIN(12)		EQU(1)	
MSD_u				LIN(1)	LIN(12)			
การแจกแจงคะแนนสะสม	ไม่ แตกต่าง	ไม่ แตกต่าง	ไม่ แตกต่าง	ไม่ แตกต่าง	ไม่ แตกต่าง	ไม่ แตกต่าง	ไม่ แตกต่าง	ไม่ แตกต่าง
ค่าดัชนีความแตกต่าง	0.912(8)	0.719(4)	0.714(3)	0.181(5)	0.173(1)	0.182(6)	0.173(1)	0.184(7)
ระดับคุณภาพ	1	1	1	1	1	1	1	1

อภิปรายผลการวิจัย

การศึกษาผลการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า ด้วยวิธีการเทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์ และตามทฤษฎีการทดสอบแบบมาตรฐานเดิม โดยผลการเทียบคะแนนพิจารณาจาก ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ค่าดัชนีคะแนนความแตกต่าง และดัชนีความแตกต่าง (c) จากกลุ่มสอบทานผลเกณฑ์ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม เป็นเกณฑ์ที่แสดงถึงคุณสมบัติของการเทียบคะแนน โดยพิจารณาว่าการแจกแจงคะแนนสะสมของคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับการแจกแจงคะแนนสะสมของคะแนนจากแบบทดสอบฉบับเป้าหมายเป็นอย่างไร หากคล้ายคลึงกันแสดงว่าวิธีการเทียบคะแนนนั้นมีคุณสมบัติของการเทียบคะแนน ขณะเดียวกันก็ใช้เกณฑ์ดัชนีคะแนนความแตกต่าง ซึ่งจะแสดงให้เห็นว่าฟังก์ชันของการเทียบคะแนนวิธีใดคล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐาน ด้วยแนวคิดที่ว่า วิธีการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ซึ่งใช้กับการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนสองค่า เมื่อนำมาดำเนินการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่าเป็นวิธีการที่ไม่ต้องปรับกระบวนการใด ๆ มากนัก ในขณะที่วิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบจำเป็นต้องมีการแก้ไขปรับสูตรต่าง ๆ เช่น การเทียบคะแนนสังเกต จะต้องมีการปรับสูตรที่ใช้สร้างการแจกแจงคะแนนสังเกตเชิงทฤษฎี โดยอาศัยการแจกแจง Multinomial แทนการแจกแจงแบบ binomial หรือการเทียบคะแนนจริงก็คำนวณจากสูตรที่แตกต่างจากเดิม รวมถึงการแปลงพารามิเตอร์ความสามารถให้อยู่ในมาตรวัดเดียวกันก็คำนวณด้วยสูตรที่แตกต่างจากการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนแบบสองค่า ดังนั้น จึงใช้คะแนนแปลงที่ได้จากการเทียบคะแนนเชิงเส้นตรงและการเทียบคะแนนอีควิเปอร์เซ็นไทล์เป็นคะแนนเกณฑ์ และดัชนีคะแนนความแตกต่างที่นำมาใช้มี 6 ดัชนี เพื่อเปรียบเทียบผลที่ได้ในแต่ละดัชนีว่ามีความสอดคล้องกันหรือแตกต่างกันอย่างไร ทั้งความคล้ายคลึงกันของการแจกแจงคะแนนสะสม และค่าดัชนีคะแนนความแตกต่างถือเป็นเกณฑ์ภายในที่ใช้ตรวจสอบผลการเทียบคะแนน ขณะเดียวกันก็ใช้ดัชนีความแตกต่าง (c) จากกลุ่มสอบทานผลซึ่งเป็นดัชนีที่แสดงให้เห็นว่าคะแนนแปลงจากการเทียบคะแนนจากแบบทดสอบฉบับ Y สู่สเกลฉบับ X นั้น แตกต่างจากคะแนนที่ผู้สอบทำได้จริงในฉบับ X เท่าใด ความแตกต่างที่ได้ถือเป็นความคลาดเคลื่อนรวมที่เกิดจากการเทียบคะแนนแต่ละวิธี ดังนั้นการประเมินคุณภาพการเทียบคะแนนโดยใช้ดัชนีความแตกต่าง (c) ซึ่งเป็นเกณฑ์ภายนอก แสดงให้เห็นความคงที่ของการเทียบคะแนน ดังมีประเด็นสำคัญของผลการเทียบคะแนนที่นำมาอภิปรายดังนี้

1. **คะแนนแปลงจากการเทียบคะแนนในแต่ละวิธี** เป็นการเทียบคะแนนจากแบบทดสอบฉบับ Y ปรับสู่สเกลฉบับ X ส่วนใหญ่มีค่าต่ำกว่าคะแนนสมมูลของแบบทดสอบฉบับ Y หรือกล่าวอีกนัยหนึ่งได้ว่าคะแนนสมมูลส่วนใหญ่ของแบบทดสอบฉบับ Y มีค่าสูงกว่าคะแนนของแบบทดสอบฉบับ X ในทุกวิธีการเทียบคะแนน ไม่ว่าจะเทียบคะแนนดิบ คะแนนสังเกต หรือคะแนนจริง โดยผลดังกล่าวที่ได้จะอยู่ในช่วงคะแนนระหว่าง ตำแหน่งเปอร์เซ็นต์ไทล์ที่ 5 ถึง 95 ของคะแนนแบบทดสอบฉบับ Y เมื่อวิเคราะห์ในเชิงเหตุผลพบว่าแบบทดสอบเทียบคะแนนทั้งสองฉบับสร้างจากจุดประสงค์เดียวกัน และกลุ่มตัวอย่างที่ทำแบบทดสอบเป็นกลุ่มสุ่มจากประชากรเดียวกัน แต่ได้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนดิบจากแบบทดสอบฉบับ Y สูงกว่าฉบับ X ขณะเดียวกันค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ได้จากคะแนนสังเกต ซึ่งสร้างตามวิธีการในทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์ของแบบทดสอบฉบับ Y มีค่าสูงกว่าแบบทดสอบฉบับ X เช่นเดียวกันกับค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ได้จากการประมาณค่าคะแนนจริงของผู้สอบตามวิธีการในทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์ในแบบทดสอบฉบับ Y มีค่าสูงกว่าฉบับ X ดังนั้นเมื่อนำคะแนนดิบมาเทียบคะแนนเชิงเส้นตรง ซึ่งเป็นการแปลงคะแนนของแบบทดสอบสองชุดให้เท่าเทียมกัน ณ คะแนนมาตรฐานค่าเดียวกัน (Petersen and others.1982 : 73) นั่นคือเป็นการแปลงคะแนนของแบบทดสอบสองชุดต่อจำนวนเดียวกันของส่วนเบี่ยงเบนมาตรฐานที่อยู่เหนือหรือต่ำกว่าค่าเฉลี่ย (Petersen and others.1989:251) โดยการวิเคราะห์คะแนนทั้งหมด (Angoff.1984) และเป็นการแปลงคะแนนอย่างคงที่ตลอดการแจกแจง (Flanagan. 1951:747-748) การแปลงคะแนนเชิงเส้นตรงจึงเป็นการเปลี่ยนจุดเริ่มต้นของมาตรวัด และหน่วยวัด ซึ่งเป็นการปรับค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน คะแนนที่แปลงแล้วจะไม่เปลี่ยนรูปร่างในการแจกแจงคะแนน จึงทำให้คะแนนสมมูลของแบบทดสอบฉบับ Y สูงกว่าฉบับ X ถึงแม้ว่าในการศึกษาคครั้งนี้ค่าเฉลี่ยของคะแนนแปลงจะแตกต่างจากค่าเฉลี่ยของแบบทดสอบเป้าหมายก็ตาม แต่ส่วนเบี่ยงเบนมาตรฐานเท่ากันกับแบบทดสอบเป้าหมาย ในขณะเดียวกันการเทียบคะแนนเชิงเส้นตรงเป็นวิธีการประมาณคะแนนสมมูลตามนิยามของการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์ที่ดีที่สุด เมื่อการแจกแจงคะแนนของแบบทดสอบเทียบคะแนนมีความคล้ายคลึงกัน (Angoff.1984:93) ดังนั้นคะแนนสมมูลส่วนใหญ่ของแบบทดสอบฉบับ Y จึงมีค่าสูงกว่าแบบทดสอบฉบับ X ซึ่งไม่ตลอดมาตรวัดดังเช่นการเทียบคะแนนเชิงเส้นตรงเพราะวิธีการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์เป็นการเทียบคะแนนเชิงเส้นโค้ง จึงมีความไม่สม่ำเสมอในคะแนนแปลงที่ได้ และเมื่อพิจารณาผลการเทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบ ซึ่งเทียบคะแนนทั้งคะแนนจริงและคะแนนสังเกต โดยการเทียบคะแนนสังเกตทุกวิธีต้องอาศัยหลักการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์จึงส่งผลให้คะแนนสมมูลส่วนใหญ่ของแบบทดสอบฉบับ Y สูงกว่าคะแนนของแบบทดสอบฉบับ X และสำหรับการเทียบคะแนนจริงซึ่งได้ผลเช่นเดียวกันอาจจะเนื่องจากการใช้ข้อสนเทศของข้อสอบและผู้สอบ

ประมาณค่าคะแนนจริงของผู้สอบแต่ละฉบับ จึงทำให้ผลที่ได้สอดคล้องกับการเทียบคะแนนสังเกต

2. การประเมินผลการเทียบคะแนนด้วยความคล้ายคลึงกันของการแจกแจงคะแนนสะสม เป็นเกณฑ์ประเมินผลการเทียบคะแนนรูปแบบหนึ่ง (Kolen and Brennan. 1995:24) ที่ใช้คุณสมบัติของการเทียบคะแนนสังเกต (Observed Score Equating Property) ที่ว่าคะแนนแปลงจากการเทียบคะแนนสังเกตของแบบทดสอบฉบับ Y ปรับสู่สเกลฉบับ X มีการแจกแจงคะแนนสะสมที่เหมือนกันกับการแจกแจงคะแนนสังเกตสะสมของแบบทดสอบฉบับเป้าหมาย (ฉบับ X) และคุณสมบัติ Weak Equity ที่ใช้กับการเทียบคะแนนจริงที่ว่าคะแนนแปลงจากการเทียบคะแนนจริงจากแบบทดสอบฉบับ Y ปรับสู่สเกลฉบับ X มีการแจกแจงคะแนนสะสมที่เหมือนกันกับการแจกแจงคะแนนจริงสะสมของแบบทดสอบเป้าหมาย (ฉบับ X) จากการตรวจสอบพบว่า การแจกแจงคะแนนสะสมของคะแนนแปลงจากการเทียบคะแนนทุกวิธีมีความคล้ายคลึงกับการแจกแจงคะแนนสะสมของแบบทดสอบฉบับเป้าหมาย คือ การแจกแจงคะแนนสะสมของคะแนนแปลงจากการเทียบคะแนนเชิงเส้นตรง อีควิเปอร์เซ็นไทล์ และการเทียบคะแนนสังเกตตามวิธีในทฤษฎีการตอบข้อสอบมีความคล้ายคลึงกับการแจกแจงคะแนนสะสมของแบบทดสอบฉบับ X และคะแนนแปลงซึ่งเป็นคะแนนจริงที่เทียบจากแบบทดสอบฉบับ Y ปรับสู่สเกลฉบับ X ด้วยการเทียบคะแนนจริงทุกวิธีมีการแจกแจงคะแนนสะสมคล้ายคลึงกับการแจกแจงคะแนนจริงในแบบทดสอบฉบับ X ผลที่ได้สอดคล้องกับสมมุติฐานในข้อ 1 ที่ว่า การแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนของแบบทดสอบฉบับเป้าหมายไม่แตกต่างกัน ที่เป็นเช่นนี้เพราะว่าแบบแผนการเทียบคะแนนแบบกลุ่มสุ่ม ซึ่งหน่วยตัวอย่างสุ่มจากประชากรเดียวกัน และในการสอบได้ดำเนินการจัดให้ผู้สอบได้รับแบบทดสอบคนละฉบับแบบสุ่ม ทำให้กลุ่มตัวอย่างเทียบคะแนนที่ทำแบบทดสอบคนละฉบับมีความสามารถใกล้เคียงกัน และแบบทดสอบที่ใช้ในการเทียบคะแนนทั้งสองฉบับสร้างจากตารางวิเคราะห์เนื้อหาเดียวกันและมีค่าสถิติของแบบทดสอบใกล้เคียงกัน โดยเฉพาะความเชื่อมั่นของแบบทดสอบทั้งสองฉบับมีค่าสูงเท่าเทียมกัน คือ 0.937 และ 0.940 ทำให้คะแนนแปลงของแบบทดสอบฉบับ Y ที่ได้จากการเทียบคะแนนอีควิเปอร์เซ็นไทล์มีการแจกแจงคะแนนสะสมที่เหมือนกันกับการแจกแจงคะแนนสะสมฉบับ X ซึ่งเป็นคุณสมบัติของการเทียบคะแนนอีควิเปอร์เซ็นไทล์ และเมื่อวิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์มีคุณสมบัติดังกล่าว วิธีการเทียบคะแนนเชิงเส้นตรงมี คุณสมบัติดังกล่าวด้วย (Kolen and Brennan.1995:12) นอกจากนี้แบบทดสอบเทียบคะแนนทั้งสองฉบับซึ่งเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ที่วัดความสามารถด้านเหตุผลเป็นแบบทดสอบที่วัดลักษณะเด่นเพียงหนึ่งลักษณะทำให้มีคุณสมบัติพื้นฐานทางการวัดสอดคล้องกับคุณสมบัติตามทฤษฎีการตอบข้อสอบ และเมื่อแบบทดสอบเทียบคะแนนทั้งสองฉบับมีความเป็นเอกมิติแล้วก็เป็นเงื่อนไขที่เพียงพอว่าแบบทดสอบนั้นมีความเป็นอิสระต่อกันในการตอบข้อสอบ (Warm.1978:107) มีผลให้การเทียบคะแนน

สังเกตตามวิธีการในทฤษฎีนี้มี คุณสมบัติของการเทียบคะแนนสังเกตด้วย และประกอบกับการเทียบคะแนนจริงเป็นการเทียบคะแนนที่มีคุณสมบัติ Weak equity อยู่แล้ว (Kolen and Brennan .1995:11) จึงทำให้เกิดความคล้อยคลึงกันของการแจกแจงคะแนนสะสมทุกวิธีการเทียบคะแนน

3. การประเมินผลการเทียบคะแนนด้วยดัชนีคะแนนความแตกต่างเป็นเกณฑ์ที่ แองกออฟฟี (Harris and Crouse. 1993 : 198) ได้เสนอไว้ ดัชนีนี้เมื่อใช้ในงานวิจัยที่จำลอง ข้อมูลจะใช้คะแนนที่แท้จริง (true value) ซึ่งเป็นค่าที่ได้จากประชากรข้อมูลเป็นคะแนนเกณฑ์ หรือเมื่อเทียบคะแนนกลับสู่ฉบับเดิม (equating test to itself) คะแนนเดิมที่ไม่ผ่านการเทียบ คะแนนจะเป็นคะแนนเกณฑ์ ดัชนีนี้จึงเป็นดัชนีที่ใช้ประเมินความคลาดเคลื่อนในเชิงสถิติ (Harris and Crouse. 1993 : 203 ;citing Skaggs and Lissitz.1986) และยังใช้เปรียบเทียบ ความแตกต่าง ระหว่างคะแนนแปลงที่ได้จากการเทียบคะแนนสองวิธีใดๆ (Harris and Crouse.1993:203) ซึ่งดัชนีคะแนนความแตกต่างที่ใช้ในการศึกษานี้เป็นการเปรียบเทียบความแตกต่างระหว่างคะแนนแปลงจากวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบกับวิธีการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมซึ่งเป็นวิธีการเทียบคะแนนพื้นฐานที่ใช้ เป็นเกณฑ์ว่าความแตกต่างในแต่ละวิธีมีมากน้อยเพียงใดโดยใช้ดัชนีทั้งสิ้น 6 ดัชนี เป็นดัชนีที่ ถ่วงน้ำหนักด้วยความถี่ของคะแนนจากแบบทดสอบฉบับ Y และดัชนีไม่ถ่วงน้ำหนักสามารถ แยกอภิปรายได้ 3 ประเด็น คือ

3.1 ความแตกต่างระหว่างคะแนนแปลงผลที่ได้มีความสอดคล้องกันทั้ง 6 ดัชนี คือฟังก์ชันการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบมีความแตกต่างจากฟังก์ชัน การเทียบคะแนนเชิงเส้นตรงมากที่สุด แต่ผลที่ได้จากการเทียบคะแนนจริงไม่ส่งผลเช่นนั้นและ ความแตกต่างระหว่างวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบทุกวิธีกับการเทียบคะแนน เชิงเส้นตรงมีค่ามากกว่าความแตกต่างเมื่อใช้คะแนนแปลงจากวิธีการเทียบคะแนนอีควิเปอร์- เซนไทล์เป็นเกณฑ์ในทุกดัชนี ที่เป็นเช่นนั้นเพราะว่ามีตัวแปรเกี่ยวกับคะแนนที่ใช้ในการเทียบ เข้ามาเกี่ยวข้อง คือการรายงานผลการเทียบคะแนนทั้งคะแนนจริงและคะแนนสังเกต ซึ่งนิยม ในการเทียบคะแนนของทั้งสองรูปแบบแตกต่างกันคือการเทียบคะแนนสังเกตนั้นเป็นการ รายงานผลของคะแนนที่เทียบเท่ากันโดยอาศัยหลักการเทียบคะแนนของอีควิเปอร์เซนไทล์ โดย ปรับสเกลของมาตรฐานวัดความสามารถของผู้สอบทั้งสองฉบับให้อยู่บนมาตรฐานเดียวกันด้วยวิธี SIM MS และ TRF ก่อนแล้วจึงนำผลที่ได้ไปสร้างการแจกแจงคะแนนสังเกตเชิงทฤษฎีของ แบบทดสอบทั้งสองฉบับ ส่วนการเทียบคะแนนจริงก็ดำเนินการประมาณค่าคะแนนจริงของ ผู้สอบที่เทียบเท่ากัน หลังจากแปลงคะแนนความสามารถให้อยู่บนสเกลเดียวกันตามวิธี SIM MS และ TRF ดังนั้นรูปแบบของการเทียบคะแนนจึงส่งผลต่อคะแนนแปลงที่ได้ทำให้คะแนน แปลงของการเทียบคะแนนสังเกตในทั้ง 3 วิธีมีความใกล้เคียงกันกับคะแนนแปลงของอีควิเปอร์- เซนไทล์มากกว่าเชิงเส้นตรง

3.2 ความคล้ายคลึงกันระหว่างคะแนนแปลงจากผลการวิจัยนี้พบว่าวิธี MS เมื่อเทียบคะแนนจริงให้ฟังก์ชันการเทียบคะแนนที่คล้ายคลึงกับฟังก์ชันการเทียบคะแนนอีควิเปอร์-เซ็นไทล์มากที่สุด ซึ่งให้ผลที่สอดคล้องกันทั้ง 4 ดัชนี คือดัชนี RMSD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก ดัชนี MAD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก และคล้ายคลึงกับการเทียบคะแนนเชิงเส้นตรงในดัชนี MSD ไม่ถ่วงน้ำหนัก การเทียบคะแนนสังเกตวิธี SIM ให้ฟังก์ชันการเทียบคะแนนที่คล้ายคลึงกับวิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์มากที่สุด ในดัชนี MSD ถ่วงน้ำหนักจะเห็นได้ว่าผลที่ได้ยังไม่สอดคล้องกันทั้ง 6 ดัชนี ที่เป็นเช่นนี้เพราะว่าการเทียบคะแนนเชิงเส้นตรงเป็นการเทียบคะแนนที่ให้ผลคงที่ตลอดมาตรวัดคือคะแนนแปลงที่ได้มีค่าต่ำกว่าคะแนนสมมูลของฉบับ Y แต่การเทียบคะแนนจริงให้คะแนนแปลงที่ไม่คงที่ตลอดมาตรวัด ดังนั้นเมื่อใช้ดัชนี MSD ซึ่งคำนึงถึงเครื่องหมายของความแตกต่าง เมื่อนำมารวมกันทำให้ผลรวมที่ได้มีขนาดเล็กลงและอาจเนื่องจากบริบทเกี่ยวกับรูปแบบของการเทียบคะแนนจริงและคะแนนสังเกตเข้ามาเกี่ยวข้องดังกล่าวไว้ใน 3.1 ทำให้สรุปได้ไม่ชัดเจนว่าคะแนนแปลงจากวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบวิธีใดให้ผลสอดคล้องกับคะแนนพื้นฐานมากที่สุด

3.3 ความสอดคล้องระหว่างดัชนีคะแนนความแตกต่างเมื่อพิจารณาค่าดัชนี RMSD และ MAD พบว่าดัชนีแต่ละตัวให้ผลที่สอดคล้องกันทั้งดัชนีที่ถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก เหตุที่ดัชนีแต่ละตัวให้ผลที่สอดคล้องกันระหว่างถ่วงน้ำหนักและไม่ถ่วงน้ำหนักอาจเป็นเพราะคะแนนดิบที่ได้จากสถานการณ์การทดสอบจริงทั้งฉบับ X และ Y มีคะแนนไม่ตลอดมาตรวัดคือมีค่าคะแนนต่ำสุดอยู่ที่ 60 และ 59 คะแนนตามลำดับ จนถึงคะแนนสูงสุดคือ 150 ทั้งสองฉบับ และเมื่อสร้างการแจกแจงคะแนนสังเกตเชิงทฤษฎีตามวิธีการในทฤษฎีการตอบข้อสอบก็ยังพบว่าคะแนนต่ำสุดของแบบทดสอบทั้งสองฉบับที่สามารถคำนวณค่าความน่าจะเป็นที่จะเกิดระดับคะแนนนี้ได้อยู่ที่ระดับคะแนน 44 และ 45 ทำให้ได้คะแนนไม่ตลอดมาตรวัดเช่นกันที่เป็นเช่นนี้อาจจะเป็นเพราะเกณฑ์การตรวจให้คะแนนหลายค่าที่กำหนดขึ้น ซึ่งมีลำดับคะแนนที่ใช้ในการศึกษาครั้งนี้ถึง 6 ลำดับขั้น ซึ่งในลำดับขั้นคะแนนศูนย์ของทุกข้อมีสัดส่วนความถี่ค่อนข้างน้อยและอาจจะเป็นด้วยธรรมชาติของแบบทดสอบลักษณะนี้ด้วย ดังนั้นการศึกษาดัชนีคะแนนความแตกต่างชนิดไม่ถ่วงน้ำหนักจึงกำหนดคะแนนต่ำสุดในการพิจารณาไว้ที่คะแนนต่ำสุดของแบบทดสอบฉบับ Y คือ 59 คะแนนจนถึงคะแนน 150 เพราะการเทียบคะแนนอีควิเปอร์เซ็นไทล์ในกรณีนี้จะต้องทำ extrapolation ในช่วงคะแนน 1-58 ซึ่งเป็นช่วงคะแนนที่ไม่มีความถี่และสำหรับการเทียบคะแนนสังเกตตามวิธีการในทฤษฎีการตอบข้อสอบก็เช่นเดียวกันต้องทำ extrapolation จาก 1-44 คะแนน ซึ่งจะมีผลต่อคะแนนแปลงที่ได้จากวิธีการเหล่านี้มีค่าความคลาดเคลื่อนสูงเกินกว่าที่เป็นจริง ทำให้ผลที่ได้จากดัชนีคะแนนความแตกต่างทั้งถ่วงน้ำหนัก และไม่ถ่วงน้ำหนักสอดคล้องกัน สำหรับดัชนี MSD ที่ให้ผลไม่สอดคล้องกันระหว่างถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก อาจเป็นเพราะการนำเอาเครื่องหมายของความแตกต่างมาคิดทำให้ขนาดความแตกต่างมีค่าเล็กลง

ประเด็นที่น่าสังเกตในการวิจัยนี้คือคะแนนที่ใช้ในการเทียบคะแนนครั้งนี้มีไม่ตลอดมาตรวัด (59-150) คะแนน ซึ่งส่งผลต่อวิธีการเทียบคะแนนที่อาศัยการแจกแจงคะแนนเป็นพื้นฐานในการเทียบคะแนนได้แก่ การเทียบคะแนนอีควิเปอร์เซ็นไทล์ และการเทียบคะแนนสังเกตตามการเทียบคะแนนในทฤษฎีการตอบข้อสอบที่ทำให้ตารางเทียบคะแนนจากวิธีการดังกล่าว ไม่สามารถระบุคะแนนแปลงที่ถูกต้องในช่วงคะแนนดิบ 1-59 และช่วงคะแนนสังเกต 1-44 ได้ ซึ่งเป็นจุดด้อยของการเทียบคะแนนตามวิธีดังกล่าว (แต่คะแนนแปลงที่ได้จากวิธีการเทียบคะแนนอื่น ๆ ยังคงใช้ได้ตลอดมาตรวัด) แต่ตารางเทียบคะแนนดังกล่าวสามารถนำไปใช้อ้างอิงได้ เพราะผู้สอบส่วนใหญ่มักจะได้คะแนนอยู่ในช่วง 60-150 คะแนน ก็สามารถนำคะแนนดังกล่าวมาเทียบคะแนนจากตารางได้ และเมื่อตัดสินผลการสอบ เช่น การตัดสินให้ผ่านตั้งแต่เปอร์เซ็นไทล์ 25 ก็สามารถนำคะแนนแปลงที่ได้มากำหนดตำแหน่งเปอร์เซ็นไทล์ให้ตัดสินผลได้อย่างเที่ยงตรงในการสอบแต่ละครั้งได้เช่นกัน แต่ถ้าหากต้องการสร้างตารางเทียบคะแนนที่มีคะแนนตลอดมาตรวัดการแก้ไขปัญหาดังกล่าวอาจดำเนินการโดยใช้การแปลงคะแนนแบบทดสอบแต่ละฉบับให้เป็นคะแนนปกติ เช่นคะแนนซีปกติ (normalized-score) หรือคะแนนทีปกติ (normalize T score) ก่อนแล้วจึงนำคะแนนปกติที่ได้มาเทียบคะแนนต่อไป โดยจะต้องศึกษาว่าวิธีการแปลงคะแนนให้เป็นคะแนนปกติวิธีใดมีความสอดคล้องกับข้อมูลผลการสอบเพื่อมิให้เกิดความคลาดเคลื่อนของคะแนนแปลงมีค่าสูงเกินไปการใช้คะแนนปกติมาเทียบคะแนนไม่ได้อยู่ในขอบข่ายการศึกษาครั้งนี้ ซึ่งผลที่ได้อาจจะมีความคลาดเคลื่อนของการเทียบคะแนนมากขึ้นกว่ามาเติมหรือไม่ต้องศึกษาต่อไป

4. การประเมินผลการเทียบคะแนนด้วยค่าดัชนีความแตกต่าง (C) ที่วิเคราะห์จากกลุ่มสอบทานผล ประเมินโดยใช้เกณฑ์ของปีเตอร์เซนและคณะ (Petersen and others. 1982:93-94) พบว่าความคลาดเคลื่อนรวมของการเทียบคะแนนแต่ละวิธีเมื่อเทียบกับเกณฑ์แล้วน้อยกว่ากำลังสองของร้อยละ 5 ของส่วนเบี่ยงเบนมาตรฐานของแบบทดสอบฉบับ X ซึ่งอยู่ในระดับที่น่าพอใจอย่างมากทุกวิธี และความคลาดเคลื่อนของการเทียบคะแนนทุกวิธีแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมุติฐานการวิจัยข้อ 2 และข้อ 3 ที่ว่า ดัชนีความแตกต่างที่ได้จากการเทียบคะแนน ตามทฤษฎีการตอบข้อสอบมีคุณภาพอยู่ในระดับที่สูงกว่าที่ได้จากการเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิมและความคลาดเคลื่อนของการเทียบคะแนนแต่ละวิธีมีความแตกต่างกัน ที่เป็นเช่นนี้หากวิเคราะห์ในเชิงเหตุผลพบว่าองค์ประกอบของการเทียบคะแนนในแนวนอนแบบแผนกลุ่มสัมพันธ์ทำให้เกิดความคล้ายคลึงกันความสามารถของกลุ่มตัวอย่างเทียบคะแนน และแบบทดสอบเทียบคะแนนทั้งสองฉบับยังมีความคล้ายคลึงกันในเรื่องเนื้อหาและคำสถิติทำให้การแจกแจงคะแนนของแบบทดสอบเทียบคะแนนทั้งสองฉบับคล้ายคลึงกันส่งผลให้การเทียบคะแนนเชิงเส้นตรง การเทียบคะแนนอีควิเปอร์เซ็นไทล์ และการเทียบคะแนนสังเกตทุกวิธี ซึ่งเทียบคะแนนโดยอาศัยวิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์มีคุณภาพอยู่ในระดับที่น่าพอใจอย่างมาก และเมื่อพิจารณาองค์ประกอบ

เกี่ยวกับข้อสอบรวมยังพบว่าข้อสอบรวมนั้นถูกสร้างให้วัดในเนื้อหาเดียวกันกับแบบทดสอบเทียบคะแนน และเป็นฉบับย่อส่วนรวมทั้งจำนวนข้อสอบรวมมีจำนวน 10 ข้อ คิดเป็นร้อยละ 33 ของจำนวนข้อสอบทั้งฉบับ ถึงแม้จะน้อยกว่าจำนวนที่แองกอฟฟ์ กำหนดไว้ (Angoff, 1984:107) แต่เป็นจำนวนที่สอดคล้องกับไรท์ และ สโตน (Wright and Stone, 1979:98) ได้เสนอไว้ และคิดว่่าน่าจะเป็นจำนวนที่ให้สารสนเทศได้อย่างพอเพียง เพราะข้อสอบแต่ละข้อมีคะแนนเต็ม 5 คะแนน สารสนเทศในแต่ละข้อจึงมากกว่าข้อสอบที่ตรวจให้คะแนนสองค่า นอกจากนี้ข้อสอบรวมยังวางในตำแหน่งเดียวกัน ในแบบทดสอบเทียบคะแนนทั้งสองฉบับคือเป็นข้อสอบข้อที่ 2 3 4 12 16 17 22 25 29 และ 30 เป็นการหลีกเลี่ยงการทำหน้าที่ต่างกันของข้อสอบ และค่าเฉลี่ยของคะแนนข้อสอบที่ได้จากผู้สอบทั้ง 2 กลุ่มมีค่าแตกต่าง 0.11 ซึ่งสอดคล้องกับประสบการณ์ของโคเลน (Kolen and Brennan, 1995:263) ที่พบว่าผลการเทียบคะแนนในวิธีการที่แตกต่างกันจะไม่มี ความแตกต่างกันเมื่อความแตกต่างของค่าเฉลี่ยของข้อสอบรวมจากผู้สอบทั้งสองกลุ่มมีค่าไม่เกิน 0.1 และเมื่อพิจารณากลุ่มสอบทานผล ซึ่งเป็นกลุ่มตัวอย่างที่ใช้ประเมินคุณภาพเทียบคะแนนในครั้งนี้ ที่โคเลนและวิทนี (Kolen and Whitney, 1982 : 287 - 291) ได้แนะนำให้ใช้ ข้อมูลคะแนนจากผู้สอบเป็นเกณฑ์ในการหาความแตกต่าง ซึ่งสุ่มจากประชากรชุดเดียวกันกับกลุ่มตัวอย่างเทียบคะแนนในจำนวนไม่น้อยกว่าร้อยละ 10 ของกลุ่มตัวอย่างเทียบคะแนน ซึ่งเป็นจำนวนที่เหมาะสม (Kolen and Whitney, 1982 : 284) และได้รับการทดสอบจากแบบทดสอบเทียบคะแนนทั้งสองฉบับโดยสุ่มอย่างมีระบบ ให้ผู้สอบแต่ละคนทำสอบแบบทดสอบฉบับ X แล้วตามด้วยฉบับ Y หรือสอบด้วยฉบับ Y ตามด้วยฉบับ X เป็นการป้องกันความคลาดเคลื่อนในการสอบ เพื่อให้คะแนนของแบบทดสอบทั้งสองฉบับได้รับผลกระทบอย่างเท่าเทียมกันในปัจจัยของการเรียนรู้ และความเหนื่อยล้า และค่าสถิติพื้นฐานของแบบทดสอบอันได้แก่ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแบบทดสอบฉบับ Y สูงกว่าฉบับ X เป็นลักษณะที่สอดคล้องกับค่าสถิติพื้นฐานของกลุ่มเทียบคะแนน ทำให้คะแนนของกลุ่มสอบทานผลถือเป็นคะแนนเกณฑ์ที่ดีที่จะนำมาใช้ในการตรวจสอบคุณภาพ

5. ความสอดคล้องของผลการเทียบคะแนน ในวิจัยครั้งนี้ได้ใช้เกณฑ์การประเมินผลการเทียบคะแนน 3 เกณฑ์ คือ ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ดัชนีคะแนนความแตกต่าง ซึ่งทั้งสองเป็นเกณฑ์ภายใน และดัชนีความแตกต่างซึ่งวิเคราะห์จากกลุ่มสอบทานผลเป็นเกณฑ์ภายนอก ซึ่งให้ผลการประเมินที่สอดคล้องกัน คือมีความคล้ายคลึงกันของการแจกแจงคะแนนสะสมจากการเทียบคะแนนทุกวิธีไม่ว่าจะเป็นการเทียบคะแนนจริง คะแนนสังเกต มีคุณสมบัติของการเทียบคะแนน คือการแจกแจงคะแนนสะสมระหว่างคะแนนแปลงกับคะแนนจากแบบทดสอบเป้าหมายคล้ายคลึงกัน สอดคล้องกับผลที่ได้จากการประเมินด้วยดัชนีความแตกต่างคือผลการเทียบคะแนนแต่ละวิธีมีคุณภาพอยู่ในระดับเดียวกัน คือระดับน่าพอใจอย่างมากและเมื่อทดสอบความแตกต่างของความคลาดเคลื่อนสัมบูรณ์ในแต่ละระดับคะแนนของแต่ละวิธีพบว่าไม่แตกต่างกัน สอดคล้องกับเกณฑ์ประเมินคุณภาพของปีเตอร์เซ็น

และคณะ และเมื่อพิจารณาค่าดัชนีความแตกต่างแต่ละค่าจะพบว่า การเทียบคะแนนเชิงเส้นตรง ให้ความคลาดเคลื่อนในการเทียบคะแนนสูงสุดคือ 0.192 ในขณะที่การเทียบคะแนนสังเกตทั้ง 3 วิธีคือวิธี MS TRF และ SIM มีค่าความคลาดเคลื่อนในการเทียบคะแนนเท่ากับ 0.174 0.173 และ 0.173 ใกล้เคียงกัน ค่าดัชนีความแตกต่างที่ได้จากการเทียบคะแนนอีควิเปอร์เซ็นไทล์ คือ 0.179 สำหรับการเทียบคะแนนจริงทั้ง 3 วิธี ได้แก่ วิธี MS TRF และ SIM ให้ผลที่ใกล้เคียงกัน ทุกวิธี คือ มีค่าดัชนีความแตกต่าง 0.181 0.182 และ 0.184 จะเห็นว่า ผลที่ได้จากดัชนีความแตกต่าง ให้ความสอดคล้องกับผลที่ได้จากการวิเคราะห์ด้วยดัชนีคะแนนความแตกต่าง ที่สรุปว่า การเทียบคะแนนสังเกตวิธี TRF มีความแตกต่างจากวิธีการเทียบคะแนนเชิงเส้นตรงมากที่สุด และการเทียบคะแนนจริงวิธี MS มีฟังก์ชันการเทียบคะแนนที่คล้ายคลึงกับฟังก์ชันการเทียบคะแนน อีควิเปอร์เซ็นไทล์ สรุปได้ผลว่าการวิจัยครั้งนี้การเทียบคะแนนแต่ละวิธีให้ผลที่คล้ายคลึงกันที่เป็นเช่นนี้สามารถอธิบายได้ว่า เพราะการตรวจสอบความคลาดเคลื่อนกระทำอยู่บนช่วงคะแนนสังเกตที่วัดได้คือ ช่วง 59-150 คะแนน ไม่ได้ตรวจสอบความคลาดเคลื่อนที่อยู่ในช่วงคะแนนต่ำ (1-58) ซึ่งไม่มีผู้สอบทำได้จึงทำให้ผลที่ได้จากเกณฑ์ทั้งสามสอดคล้องกัน

นอกจากนี้ ผลการเทียบคะแนนที่ได้จากการศึกษาครั้งนี้ยังให้ผลที่สอดคล้องกับงานวิจัยเกี่ยวกับการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่าของคิมและโคเฮน (Kim and Cohen.1995 : 167-176) ซึ่งเป็นการศึกษาในสถานการณ์จำลองข้อมูล เทียบคะแนนความสามารถด้วยโมเดลเกรต เรสพอนส์ ผลที่ได้สอดคล้องกันคือการเทียบคะแนนวิธี MS และ TRF ให้ผลที่คล้ายคลึงกัน และสอดคล้องกับงานวิจัยของลีและคณะ (Lee and others. 2001 : 1-30) ซึ่งเป็นการศึกษาในสถานการณ์ทดสอบจริงในแบบทดสอบที่ประกอบด้วยแบบทดสอบฉบับย่อย แบบแผนกลุ่มสุ่ม โดยเทียบทั้งคะแนนจริงและคะแนนสังเกต ในโมเดลเกรต เรสพอนส์ และโมเดลนอมีนอล เรสพอนส์ วิธีโค้งลักษณะแบบทดสอบและเทียบคะแนนตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้แก่การเทียบคะแนนเชิงเส้นตรง อีควิเปอร์เซ็นไทล์ และค่าเฉลี่ย พบว่า การเทียบคะแนนในวิธีต่างกันให้ผลที่คล้ายคลึงกันเมื่อแบบทดสอบมีการฝ่าฝืนข้อตกลงเบื้องต้นของทฤษฎี IRT น้อยที่สุด และยังสอดคล้องกับงานวิจัยของฮันและเฟอรรารา (HuYuh and Ferrara.1994:125-141) ที่เทียบคะแนนในสถานการณ์ทดสอบจริงผลที่ได้จากการเทียบคะแนนจริงในโมเดลพาสเซียล เครดิตและปรับสเกลด้วยวิธีโค้งลักษณะแบบทดสอบ กับการเทียบคะแนน อีควิเปอร์เซ็นไทล์ให้ผลที่เหมือนกัน เมื่อแบบทดสอบมีความยากง่ายปานกลาง และกลุ่มผู้สอบมีความสามารถใกล้เคียงกัน ถึงแม้ว่าผลที่ได้ในครั้งนี้อาจจะไม่สอดคล้องกับงานวิจัยของบาสทารี (Bastari.2000 : 1-78) ซึ่งเป็นงานวิจัยในสถานการณ์จำลองข้อมูลเทียบคะแนนความสามารถในแบบทดสอบข้อสอบผสมระหว่างข้อสอบเลือกตอบและข้อสอบที่ตรวจให้คะแนนหลายค่าพบว่า วิธีการปรับค่าพารามิเตอร์พร้อมกันให้ผลการเทียบคะแนนที่ดีกว่าการเทียบโดยใช้โค้งลักษณะแบบทดสอบในหลายองค์ประกอบที่ศึกษา ซึ่งบาสทารีได้ตั้งข้อสังเกตเกี่ยวกับสัดส่วนระหว่าง

ข้อสอบเลือกตอบกับข้อสอบที่ตรวจให้คะแนนหลายค่าว่าอาจไม่เหมาะสมและอาจจะเป็นตัวแปรหนึ่งที่ส่งผลให้การเทียบคะแนนวิธีปรับค่าพารามิเตอร์พร้อมกันดีกว่า

ข้อเสนอแนะ

จากการศึกษาผลการเทียบคะแนนที่ข้อสอบตรวจให้คะแนนหลายค่าเมื่อใช้วิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบและวิธีการตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ผู้วิจัยมีข้อเสนอแนะในการนำผลการวิจัยไปใช้และในการทำวิจัย ดังนี้

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 ข้อเสนอแนะเชิงนโยบาย

ผลการวิจัยครั้งนี้พบว่า การเทียบคะแนนแบบทดสอบตามทฤษฎีการทดสอบแบบมาตรฐานเดิมได้แก่การเทียบคะแนนเชิงเส้นตรง การเทียบคะแนนอีควิเปอร์เซ็นไทล์ และการเทียบคะแนนตามวิธีการในทฤษฎีการตอบข้อสอบ ได้แก่ วิธีปรับค่าพารามิเตอร์พร้อมกัน (SIM) วิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก (MS) และวิธีโค้งลักษณะแบบทดสอบ (TRF) ซึ่งเทียบทั้งคะแนนจริงและคะแนนสังเกตมีคุณภาพอยู่ในระดับเดียวกันคือระดับน่าพอใจอย่างมากทุกวิธี ในบริบทของการเทียบคะแนนแนวนอนของแบบแผนการรวบรวมข้อมูลที่ใช้ข้อสอบรวมในกลุ่มสุ่ม ซึ่งเป็นกลุ่มตัวอย่างจากประชากรเดียวกัน เช่น นักเรียนในชั้นเรียนเดียวกัน ปีการศึกษาเดียวกัน แต่สอบด้วยแบบทดสอบคนละฉบับหรือในการทดสอบที่มีหลายครั้งในแต่ละปีการศึกษา จำเป็นต้องใช้แบบทดสอบหลายฉบับ และแบบทดสอบที่นำมาใช้แต่ละฉบับจะต้องวัดในเนื้อหาและมีค่าสถิติที่ใกล้เคียงกัน ดังนั้นในทางทฤษฎีถือว่าวิธีการเทียบคะแนนทั้ง 5 วิธี 8 รูปแบบให้ผลที่เหมือนกันโดยส่วนใหญ่ไม่มีเหตุผลที่ต้องเลือกใช้ แต่ในทางปฏิบัติวิธีการที่ง่ายที่สุดในการดำเนินการคือวิธีการเทียบคะแนนเชิงเส้นตรง รองลงมาได้แก่วิธีการเทียบคะแนนอีควิเปอร์เซ็นไทล์ และสุดท้ายคือวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ ซึ่งวิธีการที่ง่ายที่สุดก็คือวิธีการปรับค่าพารามิเตอร์พร้อมกัน แต่การเทียบคะแนนในทฤษฎี IRT มีรูปแบบในการรายงาน 2 รูปแบบคือ การเทียบคะแนนจริง ซึ่งประมาณค่าได้จากสูตรที่ใช้ในการคำนวณซึ่งอาจจะคลาดเคลื่อนอยู่บ้าง และการเทียบคะแนนสังเกตเป็นการเทียบคะแนนที่ต้องสร้างการแจกแจงคะแนนสังเกตในเชิงทฤษฎีตามวิธีการในทฤษฎีการตอบข้อสอบ ซึ่งมีความซับซ้อนมากกว่าแต่สามารถสื่อสารกับผู้นำผลไปใช้ได้ง่ายกว่า การใช้วิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ เมื่อเลือกวิธีการแล้วควรเลือกรูปแบบการรายงานผลการเทียบคะแนนด้วย ดังนั้นการเลือกวิธีการเทียบคะแนนวิธีใด จะต้องพิจารณาว่าสถานการณ์ทดสอบที่จะนำไปใช้สอบมีบริบทที่สอดคล้องกับบริบทของการเทียบคะแนนดังกล่าวข้างต้น

การเลือกใช้วิธีการเทียบคะแนนวิธีใด เป็นการทำให้กระบวนการเทียบคะแนนมีคุณภาพมากขึ้นแต่เหตุผลสำคัญกว่านั้นก็คือการนำกระบวนการเทียบคะแนนไปใช้ก่อนการตัดสินผลเมื่อคะแนนที่ได้จากการสอบมาจากคะแนนแบบทดสอบต่างฉบับกันมิใช่คะแนนดิบที่ได้แต่ละครั้งมาเปรียบเทียบกันโดยตรง โดยมีได้คำนึงถึงความแตกต่างของแบบทดสอบที่ใช้แต่ละครั้งทำให้เกิดความได้เปรียบเสียเปรียบของผู้สอบ ดังนั้นผู้มีหน้าที่จัดการศึกษาควรจะต้องคำนึงถึงความยุติธรรมที่ต้องมีให้แก่ผู้สอบทุกคนและความถูกต้องในการนำผลไปใช้เพื่อเป็นมาตรฐานให้การตัดสินผลมีความแม่นยำและเที่ยงตรงถึงแม้ว่าแบบทดสอบที่นำมาใช้ในแต่ละครั้งจะวัดในเนื้อหาเดียวกันและมีค่าความยากทางสถิติที่เท่าเทียมกันก็ตาม แต่ความยากของข้อสอบมิได้มีความหมายเพียงแค่ว่าสถิติแต่ยังรวมถึงการเผชิญหน้ากับข้อสอบที่วัดเนื้อหาเดียวกันแต่ข้อคำถามต่างกันด้วย

1.2 ข้อเสนอแนะในการนำไปปฏิบัติ

1.2.1 ผลการวิจัยที่ได้ศึกษาในครั้งนี้หากนำไปใช้กับโครงการทดสอบขนาดใหญ่ที่ต้องประเมินความรู้ของนักเรียนในแต่ละระดับช่วงชั้น เพื่อเป็นส่วนหนึ่งของคะแนนที่จะพิจารณาบุคคลเข้าศึกษาต่อในระดับที่สูงขึ้นต่อไป การทดสอบลักษณะนี้ไม่ควรใช้แบบทดสอบเพียงฉบับใดฉบับหนึ่ง เพราะทำให้การควบคุมการสอบในห้องสอบเดียวกันดำเนินการได้ลำบากและต้องทำในช่วงระยะเวลาเดียวกัน จึงควรใช้แบบทดสอบหลายฉบับที่วัดในโครงสร้างเดียวกันมีแบบทดสอบร่วมในแต่ละฉบับ ซึ่งเป็นรูปแบบที่ทำให้ผลการเทียบคะแนนมีความแม่นยำมากที่สุด หากไม่สามารถทำได้เนื่องจากปัญหาการติวข้อสอบก็ใช้แบบทดสอบที่มีความคู่ขนานในโครงสร้างเนื้อหา และดำเนินการสอบในห้องสอบอย่างรัดกุมเพื่อกระจายผู้สอบในการทำแบบทดสอบต่างฉบับ วิธีการเทียบคะแนนที่เหมาะสมควรนำมาใช้คือวิธีการเทียบคะแนน อีคิวเปอร์เซ็นไทล์เพราะการเทียบคะแนนในกลุ่มตัวอย่างขนาดใหญ่เกิน 1,500 คน จะทำให้การเทียบคะแนนด้วยวิธีอีคิวเปอร์เซ็นไทล์ส่งผลที่ดีขณะเดียวกันควรดำเนินการเทียบคะแนนสังเกตตามทฤษฎีการตอบข้อสอบด้วยวิธีการปรับค่าพารามิเตอร์พร้อมกันควบคู่ไปด้วยเพื่อตรวจสอบให้แน่ชัดในกรณีที่แบบทดสอบมีความยากต่างกัน และการเทียบคะแนนใน 2 วิธีดังกล่าว สามารถรายงานผลคะแนนให้ผู้ที่เกี่ยวข้องเข้าใจได้มากกว่าการใช้คะแนนจริง สำหรับในสถานการณ์ที่ไม่สามารถใช้ข้อสอบร่วมได้จำเป็นต้องปรับกระบวนการเทียบคะแนนอีคิวเปอร์เซ็นไทล์ คือไม่ต้องปรับการแจกแจงความถี่ ของแบบทดสอบเทียบคะแนนสองฉบับ และการทดสอบในสถานการณ์จริง หากใช้แบบแผนแบบทดสอบร่วมสามารถตรวจสอบก่อนดำเนินการเทียบคะแนนได้ว่าวิธีการเทียบคะแนนส่งผลต่อการเทียบคะแนนหรือไม่โดยพิจารณาจากความแตกต่างของค่าเฉลี่ยของข้อสอบร่วมในแต่ละกลุ่มผู้สอบเป็นเบื้องต้นว่าถ้าแตกต่างกันไม่เกิน 0.1 การเทียบคะแนนไม่ว่าวิธีใดก็จะให้ผลที่คล้ายคลึงกัน

1.2.2 ผลการวิจัยในครั้งนี้ หากนำไปใช้กับโครงการคัดเลือกบุคคลเข้าศึกษาต่อ นอกจากจะต้องพิจารณาปรับคะแนนเฉลี่ยระดับมัธยมศึกษาตอนปลายให้เท่าเทียมกันซึ่งเป็นองค์ประกอบหนึ่งแล้ว คะแนนจากการสอบวิชาหลักและวิชาเฉพาะที่ให้ใช้ได้ถึง 2 ปี โดยให้ผู้สอบเลือกคะแนนดิบครั้งที่สูงสุดมาใช้ ซึ่งความยากของข้อสอบในแบบทดสอบแต่ละครั้งไม่เท่ากัน ต้องทำการเทียบคะแนนผู้สอบในแต่ละครั้ง โดยถือเสมือนว่าประชากรผู้สอบในช่วงระยะเวลา 2 ปี เป็นประชากรกลุ่มเดียวกัน ก็สามารถนำผลการศึกษานี้ไปใช้ได้และแบบแผนที่ดีที่สุดก็คือแบบทดสอบร่วมแต่หากเป็นแนวทางที่ไม่สามารถดำเนินการได้ในสถานการณ์การสอบคัดเลือกก็ควรใช้แบบทดสอบที่มีความใกล้เคียงทางด้านเนื้อหาให้มากที่สุด วิธีการที่เหมาะสมที่ควรนำไปใช้คือการเทียบคะแนนอีควิเปอร์เซ็นไทล์ ซึ่งให้ผลดีในกลุ่มผู้สอบด้วยตัวอย่างขนาดใหญ่ ปรับกระบวนการดัง 1.2.1 แต่ไม่ต้องเปลี่ยนสูตรเหมือนการเทียบคะแนนเชิงเส้นตรงสำหรับการดำเนินการเทียบคะแนนตามทฤษฎีการตอบข้อสอบอาจมีความไม่เหมาะสมเพราะไม่มีข้อสอบร่วมทำให้การแปลงพารามิเตอร์ให้อยู่ในมาตรฐานเดียวกันไม่แม่นยำ และวิธีปรับค่าพารามิเตอร์พร้อมกันก็อาจไม่เหมาะสมเพราะวิเคราะห์ข้อมูลคนละครั้ง

1.2.3 ผลการวิจัยครั้งนี้หากนำไปใช้ในการบริหารการศึกษาเพื่อกำหนดมาตรฐานร่วมกันของหน่วยงาน โดยตรวจสอบมาตรฐานการศึกษาของสถานศึกษาในเขตเดียวกันโดยการทดสอบด้วยเครื่องมือวัดหลายฉบับทำให้สามารถดำเนินการในช่วงเวลาที่แตกต่างกันได้ โดยเครื่องมือวัดควรมีแบบทดสอบร่วมเป็นตัวปรับความแตกต่างของกลุ่มตัวอย่าง ก็จะส่งผลให้มีการพัฒนามาตรฐานการศึกษาร่วมกัน โดยผลการสอบที่ได้ต้องนำมาเทียบคะแนนก่อนตัดสินผล ซึ่งวิธีที่ควรนำไปใช้หากนักเรียนมีจำนวนไม่มากนักควรใช้วิธีการเชิงเส้นตรง ซึ่งเป็นวิธีที่ง่ายและสะดวกที่สุด หากผู้สอบมีจำนวนมากอาจใช้การเทียบคะแนนจริงวิธีปรับค่าพารามิเตอร์พร้อมกัน ซึ่งการดำเนินการไม่ยุ่งยากมากนักเมื่อเทียบกับวิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบอื่น ๆ และผลที่ได้นำมาใช้ในการพัฒนามาตรฐาน ซึ่งไม่ต้องสื่อสารถึงตัวผู้สอบโดยตรงการใช้การเทียบคะแนนจริงจึงมีความเหมาะสม โดยสามารถดำเนินการด้วยคอมพิวเตอร์ส่วนบุคคลได้ ซึ่งภายในหน่วยงานมีอยู่แล้ว หากนำผลการวิจัยนี้ไปใช้กับระบบการเรียนการสอนในระบบโรงเรียนซึ่งมีการทดสอบวัดผลสัมฤทธิ์ปลายปีของนักเรียนในชั้นเรียนเดียวกันซึ่งอาจจะประเมินด้วยเครื่องมือต่างฉบับในแต่ละห้องสอบ เนื่องจากครูผู้สอนหลายคน โดยเครื่องมือแต่ละฉบับควรมีข้อสอบร่วมอยู่ด้วยเพื่อใช้เป็นตัวปรับความแตกต่างของกลุ่มตัวอย่าง เพื่อให้การสอบนั้นเกิดความยุติธรรมแก่ผู้สอบทุกคน จึงควรปรับและเทียบคะแนนก่อนนำคะแนนไปประเมินผลวิธีการเทียบคะแนนเชิงเส้นตรงเป็นวิธีที่ง่ายและสะดวกในการนำมาใช้ในกลุ่มตัวอย่างขนาดเล็ก

2. ข้อเสนอแนะในการทำวิจัย

2.1 ควรศึกษาผลของการเทียบคะแนนแบบทดสอบในแนวนอนที่ข้อสอบตรวจให้คะแนนหลายค่าในแบบแผนกลุ่มสุ่มที่ใช้แบบทดสอบร่วมในองค์ประกอบอื่น เพื่อขยายองค์ความรู้ของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่าออกไปให้กว้างขวางขึ้น องค์ประกอบที่ควรศึกษา คือ วิธีการเทียบคะแนนตามทฤษฎีการตอบข้อสอบใน โมเดล เจเนอรอลไลซ์ พาเซียล เกรดดิ้ง ซึ่งเป็นโมเดลที่มีการใช้อย่างแพร่หลาย แบบทดสอบในรูปแบบเขียนตอบที่ข้อสอบตรวจให้คะแนนหลายค่าในลำดับขั้นคะแนนที่แตกต่างกันกับการศึกษาครั้งนี้ คือ 4 และ 5 ลำดับขั้น ความยาวของแบบทดสอบร่วม ความยาวของแบบทดสอบเทียบคะแนนที่เหมาะสม และขนาดของหน่วยตัวอย่างต่ำสุดที่ยังคงให้ประสิทธิภาพในการเทียบคะแนน

2.2 ควรศึกษาวิธีการเทียบคะแนนแบบทดสอบในแนวนอนที่ข้อสอบตรวจให้คะแนนหลายค่าในแบบแผนแบบทดสอบร่วมของกลุ่มไม่เท่าเทียมกัน (Common-Item Nonequivalent Groups Design) ซึ่งเป็นกลุ่มที่ไม่ได้มาจากประชากรเดียวกัน ได้แก่ ผู้สอบกลุ่มใหม่ในโปรแกรมการสอบเดียวกันหรือกลุ่มนักเรียนระดับชั้นเดียวกันคนละปีการศึกษา ซึ่งอาจให้ผลการเทียบคะแนนที่แตกต่างจากแบบแผนกลุ่มสุ่ม

2.3 ควรมีการศึกษา วิธีการเทียบคะแนนในแบบทดสอบรูปแบบผสมที่ข้อสอบตรวจให้คะแนนแบบหลายค่าและสองค่ารวมกันว่าวิธีการเทียบคะแนนวิธีใดมีความเหมาะสมในองค์ประกอบการเทียบคะแนนที่สำคัญอื่น ๆ คือ สัดส่วนจำนวนข้อสอบระหว่างข้อสอบที่ตรวจให้คะแนนแบบหลายค่าและสองค่ารวมกัน ลักษณะและจำนวนข้อของข้อสอบร่วมในแบบทดสอบรูปแบบผสมว่าควรเป็นข้อสอบชนิดใดอาจเป็นทั้งข้อสอบที่ให้คะแนนแบบหลายค่าและสองค่า รวมกันหรืออย่างใดอย่างหนึ่ง เพื่อเป็นการขยายองค์ความรู้เกี่ยวกับกระบวนการเทียบคะแนนให้เหมาะสมกับสภาพการวัดในปัจจุบันที่มีความหลากหลายและมีการนำข้อสอบที่ตรวจให้คะแนนแบบหลายค่า โดยเฉพาะข้อสอบเขียนตอบมาใช้มากขึ้น

บรรณานุกรม

บรรณานุกรม

- กัลยา วาณิชย์บัญชา. (2542). *การวิเคราะห์ข้อมูลด้วย SPSS for Windows*. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- จักรกฤษณ์ สำราญใจ. (2530). *ประสิทธิภาพเชิงสัมพัทธ์ของข้อสอบเลือกตอบชนิดตัดสินคำตอบทุกตัวเลือกเทียบกับข้อสอบเลือกตอบชนิดแบบฉบับในแบบสอบผลสัมฤทธิ์*. วิทยานิพนธ์ ค.ด.(การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ชวลิต โพธิ์นคร. (2540). *การศึกษาอิทธิพลของการให้คะแนนและการคัดเลือกข้อสอบที่มีต่อคุณภาพแบบทดสอบเลือกตอบ*. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ชูชีพ พงษ์สมบุรณ์. (2528). *การเปรียบเทียบการเทียบมาตรฐานระหว่างรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบรวม*. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ชูศักดิ์ ชัมภลิจิต. (2529). "การเปรียบเทียบคะแนนทดสอบ," ในโครงการพัฒนศึกษาอาเซียน, บทความทางวิชาการทดสอบ. หน้า 143 – 150. นนทบุรี : มหาวิทยาลัย-สุโขทัยธรรมมาธิราช.
- ดรุณี พรประเสริฐ. (2541). *การเทียบคะแนนระหว่างแบบทดสอบระดับประเทศและแบบทดสอบระดับจังหวัดกลุ่มทักษะคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2540*. วิทยานิพนธ์ กศ.ม.(การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยรามคำแหง. ถ่ายเอกสาร.
- ทิพาภรณ์ ทุงฤทธิ์. (2544). *การเปรียบเทียบผลการเทียบคะแนนโดยวิธีเชิงเส้นตรง เมื่อใช้วิธีการแปลงคะแนนที่แตกต่างกันในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์*. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ธนวัฒน์ แสนสุข. (2538). *การใช้จีอาร์เอ็ม จีพีซีเอ็ม และโมเดลโลจิสติกในการเปรียบเทียบฟังก์ชันสารสนเทศของแบบวัดที่มีวิธีการให้คะแนนต่างกัน*. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

- นิภาพร ยอดเมือง. (2539). ความคลาดเคลื่อนและความลำเอียงทางสถิติในการเทียบคะแนน
เชิงเส้นตรงแบบคะแนนจริงสัมพันธ์ในกลุ่มตัวอย่างขนาดเล็ก. ปรินูญานิพนธ์ กศ.ม.
(การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัย-
ศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- บุญเชิด ภิญโญอนันตพงษ์. (2527). การทดสอบแบบอิงเกณฑ์ : แนวคิดและวิธีการ.
กรุงเทพฯ : โอเดียนสโตร์.
- _____. (2544). การประเมินการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ : แนวคิดและวิธีการ.
กรุงเทพฯ : สำนักงานคณะกรรมการการศึกษาแห่งชาติ.
- _____. (2545). "คุณภาพเครื่องมือวัด" หน่วยที่ 3. ประมวลสาระชุดวิชาการพัฒนาเครื่องมือ
สำหรับการประเมินการศึกษา. นนทบุรี : มหาวิทยาลัยสุโขทัยธรรมาธิราช.
- พรพิมล นาคเวช. (2537). การศึกษาคุณภาพของการเทียบคะแนนในแนวตั้งโดยใช้ทฤษฎี
การตอบสนองข้อสอบ. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา).
กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- พรโสภา พินิจผล. (2539). ผลการเทียบคะแนนในแบบทดสอบวิชาการศึกษาบาลสฤติศาสตร์
ตามทฤษฎีการตอบสนองข้อสอบที่ใช้วิธีการวิเคราะห์ต่างกัน. ปรินูญานิพนธ์ กศ.ม.
(การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัย
ศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- พิชัย ละแมนชัย. (2538). ขนาดกลุ่มตัวอย่างขั้นต่ำสำหรับการปรับเทียบคะแนนในแนวระดับ
ตามแนวทฤษฎีการตอบสนองรายข้อ ระหว่างแบบสอบที่มีแบบแผนการปรับเทียบ
และความยาวแบบสอบที่แตกต่างกัน. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผล
การศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ภัทรพร เกษสังข์. (2546). การศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียน
ชั้นมัธยมศึกษาตอนต้นโดยการเทียบคะแนนในแนวตั้งที่ใช้วิธีการที่เหมาะสม.
ปรินูญานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ :
บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ภาวิณี ศรีสุขวัฒนานนท์. (2529). การเปรียบเทียบผลจากการใช้รูปแบบการเทียบมาตราที่
ต่างกันเมื่อแบบสอบร่วมมีความยาวต่างกัน. วิทยานิพนธ์ ค.ด. (การวัดและ
ประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
ถ่ายเอกสาร.
- เรวดี อินทสระ. (2530). การเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตรระหว่าง
รูปแบบอิงทฤษฎีการตอบสนองกับรูปแบบการใช้เทคนิคการวิเคราะห์
องค์ประกอบ. ปรินูญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ :
บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

ล้วน สายยศ และอังคณา สายยศ. (2540). สถิติวิทยาทางการวิจัย. พิมพ์ครั้งที่ 3.

กรุงเทพฯ : สุวีริยาสาส์น.

วรรณตร พิชิตเกริกพล. (2536). การเปรียบเทียบคุณภาพของวิธีการเทียบมาตรฐานระหว่างแบบอิกวิเปอร์เซนไทล์กับแบบเชิงเส้นตรง. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

วรรณช แหลมแสง. (2536). การพัฒนากระบวนการตรวจสอบความเป็นเอกมิติของแบบสอบ. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

วรรณดี แสงประทีปทอง. (2538). การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้นมัธยมศึกษาตอนต้น โดยการเทียบคะแนนตามแนวตั้ง ตามทฤษฎีการตอบสนองข้อสอบ แบบจำลองโลจิสติกสามพารามิเตอร์. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

วัชรารณ จิตรมาศ. (2538). การเปรียบเทียบคุณภาพแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือกที่มีจำนวนตัวเลือกและวิธีการตรวจให้คะแนนต่างกัน. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

วัฒนาพร ระวังทุกข์. (2542). แผนการสอนที่เน้นผู้เรียนเป็นศูนย์กลาง. กรุงเทพฯ : แอล ที เพรส จำกัด.

ศิริชัย กาญจนวาสี. (2541). การปรับเทียบคะแนนระหว่างแบบสอบ. กรุงเทพฯ : คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

_____. (2545). ทฤษฎีการทดสอบแนวใหม่. กรุงเทพฯ : คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศิริเดช สุชีวะ. (2539, กรกฎาคม – ธันวาคม). "โมเดลการตอบสนองข้อสอบแบบพหุภาค (Polytomous IRT Model) และการประยุกต์ใช้กับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ (Computerized Adaptive Testing)," วารสารวิธีวิทยาการวิจัย. 8 (2) : 17 – 30.

สงบ ลักษณะ. (2525, กันยายน - ธันวาคม). "การเทียบระดับคะแนนระหว่างแบบทดสอบ," วารสารการวัดผลการศึกษา. 24 (1) : 21-32.

- สมพงษ์ แก้วแจ่ม. (2537). การพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการเทียบคะแนนเชิงเส้นตรงตามแนวนอน โดยใช้แบบสอบรวม. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- สมพงษ์ อัครคุณฤกษ์. (2542). การเปรียบเทียบความเชื่อมั่นของแบบทดสอบที่มีวิธีการตรวจและความยาวของแบบทดสอบต่างกัน. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- สุจินดา ผ่องอักษร. (2533). การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษาพุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบคะแนนรูปแบบราสช์. วิทยานิพนธ์ กศ.ด. (วิจัยและพัฒนาหลักสูตร). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- สุชาติ กิระนันท์. (2538). ทฤษฎีและวิธีการสำรวจตัวอย่าง. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- สุนิสา จัยม่วงศรี. (2536). ผลของความยาวของแบบสอบรวมที่มีต่อคุณภาพของวิธีการเทียบมาตราเชิงเส้นตรง. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. อัดสำเนา.
- สุมนา โสติดิผลอนันต์. (2539). การศึกษาวิธีการเทียบคะแนนแบบเส้นตรงตามแบบจำลองคะแนนจริงสัมพันธ์. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- สุรัช มีชาญ. (2539). การเปรียบเทียบความยากประจำชั้น และประสิทธิภาพในการประมาณค่าเจตคติของมาตรวัดแบบลิเคอร์ที่มีรูปแบบของการตอบ จำนวนของลำดับขั้นและทิศทางของการเรียงลำดับขั้นแตกต่างกันด้วยพาเซิล เครดิต โมเดลของมาสเตอร์. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- สุรัตน์ ขวัญบุญจันทร์. (2531). การสร้างแบบทดสอบและตารางเทียบคะแนนในแนวนอน วิชาวิทยาศาสตร์ที่วิเคราะห์ด้วยราสช์โมเดล. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- สำนักงานคณะกรรมการการศึกษาแห่งชาติ. (2542). พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542. กรุงเทพฯ : พริกหวานกราฟฟิค.

- อวยพร เรืองตระกูล. (2544). การพัฒนาและวิเคราะห์คุณภาพของวิธีการวัดคะแนนพัฒนาการตามทฤษฎีการทดสอบแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบ. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- อดิสร ศรีบุญวงศ์. (2545). การพัฒนาเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- อาภรณ์ กาญจนกิจโสภณ. (2531). การสร้างแบบทดสอบและตารางเทียบคะแนนในแนวตั้งในวิชาคณิตศาสตร์. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- อุทัยวรรณ พงศ์อร่าม. (2545). การศึกษาขนาดกลุ่มตัวอย่างที่เหมาะสมสำหรับการเปรียบเทียบคะแนนด้วยวิธีอีควิเปอร์เซ็นไทล์และวิธีเชิงเส้นตรงตามแบบจำลองคะแนนจริงสัมพันธ์ที่มีแบบแผนการเปรียบเทียบและความยาวของแบบทดสอบแตกต่างกัน. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- Andersen, E.B. (1977, March). "Sufficient Statistics and Latent Trait Models." in *Psychometrika*, 42(1) : 69-81.
- Andrich, D. (1978, December). "A Rating Formulation for Ordered Response Categories," in *Psychometrika*, 43(4) : 561-573.
- Angoff, W.H. (1982). "Summary and Derivation of Equating Methods," in *Test Equating*. Edited by Holland P.W. and Rubin D.B. P. 55 – 69. New York : ACADEMIC PRESS.
- _____. (1984). *Scales, Norms, and Equivalent Scores*. New Jersey: Princeton, Educational Testing Service. (Reprint of chapter 15 in *Educational Measurement* (1971). 2nd ed. Edited by Thronrdike R.L) Washington DC : American Council on Education.
- Baker, F.B. (1992a, March). "Equating Tests Under the Graded Response Model," in *Applied Psychological Measurement*, 16 (1) : 87-96.
- _____. (1992b). *Item Response Theory : Parameter Estimation Techniques*. New York : Marcel Dekker.
- _____. (1993a, September). "Equating Tests Under the Nominal Response Model," in *Applied Psychological Measurement*, 17 (3) : 239-251.

- _____. (1993b, March), "Equate 2.0 : A Computer Program for the Characteristic curve method of IRT Equating," in *Applied Psychological Measurement*, 17(1) : 20.
- _____. (1996, March). "An Investigation of the Sampling Distributions of Equating Coefficients," in *Applied Psychological Measurement*, 20 (1) : 45-57.
- _____. (1997, June). "Empirical Sampling Distributions of Equating Coefficients for Graded and Nominal Response Instruments," in *Applied Psychological Measurement*, 21 (2) : 157-172.
- Baker, F.B. and Al-Karni, A. (1991, Summer). " A Comparison of Two Procedures for Computing IRT Equating Coefficients," in *Journal of Educational Measurement*, 28(2) : 147-162.
- Bastari, B. (2000). *Linking Multiple – Choice and Constructed – Response Items to a Common Proficiency Scale*. (online) Dissertation, Ed. D. USA: Graduate School, Massachusetts Amherst University. Available : UMI/ Dissertation (2001) P. 1 – 78.
- Bock, R. D. (1972, March). "Estimating Item Parameters and Latent Ability when Responses are Scored in Two or More Nominal Categories," in *Psychometrika*, 37(1): 29-51.
- Brennan, R.L. (1987, September) "Introduction to Problems, Perspectives, and Practical Issues in Equating," in *Applied Psychological Measurement*, 11(3) : 221 - 224.
- Brennan, R.L. and Kolen, M.J. (1987, September). "Some Practical Issues in Equating," in *Applied Psychological Measurement*, 11(3) : 279 – 290,
- Cohen, Allan S. and Kim, Seock Ho. (1998, June). "An Investigation of Linking Methods under the Graded Response Model," in *Applied Psychological Measurement*, 22(2) : 116-130.
- DeMars, C. E. (1998). "Gender Differences in Mathematics and Science on a High School Proficiency Exam: The Role of Response Format," in *Applied Measurement in Education*, 11(3) : 279-299.
- Divgi, D.R. (1985, December). "A Minimum Chi-Square Method for Developing a Common Metric in Item Response Theory," in *Applied Psychological Measurement*, 9(4) : 413-415.

- Dodd, B.G; Ayala, R.J. and Koch, W.R. (1995,March). "Computerized Adaptive Testing With Polytomous Items," in *Applied Psychological Measurement*, 19(1) : 5–22 .
- Donoghue, John R. (1994, Winter). "An Empirical Examination of the IRT Information of Polytomously Scored Reading Items Under the Generalized Partial Credit Model," in *Journal of Educational Measurement*, 31 (4) : 295 – 311.
- Dorans, N.J. and Holland , P.W. (2000, Winter) "Population Invariance and the Equatability of Tests : Basic Theory and the Linear Case," in *Journal of Educational Measurement*, 37(4) : 281-306.
- Embretson, S.E. and Reise, S.P. (2000). *Item Response Theory for Psychologists*. New Jersey : Lawrence Erlbaum Associates.
- Ercikan, K. and others. (1998,Summer). "Calibration and Scoring of Tests with Multiple-Choice and Constructed Response Item Types," in *Journal of Educational Measurement*, 35(2) : P137-154.
- Flanagan, T.C. (1951). "Units, Scores and Norms," In *Educational Measurement*. Edited by Lindquist E.F. P. 695 – 763. Washington D.C. : American Council on Education.
- Feldt, L.S. (1969,September). "A test of the hypothesis that Cronbach's alpha or Kuder-Richardson.Coefficient twenty is the same for two test," in *Psychometrika*, 34(3) : 363 – 373.
- Fitzpatrick. A.R., and others. (1996, fall). "Scaling Performance Assessments: A Comparison of One-parameter and Two-Parameter Partial Credit Model," in *Journal of Educational Measurement*, 33(3) : 291-314.
- Garner, M., and Engelhard, G, Jr. (1999). "Gender Differences in Performance on Multiple-Choice and Constructed-Response Mathematics Items," in *Applied Measurement in Education*, 12(1) : 29-51.
- Haladyna, T.M. (1999). *Developing and Validating Multiple-Choice Test Items*. 2nd ed. New Jersey : Lawrence Erlbaum Associates.
- Hambleton, R.K. (1996). "Advances in Assessment Models, Methods, And Practices," in *Handbook of Educational Psychology*, Edited by Berliner D.C and Calfee R.C. P.899-925. New York : Simon & Schuster Maemillan.
- Hambleton, R.K., and Swaminathan, H.(1985). *Item Response Theory : Principles and Application*. Boston : Kluwer Nijhoff.

- Hambleton, R.K, Swaminathan, H & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park : C.A. SAGE Publications, Inc.
- Han, T. , Kolen, M. and Pohlmann, J. (1997). "A Comparison Among IRT True – and Observed – Score Equatings and Traditional Equipercentile Equating," in *Applied Measurement in Education* 10(2) : 105-121.
- Hanson, Bradley A. and Feinstein, Zachary S. (1997). "Application of a Polynomial Loglinear Model to Assessing Differential Item Functioning for Common Items in the Common-Item Equating Design," in *American Coll. Testing Program*, Iowa City : IA .
- Harris, D.J. and Crouse, J.D. (1993). "A Study of Criteria Used in Equating," in *Applied Measurement in Education*, 6(3) : 195 – 240.
- Harris, D.J. and Hoover, H.D. (1987, June). "An Application of the Three-Parameter IRT Model to Vertical Equating," in *Applied Psychological Measurement*, 11(2) : 151 – 159.
- Hennings, Sara S., Hirsch, T.M. and Zhang, L. (1996, April). " A Comparison in of Equating Methods Applied to Performance-Based Assessments," in Paper Presented at the *Annual Meeting of the National Council on Measurement in Education*. (CD-ROM). New York City, NY. P. 1-13. Available : ERIC (1992 – 2000/9) Acc. No ED400295
- Holland, P.W.and Rubin,D.B. (Eds.) (1982). *Test Equating*. New Jersey : Educational Testing Service, Princeton Academic Press.
- Hsu, Yaowen . and Ackerman, Terry A. (1994, April). "Equating Reading Test Scores that Combine Narrative and Expository Test Formats," in Paper presented at the *Annual Meeting of the American Educational Research Association* (abstract). New Orleans, LA, P. 4-8. Available : ERIC (1992 – 2001/09) Acc. No ED373087.
- Huynh, Huynh. And Ferrara, Steven. (1994, Summer). "A Comparison of Equal Percentile and Partial Credit Equatings for Performance-Based Assessments Composed of Free-Response Items," in *Journal of Educational Measurement*, 31(2) : 125-141.
- Jaeger, R.M. (1981, Spring). "Some Exploratory Indices for Selection of A Test Equating Method," in *Journal of Education Measurement*, 18(1) : 23-37.

- Johnson, E.G. and Owen, E. (1998, September). *Linking the National Assessment of Educational Progress (NAEP) and the third International Mathematics and Science Study (Timss) : A Technical Report*. (Online). U.S.A. : National Center For Education Statistics , Research and Development Report. Available : <http://www.ETS.ORG>.
- Kennedy, P. and Walstad, W.B. (1997). "Combining Multiple – Choice and Constructed Response Test Scores : An Economist's View," *in Applied Measurement in Education*, 10(4) : 359-375.
- Kim, Seock-Ho. And Cohen, Allan S. (1995, June). "A Minimum χ^2 Method for Equating Tests Under the Graded Response Model," *in Applied Psychological Measurement*, 19(2) : 167-176.
- _____. (1998a, June). " A Comparison of linking and Concurrent Calibration under Item Response Theory," *in Applied Psychological Measurement*, 22(2) : 131 – 143.
- _____. (1998b, December). "Detection of Differential Item Functioning Under the Graded Response Model With the Likelihood Ratio Test," *in Applied Psychological Measurement*, 22(4) : 345 – 355.
- Kolen, M.J. (1981, Spring). "Comparison of Traditional and Item Response Theory Methods for Equating Tests," *in Journal of Educational Measurement*, 18(1) : 1 – 11.
- Kolen, M.J. and Brennan, R.L. (1995). *Test Equating : Method and Practices*. Newyork : Springer-Verlag.
- Kolen, M.J., & Whitney, D.R.(1982, Winter). "Comparison of Four Procedures for Equating the Tests of General Educational Development," *in Journal of Educational Measurement*, 19(4) : 279 – 293.
- Lee,G.and others. (2001). "Comparison of Dichotomous and Polytomous Item Response Models in Equating Scores From Tests Composed of Testlets," *in Applied Psychological Measurement*, 25(4) : 357-372.
- Li , Y.H., Lissitz ,R.W. and Yang, Y.N. (1999, April). "Estimating IRT Equating Coefficients for Tests with Polytomously and Dichotomously Scored Items," in Paper Presented at the *Annual Meeting of the National Council on Measurement in Education (NCME)*. (CD-ROM). P. 1- 32. Available : ERIC (1992 – 2000/09); Acc. No. ED431800.

- Lord, F.M. (1980). *Application of Item Response Theory to Practical Testing Problems*. N.J. : Hillsdale, Erlbaum.
- Lord, F.M., and Wingesky, M.S. (1984, Fall). "Comparison of IRT True-Score and Equipercntile Observed-Score Equating," in *Applied Psychological Measurement*, 8(4) : 453 – 461.
- Lukhele, R., Thissen, D. and Wainer, H. (1994, Fall). "On the Relative Value of Multiple-Choice, Constructed-Response, and Examinee-Selected Items on Two Achievement Tests," in *Journal of Educational Measurement*, 31(3) : 234 – 250.
- Marascuilo, L.A. and McSweeney, M.(1977). *Nonparametric and distribution free methods for the social science*. Monterey, C.A.: Brooks/ Cole.
- Marco, G.L., (1977, Summer). "Item Characteristic Curve Solutions to Three Intractable Testing Problems," in *Journal of Educational Measurement*, 14(2) : 139 – 160.
- Masters, G.N. (1982,June). "A Rasch model for partial credit scoring," in *Psychometrika*, 47(2) : 149– 174.
- Masters. G.N. and Wright, B.D. (1997) "The Partial Credit Model," in *Handbook of Modern Item Response Theory*. Edited by Van der Linden W.J. and Hambleton R.D. P. 101 - 121. New York : Springer-Verlag,
- Muraki, E. (1990,March). "Fitting a polytomous item response model to Likert-type data," in *Applied Psychological Measurement*, 14 (1) : 59 – 71.
- _____. (1992, June). "A Generalized Partial Credit Model: Application of an EM Algorithm," in *Applied Psychological Measurement*, 16(2) : 159 – 176.
- _____. (1997). "A Generalized Partial Credit Model," in *Handbook of Modern Item Response Theory*. Edited by Van der Linden, W.J. and Hambleton, R.K. P. 153 – 164. New York : Springer-Verlag.
- Muraki, E., Hombo, C.M. and Lee, Yong – Won. (2000, December) "Equating and Linking of Performance Assessments," in *Applied Psychological Measurement*, 24(4) : 325-337.
- Olivares, M.A. Drasgow, F. and Mead, A.D.(1994, September). "Distinguishing Among Parametric Item Response Models for Polychotomous Ordered data," in *Applied Psychological Measurement*, 18(3) : 245 – 256.

- Osterlind, S.J.(1998). *Constructing Test Items : Multiple – Choice, Constructed – Response, Performance, and other Formats*. 2nd ed. Massachusetts : Kluwer Academic.
- Petersen, N.S., Cook, L.L. & Stocking, M.L. (1983,Summer). "IRT versus conventional equating methods: a comparative study of scale stability," in *Journal of Educational Statistics*, 8 (2) : 137 – 156.
- Petersen, N.S. , Kolen, M.J. and Hoover , H.D. (1989). "Scaling, Norming, and Equating," In *Educational Measurement*. 3rd Ed. Republication of Chapter 6. Edited by Linn, R.L. P. 221 – 262. New York : Macmillan.
- Petersen, N.S. , Marco, G.L. and Stewart, E.E. (1982). "A Test of the Adequacy of Linear Score Equating Models," in *Test Equating*. Edited by Holland, P.W. and Rubin, D.B. P. 71 – 135. New Jersey: Academic Press.
- Raju, N.S., Drasgow, F. and Slinde, J.A. (1993, Summer). "An Empiral Comparison of the Area Methods, Lord's Chi-square Test, and the Mantel-Hanszel Technique for Assessing Differential Item Functioning," in *Educational and Psychological Measurement*, 53 (2) : 301 – 314.
- Rangchikul Yavadee. (1975). *Equating and Calibration Technique For Test Program in Thailand*. A Dissertation Submitted to the Area of Educational Management System the Florida State University.
- Reise, S.P., & Yu, J. (1990, Summer). "Parameter Recovery in the Graded Response Model Using MULTILOG," in *Journal of Educational Measurement*, 27(2) : 133 – 144.
- Samejima, F. (1997). "Graded Response Model," in *Handbook of Modern Item Response Theory*. Edited by Van der Linder W.J. and Hambleton R.K. P 85-100. New York : Springer – Verlag.
- Simon, A.B., Budescu, D.V. and Nevo, B. (1997,March). " A Comparative Study of Measurement of Partial Knowledge in Multiple – Choice Tests, " in *Applied Psychological Measurement*, 21(1) : 65-88.
- Skaggs. G. and Lissitz, R.W. (1986, September). "An Exploration of the Robustness of Four Test Equating Models," in *Applied Psychological Measurement*, 10(3) : 303 – 317.
- _____. (1986, Winter). "IRT Test Equating: Relevant Issues and a Review of Recent Research," in *Review of Educational Research*. 56(4) : 495 – 529.

- Slinde, J. A. & Linn, R.L. (1977, Spring). "Vertically equated tests: Fact or *phantom*?" in *Journal of Educational Measurement*, 14 (1) : 23 – 32.
- _____. (1978, Spring). "An exploration of the adequacy of the Rasch model for the problem of vertical equating," in *Journal of Educational Measurement*, 15 (1) : 23 – 35.
- Stocking, M.L. and Lord, F.M. (1983, Spring). "Developing a Common Metric in Item Response Theory," in *Applied Psychological Measurement*, 7(2) : 201 – 210.
- Sykes, R.C. and Yen W.M. (2000, Fall). "The Scaling of Mixed-Item-Format Test with the One- Parameter and Two Parameter Partial Credit Models," in *Journal of Educational Measurement*, 37 (3) : 221-244.
- Tate, R. (1999, Winter). "A Cautionary Note on IRT – Based Linking of Tests with Polytomous Items," in *Journal of Educational Measurement*, 36(4) : 336 - 346.
- _____. (2000, Winter). "Performance of a Proposed Method for the Linking of Mixed Format Tests with Constructed Response and Multiple Choice Items," in *Journal of Educational Measurement*, 37(4) : 329-346.
- Thissen, D. (2002). *MULTILOG for Windows 7.03 User's Guide : Multiple Categorical Item Analysis and Test Scoring Using Item Response Theory*. Chicago, IL : Scientific Software Inc.
- Thissen, D. and Steinberg, L. (1984, December). "A Response Model for Multiple Choice items," in *Psychometrika*, 49(4) : 510 – 519.
- Thissen, D. and others. (1995). "Item Response Theory for Scores on Tests Including Polytomous Items with Ordered Responses," in *Applied Psychological Measurement*, 19(1) : 39 - 49.
- Triscari, Robert S. (1990,). "A Comparison of Classical and Item Response Theory Equating Methods Using a Compositive Score of Direct and Indirect Writing Assessment" *Dissertation Abstracts International*. (CD-ROM). 51(6) : 118p. Available : UMI; Dissertation Abstracts (51/06) AAC 9031731.
- Tutz, G. (1997). "Sequential Models for Ordered Responses," in *Handbook of Modern Item Response*. Edited by Van der Linden W.J. and Hambleton R.K. P. 139 – 152 Theory. New York: Springer-Verlag.
- Tzou, Huey-Ing. (1995). "An Examination of Equating Strategies Using Polytomous Response Data (Performance Assessment)," *Dissertation Abstracts International*. (CD-ROM). 57(1) :185p. Available : UMI ; Dissertation Abstracts (57/01) AAC. 9614688.

- Vale, D. (1986, December). "Linking Item Parameters onto a Common Scale," in *Applied Psychological Measurement*, 10(4) : 333 – 344.
- Van der Linden, W.J., and Eggen, T.J.H.M. (1986, December). "An Empirical Bayesian Approach to Item Banking," in *Applied Psychological Measurement*, 10(4) : 345 – 354.
- Van der Linden, W.J. and Hambleton, R.K. (Eds.) (1997). *Handbook of Modern Item Response Theory*. New York : Springer – Verlag.
- Verhelst, N.D., Glas C.A.W. and De Vries, H.H. (1997). "A Steps Model to Analyze Partial Credit," in *Handbook of Modern Item Response Theory*. Edited by Van Der Linden W.J. and Hambleton R.K. P. 123 – 137. New York : Springer – Verlag.
- Waltman, K.K.(1997, Summer). "Using Performance Standards to Link Statewide Achievement Results to NAEP," in *Journal of Educational Measurement*, 34(2) : 101 – 121.
- Wilson, J.W. (1971). "An Evaluation of Learning in Secondary school Mathematics," in *Handbook on Formative Evaluation of Student Learning*. USA : McGraw – Hill.
- Wingersky, M.S., & Lord, F.M. (1984,Summer). "An investigation of methods for reducing sampling error in certain IRT procedures," in *Applied Psychological Measurement*, 8(3) : 347 – 364.
- Wright, B.D. and Masters G.N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.
- Wright, B.D. and Stone, M.H. (1979). *Best test design : Rasch Measurement*. Chicago: Mesa Press.

ภาคผนวก

ภาคผนวก ก

- ผลการวิเคราะห์คำอำนาจจำแนก ความยากของข้อสอบในแบบทดสอบเทียบคะแนนตามทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์ และตามทฤษฎีการทดสอบแบบมาตรฐานเดิม
- ผลการวิเคราะห์คำสารสนเทศข้อสอบและแบบทดสอบตามทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์

ตาราง 52 ค่าอำนาจจำแนกและความยากของข้อสอบในแบบทดสอบเทียบคะแนนวิเคราะห์ตามทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์ และวิเคราะห์ตามทฤษฎีการทดสอบแบบมาตรฐานเดิม

เนื้อหา/ จุดประสงค์	ข้อ	ระดับ พฤติกรรม	ฉบับ	P	R	A	B ₁	B ₂	B ₃	B ₄	B ₅
<u>จำนวนเต็ม</u>											
1.1	1	เข้าใจ	X	0.822	0.499	1.369	-4.473	-2.994	-1.883	-0.743	0.436
	1	จำ	Y	0.761	0.550	1.520	-2.911	-2.022	-1.076	-0.222	0.777
1.2	2(รวม)	เข้าใจ	X	0.689	0.524	1.475	-5.817	-2.605	-0.963	0.070	2.129
	2(รวม)	เข้าใจ	Y	0.689	0.526	1.474	-4.872	-2.484	-0.883	0.062	2.172
2.1	3(รวม)	เข้าใจ	X	0.657	0.583	1.553	-3.334	-1.620	-0.777	0.485	1.710
	3(รวม)	เข้าใจ	Y	0.673	0.599	1.658	-3.966	-1.676	-0.788	0.409	1.697
2.2	4(รวม)	เข้าใจ	X	0.731	0.549	1.497	-4.109	-2.540	-1.286	-0.042	1.341
	4(รวม)	เข้าใจ	Y	0.725	0.536	1.453	-4.454	-2.400	-1.098	0.035	1.327
2.3	5	นำไปใช้	X	0.756	0.590	2.159	-2.246	-1.555	-0.911	-0.101	0.452
	6	คิดคำนวณ	X	0.724	0.645	2.061	-2.574	-1.705	-0.899	0.118	0.863
	5	วิเคราะห์	Y	0.670	0.597	1.740	-2.180	-1.265	-0.353	0.242	0.928
	6	เข้าใจ	Y	0.722	0.663	2.143	-2.529	-1.304	-0.582	0.045	0.724
3.1	7	เข้าใจ	X	0.758	0.633	2.414	-2.050	-1.414	-0.615	-0.162	0.285
	7	จำ	Y	0.789	0.581	2.116	-2.214	-1.423	-0.833	-0.302	0.214
3.2	8	เข้าใจ	X	0.788	0.564	1.749	-2.850	-1.832	-1.066	-0.510	0.506
	9	นำไปใช้	X	0.631	0.676	2.072	-2.840	-0.963	-0.191	0.516	1.018
	8	วิเคราะห์	Y	0.713	0.629	1.963	-2.381	-1.490	-0.645	0.094	0.873
	9	เข้าใจ	Y	0.798	0.553	1.612	-2.970	-2.052	-1.119	-0.510	0.438
3.3	10	เข้าใจ	X	0.626	0.547	1.395	-3.062	-1.541	-0.515	0.668	1.743
	11	เข้าใจ	X	0.738	0.663	2.403	-2.628	-1.454	-0.801	-0.179	0.920
	10	วิเคราะห์	Y	0.648	0.626	1.876	-1.885	-1.097	-0.296	0.425	0.951
	11	เข้าใจ	Y	0.636	0.486	1.279	-3.540	-1.986	-0.883	0.743	2.714
4.1	12(รวม)	จำ	X	0.669	0.529	1.325	-4.066	-2.313	-0.855	0.372	2.056
	12(รวม)	จำ	Y	0.655	0.550	1.288	-4.028	-2.084	-0.670	0.508	2.170
4.2	13	เข้าใจ	X	0.736	0.605	1.797	-3.222	-1.941	-1.018	-0.069	1.032
	13	นำไปใช้	Y	0.812	0.604	2.309	-2.222	-1.498	-0.864	-0.454	0.154
	14	วิเคราะห์	X	0.692	0.585	1.574	-3.166	-1.748	-0.796	0.212	1.282
5.1	16	จำ(รวม)	X	0.743	0.560	1.487	-3.809	-2.222	-1.176	-0.159	1.028
	16	จำ(รวม)	Y	0.723	0.583	1.520	-4.302	-2.005	-0.934	-0.020	1.236
5.2	15	วิเคราะห์	X	0.689	0.582	1.504	-2.892	-1.709	-0.519	0.120	1.053
	14	นำไปใช้	Y	0.695	0.668	2.404	-1.499	-0.921	-0.455	-0.006	0.763
	15	เข้าใจ	Y	0.721	0.573	1.699	-3.130	-1.624	-0.875	-0.023	1.086

ตาราง 52 (ต่อ)

เนื้อหา/ จุดประสงค์	ข้อ	ระดับ พฤติกรรม	ฉบับ	P	R	A	B ₁	B ₂	B ₃	B ₄	B ₅
เคยสัณ											
6.1	17	เข้าใจ(รวม)	X	0.839	0.510	2.002	-2.715	-1.568	-1.123	-0.707	-0.104
	17	เข้าใจ(รวม)	Y	0.844	0.521	2.051	-2.521	-1.569	-1.063	-0.714	-0.082
6.2	18	เข้าใจ	X	0.714	0.569	1.515	-3.091	-1.828	-0.913	0.018	1.101
	18	วิเคราะห์	Y	0.649	0.524	1.347	-3.644	-1.829	-0.523	0.555	1.653
7.1	19	จำ	X	0.636	0.462	1.059	-4.283	-2.221	-0.687	0.495	2.444
	19	เข้าใจ	Y	0.706	0.550	1.680	-2.462	-1.518	-0.711	-0.044	1.142
7.2	20	วิเคราะห์	X	0.653	0.622	1.811	-2.673	-1.138	-0.315	0.294	0.893
	20	วิเคราะห์	Y	0.715	0.599	1.660	-3.548	-1.770	-0.796	0.041	1.152
	21	วิเคราะห์	X	0.677	0.590	1.613	-2.365	-1.424	-0.578	0.237	1.050
	21	นำไปใช้	Y	0.700	0.632	1.863	-2.903	-1.404	-0.636	0.192	0.975
7.3	22	เข้าใจ(รวม)	X	0.673	0.479	1.159	-4.337	-2.070	-0.861	0.313	1.637
	22	เข้าใจ(รวม)	Y	0.671	0.513	1.282	-3.928	-1.823	-0.688	0.373	1.529
7.4	23	เข้าใจ	X	0.651	0.550	1.483	-2.983	-1.581	-0.443	0.464	1.259
	23	เข้าใจ	Y	0.641	0.479	1.197	-3.420	-1.752	-0.378	0.557	1.398
ทศนิยม											
8.1	24	เข้าใจ	X	0.583	0.525	1.307	-2.752	-1.171	-0.031	0.744	1.684
	24	เข้าใจ	Y	0.791	0.568	1.788	-3.213	-1.637	-0.960	-0.393	0.328
8.2	25	เข้าใจ(รวม)	X	0.635	0.568	1.538	-2.773	-1.364	-0.362	0.471	1.382
	25	เข้าใจ(รวม)	Y	0.640	0.577	1.568	-2.735	-1.250	-0.257	0.442	1.297
9.1	26	เข้าใจ	X	0.686	0.544	1.612	-2.177	-1.198	-0.509	0.068	0.625
	26	เข้าใจ	Y	0.615	0.557	1.509	-3.034	-1.541	-0.357	0.763	1.986
9.2	27	เข้าใจ	X	0.589	0.529	1.324	-2.674	-1.277	-0.157	0.744	1.861
	28	วิเคราะห์	X	0.639	0.540	1.400	-3.195	-1.651	-0.568	0.515	1.744
	27	เข้าใจ	Y	0.758	0.592	2.027	-2.483	-1.216	-0.676	-0.185	0.320
	28	วิเคราะห์	Y	0.649	0.547	1.440	-3.179	-1.383	-0.463	0.405	1.515
9.3	29	เข้าใจ(รวม)	X	0.581	0.518	1.378	-2.461	-1.176	-0.181	0.792	1.861
	29	เข้าใจ(รวม)	Y	0.57	0.559	1.425	-2.288	-1.017	0.019	0.895	1.825
9.4	30	เข้าใจ(รวม)	X	0.661	0.460	1.113	-3.240	-1.973	-0.796	0.428	1.472
	30	เข้าใจ(รวม)	Y	0.667	0.508	1.298	-2.926	-1.714	-0.583	0.320	1.326

ตาราง 53 สารสนเทศของข้อสอบและแบบทดสอบเทียบคะแนนที่วิเคราะห์ตามทฤษฎีการตอบ
ข้อสอบ โมเดลเกรด เรสพอนส์

		$I_i(\theta)$												
ข้อ	ฉบับ	-3.0	-2.6	-2.0	-1.6	-1	-0.60	0.0	0.6	1.0	1.6	2.0	2.6	3.0
1	X	0.566	0.574	0.579	0.578	0.577	0.574	0.558	0.499	0.417	0.265	0.177	0.088	0.053
1	Y	0.634	0.694	0.720	0.722	0.725	0.724	0.714	0.673	0.599	0.405	0.271	0.128	0.074
2*	X	0.525	0.587	0.598	0.604	0.642	0.656	0.637	0.567	0.531	0.553	0.572	0.489	0.371
2*	Y	0.535	0.588	0.605	0.607	0.641	0.658	0.637	0.561	0.522	0.544	0.569	0.498	0.384
3*	X	0.648	0.635	0.672	0.716	0.740	0.729	0.714	0.717	0.710	0.677	0.598	0.390	0.253
3*	Y	0.552	0.560	0.709	0.797	0.836	0.824	0.809	0.803	0.787	0.759	0.667	0.413	0.255
4*	X	0.633	0.654	0.664	0.668	0.670	0.667	0.664	0.647	0.637	0.563	0.450	0.257	0.160
4*	Y	0.544	0.586	0.620	0.628	0.640	0.642	0.639	0.625	0.610	0.533	0.426	0.249	0.157
5	X	0.643	1.042	1.395	1.444	1.441	1.421	1.434	1.225	0.851	0.333	0.154	0.044	0.019
5	Y	0.477	0.686	0.882	0.920	0.934	0.943	0.960	0.929	0.839	0.557	0.353	0.148	0.078
6	X	0.900	1.161	1.272	1.290	1.278	1.245	1.264	1.261	1.116	0.631	0.340	0.112	0.051
6	Y	0.907	1.192	1.217	1.273	1.399	1.428	1.432	1.343	1.100	0.531	0.263	0.080	0.034
7	X	0.486	0.976	1.660	1.767	1.742	1.804	1.807	1.315	0.752	0.225	0.090	0.022	0.008
7	Y	0.602	0.976	1.315	1.374	1.416	1.424	1.363	0.985	0.604	0.215	0.098	0.028	0.012
8	X	0.801	0.889	0.933	0.953	0.967	0.957	0.915	0.819	0.652	0.344	0.195	0.075	0.038
8	Y	0.687	0.965	1.142	1.168	1.178	1.190	1.193	1.150	1.022	0.607	0.343	0.122	0.057
9	X	1.055	1.061	0.831	0.895	1.204	1.296	1.333	1.333	1.231	0.774	0.440	0.150	0.068
9	Y	0.722	0.782	0.806	0.810	0.821	0.818	0.788	0.688	0.546	0.301	0.180	0.075	0.040
10	X	0.534	0.553	0.569	0.586	0.600	0.601	0.597	0.598	0.595	0.562	0.502	0.352	0.246
10	Y	0.345	0.585	0.943	1.054	1.091	1.094	1.105	1.088	0.986	0.632	0.380	0.146	0.072
11	X	1.200	1.523	1.476	1.624	1.769	1.774	1.633	1.541	1.488	0.791	0.374	0.099	0.038
11	Y	0.473	0.482	0.497	0.502	0.498	0.487	0.475	0.473	0.464	0.443	0.442	0.441	0.408
12*	X	0.492	0.505	0.517	0.519	0.528	0.532	0.530	0.515	0.501	0.492	0.478	0.394	0.306
12*	Y	0.450	0.462	0.486	0.494	0.501	0.507	0.507	0.496	0.483	0.472	0.461	0.396	0.319
13	X	0.895	0.904	0.953	0.976	0.984	0.979	0.967	0.936	0.890	0.638	0.412	0.172	0.089
13	Y	0.653	1.134	1.572	1.634	1.676	1.689	1.584	1.056	0.582	0.177	0.073	0.019	0.007
14	X	0.685	0.698	0.726	0.750	0.764	0.763	0.760	0.748	0.725	0.609	0.464	0.246	0.146
14	Y	0.148	0.357	1.041	1.544	1.810	1.839	1.791	1.666	1.382	0.603	0.268	0.068	0.026
15	X	0.606	0.657	0.679	0.683	0.688	0.701	0.713	0.691	0.646	0.494	0.357	0.184	0.109
15	Y	0.779	0.777	0.809	0.864	0.901	0.899	0.880	0.848	0.810	0.612	0.419	0.190	0.104
16*	X	0.614	0.633	0.667	0.678	0.683	0.681	0.672	0.651	0.615	0.474	0.344	0.178	0.106
16*	Y	0.505	0.559	0.660	0.693	0.713	0.716	0.705	0.681	0.661	0.555	0.425	0.230	0.139
17*	X	0.949	1.102	1.143	1.220	1.285	1.265	1.088	0.639	0.358	0.124	0.058	0.018	0.008
17*	Y	0.846	1.119	1.242	1.298	1.347	1.326	1.149	0.675	0.373	0.125	0.057	0.017	0.008
18	X	0.641	0.670	0.694	0.708	0.716	0.716	0.709	0.689	0.654	0.514	0.376	0.195	0.116
18	Y	0.496	0.499	0.525	0.539	0.549	0.556	0.562	0.561	0.554	0.515	0.453	0.313	0.220
19	X	0.320	0.326	0.336	0.339	0.344	0.346	0.345	0.337	0.330	0.321	0.316	0.295	0.263
19	Y	0.593	0.751	0.853	0.873	0.887	0.890	0.864	0.823	0.794	0.625	0.440	0.207	0.114
20	X	0.765	0.871	0.844	0.875	0.982	1.017	1.040	1.006	0.896	0.567	0.345	0.136	0.069
20	Y	0.699	0.684	0.764	0.816	0.846	0.854	0.845	0.816	0.784	0.617	0.438	0.210	0.117
21	X	0.516	0.667	0.781	0.804	0.815	0.818	0.818	0.799	0.746	0.553	0.383	0.183	0.103
21	Y	0.892	0.919	0.902	0.980	1.065	1.075	1.073	1.051	0.966	0.639	0.392	0.153	0.076
22*	X	0.360	0.374	0.401	0.412	0.419	0.420	0.419	0.414	0.407	0.379	0.340	0.254	0.192
22*	Y	0.429	0.437	0.474	0.494	0.509	0.513	0.513	0.508	0.497	0.453	0.391	0.268	0.189

ตาราง 53 (ต่อ)

		$I_r(\theta)$												
ข้อ	ฉบับ	-3.0	-2.6	-2.0	-1.6	-1	-0.60	0.0	0.6	1.0	1.6	2.0	2.6	3.0
23	X	0.595	0.626	0.643	0.657	0.669	0.677	0.688	0.685	0.660	0.546	0.421	0.234	0.144
23	Y	0.410	0.417	0.427	0.433	0.440	0.445	0.452	0.451	0.438	0.387	0.329	0.225	0.161
24	X	0.437	0.476	0.494	0.504	0.521	0.530	0.539	0.542	0.534	0.495	0.438	0.310	0.221
24	Y	0.850	0.820	0.873	0.955	1.013	1.017	0.978	0.798	0.577	0.271	0.146	0.053	0.026
25*	X	0.598	0.658	0.677	0.697	0.723	0.733	0.742	0.736	0.716	0.616	0.487	0.274	0.168
25*	Y	0.609	0.673	0.686	0.707	0.746	0.761	0.776	0.769	0.741	0.615	0.469	0.251	0.149
26	X	0.435	0.600	0.757	0.794	0.822	0.832	0.827	0.752	0.626	0.373	0.231	0.099	0.054
26	Y	0.617	0.638	0.651	0.671	0.683	0.688	0.691	0.690	0.684	0.664	0.625	0.471	0.335
27	X	0.438	0.490	0.519	0.528	0.540	0.545	0.551	0.550	0.544	0.517	0.474	0.356	0.262
27	Y	0.796	1.052	1.102	1.138	1.277	1.311	1.277	1.001	0.668	0.266	0.128	0.040	0.018
28	X	0.546	0.557	0.574	0.591	0.603	0.606	0.605	0.599	0.590	0.559	0.502	0.353	0.247
28	Y	0.559	0.556	0.566	0.599	0.639	0.650	0.651	0.640	0.625	0.565	0.475	0.299	0.196
29*	X	0.427	0.506	0.560	0.573	0.588	0.593	0.596	0.594	0.588	0.561	0.512	0.378	0.273
29*	Y	0.403	0.506	0.587	0.605	0.621	0.628	0.636	0.639	0.635	0.602	0.543	0.387	0.272
30*	X	0.363	0.378	0.387	0.389	0.390	0.390	0.390	0.386	0.377	0.342	0.299	0.217	0.163
30*	Y	0.462	0.497	0.517	0.522	0.526	0.530	0.532	0.524	0.506	0.438	0.360	0.229	0.155
$I_r(\theta)$	X	19.673	22.352	23.998	24.821	25.696	25.871	25.554	23.794	21.385	15.891	12.125	7.551	5.344
$SE(\theta)$	X	0.225	0.212	0.204	2.201	0.197	0.197	0.198	0.205	0.216	0.251	0.287	0.364	0.433
$I_r(\theta)$	Y	18.676	21.952	25.193	26.713	27.929	28.127	27.585	25.012	21.838	15.726	11.882	7.353	5.204
$SE(\theta)$	Y	0.231	0.213	0.199	0.193	0.189	0.189	0.190	0.200	0.214	0.252	0.290	0.369	0.438
$RE(\theta) = \frac{I_r(\theta)}{I_x(\theta)}$														
		0.949	0.982	1.05	1.706	1.086	1.087	1.079	1.051	1.021	0.990	0.980	0.974	0.974

* หมายถึง ข้อสอบร่วมในฉบับ X และ Y

$$RAI(\theta; Y, X) = \frac{AI(\theta, Y)}{AI(\theta, X)} = \frac{20.245}{19.543} = 1.036$$

ภาคผนวก ข

- ผลการแจกแจงคะแนนความสามารถของผู้สอบฉบับ X และ Y วิธีปรับค่าพารามิเตอร์แยกฉบับ
- ผลการแจกแจงคะแนนความสามารถของผู้สอบฉบับ X และ Y วิธีปรับค่าพารามิเตอร์พร้อมกัน
- ผลการวิเคราะห์ ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และตำแหน่งเปอร์เซ็นต์ไทล์ในการเทียบคะแนนสังเกตวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก
- ผลการวิเคราะห์ ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และตำแหน่งเปอร์เซ็นต์ไทล์ในการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบ
- ผลการวิเคราะห์ ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และตำแหน่งเปอร์เซ็นต์ไทล์ในการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน
- ผลการวิเคราะห์ ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และ ตำแหน่งเปอร์เซ็นต์ไทล์ในการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์

ตาราง 54 ผลการแจกแจงคะแนนความสามารถของผู้สอบฉบับ X และ Y วิธีปรับค่า
พารามิเตอร์แยกฉบับ

ลำดับ ที่	ฉบับ Y		Y → X (วิธี MS)		Y → X (วิธี TRF)		ฉบับ X	
	θ_i	$\psi_i(\theta_i)$	θ_n	$\psi_n(\theta_n)$	θ_n	$\psi_n(\theta_n)$	θ_i	$\psi_i(\theta_i)$
1	-1.250	0.000000	-1.333	0.000000	-1.343	0.000000	-1.250	0.000866
2	-1.125	0.003457	-1.206	0.003457	-1.215	0.003457	-1.125	0.004329
3	-1.000	0.007779	-1.078	0.007779	-1.087	0.007779	-1.000	0.009524
4	-0.875	0.029386	-0.950	0.029386	-0.959	0.029386	-0.875	0.027706
5	-0.750	0.044080	-0.823	0.044080	-0.831	0.044080	-0.750	0.048485
6	-0.625	0.063094	-0.695	0.063094	-0.702	0.063094	-0.625	0.080519
7	-0.500	0.076059	-0.568	0.076059	-0.574	0.076059	-0.500	0.077922
8	-0.375	0.076059	-0.440	0.076059	-0.446	0.076059	-0.375	0.079654
9	-0.250	0.076923	-0.312	0.076923	-0.318	0.076923	-0.250	0.075325
10	-0.125	0.064823	-0.185	0.064823	-0.190	0.064823	-0.125	0.059740
11	0.000	0.055315	-0.057	0.055315	-0.062	0.055315	0.000	0.043290
12	0.125	0.039758	0.070	0.039758	0.066	0.039758	0.125	0.048485
13	0.250	0.038894	0.198	0.038894	0.195	0.038894	0.250	0.033766
14	0.375	0.038029	0.326	0.038029	0.323	0.038029	0.375	0.032900
15	0.500	0.023336	0.453	0.023336	0.451	0.023336	0.500	0.038961
16	0.625	0.032844	0.581	0.032844	0.579	0.032844	0.625	0.029437
17	0.750	0.033708	0.708	0.033708	0.707	0.033708	0.750	0.024242
18	0.875	0.032844	0.836	0.032844	0.835	0.032844	0.875	0.032035
19	1.000	0.033708	0.964	0.033708	0.964	0.033708	1.000	0.032900
20	1.125	0.031979	1.091	0.031979	1.092	0.031979	1.125	0.023377
21	1.250	0.026793	1.219	0.026793	1.220	0.026793	1.250	0.024242
22	1.375	0.018150	1.346	0.018150	1.348	0.018150	1.375	0.023377
23	1.500	0.025065	1.474	0.025065	1.476	0.025065	1.500	0.024242
24	1.625	0.019015	1.602	0.019015	1.604	0.019015	1.625	0.019048
25	1.750	0.017286	1.729	0.017286	1.732	0.017286	1.750	0.022511
26	1.875	0.019879	1.857	0.019879	1.861	0.019879	1.875	0.016450
27	2.000	0.014693	1.984	0.014693	1.989	0.014693	2.000	0.015584
28	2.125	0.009507	2.112	0.009507	2.117	0.009507	2.125	0.008658
29	2.250	0.007779	2.239	0.007779	2.245	0.007779	2.250	0.012121
30	2.375	0.015557	2.367	0.015557	2.373	0.015557	2.375	0.008658
31	2.500	0.008643	2.495	0.008643	2.501	0.008643	2.500	0.006926
32	2.625	0.009507	2.622	0.009507	2.629	0.009507	2.625	0.006061
33	2.750	0.002593	2.750	0.002593	2.758	0.002593	2.750	0.002597
34	2.875	0.001729	2.877	0.001729	2.886	0.001729	2.875	0.001732
35	3.000	0.000000	3.005	0.000000	3.014	0.000000	3.000	0.001732
36	3.125	0.001729	3.133	0.001729	3.142	0.001729	3.125	0.002597

ตาราง 55 ผลการแจกแจงคะแนนความสามารถของผู้สอบฉบับ X และ Y วิธีปรับค่าพารามิเตอร์
พร้อมกัน

ลำดับ ที่	ฉบับ Y		ฉบับ X	
	θ_i	$\psi_i(\theta_i)$	θ_i	$\psi_i(\theta_i)$
1	-1.250	0.00086	-1.250	0.00000
2	-1.125	0.00432	-1.125	0.00260
3	-1.000	0.01642	-1.000	0.00693
4	-0.875	0.03198	-0.875	0.02078
5	-0.750	0.07952	-0.750	0.04069
6	-0.625	0.03630	-0.625	0.06667
7	-0.500	0.08557	-0.500	0.08918
8	-0.375	0.07087	-0.375	0.08485
9	-0.250	0.06914	-0.250	0.08225
10	-0.125	0.06223	-0.125	0.05801
11	0.000	0.04927	0.000	0.04502
12	0.125	0.03717	0.125	0.05281
13	0.250	0.04235	0.250	0.03290
14	0.375	0.03111	0.375	0.03377
15	0.500	0.02593	0.500	0.04156
16	0.625	0.03371	0.625	0.03030
17	0.750	0.03025	0.750	0.02251
18	0.875	0.03111	0.875	0.03290
19	1.000	0.03544	1.000	0.03377
20	1.125	0.02939	1.125	0.02597
21	1.250	0.02679	1.250	0.02251
22	1.375	0.01815	1.375	0.02424
23	1.500	0.02506	1.500	0.02511
24	1.625	0.01901	1.625	0.01991
25	1.750	0.01469	1.750	0.01991
26	1.875	0.02161	1.875	0.01818
27	2.000	0.01469	2.000	0.01472
28	2.125	0.00951	2.125	0.01039
29	2.250	0.00691	2.250	0.01299
30	2.375	0.01556	2.375	0.00693
31	2.500	0.00951	2.500	0.00693
32	2.625	0.00951	2.625	0.00606
33	2.750	0.00173	2.750	0.00260
34	2.875	0.00259	2.875	0.00346
35	3.000	0.00000	3.000	0.00000
36	3.125	0.00173	3.125	0.00260

ตาราง 56 ผลการวิเคราะห์ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และตำแหน่งเปอร์เซ็นต์ไทล์
ในการเทียบคะแนนสังเกต วิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์
ความยาก

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
0	0.000001	0.000001	0.00005	0.000001	0.000001	0.0000
1	0.000001	0.000002	0.00015	0.000001	0.000002	1.0000
2	0.000001	0.000003	0.00025	0.000001	0.000003	2.0000
3	0.000001	0.000004	0.00035	0.000001	0.000004	3.0000
4	0.000001	0.000005	0.00045	0.000001	0.000005	4.0000
5	0.000001	0.000006	0.00055	0.000001	0.000006	5.0000
6	0.000001	0.000007	0.00065	0.000001	0.000007	6.0000
7	0.000001	0.000008	0.00075	0.000001	0.000008	7.0000
8	0.000001	0.000009	0.00085	0.000001	0.000009	8.0000
9	0.000001	0.000010	0.00095	0.000001	0.000010	9.0000
10	0.000001	0.000011	0.00105	0.000001	0.000011	10.0000
11	0.000001	0.000012	0.00115	0.000001	0.000012	11.0000
12	0.000001	0.000013	0.00125	0.000001	0.000013	12.0000
13	0.000001	0.000014	0.00135	0.000001	0.000014	13.0000
14	0.000001	0.000015	0.00145	0.000001	0.000015	14.0000
15	0.000001	0.000016	0.00155	0.000001	0.000016	15.0000
16	0.000001	0.000017	0.00165	0.000001	0.000017	16.0000
17	0.000001	0.000018	0.00175	0.000001	0.000018	17.0000
18	0.000001	0.000019	0.00185	0.000001	0.000019	18.0000
19	0.000001	0.000020	0.00195	0.000001	0.000020	19.0000
20	0.000001	0.000021	0.00205	0.000001	0.000021	20.0000
21	0.000001	0.000022	0.00215	0.000001	0.000022	21.0000
22	0.000001	0.000023	0.00225	0.000001	0.000023	22.0000
23	0.000001	0.000024	0.00235	0.000001	0.000024	23.0000
24	0.000001	0.000025	0.00245	0.000001	0.000025	24.0000
25	0.000001	0.000026	0.00255	0.000001	0.000026	25.0000
26	0.000001	0.000027	0.00265	0.000001	0.000027	26.0000
27	0.000001	0.000028	0.00275	0.000001	0.000028	27.0000
28	0.000001	0.000029	0.00285	0.000001	0.000029	28.0000
29	0.000001	0.000030	0.00295	0.000001	0.000030	29.0000
30	0.000001	0.000031	0.00305	0.000001	0.000031	30.0000
31	0.000001	0.000032	0.00315	0.000001	0.000032	31.0000
32	0.000001	0.000033	0.00325	0.000001	0.000033	32.0000
33	0.000001	0.000034	0.00335	0.000001	0.000034	33.0000
34	0.000001	0.000035	0.00345	0.000001	0.000035	34.0000

ตาราง 56 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
35	0.000001	0.000036	0.00355	0.000001	0.000036	35.0000
36	0.000001	0.000037	0.00365	0.000001	0.000037	36.0000
37	0.000001	0.000038	0.00375	0.000001	0.000038	37.0000
38	0.000001	0.000039	0.00385	0.000001	0.000039	38.0000
39	0.000001	0.000040	0.00395	0.000001	0.000040	39.0000
40	0.000001	0.000041	0.00405	0.000001	0.000041	40.0000
41	0.000001	0.000042	0.00415	0.000001	0.000042	41.0000
42	0.000001	0.000043	0.00425	0.000001	0.000043	42.0000
43	0.000001	0.000044	0.00435	0.000001	0.000044	43.0000
44	0.000001	0.000045	0.00445	0.000001	0.000045	44.0000
45	0.000002	0.000047	0.00460	0.000001	0.000046	45.5000
46	0.000004	0.000051	0.00490	0.000001	0.000047	48.0000
47	0.000008	0.000059	0.00550	0.000001	0.000048	49.5000
48	0.000016	0.000075	0.00670	0.000002	0.000050	50.7500
49	0.000030	0.000105	0.00900	0.000005	0.000055	51.8929
50	0.000054	0.000159	0.01320	0.000008	0.000063	53.0319
51	0.000090	0.000249	0.02040	0.000016	0.000079	54.1410
52	0.000144	0.000393	0.03210	0.000028	0.000107	55.2236
53	0.000220	0.000613	0.05030	0.000047	0.000154	56.2957
54	0.000322	0.000935	0.07740	0.000078	0.000232	57.3566
55	0.000454	0.001389	0.11620	0.000123	0.000355	58.4065
56	0.000622	0.002011	0.17000	0.000186	0.000541	59.4454
57	0.000831	0.002842	0.24265	0.000272	0.000813	60.4755
58	0.001088	0.003930	0.33860	0.000385	0.001198	61.4979
59	0.001404	0.005334	0.46320	0.000531	0.001729	62.5108
60	0.001784	0.007118	0.62260	0.000715	0.002444	63.5201
61	0.002238	0.009356	0.82370	0.000944	0.003388	64.5240
62	0.002767	0.012123	1.07395	0.001227	0.004615	65.5218
63	0.003369	0.015492	1.38075	0.001571	0.006186	66.5109
64	0.004037	0.019529	1.75105	0.001991	0.008177	67.4864
65	0.004756	0.024285	2.19070	0.002495	0.010672	68.4437
66	0.005513	0.029798	2.70415	0.003094	0.013766	69.3836
67	0.006293	0.036091	3.29445	0.003796	0.017562	70.3053
68	0.007083	0.043174	3.96325	0.004604	0.022166	71.2092
69	0.007873	0.051047	4.71105	0.005518	0.027684	72.0959
70	0.008659	0.059706	5.53765	0.006532	0.034216	72.9673
71	0.009437	0.069143	6.44245	0.007638	0.041854	73.8253
72	0.010203	0.079346	7.42445	0.008821	0.050675	74.6723

ตาราง 56 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
73	0.010954	0.090300	8.48230	0.010062	0.060737	75.5104
74	0.011683	0.101983	9.61415	0.011335	0.072072	76.3282
75	0.012382	0.114365	10.81740	0.012607	0.084679	77.1438
76	0.013042	0.127407	12.08860	0.013841	0.098520	77.9596
77	0.013654	0.141061	13.42340	0.014996	0.113516	78.7766
78	0.014208	0.155269	14.81650	0.016035	0.129551	79.5955
79	0.014697	0.169966	16.26175	0.016928	0.146479	80.4139
80	0.015116	0.185082	17.75240	0.017658	0.164137	81.2347
81	0.015464	0.200546	19.28140	0.018221	0.182358	82.0613
82	0.015742	0.216288	20.84170	0.018627	0.200985	82.8934
83	0.015954	0.232242	22.42650	0.018891	0.219876	83.7306
84	0.016104	0.248346	24.02940	0.019035	0.238911	84.5725
85	0.016196	0.264542	25.64440	0.019074	0.257985	85.4192
86	0.016233	0.280775	27.26585	0.019020	0.277005	86.2715
87	0.016216	0.296991	28.88830	0.018881	0.295886	87.1219
88	0.016144	0.313135	30.50630	0.018656	0.314542	87.9919
89	0.016014	0.329149	32.11420	0.018345	0.332887	88.8598
90	0.015826	0.344975	33.70620	0.017947	0.350834	89.7326
91	0.015580	0.360555	35.27650	0.017465	0.368299	90.6106
92	0.015279	0.375834	36.81945	0.016907	0.385206	91.4940
93	0.014929	0.390763	38.32985	0.016288	0.401494	92.3872
94	0.014537	0.405300	39.80315	0.015629	0.417123	93.2874
95	0.014111	0.419411	41.23555	0.014954	0.432077	94.1950
96	0.013658	0.433069	42.62400	0.014291	0.446368	95.1097
97	0.013185	0.446254	43.96615	0.013666	0.460034	96.0307
98	0.012700	0.458954	45.26040	0.013096	0.473130	96.9563
99	0.012213	0.471167	46.50605	0.012593	0.485723	97.8838
100	0.011733	0.482900	47.70335	0.012155	0.497878	98.8100
101	0.011276	0.494176	48.85380	0.011772	0.509650	99.7316
102	0.010854	0.505030	49.96030	0.011429	0.521079	100.6465
103	0.010479	0.515509	51.02695	0.011109	0.532188	101.5542
104	0.010160	0.525669	52.05890	0.010801	0.542989	102.4571
105	0.009899	0.535568	53.06185	0.010503	0.553492	103.3587
106	0.009688	0.545256	54.04120	0.010221	0.563713	104.2614
107	0.009516	0.554772	55.00140	0.009969	0.573682	105.1689
108	0.009366	0.564138	55.94550	0.009764	0.583446	106.0834
109	0.009220	0.573358	56.87480	0.009615	0.593061	107.0051
110	0.009064	0.582422	57.78900	0.009524	0.602585	107.9310

ตาราง 56 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
111	0.008894	0.591316	58.68690	0.009483	0.612068	108.8560
112	0.008715	0.600031	59.56735	0.009472	0.621540	109.7743
113	0.008546	0.608577	60.43040	0.009465	0.631005	110.6813
114	0.008413	0.616990	61.27835	0.009440	0.640445	111.5755
115	0.008342	0.625332	62.11610	0.009378	0.649823	112.4600
116	0.008355	0.633687	62.95095	0.009276	0.659099	113.3420
117	0.008461	0.642148	63.79175	0.009143	0.668242	114.2323
118	0.008654	0.650802	64.64750	0.009002	0.677244	115.1430
119	0.008912	0.659714	65.52580	0.008885	0.686129	116.0859
120	0.009211	0.668925	66.43195	0.008821	0.694950	117.0710
121	0.009523	0.678448	67.36865	0.008834	0.703784	118.1048
122	0.009829	0.688277	68.33625	0.008931	0.712715	119.1886
123	0.010120	0.698397	69.33370	0.009101	0.721816	120.3171
124	0.010392	0.708789	70.35930	0.009318	0.731134	121.4784
125	0.010647	0.719436	71.41125	0.009541	0.740675	122.6536
126	0.010887	0.730323	72.48795	0.009733	0.750408	123.8288
127	0.011107	0.741430	73.58765	0.009864	0.760272	124.9971
128	0.011298	0.752728	74.70790	0.009926	0.770198	126.1580
129	0.011442	0.764170	75.84490	0.009931	0.780129	127.3152
130	0.011523	0.775693	76.99315	0.009910	0.790039	128.4732
131	0.011530	0.787223	78.14580	0.009905	0.799944	129.6341
132	0.011463	0.798686	79.29545	0.009951	0.809895	130.7943
133	0.011346	0.810032	80.43590	0.010071	0.819966	131.9437
134	0.011224	0.821256	81.56440	0.010269	0.830235	133.0708
135	0.011151	0.832407	82.68315	0.010531	0.840766	134.1686
136	0.011176	0.843583	83.79950	0.010834	0.851600	135.2369
137	0.011320	0.854903	84.92430	0.011151	0.862751	136.2824
138	0.011568	0.866471	86.06870	0.011462	0.874213	137.3149
139	0.011884	0.878355	87.24130	0.011745	0.885958	138.3430
140	0.012224	0.890579	88.44670	0.011974	0.897932	139.3731
141	0.012544	0.903123	89.68510	0.012108	0.910040	140.4097
142	0.012799	0.915922	90.95225	0.012108	0.922148	141.4573
143	0.012970	0.928892	92.24070	0.011964	0.934112	142.5216
144	0.013130	0.942022	93.54570	0.011717	0.945829	143.6148
145	0.013415	0.955437	94.87295	0.011435	0.957264	144.7537
146	0.013708	0.969145	96.22910	0.011121	0.968385	145.9520
147	0.013173	0.982318	97.57315	0.010600	0.978985	147.1931
148	0.010544	0.992862	98.75900	0.009532	0.988517	148.4027

ตาราง 56 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
149	0.005756	0.998618	99.57400	0.007506	0.996023	149.4623
150	0.001382	1.000000	99.93090	0.003977	1.000000	150.3300

ตาราง 57 ผลการวิเคราะห์ค่าความถี่สัมพัทธ์ ความถี่สะสมสัมพัทธ์ และตำแหน่งเปอร์เซ็นต์ไทล์
ในการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบ

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
0	0.000001	0.000001	0.000050	0.000001	0.000001	0.0000
1	0.000001	0.000002	0.000150	0.000001	0.000002	1.0000
2	0.000001	0.000003	0.000250	0.000001	0.000003	2.0000
3	0.000001	0.000004	0.000350	0.000001	0.000004	3.0000
4	0.000001	0.000005	0.000450	0.000001	0.000005	4.0000
5	0.000001	0.000006	0.000550	0.000001	0.000006	5.0000
6	0.000001	0.000007	0.000650	0.000001	0.000007	6.0000
7	0.000001	0.000008	0.000750	0.000001	0.000008	7.0000
8	0.000001	0.000009	0.000850	0.000001	0.000009	8.0000
9	0.000001	0.000010	0.000950	0.000001	0.000010	9.0000
10	0.000001	0.000011	0.001050	0.000001	0.000011	10.0000
11	0.000001	0.000012	0.001150	0.000001	0.000012	11.0000
12	0.000001	0.000013	0.001250	0.000001	0.000013	12.0000
13	0.000001	0.000014	0.001350	0.000001	0.000014	13.0000
14	0.000001	0.000015	0.001450	0.000001	0.000015	14.0000
15	0.000001	0.000016	0.001550	0.000001	0.000016	15.0000
16	0.000001	0.000017	0.001650	0.000001	0.000017	16.0000
17	0.000001	0.000018	0.001750	0.000001	0.000018	17.0000
18	0.000001	0.000019	0.001850	0.000001	0.000019	18.0000
19	0.000001	0.000020	0.001950	0.000001	0.000020	19.0000
20	0.000001	0.000021	0.002050	0.000001	0.000021	20.0000
21	0.000001	0.000022	0.002150	0.000001	0.000022	21.0000
22	0.000001	0.000023	0.002250	0.000001	0.000023	22.0000
23	0.000001	0.000024	0.002350	0.000001	0.000024	23.0000
24	0.000001	0.000025	0.002450	0.000001	0.000025	24.0000
25	0.000001	0.000026	0.002550	0.000001	0.000026	25.0000
26	0.000001	0.000027	0.002650	0.000001	0.000027	26.0000
27	0.000001	0.000028	0.002750	0.000001	0.000028	27.0000
28	0.000001	0.000029	0.002850	0.000001	0.000029	28.0000
29	0.000001	0.000030	0.002950	0.000001	0.000030	29.0000
30	0.000001	0.000031	0.003050	0.000001	0.000031	30.0000
31	0.000001	0.000032	0.003150	0.000001	0.000032	31.0000
32	0.000001	0.000033	0.003250	0.000001	0.000033	32.0000
33	0.000001	0.000034	0.003350	0.000001	0.000034	33.0000

ตาราง 57 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
34	0.000001	0.000035	0.003450	0.000001	0.000035	34.0000
35	0.000001	0.000036	0.003550	0.000001	0.000036	35.0000
36	0.000001	0.000037	0.003650	0.000001	0.000037	36.0000
37	0.000001	0.000038	0.003750	0.000001	0.000038	37.0000
38	0.000001	0.000039	0.003850	0.000001	0.000039	38.0000
39	0.000001	0.000040	0.003950	0.000001	0.000040	39.0000
40	0.000001	0.000041	0.004050	0.000001	0.000041	40.0000
41	0.000001	0.000042	0.004150	0.000001	0.000042	41.0000
42	0.000001	0.000043	0.004250	0.000001	0.000043	42.0000
43	0.000001	0.000044	0.004350	0.000001	0.000044	43.0000
44	0.000001	0.000045	0.004450	0.000001	0.000045	44.0000
45	0.000003	0.000048	0.004650	0.000001	0.000046	46.0000
46	0.000005	0.000053	0.005050	0.000001	0.000047	48.6000
47	0.000010	0.000063	0.005800	0.000001	0.000048	49.8750
48	0.000020	0.000083	0.007300	0.000002	0.000050	51.1250
49	0.000036	0.000119	0.010100	0.000005	0.000055	52.2857
50	0.000063	0.000182	0.015050	0.000008	0.000063	53.4255
51	0.000104	0.000286	0.023400	0.000016	0.000079	54.5163
52	0.000164	0.000450	0.036800	0.000028	0.000107	55.5699
53	0.000247	0.000697	0.057350	0.000047	0.000154	56.6195
54	0.000358	0.001055	0.087600	0.000078	0.000232	57.6636
55	0.000500	0.001555	0.130500	0.000123	0.000355	58.7015
56	0.000679	0.002234	0.189450	0.000186	0.000541	59.7315
57	0.000901	0.003135	0.268450	0.000272	0.000813	60.7548
58	0.001174	0.004309	0.372200	0.000385	0.001198	61.7722
59	0.001506	0.005815	0.506200	0.000531	0.001729	62.7845
60	0.001907	0.007722	0.676850	0.000715	0.002444	63.7926
61	0.002381	0.010103	0.891250	0.000944	0.003388	64.7948
62	0.002930	0.013033	1.156800	0.001227	0.004615	65.7896
63	0.003550	0.016583	1.480800	0.001571	0.006186	66.7745
64	0.004232	0.020815	1.869900	0.001991	0.008177	67.7470
65	0.004961	0.025776	2.329550	0.002495	0.010672	68.7047
66	0.005722	0.031498	2.863700	0.003094	0.013766	69.6459
67	0.006502	0.038000	3.474900	0.003796	0.017562	70.5698
68	0.007288	0.045288	4.164400	0.004604	0.022166	71.4725
69	0.008072	0.053360	4.932400	0.005518	0.027684	72.3468

ตาราง 57 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
70	0.008851	0.062211	5.778550	0.006532	0.034216	73.2067
71	0.009620	0.071831	6.702100	0.007638	0.041854	74.0544
72	0.010376	0.082207	7.701900	0.008821	0.050675	74.8924
73	0.011115	0.093322	8.776450	0.010062	0.060737	75.7229
74	0.011832	0.105154	9.923800	0.011335	0.072072	76.5479
75	0.012516	0.117670	11.141200	0.012607	0.084679	77.3597
76	0.013159	0.130829	12.424950	0.013841	0.098520	78.1694
77	0.013752	0.144581	13.770500	0.014996	0.113516	78.9817
78	0.014285	0.158866	15.172350	0.016035	0.129551	79.7970
79	0.014752	0.173618	16.624200	0.016928	0.146479	80.6155
80	0.015149	0.188767	18.119250	0.017658	0.164137	81.4360
81	0.015476	0.204243	19.650500	0.018221	0.182358	82.2595
82	0.015734	0.219977	21.211000	0.018627	0.200985	83.0889
83	0.015927	0.235904	22.794050	0.018891	0.219876	83.9237
84	0.016061	0.251965	24.393450	0.019035	0.238911	84.7634
85	0.016139	0.268104	26.003450	0.019074	0.257985	85.6078
86	0.016164	0.284268	27.618600	0.019020	0.277005	86.4569
87	0.016134	0.300402	29.233500	0.018881	0.295886	87.3119
88	0.016050	0.316452	30.842700	0.018656	0.314542	88.1722
89	0.015909	0.332361	32.440650	0.018345	0.332887	89.0377
90	0.015710	0.348071	34.021600	0.017947	0.350834	89.9084
91	0.015455	0.363526	35.579850	0.017465	0.368299	90.7843
92	0.015146	0.378672	37.109900	0.016907	0.385206	91.6656
93	0.014791	0.393463	38.606750	0.016288	0.401494	92.5529
94	0.014395	0.407858	40.066050	0.015629	0.417123	93.4488
95	0.013966	0.421824	41.484100	0.014954	0.432077	94.3540
96	0.013513	0.435337	42.858050	0.014291	0.446368	95.2662
97	0.013041	0.448378	44.185750	0.013666	0.460034	96.1844
98	0.012558	0.460936	45.465700	0.013096	0.473130	97.1065
99	0.012075	0.473011	46.697350	0.012593	0.485723	98.0299
100	0.011601	0.484612	47.881150	0.012155	0.497878	98.9512
101	0.011151	0.495763	49.018750	0.011772	0.509650	99.8673
102	0.010738	0.506501	50.113200	0.011429	0.521079	100.7764
103	0.010373	0.516874	51.168750	0.011109	0.532188	101.6783
104	0.010065	0.526939	52.190650	0.010801	0.542989	102.5745
105	0.009813	0.536752	53.184550	0.010503	0.553492	103.4692
106	0.009611	0.546363	54.155750	0.010221	0.563713	104.3675
107	0.009446	0.555809	55.108600	0.009969	0.573682	105.2709

ตาราง 57 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
108	0.009300	0.565109	56.045900	0.009764	0.583446	106.1816
109	0.009155	0.574264	56.968650	0.009615	0.593061	107.0992
110	0.009000	0.583264	57.876400	0.009524	0.602585	108.0205
111	0.008830	0.592094	58.767900	0.009483	0.612068	108.9402
112	0.008652	0.600746	59.642000	0.009472	0.621540	109.8527
113	0.008485	0.609231	60.498850	0.009465	0.631005	110.7535
114	0.008355	0.617586	61.340850	0.009440	0.640445	111.6415
115	0.008289	0.625875	62.173050	0.009378	0.649823	112.5201
116	0.008308	0.634183	63.002900	0.009276	0.659099	113.3969
117	0.008420	0.642603	63.839300	0.009143	0.668242	114.2826
118	0.008617	0.651220	64.691150	0.009002	0.677244	115.1895
119	0.008878	0.660098	65.565900	0.008885	0.686129	116.1292
120	0.009177	0.669275	66.468650	0.008821	0.694950	117.1111
121	0.009488	0.678763	67.401900	0.008834	0.703784	118.1417
122	0.009792	0.688555	68.365900	0.008931	0.712715	119.2220
123	0.010080	0.698635	69.359500	0.009101	0.721816	120.3464
124	0.010350	0.708985	70.381000	0.009318	0.731134	121.5029
125	0.010603	0.719588	71.428650	0.009541	0.740675	122.6727
126	0.010842	0.730430	72.500900	0.009733	0.750408	123.8427
127	0.011061	0.741491	73.596050	0.009864	0.760272	125.0059
128	0.011252	0.752743	74.711700	0.009926	0.770198	126.1619
129	0.011397	0.764140	75.844150	0.009931	0.780129	127.3144
130	0.011480	0.775620	76.988000	0.009910	0.790039	128.4680
131	0.011488	0.787108	78.136400	0.009905	0.799944	129.6246
132	0.011423	0.798531	79.281950	0.009951	0.809895	130.7807
133	0.011308	0.809839	80.418500	0.010071	0.819966	131.9262
134	0.011187	0.821026	81.543250	0.010269	0.830235	133.0498
135	0.011114	0.832140	82.658300	0.010531	0.840766	134.1444
136	0.011139	0.843279	83.770950	0.010834	0.851600	135.2098
137	0.011283	0.854562	84.892050	0.011151	0.862751	136.2527
138	0.011533	0.866095	86.032850	0.011462	0.874213	137.2828
139	0.011851	0.877946	87.202050	0.011745	0.885958	138.3087
140	0.012195	0.890141	88.404350	0.011974	0.897932	139.3370
141	0.012522	0.902663	89.640200	0.012108	0.910040	140.3722
142	0.012786	0.915449	90.905600	0.012108	0.922148	141.4187
143	0.012968	0.928417	92.193300	0.011964	0.934112	142.4822
144	0.013134	0.941551	93.498400	0.011717	0.945829	143.5744
145	0.013428	0.954979	94.826500	0.011435	0.957264	144.7130

ตาราง 57 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
146	0.013749	0.968728	96.185350	0.011121	0.968385	145.9127
147	0.013271	0.981999	97.536350	0.010600	0.978985	147.1583
148	0.010696	0.992695	98.734700	0.009532	0.988517	148.3773
149	0.005885	0.998580	99.563750	0.007506	0.996023	149.4486
150	0.001420	1.000000	99.929000	0.003977	1.000000	150.3215

ตาราง 58 ผลการวิเคราะห์ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสม และตำแหน่งเปอร์เซ็นต์ไทล์
ในการเทียบคะแนนสังเกตวิธีปรับค่าพารามิเตอร์พร้อมกัน

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
0	0.000001	0.000001	0.00005	0.000001	0.000001	0.0000
1	0.000001	0.000002	0.00015	0.000001	0.000002	1.0000
2	0.000001	0.000003	0.00025	0.000001	0.000003	2.0000
3	0.000001	0.000004	0.00035	0.000001	0.000004	3.0000
4	0.000001	0.000005	0.00045	0.000001	0.000005	4.0000
5	0.000001	0.000006	0.00055	0.000001	0.000006	5.0000
6	0.000001	0.000007	0.00065	0.000001	0.000007	6.0000
7	0.000001	0.000008	0.00075	0.000001	0.000008	7.0000
8	0.000001	0.000009	0.00085	0.000001	0.000009	8.0000
9	0.000001	0.000010	0.00095	0.000001	0.000010	9.0000
10	0.000001	0.000011	0.00105	0.000001	0.000011	10.0000
11	0.000001	0.000012	0.00115	0.000001	0.000012	11.0000
12	0.000001	0.000013	0.00125	0.000001	0.000013	12.0000
13	0.000001	0.000014	0.00135	0.000001	0.000014	13.0000
14	0.000001	0.000015	0.00145	0.000001	0.000015	14.0000
15	0.000001	0.000016	0.00155	0.000001	0.000016	15.0000
16	0.000001	0.000017	0.00165	0.000001	0.000017	16.0000
17	0.000001	0.000018	0.00175	0.000001	0.000018	17.0000
18	0.000001	0.000019	0.00185	0.000001	0.000019	18.0000
19	0.000001	0.000020	0.00195	0.000001	0.000020	19.0000
20	0.000001	0.000021	0.00205	0.000001	0.000021	20.0000
21	0.000001	0.000022	0.00215	0.000001	0.000022	21.0000
22	0.000001	0.000023	0.00225	0.000001	0.000023	22.0000
23	0.000001	0.000024	0.00235	0.000001	0.000024	23.0000
24	0.000001	0.000025	0.00245	0.000001	0.000025	24.0000
25	0.000001	0.000026	0.00255	0.000001	0.000026	25.0000
26	0.000001	0.000027	0.00265	0.000001	0.000027	26.0000
27	0.000001	0.000028	0.00275	0.000001	0.000028	27.0000
28	0.000001	0.000029	0.00285	0.000001	0.000029	28.0000
29	0.000001	0.000030	0.00295	0.000001	0.000030	29.0000
30	0.000001	0.000031	0.00305	0.000001	0.000031	30.0000
31	0.000001	0.000032	0.00315	0.000001	0.000032	31.0000
32	0.000001	0.000033	0.00325	0.000001	0.000033	32.0000
33	0.000001	0.000034	0.00335	0.000001	0.000034	33.0000
34	0.000001	0.000035	0.00345	0.000001	0.000035	34.0000
35	0.000001	0.000036	0.00355	0.000001	0.000036	35.0000

ตาราง 58 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
36	0.000001	0.000037	0.00365	0.000001	0.000037	36.0000
37	0.000001	0.000038	0.00375	0.000001	0.000038	37.0000
38	0.000001	0.000039	0.00385	0.000001	0.000039	38.0000
39	0.000001	0.000040	0.00395	0.000001	0.000040	39.0000
40	0.000001	0.000041	0.00405	0.000001	0.000041	40.0000
41	0.000001	0.000042	0.00415	0.000001	0.000042	41.0000
42	0.000001	0.000043	0.00425	0.000001	0.000043	42.0000
43	0.000001	0.000044	0.00435	0.000001	0.000044	43.0000
44	0.000001	0.000045	0.00445	0.000001	0.000045	44.0000
45	0.000001	0.000046	0.00455	0.000001	0.000046	45.0000
46	0.000002	0.000048	0.00470	0.000001	0.000047	46.5000
47	0.000003	0.000051	0.00495	0.000001	0.000048	48.2500
48	0.000005	0.000056	0.00535	0.000002	0.000050	49.6000
49	0.000008	0.000064	0.00600	0.000003	0.000053	50.7000
50	0.000014	0.000078	0.00710	0.000005	0.000058	51.6579
51	0.000025	0.000103	0.00905	0.000010	0.000068	52.6029
52	0.000042	0.000145	0.01240	0.000019	0.000087	53.5508
53	0.000069	0.000214	0.01795	0.000034	0.000121	54.4915
54	0.000108	0.000322	0.02680	0.000059	0.000180	55.3889
55	0.000165	0.000487	0.04044	0.000099	0.000279	56.2994
56	0.000244	0.000731	0.06089	0.000157	0.000436	57.2269
57	0.000352	0.001083	0.09069	0.000238	0.000674	58.1676
58	0.000495	0.001578	0.13303	0.000349	0.001023	59.1212
59	0.000682	0.002260	0.19187	0.000495	0.001518	60.0888
60	0.000920	0.003180	0.27196	0.000681	0.002199	61.0688
61	0.001216	0.004395	0.37874	0.000916	0.003115	62.0571
62	0.001578	0.005973	0.51842	0.001208	0.004322	63.0508
63	0.002011	0.007984	0.69784	0.001565	0.005887	64.0466
64	0.002523	0.010506	0.92451	0.001997	0.007884	65.0422
65	0.003117	0.013623	1.20647	0.002511	0.010394	66.0361
66	0.003798	0.017421	1.55222	0.003116	0.013510	67.0277
67	0.004570	0.021992	1.97065	0.003813	0.017323	68.0174
68	0.005431	0.027423	2.47073	0.004606	0.021930	69.0056
69	0.006371	0.033794	3.06084	0.005494	0.027424	69.9922
70	0.007370	0.041164	3.74788	0.006470	0.033894	70.9761
71	0.008395	0.049559	4.53612	0.007530	0.041424	71.9545

ตาราง 58 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
72	0.009398	0.058956	5.42573	0.008663	0.050086	72.9232
73	0.010329	0.069285	6.41206	0.009856	0.059942	73.8767
74	0.011138	0.080423	7.48543	0.011091	0.071033	74.8094
75	0.011795	0.092219	8.63210	0.012348	0.083381	75.7161
76	0.012294	0.104513	9.83657	0.013601	0.096982	76.5934
77	0.012657	0.117170	11.08413	0.014818	0.111800	77.4353
78	0.012931	0.130101	12.36354	0.015966	0.127766	78.2413
79	0.013178	0.143279	13.66899	0.017013	0.144779	79.0246
80	0.013454	0.156733	15.00059	0.017927	0.162706	79.7916
81	0.013794	0.170527	16.36299	0.018684	0.181390	80.5494
82	0.014204	0.184731	17.76287	0.019270	0.200660	81.2987
83	0.014664	0.199394	19.20626	0.019677	0.220337	82.0538
84	0.015131	0.214525	20.69598	0.019907	0.240244	82.8201
85	0.015556	0.230081	22.23030	0.019967	0.260211	83.5987
86	0.015897	0.245977	23.80291	0.019870	0.280081	84.3887
87	0.016129	0.262106	25.40417	0.019626	0.299707	85.1910
88	0.016245	0.278351	27.02283	0.019247	0.318954	86.0041
89	0.016256	0.294606	28.64783	0.018744	0.337699	86.8259
90	0.016179	0.310785	30.26954	0.018133	0.355832	87.6552
91	0.016036	0.326820	31.88024	0.017436	0.373268	88.4921
92	0.015844	0.342664	33.47420	0.016680	0.389949	89.3423
93	0.015613	0.358276	35.04701	0.015900	0.405848	90.2043
94	0.015349	0.373625	36.59510	0.015131	0.420979	91.0803
95	0.015048	0.388673	38.11493	0.014407	0.435386	91.9725
96	0.014710	0.403383	39.60281	0.013757	0.449143	92.8824
97	0.014333	0.417716	41.05494	0.013194	0.462337	93.8107
98	0.013919	0.431635	42.46753	0.012721	0.475058	94.7566
99	0.013471	0.445106	43.83702	0.012325	0.487383	95.7169
100	0.013001	0.458107	45.16062	0.011983	0.499366	96.6867
101	0.012522	0.470629	46.43678	0.011668	0.511034	97.6597
102	0.012051	0.482680	47.66544	0.011357	0.522392	98.6295
103	0.011607	0.494287	48.84836	0.011036	0.533428	99.5918
104	0.011206	0.505494	49.98904	0.010706	0.544134	100.5449
105	0.010860	0.516354	51.09237	0.010379	0.554514	100.9887
106	0.010572	0.526926	52.16401	0.010080	0.564594	102.4338
107	0.010337	0.537264	53.20950	0.009834	0.574428	103.3792

ตาราง 58 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
108	0.010142	0.547406	54.23350	0.009657	0.584084	104.3319
109	0.009969	0.557375	55.23907	0.009556	0.593640	105.2954
110	0.009796	0.567171	56.22730	0.009524	0.603163	106.2697
111	0.009611	0.576781	57.19760	0.009536	0.612699	107.2507
112	0.009411	0.586192	58.14866	0.009562	0.622261	108.2310
113	0.009205	0.595396	59.07942	0.009567	0.631827	109.2022
114	0.009013	0.604409	59.99028	0.009525	0.641352	110.1576
115	0.008859	0.613268	60.88384	0.009422	0.650773	111.0951
116	0.008768	0.622035	61.76516	0.009262	0.660035	112.0180
117	0.008753	0.630788	62.64118	0.009065	0.669100	112.9339
118	0.008820	0.639608	63.51980	0.008868	0.677967	113.8539
119	0.008958	0.648565	64.40866	0.008711	0.686678	114.7903
120	0.009151	0.657716	65.31408	0.008632	0.695310	115.7556
121	0.009377	0.667093	66.24044	0.008654	0.703963	116.7614
122	0.009620	0.676712	67.19024	0.008783	0.712746	117.8161
123	0.009871	0.686583	68.16475	0.009000	0.721746	118.9225
124	0.010131	0.696714	69.16484	0.009269	0.731014	120.0759
125	0.010407	0.707122	70.19179	0.009541	0.740555	121.2636
126	0.010703	0.717825	71.24733	0.009772	0.750326	122.4690
127	0.011013	0.728838	72.33317	0.009928	0.760254	123.6711
128	0.011320	0.740159	73.44985	0.009997	0.770251	124.8652
129	0.011595	0.751754	74.59562	0.009991	0.780243	126.0528
130	0.011806	0.763560	75.76570	0.009944	0.790187	127.2384
131	0.011931	0.775491	76.95257	0.009902	0.800089	128.4274
132	0.011959	0.787450	78.14709	0.009907	0.809995	129.6235
133	0.011905	0.799356	79.34031	0.009995	0.819991	130.8248
134	0.011802	0.811158	80.52568	0.010179	0.830170	132.0217
135	0.011702	0.822860	81.70090	0.010444	0.840615	133.2017
136	0.011657	0.834517	82.86888	0.010757	0.851372	134.3545
137	0.011703	0.846221	84.03690	0.011077	0.862449	135.4765
138	0.011849	0.858070	85.21451	0.011370	0.873820	136.5698
139	0.012070	0.870140	86.41047	0.011617	0.885437	137.6456
140	0.012340	0.882480	87.63098	0.011814	0.897251	138.7144
141	0.012633	0.895113	88.87965	0.011953	0.909204	139.7844
142	0.012919	0.908032	90.15725	0.012021	0.921225	140.8615
143	0.013140	0.921172	91.46021	0.011997	0.933223	141.9490

ตาราง 58 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
144	0.013263	0.934435	92.78036	0.011875	0.945098	143.0483
145	0.013377	0.947812	94.11236	0.011661	0.956759	144.1653
146	0.013635	0.961447	95.46295	0.011326	0.968085	145.3174
147	0.013757	0.975204	96.83255	0.010746	0.978832	146.5223
148	0.012559	0.987763	98.14835	0.009654	0.988485	147.7747
149	0.008828	0.996591	99.21769	0.007577	0.996062	148.9872
150	0.003409	1.000000	99.82954	0.003937	1.000000	150.0700

ตาราง 59 ผลการวิเคราะห์ค่าความถี่สัมพัทธ์ ความถี่สัมพัทธ์สะสมและตำแหน่งเปอร์เซ็นต์ไทล์
ของการเทียบคะแนนอีควิเปอร์เซ็นต์ไทล์

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
0	0.000001	0.000001	0.00005	0.000001	0.000001	0.0000
1	0.000001	0.000002	0.00015	0.000001	0.000002	1.0000
2	0.000001	0.000003	0.00025	0.000001	0.000003	2.0000
3	0.000001	0.000004	0.00035	0.000001	0.000004	3.0000
4	0.000001	0.000005	0.00045	0.000001	0.000005	4.0000
5	0.000001	0.000006	0.00055	0.000001	0.000006	5.0000
6	0.000001	0.000007	0.00065	0.000001	0.000007	6.0000
7	0.000001	0.000008	0.00075	0.000001	0.000008	7.0000
8	0.000001	0.000009	0.00085	0.000001	0.000009	8.0000
9	0.000001	0.000010	0.00095	0.000001	0.000010	9.0000
10	0.000001	0.000011	0.00105	0.000001	0.000011	10.0000
11	0.000001	0.000012	0.00115	0.000001	0.000012	11.0000
12	0.000001	0.000013	0.00125	0.000001	0.000013	12.0000
13	0.000001	0.000014	0.00135	0.000001	0.000014	13.0000
14	0.000001	0.000015	0.00145	0.000001	0.000015	14.0000
15	0.000001	0.000016	0.00155	0.000001	0.000016	15.0000
16	0.000001	0.000017	0.00165	0.000001	0.000017	16.0000
17	0.000001	0.000018	0.00175	0.000001	0.000018	17.0000
18	0.000001	0.000019	0.00185	0.000001	0.000019	18.0000
19	0.000001	0.000020	0.00195	0.000001	0.000020	19.0000
20	0.000001	0.000021	0.00205	0.000001	0.000021	20.0000
21	0.000001	0.000022	0.00215	0.000001	0.000022	21.0000
22	0.000001	0.000023	0.00225	0.000001	0.000023	22.0000
23	0.000001	0.000024	0.00235	0.000001	0.000024	23.0000
24	0.000001	0.000025	0.00245	0.000001	0.000025	24.0000
25	0.000001	0.000026	0.00255	0.000001	0.000026	25.0000
26	0.000001	0.000027	0.00265	0.000001	0.000027	26.0000
27	0.000001	0.000028	0.00275	0.000001	0.000028	27.0000
28	0.000001	0.000029	0.00285	0.000001	0.000029	28.0000
29	0.000001	0.000030	0.00295	0.000001	0.000030	29.0000
30	0.000001	0.000031	0.00305	0.000001	0.000031	30.0000
31	0.000001	0.000032	0.00315	0.000001	0.000032	31.0000
32	0.000001	0.000033	0.00325	0.000001	0.000033	32.0000
33	0.000001	0.000034	0.00335	0.000001	0.000034	33.0000
34	0.000001	0.000035	0.00345	0.000001	0.000035	34.0000
35	0.000001	0.000036	0.00355	0.000001	0.000036	35.0000

ตาราง 59 (ต่อ)

คะแนน ฉบับ Y	ความดีสัมพัทธ์ ฉบับ Y	ความดีสัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความดีสัมพัทธ์ ฉบับ X	ความดีสัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
36	0.000001	0.000037	0.00365	0.000001	0.000037	36.0000
37	0.000001	0.000038	0.00375	0.000001	0.000038	37.0000
38	0.000001	0.000039	0.00385	0.000001	0.000039	38.0000
39	0.000001	0.000040	0.00395	0.000001	0.000040	39.0000
40	0.000001	0.000041	0.00405	0.000001	0.000041	40.0000
41	0.000001	0.000042	0.00415	0.000001	0.000042	41.0000
42	0.000001	0.000043	0.00425	0.000001	0.000043	42.0000
43	0.000001	0.000044	0.00435	0.000001	0.000044	43.0000
44	0.000001	0.000045	0.00445	0.000001	0.000045	44.0000
45	0.000001	0.000046	0.00455	0.000001	0.000046	45.0000
46	0.000001	0.000047	0.00465	0.000001	0.000047	46.0000
47	0.000001	0.000048	0.00475	0.000001	0.000048	47.0000
48	0.000001	0.000049	0.00485	0.000001	0.000049	48.0000
49	0.000001	0.000050	0.00495	0.000001	0.000050	49.0000
50	0.000001	0.000051	0.00505	0.000001	0.000051	50.0000
51	0.000001	0.000052	0.00515	0.000001	0.000052	51.0000
52	0.000001	0.000053	0.00525	0.000001	0.000053	52.0000
53	0.000001	0.000054	0.00535	0.000001	0.000054	53.0000
54	0.000001	0.000055	0.00545	0.000001	0.000055	54.0000
55	0.000001	0.000056	0.00555	0.000001	0.000056	55.0000
56	0.000001	0.000057	0.00565	0.000001	0.000057	56.0000
57	0.000001	0.000058	0.00575	0.000001	0.000058	57.0000
58	0.000001	0.000059	0.00585	0.000001	0.000059	58.0000
59	0.000596	0.000655	0.03568	0.000001	0.000060	59.8263
60	0.000001	0.000656	0.06551	0.000910	0.000970	60.1543
61	0.001947	0.002603	0.16292	0.000001	0.000971	61.7087
62	0.001334	0.003937	0.32699	0.003156	0.004127	62.2285
63	0.000993	0.004930	0.44337	0.000001	0.004128	63.6479
64	0.001875	0.006805	0.58677	0.002069	0.006197	64.3409
65	0.002003	0.008809	0.78069	0.000001	0.006198	65.8649
66	0.003912	0.012721	1.07646	0.004409	0.010607	67.5302
67	0.003425	0.016145	1.44330	0.000001	0.010608	68.2352
68	0.003840	0.019985	1.80654	0.005204	0.015811	68.8531
69	0.002722	0.022707	2.13464	0.006384	0.022195	69.3671
70	0.004922	0.027630	2.51686	0.011548	0.033743	69.7575
71	0.008744	0.036374	3.20018	0.006371	0.040114	70.3492
72	0.007794	0.044168	4.02708	0.007649	0.047763	71.5204
73	0.009363	0.053530	4.88490	0.008776	0.056539	72.6238

ตาราง 59 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์ สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
74	0.011585	0.065115	5.93228	0.017481	0.074020	73.6593
75	0.014720	0.079835	7.24754	0.010567	0.084586	74.4117
76	0.010194	0.090030	8.49325	0.016020	0.100606	75.5216
77	0.010215	0.100244	9.51369	0.014871	0.115478	76.1586
78	0.020145	0.120389	11.03168	0.022336	0.137814	77.1530
79	0.012270	0.132659	12.65243	0.023700	0.161514	77.9946
80	0.012492	0.145152	13.89055	0.017475	0.178989	78.5461
81	0.019620	0.164772	15.49617	0.020971	0.199960	79.2235
82	0.015202	0.179974	17.23729	0.022013	0.221973	80.1214
83	0.020330	0.200304	19.01390	0.021047	0.243020	81.0317
84	0.020484	0.220788	21.05458	0.018917	0.261937	81.9809
85	0.020555	0.241343	23.10651	0.027981	0.289918	82.9320
86	0.020780	0.262123	25.17325	0.021269	0.311188	83.9606
87	0.012353	0.274476	26.82991	0.011889	0.323077	84.7274
88	0.014013	0.288489	28.14822	0.017413	0.340490	85.1985
89	0.014274	0.302763	29.56258	0.024266	0.364756	85.7683
90	0.021693	0.324456	31.36095	0.020781	0.385537	86.7037
91	0.020729	0.345185	33.48207	0.015777	0.401314	88.1744
92	0.024140	0.369325	35.72551	0.016978	0.418292	89.1909
93	0.012704	0.382030	37.56774	0.019633	0.437925	90.0255
94	0.019340	0.401369	39.16994	0.015293	0.453218	90.8906
95	0.012862	0.414231	40.78003	0.016508	0.469726	91.8820
96	0.023566	0.437797	42.60141	0.012059	0.481785	92.8933
97	0.014913	0.452710	44.52536	0.011626	0.493410	93.9792
98	0.016181	0.468891	46.08006	0.020613	0.514023	94.9593
99	0.010769	0.479660	47.42754	0.009989	0.524012	95.8773
100	0.013370	0.493030	48.63450	0.011293	0.535305	96.8923
101	0.013719	0.506749	49.98895	0.006609	0.541914	97.8143
102	0.012295	0.519044	51.28962	0.013066	0.554980	98.4453
103	0.007971	0.527015	52.30292	0.007501	0.562481	99.4016
104	0.010032	0.537047	53.20308	0.005082	0.567563	100.2101
105	0.009489	0.546536	54.17916	0.009974	0.577537	101.4815
106	0.011522	0.558059	55.22973	0.007231	0.584767	102.2947
107	0.018278	0.576337	56.71975	0.011304	0.596071	104.4281
108	0.007434	0.583771	58.00537	0.012002	0.608073	105.8481
109	0.013712	0.597483	59.06270	0.012192	0.620265	107.0184
110	0.007832	0.605316	60.13994	0.012623	0.632887	107.9440

ตาราง 59 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
111	0.011348	0.616664	61.09896	0.014357	0.647244	108.7392
112	0.008697	0.625361	62.10121	0.007418	0.654662	109.5592
113	0.006903	0.632264	62.88124	0.007627	0.662289	110.1772
114	0.005327	0.637591	63.49277	0.007906	0.670196	110.6421
115	0.009707	0.647298	64.24449	0.006491	0.676686	111.1657
116	0.009119	0.656418	65.18581	0.008628	0.685314	112.1220
117	0.014197	0.670614	66.35160	0.009524	0.694838	113.6552
118	0.012338	0.682952	67.67833	0.009638	0.704477	115.5112
119	0.005061	0.688014	68.54829	0.010386	0.714863	116.5177
120	0.009405	0.697418	69.27158	0.004529	0.719392	117.2771
121	0.014039	0.711457	70.44375	0.003336	0.722727	118.4959
122	0.009153	0.720610	71.60337	0.006977	0.729704	119.7585
123	0.005063	0.725673	72.31417	0.018837	0.748541	121.5594
124	0.012526	0.738199	73.19362	0.014184	0.762725	122.6185
125	0.012128	0.750327	74.42631	0.013657	0.776381	123.2729
126	0.009397	0.759724	75.50256	0.008582	0.784963	123.9572
127	0.011308	0.771032	76.53781	0.011001	0.795964	124.6943
128	0.017201	0.788233	77.96325	0.008758	0.804723	125.8788
129	0.006555	0.794789	79.15108	0.010739	0.815462	127.0952
130	0.013134	0.807923	80.13555	0.008961	0.824423	128.1155
131	0.010276	0.818198	81.30603	0.005185	0.829607	129.2764
132	0.009020	0.827218	82.27081	0.008303	0.837910	130.3087
133	0.011097	0.838315	83.27667	0.010951	0.848861	131.8805
134	0.011734	0.850049	84.41822	0.010648	0.859509	133.0727
135	0.007689	0.857738	85.38936	0.006453	0.865963	133.9726
136	0.010168	0.867906	86.28221	0.014576	0.880539	135.0134
137	0.008275	0.876181	87.20433	0.009662	0.890201	135.9172
138	0.009144	0.885325	88.07528	0.012603	0.902804	136.5222
139	0.014243	0.899568	89.24465	0.013754	0.916558	137.6782
140	0.014674	0.914242	90.69052	0.009684	0.926241	138.7982
141	0.008450	0.922692	91.84671	0.005020	0.931262	139.6972
142	0.011773	0.934464	92.85782	0.011500	0.942762	140.9655
143	0.008563	0.943028	93.87461	0.013951	0.956713	142.1508
144	0.008325	0.951353	94.71902	0.007003	0.963715	142.8174
145	0.011106	0.962459	95.69059	0.009316	0.973032	143.5276
146	0.014842	0.977301	96.98799	0.007297	0.980329	145.1617
147	0.011194	0.988495	98.28978	0.007512	0.987841	146.8420
148	0.007521	0.996016	99.22555	0.005590	0.993430	148.2898

ตาราง 59 (ต่อ)

คะแนน ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ Y	ความถี่สัมพัทธ์สะสม ฉบับ Y	ตำแหน่งเปอร์เซ็นต์ไทล์ ฉบับ Y	ความถี่สัมพัทธ์ ฉบับ X	ความถี่สัมพัทธ์สะสม ฉบับ X	คะแนนแปลง ฉบับ Y
149	0.002377	0.998393	99.72046	0.003753	0.997184	149.5074
150	0.001607	1.000000	99.91966	0.002816	1.000000	150.2100

ภาคผนวก ค

- ตัวอย่างแบบพิจารณาความสอดคล้องระหว่างเนื้อหา จุดประสงค์กับระดับพฤติกรรมการวัด
- ตัวอย่างแบบประเมินความสอดคล้องระหว่างเนื้อหา ระดับพฤติกรรมกับข้อสอบ
- รายชื่อผู้เชี่ยวชาญ
- หนังสือขอความร่วมมือในการเก็บรวบรวมข้อมูล

ตัวอย่างแบบพิจารณาความสอดคล้องระหว่างเนื้อหา จุดประสงค์ กับระดับพฤติกรรมการวัด

คำชี้แจง

ขอให้ท่านพิจารณาว่า

1. การกำหนดจุดประสงค์เชิงพฤติกรรมในแต่ละเนื้อหาที่วิเคราะห์ไว้มีความเหมาะสมกับสมรรถภาพที่ต้องการวัดเพียงใด
2. ระดับพฤติกรรมที่พึงประสงค์ที่กำหนดไว้ในแต่ละเนื้อหา แต่ละจุดประสงค์มีความเหมาะสมเพียงใด และน้ำหนักที่กำหนดไว้เหมาะสมหรือไม่ (โดยคิดจากฐาน 100 เปอร์เซนต์ เพื่อให้ได้ข้อสอบที่เมื่อวิเคราะห์แล้วมีคุณภาพ จำแนกออกเป็น 2 ฉบับที่คู่ขนานกันตามเนื้อหา ฉบับละ 30 ข้อ)
3. ขอให้ท่านโปรดแก้ไขลงในรายละเอียด หรือในข้อวิจารณ์ข้างท้าย

ตารางวิเคราะห์เนื้อหา จุดประสงค์ และระดับพฤติกรรมการวัด

เนื้อหา	จุดประสงค์ เชิงพฤติกรรม	ระดับพฤติกรรม (จำนวนข้อ)				
		ความรู้ ความจำ	ความ เข้าใจ	การนำ ไปใช้	การ วิเคราะห์	รวม

ข้อวิจารณ์

เนื้อหาย่อที่ 1

เนื้อหาย่อที่ 2

เนื้อหาย่อที่ 3

แบบประเมินความสอดคล้องระหว่างเนื้อหา ระดับพฤติกรรมกับข้อสอบ

คำชี้แจง

ให้ท่านพิจารณาว่าข้อสอบแต่ละข้อวัดในเนื้อหาและพฤติกรรมที่ต้องการวัดตามที่กำหนดไว้หรือไม่ โดยพิจารณาให้น้ำหนัก ดังนี้

- +1 เท่ากับ แน่ใจว่าข้อสอบวัดได้ตรงตามเนื้อหาและพฤติกรรมที่ต้องการวัด
- 0 เท่ากับ ไม่แนใจว่าข้อสอบวัดได้ตรงตามเนื้อหาและพฤติกรรมที่ต้องการวัด
- 1 เท่ากับ แน่ใจว่าข้อสอบวัดได้ไม่ตรงตามเนื้อหาและพฤติกรรมที่ต้องการวัด

ลำดับ ที่	เนื้อหา/ จุดประสงค์	ข้อสอบ/ระดับพฤติกรรมการวัด	ผลการ ประเมิน			ความเห็นของ ผู้เชี่ยวชาญ
			-1	0	+1	

รายละเอียดขั้นตอนการสร้างแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบ

1. กำหนดจุดมุ่งหมายของการสอบ

เป็นการวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 เพื่อนำคะแนนที่ได้ไปทำการศึกษาเปรียบเทียบผลการเทียบคะแนนที่ได้จากวิธีการเทียบคะแนนที่แตกต่างกัน

2. ลักษณะของแบบทดสอบ

เป็นแบบทดสอบเลือกตอบชนิด 5 ตัวเลือก ที่มีคำตอบถูกหลายตัวเลือก ตั้งแต่ 2-4 ตัวเลือก โดยผู้สอบต้องระบุทุกตัวเลือกว่าถูกหรือผิด (MTF) เป็นการวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ ด้านเหตุผล ซึ่งวัดความสามารถด้านความคิดรวบยอดทางคณิตศาสตร์ ที่เกี่ยวข้องกับหลักการ วิธีการ และมโนคติในการคิดแบบนามธรรม ในเรื่องของระบบจำนวนเต็ม เศษส่วนและทศนิยม ก็คือเมื่อนักเรียนได้เรียนรู้กฎเกณฑ์ต่าง ๆ แล้วสามารถขยายความรู้ออกมาในรูปที่กว้างขวางกว่าเดิมได้หรือไม่ แบบทดสอบมีลักษณะไม่ได้มุ่งวัดทางด้านทักษะการคำนวณ แต่ต้องทราบวิธีการ ขั้นตอนการคำนวณ ความเข้าใจในการแปลความ ขยายความ และคิดหาเหตุผล

3. การวิเคราะห์เนื้อหา

กำหนดขอบเขตเนื้อหา เรื่องระบบจำนวนเต็ม เศษส่วนและทศนิยมให้ครอบคลุมพื้นฐานทางคณิตศาสตร์ที่ได้เรียนไปแล้ว ทั้งนี้ไม่รวมเศษส่วนและทศนิยมที่เป็นจำนวนลบโดยวิเคราะห์เนื้อหาตามหลักสูตรมัธยมศึกษาตอนต้นในวิชา ค 101 ค 102 และ ค 203 โดยกำหนดมโนคติที่สอดคล้องกับเนื้อหาและพฤติกรรมย่อยที่ใช้ในการวัด

4. การวิเคราะห์พฤติกรรม

วิเคราะห์พฤติกรรมในแต่ละเนื้อหาย่อย ที่สอดคล้องกับมโนคติ โดยยึดพฤติกรรมการเรียนรู้ที่พึงประสงค์ด้านสติปัญญาของวิลสันซึ่งแบ่งเป็น 4 ชั้น คือ

1. ความรู้ความจำเกี่ยวกับการคิดคำนวณ (Computation) ประกอบด้วย ความรู้ความจำเกี่ยวข้องกับข้อเท็จจริง ศัพท์ นิยาม และความสามารถในการใช้กระบวนการคิดคำนวณตามลำดับขั้นตอนที่ได้เคยเรียนรู้มาแล้ว

2. ความเข้าใจ (Comprehension) ประกอบด้วยความเข้าใจเกี่ยวกับมโนคติ หลักการทางคณิตศาสตร์ และการสรุปอ้างอิงเป็นกรณีทั่วไป ความเข้าใจเกี่ยวกับโครงสร้างทางคณิตศาสตร์ ความสามารถในการเปลี่ยนโจทย์ปัญหาจากแบบหนึ่งไปอีกรูปแบบหนึ่ง ความสามารถในการศึกษาตามแนวทางของเหตุผล และความสามารถในระหว่างการทำอ่านและตีความโจทย์คณิตศาสตร์

3. การนำไปใช้ (Application) ประกอบด้วยความสามารถในการแก้ปัญหาต่าง ๆ คล้ายคลึงกับที่เรียนมา ตลอดจนความสามารถในการเปรียบเทียบ วิเคราะห์ข้อมูล และความสามารถมองเห็นแบบลักษณะโครงสร้างที่เหมือนกันและสมมาตรกัน

4. การวิเคราะห์ (Analysis) ประกอบด้วยความสามารถในการแก้ปัญหาที่ไม่เคยประสบมาก่อน ซึ่งเป็นปัญหาที่ซับซ้อนไม่มีในแบบฝึกหัดหรือตัวอย่าง แต่อยู่ในขอบข่ายของเนื้อหาวิชาที่เรียน และความสามารถในการค้นหาความสัมพันธ์โดยการจัดส่วนต่าง ๆ ที่โจทย์กำหนดให้ใหม่เพื่อใช้ในการแก้ปัญหา

5. สร้างตารางวิเคราะห์เนื้อหา

สร้างตารางวิเคราะห์เนื้อหาและพฤติกรรมที่สอดคล้องกับจุดมุ่งหมายของหลักสูตร โดยกำหนดน้ำหนักของการวัดในแต่ละพฤติกรรมรวมทั้งสิ้นจำนวน 100 ข้อ

6. สร้างข้อสอบ

สร้างข้อสอบตามโครงสร้างตารางวิเคราะห์เนื้อหาจำนวน 100 ข้อ เพื่อนำไปใช้จริงจำนวน 50 ข้อ โดยแยกเป็นแบบทดสอบเทียบคะแนน 2 ฉบับที่วัดในเนื้อหาเดียวกัน ฉบับละ 30 ข้อ (แต่ละฉบับมีข้อสอบรวม 10 ข้อ)

รายนามผู้เชี่ยวชาญ

รายนามผู้มีความเชี่ยวชาญในการสอนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาที่ 2
พิจารณาการกำหนดน้ำหนักข้อสอบในแต่ละเนื้อหา และระดับพฤติกรรมการวัด (ลำดับ 1-4)
และตรวจสอบความเที่ยงตรงเชิงเนื้อหา (ลำดับ 1-6)

อาจารย์นิศา ด้านวิริยะกุล	อาจารย์โรงเรียนราชวินิตบางเขน เชี่ยวชาญด้านการวัดและประเมินผลการศึกษา และการสอน คณิตศาสตร์
อาจารย์เบญจา เขียวสม	อาจารย์โรงเรียนปากน้ำหลังสวนวิทยา เชี่ยวชาญด้านการวัดและประเมินผลการศึกษา และการสอน คณิตศาสตร์
อาจารย์สุชาติ สิริมินนนท์	อาจารย์โรงเรียนบางกะปิ เชี่ยวชาญด้านการวัดและประเมินผลการศึกษาและการสอน คณิตศาสตร์
ดร.จรรยา ภูอุดม	อาจารย์โรงเรียนดอนเมืองจตุรจินดา เชี่ยวชาญด้านการสอนคณิตศาสตร์
รศ.อุษาวดี จันทรสนธิ	อาจารย์ประจำสาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมมาธิราช เชี่ยวชาญด้านการสอนคณิตศาสตร์
รศ.ดร.ปรีชา เนาว์เย็นผล	อาจารย์ประจำสาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมมาธิราช เชี่ยวชาญด้านการสอนคณิตศาสตร์



ที่ ศธ 0812/

4232

สำนักงานสามัญศึกษากรุงเทพมหานคร
กรมสามัญศึกษา กระทรวงศึกษาธิการ
เขตดุสิต กรุงเทพมหานคร 10300

๒๖ กรกฎาคม 2545

เรื่อง ขอความร่วมมือในการวิจัย

เรียน ผู้อำนวยการโรงเรียน

สิ่งที่ส่งมาด้วย แบบทดสอบและรายชื่อโรงเรียนในสังกัด

ด้วย นางสาวสุนิสา จุยม่วงศรี นิสิตระดับปริญญาเอก สาขาการทดสอบและวัดผล การศึกษามหาวิทยาลัยศรีนครินทรวิโรฒ ได้รับอนุมัติให้ดำเนินการจัดทำปฏิญานิพนธ์เรื่อง "การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า" ในการนี้ ผู้วิจัยมีความประสงค์จะเก็บข้อมูล เพื่อการวิจัยโดยแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ เรื่อง จำนวนเต็มเศษส่วนและทศนิยม นักเรียน ระดับมัธยมศึกษาชั้นปีที่ 2 ในโรงเรียนสังกัดกรมสามัญศึกษา กรุงเทพมหานคร (รายชื่อโรงเรียนดังแนบ)

สำนักงานสามัญศึกษากรุงเทพมหานคร พิจารณาแล้วเห็นว่า การวิจัยดังกล่าว เป็นประโยชน์ต่อการพัฒนาคุณภาพการศึกษาและวงการศึกษาโดยรวม สมควรให้การสนับสนุน

จึงเรียนมาเพื่อทราบและโปรดให้ความอนุเคราะห์ในครั้งนี้นี้ด้วย

ขอแสดงความนับถือ

(นายเฉลิมชัย รัตนประยูร)

ผู้อำนวยการสามัญศึกษากรุงเทพมหานคร

กลุ่มงานส่งเสริมและพัฒนาคุณภาพการศึกษา

โทร. 0 2281-9996-7, 0 2282-3834 ต่อ 108,109

โทรสาร ต่อ 113,115 เว็บไซต์ <http://bkk.ge.go.th>

ภาคผนวก ง

- แบบทดสอบเทียบคะแนนฉบับ x
- แบบทดสอบเทียบคะแนนฉบับ y
- ตัวอย่างกระดาษคำตอบของผู้สอบ

ฉบับ x

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์
เรื่อง จำนวนเต็ม เศษส่วน และทศนิยม

คำชี้แจงในการตอบแบบทดสอบ

1. แบบทดสอบฉบับนี้เป็นแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบ โดยในหนึ่งข้อจะมีตัวเลือกที่เป็นคำตอบถูกต้องตั้งแต่สองตัวเลือกจนถึงสี่ตัวเลือก ในการตอบนักเรียนจะต้องพิจารณา ตอบทุกตัวเลือก โดยทำเครื่องหมายถูก (✓) หน้าตัวเลือกที่เห็นว่าถูก และทำเครื่องหมายผิด (X) หน้าตัวเลือกที่เห็นว่าผิด ลงในกระดาษคำตอบ
2. ข้อสอบมีทั้งหมด 30 ข้อ จำนวน 6 หน้า เวลาในการทำข้อสอบ 75 นาที
3. ให้นักเรียนกรอกรายละเอียดลงบนกระดาษคำตอบ แต่กรุณาอย่าทำเครื่องหมายใด ๆ ลงบนแบบทดสอบ ถ้าต้องการทดให้ทดในกระดาษที่สำรองไว้
4. ถ้านักเรียนมีข้อสงสัยประการใดให้ถามผู้ดำเนินการสอบก่อนลงมือทำ
5. ให้นักเรียนตรวจทานก่อนส่งกระดาษคำตอบว่าได้ทำเครื่องหมายลงในกระดาษคำตอบครบทุกตัวเลือก

ตัวอย่างการตอบแบบทดสอบ

<u>คำถาม</u>		<u>กระดาษคำตอบ</u>					
ข้อ 0	ข้อสรุปเกี่ยวกับจำนวนในข้อใดเป็นจริง	ข้อ	ก	ข	ค	ง	จ
ก.	1 เป็นจำนวนนับที่มีค่าน้อยที่สุด	0	✓	X	✓	X	X
ข.	0 เป็นจำนวนนับ	00					
ค.	1 เป็นจำนวนตรงข้ามของ -1						
ง.	จำนวนนับก็คือจำนวนเต็ม						
จ.	จำนวนเต็มแบ่งเป็นจำนวนเต็มบวก และจำนวนเต็มลบ						

1. ข้อสรุปเกี่ยวกับจำนวนเต็มในข้อใดเป็นจริง
 - ก. ถ้า a เป็นจำนวนเต็มบวกแล้ว a มากกว่าศูนย์
 - ข. ถ้า a เป็นจำนวนเต็มลบ และ b เป็นจำนวนเต็มบวก แล้ว a น้อยกว่า b
 - ค. ถ้า a เป็นจำนวนเต็มลบ แล้วจำนวนตรงข้ามของ a มีค่ามากกว่าศูนย์
 - ง. ถ้า b เป็นจำนวนตรงข้ามของ a แล้ว a อยู่ห่างจากศูนย์น้อยกว่า b
 - จ. ถ้า a และ b เป็นจำนวนเต็มลบ และ a น้อยกว่า b แล้ว a อยู่ห่างจากศูนย์น้อยกว่า b
2. ข้อใดเป็นจำนวนเต็มทุกจำนวน
 - ก. 1, 0, -1, -2
 - ข. 0, 2, 3, 4
 - ค. -1, -2, -3, -4
 - ง. 0.0, -1.5, -2.0, 3.0
 - จ. $\frac{-4}{2}, \frac{0}{2}, \frac{4}{2}, \frac{6}{2}$
3. ให้ a และ b เป็นจำนวนเต็มใด ๆ ค่าของ a, b ในข้อใดทำให้ผลลัพธ์ของ $a - b$ เป็นจำนวนเต็มลบ
 - ก. a เป็นจำนวนเต็มลบ และ b เป็นศูนย์
 - ข. a และ b เป็นจำนวนเต็มลบ โดยที่ a อยู่ห่างจากศูนย์มากกว่า b
 - ค. a และ b เป็นจำนวนเต็มบวก โดยที่ a อยู่ห่างจากศูนย์มากกว่า b
 - ง. a เป็นจำนวนเต็มลบ และ b เป็นจำนวนเต็มบวก
 - จ. a เป็นศูนย์ และ b เป็นจำนวนเต็มบวก
4. กำหนด x และ y เป็นจำนวนเต็มใด ๆ จงพิจารณาว่าข้อใดเป็นจริง
 - ก. $x + y = y + x$
 - ข. $x + y = -(-x - y)$
 - ค. $x + y = -x - y$
 - ง. $x + y = -(x - y)$
 - จ. $x + y = x + (-y)$
5. กำหนด $P + Q = 10$ และ $R + S = -12$ ดังนั้น $(P + Q) - (R + S)$ มีค่าเท่ากับข้อใด
 - ก. $10 - 12$
 - ข. $10 - (-12)$
 - ค. $10 + 12$
 - ง. $(-12) + 10$
 - จ. $-10 - 12$
6. $[(-10) + 8] - [(-9) + 2]$ มีค่าเท่ากับข้อใด
 - ก. $(-2) + 7$
 - ข. $(-17) + 10$
 - ค. $2 + (-7)$
 - ง. $5 + 0$
 - จ. $2 + 11$
7. ให้ a, b, c เป็นจำนวนเต็มใด ๆ ค่าของ a, b, c ในข้อใดทำให้ผลลัพธ์ของ $a \times b \times c$ เป็นจำนวนเต็มบวก
 - ก. a และ b เป็นจำนวนเต็มบวก c เป็นจำนวนเต็มลบ
 - ข. a เป็นจำนวนเต็มบวก b และ c เป็นจำนวนเต็มลบ
 - ค. a และ b เป็นจำนวนเต็มลบ c เป็นจำนวนเต็มบวก
 - ง. a, b และ c เป็นจำนวนเต็มบวก
 - จ. a, b และ c เป็นจำนวนเต็มลบ

8. ข้อใดสรุปผลการคูณ หรือหารจำนวนเต็มได้ถูกต้อง

- ก. $(-10) \times 5 = -50$ เป็นจริง เพราะว่าจำนวนเต็มลบคูณกับจำนวนเต็มบวกให้ผลลัพธ์เป็นจำนวนเต็มลบ
- ข. $10 \times (-5) = -50$ เป็นจริง เพราะว่าจำนวนเต็มบวกคูณกับจำนวนเต็มลบให้ผลลัพธ์เป็นจำนวนเต็มลบ
- ค. $(-10) \times (-5) = -50$ เป็นจริง เพราะว่าจำนวนเต็มลบคูณกับจำนวนเต็มลบให้ผลลัพธ์เป็นจำนวนเต็มลบ
- ง. $(-10) \div (-5) = -2$ เป็นจริง เพราะว่าจำนวนเต็มลบหารด้วยจำนวนเต็มลบให้ผลลัพธ์เป็นจำนวนเต็มลบ
- จ. $10 \div (-5) = 2$ ไม่จริง เพราะว่าจำนวนเต็มบวกหารด้วยจำนวนเต็มลบให้ผลลัพธ์เป็นจำนวนเต็มลบ

9. กำหนด $(a \times b) = -15$ และ $(c \times d) = 3$ ดังนั้นผลลัพธ์ของ $(a \times b) \div (c \times d)$ จะมีค่าเท่ากับข้อใด

- ก. $(-15) \div 3$
- ข. $15 \div (-3)$
- ค. $3 \div (-15)$
- ง. $(-15) \div [(-1) \times (-3)]$
- จ. $[(-15) \times (-1)] \div (-3)$

10. กำหนดให้ a และ b เป็นจำนวนเต็มลบ โดยที่ a มากกว่า b และ c เป็นจำนวนเต็มบวก ข้อใดสรุปผลลัพธ์ได้ถูกต้อง

- ก. $a - b + c$ เป็นจำนวนเต็มบวก
- ข. $a + b - c$ เป็นจำนวนเต็มบวก
- ค. $c - b - a$ เป็นจำนวนเต็มบวก
- ง. $b - a - c$ เป็นจำนวนเต็มลบ
- จ. $a - c + b$ เป็นจำนวนเต็มลบ

11. กำหนด a, b, c เป็นจำนวนเต็มลบใด ๆ และ $a = b$ ข้อใดสรุปผลลัพธ์ได้ถูกต้อง

- ก. $ac = bc$
- ข. ab มากกว่า 0
- ค. abc มากกว่า 0
- ง. $a^2 = b^2$
- จ. $a + c = b + c$

12. กำหนด a เป็นจำนวนเต็มใด ๆ ประโยคในข้อใดเป็นจริง

- ก. $a + 1 = 1 + a$
- ข. $a - 1 = 1 - a$
- ค. $a \times 1 = 1 \times a$
- ง. $\frac{1}{a} = \frac{a}{1}$
- จ. $\frac{a}{1} = a \times 1$

13. ให้ a, b และ c เป็นจำนวนเต็มใด ๆ ข้อกำหนดในข้อใดทำให้ $(a + b) \times c = 0$

- ก. c มีค่าเป็น 0
- ข. a เป็นจำนวนตรงข้ามของ b
- ค. $a + b$ มีค่าเท่ากับ 0
- ง. $a + c$ มีค่าเท่ากับ 0
- จ. $b \div c$ มีค่าเท่ากับ 1

14. กำหนด a เป็นจำนวนเต็มใด ๆ ข้อใดสรุปได้ถูกต้องตามสมบัติของศูนย์ และหนึ่ง

ก. $a \times [(-5) + 5] = 0$ เป็นจริง เพราะว่า

$(-5) + 5 = 0$ และ $a \times 0 = 0$

ข. $a + [(-3) + 3] = a$ เป็นจริง เพราะว่า

$(-3) + 3 = 0$ และ $a + 0 = a$

ค. $a \div [(-1) \times 1] = a$ เป็นจริง เพราะว่า

$(-1) \times 1 = -1$ และ $a \div -1 = -a$

ง. $a \div [(-11) + 11] = 0$ ไม่จริง เพราะว่า

$(-11) + 11 = 0$ และ $a \div 0 =$ หาค่าไม่ได้

จ. $a \times [(-5) + 6] = 0$ ไม่จริง เพราะว่า

$(-5) + 6 = 1$ และ $a \times 1 = a$

15. ถ้า $(a \times 10^2) + (b \times 10^2) = c \times 10^2$ โดย

ที่ a และ b เป็นจำนวนเต็มบวก แล้ว

ข้อใดกล่าวถูกต้อง

ก. b น้อยกว่า c

ข. a น้อยกว่า c

ค. a มากกว่า b

ง. $ab = c$

จ. $a + b = c$

16. กำหนด x, y, z เป็นจำนวนเต็มใด ๆ

ข้อใดกล่าวถึงสมบัติของจำนวนเต็มได้ถูกต้อง

ก. $x(yz) = x(zy)$ ตามสมบัติการสลับที่ของ

การคูณ

ข. $(yz)x = x(yz)$ ตามสมบัติการสลับที่ของ

การคูณ

ค. $x(y+z) = xy + xz$ ตามสมบัติการแจกแจง

ง. $x + y = y + x$ ตามสมบัติการสลับที่ของ

การบวก

จ. $(y + z)x = xy + xz$ ตามสมบัติการ

เปลี่ยนกลุ่มของการบวก

17. $\frac{a}{b}$ จะเป็นเศษส่วนจำนวนคละเมื่อใด

ก. $a = 10, b = 7$

ข. $a = 8, b = 0$

ค. $a = 5, b = 7$

ง. $a = 7, b = 4$

จ. $a = 0, b = 5$

18. เศษส่วนในข้อใดมีค่าเท่ากับทุกจำนวน

ก. $\frac{1}{2}, \frac{2}{4}, \frac{3}{6}$

ข. $\frac{1}{3}, \frac{2}{6}, \frac{3}{9}$

ค. $\frac{1}{4}, \frac{2}{8}, \frac{4}{12}$

ง. $\frac{2}{3}, \frac{4}{6}, \frac{6}{9}$

จ. $\frac{16}{24}, \frac{24}{36}, \frac{32}{48}$

19. ในการคูณเศษส่วนเกี่ยวข้องกับวิธีการในข้อใด

ก. การหา ค.ร.น.

ข. การทอนเศษส่วน

ค. กลับตัวเศษและตัวส่วนของเศษส่วนที่เป็นตัวหาร

ง. การคูณเศษกับเศษ และส่วนกับส่วน

จ. การทำเศษส่วนจำนวนคละให้เป็นเศษเกิน

20. ให้ $b = \frac{1}{2} - \frac{1}{4}$ ค่าของ $12b$ ตรงกับข้อใด

ก. $\frac{1}{2} \times 6$

ข. $\frac{15}{3} \times \frac{3}{5}$

ค. $\frac{3}{5} \times \frac{5}{3}$

ง. $1 \div \frac{1}{3}$

จ. $\frac{6}{5} \div \frac{2}{5}$

21. ข้อใดสรุปผลการคูณหรือหารเศษส่วนได้ถูกต้อง

ก. $\frac{2}{8} \times \frac{16}{5} = \frac{8}{40}$ ไม่จริง เพราะว่าทอน

ตัวเศษกับตัวเศษ

ข. $\frac{2}{5} \times \frac{14}{3} = \frac{16}{8}$ ไม่จริง เพราะว่า

นำตัวเศษรวมกับตัวเศษและตัวส่วน

รวมกับตัวส่วน

ค. $\frac{2}{3} \div \frac{4}{15} = \frac{5}{2}$ ไม่จริง เพราะว่า

$$\frac{2}{3} \div \frac{4}{15} = \frac{2}{3} \times \frac{15}{4} = \frac{8}{45}$$

ง. $\frac{2}{7} \div \frac{5}{9} = \frac{18}{35}$ ไม่จริง เพราะว่า

$$\frac{2}{7} \div \frac{5}{9} = \frac{2}{7} \times \frac{9}{5} = \frac{35}{18}$$

จ. $\frac{3}{7} \div \frac{21}{5} = \frac{21}{5} \div \frac{3}{7}$ ไม่จริง เพราะว่า

จำนวนเศษส่วนไม่มีสมบัติการสลับที่

สำหรับการหาร

22. นายสุชาติแบ่งเงิน x บาท ให้บุตร 4 คน โดย บุตรคนโตได้รับส่วนแบ่งเป็นครึ่งหนึ่งของเงิน ทั้งหมด ที่เหลือแบ่งให้น้องอีก 3 คน ๆ ละ เท่ากันจงพิจารณาว่าข้อใดสรุปได้ถูกต้อง

ก. จำนวนเงินที่บุตรคนโตได้รับเป็น $\frac{x}{2}$ บาท

ข. จำนวนเงินที่น้องทั้งสามคนได้รับรวมเป็น $\frac{x}{2}$ บาท

ค. น้อง ๆ ได้รับเงินคนละ $\frac{x}{2} \div 3$ บาท

ง. จำนวนเงินที่เหลือจากให้บุตรคนโตเป็น $x - \frac{1}{2}$ บาท

จ. จำนวนเงินที่บุตรทั้ง 4 คน ได้รับรวมเป็น $\frac{x}{2} + \frac{x}{2}$ บาท

23. ค่าของเศษส่วนในข้อใด เท่ากับ $\frac{a}{b}$

ก. $\frac{a}{b} \times \frac{5}{5}$

ข. $\frac{a}{b} \div \frac{5}{5}$

ค. $\frac{a}{b} \div \frac{a}{b}$

ง. $\frac{a}{b} + \frac{5}{5}$

จ. $\frac{a}{b} - \frac{5}{5}$

24. จำนวนทศนิยมในข้อใดเรียงลำดับจากน้อยไปหามากได้ถูกต้อง

ก. 2.578 2.578 2.578

ข. 2.578 2.578 2.578

ค. 2.578 2.578 2.578

ง. 2.5708 2.578 2.578

จ. 2.579 2.578 2.578

25. ข้อใดเรียงลำดับจำนวนจากน้อยไปมากได้ถูกต้อง

ก. $\frac{1}{2} < 0.56 < \frac{1}{4}$

ข. $\frac{7}{8} < 1.2 < 1.3$

ค. $\frac{1}{10} < 0.2 < \frac{3}{5}$

ง. $0.34 < \frac{1}{4} < \frac{2}{3}$

จ. $0.78 < \frac{9}{10} < 1.1$

26. A และ B เป็นจำนวนที่อยู่ในรูปทศนิยม ถ้าผลคูณของ A กับ B มีทศนิยม 6 ตำแหน่ง แล้วจงพิจารณาว่าข้อสรุปใดถูกต้อง

ก. A มีทศนิยม 1 ตำแหน่ง และ B มีทศนิยม 6 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

ข. A มีทศนิยม 2 ตำแหน่ง และ B มีทศนิยม 4 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

ค. A มีทศนิยม 2 ตำแหน่ง และ B มีทศนิยม 3 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

ง. A มีทศนิยม 3 ตำแหน่ง และ B มีทศนิยม 3 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

จ. A มีทศนิยม 5 ตำแหน่ง และ B มีทศนิยม 1 ตำแหน่ง โดยที่ผลคูณตัวสุดท้ายไม่เป็น 0

27. ข้อใดสรุปผลลัพธ์ที่ได้จากการคูณหรือหารจำนวนทศนิยมได้ถูกต้อง

ก. 25.3×0.4 มีค่ามากกว่า 25.3

ข. 0.8×0.9 มีค่ามากกว่า 1

ค. $4.6 \div 0.6$ มีค่ามากกว่า 4.6

ง. $3.7 \div 4.7$ มีค่าน้อยกว่า 1

จ. $5.2 \div 2.2$ มีค่ามากกว่า 1

28. ข้อใดสรุปผลการคูณหรือหารทศนิยม
ได้ถูกต้อง

- ก. $2.35 \times 4.2 = 98.7$ ไม่จริง เพราะว่า
ผลคูณของ 2.35 ซึ่งมีค่าน้อยกว่า 3 กับ
4.2 ซึ่งมีค่าน้อยกว่า 5 ต้องมีค่าน้อยกว่า 15
- ข. $10.2 \times 0.12 = 1.224$ เป็นจริง เพราะว่า
จำนวนที่มีทศนิยม 1 ตำแหน่งคูณด้วย
จำนวนที่มีทศนิยม 2 ตำแหน่ง จะได้ผล
คูณเป็นจำนวนที่มีทศนิยม 3 ตำแหน่ง
- ค. $10.5 \div 0.4 = 26.25$ เป็นจริง เพราะว่า
ตัวหารมีค่าน้อยกว่า 1 เมื่อหาร 10.5 จะทำ
ให้ผลลัพธ์ที่ได้มีค่ามากกว่า 10.5
- ง. $20.4 \div 1.2 = 204 \div 12$ เป็นจริง เพราะว่า
การหารด้วยจำนวนทศนิยมจะเลื่อน
ตำแหน่งของตัวหารจนเป็นจำนวนเต็มก่อน
หาผลหาร
- จ. $14.5 \div 2.4 = 2.4 \div 14.5$ เป็นจริง
 เพราะว่าจำนวนทศนิยมมีสมบัติการ
สลับที่สำหรับการหาร

29. เชือกเส้นหนึ่งยาว y เมตร ดัดแบ่งออกเป็น
4 เส้น สองเส้นแรกยาวเส้นละ 1.25 เมตร
เส้นที่สามยาว 1.50 เมตร ส่วนที่เหลือยาว
เท่าใด

- ก. $[(2 \times 1.25) + 1.50] - y$
- ข. $y - [(2 \times 1.25) + 1.50]$
- ค. $y - (1.25 + 1.50)$
- ง. $y - 2.50 - 1.50$
- จ. $y - 4.00$

30. ให้ a เป็นจำนวนทศนิยมที่มีค่ามากกว่า 1
ข้อใดสรุปผลคูณหรือผลหารของ a ได้ถูกต้อง

- ก. $\frac{1}{a} > 1$
- ข. $\frac{1}{a} < a$
- ค. $a \times a > 1$
- ง. $a \times a < a$
- จ. $a \times a > \frac{1}{a}$

ขอขอบคุณนักเรียนที่ตั้งใจทำ

ฉบับ y

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์

เรื่อง จำนวนเต็ม เศษส่วน และทศนิยม

คำชี้แจงในการตอบแบบทดสอบ

1. แบบทดสอบฉบับนี้เป็นแบบทดสอบเลือกตอบที่มีคำตอบถูกหลายคำตอบ โดยในหนึ่งข้อจะมีตัวเลือกที่เป็นคำตอบถูกต้องตั้งแต่สองตัวเลือกจนถึงสี่ตัวเลือก ในการตอบนักเรียนจะต้องพิจารณาตอบทุกตัวเลือก โดยทำเครื่องหมายถูก (✓) หน้าตัวเลือกที่เห็นว่าถูก และทำเครื่องหมายผิด (X) หน้าตัวเลือกที่เห็นว่าผิด ลงในกระดาษคำตอบ
2. ข้อสอบมีทั้งหมด 30 ข้อ จำนวน 6 หน้า เวลาในการทำข้อสอบ 75 นาที
3. ให้นักเรียนกรอกรายละเอียดลงบนกระดาษคำตอบ แต่กรุณาอย่าทำเครื่องหมายใด ๆ ลงบนแบบทดสอบ ถ้าต้องการทดให้ทดในกระดาษที่สำรองไว้
4. ถ้านักเรียนมีข้อสงสัยประการใดให้ถามผู้ดำเนินการสอบก่อนลงมือทำ
5. ให้นักเรียนตรวจทานก่อนส่งกระดาษคำตอบว่าได้ทำเครื่องหมายลงในกระดาษคำตอบครบทุกตัวเลือก

ตัวอย่างการตอบแบบทดสอบ

คำถาม		กระดาษคำตอบ					
ข้อ 0	ข้อสรุปเกี่ยวกับจำนวนในข้อใดเป็นจริง	ข้อ	ก	ข	ค	ง	จ
ก.	1 เป็นจำนวนนับที่มีค่าน้อยที่สุด	0	✓	X	✓	X	X
ข.	0 เป็นจำนวนนับ	00					
ค.	1 เป็นจำนวนตรงข้ามของ -1						
ง.	จำนวนนับก็คือจำนวนเต็ม						
จ.	จำนวนเต็มแบ่งเป็นจำนวนเต็มบวก และจำนวนเต็มลบ						

1. ข้อสรุปเกี่ยวกับจำนวนเต็มในข้อใดเป็นจริง

- ก. 0 เป็นจำนวนเต็มที่น้อยที่สุด
- ข. จำนวนตรงข้ามของศูนย์เท่ากับศูนย์
- ค. 1 เป็นจำนวนเต็มบวกที่มีค่าน้อยที่สุด
- ง. -1 เป็นจำนวนเต็มลบที่มีค่าน้อยที่สุด
- จ. -1 อยู่ห่างจากศูนย์เท่ากับ 1

2. ข้อใดเป็นจำนวนเต็มทุกจำนวน

- ก. 1, 0, -1, -2
- ข. 0, 2, 3, 4
- ค. -1, -2, -3, -4
- ง. 0.0, -1.5, -2.0, 3.0
- จ. $\frac{-4}{2}, \frac{0}{2}, \frac{4}{2}, \frac{6}{2}$

3. ให้ a และ b เป็นจำนวนเต็ม ค่าของ a, b ในข้อใดทำให้ผลลัพธ์ของ $a - b$ เป็นจำนวนเต็มลบ

- ก. a เป็นจำนวนเต็มลบ และ b เป็นศูนย์
- ข. a และ b เป็นจำนวนเต็มลบ โดยที่ a อยู่ห่างจากศูนย์มากกว่า b
- ค. a และ b เป็นจำนวนเต็มบวก โดยที่ a อยู่ห่างจากศูนย์มากกว่า b
- ง. a เป็นจำนวนเต็มลบ และ b เป็นจำนวนเต็มบวก
- จ. a เป็นศูนย์ และ b เป็นจำนวนเต็มบวก

4. กำหนดให้ x และ y เป็นจำนวนเต็ม จงพิจารณาว่าข้อใดเป็นจริง

- ก. $x + y = y + x$
- ข. $x + y = -(-x - y)$
- ค. $x + y = -x - y$
- ง. $x + y = -(x - y)$
- จ. $x + y = x + (-y)$

5. $[(-a) + b] - [(-c) + d]$ มีค่าเท่าใด

เมื่อ $a = 5, b = 4, c = 3$ และ $d = -2$

- ก. $a + (-d)$
- ข. $a - b + c$
- ค. $b + c + d$
- ง. $1 \times b$
- จ. $2c + d$

6. ข้อใดสรุปผลการบวกหรือลบจำนวนเต็มได้ถูกต้อง

- ก. $(-7) - (-8) = 1$ เป็นจริง เพราะว่า $(-7) - (-8) = -7 + (8) = 1$
- ข. $8 - (-7) = 1$ เป็นจริง เพราะว่า $8 - (-7) = 8 - 7 = 1$
- ค. $7 - 8 = -1$ เป็นจริง เพราะว่า $7 - 8 = 7 + (-8) = -1$
- ง. $-8 + 7 = 7 + (-8)$ เป็นจริง เพราะว่าจำนวนเต็มมีสมบัติการสลับที่สำหรับการบวก
- จ. $-7 - 8 = 1$ ไม่จริง เพราะว่า $-7 - 8 = (-7) + (-8) = -15$

7. ข้อใดกล่าวได้ถูกต้องเกี่ยวกับผลลัพธ์ของการคูณหรือหารจำนวนเต็ม

- ก. ผลคูณของจำนวนเต็มลบกับจำนวนเต็มบวกเป็นจำนวนเต็มบวก
- ข. ผลคูณของจำนวนเต็มลบกับจำนวนเต็มลบเป็นจำนวนเต็มบวก
- ค. ผลคูณของจำนวนเต็มบวกกับจำนวนเต็มลบเป็นจำนวนเต็มลบ
- ง. ผลหารของจำนวนเต็มลบกับจำนวนเต็มลบที่หารลงตัวเป็นจำนวนเต็มลบ
- จ. ผลหารของจำนวนเต็มลบกับจำนวนเต็มบวกที่หารลงตัวจะเป็นจำนวนเต็มลบ

8. ให้ $a = -8$, $b = 4$, $c = -2$ ดังนั้น $(a \div b) \times c$ มีค่าเท่ากับข้อใด

ก. $(-2) \times c$
 ข. $b \times c$
 ค. $a \div c$
 ง. $(2 \times b) \div c$
 จ. $(a \times c) \div b$

9. ผลลัพธ์ในข้อใดมีค่าเป็นจำนวนเต็มบวก

ก. $[0 \times (-5)] \div 2$
 ข. $[(-2) \times (-4)] \div 4$
 ค. $[10 \times (-5)] \times (-7)$
 ง. $[20 \div 10] \times (-8)$
 จ. $[20 \div 5] \div 2$

10. กำหนด $\frac{a}{-5} = b$ และ $\frac{b}{c} = 3$ โดยที่ $c = 2$ ข้อใดสรุปผลลัพธ์ของ a และ b ได้ถูกต้อง

ก. $a + b = -24$
 ข. $a - b = -36$
 ค. $b - a = 24$
 ง. $a \times b = 180$
 จ. $a \div b = -5$

11. จำนวนในข้อใดเมื่อนำมาคูณกับจำนวนเต็มบวก แล้วผลลัพธ์มีค่าไม่น้อยกว่าเดิม

ก. จำนวนเต็มลบ
 ข. จำนวนเต็มบวก
 ค. จำนวนเต็มยกกำลังสอง
 ง. จำนวนเต็มบวกยกกำลังสอง
 จ. จำนวนเต็มลบยกกำลังสอง

12. กำหนด a เป็นจำนวนเต็มใด ๆ ประโยคในข้อใดเป็นจริง

ก. $a + 1 = 1 + a$
 ข. $a - 1 = 1 - a$
 ค. $a \times 1 = 1 \times a$
 ง. $\frac{1}{a} = \frac{a}{1}$
 จ. $\frac{a}{1} = a \times 1$

13. กำหนด x และ y เป็นจำนวนเต็มใด ๆ ข้อกำหนดในข้อใดทำให้ $x + y = 1$

ก. x เป็นจำนวนตรงข้ามของ -1 และ y มีค่าเป็นศูนย์
 ข. x มีค่าเป็นหนึ่งและ y มีค่าเป็นศูนย์
 ค. x มีค่าเป็นศูนย์ และ y มีค่าเป็นหนึ่ง
 ง. x และ y มีค่าเท่ากับ -1
 จ. x และ y มีค่าเท่ากับ 1

14. กำหนด a และ b เป็นจำนวนเต็มใด ๆ ค่า a, b ในข้อใดที่ทำให้ $4a + 4b = 16$ เป็นจริง

ก. $a = 8, b = -4$
 ข. $a = 5, b = 11$
 ค. $a = 2, b = 2$
 ง. $a = 0, b = 4$
 จ. $a = 4, b = 4$

15. กำหนด a, b และ c เป็นจำนวนเต็มใด ๆ
ที่ไม่เท่ากับศูนย์ โดยที่ $a = b = c$

ข้อใดสรุปได้ถูกต้อง

ก. $a + b = b + c$

ข. $a - c = b - c$

ค. $\frac{a}{b} = 1$

ง. $\frac{ab}{c} = 2$

จ. $ab = c^2$

16. กำหนด x, y, z เป็นจำนวนเต็มใด ๆ
ข้อใดกล่าวถึงสมบัติของจำนวนเต็มได้ถูกต้อง

ก. $x(yz) = x(z y)$ ตามสมบัติการสลับที่ของการคูณ

ข. $(yz)x = x(yz)$ ตามสมบัติการสลับที่ของการคูณ

ค. $x(y+z) = xy + xz$ ตามสมบัติการแจกแจง

ง. $x + y = y + x$ ตามสมบัติการสลับที่ของการบวก

จ. $(y + z)x = xy + xz$ ตามสมบัติการเปลี่ยนกลุ่มของการบวก

17. $\frac{a}{b}$ จะเป็นเศษส่วนจำนวนคละเมื่อใด

ก. $a = 10, b = 7$

ข. $a = 8, b = 0$

ค. $a = 5, b = 7$

ง. $a = 7, b = 4$

จ. $a = 0, b = 5$

18. ข้อใดสรุปผลการเปรียบเทียบเศษส่วนได้ถูกต้อง

ก. $\frac{19}{30} > \frac{13}{32}$ เป็นจริง เพราะว่า $\frac{19}{30} > \frac{1}{2}$ และ

$\frac{13}{32} < \frac{1}{2}$ ดังนั้น $\frac{19}{30} > \frac{13}{32}$

ข. $\frac{5}{12} < \frac{9}{10}$ เป็นจริง เพราะว่า $\frac{5}{12} < \frac{5}{10}$ และ

$\frac{5}{10} < \frac{9}{10}$ ดังนั้น $\frac{5}{12} < \frac{9}{10}$

ค. $\frac{1}{2} = \frac{4}{8}$ เป็นจริง เพราะว่า เมื่อนำ 4 ไปหารทั้ง

ตัวเศษ และตัวส่วนของ $\frac{4}{8}$ จะได้ $\frac{1}{2}$

ง. $\frac{3}{10} = \frac{4}{11}$ ไม่จริง เพราะว่า เมื่อทำส่วนของ

เศษส่วนทั้งสองให้เท่ากันจะได้ตัวเศษของทั้งสองจำนวนเป็น 3×11 กับ 4×10 ซึ่งมีค่าไม่เท่ากัน

จ. $\frac{7}{10} > \frac{7}{12}$ ไม่จริง เพราะว่า ตัวเศษของ

ทั้งสองจำนวนเท่ากันแต่ตัวส่วน $10 < 12$

ดังนั้น ผลหารของ 10 ย่อมน้อยกว่าผลหารของ 12

19. ผลลบของ $7\frac{2}{3}$ กับ $2\frac{1}{3}$ มีค่าตรงกับข้อใด

ก. $(7 - 2) + \left(-\frac{2}{3} - \frac{1}{3}\right)$

ข. $(7 - 2) + \left(\frac{2}{3} - \frac{1}{3}\right)$

ค. $(7 - 2) - \left(\frac{2}{3} + \frac{1}{3}\right)$

ง. $\frac{23}{3} + \left(-\frac{7}{3}\right)$

จ. $\frac{23}{3} - \frac{7}{3}$

20. ข้อใดสรุปผลการบวกหรือลบเศษส่วนได้ถูกต้อง

ก. $\frac{3}{7} + \frac{2}{7} = \frac{5}{7}$ เป็นจริง เพราะว่า
การบวกเศษส่วนที่ตัวส่วนเท่ากัน
สามารถรวมตัวเศษของสองจำนวนได้
โดยตรง

ข. $\frac{1}{5} + \frac{2}{5} = \frac{1+2}{5+5}$ เป็นจริง เพราะว่า
การบวกเศษส่วนต้องบวกทั้งตัวเศษ
และตัวส่วน

ค. $\frac{3}{4} + \frac{1}{4} = \frac{3}{4 \times 4} + \frac{1}{4 \times 4}$ ไม่จริง เพราะว่า
 $\frac{3}{4}$ ไม่เท่ากับ $\frac{3}{4 \times 4}$ และ $\frac{1}{4}$ ไม่เท่ากับ
 $\frac{1}{4 \times 4}$

ง. $\frac{3}{5} - \frac{2}{7} = \frac{3 \times 5}{5 \times 7} - \frac{2 \times 7}{5 \times 7}$ ไม่จริง
เพราะว่า $\frac{3}{5}$ ไม่เท่ากับ $\frac{3 \times 5}{5 \times 7}$ และ $\frac{2}{7}$
ไม่เท่ากับ $\frac{2 \times 7}{5 \times 7}$

จ. $\frac{3}{4} - \frac{2}{5} = \frac{2}{5} - \frac{3}{4}$ ไม่จริง เพราะว่า จำนวน
เศษส่วนไม่มีสมบัติการสลับที่สำหรับการลบ

21. ข้อใดมีผลลัพธ์เท่ากับ $\frac{4}{7}$

ก. $\frac{2}{7} \times 2$

ข. $\frac{2}{7} \div \frac{1}{2}$

ค. $\frac{1}{7} \div \frac{4}{28}$

ง. $\frac{4}{7} \times \frac{7}{8} \times \frac{8}{7}$

จ. $\left(\frac{14}{3} \div \frac{7}{3}\right) \times \frac{2}{7}$

22. นายสุชาติแบ่งเงิน x บาท ให้บุตร 4 คน โดย
บุตรคนโตได้รับส่วนแบ่งเป็นครึ่งหนึ่งของเงิน
ทั้งหมด ที่เหลือแบ่งให้น้องอีก 3 คน ๆ ละ
เท่า ๆ กัน จงพิจารณาว่าข้อใดสรุปได้ถูกต้อง

ก. จำนวนเงินที่บุตรคนโตได้รับเป็น $\frac{x}{2}$ บาท

ข. จำนวนเงินที่น้องทั้งสามคนได้รับรวมเป็น
 $\frac{x}{2}$ บาท

ค. น้อง ๆ ได้รับเงินคนละ $\frac{x}{2} \div 3$ บาท

ง. จำนวนเงินที่เหลือจากให้บุตรคนโตเป็น
 $x - \frac{1}{2}$ บาท

จ. จำนวนเงินที่บุตรทั้ง 4 คน ได้รับรวมเป็น
 $\frac{x}{2} + \frac{x}{2}$ บาท

23. ให้ p , q , และ r เป็นจำนวนเต็มบวก โดยที่
 $p < q$ และ $q < r$ ข้อใดสรุปค่าเศษส่วนได้
ถูกต้อง

ก. $\frac{r}{p}$ มากกว่า 1

ข. $\frac{p}{q}$ มากกว่า 1

ค. $\frac{q}{r}$ น้อยกว่า 1

ง. $\frac{r+q}{p}$ น้อยกว่า 1

จ. $\frac{r}{q+r}$ น้อยกว่า 1

24. ให้ $a = 20.8759$, $b = 20.5879$, $c = 20.7589$ และ $d = 20.9578$ ข้อใดเรียงลำดับจำนวนทศนิยมจากมากไปน้อยได้ถูกต้อง

ก. $a > c > b$
 ข. $a > b > c$
 ค. $d > a > c$
 ง. $d > a > b$
 จ. $d > c > b$

25. ข้อใดเรียงลำดับจำนวนจากน้อยไปมากได้ถูกต้อง

ก. $\frac{1}{2} < 0.56 < \frac{1}{4}$
 ข. $\frac{7}{8} < 1.2 < 1.3$
 ค. $\frac{1}{10} < 0.2 < \frac{3}{5}$
 ง. $0.34 < \frac{1}{4} < \frac{2}{3}$
 จ. $0.78 < \frac{9}{10} < 1.1$

26. 0.01 หารด้วย 0.1 มีค่าเท่ากับข้อใด

ก. $\frac{1}{100}$
 ข. $\frac{1}{10}$
 ค. $\frac{1}{100} \times \frac{10}{1}$
 ง. $\frac{0.01}{1}$
 จ. $\frac{0.01}{10}$
 ข. $\frac{1}{100} \times \frac{10}{1}$

27. ผลลัพธ์ที่ได้จากการบวกหรือลบจำนวนทศนิยมในข้อใด มีทศนิยม 4 ตำแหน่ง

ก. $130.2 + 1.4568$
 ข. $10.78 + 1.304$
 ค. $15.347 + 10.1249$
 ง. $1.7856 - 0.49$
 จ. $30.18 - 15.2784$

28. ข้อใดสรุปผลการบวกหรือลบทศนิยมได้ถูกต้อง

ก. $10.2 + 1.95 = 1.215$ ไม่จริง เพราะว่าผลบวกของ 10.2 ซึ่งมีค่ามากกว่า 10 กับ 1.95 ซึ่งมีค่ามากกว่า 1 ต้องมีค่ามากกว่าผลบวกของ 10 กับ 1 ซึ่งเท่ากับ 11
 ข. $11.2 + 5.8 = \frac{170}{10}$ ไม่จริง เพราะว่า 11.2 เท่ากับ $\frac{112}{10}$ และ 5.8 เท่ากับ $\frac{58}{10}$ เมื่อนำมารวมกันมีค่าเท่ากับ $\frac{112}{10 \times 10} + \frac{58}{10 \times 10} = \frac{170}{100}$
 ค. $14.2 + 1.55 = 29.7$ ไม่จริง เพราะว่าผลบวกของ 14.2 ซึ่งน้อยกว่า 15 กับ 1.55 ซึ่งน้อยกว่า 2 ต้องมีค่าน้อยกว่าผลบวกของ 15 กับ 2 ซึ่งเท่ากับ 17
 ง. $5.35 - 2.1 = 5.14$ ไม่จริง เพราะว่า ผลลบของ 5.35 ซึ่งมีค่าน้อยกว่า 6 กับ 2.1 ซึ่งมีค่ามากกว่า 2 ต้องมีค่าน้อยกว่าผลลบของ 6 กับ 2 ซึ่งเท่ากับ 4
 จ. $5.35 - 2.5 = 2.5 - 5.35$ ไม่จริง เพราะว่าจำนวนทศนิยมไม่มีสมบัติการสลับที่สำหรับการลบ

29. เชือกเส้นหนึ่งยาว y เมตร ตัดแบ่งออกเป็น 4 เส้น สองเส้นแรกยาวเส้นละ 1.25 เมตร เส้นที่สามยาว 1.50 เมตร ส่วนที่เหลือยาวเท่าใด

ก. $[(2 \times 1.25) + 1.50] - y$

ข. $y - [(2 \times 1.25) + 1.50]$

ค. $y - (1.25 + 1.50)$

ง. $y - 2.50 - 1.50$

จ. $y - 4.00$

30. ให้ a เป็นจำนวนทศนิยมที่มีค่ามากกว่า 1 ข้อใดสรุปผลคูณหรือผลหารของ a ได้ถูกต้อง

ก. $\frac{1}{a} > 1$

ข. $\frac{1}{a} < a$

ค. $a \times a > 1$

ง. $a \times a < a$

จ. $a \times a > \frac{1}{a}$

ขอขอบคุณนักเรียนที่ตั้งใจทำ

ฉบับ x

กระดาษคำตอบ

ชื่อ-สกุล

เลขที่

โรงเรียน

ห้อง

ข้อ	ก	ข	ค	ง	จ	ข้อ	ก	ข	ค	ง	จ
1						16					
2						17					
3						18					
4						19					
5						20					
6						21					
7						22					
8						23					
9						24					
10						25					
11						26					
12						27					
13						28					
14						29					
15						30					

ฉบับ y

กระดาษคำตอบ

ชื่อ-สกุล
โรงเรียน

เลขที่
ห้อง

ข้อ	ก	ข	ค	ง	จ	ข้อ	ก	ข	ค	ง	จ
1						16					
2						17					
3						18					
4						19					
5						20					
6						21					
7						22					
8						23					
9						24					
10						25					
11						26					
12						27					
13						28					
14						29					
15						30					

ภาคผนวก จ

- ตัวอย่างผลการวิเคราะห์พารามิเตอร์ความยากและอำนาจจำแนกด้วยโปรแกรม MULTILOG Version 7.03 for window
- ตัวอย่างผลการวิเคราะห์พารามิเตอร์ความสามารถด้วยโปรแกรม MULTILOG Version 7.03 for window
- ตัวอย่างผลการวิเคราะห์องค์ประกอบเชิงยืนยันด้วยโปรแกรม LISREL Version 8.30
- ผลการวิเคราะห์สัมประสิทธิ์การเทียบคะแนนด้วยโปรแกรม Equate 2.1
- ตัวอย่างคำสั่งบนโปรแกรม SPSS Version 7.5 ที่ใช้คำนวณค่าความน่าจะเป็นของผู้สอบที่มีความสามารถระดับ θ จะตอบข้อสอบข้อที่ j ได้คะแนนในลำดับชั้น k หรือสูงกว่า $(P_{jk}^+(\theta))$
- ตัวอย่างคำสั่งบนโปรแกรม SPSS Version 7.5 ที่ใช้คำนวณค่าความน่าจะเป็นของผู้สอบที่มีความสามารถระดับ θ จะตอบข้อสอบข้อที่ j ได้คะแนนลำดับชั้น k $(P_{jk}(\theta))$
- ตัวอย่างคำสั่งบนโปรแกรม SPSS Version 7.5 ที่ใช้คำนวณคะแนนจริง
- ตัวอย่างคำสั่งบนโปรแกรม SPSS Version 7.5 ที่ใช้คำนวณค่าความน่าจะเป็นในแต่ละระดับคะแนนแต่ละระดับความสามารถตามสูตร 85-86

ตัวอย่างผลการวิเคราะห์พารามิเตอร์ความยากและอำนาจจำแนกด้วยโปรแกรม MULTILOG
Version 7.03 for Windows

MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION 7.0.3
MULTILOG for Windows 7.00.2327.2
Created on: 10 March 2003, 19:25:18

```
>PROBLEM RANDOM,
    INDIVIDUAL,
    DATA = 'D:\mlwin\strey.DAT',
    NITEMS = 30,
    NGROUPS = 1,
    NEXAMINEES = 1157,
    NCHARS = 4;

DATA FILE NAME IS
    D:\MLWIN\STREAY.DAT

TYPE OF INPUT:
    INDIVIDUAL RESPONSE VECTORS

>TEST ALL,
    GRADED,
    NC = (6(0)30);

>SAV;

NUMBER OF CODES 6
012345
    VECTOR OF CATEGORIES FOR CODE=0
1111111111111111111111111111111111
    VECTOR OF CATEGORIES FOR CODE=1
2222222222222222222222222222222222
    VECTOR OF CATEGORIES FOR CODE=2
3333333333333333333333333333333333
    VECTOR OF CATEGORIES FOR CODE=3
4444444444444444444444444444444444
    VECTOR OF CATEGORIES FOR CODE=4
5555555555555555555555555555555555
    VECTOR OF CATEGORIES FOR CODE=5
6666666666666666666666666666666666
(4a1,3x,30a1)
```

MULTILOG--FOR MULTIPLE CATEGORICAL ITEM RESPONSE DATA--VERSION 7.0.3
 MULTILOG for Windows 7.00.2327.2
 Created on: 10 March 2003, 19:25:18

DATA PARAMETERS:

NUMBER OF LINES IN THE DATA FILE: 1157
 NUMBER OF CATEGORICAL-RESPONSE ITEMS: 30
 NUMBER OF CONTINUOUS-RESPONSE ITEMS, AND/OR GROUPS: 1
 TOTAL NUMBER OF "ITEMS" (INCLUDING GROUPS): 31

NUMBER OF CHARACTERS IN ID FIELDS: 4
 MAXIMUM NUMBER OF RESPONSE-CODES FOR ANY ITEM: 6
 THE MISSING VALUE CODE FOR CONTINUOUS DATA: 9.0000
 THE DATA WILL BE STORED IN MEMORY

ESTIMATION PARAMETERS:

THE ITEMS WILL BE CALIBRATED--
 BY MARGINAL MAXIMUM LIKELIHOOD ESTIMATION
 MAXIMUM NUMBER OF EM CYCLES PERMITTED: 25
 NUMBER OF PARAMETER-SEGMENTS USED IS: 30
 NUMBER OF FREE PARAMETERS IS: 180
 MAXIMUM NUMBER OF M-STEP ITERATIONS IS 4 TIMES
 THE NUMBER OF PARAMETERS IN THE SEGMENT
 NUMBER OF QUADRATURE POINTS IS: 19
 THE M-STEP CONVERGENCE CRITERION IS: 0.000100
 THE EM-CYCLE CONVERGENCE CRITERION IS: 0.001000
 THE RK CONTROL PARAMETER (FOR THE M-STEPS) IS: 0.9000
 THE RM CONTROL PARAMETER (FOR THE M-STEPS) IS: 1.0000
 THE MAXIMUM ACCELERATION PERMITTED IS: 0.0000
 THETA-GROUP LOCATIONS WILL REMAIN UNCHANGED

OUTPUT-CONTROL PARAMETERS

THE ITEM PARAMETERS WILL BE SAVED ON FILE <jobname>.PAR

QUADRATURE POINTS FOR MML, AT THETA:

-4.500
 -4.000
 -3.500
 -3.000
 -2.500
 -2.000
 -1.500
 -1.000
 -0.500
 0.000
 0.500
 1.000
 1.500
 2.000
 2.500
 3.000
 3.500
 4.000
 4.500

MULTILOG for Windows 7.00.2327.2

READING DATA...

KEY-

CODE CATEGORY

```

0  11111111111111111111111111111111
1  22222222222222222222222222222222
2  33333333333333333333333333333333
3  44444444444444444444444444444444
4  55555555555555555555555555555555
5  66666666666666666666666666666666

```

FORMAT FOR DATA-

(4a1,3x,30a1)

FIRST OBSERVATION AS READ-

```

ID  0001
ITEMS 544555555534553555554145535144
NORML  0.000

```

FINISHED CYCLE 25

MAXIMUM INTERCYCLE PARAMETER CHANGE= 0.02892 P(8)

ITEM SUMMARY

MULTILOG for Windows 7.00.2327.2

ITEM 1: 6 GRADED CATEGORIES
P(#) ESTIMATE (S.E.)

```

A    1  1.52 (0.11)
B( 1)  2 -2.91 (0.30)
B( 2)  3 -2.02 (0.18)
B( 3)  4 -1.08 (0.10)
B( 4)  5 -0.22 (0.07)
B( 5)  6  0.78 (0.07)

```

@THETA: INFORMATION: (Theta values increase in steps of 0.2)

```

-3.0 - -1.6 0.634 0.670 0.694 0.708 0.716 0.720 0.721 0.722
-1.4 -  0.0 0.723 0.724 0.725 0.725 0.724 0.723 0.719 0.714
 0.2 -  1.6 0.706 0.693 0.673 0.643 0.599 0.542 0.476 0.405
 1.8 -  3.0 0.335 0.271 0.214 0.167 0.128 0.098 0.074

```

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
CATEGORY(K): 1 2 3 4 5 6

```

OBS. FREQ.  16  41 129 223 308 440
OBS. PROP.  0.0138 0.0354 0.1115 0.1927 0.2662 0.3803
EXP. PROP.  0.0312 0.0615 0.1443 0.2038 0.2570 0.3021

```

ITEM 2: 6 GRADED CATEGORIES
P(#) ESTIMATE (S.E.)

```

A    7  1.47 (0.11)
B( 1)  8 -4.87 (1.01)
B( 2)  9 -2.48 (0.23)
B( 3) 10 -0.88 (0.09)
B( 4) 11  0.06 (0.07)
B( 5) 12  2.17 (0.13)

```

@THETA: INFORMATION: (Theta values increase in steps of 0.2)
 -3.0 - -1.6 0.535 0.565 0.588 0.602 0.606 0.605 0.604 0.607
 -1.4 - 0.0 0.615 0.628 0.641 0.652 0.658 0.658 0.651 0.637
 0.2 - 1.6 0.616 0.589 0.561 0.537 0.522 0.519 0.528 0.544
 1.8 - 3.0 0.560 0.569 0.563 0.539 0.498 0.445 0.384

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K): 1 2 3 4 5 6
 OBS. FREQ. 1 31 208 274 501 142
 OBS. PROP. 0.0009 0.0268 0.1798 0.2368 0.4330 0.1227
 EXP. PROP. 0.0022 0.0545 0.2245 0.2350 0.4022 0.0815

ITEM 3: 6 GRADED CATEGORIES
 P(ESTIMATE (S.E.))

A 13 1.66 (0.11)
 B(1) 14 -3.97 (0.64)
 B(2) 15 -1.68 (0.14)
 B(3) 16 -0.79 (0.08)
 B(4) 17 0.41 (0.06)
 B(5) 18 1.70 (0.10)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)
 -3.0 - -1.6 0.552 0.543 0.560 0.598 0.652 0.709 0.759 0.797
 -1.4 - 0.0 0.821 0.833 0.836 0.832 0.824 0.815 0.810 0.809
 0.2 - 1.6 0.810 0.808 0.803 0.794 0.787 0.782 0.775 0.759
 1.8 - 3.0 0.724 0.667 0.591 0.503 0.413 0.329 0.255

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K): 1 2 3 4 5 6
 OBS. FREQ. 2 74 160 385 334 202
 OBS. PROP. 0.0017 0.0640 0.1383 0.3328 0.2887 0.1746
 EXP. PROP. 0.0052 0.1188 0.1674 0.3210 0.2665 0.1211

ITEM 30: 6 GRADED CATEGORIES
 P(ESTIMATE (S.E.))

A 175 1.30 (0.10)
 B(1) 176 -2.93 (0.32)
 B(2) 177 -1.71 (0.17)
 B(3) 178 -0.58 (0.09)
 B(4) 179 0.32 (0.07)
 B(5) 180 1.33 (0.10)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)
 -3.0 - -1.6 0.462 0.482 0.497 0.506 0.512 0.517 0.520 0.522
 -1.4 - 0.0 0.523 0.525 0.526 0.528 0.530 0.531 0.532 0.532
 0.2 - 1.6 0.531 0.528 0.524 0.517 0.506 0.490 0.467 0.438
 1.8 - 3.0 0.401 0.360 0.316 0.271 0.229 0.190 0.155

OBSERVED AND EXPECTED COUNTS/PROPORTIONS IN
 CATEGORY(K): 1 2 3 4 5 6
 OBS. FREQ. 27 88 225 256 252 309
 OBS. PROP. 0.0233 0.0761 0.1945 0.2213 0.2178 0.2671
 EXP. PROP. 0.0436 0.1072 0.2100 0.2169 0.2118 0.2106

ITEM 31: GRP1, N[MU: 0.00 SIGMA: 1.00]
P(#):(S.E.): 182; (0.00) 183; (0.00)

@THETA: INFORMATION: (Theta values increase in steps of 0.2)
-3.0 - -1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
-1.4 - 0.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
0.2 - 1.6 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000
1.8 - 3.0 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000

TOTAL TEST INFORMATION

@THETA: INFORMATION:
-3.0 - -1.6 18.676 20.399 21.952 23.252 24.305 25.193 25.997 26.713
-1.4 - 0.0 27.282 27.677 27.929 28.074 28.127 28.083 27.919 27.585
0.2 - 1.6 27.012 26.158 25.012 23.569 21.838 19.880 17.804 15.726
1.8 - 3.0 13.734 11.882 10.196 8.686 7.353 6.195 5.204

@THETA: POSTERIOR STANDARD DEVIATION:
-3.0 - -1.6 0.231 0.221 0.213 0.207 0.203 0.199 0.196 0.193
-1.4 - 0.0 0.191 0.190 0.189 0.189 0.189 0.189 0.189 0.190
0.2 - 1.6 0.192 0.196 0.200 0.206 0.214 0.224 0.237 0.252
1.8 - 3.0 0.270 0.290 0.313 0.339 0.369 0.402 0.438

MARGINAL RELIABILITY: 0.9582

NEGATIVE TWICE THE LOGLIKELIHOOD= 75357.8
(CHI-SQUARE FOR SEVERAL TIMES MORE EXAMINEES THAN CELLS)

NORMAL PROGRAM TERMINATION

START DATE: 03-17-2003
START TIME: 18:03:09
END TIME: 18:03:33

```
>PROBLEM SCORE,  
    INDIVIDUAL,  
    DATA = 'D:\mlwin\strey.DAT',  
    NITEMS = 30,  
    NGROUPS = 1,  
    NEXAMINEES = 1157,  
    NCHARS = 4;  
>TEST ALL,  
    GRADED,  
    NC = (6(0)30);  
>START ALL, PARAM = 'D:\MLWIN\STREAY.PAR';  
>SAV;  
>END ;
```

6

012345

111111111111111111111111111111111111

222222222222222222222222222222222222

333333333333333333333333333333333333

444444444444444444444444444444444444

555555555555555555555555555555555555

666666666666666666666666666666666666

(4a1,3x,30a1)

1.306	0.233	0001
1.238	0.229	0002
0.675	0.203	0003
0.879	0.218	0004
0.027	0.186	0005
-0.454	0.203	0006
-0.601	0.178	0007
-0.131	0.196	0008
0.025	0.186	0009
0.708	0.203	0010
0.359	0.196	0011
0.622	0.193	0012
-0.034	0.191	0013
-0.599	0.178	0014
0.087	0.201	0015
1.093	0.213	0016
-0.161	0.177	0017
0.547	0.196	0018
-0.311	0.180	0019
0.003	0.186	0020
-0.293	0.190	0021
-0.025	0.196	0022
-0.183	0.202	0023
0.690	0.198	0024
0.709	0.203	0025
0.937	0.211	0026
0.910	0.202	0027
0.514	0.199	0028
0.958	0.202	0029
0.857	0.196	0030
1.085	0.224	0031
0.945	0.206	0032
0.079	0.198	0033
-0.426	0.176	0034
0.907	0.221	0035
1.151	0.232	0036
0.800	0.200	0037
0.420	0.196	0038
-0.291	0.208	0039
0.345	0.191	0040
-0.092	0.189	0041
0.589	0.198	0042
-0.724	0.206	0043
0.906	0.216	0044
1.236	0.226	0045
1.390	0.241	0046
0.354	0.198	0047
1.325	0.235	0048
1.250	0.240	0049
0.522	0.193	0050
0.136	0.190	0051
0.787	0.219	0052
-0.591	0.180	0053
-0.370	0.195	0054
-0.236	0.180	0055
0.328	0.206	0056
1.479	0.250	0057
2.358	0.327	0058
1.635	0.252	0059
1.341	0.232	0060
2.050	0.291	0061
1.278	0.235	0062
1.113	0.211	0063
1.523	0.240	0064
1.240	0.221	0065
1.894	0.277	0066
2.143	0.301	0067
1.259	0.225	0068
1.499	0.243	0069
0.856	0.213	0070
0.159	0.197	0071
1.643	0.266	0072

-0.453	0.195	1081
1.951	0.295	1082
1.707	0.251	1083
0.822	0.201	1084
1.305	0.234	1085
-0.271	0.191	1086
1.253	0.234	1087
1.511	0.247	1088
2.144	0.312	1089
2.282	0.318	1090
1.568	0.246	1091
1.893	0.283	1092
1.565	0.250	1093
2.574	0.365	1094
1.659	0.263	1095
2.811	0.403	1096
2.134	0.305	1097
2.553	0.369	1098
2.209	0.312	1099
1.601	0.248	1100
2.752	0.408	1101
2.136	0.301	1102
2.867	0.412	1103
2.312	0.325	1104
2.022	0.287	1105
3.095	0.458	1106
1.629	0.261	1107
3.095	0.458	1108
2.462	0.345	1109
2.491	0.359	1110
0.891	0.220	1111
0.533	0.205	1112
0.679	0.215	1113
0.377	0.201	1114
0.889	0.221	1115
0.706	0.215	1116
1.659	0.249	1117
0.607	0.201	1118
0.740	0.199	1119
0.286	0.211	1120
1.273	0.222	1121
1.275	0.232	1122
1.119	0.233	1123
0.062	0.194	1124
0.478	0.203	1125
-0.113	0.207	1126
0.842	0.217	1127
-0.218	0.184	1128
0.230	0.203	1129
0.520	0.215	1130
0.670	0.205	1131
0.504	0.199	1132
1.338	0.233	1133
1.659	0.256	1134
0.989	0.213	1135
-0.456	0.184	1136
0.060	0.198	1137
-0.196	0.201	1138
-0.667	0.186	1139
-0.801	0.192	1140
-0.005	0.197	1141
0.103	0.204	1142
0.315	0.201	1143
-0.885	0.193	1144
-0.281	0.188	1145
-0.126	0.187	1146
-0.427	0.197	1147
-0.374	0.179	1148
-0.819	0.203	1149
-0.903	0.184	1150
-0.436	0.182	1151
-0.122	0.200	1152

-0.283	0.191	1153
-0.613	0.182	1154
-0.182	0.199	1155
-0.450	0.173	1156
-0.880	0.175	1157

ตัวอย่างผลการวิเคราะห์องค์ประกอบเชิงยืนยันด้วยโปรแกรม LISREL Version 8.30

DATE: 10/15/2003
TIME: 21:14

L I S R E L 8.30

BY

Karl G. Jöreskog & Dag Sörbom

This program is published exclusively by
Scientific Software International, Inc.
7383 N. Lincoln Avenue, Suite 100
Chicago, IL 60646-1704, U.S.A.
Phone: (800)247-6113, (847)675-0720, Fax: (847)675-2140
Copyright by Scientific Software International, Inc., 1981-99
Use of this program is subject to the terms specified in the
Universal Copyright Convention.
Website: www.ssicentral.com

The following lines were read from file A:\DATA2.LS8:

confirmatory
da ni=30 no=1157 ma=cm
la
sco1 sco2 sco3 sco4 sco5 sco6 sco7 sco8 sco9 sco10 sco11 sco12 sco13 sco14 sco15
sco16 sco17 sco18 sco19 sco20 sco21 sco22 sco23 sco24 sco25 sco26 sco27 sco28
sco29 sco30
CM
0.864
0.297 0.895
0.327 0.381 0.913
0.290 0.261 0.318 0.897
0.311 0.342 0.367 0.292 0.889
0.364 0.353 0.384 0.335 0.414 0.872
0.307 0.297 0.317 0.295 0.334 0.416
0.324 0.316 0.357 0.336 0.453 0.417
0.313 0.270 0.272 0.288 0.288 0.349
0.321 0.305 0.384 0.337 0.382 0.422
0.289 0.298 0.307 0.288 0.324 0.306
0.304 0.275 0.272 0.307 0.301 0.369
0.339 0.293 0.337 0.342 0.306 0.375
0.369 0.366 0.401 0.362 0.431 0.452
0.321 0.326 0.364 0.321 0.351 0.387
0.312 0.321 0.331 0.285 0.322 0.385
0.289 0.238 0.255 0.268 0.284 0.329
0.301 0.279 0.318 0.264 0.299 0.341
0.289 0.290 0.344 0.310 0.330 0.373
0.327 0.295 0.351 0.315 0.348 0.370
0.327 0.334 0.373 0.342 0.359 0.408
0.253 0.287 0.349 0.275 0.287 0.361
0.228 0.239 0.305 0.272 0.263 0.321
0.290 0.273 0.309 0.268 0.300 0.352
0.303 0.290 0.349 0.336 0.326 0.334
0.284 0.304 0.396 0.315 0.319 0.328
0.301 0.271 0.331 0.303 0.322 0.375
0.280 0.290 0.350 0.275 0.295 0.339
0.304 0.305 0.367 0.324 0.324 0.311
0.262 0.295 0.298 0.261 0.309 0.289
0.798

0.361	0.885				
0.397	0.356	0.822			
0.342	0.400	0.289	0.894		
0.288	0.299	0.269	0.318	0.916	
0.283	0.313	0.310	0.354	0.260	0.921
0.365	0.366	0.326	0.336	0.278	0.333
0.364	0.463	0.368	0.441	0.345	0.389
0.331	0.361	0.305	0.351	0.281	0.308
0.343	0.349	0.287	0.346	0.238	0.324
0.336	0.301	0.322	0.266	0.230	0.274
0.250	0.319	0.260	0.316	0.297	0.301
0.309	0.381	0.295	0.374	0.271	0.294
0.319	0.363	0.336	0.352	0.275	0.334
0.331	0.403	0.329	0.366	0.299	0.322
0.266	0.336	0.251	0.302	0.258	0.284
0.261	0.276	0.204	0.320	0.213	0.254
0.300	0.342	0.317	0.304	0.255	0.276
0.300	0.318	0.279	0.373	0.297	0.287
0.234	0.306	0.238	0.371	0.242	0.295
0.352	0.352	0.306	0.333	0.262	0.284
0.271	0.317	0.282	0.336	0.258	0.292
0.253	0.289	0.246	0.345	0.275	0.304
0.289	0.292	0.265	0.317	0.270	0.287
0.783					
0.404	0.881				
0.361	0.388	0.895			
0.372	0.383	0.344	0.896		
0.345	0.327	0.276	0.301	0.728	
0.263	0.359	0.335	0.317	0.205	0.918
0.325	0.375	0.372	0.317	0.302	0.359
0.347	0.400	0.360	0.376	0.293	0.333
0.344	0.430	0.332	0.349	0.312	0.324
0.291	0.312	0.321	0.320	0.253	0.280
0.255	0.300	0.292	0.275	0.195	0.302
0.329	0.358	0.316	0.337	0.317	0.234
0.320	0.375	0.298	0.356	0.260	0.311
0.272	0.374	0.347	0.337	0.198	0.306
0.337	0.376	0.297	0.338	0.293	0.276
0.246	0.332	0.332	0.307	0.224	0.319
0.243	0.339	0.329	0.298	0.216	0.282
0.272	0.340	0.277	0.268	0.247	0.283
0.901					
0.328	0.897				
0.369	0.372	0.891			
0.299	0.309	0.330	0.913		
0.289	0.282	0.293	0.251	0.907	
0.303	0.335	0.336	0.285	0.214	0.811
0.331	0.346	0.346	0.241	0.295	0.314
0.336	0.322	0.346	0.306	0.305	0.287
0.320	0.341	0.394	0.301	0.256	0.385
0.346	0.371	0.359	0.298	0.279	0.302
0.326	0.310	0.337	0.270	0.317	0.301
0.268	0.274	0.350	0.247	0.237	0.310
0.910					
0.388	0.928				
0.354	0.324	0.817			
0.333	0.386	0.326	0.918		
0.395	0.413	0.289	0.405	0.930	
0.299	0.304	0.323	0.317	0.293	0.908

me
0.000 0.000 0.000 0.000 0.000 0.000
0.000 0.000 0.000 0.000 0.000 0.000
0.000 0.000 0.000 0.000 0.000 0.000
0.000 0.000 0.000 0.000 0.000 0.000
0.000 0.000 0.000 0.000 0.000 0.000
MO NX=30 NK=1 LX=FU,FR PH=FR TD=FU,FI
FI LX 1 1
VA 1.00 LX 1 1
FR TD 1 1 TD 2 2 TD 3 3 TD 4 4 TD 5 5 TD 6 6 TD 7 7 TD 8 8 TD 9 9 TD 10 10 TD 11 11 TD 12 12
FR TD 13 13 TD 14 14 TD 15 15 TD 16 16 TD 17 17 TD 18 18 TD 19 19 TD 20 20 TD 21 21 TD 22 22
FR TD 23 23 TD 24 24 TD 25 25 TD 26 26 TD 27 27 TD 28 28 TD 29 29 TD 30 30
LK
FACTOR
OU SE TV SS MI RS FS

confirmatory

Number of Input Variables 30
Number of Y - Variables 0
Number of X - Variables 30
Number of ETA - Variables 0
Number of KSI - Variables 1
Number of Observations 1157

W_A_R_N_I_N_G: Matrix to be analyzed is not positive definite,
ridge option taken with ridge constant = 10.000

confirmatory

Covariance Matrix to be Analyzed

	sco1	sco2	sco3	sco4	sco5	sco6
sco1	9.5040					
sco2	0.2970	9.8450				
sco3	0.3270	0.3810	10.0430			
sco4	0.2900	0.2610	0.3180	9.8670		
sco5	0.3110	0.3420	0.3670	0.2920	9.7790	
sco6	0.3640	0.3530	0.3840	0.3350	0.4140	9.5920
sco7	0.3070	0.2970	0.3170	0.2950	0.3340	0.4160
sco8	0.3160	0.3570	0.3360	0.4530	0.4170	0.3130
sco9	0.2880	0.2880	0.3490	0.3210	0.3050	0.3840
sco10	0.2890	0.2980	0.3070	0.2880	0.3240	0.3060
sco11	0.3010	0.3690	0.3390	0.2930	0.3370	0.3420
sco12	0.3620	0.4310	0.4520	0.3210	0.3260	0.3640
sco13	0.2850	0.3220	0.3850	0.2890	0.2380	0.2550
sco14	0.2990	0.3410	0.2890	0.2900	0.3440	0.3100
sco15	0.3270	0.3340	0.3730	0.3420	0.3590	0.4080
sco16	0.2720	0.2630	0.3210	0.2900	0.2730	0.3090
sco17	0.3040	0.3960	0.3150	0.3190	0.3280	0.3010
sco18	0.3040	0.3050	0.3670	0.3240	0.3240	0.3110
sco19	0.3420	0.4000	0.2890	0.8940	0.2880	0.2990
sco20	0.2780	0.3330	0.3640	0.4630	0.3680	0.4410
sco21	0.3360	0.3010	0.3220	0.2660	0.2300	0.2740
sco22	0.3520	0.2750	0.3340	0.3310	0.4030	0.3290
sco23	0.3420	0.3170	0.3040	0.2550	0.2760	0.3000
sco24	0.2710	0.3170	0.2820	0.3360	0.2580	0.2920
sco25	0.3720	0.3830	0.3440	0.8960	0.3450	0.3270
sco26	0.2930	0.3330	0.3440	0.4300	0.3320	0.3490

sco27	0.3200	0.3750	0.2980	0.3560	0.2600	0.3110
sco28	0.2980	0.2160	0.2820	0.2720	0.3400	0.2770
sco29	0.2140	0.8110	0.3310	0.3460	0.3460	0.2410
sco30	0.2700	0.3170	0.3010	0.2680	0.2740	0.3500

Covariance Matrix to be Analyzed

	sco7	sco8	sco9	sco10	sco11	sco12
sco7	3.5640					
sco8	0.2700	2.9920				
sco9	0.3370	0.3820	4.6420			
sco10	0.3040	0.2750	0.2720	3.3770		
sco11	0.3060	0.3750	0.3690	0.3660	4.4110	
sco12	0.3210	0.3510	0.3870	0.3120	0.3210	3.6410
sco13	0.2680	0.2840	0.3290	0.3010	0.2790	0.3180
sco14	0.3300	0.3730	0.3270	0.2950	0.3510	0.3150
sco15	0.2530	0.2870	0.3490	0.2750	0.2870	0.3610
sco16	0.2680	0.3000	0.3520	0.3030	0.2900	0.3490
sco17	0.2710	0.3310	0.3030	0.3220	0.3750	0.2800
sco18	0.2620	0.2950	0.2980	0.2610	0.3090	0.2890
sco19	0.2690	0.3180	0.9160	0.2830	0.3130	0.3100
sco20	0.3450	0.3890	0.3310	0.3610	0.3050	0.3510
sco21	0.2500	0.3190	0.2600	0.3160	0.2970	0.3010
sco22	0.3660	0.2990	0.3220	0.2660	0.3360	0.2510
sco23	0.3180	0.2790	0.3730	0.2970	0.2870	0.2340
sco24	0.2530	0.2890	0.2460	0.3450	0.2750	0.3040
sco25	0.2760	0.3010	0.7280	0.2630	0.3590	0.3350
sco26	0.3120	0.3240	0.2910	0.3120	0.3210	0.3200
sco27	0.2720	0.3740	0.3470	0.3370	0.1980	0.3060
sco28	0.2680	0.2470	0.2830	0.9010	0.3280	0.8970
sco29	0.2950	0.3140	0.3360	0.3220	0.3460	0.3060
sco30	0.2470	0.2370	0.3100	0.9100	0.3880	0.9280

Covariance Matrix to be Analyzed

	sco13	sco14	sco15	sco16	sco17	sco18
sco13	2.9040					
sco14	0.3480	4.0700				
sco15	0.2280	0.2390	3.3550			
sco16	0.3360	0.3260	0.3340	3.1240		
sco17	0.2900	0.3500	0.2750	0.2950	3.7290	
sco18	0.7980	0.3610	0.8850	0.3970	0.3560	9.0420
sco19	0.3540	0.2600	0.9210	0.3650	0.3660	0.3260
sco20	0.2810	0.3080	0.3430	0.3490	0.2870	0.3460
sco21	0.3090	0.3810	0.2950	0.3740	0.2710	0.2940
sco22	0.3020	0.2580	0.2840	0.2610	0.2760	0.2040
sco23	0.3060	0.2380	0.3710	0.2420	0.2950	0.3520
sco24	0.2890	0.2920	0.2650	0.3170	0.2700	0.2870
sco25	0.3170	0.2050	0.9180	0.3250	0.3750	0.3720
sco26	0.2530	0.2800	0.2550	0.3000	0.2920	0.2750
sco27	0.3370	0.3760	0.2970	0.3380	0.2930	0.2760
sco28	0.3690	0.3720	0.8910	0.2990	0.3090	0.3300
sco29	0.3050	0.2870	0.3200	0.3410	0.3940	0.3010
sco30	0.3540	0.3240	0.8170	0.3330	0.3860	0.3260

Covariance Matrix to be Analyzed

	sco19	sco20	sco21	sco22	sco23	sco24
sco19	3.6960					
sco20	0.2380	3.5640				
sco21	0.3190	0.3630	3.6960			
sco22	0.3200	0.2130	0.2540	3.3000		
sco23	0.3520	0.3060	0.3330	0.2620	3.1240	
sco24	0.7830	0.4040	0.8810	0.3610	0.3880	9.8450
sco25	0.3170	0.3020	0.3590	0.3470	0.4000	0.3600
sco26	0.1950	0.3020	0.3290	0.3580	0.3160	0.3370
sco27	0.2460	0.3320	0.3320	0.3070	0.2240	0.3190
sco28	0.9130	0.2890	0.2820	0.2930	0.2510	0.9070
sco29	0.2560	0.3850	0.3460	0.3710	0.3590	0.2980
sco30	0.9180	0.3950	0.4130	0.2890	0.4050	0.9300

Covariance Matrix to be Analyzed

	sco25	sco26	sco27	sco28	sco29	sco30
sco25	4.1360					
sco26	0.3170	2.5740				
sco27	0.2430	0.3390	3.6190			
sco28	0.3030	0.3350	0.3360	3.1350		
sco29	0.2790	0.3020	0.3260	0.3100	3.7070	
sco30	0.2990	0.3040	0.3230	0.3170	0.2930	9.9880

confirmatory

Parameter Specifications

LAMBDA-X

FACTOR

sco1	0
sco2	1
sco3	2
sco4	3
sco5	4
sco6	5
sco7	6
sco8	7
sco9	8
sco10	9
sco11	10
sco12	11
sco13	12
sco14	13
sco15	14
sco16	15
sco17	16
sco18	17
sco19	18
sco20	19
sco21	20
sco22	21
sco23	22
sco24	23
sco25	24

sco26 25
sco27 26
sco28 27
sco29 28
sco30 29

PHI

FACTOR

30

THETA-DELTA

sco1	sco2	sco3	sco4	sco5	sco6
-----	-----	-----	-----	-----	-----
31	32	33	34	35	36

THETA-DELTA

sco7	sco8	sco9	sco10	sco11	sco12
-----	-----	-----	-----	-----	-----
37	38	39	40	41	42

THETA-DELTA

sco13	sco14	sco15	sco16	sco17	sco18
-----	-----	-----	-----	-----	-----
43	44	45	46	47	48

THETA-DELTA

sco19	sco20	sco21	sco22	sco23	sco24
-----	-----	-----	-----	-----	-----
49	50	51	52	53	54

THETA-DELTA

sco25	sco26	sco27	sco28	sco29	sco30
-----	-----	-----	-----	-----	-----
55	56	57	58	59	60

confirmatory

Number of Iterations = 9

LISREL Estimates (Maximum Likelihood)

LAMBDA-X

FACTOR

sco1 1.0000

sco2 1.1186
(0.3053)
3.6640

sco3 1.0807
(0.3012)
3.5874

sco4 1.2252
(0.3214)
3.8125

sco5 1.0440
(0.2941)
3.5502

sco6 1.0619
(0.2952)
3.5969

sco7 0.9440
(0.2263)
4.1706

sco8 1.0211
(0.2340)
4.3630

sco9 1.2154
(0.2817)
4.3142

sco10 1.1530
(0.2606)
4.4244

sco11 1.0371
(0.2496)
4.1545

sco12 1.2170
(0.2741)
4.4399

sco13 1.0129
(0.2318)
4.3702

sco14 1.0022
(0.2408)
4.1622

sco15 1.3939
(0.3039)
4.5871

sco16 1.0224
(0.2355)
4.3416

sco17 1.0137
(0.2396)
4.2309

sco18 1.1929
(0.3107)
3.8395

sco19 1.4257
(0.3121)
4.5679

sco20 1.0536
(0.2449)
4.3022

sco21 1.0432
(0.2443)
4.2702

sco22 0.9648
(0.2273)
4.2449

sco23 0.9875
(0.2295)
4.3035

sco24 1.3271
(0.3370)
3.9382

sco25 1.2258
(0.2796)
4.3845

sco26 0.9807
(0.2230)
4.3987

sco27 1.0009
(0.2364)
4.2336

sco28 1.4576
(0.3141)
4.6409

sco29 1.0542
(0.2463)
4.2805

sco30 1.4656
(0.3600)
4.0707

PHI

FACTOR

0.2713
(0.1102)
2.4630

THETA-DELTA

sco1	sco2	sco3	sco4	sco5	sco6
9.2327	9.5055	9.7261	9.4598	9.4833	9.2861
(0.3884)	(0.4008)	(0.4096)	(0.4000)	(0.3992)	(0.3911)
23.7740	23.7163	23.7449	23.6495	23.7576	23.7416

THETA-DELTA

sco7	sco8	sco9	sco10	sco11	sco12
3.3222	2.7091	4.2412	3.0163	4.1192	3.2392
(0.1421)	(0.1173)	(0.1830)	(0.1314)	(0.1761)	(0.1414)
23.3786	23.0902	23.1805	22.9517	23.3962	22.9110

THETA-DELTA

sco13	sco14	sco15	sco16	sco17	sco18
2.6256	3.7975	2.8278	2.8404	3.4502	8.6559
(0.1138)	(0.1624)	(0.1266)	(0.1228)	(0.1480)	(0.3662)
23.0755	23.3879	22.3417	23.1317	23.3054	23.6354

THETA-DELTA

sco19	sco20	sco21	sco22	sco23	sco24
3.1445	3.2628	3.4007	3.0475	2.8594	9.3671
(0.1401)	(0.1406)	(0.1463)	(0.1309)	(0.1233)	(0.3973)
22.4426	23.2006	23.2505	23.2866	23.1985	23.5770

THETA-DELTA

sco25	sco26	sco27	sco28	sco29	sco30
3.7284	2.3130	3.3472	2.5585	3.4055	9.4053
(0.1618)	(0.1005)	(0.1436)	(0.1164)	(0.1466)	(0.4006)
23.0452	23.0135	23.3018	21.9867	23.2349	23.4772

Squared Multiple Correlations for X - Variables

sco1	sco2	sco3	sco4	sco5	sco6
0.0285	0.0345	0.0316	0.0413	0.0302	0.0319

Squared Multiple Correlations for X - Variables

sco7	sco8	sco9	sco10	sco11	sco12
0.0678	0.0945	0.0863	0.1068	0.0662	0.1104

Squared Multiple Correlations for X - Variables

sco13	sco14	sco15	sco16	sco17	sco18
-------	-------	-------	-------	-------	-------

0.0959	0.0669	0.1571	0.0908	0.0748	0.0427
--------	--------	--------	--------	--------	--------

Squared Multiple Correlations for X - Variables

sco19	sco20	sco21	sco22	sco23	sco24
0.1492	0.0845	0.0799	0.0765	0.0847	0.0485

Squared Multiple Correlations for X - Variables

sco25	sco26	sco27	sco28	sco29	sco30
0.0986	0.1014	0.0751	0.1839	0.0813	0.0583

Goodness of Fit Statistics

Degrees of Freedom = 405
Minimum Fit Function Chi-Square = 435.6870 (P = 0.1411)
Normal Theory Weighted Least Squares Chi-Square = 413.2592 (P = 0.3777)
Estimated Non-centrality Parameter (NCP) = 8.2592
90 Percent Confidence Interval for NCP = (0.0 ; 60.2909)

Minimum Fit Function Value = 0.3769
Population Discrepancy Function Value (F0) = 0.007145
90 Percent Confidence Interval for F0 = (0.0 ; 0.05215)
Root Mean Square Error of Approximation (RMSEA) = 0.004200
90 Percent Confidence Interval for RMSEA = (0.0 ; 0.01135)
P-Value for Test of Close Fit (RMSEA < 0.05) = 1.0000

Expected Cross-Validation Index (ECVI) = 0.4613
90 Percent Confidence Interval for ECVI = (0.4542 ; 0.5063)
ECVI for Saturated Model = 0.8045
ECVI for Independence Model = 1.6416

Chi-Square for Independence Model with 435 Degrees of Freedom = 1837.7335
Independence AIC = 1897.7335
Model AIC = 533.2592
Saturated AIC = 930.0000
Independence CAIC = 2079.3411
Model CAIC = 896.4744
Saturated CAIC = 3744.9174

Root Mean Square Residual (RMR) = 0.1127
Standardized RMR = 0.02456
Goodness of Fit Index (GFI) = 0.9767
Adjusted Goodness of Fit Index (AGFI) = 0.9733
Parsimony Goodness of Fit Index (PGFI) = 0.8507

Normed Fit Index (NFI) = 0.7629
Non-Normed Fit Index (NNFI) = 0.9765
Parsimony Normed Fit Index (PNFI) = 0.7103
Comparative Fit Index (CFI) = 0.9781
Incremental Fit Index (IFI) = 0.9786
Relative Fit Index (RFI) = 0.7454

Critical N (CN) = 1259.0168

***** EQUATE VERSION 2.1 *****
 DATE: DAY= 12 MONTH= 3 YEAR= 2003
 START TIME= 14:59: 5
 EQUATING TEST Y TO TEST X
 NUMBER OF ABILITY SCALE POINTS= 21
 GRADED RESPONSE MODEL
 FROM METRIC ITEM PARAMETER FILE NAME IS
 D:\EQUATE\STREAY.PAR
 FILE FORMAT IS (/6F10.3)
 NUMBER OF ITEMS IN "FROM" TEST IS 30
 FROM TEST IS IN LOGISTIC OGIVE METRIC

317

TO METRIC ITEM PARAMETER FILE NAME IS
 D:\EQUATE\STREAX.PAR
 FILE FORMAT IS (/6F10.3)
 NUMBER OF ITEMS IN "TO" TEST IS 30
 TO TEST IS IN LOGISTIC OGIVE METRIC.

TRANSFORMED ITEM PARAMETER FILE NAME IS
 D:\EQUATE\NIPW.DAT

TRANSFORMED PARAMETERS WILL BE IN
 LOGISTIC OGIVE METRIC

NAME OF "FROM" THETA FILE IS
 D:\EQUATE\STREAY.SC
 FILE FORMAT IS (4X,F6.3)
 NUMBER OF EXAMINEES IN "FROM" THETA FILE IS 1157

NAME OF FILE FOR TRANSFORMED THETAS IS
 D:\EQUATE\NTHYW.DAT

ANCHOR ITEM IDs FOR "FROM" INSTRUMENT ARE
 2-4,12,16-17,22,25,29-30:

ITEM	N CATEGORIES	CATEGORY WITH HIGH WEIGHT
1	6	6
2	6	6
3	6	6
4	6	6
5	6	6
6	6	6
7	6	6
8	6	6
9	6	6
10	6	6
11	6	6
12	6	6
13	6	6
14	6	6
15	6	6
16	6	6
17	6	6
18	6	6
19	6	6
20	6	6
21	6	6
22	6	6
23	6	6
24	6	6
25	6	6

26	6	6
27	6	6
28	6	6
29	6	6
30	6	6

318

ANCHOR ITEM IDS FOR "TO" INSTRUMENT ARE
2-4,12,16-17,22,25,29-30:

ITEM N CATEGORIES CATEGORY WITH HIGH WEIGHT

1	6	6
2	6	6
3	6	6
4	6	6
5	6	6
6	6	6
7	6	6
8	6	6
9	6	6
10	6	6
11	6	6
12	6	6
13	6	6
14	6	6
15	6	6
16	6	6
17	6	6
18	6	6
19	6	6
20	6	6
21	6	6
22	6	6
23	6	6
24	6	6
25	6	6
26	6	6
27	6	6
28	6	6
29	6	6
30	6	6

INITIAL VALUE FOR A= 1.0330 INITIAL VALUE FOR K= -0.0468
FUNCTION AT INITIAL VALUES = 0.033878

NUMBER OF ITERATIONS PERFORMED = 5
METRIC TRANSFORMATION COEFFICIENTS ARE

A= 1.0252 K= -0.0617

FUNCTION VALUE = 0.025858

SUMMARY STATISTICS OF TRANSFORMED ABILITIES

MEAN = 0.290058
VARIANCE= 0.897033
STANDARD DEVIATION= 0.947118

LABORATORY OF EXPERIMENTAL DESIGN
DEPARTMENT OF EDUCATIONAL PSYCHOLOGY
UNIVERSITY OF WISCONSIN
PROGRAM DATED 14 JULY 1995
END TIME= 15: 8:29

***** EQUATE VERSION 2.1 *****

ตัวอย่างคำสั่งบนโปรแกรม SPSS Version 7.5 ที่ใช้คำนวณค่าความน่าจะเป็นของผู้สอบที่มีความสามารถระดับ θ จะตอบข้อสอบข้อที่ j ได้คะแนนในลำดับขั้น k หรือสูงกว่า ($P_{jk}^*(\theta)$)

```
compute phy11 = (2.718**((1.7*1.52005)*(thy + 2.91123)))/(1+(2.718**((1.7*1.52005)*(thy + 2.91123)))).
compute phy12 = (2.718**((1.7*1.52005)*(thy + 2.02208)))/(1+(2.718**((1.7*1.52005)*(thy + 2.02208)))).
compute phy13 = (2.718**((1.7*1.52005)*(thy + 1.07634)))/(1+(2.718**((1.7*1.52005)*(thy + 1.07634)))).
compute phy14 = (2.718**((1.7*1.52005)*(thy + 0.22239)))/(1+(2.718**((1.7*1.52005)*(thy + 0.22239)))).
compute phy15 = (2.718**((1.7*1.52005)*(thy - 0.77716)))/(1+(2.718**((1.7*1.52005)*(thy - 0.77716)))).
compute phy21 = (2.718**((1.7*1.47434)*(thy + 4.87186)))/(1+(2.718**((1.7*1.47434)*(thy + 4.87186)))).
compute phy22 = (2.718**((1.7*1.47434)*(thy + 2.48415)))/(1+(2.718**((1.7*1.47434)*(thy + 2.48415)))).
compute phy23 = (2.718**((1.7*1.47434)*(thy + 0.88316)))/(1+(2.718**((1.7*1.47434)*(thy + 0.88316)))).
compute phy24 = (2.718**((1.7*1.47434)*(thy - 0.06188)))/(1+(2.718**((1.7*1.47434)*(thy - 0.06188)))).
compute phy25 = (2.718**((1.7*1.47434)*(thy - 2.17238)))/(1+(2.718**((1.7*1.47434)*(thy - 2.17238)))).
compute phy31 = (2.718**((1.7*1.65788)*(thy + 3.96560)))/(1+(2.718**((1.7*1.65788)*(thy + 3.96560)))).
compute phy32 = (2.718**((1.7*1.65788)*(thy + 1.67628)))/(1+(2.718**((1.7*1.65788)*(thy + 1.67628)))).
compute phy33 = (2.718**((1.7*1.65788)*(thy + 0.78831)))/(1+(2.718**((1.7*1.65788)*(thy + 0.78831)))).
compute phy34 = (2.718**((1.7*1.65788)*(thy - 0.40888)))/(1+(2.718**((1.7*1.65788)*(thy - 0.40888)))).
compute phy35 = (2.718**((1.7*1.65788)*(thy - 1.69728)))/(1+(2.718**((1.7*1.65788)*(thy - 1.69728)))).
compute phy41 = (2.718**((1.7*1.45341)*(thy + 4.45427)))/(1+(2.718**((1.7*1.45341)*(thy + 4.45427)))).
compute phy42 = (2.718**((1.7*1.45341)*(thy + 2.39965)))/(1+(2.718**((1.7*1.45341)*(thy + 2.39965)))).
compute phy43 = (2.718**((1.7*1.45341)*(thy + 1.09776)))/(1+(2.718**((1.7*1.45341)*(thy + 1.09776)))).
compute phy44 = (2.718**((1.7*1.45341)*(thy - 0.03543)))/(1+(2.718**((1.7*1.45341)*(thy - 0.03543)))).
compute phy45 = (2.718**((1.7*1.45341)*(thy - 1.32706)))/(1+(2.718**((1.7*1.45341)*(thy - 1.32706)))).
compute phy51 = (2.718**((1.7*1.74034)*(thy + 2.18011)))/(1+(2.718**((1.7*1.74034)*(thy + 2.18011)))).
compute phy52 = (2.718**((1.7*1.74034)*(thy + 1.26456)))/(1+(2.718**((1.7*1.74034)*(thy + 1.26456)))).
compute phy53 = (2.718**((1.7*1.74034)*(thy + 0.35312)))/(1+(2.718**((1.7*1.74034)*(thy + 0.35312)))).
compute phy54 = (2.718**((1.7*1.74034)*(thy - 0.24165)))/(1+(2.718**((1.7*1.74034)*(thy - 0.24165)))).
compute phy55 = (2.718**((1.7*1.74034)*(thy - 0.92776)))/(1+(2.718**((1.7*1.74034)*(thy - 0.92776)))).
compute phy61 = (2.718**((1.7*2.14315)*(thy + 2.52926)))/(1+(2.718**((1.7*2.14315)*(thy + 2.52926)))).
compute phy62 = (2.718**((1.7*2.14315)*(thy + 1.30411)))/(1+(2.718**((1.7*2.14315)*(thy + 1.30411)))).
compute phy63 = (2.718**((1.7*2.14315)*(thy + 0.58151)))/(1+(2.718**((1.7*2.14315)*(thy + 0.58151)))).
compute phy64 = (2.718**((1.7*2.14315)*(thy - 0.04506)))/(1+(2.718**((1.7*2.14315)*(thy - 0.04506)))).
compute phy65 = (2.718**((1.7*2.14315)*(thy - 0.72388)))/(1+(2.718**((1.7*2.14315)*(thy - 0.72388)))).
compute phy71 = (2.718**((1.7*2.11564)*(thy + 2.21358)))/(1+(2.718**((1.7*2.11564)*(thy + 2.21358)))).
compute phy72 = (2.718**((1.7*2.11564)*(thy + 1.42308)))/(1+(2.718**((1.7*2.11564)*(thy + 1.42308)))).
compute phy73 = (2.718**((1.7*2.11564)*(thy + 0.83271)))/(1+(2.718**((1.7*2.11564)*(thy + 0.83271)))).
compute phy74 = (2.718**((1.7*2.11564)*(thy + 0.30186)))/(1+(2.718**((1.7*2.11564)*(thy + 0.30186)))).
compute phy75 = (2.718**((1.7*2.11564)*(thy - 0.21357)))/(1+(2.718**((1.7*2.11564)*(thy - 0.21357)))).
compute phy81 = (2.718**((1.7*1.96250)*(thy + 2.38051)))/(1+(2.718**((1.7*1.96250)*(thy + 2.38051)))).
compute phy82 = (2.718**((1.7*1.96250)*(thy + 1.49027)))/(1+(2.718**((1.7*1.96250)*(thy + 1.49027)))).
compute phy83 = (2.718**((1.7*1.96250)*(thy + 0.64546)))/(1+(2.718**((1.7*1.96250)*(thy + 0.64546)))).
compute phy84 = (2.718**((1.7*1.96250)*(thy - 0.09428)))/(1+(2.718**((1.7*1.96250)*(thy - 0.09428)))).
compute phy85 = (2.718**((1.7*1.96250)*(thy - 0.87294)))/(1+(2.718**((1.7*1.96250)*(thy - 0.87294)))).
compute phy91 = (2.718**((1.7*1.61225)*(thy + 2.96988)))/(1+(2.718**((1.7*1.61225)*(thy + 2.96988)))).
compute phy92 = (2.718**((1.7*1.61225)*(thy + 2.05191)))/(1+(2.718**((1.7*1.61225)*(thy + 2.05191)))).
compute phy93 = (2.718**((1.7*1.61225)*(thy + 1.11865)))/(1+(2.718**((1.7*1.61225)*(thy + 1.11865)))).
compute phy94 = (2.718**((1.7*1.61225)*(thy + 0.51015)))/(1+(2.718**((1.7*1.61225)*(thy + 0.51015)))).
compute phy95 = (2.718**((1.7*1.61225)*(thy - 0.43815)))/(1+(2.718**((1.7*1.61225)*(thy - 0.43815)))).
compute phy101 = (2.718**((1.7*1.87621)*(thy + 1.88506)))/(1+(2.718**((1.7*1.87621)*(thy + 1.88506)))).
compute phy102 = (2.718**((1.7*1.87621)*(thy + 1.09735)))/(1+(2.718**((1.7*1.87621)*(thy + 1.09735)))).
compute phy103 = (2.718**((1.7*1.87621)*(thy + 0.29557)))/(1+(2.718**((1.7*1.87621)*(thy + 0.29557)))).
compute phy104 = (2.718**((1.7*1.87621)*(thy - 0.42531)))/(1+(2.718**((1.7*1.87621)*(thy - 0.42531)))).
compute phy105 = (2.718**((1.7*1.87621)*(thy - 0.95138)))/(1+(2.718**((1.7*1.87621)*(thy - 0.95138)))).
compute phy111 = (2.718**((1.7*1.27943)*(thy + 3.53997)))/(1+(2.718**((1.7*1.27943)*(thy + 3.53997)))).
compute phy112 = (2.718**((1.7*1.27943)*(thy + 1.98560)))/(1+(2.718**((1.7*1.27943)*(thy + 1.98560)))).
compute phy113 = (2.718**((1.7*1.27943)*(thy + 0.88286)))/(1+(2.718**((1.7*1.27943)*(thy + 0.88286)))).
compute phy114 = (2.718**((1.7*1.27943)*(thy - 0.74337)))/(1+(2.718**((1.7*1.27943)*(thy - 0.74337)))).
compute phy115 = (2.718**((1.7*1.27943)*(thy - 2.71400)))/(1+(2.718**((1.7*1.27943)*(thy - 2.71400)))).
compute phy121 = (2.718**((1.7*1.28815)*(thy + 4.02841)))/(1+(2.718**((1.7*1.28815)*(thy + 4.02841)))).
compute phy122 = (2.718**((1.7*1.28815)*(thy + 2.08449)))/(1+(2.718**((1.7*1.28815)*(thy + 2.08449)))).
compute phy123 = (2.718**((1.7*1.28815)*(thy + 0.67047)))/(1+(2.718**((1.7*1.28815)*(thy + 0.67047)))).
compute phy124 = (2.718**((1.7*1.28815)*(thy - 0.50829)))/(1+(2.718**((1.7*1.28815)*(thy - 0.50829)))).
compute phy125 = (2.718**((1.7*1.28815)*(thy - 2.17003)))/(1+(2.718**((1.7*1.28815)*(thy - 2.17003)))).
compute phy131 = (2.718**((1.7*2.30909)*(thy + 2.22187)))/(1+(2.718**((1.7*2.30909)*(thy + 2.22187)))).
compute phy132 = (2.718**((1.7*2.30909)*(thy + 1.49799)))/(1+(2.718**((1.7*2.30909)*(thy + 1.49799)))).
```



```

compute phy265 = (2.718**((1.7*1.50931)*(thy - 1.98554)))/(1+(2.718**((1.7*1.50931)*(thy - 1.98554)))).
compute phy271 = (2.718**((1.7*2.02654)*(thy + 2.48263)))/(1+(2.718**((1.7*2.02654)*(thy + 2.48263)))).
compute phy272 = (2.718**((1.7*2.02654)*(thy + 1.21584)))/(1+(2.718**((1.7*2.02654)*(thy + 1.21584)))).
compute phy273 = (2.718**((1.7*2.02654)*(thy + 0.67602)))/(1+(2.718**((1.7*2.02654)*(thy + 0.67602)))).
compute phy274 = (2.718**((1.7*2.02654)*(thy + 0.18455)))/(1+(2.718**((1.7*2.02654)*(thy + 0.18455)))).
compute phy275 = (2.718**((1.7*2.02654)*(thy - 0.31908)))/(1+(2.718**((1.7*2.02654)*(thy - 0.31908)))).
compute phy281 = (2.718**((1.7*1.43976)*(thy + 3.17880)))/(1+(2.718**((1.7*1.43976)*(thy + 3.17880)))).
compute phy282 = (2.718**((1.7*1.43976)*(thy + 1.38321)))/(1+(2.718**((1.7*1.43976)*(thy + 1.38321)))).
compute phy283 = (2.718**((1.7*1.43976)*(thy + 0.46335)))/(1+(2.718**((1.7*1.43976)*(thy + 0.46335)))).
compute phy284 = (2.718**((1.7*1.43976)*(thy - 0.40491)))/(1+(2.718**((1.7*1.43976)*(thy - 0.40491)))).
compute phy285 = (2.718**((1.7*1.43976)*(thy - 1.51515)))/(1+(2.718**((1.7*1.43976)*(thy - 1.51515)))).
compute phy291 = (2.718**((1.7*1.42479)*(thy + 2.28789)))/(1+(2.718**((1.7*1.42479)*(thy + 2.28789)))).
compute phy292 = (2.718**((1.7*1.42479)*(thy + 1.01688)))/(1+(2.718**((1.7*1.42479)*(thy + 1.01688)))).
compute phy293 = (2.718**((1.7*1.42479)*(thy - 0.01888)))/(1+(2.718**((1.7*1.42479)*(thy - 0.01888)))).
compute phy294 = (2.718**((1.7*1.42479)*(thy - 0.89519)))/(1+(2.718**((1.7*1.42479)*(thy - 0.89519)))).
compute phy295 = (2.718**((1.7*1.42479)*(thy - 1.82485)))/(1+(2.718**((1.7*1.42479)*(thy - 1.82485)))).
compute phy301 = (2.718**((1.7*1.29781)*(thy + 2.92568)))/(1+(2.718**((1.7*1.29781)*(thy + 2.92568)))).
compute phy302 = (2.718**((1.7*1.29781)*(thy + 1.71367)))/(1+(2.718**((1.7*1.29781)*(thy + 1.71367)))).
compute phy303 = (2.718**((1.7*1.29781)*(thy + 0.58268)))/(1+(2.718**((1.7*1.29781)*(thy + 0.58268)))).
compute phy304 = (2.718**((1.7*1.29781)*(thy - 0.31982)))/(1+(2.718**((1.7*1.29781)*(thy - 0.31982)))).
compute phy305 = (2.718**((1.7*1.29781)*(thy - 1.32558)))/(1+(2.718**((1.7*1.29781)*(thy - 1.32558)))).
execute.

```

ตัวอย่างคำสั่งบนโปรแกรม SPSS Version 7.5 ที่ใช้คำนวณค่าความน่าจะเป็นของผู้สอบที่มีความสามารถระดับ θ จะตอบข้อสอบข้อที่ j ได้คะแนนลำดับขั้น k ($P_{jk}(\theta)$)

```
compute py10 = 1- phy11.
compute py11 = phy11 - phy12.
compute py12 = phy12 - phy13.
compute py13 = phy13 - phy14.
compute py14 = phy14 - phy15.
compute py15 = phy15.
compute py20 = 1- phy21.
compute py21 = phy21 - phy22.
compute py22 = phy22 - phy23.
compute py23 = phy23 - phy24.
compute py24 = phy24 - phy25.
compute py25 = phy25.
compute py30 = 1- phy31.
compute py31 = phy31 - phy32.
compute py32 = phy32 - phy33.
compute py33 = phy33 - phy34.
compute py34 = phy34 - phy35.
compute py35 = phy35.
compute py40 = 1- phy41.
compute py41 = phy41 - phy42.
compute py42 = phy42 - phy43.
compute py43 = phy43 - phy44.
compute py44 = phy44 - phy45.
compute py45 = phy45.
compute py50 = 1- phy51.
compute py51 = phy51 - phy52.
compute py52 = phy52 - phy53.
compute py53 = phy53 - phy54.
compute py54 = phy54 - phy55.
compute py55 = phy55.
compute py60 = 1- phy61.
compute py61 = phy61 - phy62.
compute py62 = phy62 - phy63.
compute py63 = phy63 - phy64.
compute py64 = phy64 - phy65.
compute py65 = phy65.
compute py70 = 1- phy71.
compute py71 = phy71 - phy72.
compute py72 = phy72 - phy73.
compute py73 = phy73 - phy74.
compute py74 = phy74 - phy75.
compute py75 = phy75.
compute py80 = 1- phy81.
compute py81 = phy81 - phy82.
compute py82 = phy82 - phy83.
compute py83 = phy83 - phy84.
compute py84 = phy84 - phy85.
compute py85 = phy85.
compute py90 = 1- phy91.
compute py91 = phy91 - phy92.
compute py92 = phy92 - phy93.
compute py93 = phy93 - phy94.
compute py94 = phy94 - phy95.
compute py95 = phy95.
compute py100 = 1- phy101.
compute py101 = phy101 - phy102.
compute py102 = phy102 - phy103.
compute py103 = phy103 - phy104.
compute py104 = phy104 - phy105.
compute py105 = phy105.
compute py110 = 1- phy111.
compute py111 = phy111 - phy112.
```

```

compute py112 = phy112 - phy113.
compute py113 = phy113 - phy114.
compute py114 = phy114 - phy115.
compute py115 = phy115.
compute py120 = 1- phy121.
compute py121 = phy121 - phy122.
compute py122 = phy122 - phy123.
compute py123 = phy123 - phy124.
compute py124 = phy124 - phy125.
compute py125 = phy125.
compute py130 = 1- phy131.
compute py131 = phy131 - phy132.
compute py132 = phy132 - phy133.
compute py133 = phy133 - phy134.
compute py134 = phy134 - phy135.
compute py135 = phy135.
compute py140 = 1- phy141.
compute py141 = phy141 - phy142.
compute py142 = phy142 - phy143.
compute py143 = phy143 - phy144.
compute py144 = phy144 - phy145.
compute py145 = phy145.
compute py150 = 1- phy151.
compute py151 = phy151 - phy152.
compute py152 = phy152 - phy153.
compute py153 = phy153 - phy154.
compute py154 = phy154 - phy155.
compute py155 = phy155.
compute py160 = 1- phy161.
compute py161 = phy161 - phy162.
compute py162 = phy162 - phy163.
compute py163 = phy163 - phy164.
compute py164 = phy164 - phy165.
compute py165 = phy165.
compute py170 = 1- phy171.
compute py171 = phy171 - phy172.
compute py172 = phy172 - phy173.
compute py173 = phy173 - phy174.
compute py174 = phy174 - phy175.
compute py175 = phy175.
compute py180 = 1- phy181.
compute py181 = phy181 - phy182.
compute py182 = phy182 - phy183.
compute py183 = phy183 - phy184.
compute py184 = phy184 - phy185.
compute py185 = phy185.
compute py190 = 1- phy191.
compute py191 = phy191 - phy192.
compute py192 = phy192 - phy193.
compute py193 = phy193 - phy194.
compute py194 = phy194 - phy195.
compute py195 = phy195.
compute py200 = 1- phy201.
compute py201 = phy201 - phy202.
compute py202 = phy202 - phy203.
compute py203 = phy203 - phy204.
compute py204 = phy204 - phy205.
compute py205 = phy205.
compute py210 = 1- phy211.
compute py211 = phy211 - phy212.
compute py212 = phy212 - phy213.
compute py213 = phy213 - phy214.
compute py214 = phy214 - phy215.
compute py215 = phy215.
compute py220 = 1- phy221.
compute py221 = phy221 - phy222.
compute py222 = phy222 - phy223.

```

```
compute py223 = phy223 - phy224.  
compute py224 = phy224 - phy225.  
compute py225 = phy225.  
compute py230 = 1- phy231.  
compute py231 = phy231 - phy232.  
compute py232 = phy232 - phy233.  
compute py233 = phy233 - phy234.  
compute py234 = phy234 - phy235.  
compute py235 = phy235.  
compute py240 = 1- phy241.  
compute py241 = phy241 - phy242.  
compute py242 = phy242 - phy243.  
compute py243 = phy243 - phy244.  
compute py244 = phy244 - phy245.  
compute py245 = phy245.  
compute py250 = 1- phy251.  
compute py251 = phy251 - phy252.  
compute py252 = phy252 - phy253.  
compute py253 = phy253 - phy254.  
compute py254 = phy254 - phy255.  
compute py255 = phy255.  
compute py260 = 1- phy261.  
compute py261 = phy261 - phy262.  
compute py262 = phy262 - phy263.  
compute py263 = phy263 - phy264.  
compute py264 = phy264 - phy265.  
compute py265 = phy265.  
compute py270 = 1- phy271.  
compute py271 = phy271 - phy272.  
compute py272 = phy272 - phy273.  
compute py273 = phy273 - phy274.  
compute py274 = phy274 - phy275.  
compute py275 = phy275.  
compute py280 = 1- phy281.  
compute py281 = phy281 - phy282.  
compute py282 = phy282 - phy283.  
compute py283 = phy283 - phy284.  
compute py284 = phy284 - phy285.  
compute py285 = phy285.  
compute py290 = 1- phy291.  
compute py291 = phy291 - phy292.  
compute py292 = phy292 - phy293.  
compute py293 = phy293 - phy294.  
compute py294 = phy294 - phy295.  
compute py295 = phy295.  
compute py300 = 1- phy301.  
compute py301 = phy301 - phy302.  
compute py302 = phy302 - phy303.  
compute py303 = phy303 - phy304.  
compute py304 = phy304 - phy305.  
compute py305 = phy305.  
execute.
```

ตัวอย่างคำสั่งบนโปรแกรม SPSS Version 7.5 ที่ใช้คำนวณคะแนนจริง

```
compute upy1 = py11 + (2*py12) + (3*py13) + (4*py14) + (5*py15).
compute upy2 = py21 + (2*py22) + (3*py23) + (4*py24) + (5*py25).
compute upy3 = py31 + (2*py32) + (3*py33) + (4*py34) + (5*py35).
compute upy4 = py41 + (2*py42) + (3*py43) + (4*py44) + (5*py45).
compute upy5 = py51 + (2*py52) + (3*py53) + (4*py54) + (5*py55).
compute upy6 = py61 + (2*py62) + (3*py63) + (4*py64) + (5*py65).
compute upy7 = py71 + (2*py72) + (3*py73) + (4*py74) + (5*py75).
compute upy8 = py81 + (2*py82) + (3*py83) + (4*py84) + (5*py85).
compute upy9 = py91 + (2*py92) + (3*py93) + (4*py94) + (5*py95).
compute upy10 = py101 + (2*py102) + (3*py103) + (4*py104) + (5*py105).
compute upy11 = py111 + (2*py112) + (3*py113) + (4*py114) + (5*py115).
compute upy12 = py121 + (2*py122) + (3*py123) + (4*py124) + (5*py125).
compute upy13 = py131 + (2*py132) + (3*py133) + (4*py134) + (5*py135).
compute upy14 = py141 + (2*py142) + (3*py143) + (4*py144) + (5*py145).
compute upy15 = py151 + (2*py152) + (3*py153) + (4*py154) + (5*py155).
compute upy16 = py161 + (2*py162) + (3*py163) + (4*py164) + (5*py165).
compute upy17 = py171 + (2*py172) + (3*py173) + (4*py174) + (5*py175).
compute upy18 = py181 + (2*py182) + (3*py183) + (4*py184) + (5*py185).
compute upy19 = py191 + (2*py192) + (3*py193) + (4*py194) + (5*py195).
compute upy20 = py201 + (2*py202) + (3*py203) + (4*py204) + (5*py205).
compute upy21 = py211 + (2*py212) + (3*py213) + (4*py214) + (5*py215).
compute upy22 = py221 + (2*py222) + (3*py223) + (4*py224) + (5*py225).
compute upy23 = py231 + (2*py232) + (3*py233) + (4*py234) + (5*py235).
compute upy24 = py241 + (2*py242) + (3*py243) + (4*py244) + (5*py245).
compute upy25 = py251 + (2*py252) + (3*py253) + (4*py254) + (5*py255).
compute upy26 = py261 + (2*py262) + (3*py263) + (4*py264) + (5*py265).
compute upy27 = py271 + (2*py272) + (3*py273) + (4*py274) + (5*py275).
compute upy28 = py281 + (2*py282) + (3*py283) + (4*py284) + (5*py285).
compute upy29 = py291 + (2*py292) + (3*py293) + (4*py294) + (5*py295).
compute upy30 = py301 + (2*py302) + (3*py303) + (4*py304) + (5*py305).
execute.
compute truey = upy1 + upy2 + upy3 + upy4 + upy5 + upy6 + upy7 + upy8 + upy9 + upy10 + upy11 + upy12 +
upy13 + upy14 + upy15 + upy16 + upy17 + upy18 + upy19 + upy20 + upy21 + upy22 + upy23 + upy24 +
upy25 + upy26 + upy27 + upy28 + upy29 + upy30.
execute.
```

ตัวอย่างคำสั่งบนโปรแกรม SPSS Version 7.5 ที่ใช้คำนวณค่าความน่าจะเป็นในแต่ละระดับ

คะแนนแต่ละระดับความสามารถตามสูตรที่ 85-86

```
compute psy10 = py10.
compute psy11 = py11.
compute psy12 = py12.
compute psy13 = py13.
compute psy14 = py14.
compute psy15 = py15.
compute psy20 = py20*psy10.
compute psy21 = (py20*psy11) + (py21*psy10).
compute psy22 = (py20*psy12) + (py21*psy11) + (py22*psy10).
compute psy23 = (py20*psy13) + (py21*psy12) + (py22*psy11) + (py23*psy10).
compute psy24 = (py20*psy14) + (py21*psy13) + (py22*psy12) + (py23*psy11) + (py24*psy10).
compute psy25 = (py20*psy15) + (py21*psy14) + (py22*psy13) + (py23*psy12) + (py24*psy11) +
(py25*psy10).
compute psy26 = (py21*psy15) + (py22*psy14) + (py23*psy13) + (py24*psy12) + (py25*psy11).
compute psy27 = (py22*psy15) + (py23*psy14) + (py24*psy13) + (py25*psy12).
compute psy28 = (py23*psy15) + (py24*psy14) + (py25*psy13).
compute psy29 = (py24*psy15) + (py25*psy14).
compute psy210 = (py25*psy15).
compute psy30 = py30*psy20.
compute psy31 = (py30*psy21) + (py31*psy20).
compute psy32 = (py30*psy22) + (py31*psy21) + (py32*psy20).
compute psy33 = (py30*psy23) + (py31*psy22) + (py32*psy21) + (py33*psy20).
compute psy34 = (py30*psy24) + (py31*psy23) + (py32*psy22) + (py33*psy21) + (py34*psy20).
compute psy35 = (py30*psy25) + (py31*psy24) + (py32*psy23) + (py33*psy22) + (py34*psy21) +
(py35*psy20).
compute psy36 = (py30*psy26) + (py31*psy25) + (py32*psy24) + (py33*psy23) + (py34*psy22) +
(py35*psy21).
compute psy37 = (py30*psy27) + (py31*psy26) + (py32*psy25) + (py33*psy24) + (py34*psy23) +
(py35*psy22).
compute psy38 = (py30*psy28) + (py31*psy27) + (py32*psy26) + (py33*psy25) + (py34*psy24) +
(py35*psy23).
compute psy39 = (py30*psy29) + (py31*psy28) + (py32*psy27) + (py33*psy26) + (py34*psy25) +
(py35*psy24).
compute psy310 = (py30*psy210) + (py31*psy29) + (py32*psy28) + (py33*psy27) + (py34*psy26) +
(py35*psy25).
compute psy311 = (py31*psy210) + (py32*psy29) + (py33*psy28) + (py34*psy27) + (py35*psy26).
compute psy312 = (py32*psy210) + (py33*psy29) + (py34*psy28) + (py35*psy27).
compute psy313 = (py33*psy210) + (py34*psy29) + (py35*psy28).
compute psy314 = (py34*psy210) + (py35*psy29).
compute psy315 = (py35*psy210).

:
:

compute psy300 = py300*psy290.
compute psy301 = (py300*psy291) + (py301*psy290).
compute psy302 = (py300*psy292) + (py301*psy291) + (py302*psy290).
compute psy303 = (py300*psy293) + (py301*psy292) + (py302*psy291) + (py303*psy290).
compute psy304 = (py300*psy294) + (py301*psy293) + (py302*psy292) + (py303*psy291) + (py304*psy290).
compute psy305 = (py300*psy295) + (py301*psy294) + (py302*psy293) + (py303*psy292) + (py304*psy291) +
(py305*psy290).
compute psy306 = (py300*psy296) + (py301*psy295) + (py302*psy294) + (py303*psy293) + (py304*psy292) +
(py305*psy291).
compute psy307 = (py300*psy297) + (py301*psy296) + (py302*psy295) + (py303*psy294) + (py304*psy293) +
(py305*psy292).
compute psy308 = (py300*psy298) + (py301*psy297) + (py302*psy296) + (py303*psy295) + (py304*psy294) +
(py305*psy293).
compute psy309 = (py300*psy299) + (py301*psy298) + (py302*psy297) + (py303*psy296) + (py304*psy295) +
(py305*psy294).
compute psy3010 = (py300*psy2910) + (py301*psy299) + (py302*psy298) + (py303*psy297) +
(py304*psy296) + (py305*psy295).
```

[illegible]

[illegible]

[illegible]

[illegible]

```
compute psy30143 = (py300*psy29143) + (py301*psy29142) + (py302*psy29141) + (py303*psy29140) +  
(py304*psy29139) + (py305*psy29138).  
compute psy30144 = (py300*psy29144) + (py301*psy29143) + (py302*psy29142) + (py303*psy29141) +  
(py304*psy29140) + (py305*psy29139).  
compute psy30145 = (py300*psy29145) + (py301*psy29144) + (py302*psy29143) + (py303*psy29142) +  
(py304*psy29141) + (py305*psy29140).  
compute psy30146 = (py301*psy29145) + (py302*psy29144) + (py303*psy29143) + (py304*psy29142) +  
(py305*psy29141).  
compute psy30147 = (py302*psy29145) + (py303*psy29144) + (py304*psy29143) + (py305*psy29142).  
compute psy30148 = (py303*psy29145) + (py304*psy29144) + (py305*psy29143).  
compute psy30149 = (py304*psy29145) + (py305*psy29144).  
compute psy30150 = (py305*psy29145).  
execute.
```

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ-นามสกุล	นางสาวสุนิสา จุยม่วงศรี
สถานที่เกิด	อำเภอพญาไท จังหวัดกรุงเทพมหานคร
สถานที่อยู่ปัจจุบัน	55/380 หมู่บ้านเมืองทองธานี ถนนแจ้งวัฒนะ อำเภอปากเกร็ด จังหวัดนนทบุรี 11120
ตำแหน่งหน้าที่การงานในปัจจุบัน	อาจารย์ประจำสำนักทะเบียนและวัดผล ระดับ 7
สถานที่ทำงานปัจจุบัน	สำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช ถนนแจ้งวัฒนะ ตำบลบางพูด อำเภอปากเกร็ด จังหวัดนนทบุรี 11120
ประวัติการศึกษา	
พ.ศ.2524	ศ.บ. วิชาเอกสถิติ มหาวิทยาลัยธรรมศาสตร์
พ.ศ. 2537	ค.ม. สาขาการวัดและประเมินผลการศึกษา จุฬาลงกรณ์มหาวิทยาลัย
พ.ศ.2546	กศ.ด. สาขาการทดสอบและวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ