

๑๙๖๖/๒๖๖

๙๖๖๖๖

๙. ๓

การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6
ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา
(ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบคะแนนรูปแบบราสส์

ปริณิธาน

ของ

สุจินดา ผ่องอักษร

๒๗ พ.ย. ๒๕๓๔

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาคุษภักดิ์บัณฑิต สาขาการวิจัยและพัฒนาหลักสูตร

ธันวาคม ๒๕๓๓

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

๖๖๖๖๖

174716

คณะกรรมการที่ปรึกษาประจำตัวนิสิต และคณะกรรมการสอบได้พิจารณาประเด็นที่
ฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษาคุณวุฒิบัณฑิต
สาขาการวิจัยและพัฒนาหลักสูตร ของมหาวิทยาลัยศรีนครินทรวิโรฒ

คณะกรรมการที่ปรึกษา

..... ประธาน
(รศ.ดร.สมบูรณ์ ชิตพงศ์)

..... กรรมการ
(ดร.สงวน ลักษณะ)

..... กรรมการ
(ดร.ชูศักดิ์ ชัมภลชิต)

คณะกรรมการสอบ

..... ประธาน
(รศ.ดร.สมบูรณ์ ชิตพงศ์)

..... กรรมการ
(ดร.สงวน ลักษณะ)

..... กรรมการ
(ดร.ชูศักดิ์ ชัมภลชิต)

..... กรรมการที่แต่งตั้งเพิ่มเติม
(ดร.กมล ภูประเสริฐ)

บัณฑิตวิทยาลัยอนุมัติให้รับประเด็นที่ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญา
การศึกษาคุณวุฒิบัณฑิต สาขาการวิจัยและพัฒนาหลักสูตร ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีบัณฑิตวิทยาลัย
(ศ.ดร.สมพร บัวทอง)

วันที่ 31.. เดือน ..พฤษภาคม..... พ.ศ. 2533

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลงได้โดยได้รับความช่วยเหลืออย่างดียิ่งจาก รองศาสตราจารย์
ดร.สมบูรณ์ ชิตพงศ์ ดร.สงบ ลักษณะ และ ดร.ชูศักดิ์ ชัมภลชิต ที่ได้กรุณาให้ความรู้และคำแนะนำ
ในการวิจัย ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ ดร.กมล ภูประเสริฐ ที่ได้กรุณาให้การสนับสนุนด้านการศึกษาแก่ผู้วิจัย
และกรุณาให้คำแนะนำในการทำวิจัยตลอดมา และขอกราบขอบพระคุณ ดร.ภาวดี ศรีสุขวัฒนานนท์
ที่ได้กรุณาให้ความรู้ คำแนะนำตลอดจนเอกสารที่มีประโยชน์ในการทำวิจัยครั้งนี้

ขอขอบพระคุณผู้อำนวยการการประถมศึกษาจังหวัดพุมธานี หัวหน้าการประถมศึกษาอำเภอ
ผู้ช่วยหัวหน้าการประถมศึกษาอำเภอ และขอขอบคุณ คุณสิริ พึ่งจิตต์ตน และศึกษานิเทศก์อำเภอที่ได้
ให้ความอนุเคราะห์ในการให้ข้อมูลผลการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศของ
จังหวัดพุมธานี ตั้งแต่ปีการศึกษา 2529 ถึง 2531

ขอขอบคุณผู้บริหารโรงเรียน และคณะครูอาจารย์ของโรงเรียนในกลุ่มตัวอย่าง ในอำเภอ
ธัญบุรี จังหวัดพุมธานี ที่ให้ความช่วยเหลือในการเก็บรวบรวมข้อมูลครั้งนี้

ขอขอบคุณ คุณอมรเชษฐ์ ผ่องอักษร คุณชาลี สุปารา คุณเจริญ สุปารา คุณละชิด สุปารา
คุณเสรี ปาระแก้ว และคุณศิริธร ปาระแก้ว ตลอดจนเพื่อน ๆ น้อง ๆ ที่ได้ให้ความช่วยเหลือในการ
เก็บรวบรวมข้อมูล วิเคราะห์ข้อมูล และเป็นกำลังใจในการทำปริญญานิพนธ์นี้

สุดท้ายนี้ ผู้วิจัยขอน้อมรำลึกถึงพระคุณของบิดา มารดา ครู อาจารย์ ที่ได้อบรมสั่งสอนให้
การช่วยเหลือ และเป็นกำลังใจแก่ผู้วิจัยตลอดมา

สุจินดา ผ่องอักษร

ปริญญานิพนธ์ฉบับนี้ได้รับทุนอุดหนุนจากทุนศาสตราจารย์ ดร.สาโรช บัวศรี
ซึ่งมอบให้แก่บัณฑิตในการทำวิจัย ผู้วิจัยรู้สึกซาบซึ้งในพระคุณยิ่ง

สารบัญ

บทที่

หน้า

1	บทนำ	1
	ภูมิหลัง	1
	จุดมุ่งหมายของการวิจัย	15
	สมมติฐานของการวิจัย	15
	ความสำคัญของการวิจัย	16
	ขอบเขตของการวิจัย	17
	นิยามศัพท์เฉพาะ	19
2	เอกสารและงานวิจัยที่เกี่ยวข้อง	23
	แนวคิดเชิงทฤษฎีและนิยามของการเทียบคะแนน	23
	แนวคิดเชิงทฤษฎีของรูปแบบราสส์	30
	ข้อตกลงเบื้องต้นของรูปแบบราสส์	33
	ความแกร่งของรูปแบบราสส์	33
	การประมาณค่าพารามิเตอร์ของความสามารถและของข้อสอบ	34
	รูปแบบในการเก็บรวบรวมข้อมูล	36
	การหาค่าคงที่ในการแปลงค่าประมาณพารามิเตอร์	37
	การเทียบค่าประมาณความสามารถของผู้สอบ	40
	การเทียบคะแนนโดยให้คะแนนจริง	42
	การประเมินความเพียงพอของวิธีการเทียบคะแนน	44
	งานวิจัยที่เกี่ยวข้อง	46

3	วิธีดำเนินการวิจัย	68
	ประชากรและกลุ่มตัวอย่าง	68
	เครื่องมือที่ใช้ในการวิจัย	71
	คุณภาพของแบบสอบถาม	71
	การเก็บรวบรวมข้อมูล	72
	การวิเคราะห์ข้อมูล	74
	สรุปแบบแผนในการดำเนินการวิจัย	80
4	ผลการวิเคราะห์ข้อมูล	81
	สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล	81
	การนำเสนอผลการวิเคราะห์ข้อมูล	82
	ค่าสถิติพื้นฐานของคะแนนผลการสอบก่อนทำการเทียบคะแนน	83
	ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าประมาณพารามิเตอร์ก่อนการปรับให้อยู่ ในสเกลเดียวกัน	91
	ค่าสถิติพื้นฐานของคะแนนผลการสอบจากกลุ่มผู้สอบร่วม	93
	ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าประมาณพารามิเตอร์จากกลุ่มผู้สอบร่วม	95
	ค่าคงที่ α และ β และสมการเชิงเส้นตรงในการแปลงค่าประมาณพารามิเตอร์ จากแบบสอบต่างชุด ให้อยู่ในสเกลเดียวกัน	97
	ผลการแปลงค่าประมาณพารามิเตอร์จากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน	99
	ผลการเทียบคะแนนแบบสอบสามชุดในกลุ่มทักษะวิชาภาษาไทย	110
	ผลการเทียบคะแนนแบบสอบสามชุดในกลุ่มทักษะวิชาคณิตศาสตร์	113
	ผลสัมฤทธิ์ทางการเรียนในแต่ละกลุ่มวิชา เมื่อเทียบคะแนนแล้ว	121

ผลสัมฤทธิ์ทางการเรียนและความก้าวหน้าในกลุ่มทักษะวิชาภาษาไทย เมื่อเทียบ คะแนนแล้ว จำแนกเป็นรายอำเภอ	123
ผลสัมฤทธิ์ทางการเรียนและความก้าวหน้าในกลุ่มทักษะวิชาคณิตศาสตร์ เมื่อเทียบ คะแนนแล้ว จำแนกเป็นรายอำเภอ	126
ผลการเปรียบเทียบความแตกต่างระหว่างผลสัมฤทธิ์ทางการเรียนในช่วง 3 ปีการศึกษา	128
ผลการเปรียบเทียบความแตกต่างของสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ระหว่าง การเทียบคะแนนกับการไม่เทียบคะแนน	137
5 สรุปผล อภิปรายผลและข้อเสนอแนะ	143
บรรณานุกรม	161
ภาคผนวก	168
ค่าประมาณพหาวามิตอร์ความยากของข้อสอบ และความสามารถของผู้สอบในแต่ละ กลุ่มวิชาที่ได้จากผลการสอบ เมื่อปีการศึกษา 2529 ถึง 2531	169
ค่าประมาณพหาวามิตอร์ความยากของข้อสอบ และความสามารถของผู้สอบในแต่ละ กลุ่มวิชา ที่ได้จากผลการสอบโดยให้กลุ่มผู้สอบร่วม	187
ภาพแสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบ แบบสอบแต่ละชุดในแต่ละกลุ่มวิชา	301
ตารางเทียบคะแนนจริงในกลุ่มทักษะวิชาภาษาไทย ตั้งแต่ปีการศึกษา 2529 ถึง 2531	307
ตารางเทียบคะแนนจริงในกลุ่มทักษะวิชาคณิตศาสตร์ ตั้งแต่ปีการศึกษา 2529 ถึง 2531	309

บัญชีตาราง

ตาราง	หน้า
1 คะแนนเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจในกลุ่มทักษะ (ภาษาไทย คณิตศาสตร์) ปีการศึกษา 2529 – 2531 และเปรียบเทียบความก้าวหน้าระหว่างปี	10
2 จำนวนโรงเรียนและนักเรียนกลุ่มตัวอย่างในแต่ละปีการศึกษา จำแนกเป็นอำเภอ	70
3 แบบสอบที่ใช้สอบในโรงเรียนกลุ่มตัวอย่างที่ใช้หาค่าคงที่ จำแนกตามกลุ่มวิชาและห้องเรียน	73
4 ค่าสถิติพื้นฐานของคะแนนผลการสอบในแต่ละปีการศึกษา ก่อนการเทียบคะแนนจำแนกเป็นรายกลุ่มวิชา	83
5 ค่าสถิติพื้นฐานของคะแนนผลการสอบ (คะแนนดิบ) ในกลุ่มทักษะวิชาภาษาไทยก่อนทำการเทียบคะแนน จำแนกเป็นรายอำเภอ	85
6 ค่าสถิติพื้นฐานของคะแนนผลการสอบ (คะแนนดิบ) ในกลุ่มทักษะวิชาคณิตศาสตร์ในแต่ละปีการศึกษาก่อนการเทียบคะแนน จำแนกเป็นรายอำเภอ	88
7 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าประมาณพารามิเตอร์ความยากของข้อสอบและค่าประมาณความสามารถของผู้สอบที่สอบแบบสอบในปีการศึกษา 2529 ถึง 2531	91
8 ค่าสถิติพื้นฐานของคะแนนผลการสอบของผู้สอบกลุ่มที่ 1 และ กลุ่มที่ 2	93
9 สรุปค่าประมาณพารามิเตอร์ความยากของข้อสอบ และความสามารถของผู้สอบร่วม	95
10 ค่าคงที่ α และ β และสมการเชิงเส้นตรงในการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน	97
11 ผลการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบร่วมในกลุ่มทักษะวิชาภาษาไทยให้อยู่ในสเกลเดียวกัน	99

12	ผลการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบร่วมในกลุ่มทักษะวิชา คณิตศาสตร์ให้อยู่ในสเกลเดียวกัน	102
13	ผลการใช้ค่าคงที่ และสมการในการแปลงค่าประมาณพารามิเตอร์ความสามารถของ ผู้สอบในกลุ่มทักษะวิชาภาษาไทย เมื่อปีการศึกษา 2529 ถึง 2531 ให้อยู่ใน สเกลเดียวกัน	105
14	ผลการใช้ค่าคงที่และสมการในการแปลงค่าประมาณพารามิเตอร์ความสามารถของ ผู้สอบในกลุ่มทักษะวิชาคณิตศาสตร์ เมื่อปีการศึกษา 2529 ถึง 2531 ให้อยู่ใน สเกลเดียวกัน	107
15	ผลการเทียบคะแนนจริงในกลุ่มทักษะวิชาภาษาไทย ตั้งแต่ปีการศึกษา 2529 ถึง 2531	110
16	ผลการเทียบคะแนนจริงในกลุ่มทักษะวิชาคณิตศาสตร์ ตั้งแต่ปีการศึกษา 2529 ถึง 2531	115
17	ผลการแจกแจงความถี่ของคะแนนที่ได้จากการสอบเมื่อปีการศึกษา 2529 ถึง 2531 และคะแนนเทียบในแต่ละปีการศึกษา ในกลุ่มทักษะวิชาภาษาไทย	116
18	ผลการแจกแจงความถี่ของคะแนนที่ได้จากการสอบเมื่อปีการศึกษา 2529 ถึง 2531 และคะแนนเทียบในแต่ละปีการศึกษา ในกลุ่มทักษะวิชาคณิตศาสตร์	118
19	ค่าเฉลี่ย ค่าเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจใน แต่ละกลุ่มวิชา และในแต่ละปีการศึกษา เมื่อเทียบคะแนนแล้ว	121
20	ค่าเฉลี่ย ค่าเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจในกลุ่มทักษะ วิชาภาษาไทย เมื่อเทียบคะแนนเป็นรายอำเภอ	123
21	ค่าเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจในกลุ่มทักษะวิชาคณิตศาสตร์ เมื่อเทียบคะแนนเป็นรายอำเภอ	126
22	ผลการวิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบความแตกต่างของคะแนนเฉลี่ย ร้อยละระหว่าง 3 ปีการศึกษา ในแต่ละกลุ่มวิชา	129

23	ผลการวิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบความแตกต่างของคะแนนเฉลี่ย ร้อยละระหว่าง 3 ปีการศึกษา ในกลุ่มวิชาภาษาไทย จำแนกเป็นรายอำเภอ	130
24	ผลการวิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบความแตกต่างของคะแนนเฉลี่ย ร้อยละระหว่าง 3 ปีการศึกษา ในกลุ่มวิชาคณิตศาสตร์ จำแนกเป็นรายอำเภอ	132
25	ผลการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ในกลุ่มทักษะ วิชาภาษาไทย	134
26	ผลการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ในกลุ่มทักษะวิชา คณิตศาสตร์	135
27	ผลการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ของแต่ละกลุ่มวิชา จำแนกเป็นรายอำเภอ	136
28	ผลการเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ ระหว่างเมื่อเทียบ คะแนนแล้วกับเมื่อยังไม่ได้เทียบคะแนน จำแนกเป็นรายกลุ่มวิชาและรายอำเภอ	138
29	สัดส่วนจำนวนนักเรียนที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจในแต่ละปีการศึกษา เมื่อ เทียบคะแนนแล้ว และเมื่อไม่เทียบคะแนน	140
30	ค่าเฉลี่ย ค่าเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจใน แต่ละกลุ่มวิชา และในแต่ละปีการศึกษา ระหว่างเมื่อเทียบคะแนนแล้ว และเมื่อ ไม่ได้เทียบคะแนน (ตัวเลขในวงเล็บบรรทัดแรกคือ อัตราความก้าวหน้าในแต่ละปี การศึกษา ตัวเลขในวงเล็บบรรทัดที่สองคืออัตราความก้าวหน้าในฟ่งสามปี การศึกษา)	142
31	ตารางเทียบคะแนนจริงในกลุ่มทักษะวิชาภาษาไทย ตั้งแต่ปีการศึกษา 2529 ถึง 2531	168
32	ตารางเทียบคะแนนจริงในกลุ่มทักษะวิชาคณิตศาสตร์ ตั้งแต่ปีการศึกษา 2529 ถึง 2531	170

บัญชีภาพประกอบ

ภาพประกอบ		หน้า
1	แสดงโครงการตอบข้อสอบ	31
2	แสดงการเทียบคะแนนจริงตามทฤษฎีการตอบข้อสอบ ระหว่างแบบสอบสองชุด ...	43
3	แสดงแบบแผนในการดำเนินการวิจัย	80
4	แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ ในกลุ่มทักษะวิชาภาษาไทย เมื่อปีการศึกษา 2529	301
5	แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ ในกลุ่มทักษะวิชาภาษาไทย เมื่อปีการศึกษา 2530	302
6	แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ ในกลุ่มทักษะวิชาภาษาไทย เมื่อปีการศึกษา 2531	303
7	แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ ในกลุ่มทักษะวิทยาศาสตร์ เมื่อปีการศึกษา 2529	304
8	แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ ในกลุ่มทักษะวิทยาศาสตร์ เมื่อปีการศึกษา 2530	305
9	แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ ในกลุ่มทักษะวิทยาศาสตร์ เมื่อปีการศึกษา 2531	306

บทนำ

ภูมิหลัง

การตรวจสอบคุณภาพการศึกษาโดยการสอบวัดผลสัมฤทธิ์ทางการเรียนของผู้เรียนเป็นวิธีหนึ่งในการติดตามผลของการจัดการศึกษา เพราะข้อมูลที่ได้จากการสอบดังกล่าวจะช่วยให้ครูผู้สอนทราบถึงผลการเรียนและความก้าวหน้าของผู้เรียนโดยสรุป หลังจากที่ได้ผ่านการเรียนการสอนในช่วงระยะเวลาหนึ่งมาแล้ว ซึ่งจะช่วยให้อาจสามารถวินิจฉัยความสามารถทางการเรียนของผู้เรียน และสามารถจัดการเรียนการสอนให้กับผู้เรียนแตกต่างกันไปตามปัญหาที่พบนั้นได้อย่างเหมาะสม (Rudman. 1987 : 74 - 79) นอกจากนี้จะช่วยให้ผู้สร้างและพัฒนาหลักสูตรได้ทราบถึงผลที่ได้รับจากการนำหลักสูตรไปใช้ ซึ่งจะเป็นข้อมูลย้อนกลับสำหรับการพิจารณาปรับปรุงแก้ไขหลักสูตรต่อไป (Worthen and Sanders. 1973 : 112 - 113) ทั้งนี้หากการติดตามผลการจัดการศึกษาดังกล่าวหากกระทำอย่างต่อเนื่องโดยทำการสอบวัดกับนักเรียนในกลุ่มอายุและระดับชั้นต่าง ๆ กัน ตลอดจนการวัดความเปลี่ยนแปลงในทางองศาหรือเสื่อมโทรมของผลการศึกษาในช่วงเวลาต่าง ๆ กันแล้ว จะทำให้ได้ข้อมูลที่มีประโยชน์ต่อการกำหนดนโยบาย และแนวปฏิบัติในการจัดการศึกษาให้มีความเหมาะสมยิ่งขึ้น (สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ. 2528 : 1)

การศึกษาถึงความเปลี่ยนแปลงของผลการศึกษาในช่วงระยะเวลาต่าง ๆ กัน เช่น การศึกษาทุกปีการศึกษาอย่างต่อเนื่องนั้น เมื่อนำคะแนนที่ได้จากการสอบในแต่ละปีการศึกษา มาเปรียบเทียบกันเพื่อให้เห็นแนวโน้มที่เพิ่มขึ้นหรือลดลงของผลการศึกษาในปีการศึกษาต่าง ๆ ที่ผ่านมา ถ้าสามารถใช้แบบสอบชุดเดียวกันดำเนินการสอบในทุกปีการศึกษาแล้ว ก็จะสามารถนำคะแนนดิบที่ได้จากการสอบต่างปีการศึกษานั้นมาเปรียบเทียบกันได้โดยตรง แต่ในทางปฏิบัติ

แบบสอบชุดใดก็ตาม เมื่อได้นำมาใช้สอบไปแล้วครั้งหนึ่งจะไม่สามารถนำกลับมาใช้ให้ได้ผลสมบูรณ์ เช่นเดียวกับการสอบในครั้งแรก ทั้งนี้เพราะสถานการณ์ในการสอบบางประการอาจเปลี่ยนแปลงไป (Klein and Jarjoura. 1985 : 197) และอาจมีปัญหาในเรื่องการรักษาความลับของแบบสอบ ให้สามารถนำกลับมาใช้ได้ตลอด เนื่องจากอาจมีผู้สอบบางคนได้รับคำแนะนำหรือได้รับการฝึกหัดตอบคำถามในแบบสอบที่เคยใช้มาแล้ว ทำให้ผู้ที่สอบในปีต่อ ๆ มาได้รับประโยชน์หรือได้คะแนนจากการสอบมากกว่าผู้สอบในปีก่อน ๆ (Lord. 1980 : 205; Holland and Rubin. 1982 : 2) ดังนั้น การเปรียบเทียบผลการสอบ โดยใช้แบบสอบชุดเดียวกันทุกปีจึงอาจไม่ได้ข้อมูลที่แท้จริง

ด้วยเหตุนี้ผู้สร้างแบบสอบจึงพยายามที่จะสร้างแบบสอบชุดใหม่ขึ้นทุกปี โดยคาดหวังว่าจะได้แบบสอบที่มีความคล้ายคลึงกับแบบสอบชุดเดิมมากที่สุด (Klein and Jarjourns. 1985 : 197) จนสามารถนำคะแนนผลการสอบจากแบบสอบชุดเดิม และชุดใหม่มาเปรียบเทียบกันได้โดยตรงราวกับว่าเป็นแบบสอบชุดเดียวกันหรือสามารถใช้แทนกันได้ (Holland and Rubin. 1982 : 1) นั่นคือ การพยายามสร้างแบบสอบให้คู่ขนานกับชุดเดิมอย่างแท้จริง จึงจะมีความเป็นไปได้ที่จะสามารถเปรียบเทียบคะแนนผลการสอบจากแบบสอบต่างชุดนั้นได้ไม่ว่าผู้สอบจะสอบด้วยแบบสอบชุดใด ก็ไม่มีผลแตกต่างกัน (Lord. 1980 : 198)

สำหรับการสร้างแบบสอบให้คู่ขนานกันอย่างแท้จริงนั้น จะต้องสร้างให้คู่ขนานกันทั้งในด้านโครงสร้าง เวลา ชนิดของข้อสอบ รูปแบบ และเนื้อหาวิชา (Holland and Rubin. 1982 : 10; citing Levine. 1955) ทั้งนี้โดยสร้างแบบสอบชุดใหม่ขึ้นตามตารางกำหนดลักษณะเฉพาะของข้อสอบ ดำเนินการสอบและพิจารณาค่าสถิติของข้อสอบอย่างระมัดระวัง ซึ่งจะต้องเหมือนกับแบบสอบชุดเดิมทุกประการ (Klein and Jarjoura. 1985 : 197) แต่ในทางปฏิบัตินั้น การสร้างแบบสอบให้คู่ขนานกันจนเกิดคะแนนสมมูล (equivalent scores) ในแบบสอบต่างชุดจนสามารถนำคะแนนดิบจากแบบสอบต่างชุดมาเปรียบเทียบกันได้โดยตรงนั้นมีโอกาสเป็นไปได้น้อยมาก จะทำได้ก็เพียงเพื่อให้เกิดความใกล้เคียงกันในด้านเนื้อหา ค่าสถิติของข้อสอบ และของแบบสอบต่างชุดเท่านั้น (Gulliksen. 1958 : 174) กล่าวคือ ถึงแม้ว่าจะสามารถ

สร้างแบบสอบให้คู่ขนานกันในด้านต่าง ๆ ดังกล่าวได้ แต่อย่างน้อยที่สุดแบบสอบจะยังคงมีความแตกต่างกันในความยาก ซึ่งจะทำให้โครงสร้างของแบบสอบทั้งสองชุดนั้นยังแตกต่างกัน (Holland and Rubin. 1982 : 2) เพราะการดำเนินการสอบที่จัดขึ้นต่างเวลากัน ย่อมหมายถึงการเปลี่ยนแปลงสถานการณ์ของการทดสอบ ซึ่งมีผลกระทบต่อ การแปรเปลี่ยนของการแจกแจงคะแนนดิบ จึงนำผลการสอบที่สอบต่างเวลานั้นมาเปรียบเทียบกันโดยตรงไม่ได้ (Flanagan. 1951 : 749 - 750) และถึงแม้ว่าการสร้างแบบสอบคู่ขนานนี้กระทำโดยผู้เชี่ยวชาญในการสร้างแบบสอบ ก็ไม่อาจคาดหวังได้ว่าแบบสอบสองชุดนั้นจะมีความยากเท่ากันอย่างแท้จริง (Dorans and Kingston. 1985 : 250) นอกจากนี้หากบางปีมีการเปลี่ยนแปลงที่จุดเน้นบางประการของการจัดการศึกษา ซึ่งทำให้การสร้างแบบสอบชุดใหม่ จำเป็นต้องเปลี่ยนแปลงเนื้อหาและระดับความยากไปด้วย (Wright and Stone. 1979 : IX) ดังนั้น การสร้างแบบสอบให้คู่ขนานกันอย่างแท้จริงจึงมีข้อจำกัดมาก และทำได้ยาก อาจต้องสูญเสียเวลามากและไม่คุ้มค่าต่อการลงทุน อีกทั้งการเปรียบเทียบคะแนนที่ได้จากการสอบที่สร้างขึ้นให้คู่ขนานกันนั้นก็ยังไม่เพียงพอ ทั้งนี้เพราะแบบสอบต่างชุดยังคงมีความยากหรือง่ายกว่ากัน ซึ่งไม่สามารถใช้แทนกันได้โดยไม่แตกต่างกัน

ต่อมาจึงมีผู้คิดค้นวิธีการใหม่ขึ้น โดยอาศัยวิธีการทางสถิติที่เหมาะสมมาปรับหรือชดเชยในค่าความยากที่แตกต่างกันในแบบสอบต่างชุดนั้น (Holland and Rubin. 1982 : 2) และเมื่อทำการแปลงคะแนนของแบบสอบชุดหนึ่งไปสู่สเกลคะแนนของแบบสอบอีกชุดหนึ่งแล้ว คะแนนจากแบบสอบต่างชุดจะอยู่บนสเกลการวัดเดียวกัน จึงสามารถหาคะแนนสมมูลระหว่างแบบสอบต่างชุดนั้น และนำคะแนนมาเปรียบเทียบกันได้ (Angoff. 1971 : 562; Rentz and Bashaw. 1977 : 163) ทั้งนี้ไม่ว่าแบบสอบต่างชุดนั้นจะคู่ขนานกันหรือไม่ก็ตาม ถ้ามุ่งวัดในสิ่งเดียวกันก็จะสามารถแปลงคะแนนจากแบบสอบต่างชุดให้มีคุณสมบัติที่เทียบเท่ากันได้เสมอ วิธีการดังกล่าวนี้เรียกว่า "การเทียบคะแนนแบบสอบ" (test equating) (Lord. 1977 : 128; Kolen. 1981 : 1 - 2) ซึ่งสามารถเทียบคะแนนได้ทั้งในแนวนอน (horizontal equating) และในแนวตั้ง (vertical equating) การเทียบคะแนนในแนวนอนนั้นเป็นการเทียบระหว่างแบบสอบที่มีความยาก

ใกล้เคียงกัน สอดกับประชากรซึ่งมีความสามารถใกล้เคียงกัน ได้แก่ นักเรียนระดับชั้นเดียวกัน เพื่อเทียบหาว่ากึ่งคะแนนของแบบสอบชุดหนึ่งจะเท่ากับกึ่งคะแนนของแบบสอบอีกชุดหนึ่ง ดังนั้นจึงสามารถเทียบหาคะแนนที่เท่ากันระหว่างแบบสอบต่างชุดได้ ส่วนการเทียบคะแนนในแนวตั้งเป็นการเทียบระหว่างแบบสอบที่มีความยากแตกต่างกัน ซึ่งสอดคล้องกับประชากรที่มีระดับความสามารถต่างกัน ได้แก่ นักเรียนต่างระดับชั้น เป็นต้น เพื่อเทียบหาคะแนนตามระดับความสามารถของผู้สอบ (สงขล. 2525 : 23) อย่างไรก็ตามการเทียบคะแนนไม่ว่าจะเป็นแนวตั้งหรือแนวนอนเป็นกระบวนการเชิงประจักษ์ (empirical procedure) ซึ่งเกี่ยวข้องกับรูปแบบในการเก็บรวบรวมข้อมูล นิยามในการแปลงคะแนน และวิธีการทางสถิติที่เหมาะสมสำหรับการเทียบคะแนนแบบสอบ (Dorans and Kingston. 1985 : 250; Brennen. 1987 : 221 - 222)

รูปแบบในการเก็บรวบรวมข้อมูลจัดแบ่งได้เป็น 3 แบบ คือ แบบกลุ่มผู้สอบกลุ่มเดียว แบบกลุ่มผู้สอบที่เท่าเทียมกัน และแบบที่ใช้แบบสอบร่วม (Doran and Kingston. 1985 : 198; Angoff. 1971 : 569 - 586)

สำหรับแบบแรกกลุ่มผู้สอบกลุ่มเดียวทำแบบสอบทั้งสองชุด การเทียบคะแนนทำได้โดยการปรับเฉพาะความแตกต่างของค่าเฉลี่ยความยากระหว่างแบบสอบทั้งสองชุด ในลักษณะของการเพิ่มค่าคงที่ค่าหนึ่งเพื่อให้คะแนนจากแบบสอบต่างชุดนั้นมีจุดเริ่มต้นที่เดียวกัน และเทียบคะแนนกันได้ (Rentz and Bashaw. 1977 : 163) ส่วนแบบที่สองกลุ่มผู้สอบที่เท่าเทียมกันสองกลุ่ม ซึ่งอาจได้มาจากการสุ่ม แต่ละกลุ่มสอบแบบสอบเพียงหนึ่งชุด รูปแบบนี้หลีกเลี่ยงผลของความเหนื่อยล้าในการทำแบบสอบทั้งสองชุด และผลของการสอบแบบสอบชุดใดก่อนหรือหลังแต่จะมีความลำเอียงในการเทียบคะแนนอันเนื่องมาจากอาจมีความแตกต่างในการแจกแจงความสามารถของกลุ่มผู้สอบ (Hambleton. 1985 : 199) สำหรับแบบที่สามนั้นใช้แบบสอบร่วม โดยกลุ่มผู้สอบกลุ่มหนึ่งทำแบบสอบชุดหนึ่ง ส่วนอีกกลุ่มทำแบบสอบอีกชุดหนึ่ง และทั้งสองกลุ่มต้องทำแบบสอบร่วมเหมือนกัน ซึ่งแบบสอบร่วมจะประกอบด้วยข้อสอบที่มีเนื้อหา และระดับความยากใกล้เคียงกับแบบสอบสองชุดที่จะทำการเทียบคะแนนนั้นให้มากที่สุด ประสิทธิภาพของแบบสอบร่วมขึ้นอยู่กับความสัมพันธ์ระหว่าง

แบบสอบร่วมกับแบบสอบทั้งสองชุดนั้น ซึ่งจะต้องมีความสัมพันธ์กันมากพอที่จะเทียบคะแนนอย่างได้ผล โดยอาจมีจำนวนข้อน้อยกว่าแบบสอบที่เทียบคะแนนแต่ต้องสามารถแทน (represent) แบบสอบสองชุดนั้นได้ ทั้งนี้อาจสอบรวมหรือแยกสอบต่างเวลากับแบบสอบสองชุดก็ได้ (Angoff. 1971 : 576) แบบสอบร่วมใช้เป็นตัวเชื่อมข้อมูลจากแบบสอบสองชุดเข้าด้วยกัน โดยจะทำหน้าที่ลดความลำเอียงในการเทียบคะแนนซึ่งเนื่องมาจากความแตกต่างในความสามารถระหว่างกลุ่มผู้สอบที่อาจไม่เท่าเทียมกัน หรือลดความแปรปรวนในการสุ่มตัวอย่างให้น้อยลง ผลการเทียบคะแนนจึงมีความถูกต้อง (precise) มากยิ่งขึ้น (Petersen and others. 1982 : 73; Angoff. 1971 : 577) ดังนั้นรูปแบบนี้จึงใช้เทียบคะแนนกลุ่มผู้สอบที่มีความสามารถต่างกันได้ โดยใช้ความแตกต่างของค่าความยากโดยเฉลี่ยจากแบบสอบร่วม ซึ่งสอบโดยผู้สอบทั้งสองกลุ่มนั้นเป็นตัวปรับคะแนนแบบสอบสองชุดให้อยู่ในสเกลเดียวกันและสามารถเทียบคะแนนกันได้ (Rentz and Bashaw. 1977 : 162 - 163 ; Dorans and Kingston. 1985 : 250)

ส่วนวิธีการในการเทียบคะแนนนั้นได้พัฒนามาหลายวิธีด้วยกัน โดยอาศัยแนวคิดทางทฤษฎีการสอบ การกำหนดคณิยาม การออกแบบรวบรวมข้อมูล ตลอดจนเทคนิคในการวิเคราะห์ทางสถิติที่แตกต่างกันไป แต่วิธีเทียบคะแนนที่นิยมใช้มี 3 แบบ คือ (Petersen and others. 1982 : 73)

1. การเทียบคะแนนแบบเอควิเปอร์เซนไทล์ (equipercentile equating) อาศัยทฤษฎีการสอบแบบดั้งเดิม (classical theory) ทำการเทียบคะแนนดิบระหว่างแบบสอบต่างชุดที่ตำแหน่งเปอร์เซนไทล์เดียวกัน (Angoff. 1971 : 563; citing Flanagan. 1951; Lord. 1950)

2. การเทียบคะแนนเชิงเส้นตรง (linear equating) อาศัยทฤษฎีการทดสอบแบบดั้งเดิมที่ได้พัฒนามาพร้อม ๆ กับวิธีแรก โดยทำการเทียบคะแนนดิบระหว่างแบบสอบต่างชุดที่ตรงกับคะแนนมาตรฐาน (Z-score) ค่าเดียวกัน

3. การเทียบคะแนนโดยใช้โค้งลักษณะข้อสอบ (item characteristic curve) หรือการเทียบคะแนนโดยอาศัยทฤษฎีการตอบข้อสอบ (item response theory : IRT) ซึ่งเทียบคะแนนโดยใช้คะแนนจริง (true score) ที่ระดับค่าประมาณความสามารถแท้ (latent trait) เดียวกันระหว่างแบบสอบต่างชุดนั้น (Lord. 1980 : 199 - 201)

สำหรับการศึกษาในเชิงประเมินว่าวิธีการเทียบคะแนนวิธีใดจะให้ผลการเทียบคะแนนที่ถูกต้องหรือมีประสิทธิภาพมากกว่ากันนั้น อาจกระทำได้ 2 แบบ คือ ประเมินก่อนการเทียบคะแนน โดยพิจารณาจากการแจกแจงคะแนนที่ได้จากการสอบด้วยแบบสอบต่างชุดว่าเป็นไปตามนิยามของการกำหนดคะแนนสมมูลของวิธีใดยอมแสดงว่าวิธีการนั้นเป็นวิธีที่เหมาะสมมากกว่าและควรเลือกใช้ ในการเทียบคะแนนซึ่งจากแนวคิดเชิงทฤษฎีและนิยามของการเทียบคะแนนพอสรุปได้ว่า การเทียบคะแนนเชิงเส้นตรงมีความเหมาะสมในการแปลงคะแนนอย่างมาก เมื่อการแจกแจงคะแนนของแบบสอบต่างชุดที่นำมาเทียบคะแนนนั้นมีรูปร่างเหมือนกัน (Angoff. 1971 : 568; Braun and Holland. 1982 : 23) แต่ถ้ารูปร่างของการแจกแจงคะแนนของแบบสอบต่างชุดนั้นแตกต่างกันแล้ว การเทียบคะแนนแบบอิกวิเปอร์เซนไทล์จะให้ผลดีกว่าการเทียบคะแนนเชิงเส้นตรง (Angoff. 1971 : 564) อย่างไรก็ตามหากรูปร่างของการแจกแจงคะแนนจากแบบสอบต่างชุดคล้ายคลึงกันมากควรใช้การเทียบคะแนนเชิงเส้นตรงซึ่งจะให้ผลดีกว่าการเทียบแบบอิกวิเปอร์เซนไทล์เพราะเป็นปรนัยมากกว่า และไม่มีความคลาดเคลื่อนในการปรับหรือเกลาเส้นกราฟให้เรียบตามวิธีการเทียบคะแนนแบบอิกวิเปอร์เซนไทล์ ซึ่งจะมีความคลาดเคลื่อนมากในช่วงที่มีข้อมูลน้อย (Angoff. 1971 : 564; Potthoff. 1982 : 209) ส่วนการเทียบคะแนนโดยอาศัยทฤษฎีการตอบข้อสอบ นั้นมีคุณสมบัติของความไม่แปรเปลี่ยนค่าประมาณของประชากรข้อสอบเมื่อเปลี่ยนกลุ่มผู้สอบ ดังนั้นจึงไม่มีนิยามเกี่ยวกับการแจกแจงคะแนนที่ได้จากแบบสอบ วิธีนี้จึงให้ผลดี แม้ว่าในกรณีที่การแจกแจงคะแนนจากแบบสอบต่างชุดนี้แตกต่างกัน (Kolen and Whitney. 1982 : 279 - 293; Cook, Dunbar and Eignor. 1981 : 1 - 52)

ส่วนการประเมินอีกแบบหนึ่ง คำเนิการหลังจากการเทียบคะแนนแล้ว โดยเปรียบเทียบคะแนนที่ได้จากการเทียบคะแนนด้วยวิธีการต่าง ๆ กับคะแนนเกณฑ์ คะแนนเกณฑ์ที่ใช้อาจได้มาจากการเทียบคะแนนด้วยวิธีการอื่นที่เชื่อถือได้ หรือได้จากกลุ่มตัวอย่างอื่นที่เท่าเทียมกัน นอกจากนี้อาจได้จากการเทียบคะแนนกับตัวเองโดยผ่านการเทียบกับแบบสอบชุดอื่น ๆ อย่างต่อเนื่องกันในลักษณะสายโซ่ก็ได้ ซึ่งต้องมีแบบสอบมากกว่า 2 ชุด และหากพบว่าวิธีการใดมีผลการเทียบใกล้เคียงกับเกณฑ์มากที่สุด ย่อมแสดงว่ามีความคลาดเคลื่อนในการเทียบคะแนนน้อยกว่า นั่นคือวิธีการนั้นเป็นวิธีที่มีประสิทธิภาพมากกว่า (Petersen and others. 1982 : 72; Kolen and Whitney. 1982 : 284; Jaeger. 1981 : 27 - 28)

อย่างไรก็ตาม งานวิจัยเท่าที่ผ่านมา ยังไม่อาจสรุปได้ว่าวิธีการเทียบคะแนนวิธีใดมีประสิทธิภาพมากที่สุดในทุกสถานการณ์ เพราะผลการเทียบคะแนนขึ้นอยู่กับองค์ประกอบหลายประการ ได้แก่ รูปแบบในการเก็บรวบรวมข้อมูล ลักษณะของกลุ่มตัวอย่าง และลักษณะของแบบสอบ เป็นต้น (Petersen and others. 1982 : 71 - 135; Kolen and Whitney. 1982 : 279 - 293) แต่ด้วยการเลือกใช้วิธีการในการเทียบคะแนนที่เหมาะสมแล้วจะมีความเป็นไปได้ที่จะวัดการเปลี่ยนแปลงหรือความก้าวหน้าของผู้เรียนเมื่อผ่านการเรียนการสอนในช่วงระยะเวลาต่าง ๆ เพื่อจัดทำแผนภูมิแสดงแนวโน้มและเพื่อรวบรวมข้อมูลโดยการสร้างตารางคะแนนที่เทียบเท่าระหว่างแบบสอบต่างชุดและสอบต่างเวลากันได้ (สงบ ลักษณะ. 2525 : 22; Angoff. 1971 : 562 - 563)

สำหรับการนำแนวคิดและเทคนิควิธีการของการเทียบคะแนนไปใช้นั้น ในต่างประเทศได้มีการศึกษากันอย่างกว้างขวาง ซึ่งได้มีบทความและงานวิจัยหลายชิ้นที่ศึกษาในแง่มุมต่าง ๆ กัน ได้แก่ การเปรียบเทียบผลการเทียบคะแนนด้วยวิธีการและตัวแปรต่าง ๆ กัน (Kolen. 1981 : 1 - 11; Kolen and Whitney. 1982 : 279 - 293; Lord and Wingersky. 1984 : 453 - 461; Cope. 1987 : 143 - 149) การศึกษาความคลาดเคลื่อนในการเทียบคะแนน (Lord. 1982 : 165 - 173; Angoff. 1971 : 569 - 586; citing Lord. 1950 ;

Braun and Holland . 1982 : 33 - 39) การประเมินความเพียงพอของวิธีการเทียบคะแนนแบบต่าง ๆ (Slinde and Linn . 1978 : 23 - 35; Jaeger . 1981 : 23 - 38 ; Kolen and Whitney . 1982 : 284 ; Petersen , Marco and Stewart . 1982 : 91) การศึกษาความแกร่ง (robustness) ต่อการฝ่าฝืนข้อตกลงเบื้องต้นของวิธีการเทียบคะแนนวิธีต่าง ๆ (Dorans and Kingston . 1985 : 249 - 262 ; Skaggs and Lissitz . 1986 : 303 - 317) ตลอดจนการสร้างสเกลอ้างอิงสำหรับการเทียบคะแนนที่ได้จากแบบสอบต่างชุด เป็นต้น (Rentz and Bashaw . 1977 : 161 - 179)

ส่วนในประเทศไทยนั้นพบว่า การนำแนวคิดและวิธีการในการเทียบคะแนนแบบสอบไปใช้ยังมีน้อยมาก เท่าที่ผ่านมามีหน่วยงานต่าง ๆ ตลอดจนงานวิจัยจำนวนมากที่ศึกษาผลสัมฤทธิ์ทางการเรียนของนักเรียนในระดับชั้น วิชาหรือกลุ่มประสบการณ์ต่าง ๆ ซึ่งโดยทั่วไปทำการวิเคราะห์ผลจากค่าคะแนนเฉลี่ย และเป็นที่เข้าใจกันว่าเกณฑ์ขั้นต่ำที่น่าพอใจของผลการสอบคือ นักเรียนต้องมีค่าคะแนนเฉลี่ยตั้งแต่ร้อยละ 50 ขึ้นไปของคะแนนเต็ม และการเปรียบเทียบคะแนนผลการสอบที่ได้จากแบบสอบต่างชุดก็คือ การเปรียบเทียบคะแนนเฉลี่ยหรือคะแนนดิบจากแบบสอบต่างชุดโดยตรง

สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ (สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ . 2531 : 88 - 89) ได้ทำการประเมินคุณภาพของนักเรียนนั้นประถมศึกษาปีที่ 6 ในสังกัดอย่างเป็นระบบและต่อเนื่องมาตั้งแต่ปีการศึกษา 2527 และต่อมาได้ทำการประเมินในระดับชั้นประถมศึกษาปีที่ 2 และ 4 ด้วย นอกจากนี้สำนักงานการประถมศึกษาจังหวัดทุกจังหวัดก็ได้ดำเนินการประเมินคุณภาพนักเรียนในสังกัดภายในจังหวัดของตนเอง โดยมีแนวทางในการดำเนินงานเช่นเดียวกับระดับประเทศ จุดมุ่งหมายที่สำคัญของการประเมิน คือ เพื่อให้ได้ข้อมูลเกี่ยวกับคุณภาพทางการศึกษาว่าผลผลิตของการจัดการศึกษาตามหลักสูตรประถมศึกษาพุทธศักราช 2521 ซึ่งพิจารณาจากผลสัมฤทธิ์ทางการเรียนของนักเรียนที่เรียนจบตามหลักสูตรในปีการศึกษาต่าง ๆ นั้นมีคุณภาพและมีแนวโน้มที่เพิ่มขึ้นหรือลดลงเพียงใด สำหรับเครื่องมือที่ใช้ในการประเมิน คือ แบบสอบซึ่งสร้างขึ้นใหม่ทุกปีการศึกษา ตามขอบเขตสมรรถภาพที่ต้องการวัดในแต่ละกลุ่มประสบการณ์ ทั้งทางด้าน

ความรู้ ความคิด ทักษะการปฏิบัติ ตลอดจนคุณลักษณะต่าง ๆ ของผู้เรียน ซึ่งสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติได้กำหนดสมรรถภาพที่ต้องวัดไว้เหมือนกันทุกปี เพื่อให้เป็นแนวทางในการประเมินในปีการศึกษาต่อไป และเพื่อให้สามารถเปรียบเทียบผลระหว่างปีการศึกษาต่าง ๆ ได้ โดยการสร้างแบบสอบชุดใหม่ก็คือ การสร้างข้อคำถามใหม่ในลักษณะของการสร้างแบบสอบที่คู่ขนานกันกับแบบสอบชุดที่สร้างขึ้นในปีก่อน ๆ ทุกชุด ทั้งในด้านโครงสร้างและเนื้อหา ตามขอบเขตของสมรรถภาพที่กำหนดไว้ นั้น โดยคาดหวังว่าจะได้แบบสอบที่มีความคล้ายคลึงกับแบบสอบชุดเดิมมากที่สุด และจะดำเนินการสอบให้เป็นมาตรฐานเดียวกันทุกปีการศึกษา

ในการวิเคราะห์คะแนนผลการสอบ ได้แยกพิจารณาเป็น 2 ส่วน คือ

1. หาค่าคะแนนเฉลี่ยร้อยละและร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละกลุ่มประสบการณ์ทุกปีการศึกษา โดยใช้เกณฑ์ที่น่าพอใจของผลการสอบที่ร้อยละ 50 ของคะแนนเต็มเป็นเกณฑ์ขั้นต่ำในการตัดสินผลการสอบของนักเรียนในแต่ละจังหวัดและส่วนรวมทั้งประเทศ
2. เปรียบเทียบความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนระหว่างปีการศึกษาต่าง ๆ โดยพิจารณาจากอัตราเพิ่มและลดของค่าคะแนนเฉลี่ยร้อยละและร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละกลุ่มประสบการณ์ในช่วงปีการศึกษาต่าง ๆ ได้แก่ ช่วง 1 ปี, 2 ปี และ 3 ปีการศึกษาที่ผ่านมา เพื่อจัดทำข้อมูลแสดงให้เห็นแนวโน้มของผลการศึกษาของแต่ละจังหวัดและส่วนรวมทั้งประเทศ

หลังจากนั้นจึงนำมารายงานผลการประเมินแสดงความก้าวหน้าในระหว่างปีการศึกษาต่าง ๆ ตัวอย่างรายงานผลการประเมินในกลุ่มทักษะ (วิชาคณิตศาสตร์และภาษาไทย) ตามโครงการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศ ดังตาราง 1

ตาราง 1 คะแนนเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจ ในกลุ่มทักษะ (ภาษาไทยและคณิตศาสตร์) ปีการศึกษา 2527 - 2531 และเปรียบเทียบความก้าวหน้าระหว่างปี (เลขบรรทัดล่างคือความก้าวหน้าเปรียบเทียบแต่ละปี) (สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ. 2531 : 1)

กลุ่มประสบการณ์	คะแนนเฉลี่ยร้อยละ					ร้อยละนักเรียนที่มีผลน่าพอใจ				
	2527	2528	2529	2530	2531	2527	2528	2529	2530	2531
ทักษะภาษาไทย	49.08	56.84	58.40	56.29	60.78	47	69	76	64	79
	-	7.76	1.56	-2.11	4.49	-	22	7	-10	15
ทักษะคณิตศาสตร์	33.11	36.52	47.81	46.16	44.80	11	18	41	40	33
	-	3.41	11.29	-1.65	-1.36	-	7	23	-1	-7

ผลการประเมินดังกล่าวนี้ใช้เป็นแนวทางในการวางแผนเพื่อพัฒนาคุณภาพทั้งในระดับประเทศ จังหวัด และอำเภอต่าง ๆ รวมทั้งใช้เป็นข้อมูลในการติดตามผลการดำเนินงานเร่งรัดคุณภาพของการประถมศึกษาด้วย (สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ. 2530 : 2) ดังนั้นผลการประเมินจึงมีความสำคัญยิ่งต่อการตัดสินใจของผู้ที่เกี่ยวข้องกับการจัดการศึกษาทุกระดับ

เนื่องจากผลการประเมินดังกล่าวได้มาจากการวิเคราะห์คะแนนผลการสอบซึ่งใช้แบบสอบต่างชุดกันทุกปี แม้ว่าแบบสอบชุดใหม่จะประกอบด้วยข้อสอบที่สร้างขึ้นใหม่ให้มีลักษณะต่าง ๆ ใกล้เคียงมากที่สุดกับแบบสอบที่สร้างขึ้นชุดก่อน ๆ ก็ตาม แต่ไม่สามารถยืนยันได้ว่าค่าสถิติของข้อสอบและของแบบสอบที่ได้จากแบบสอบต่างชุดเหล่านั้นจะมีความเท่าเทียมกันและสามารถใช้แทนกันได้โดย

ไม่แตกต่างกัน ซึ่งจากแนวคิดของการเทียบคะแนนดังกล่าวมาข้างต้นนั้น จะเห็นว่าการเปรียบเทียบคะแนนผลการสอบที่ได้จากแบบสอบต่างชุด เพื่อศึกษาความก้าวหน้าระหว่างปีการศึกษาต่าง ๆ ดังกล่าว ยังมีจุดอ่อนเกี่ยวกับความแตกต่างของแบบสอบแต่ละปี จึงจำเป็นต้องปรับความแตกต่างระหว่างแบบสอบเสียก่อนด้วยวิธีการเทียบคะแนน เพื่อให้สามารถรายงานคะแนนผลการสอบในปีการศึกษาต่าง ๆ ที่ผ่านมาได้อย่างเที่ยงตรง ซึ่งการประเมินคุณภาพดังกล่าวยังไม่ได้นำวิธีการเทียบคะแนนแบบสอบมาใช้ ดังนั้น การนำผลการประเมินไปใช้ต่อไปอาจเกิดความผิดพลาดขึ้นได้ ดังการศึกษาของเยาวดี รวงชัยกุล (Rangchaikul. 1975 : 87 - 88) ซึ่งทำการเทียบคะแนนผลการสอบที่กระทรวงศึกษาธิการได้จัดดำเนินการสอบให้กับนักเรียนในระบบโรงเรียน เมื่อเดือนมีนาคม พ.ศ. 2516 และ พ.ศ. 2517 และการสอบที่จัดให้กับบุคคลทั่วไปที่อยู่นอกระบบโรงเรียนเมื่อเดือนสิงหาคม พ.ศ. 2516 และ พ.ศ. 2517 เพื่อตัดสินว่านักเรียนสมควรได้รับการรับรองการผ่านชั้นมัธยมศึกษาปีที่ 5 (ม.ศ.5) หรือไม่ โดยใช้แบบสอบต่างกัน 4 ชุด ซึ่งไม่มีแบบสอบร่วม ทำการเทียบคะแนนแต่ละชุดไปสู่สเกลร่วมด้วยวิธีการเชิงเส้นตรง ผลการเทียบคะแนนเมื่อใช้คะแนนจุดตัดการผ่านที่ร้อยละ 50 ของคะแนนเต็มเหมือนกันทุกชุดพบว่า หลังจากการเทียบคะแนนแล้ว แบบสอบ 4 ชุดมีจุดตัด (cut-off point) สำหรับการผ่านแตกต่างกันในสเกลร่วมเดียวกัน ตัวอย่างเช่น ในการสอบเดือนมีนาคม พ.ศ. 2517 พบว่าคะแนนจุดตัดสำหรับการผ่านในสเกลร่วมเท่ากับ 88 แต่การสอบเมื่อเดือนมีนาคม พ.ศ. 2516 พบว่าคะแนนจุดตัดสำหรับการผ่านในสเกลร่วมเท่ากับ 78 และภายหลังจากการตรวจสอบสัดส่วนของผู้สอบโดยการตัดสินด้วยเกณฑ์ร้อยละ 50 เหมือนกัน พบว่านักเรียนที่ควรสอบได้แต่ได้รับการตัดสินให้ตกมีจำนวนน่าสนใจ ตัวอย่างเช่นการเทียบคะแนนโดยใช้แบบสอบเมื่อเดือนมีนาคม พ.ศ. 2516 และ มีนาคม พ.ศ. 2517 พบว่ามีจำนวนนักเรียนร้อยละ 37.6 ที่ตกเมื่อเดือนสิงหาคม พ.ศ. 2516 และร้อยละ 9 ที่ตกเมื่อเดือนมีนาคม พ.ศ. 2517 ซึ่งควรเป็นผู้สอบได้

จากที่กล่าวมา ผู้วิจัยมีความสนใจที่จะนำแนวคิดและวิธีการในการเทียบคะแนนแบบสอบมาใช้ปรับปรุงความเที่ยงตรงในการรายงานคะแนนผลการสอบ โดยจะทำการศึกษาความก้าวหน้า

ของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตร
 ประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษาที่ผ่านมา จากคะแนนผลทดสอบ
 ด้วยแบบสอบต่างชุดที่สร้างขึ้นตามโครงการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับ
 ประเทศ ซึ่งได้ดำเนินการสอบนักเรียนกลุ่มตัวอย่างไปแล้วเมื่อปีการศึกษา 2529, 2530 และ
 2531 และในปัจจุบันการรายงานคะแนนผลทดสอบยังไม่ได้ทำการเทียบคะแนนดังกล่าว ผู้วิจัยจะนำ
 วิธีการเทียบคะแนนโดยใช้ฟังก์ชันลักษณะข้อสอบรูปแบบราสช์ (Rasch model หรือ one parameter
 logistic model) มาใช้ศึกษาความก้าวหน้า ทั้งนี้โดยมีเหตุผลดังนี้

1. วิธีเทียบคะแนนแบบอิกวิเปอร์เซนไทล์มีข้อจำกัดที่ต้องคำนึงถึงอยู่หลายประการ ได้แก่

1.1 วิธีนี้ใช้การเขียนเส้นกราฟเพื่อหาคะแนนสมมูลซึ่งจะต้องปรับหรือเกลาเส้นโค้ง
 ให้เรียบ ดังนั้นประสบการณ์และการตัดสินใจเกลาเส้นให้เรียบด้วยข้อมูลเหล่านั้น อาจนำไปสู่ความ
 คลาดเคลื่อนอย่างมาก การเทียบคะแนนด้วยวิธีนี้จึงเป็นการเทียบไปยังจำนวนเต็มทีใกล้เคียงที่สุด

(Angoff. 1971 : 571 - 572)

1.2 รูปแบบอิกวิเปอร์เซนไทล์ มีความไวต่อความแปรปรวนเชิงสุ่มมาก โดยเฉพาะ
 กลุ่มตัวอย่างที่มีขนาดเล็ก (Angoff. 1971 : 571; Potthoff. 1982 : 210)

1.3 ในกรณีที่แบบสอบสองชุดมีความเชื่อมั่นต่างกันมาก ผลของการเทียบคะแนน
 จะขาดความคงที่ (Potthoff. 1982 : 210)

1.4 การเทียบคะแนนสมมูลของแบบสอบสองชุดทำได้เฉพาะในช่วงพิสัยของคะแนน
 ที่มีความถี่ของคะแนนสังเกตเพียงพอ ส่วนช่วงที่มีความถี่ของคะแนนน้อยจะมีความคลาดเคลื่อนสูงมาก
 (Angoff. 1971 : 571)

2. วิธีเทียบคะแนนเชิงเส้นตรงมีข้อจำกัดที่ต้องคำนึงถึงที่สำคัญ คือ

2.1 ต้องมีความเป็นคู่ขนานกันของแบบสอบซึ่งกระทำไ้ยาก จึงมีปัญหว่า ความ
 เป็นคู่ขนานกันโดยประมาณ (approximate) ของแบบสอบควรจะเป็นเท่าไรตามที่ต้องการเพื่อให้มี
 ความเพียงพอในวิธีการเทียบเชิงเส้นตรงนี้ (Jaeger. 1981 : 24)

2.2 หากมีความแตกต่างในรูปร่างของการแจกแจงคะแนนดิบจากแบบทดสอบชุดที่จะทำการเทียบคะแนนแล้ว จะกำหนดได้อย่างไรว่าความแตกต่างนั้นมีเพียงเล็กน้อยจนไม่ต้องการนึ่งได้ (Jaeger . 1981 : 24)

2.3 ค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยากและค่าอำนาจจำแนก จะมีค่าไม่คงที่ แต่จะเปลี่ยนไปตามความสามารถของกลุ่มผู้สอบเช่นเดียวกับวิธีเทียบคะแนนแบบอควิเปอร์เซนไทล์ ดังนั้น การเทียบคะแนนด้วยวิธีการเชิงเส้นตรงหรืออควิเปอร์เซนไทล์จากกลุ่มผู้สอบที่มีความสามารถแตกต่างกันก็จะให้ผลไม่คงที่โดยแปรผันไปตามกลุ่มผู้สอบที่นำมาใช้เทียบคะแนน (Hambleton . 1978 : 498 - 499)

3. คุณสมบัติของการเทียบคะแนนตามทฤษฎีการตอบข้อสอบจะให้ผลการเทียบคะแนน 3 ประการ คือ (Lord . 1980 : 199)

3.1 ความเสมอภาค (equity) หมายถึง สำหรับทุก ๆ ความสามารถ θ การแจกแจงความถี่แบบมีเงื่อนไขของคะแนนแปลง $X(Y)$ หรือ Y^* ที่ θ ที่กำหนดให้ต้องเหมือนกับ การแจกแจงความถี่แบบมีเงื่อนไขของคะแนน X

3.2 ความไม่แปรผันไปตามกลุ่มตัวอย่าง (invariance across groups) หมายถึง คะแนนแปลง $X(Y)$ จะคงที่ โดยไม่ขึ้นกับตัวแปรด้านอื่น ๆ ของประชากรที่นำมาเทียบคะแนน

3.3 ความสมมาตร (symetry) หมายถึง การเทียบคะแนนจะเหมือนกันไม่ว่าการเทียบนั้นจะเป็นการเทียบจากแบบสอบ X ไปยังแบบสอบ Y หรือจากแบบสอบ Y ไปยังแบบสอบ X ก็ตาม

ทั้งนี้ การเทียบคะแนนตามทฤษฎีข้างต้นไม่มีคุณสมบัติเช่นนี้ (Hambleton . 1980 : 498 - 499) ดังนั้น ผู้วิจัยจึงเลือกใช้วิธีเทียบคะแนนตามทฤษฎีการตอบข้อสอบ และในทั้งสามรูปแบบของการเทียบคะแนนตามทฤษฎีการตอบข้อสอบนี้ รูปแบบราสช์ซับซ้อนน้อยกว่ารูปแบบสองและสามพารามิเตอร์จึงเข้าใจได้ง่าย และเสียเวลาในการคำนวณน้อยกว่า (Vijver . 1986 : 45)

แนวความคิดของรูปแบบราสช์ เชื่อว่าโอกาสที่ผู้สอบจะทำข้อสอบได้ถูกต้องหรือไม่ ขึ้นอยู่กับผลต่างของค่าพารามิเตอร์เพียง 2 ค่าเท่านั้น คือ ระดับความสามารถของตนเอง (ability parameter : θ) และระดับความยากของข้อสอบ (difficulty parameter : b) (Hambleton and Cook. 1978 : 475) ถ้าหากความสามารถของผู้สอบน้อยกว่าระดับความยากของข้อสอบแล้ว โอกาสที่ผู้สอบจะทำข้อสอบได้ถูกต้องย่อมจะน้อยกว่า .50 และในทำนองเดียวกัน ถ้าหากความสามารถของผู้สอบมากกว่าระดับความยากของข้อสอบแล้ว โอกาสที่ผู้สอบจะทำข้อสอบได้ถูกต้องก็มีมากกว่า .50 (Wright and Stone. 1979 : 12) ทั้งนี้โดยมีข้อตกลงว่าข้อสอบที่จะนำมาใช้จะต้องมีค่าอำนาจจำแนกเท่ากัน และมีลักษณะที่จะทำให้เกิดการเดาได้น้อยที่สุด แต่ในทางปฏิบัติลักษณะทั้งสองนี้เป็นความแกร่งของรูปแบบซึ่งข้อสอบที่ไม่มีลักษณะดังกล่าวนี้ ก็อาจใช้การวิเคราะห์ตามรูปแบบนี้ได้ด้วย (สุพัตน์ สุกมลสันต์. 2530 : 32) จากงานวิจัยเพื่อศึกษาความเหมาะสมของข้อมูลจริงโดยใช้รูปแบบราสช์และรูปแบบสามพารามิเตอร์พบว่ามีความเหมาะสมกับข้อมูลประมาณร้อยละ 80 ทั้งสองรูปแบบ (Hutten. 1982 : 4799-A) ทั้งนี้เพราะรูปแบบที่มีพารามิเตอร์หลายตัวก็ไม่สามารถที่จะประมาณค่าพารามิเตอร์ได้ถูกต้องทุกตัวโดยเฉพาะในกรณีของกลุ่มตัวอย่างมีขนาดเล็ก หรือในกรณีที่กลุ่มตัวอย่างมีความสามารถสูง ความถูกต้องของค่าพารามิเตอร์การเดาก็จะต่ำ ดังนั้นหากรูปแบบราสช์เหมาะสมกับข้อมูลแล้ว การใช้รูปแบบที่ต่างกันจะทำให้เกิดผลในทางปฏิบัติอย่างมาก (Hulin. 1982; Thissen and Wainer. 1982; Traub and Lam. 1985 : 26; citing Vijver. 1986 : 45; Hambleton. 1985 : 195) นอกจากนี้รูปแบบราสช์มีความแกร่งแม้ในกรณีที่กลุ่มตัวอย่างมีขนาดเล็ก ซึ่งการแจกแจงความสามารถของกลุ่มตัวอย่างหรือความยากของข้อสอบไม่จำเป็นต้องเป็นโค้งปกติ เพียงแต่มีแนวโน้มจะเป็นโค้งปกติเท่านั้น และกลุ่มตัวอย่างไม่จำเป็นต้องมาจากการสุ่มก็ได้ (Wright and Stone. 1979 : 20) โดยขนาดของกลุ่มตัวอย่างอาจมีเพียง 100 คนก็ใช้ได้ (Wright. 1977 : 219) ในขณะที่ยังการประมาณค่าตามรูปแบบสามพารามิเตอร์ควรใช้กลุ่มตัวอย่างที่มีขนาดตั้งแต่ 1,000 คนขึ้นไป เพื่อให้ได้ค่าพารามิเตอร์คงที่ หรือมีความแม่นยำในการประมาณค่าพารามิเตอร์ต่าง ๆ (Douglass. 1981 : 4000-A; Kolen and Whitney. 1982 : 279, 293)

จุดมุ่งหมายของการวิจัย

1. เพื่อศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบคะแนนรูปแบบราสช์
2. เพื่อเปรียบเทียบความแตกต่างของผลสัมฤทธิ์ทางการเรียนระหว่างปีการศึกษาต่าง ๆ หลังจากที่ได้เทียบคะแนนแล้ว
3. เพื่อเปรียบเทียบสัดส่วนจำนวนนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนผ่านเกณฑ์ที่น่าพอใจระหว่างการเทียบคะแนนกับการไม่เทียบคะแนน

สมมุติฐานของการวิจัย

จากจุดมุ่งหมายของการวิจัยที่กำหนดไว้ ผู้วิจัยต้องการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนของนักเรียนในช่วง 3 ปีการศึกษาเมื่อเทียบคะแนนแล้ว และเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจระหว่างเมื่อเทียบคะแนนแล้วกับเมื่อยังไม่ได้เทียบคะแนน

สำหรับการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในช่วง 3 ปีการศึกษานั้น ผู้วิจัยจะทำการเปรียบเทียบความแตกต่างของค่าคะแนนเฉลี่ยร้อยละระหว่างปีการศึกษาต่าง ๆ เมื่อเทียบคะแนนแล้วว่ามีอัตราที่เพิ่มขึ้นหรือลดลงจากปีการศึกษาหรือไม่ เพียงใด ซึ่งจากผลการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศที่ผ่านมา สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติได้ใช้เป็นแนวทางในการวางแผนเพื่อพัฒนาคุณภาพระดับประเทศ จังหวัด และอำเภอต่าง ๆ รวมทั้งใช้เป็นข้อมูลในการติดตามผลการดำเนินงานเร่งรัดคุณภาพการประถมศึกษาอีกด้วย (สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ. 2530 : 2) ดังนั้นผู้วิจัยจึงกำหนดสมมุติฐานข้อที่ 1 ดังนี้

สมมติฐานข้อที่ 1 ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษาในแต่ละปีการศึกษามีความก้าวหน้าขึ้นตามลำดับ

ส่วนการศึกษาสักส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจระหว่างเมื่อเทียบคะแนนแล้วกับเมื่อยังไม่ได้เทียบคะแนนนั้น จากทฤษฎีและผลการวิจัยในอดีต ผู้วิจัยจึงกำหนดสมมติฐานข้อที่ 2 ดังนี้

สมมติฐานข้อที่ 2 สักส่วนจำนวนนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนผ่านเกณฑ์ที่น่าพอใจระหว่างการเทียบคะแนนกับการไม่เทียบคะแนนแตกต่างกัน

ความสำคัญของการวิจัย

1. ผลการวิจัยจะช่วยปรับปรุงความเที่ยงตรงในการรายงานคุณภาพทางการศึกษาในช่วงระยะเวลา 3 ปีการศึกษาว่า ผลของการจัดการศึกษาตามหลักสูตร หรือผลผลิตที่ได้จากการนำหลักสูตรไปใช้ ซึ่งพิจารณาจากผลสัมฤทธิ์ทางการเรียนของนักเรียนในแต่ละกลุ่มวิชาในกลุ่มทักษะ (คณิตศาสตร์ และภาษาไทย) นั้น ปรากฏผลอยู่ในระดับใด และมีความก้าวหน้าขึ้นกว่าเดิมหรือลดลงจากเดิมเพียงใด ทั้งนี้ผลการวิจัยนี้จะลดความคลาดเคลื่อนอันเนื่องมาจากความแตกต่างในความยากระหว่างแบบสอบต่างชุด ซึ่งเป็นปัญหาในการเปรียบเทียบคะแนนผลการสอบ และการตัดสินใจผ่านเกณฑ์ที่น่าพอใจดังกล่าว โดยหลังจากการเทียบคะแนนแล้วคะแนนแต่ละตัวจะมีคุณสมบัติที่จะเปรียบเทียบกันได้ และการกำหนดจุดตัดการผ่านจะแม่นยำมากยิ่งขึ้น ดังนั้นผู้ที่เกี่ยวข้องกับการจัดการศึกษาในระดับต่าง ๆ สามารถจะนำไปเป็นข้อมูล เพื่อใช้เป็นแนวทางในการวางแผนพัฒนาคุณภาพการศึกษา ระดับจังหวัด อำเภอ และโรงเรียน รวมทั้งใช้เป็นข้อมูลในการติดตามผลการดำเนินงานเพื่อเร่งรัดคุณภาพการประถมศึกษาในส่วนที่พบว่าผลยังไม่ถึงเกณฑ์ที่น่าพอใจต่อไป

2. ผลการวิจัยจะเป็นแนวทางในการพัฒนาการสร้างแบบสอบ การวัดผลและการประเมินผลให้มีประสิทธิภาพมากยิ่งขึ้นต่อไป

ขอบเขตของการวิจัย

1. ข้อมูลที่ใช้ในการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นประถมศึกษาปีที่ 6 ครั้งนี้ คือ ผลการตอบข้อสอบรายข้อรายบุคคลตามโครงการประเมินคุณภาพของนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศ ที่ได้ดำเนินการสอบนักเรียนกลุ่มตัวอย่างทุกอำเภอ และทุกจังหวัดทั่วประเทศ โดยสุ่มนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 แต่ละรุ่นให้ได้จำนวนนักเรียนในแต่ละอำเภอตามสัดส่วนของโรงเรียนขนาดต่าง ๆ สำหรับในการศึกษาค้างนี้ผู้วิจัยจะศึกษากับกลุ่มตัวอย่างในจังหวัดปทุมธานี เฉพาะอำเภอที่มีข้อมูลผลการตอบข้อสอบรายข้อรายบุคคลครบทั้ง 3 ปีการศึกษา ตั้งแต่ปีการศึกษา 2529 ถึง 2531 ซึ่งมีจำนวน 4 อำเภอ จากอำเภอทั้งหมด 7 อำเภอ จำนวนโรงเรียนกลุ่มตัวอย่างปีการศึกษาละ 20 ถึง 21 โรงเรียน ซึ่งมีจำนวนนักเรียนในแต่ละปีการศึกษา 523 คน, 453 คน และ 467 คน ตามลำดับ

2. แบบสอบที่ใช้เทียบคะแนน คือ แบบสอบที่สร้างขึ้นตามโครงการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศ ในกลุ่มทักษะ คือ กลุ่มวิชาคณิตศาสตร์และกลุ่มวิชาภาษาไทย กลุ่มวิชาละ 3 ชุด ที่ใช้สอบในปีการศึกษา 2529 ถึง 2531 ปีการศึกษาละ 1 ชุด ผู้วิจัยนำมาศึกษาเฉพาะภาคความรู้ (cognitive domain) สำหรับกลุ่มทักษะวิชาคณิตศาสตร์ประกอบด้วยข้อสอบแบบเลือกตอบชนิด 4 ตัวเลือก ซึ่งมีข้อสอบชุดละ 38 ข้อ แต่ละชุดวัด 3 สมรรถภาพ คือ ความรู้ ความเข้าใจทางคณิตศาสตร์ จำนวน 12 ข้อ ทักษะการคิดคำนวณจำนวน 10 ข้อ และทักษะการแก้ปัญหาโจทย์จำนวน 16 ข้อ ส่วนกลุ่มทักษะวิชาภาษาไทยประกอบด้วยข้อสอบแบบเลือกตอบชนิด 4 ตัวเลือก ซึ่งมีข้อสอบชุดละ 30 ข้อ แต่ละชุดวัด 3 สมรรถภาพ คือ ความเข้าใจในการฟังจำนวน 8 ข้อ ความเข้าใจในการอ่านจำนวน 15 ข้อ และความรู้เกี่ยวกับหลักเกณฑ์ทางภาษาจำนวน 7 ข้อ โดยแต่ละชุดไม่มีข้อสอบร่วมหรือไม่มีข้อสอบข้อใดซ้ำกันเลยทั้งสองกลุ่มวิชา

3. คุณภาพของแบบสอบ แบบสอบเน้นความเที่ยงตรงเชิงเนื้อหา กล่าวคือแบบสอบวัดได้ตรงตามสมรรถภาพที่กำหนด โดยคณะผู้สร้างแบบสอบจะสร้างข้อสอบขึ้นตามขอบเขตของสมรรถภาพ

ที่ต้องการวัดที่สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติได้กำหนดไว้เหมือนกันทุกปีการศึกษา โดยคาดหวังว่าแบบสอบชุดใหม่จะต้องมีคุณลักษณะใกล้เคียงกับแบบสอบชุดเดิมมากที่สุด และการดำเนินการสอบให้เป็นมาตรฐานเดียวกันทุกปีการศึกษา

4. วิธีการที่ใช้ในการเทียบคะแนนครั้งนี้ คือ วิธีเทียบคะแนนโดยใช้โด่งลักษณะข้อสอบ หรือวิธีเทียบคะแนนตามทฤษฎีการตอบข้อสอบโดยใช้รูปแบบราสส์ หรือรูปแบบหนึ่งพารามิเตอร์โลจิสติก ซึ่งจะทำให้การเทียบคะแนนในแนวนอน

5. สถานการณ์ในการสอบตามโครงการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศนั้น ใช้แบบสอบต่างชุดกันทุกปีการศึกษา ซึ่งสอบกับนักเรียนต่างกลุ่มหรือต่างปีการศึกษา โดยไม่มีข้อสอบร่วม และไม่มีผู้สอบร่วม ดังนั้นค่าประมาณพารามิเตอร์จากแบบสอบแต่ละชุดจะไม่อยู่ในสเกลเดียวกัน แต่การเชื่อมค่าพารามิเตอร์ที่ประมาณค่าได้จากแบบสอบต่างชุดนั้น จะต้องได้มาจากรูปแบบที่เป็นที่ยอมรับกันโดยทั่วไป 3 รูปแบบ คือ รูปแบบที่ใช้ผู้สอบร่วม (กลุ่มเดียวทำแบบสอบ 2 ชุด) รูปแบบที่ใช้กลุ่มที่เท่าเทียมกัน และรูปแบบที่ใช้ข้อสอบร่วม (Hambleton. 1985 : 198; Lord. 1982 : 145) ทั้งนี้เพื่อกำหนดค่าคงที่ในการแปลงค่าประมาณพารามิเตอร์จากแต่ละชุดให้อยู่ในสเกลเดียวกัน และสามารถเทียบคะแนนระหว่างแบบสอบต่างชุดได้ ด้วยเหตุนี้ผู้วิจัยจึงศึกษาค่าคงที่ดังกล่าวโดยใช้กลุ่มตัวอย่างอื่น โดยจะนำแบบสอบทั้ง 3 ชุด คือ แบบสอบที่ใช้ในปีการศึกษา 2529 ถึง 2531 ทั้งสองกลุ่มวิชาไปทำการสอบกับนักเรียนกลุ่มตัวอย่างชั้นประถมศึกษาปีที่ 6 ที่กำลังเรียนภาคปลาย ปีการศึกษา 2532 โดยทำการสอบระหว่างวันที่ 25 มกราคมถึง 8 ถึง กุมภาพันธ์ 2533 ทั้งนี้จะศึกษารูปแบบที่ใช้ผู้สอบร่วม กล่าวคือผู้สอบแต่ละคนจะต้องสอบแบบสอบ 2 ชุด คือ ชุดที่ใช้ในปีการศึกษา 2529 หนึ่งชุด และชุดที่ใช้ในปีการศึกษา 2530 หรือ 2531 อีกหนึ่งชุด โดยใช้กลุ่มตัวอย่างในอำเภอธัญบุรี จังหวัดปทุมธานี จำนวน 4 โรงเรียน ซึ่งมีนักเรียนชั้นประถมศึกษาปีที่ 6 อย่างน้อย 2 ห้องเรียน จำนวนนักเรียนรวมทั้งสิ้น 581 คน ในจำนวนนี้เป็นนักเรียนที่สอบแบบสอบชุดที่ใช้ในปีการศึกษา 2529 กับ 2530 จำนวน 224 คน และสอบชุดที่ใช้ในปีการศึกษา 2529 กับ 2531 จำนวน 197 คน

6. ตัวแปรที่ใช้ในการศึกษารั้งนี้ มีดังนี้

6.1 ตัวแปรต้น ได้แก่

6.1.1 แบบสอบในกลุ่มทักษะ (วิชาคณิตศาสตร์ และวิชาภาษาไทย)

กลุ่มวิชาละ 3 ชุด

6.1.2 การเทียบคะแนนรูปแบบราชย์ และการไม่เทียบคะแนน

6.2 ตัวแปรตาม ได้แก่

6.2.1 ผลสัมฤทธิ์ทางการเรียนของนักเรียนในแต่ละกลุ่มวิชาและแต่ละปีการศึกษา เมื่อได้ทำการเทียบคะแนนแล้ว

6.2.2 ผลสัมฤทธิ์ทางการเรียนของนักเรียนรายอำเภอในแต่ละกลุ่มวิชาและในแต่ละปีการศึกษา เมื่อได้ทำการเทียบคะแนนแล้ว

6.2.3 ความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในช่วง 3 ปีการศึกษาของแต่ละอำเภอ และรวมทุกอำเภอ

6.2.4 สัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจเมื่อเทียบคะแนนแล้ว

6.2.5 สัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจเมื่อไม่เทียบคะแนน

นิยามศัพท์เฉพาะ

1. การเทียบคะแนนแบบสอบ หมายถึง กระบวนการในการปรับคะแนนที่ได้จากแบบสอบต่างชุดให้เป็นคะแนนที่มีคุณสมบัติที่สามารถนำมาเปรียบเทียบกันได้

2. ค่าคงที่ของการเทียบคะแนน หมายถึง ค่าความชัน และจุดตัดแกนของสมการเชิงเส้นตรง ซึ่งใช้ในสมการการแปลงค่าประมาณพารามิเตอร์ที่ได้จากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน

3. ความแกร่ง หมายถึง คุณสมบัติของผลการวิเคราะห์ด้วยราชย์โมเดลที่ผ่านได้โดยไม่เกิดผลเสียต่อโมเดล

4. ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน หมายถึง ค่าประมาณความคลาดเคลื่อนจากการสุ่มตัวอย่าง (sampling error) ของวิธีการเทียบคะแนนแต่ละวิธี ค่านี้เป็นดัชนีสำหรับเปรียบเทียบความแม่นยำของการเทียบคะแนนภายในแต่ละวิธีของการเทียบคะแนน

5. สมรรถภาพ หมายถึง ความสามารถที่ต้องการวัดในแต่ละกลุ่มวิชาหรือกลุ่มประสบการณ์ตามหลักสูตร ซึ่งสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติได้กำหนดขอบเขตของแต่ละสมรรถภาพไว้เหมือนกันทุกปี เพื่อเป็นแนวทางในการประเมินผลการศึกษา และการเปรียบเทียบผลการศึกษาระหว่างปีการศึกษาต่าง ๆ

6. ความสามารถของผู้สอบ คือ ระดับความสามารถที่คำนวณจากคะแนนจริง แต่ปรับหน่วยให้เป็นคะแนนมาตรฐาน (สงบ ลักษณะ. 2525 : 28) ในทางปฏิบัติมีค่าอยู่ระหว่าง -3 ถึง +3 โดยค่า -3 แสดงว่ามีความสามารถต่ำ และค่า +3 แสดงว่ามีความสามารถสูง

7. ความไม่แปรผันตามกลุ่มตัวอย่าง คือ ค่าพารามิเตอร์ต่าง ๆ จะมีค่าแน่นอน แม้ว่า จะนำข้อสอบชุดเดิมไปใช้กับคนต่างกลุ่มกันจนปรากฏค่าพารามิเตอร์ต่าง ๆ เปลี่ยนไป ก็เป็นการเปลี่ยนเพราะอิทธิพลของการมีจุดตั้งต้นของความสามารถที่บนแกนราบต่างกันไปเท่านั้น ถ้ามีการปรับค่าความสามารถทำให้มีสเกลเดียวกันแล้วค่าพารามิเตอร์เหล่านี้จากการสอบต่างกลุ่มจะคงเดิมเสมอ

8. ดัชนีความแตกต่าง (discrepancy index) หมายถึง ค่าความคลาดเคลื่อนรวม (total error : TE) ในการเทียบคะแนนของวิธีการเทียบคะแนนซึ่งคำนวณได้จากค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างคะแนนที่แปลงแล้ว ด้วยวิธีการเทียบคะแนนที่ใช้กับคะแนนเกณฑ์ และถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ ดังนั้นจึงเป็นค่ามาตรฐานซึ่งสามารถเปรียบเทียบกันได้โดยตรงระหว่างค่าที่ได้จากการใช้วิธีการเทียบคะแนนที่ต่างกัน เพื่อประเมินประสิทธิภาพหรือความพอเพียงในการเทียบคะแนนของวิธีการเหล่านั้น

9. คะแนนเกณฑ์ หมายถึง คะแนนที่ได้มาจากวิธีการเทียบคะแนนที่เชื่อถือได้หรือได้จากกลุ่มตัวอย่างอื่นที่สุ่มจากประชากรเดียวกันซึ่งใช้เป็นกลุ่มตรวจสอบผล (cross validation

sample) หรือได้จากคะแนนดิบของแบบสอบชุดแรกก่อนการแปลงคะแนนซึ่งใช้ในการเทียบคะแนนแบบสายโซ่

10. การเทียบคะแนนแบบสายโซ่ หมายถึง การเทียบคะแนนแบบสอบมากกว่า 2 ชุด โดยเทียบครั้งละ 2 ชุดต่อเนื่องกันในลักษณะคล้ายสายโซ่ ซึ่งหากเทียบกลับมาสู่แบบสอบชุดแรกแล้ว คะแนนดิบจากแบบสอบชุดแรกจะใช้เป็นคะแนนเกณฑ์ สำหรับเปรียบเทียบความแตกต่างกับคะแนนที่แปลงแล้วในครั้งสุดท้าย

11. กลุ่มตรวจสอบผล หมายถึง กลุ่มตัวอย่างที่สุ่มมาจากประชากรเดียวกันกับกลุ่มตัวอย่างที่ได้ข้อมูลมาแปลงคะแนน โดยไม่มีหน่วยตัวอย่างที่ซ้ำกันเลย

12. แบบสอบร่วม หมายถึง แบบสอบที่มีข้อสอบที่มีเนื้อหาโครงสร้างและความยากเหมือนกับแบบสอบที่นำมาเทียบคะแนนทั้งสองชุด โดยอาจมีจำนวนข้อน้อยกว่าแบบสอบที่นำมาเทียบคะแนนนั้นได้

13. กลุ่มผู้สอบร่วม หมายถึง กลุ่มผู้สอบกลุ่มเดียวสอบแบบสอบมากกว่า 1 ชุด

14. ความพอเพียงของวิธีการเทียบคะแนน (adequacy of equating) หมายถึง ระดับคุณภาพของวิธีการเทียบคะแนนที่ใช้วิธีของปีเตอร์เซน มาร์โค และสจวร์ต (Petersen Marco and Stewart. 1982 : 71 - 135)

$TE \leq (.05 S_x)^2$ หมายถึง น่าพอใจอย่างมาก

$(.05 S_x)^2 < TE \leq (.10 S_x)^2$ หมายถึง น่าพอใจ

$(.10 S_x)^2 < TE \leq (.15 S_x)^2$ หมายถึง ปานกลาง

$(.15 S_x)^2 < TE \leq (.20 S_x)^2$ หมายถึง ไม่น่าพอใจ

$(.20 S_x)^2 < TE$ หมายถึง ไม่น่าพอใจ

15. ตารางเทียบคะแนนแบบสอบ หมายถึง ตารางแสดงตัวเลขคะแนนที่สมมูลกัน (equivalent scores) ระหว่างแบบสอบต่างชุด

16. คะแนนสมมูล หมายถึง คะแนนระหว่างแบบสอบต่างชุด ซึ่งเทียบคะแนนกันได้ว่า เป็นคะแนนที่เท่าเทียมกัน

17. ความก้าวหน้าของผลการศึกษา หมายถึง ความเปลี่ยนแปลงของผลการศึกษา ซึ่งพิจารณาจากจำนวนที่เพิ่มขึ้นหรือลดลงของค่าเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจ

18. ร้อยละนักเรียนที่มีผลน่าพอใจ หมายถึง จำนวนนักเรียนที่มีคะแนนผลการสอบตั้งแต่ร้อยละ 50 ขึ้นไปของคะแนนเต็ม ทั้งนี้เมื่อคิดจากจำนวนนักเรียน 100 คน

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในบทนี้ ผู้วิจัยได้เสนอแนวคิดและเทคนิควิธีในการเทียบคะแนนแบบสอบที่นำมาใช้ในการปรับคะแนนที่ได้จากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน เพื่อให้สามารถเปรียบเทียบคะแนนและศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนของนักเรียนที่เรียนจบหลักสูตรในช่วง 3 ปีการศึกษาที่ผ่านมาได้ ตามลำดับหัวข้อดังนี้

1. แนวคิดเชิงทฤษฎีและนิยามของการเทียบคะแนน
2. แนวคิดเชิงทฤษฎีของรูปแบบราสช์
3. การประมาณค่าพารามิเตอร์และการหาค่าคงที่ในการเทียบคะแนน
4. การเทียบค่าประมาณความสามารถของผู้สอบ
5. การเทียบคะแนนโดยใช้คะแนนจริง
6. การประเมินความเพียงพอของวิธีการในการเทียบคะแนน
7. งานวิจัยที่เกี่ยวข้อง

แนวคิดเชิงทฤษฎีและนิยามของการเทียบคะแนน

กิลลิคเสน (Gulliksen. 1950 : 298 – 304) ได้ให้ความหมายของการเทียบคะแนนไว้ว่า คือ วิธีการทำคะแนนจากแบบสอบสองชุดของวิชาเดียวกันให้เป็นคะแนนสมมูลที่เปรียบเทียบกันได้โดยตรง โดยเสนอวิธีการให้ผู้สอบกลุ่มเดียวทำแบบสอบสองชุด และใช้วิธีง่าย ๆ คือ แปลงคะแนนแต่ละชุดให้เป็นคะแนนมาตรฐาน แล้วนำคะแนนที่แปลงแล้วมาเทียบกันโดยตรง แต่ก่อนอื่นให้ตรวจสอบความเป็นคู่ขนานของแบบสอบโดยใช้สถิติของวิลส์ (Wilks) หรือพิจารณาจาก

ค่าสหสัมพันธ์ระหว่างแบบสอบแต่ละชุดกับเกณฑ์ ถ้ามีค่าเท่ากันโดยประมาณ ($R_{yc} = R_{xc}$) ก็จะสามารถบอกถึงความเพียงพอในการเทียบคะแนนแบบสอบทั้งสองชุด

สำหรับฟลานาแกน (Flanagan. 1951 : 747 - 748) ได้ให้นิยามการเทียบคะแนนว่าเป็นวิธีการในการหาคะแนนจากแบบสอบต่างชุดให้สามารถนำมาเปรียบเทียบกันได้คำว่า "ความสามารถในการเปรียบเทียบกันได้" มีความหมายเฉพาะที่ว่า เมื่อกำหนดประชากรให้ ถ้าการแจกแจงของคะแนนจริงจากแบบสอบทั้งสองชุดซึ่งสอบกับกลุ่มตัวอย่างที่เลือกมาขนาดใหญ่ใด ๆ มีลักษณะเหมือนกันแล้ว คะแนนดิบจากแบบสอบทั้งสองชุดจึงจะสามารถเปรียบเทียบกันได้ หรือถ้าความเชื่อมั่น (reliability) ของแบบสอบทั้งสองชุดเท่ากันในประชากรนั้นแล้ว ก็สามารถเปรียบเทียบการแจกแจงของค่าที่ได้เช่นกัน จากความหมายดังกล่าวเป็นนิยามเชิงทฤษฎี ในทางปฏิบัติได้นิยามไว้ว่า ในประชากรที่กำหนด ถ้าคะแนนจากแบบสอบสองชุดมีค่าเฉลี่ยเท่ากันหรือเกือบเท่ากันในทุก ๆ กลุ่มตัวอย่างขนาดใหญ่ใด ๆ แล้ว การเปรียบเทียบจะทำได้ อย่างไรก็ตามการจะทำให้เกิดลักษณะตามที่นิยามเป็นเรื่องที่ยากมาก สิ่งสำคัญคือแบบสอบต้องเป็นคู่ขนานกัน ฟลานาแกนจึงแนะนำว่า ควรเริ่มต้นตั้งแต่การสร้างแบบสอบให้มีคุณสมบัติของความเป็นคู่ขนานกัน แล้วเลือกวิธีการเทียบคะแนนที่เหมาะสมซึ่งเสนอไว้ 4 วิธี คือ

1. ใช้ค่าเฉลี่ย โดยคำนวณค่าเฉลี่ยของการแจกแจงคะแนนทั้งสองชุด ถ้าความแตกต่างของค่าเฉลี่ยอยู่ภายในขอบเขตของการแปรผันเชิงสุ่มแล้ว ถือว่าคะแนนทั้งสองชุดนั้นเปรียบเทียบกันได้ แต่ถ้าความแตกต่างมีนัยสำคัญให้ใช้วิธีบวกเข้าหรือลบออกเท่าจำนวนที่แตกต่างจากคะแนนชุดที่หนึ่ง เพื่อให้เกิดคะแนนสมมูลกับอีกชุดหนึ่ง
2. ให้เทคนิคของสมการถดถอยโดยหาค่าประมาณที่ดีที่สุดของคะแนนจากแบบสอบชุดหนึ่งซึ่งรู้ค่าของอีกชุดหนึ่ง
3. ใช้คะแนนมาตรฐาน ซึ่งเป็นวิธีปรับคะแนนอย่างคงที่ตลอดการแจกแจง
4. ใช้วิธีอีควิเปอร์เซนไทล์ ซึ่งเป็นวิธีหาค่าคะแนนโดยการเปรียบเทียบตามสัดส่วนของการแจกแจงคะแนน

ส่วนแองกอฟฟ์ (Angoff. 1971 : 562) ได้ให้ความหมายของการเทียบคะแนนว่าเป็นการแปลงระบบหน่วยการวัดของแบบสอบชุดหนึ่งไปสู่ระบบหน่วยการวัดของแบบสอบอีกชุดหนึ่ง ซึ่งหลังจากการแปลงแล้วจะสามารถเทียบหาคะแนนสมมูลระหว่างแบบสอบต่างชุดได้ แองกอฟฟ์ได้เสนอวิธีการในการเทียบคะแนนไว้ 2 แบบ คือ การเทียบคะแนนแบบอิกวิเปอร์เซนไทล์และการเทียบคะแนนเชิงเส้นตรง

สำหรับนิยามของการเทียบคะแนนแบบอิกวิเปอร์เซนไทล์ ลอร์ด และฟลานาแกน (Angoff. 1971 : 563; citing Lord. 1957 and Flanagan. 1951) ได้กล่าวไว้ว่า คะแนนสองคะแนน คะแนนตัวหนึ่งมาจากแบบสอบชุด X และอีกคะแนนมาจากแบบสอบชุด Y เมื่อ X และ Y วัดในสิ่งเดียวกันด้วยระดับความเชื่อมั่นเท่ากันจะสมมูลกัน ถ้าต่างก็เป็นค่าที่ตรงกับตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน จากนิยามนี้หากแบบสอบสองชุดมีความยากต่างกันแล้ว รูปร่างของการแจกแจงคะแนนดิบจากแบบสอบสองชุดจะแตกต่างกัน การเทียบคะแนนเชิงเส้นโค้ง (curvilinear) หรือการเทียบแบบอิกวิเปอร์เซนไทล์จะทำการย่อและขยายสเกลของคะแนนดิบ ในแบบสอบชุดหนึ่งเพื่อให้มีการแจกแจงคะแนนเหมือนกับอีกชุดหนึ่ง และสามารถเทียบคะแนนที่สมมูลระหว่างแบบสอบได้ (Angoff. 1971 : 564)

วิธีเทียบคะแนนแบบอิกวิเปอร์เซนไทล์นี้มีข้อจำกัดที่ต้องคำนึงถึงอยู่หลายประการ ได้แก่

1. วิธีนี้ใช้การเขียนเส้นกราฟเพื่อหาคะแนนสมมูลซึ่งจะต้องปรับหรือเกลากล้นโค้งให้เรียบ ดังนั้นประสบการณ์และการตัดสินใจเกลากล้นให้เรียบด้วยข้อมูลเหล่านั้น อาจนำไปสู่ความคลาดเคลื่อนอย่างมาก การเทียบคะแนนด้วยวิธีนี้จึงเป็นการเทียบไปยังจำนวนเต็มทีใกล้เคียงที่สุด (Angoff. 1971 : 571 - 572)

2. รูปแบบอิกวิเปอร์เซนไทล์ มีความไวต่อความแปรปรวนเชิงสูงมาก โดยเฉพาะกลุ่มตัวอย่างที่มีขนาดเล็ก (Angoff. 1971 : 571; Potthoff. 1982 : 210)

3. ในกรณีที่แบบสอบสองชุดมีค่าความเชื่อมั่นต่างกันมาก ผลของการเทียบคะแนนจะขาดความคงที่ (Potthoff. 1982 : 210)

4. การเทียบคะแนนสมมูลของแบบทดสอบสองชุดทำได้เฉพาะในช่วงพิสัยของคะแนนที่มีความถี่ของคะแนนสังเกตเพียงพอ ส่วนช่วงที่มีความถี่ของคะแนนน้อยจะมีความคลาดเคลื่อนสูงมาก

การเทียบคะแนนสมมูลระหว่างแบบสอบอีกวิธีหนึ่ง คือ วิธีเทียบคะแนนเชิงเส้นตรง วิธีนี้เป็นวิธีที่ง่ายในขั้นตอนของการแปลงคะแนนโดยมีข้อตกลงว่าแบบทดสอบชุดที่คู่ขนานกันจะมีการแจกแจงคะแนนดิบเป็นอย่างเดียวกัน ยกเว้นสำหรับความแตกต่างในค่าเฉลี่ยและความแปรปรวน (Lord and Novick. 1968 : 48) นั่นคือนอกจากความแตกต่างในสองโมเมนต์แรกแล้วโมเมนต์มาตรฐานของการแจกแจงคะแนนดิบของแบบทดสอบสองชุดของกลุ่มผู้สอบที่กำหนดให้จะต้องเหมือนกันหมด (Angoff. 1971 : 568) จากข้อตกลงนี้จึงมีความเป็นไปได้ที่จะแปลงคะแนนจากแบบทดสอบชุดหนึ่งไปยังสเกลของคะแนนมาตรฐาน โดยคะแนนของชุด Y จะแปลงในสเกลของชุด X ตามสมการการแปลงเชิงเส้นตรงในสมการ

$$\frac{Y - M_y}{S_y} = \frac{X - M_x}{S_x}$$

เมื่อ M_x และ M_y คือ ค่าเฉลี่ยของการแจกแจงคะแนนดิบจากแบบทดสอบชุด X และชุด Y ตามลำดับ

S_x และ S_y คือ ค่าเบี่ยงเบนมาตรฐานของการแจกแจงคะแนนชุด X และชุด Y ตามลำดับ

นิยามของการเทียบคะแนนเชิงเส้นตรงจึงกล่าวไว้ว่า คะแนนของแบบทดสอบสองชุดจะเท่าเทียมกัน ถ้าต่างก็ตรงกับคะแนนมาตรฐาน ค่าเดียวกัน (Angoff. 1971 : 564; Petersen and others. 1982 : 73)

วิธีเทียบคะแนนเชิงเส้นตรงนี้มีข้อจำกัดที่ต้องคำนึงถึงที่สำคัญ คือ

1. ความเป็นคู่ขนานกันของแบบสอบกระทำไดยาก จึงมีปัญหว่า ความเป็นคู่ขนานกันโดยประมาณของแบบสอบควรจะเป็นเท่าไรตามที่ต้องการเพื่อให้มีความเพียงพอในวิธีการเทียบเชิงเส้นตรงนี้ (Jaeger. 1981 : 24)

2. หากมีความแตกต่างในรูปร่างของการแจกแจงคะแนนดิบจากแบบสอบสองชุดที่จะทำการเทียบคะแนนแล้ว จะกำหนดได้อย่างไรว่าความแตกต่างนั้นมีเพียงเล็กน้อยจนไม่จำเป็นต้องคำนึงถึงได้ (Jaeger . 1981 : 24)

3. ค่าพารามิเตอร์ของข้อสอบได้แก่ค่าความยากและค่าอำนาจจำแนก จะมีค่าไม่คงที่แต่จะเปลี่ยนไปตามความสามารถของกลุ่มผู้สอบ เช่นเดียวกับวิธีเทียบคะแนนแบบอีควิเปอร์เซนไทล์ ดังนั้นการเทียบคะแนนด้วยวิธีการเชิงเส้นตรงหรืออีควิเปอร์เซนไทล์จากกลุ่มผู้สอบที่มีความสามารถแตกต่างกันก็จะให้ผลไม่คงที่โดยแปรผันไปตามกลุ่มผู้สอบที่นำมาใช้เทียบคะแนน (Hambleton . 1978 : 498 - 499)

อย่างไรก็ตามหากรูปร่างของการแจกแจงคะแนนดิบจากแบบสอบสองชุดคล้ายคลึงกันแล้ว ผลการเทียบคะแนนด้วยวิธีการเชิงเส้นตรงและวิธีอีควิเปอร์เซนไทล์จะมีความใกล้เคียงกันมาก ซึ่งในกรณีการเลือกใช้วิธีการเทียบคะแนนเชิงเส้นตรงจะให้ผลการเทียบที่น่าพอใจกว่าวิธีเทียบคะแนนแบบอีควิเปอร์เซนไทล์ เนื่องจากวิธีเทียบเชิงเส้นตรงนี้ วิเคราะห์คะแนนทั้งหมดและเป็นอิสระจากความคลาดเคลื่อนต่าง ๆ อันเนื่องมาจากการเกลาเส้นกราฟให้เรียบ ซึ่งวิธีเทียบคะแนนแบบอีควิเปอร์เซนไทล์ จะเกิดความคลาดเคลื่อนชนิดนี้ค่อนข้างรุนแรงมากในพิสัยของคะแนนที่ข้อมูลมีน้อยหรือข้อมูลไม่แน่นอน (Angoff . 1971 : 564)

เนื่องจากการเทียบคะแนนโดยใช้วิธีอีควิเปอร์เซนไทล์ และวิธีเชิงเส้นตรงมีปัญหาในการวัดดังกล่าวมา จึงมีความต้องการวิธีการสำหรับเทียบคะแนนแบบสอบวิธีอื่น ๆ ซึ่งลอร์ด (Kolen . 1981 : 1 citing; Lord . 1977) ก็ได้เสนอว่าวิธีการเทียบคะแนนแบบดั้งเดิมทั้งสองวิธีไม่เหมาะสมในการเทียบคะแนนแบบสอบที่มีความยากแตกต่างกัน ในขณะที่การเทียบคะแนนโดยใช้ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) มีความเหมาะสมในการเทียบคะแนนมากกว่า

การเทียบคะแนนตามทฤษฎีการตอบข้อสอบ อาศัยรูปแบบความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้ตอบข้อสอบข้อใด ๆ ในแบบสอบว่าขึ้นอยู่กับความสามารถของผู้สอบ ซึ่งประมาณค่า

ได้จากโค้งลักษณะการตอบข้อสอบ (item Characteristic Curve; ICC) ของแต่ละรูปแบบที่ใช้ ซึ่งอาจจะเป็นรูปแบบหนึ่ง สอง หรือสามพารามิเตอร์ ทั้งนี้โดยไม่ขึ้นกับการแจกแจงคะแนนของกลุ่มตัวอย่าง (Lord and Novick, 1968 : 359) ดังนั้นการเทียบคะแนนด้วยวิธีนี้จะสามารถประมาณค่าความสามารถของผู้สอบแต่ละคนได้ (Harris and Hoover, 1987 : 151)

นิตยการเทียบคะแนนโดยใช้ทฤษฎีการตอบข้อสอบกล่าวว่า คะแนนแปลง Y^* และคะแนนดิบ X จะเทียบกันได้ก็ต่อเมื่อไม่ว่าผู้สอบจะสอบด้วยแบบสอบชุด X หรือชุด Y ก็จะไม่มีความแตกต่างกัน (Lord, 1977 : 128) ดังนั้น นิยามของการเทียบคะแนน จึงเน้นความสำคัญของความเสมอภาค (equity) หรือความเป็นธรรมที่บุคคลจะได้รับจากการทดสอบอย่างเท่าเทียมกัน ถึงแม้จะสอบแบบสอบต่างชุดก็ตาม (Lord, 1982 : 141) ผู้สอบจะไม่ได้รับหรือสูญเสียประโยชน์เป็นพิเศษอันเนื่องมาจากการได้รับการทดสอบด้วยแบบสอบชุดใดชุดหนึ่งที่มีความยากหรือง่ายกว่ากัน (Angoff, 1971 : 563)

ต่อมา ลอร์ด (Lord, 1980 : 199) ได้ตั้งข้อกำหนดที่สำคัญ 3 ประการ ในการเทียบคะแนนแบบสอบสองชุดที่วัดคุณลักษณะเดียว ซึ่งเทียบคะแนนที่สมมูลกันที่ระดับความสามารถเดียวกันไว้ดังนี้

1. ความเสมอภาค หมายถึง สำหรับทุก ๆ ความสามารถ θ การแจกแจงความถี่แบบมีเงื่อนไขของคะแนนแปลง $X(Y)$ หรือ Y^* ที่ θ ที่กำหนดให้ต้องเหมือนกับการแจกแจงความถี่แบบมีเงื่อนไขของคะแนน X
2. ความไม่แปรผันไปตามกลุ่มตัวอย่าง หมายถึง คะแนนแปลง $X(Y)$ จะคงที่ โดยไม่ขึ้นกับตัวแปรด้านอื่น ๆ ของประชากรที่นำมาเทียบคะแนน
3. ความสมมาตร หมายถึง การเทียบคะแนนจะต้องเหมือนกัน ไม่ว่าการเทียบนั้นจะเป็นการเทียบจากแบบสอบ X ไปยังแบบสอบ Y หรือจากแบบสอบ Y ไปยังแบบสอบ X ก็ตาม

การเทียบโดยใช้ทฤษฎีการตอบข้อสอบนี้มี 2 แบบ คือ การเทียบคะแนนโดยใช้คะแนนจริง และการเทียบโดยใช้คะแนนสังเกต ทั้งสองวิธีนี้มีข้อได้เปรียบและเสียเปรียบ กล่าวคือ วิธีเทียบคะแนนจริงไม่สามารถอธิบายคะแนนที่อยู่ต่ำกว่าระดับการเดาได้ โดยจะให้ความหมายของคะแนนสมมูลเฉพาะคะแนนที่อยู่เหนือค่าเฉลี่ยของการเดา ซึ่งถึงแม้จะเป็นการเทียบโดยใช้คะแนนจริงแต่ก็ยังเป็นค่าที่ประมาณได้จากสูตรในการคำนวณ ดังนั้นจึงยังคงมีความคลาดเคลื่อนอยู่ ส่วนการเทียบคะแนนโดยใช้คะแนนสังเกตนั้นเป็นการเทียบคะแนนโดยประมาณ ซึ่งสามารถอธิบายคะแนนสมมูลจาก X และ Y ได้ครอบคลุมพิสัยของคะแนนที่สังเกตได้ ลอร์ด ได้กล่าวว่า ทั้งสองวิธีนี้มีความสอดคล้องกันมาก แต่การสรุปผลเพื่อนำไปใช้อ้างอิงต้องทำอย่างพิถีพิถัน (Lord. 1980 : 202 - 203) อย่างไรก็ตาม การเทียบคะแนนโดยใช้คะแนนสังเกตตามทฤษฎีการตอบข้อสอบนี้ยังนำไปใช้ในงานวิจัยน้อยมากอาจเป็นเพราะว่า วิธีนี้มีความซับซ้อนและลงทุนแพงกว่าวิธีการเทียบคะแนนโดยใช้คะแนนจริง (Lord and Wingersky. 1984 : 453)

ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงเลือกใช้วิธีเทียบคะแนนจริงตามทฤษฎีการตอบข้อสอบ เพราะจะสามารถเตรียมทั้งค่าประมาณพารามิเตอร์ของข้อสอบ และค่าประมาณความสามารถของผู้สอบที่ไม่แปรผันไปตามคุณลักษณะของกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ อย่างไรก็ตาม โมเดลการเทียบคะแนนด้วยวิธีนี้ยังแบ่งออกได้เป็น 3 รูปแบบ ซึ่งแตกต่างกันไปตามจำนวนพารามิเตอร์ที่ใช้ในแต่ละรูปแบบดังนี้

1. รูปแบบที่ใช้พารามิเตอร์ 2 ตัว (Two Parameters Logistic Model) ฟังก์ชันที่ใช้กับเส้นโค้งลักษณะของข้อสอบ มีรูปแบบดังนี้

$$P_g(\theta) = \frac{\ell \frac{Da_g(\theta - b_g)}{g}}{1 + \ell \frac{Da_g(\theta - b_g)}{g}} \quad ; g = 1, 2, \dots, n$$

เมื่อ $P_g(\theta)$ คือ ความน่าจะเป็นในการตอบข้อสอบข้อที่ g ถูกเมื่อผู้สอบมีความสามารถ θ

- a_g คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ g
 b_g คือ ค่าความยากของข้อสอบข้อที่ g
 D คือ scaling factor มีค่า 1.702
 e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818 ...

2. รูปแบบที่ใช้พารามิเตอร์ 3 ตัว (Three-Parameters Logistic Model)

รูปแบบ 3 พารามิเตอร์สามารถเขียนได้จากรูปแบบ 2 พารามิเตอร์ โดยเพิ่มพารามิเตอร์ตัวที่ 3 คือ ค่าการเดิมน ซึ่งมีสูตรในการคำนวณดังนี้

$$P_g(\theta) = c_g + (1 - c_g) \frac{\frac{Da_g(\theta - b_g)}{l}}{1 + \frac{Da_g(\theta - b_g)}{l}} \quad ; \quad g = 1, 2, 3, \dots, n$$

เมื่อ c_g คือ ค่าการเดิมนของข้อสอบข้อที่ g ทั้งนี้รูปแบบนี้ไม่สามารถประมาณค่าพารามิเตอร์ที่อยู่ต่ำกว่าระดับการเดิมนได้

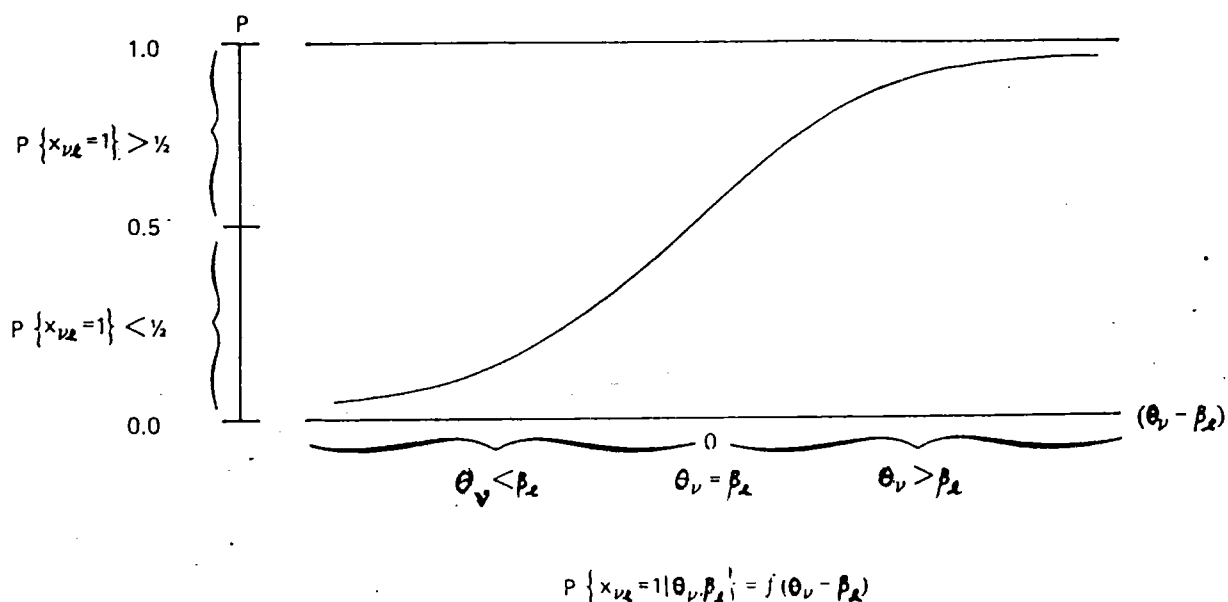
3. รูปแบบที่ใช้พารามิเตอร์ 1 ตัว (One-Parameter Logistic Model) ซึ่งราสช์ (Rasch) ได้พัฒนาขึ้นมาใช้ โดยรูปแบบนี้เป็นกรณีพิเศษของรูปแบบ 2 พารามิเตอร์ ที่เบอร์บอม (Birnbuam) ได้พัฒนาขึ้น ข้อตกลงเบื้องต้นที่เพิ่มขึ้น คือ ค่าอำนาจจำแนกของข้อสอบ (a) เป็นค่าคงที่สำหรับข้อสอบแต่ละข้อ และค่าความยากของข้อสอบแปรเปลี่ยนไป สูตรในการคำนวณดังนี้

$$P_g(\theta) = \frac{\frac{Da(\theta - b_g)}{l}}{1 + \frac{Da(\theta - b_g)}{l}} \quad ; \quad g = 1, 2, 3, \dots, n$$

แนวคิดเชิงทฤษฎีของรูปแบบราสช์

รูปแบบราสช์เป็นรูปแบบที่มีความซับซ้อนน้อยที่สุด ในโมเดลการวิเคราะห์ตามทฤษฎีการตอบข้อสอบ โดยมีแนวคิดที่ว่า โอกาสที่ผู้สอบแต่ละคนจะทำข้อสอบแต่ละข้อได้ถูกต้องหรือไม่ ขึ้นอยู่กับ

ความแตกต่างของค่าพารามิเตอร์เพียง 2 ค่าเท่านั้น คือ ผลต่างของค่าความสามารถของผู้สอบ (θ) และค่าความยากของข้อสอบ (β หรือ b) เช่น ถ้าค่าความสามารถของผู้สอบ เท่ากับ .50 และความยากของข้อสอบ เท่ากับ .50 ด้วยแล้ว โอกาสที่ผู้สอบจะทำข้อสอบได้ถูกต้องจะเท่ากับ .50 และถ้าหากค่า $\theta < \beta$ แล้ว โอกาสที่ผู้สอบจะทำข้อสอบได้ถูกต้องย่อมจะน้อยกว่า .50 และในทำนองเดียวกัน ถ้าหากค่า $\theta > \beta$ แล้ว โอกาสที่ผู้สอบจะทำข้อสอบได้ถูกต้องก็มีมากกว่า .50 ดังภาพ 1 (Wright and Stone. 1979 : 14)



ภาพประกอบ 1 แสดงโค้งการตอบข้อสอบ (the response curve)

จากภาพ 2 เส้นความสัมพันธ์ระหว่าง และ เป็นความสัมพันธ์เชิงคณิตศาสตร์ กล่าวคือ โอกาสที่บุคคล v ซึ่งมีความสามารถ θ (หรือ θ_v) จะทำข้อสอบ ℓ ที่มีระดับความยาก β (หรือ β_ℓ) ซึ่งเขียนได้ในรูป

$$\text{โอกาสแห่งความสำเร็จ (odd of success)} = (\theta - \beta_\ell)$$

ในการคำนวณโอกาสในการตอบข้อสอบถูก จะมีค่าอยู่ระหว่าง $\pm \alpha$ ดังนั้นเพื่อให้ค่าของ $(\theta_v - \beta_\ell)$ เป็นค่าที่มีหน่วยเล็กลงและคงที่ที่เหมาะสมแก่การนำมาใช้ จึงใช้ค่า exponential (exp) ของ $(\theta_v - \beta_\ell)$ แทน ซึ่งจะมีค่าอยู่ระหว่าง 0 กับ $+\infty$ โดยมีหน่วยการวัดเรียกว่า log-odds units หรือ logits และเพื่อให้ค่า $\exp(\theta_v - \beta_\ell)$ มีค่าเป็น interval scale ระหว่าง 0 และ 1 จะได้สูตรที่แสดงว่า โอกาสที่บุคคล v ที่มีความสามารถ θ จะทำข้อสอบ ℓ ที่มีความยาก β ได้ถูกต้อง (ได้คะแนน = 1 หรือ X_{v1}) คือ

$$P\{X_{v1} = 1/\theta_v, \beta_\ell\} = \exp(\theta_v - \beta_\ell) / 1 + \exp(\theta_v - \beta_\ell)$$

เมื่อ $\exp = 2.71828$

จะเห็นว่าค่า $\theta > \beta$ มาก ค่า P ก็จะสูง และในกรณีที่ข้อสอบมีทั้งหมด 1 ข้อ และกลุ่มตัวอย่างมี n คน คะแนนรวมของคนที่ v หรือ $r_v = \sum_{\ell} X_{v\ell}$ และจำนวนคนที่ทำข้อสอบแต่ละข้อถูก คือ $s_\ell = \sum_v X_{v\ell}$

ดังนั้น ถ้ากำหนดให้ $((X_{v\ell}))$ แทนเมตริก (matrix) ทั้งหมดของผลการสอบแล้ว โอกาสที่ผู้สอบทุกคนจะตอบข้อสอบทุกข้อถูกต้อง คือ (Wright and Stone, 1979 : 19 - 20)

$$P\{((X_{v\ell}))/(\theta_v)_i(\beta_\ell)\} = \frac{[\exp(\sum_v r_v \theta_v)] [\exp(-\sum_\ell s_\ell \beta_\ell)]}{\prod_v \prod_\ell [1 + \exp(\theta_v - \beta_\ell)]}$$

สมการนี้มีความสำคัญมาก เพราะแสดงว่า การประมาณค่าพารามิเตอร์ของบุคคลและของข้อสอบต่างเป็นอิสระจากกัน กล่าวคือ การคำนวณหาค่า θ ของแต่ละบุคคล (θ_v) เราอาศัยเพียงคะแนนรวมของแต่ละข้อ (s_ℓ) เท่านั้น ในการกำจัดค่า β_ℓ ออกจากสมการนี้ ดังนั้นการคำนวณหาค่า θ จึงเป็นอิสระจากกลุ่มข้อสอบ (item-free person measurement) ในทำนองเดียวกัน การคำนวณหาค่าความยากของข้อสอบแต่ละข้อ (β_ℓ) ก็อาศัยเพียงคะแนนรวมของแต่ละบุคคล (r_v) เท่านั้น ในการกำจัดค่า θ_v ออกจากสมการ ดังนั้นวิธีการคำนวณหา β จึงเป็นอิสระจากกลุ่มตัวอย่าง (sample-free item calibration)

ข้อตกลงเบื้องต้นของรูปแบบราศี

ข้อตกลงเบื้องต้นของรูปแบบราศีมีดังนี้ (Lord and Novick. 1968 : 359;
Hambleton. 1977 : 78; Tinsley. 1975 : 326)

1. แบบสอบมีมิติเดียว กล่าวคือแบบสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียวหรือมีความเป็นเอกพันธ์กัน (homogeneous)
2. ข้อสอบแต่ละข้อจะต้องเป็นอิสระจากกัน กล่าวคือการตอบข้อสอบข้อหนึ่งจะไม่มีผลต่อการตอบข้อสอบข้ออื่น
3. โอกาสที่ผู้สอบจะตอบข้อสอบแต่ละข้อได้ถูกต้องขึ้นอยู่กับความสามารถของผู้สอบ และความยากของข้อสอบ เท่านั้น
4. คะแนนเป็นแบบ dichotomous เท่านั้น คือ เป็น 1 หรือ 0
5. ข้อสอบแต่ละข้อมีค่าอำนาจจำแนกเท่ากันหรือมีค่าคงที่เพื่อความสะดวกในการคำนวณ จึงให้เป็น 1 และไม่มีการเดาเกิดขึ้น

ความแรงของรูปแบบราศี

ในการใช้รูปแบบราศีนั้น ได้มีการทดลองใช้ในสถานการณ์ที่ต่าง ๆ กัน ในกรณีข้อมูลไม่เป็นไปตามข้อตกลงของรูปแบบนี้ เพื่อหาความแรงของรูปแบบต่อการที่จะนำไปใช้ ซึ่งปรากฏผลดังนี้

1. ขนาดของกลุ่มตัวอย่างที่จะใช้กับรูปแบบราศีอาจมีขนาดเล็ก ประมาณ 100 คน เท่านั้นก็ได้ (Wright. 1977 : 219)
2. การกระจายความสามารถของกลุ่มตัวอย่างหรือความยากของข้อสอบไม่จำเป็นต้องเป็นโค้งปกติ เพียงแต่มีแนวโน้มว่าจะเป็นโค้งปกติเท่านั้น และกลุ่มตัวอย่างไม่จำเป็นต้องมาจากการสุ่มก็ได้ (wright and Stone. 1979 : 20)
3. ข้อสอบไม่จำเป็นต้องมีอำนาจจำแนกเท่านั้น และไม่จำเป็นต้องคำนึงถึงค่าการเดา ก็ได้ (Wright and Panchapakesen. 1969 : 25) ซึ่งจากการวิจัยพบว่าทั้งรูปแบบราศี

และรูปแบบสามพารามิเตอร์ต่างก็เหมาะสมกับข้อมูลจริงประมาณร้อยละ 80 (Hutten. 1982 : 4799-A) และหากข้อสอบที่วิเคราะห์ตามทฤษฎีตั้งเดิมและคัดเลือกแล้วว่าเป็นข้อสอบที่ดี จะมีอยู่ประมาณร้อยละ 5 ที่ไม่เหมาะกับการวิเคราะห์ตามรูปแบบราสส์ ซึ่งถือว่าเป็นข้อสอบที่ไม่ดี (Rentz and Rentz. 1978 : 10)

4. ข้อสอบที่มีน้ำหนักคะแนนต่างกัน (multiple point) ก็ใช้กับรูปแบบนี้ได้ (สุพัณณ์ สุกมลสันต์. 2530 : 37; อ้างอิงมาจาก Willmott. 1980 : 195)

การประมาณค่าพารามิเตอร์ของความสามารถและของข้อสอบ

การคำนวณหาค่า θ_v และ θ_d โดยวิธีการกำจัดค่าใดค่าหนึ่งออกไปจากสมการเพื่อหาค่าอีกค่าหนึ่งนั้นอาจใช้วิธี Conditional Maximum Likelihood แต่วิธีนี้มีความยุ่งยากในการคำนวณ และใช้วิเคราะห์ข้อสอบที่ยาวเกินกว่า 25 ข้อไม่ได้ ต่อมาจึงใช้วิธี Unconditional Maximum Likelihood (UCON) ซึ่งสามารถวิเคราะห์ข้อสอบที่มีขนาดยาวหรือสั้นได้ด้วย แต่จะมีความลำเอียง (bias) บ้างเล็กน้อย ซึ่งก็แก้ไขได้ อย่างไรก็ตามวิธีดังกล่าวนี้ในโปรแกรมไบคาล (BICAL) แนะนำให้ใช้เมื่อข้อสอบมีขนาดสั้น (ประมาณ 25 ข้อ) และการกระจายของคะแนนมีลักษณะเบ้ แม้ว่าจะมีขนาดของกลุ่มตัวอย่างมากหรือน้อยก็ตาม

ส่วนในกรณีข้อสอบมีขนาดยาว (มากกว่า 25 ข้อขึ้นไป) และมีกลุ่มตัวอย่างขนาดใหญ่ (มากกว่า 400 คน) และเชื่อว่าการกระจายของคะแนนมีแนวโน้มว่าเป็นโค้งปกติแล้ว ควรใช้วิธีของ Cohen's Approximation (PROX) จะให้ผลดีกว่า (สุพัณณ์ สุกมลสันต์. 2530 : 35)

สำหรับการประมาณค่าพารามิเตอร์ข้อสอบ และความสามารถของผู้สอบนั้น โปรแกรมจะใช้วิธีการกำหนดค่าเริ่มแรกของพารามิเตอร์ที่ละค่า แล้วทำการประมาณค่าซ้ำ ๆ กันอย่างต่อเนื่อง จนกระทั่งมีความเป็นไปได้อย่างสอดคล้องกันสูงสุดระหว่างค่าประมาณของความสามารถและของข้อสอบกับข้อมูลการสอบจริง (Hambleton and Swaminathan. 1985 : 14) หรือได้ค่าพารามิเตอร์

ต่าง ๆ กัน โดยมีขั้นตอนในการทำงานของโปรแกรมโดยสรุปดังนี้ (Wright and Stone .

1977 : 32 - 65)

ขั้นที่ 1 โปรแกรมจะทำการตรวจสอบเพื่อหาว่าข้อสอบแต่ละข้อมีคนทำได้ถูกต้องกี่คน และจะนับจำนวนความถี่ของคะแนนรวมของผู้สอบแต่ละคน

ขั้นที่ 2 กำจัดข้อสอบที่ไม่มีคนทำได้เลย หรือข้อสอบที่มีคนตอบถูกทุกคน และกำจัดคนที่ได้คะแนน 0 หรือได้คะแนนเต็ม ออกจากการวิเคราะห์เพราะถือว่าข้อสอบและคะแนนดังกล่าวไม่มีลักษณะจำเป็นที่จะนำมาวิเคราะห์

ขั้นที่ 3 ให้ค่าประมาณความยาก (b_i) และค่าประมาณความสามารถ (θ) ซึ่งได้จากการกำหนดค่าเริ่มแรกของ b_i , θ และกำหนดค่าของชุดของ b_i ที่ θ แล้วใช้วิธีการทำซ้ำ ๆ จนได้ค่า b ของข้อสอบข้อต่าง ๆ กันที่ และคำนวณค่า θ ของผู้สอบที่ได้คะแนนรวมแตกต่างกันได้

ขั้นที่ 4 ปรับปรุงค่า b คำนวณค่า θ และปรับปรุงค่า θ ของผู้ที่ได้คะแนนรวมต่างกัน

ขั้นที่ 5 ทดสอบความเหมาะสม (fit) ของโค้งลักษณะข้อสอบแต่ละข้อว่าเหมาะสมกับโค้งลักษณะข้อสอบตามรูปแบบหรือไม่ โปรแกรมจะแบ่งกลุ่มตัวอย่างออกเป็นกลุ่มย่อยที่มีความสามารถแตกต่างกันโดยใช้คะแนนรวมเป็นเกณฑ์ตามจำนวนกลุ่มที่ผู้วิเคราะห์ต้องการแต่ไม่เกิน 6 กลุ่ม การศึกษาครั้งนี้ผู้วิจัยจะแบ่งเป็น 6 กลุ่ม เพื่อประเมินความไม่แปรผันของค่าประมาณความยากของข้อสอบ เมื่อกลุ่มผู้สอบมีความสามารถต่างกันหรือเมื่อเปลี่ยนกลุ่มผู้สอบโดยมีหลักในการพิจารณาดังนี้

5.1 สัดส่วนของคนที่ตอบถูกจะเพิ่มขึ้นจากกลุ่มที่มีความสามารถน้อย ไปยังกลุ่มที่มีความสามารถมากกว่า

5.2 โอกาสที่ผู้สอบคนหนึ่งจะตอบถูกในข้อสอบที่ง่ายย่อมมีมากกว่าโอกาสที่จะตอบถูกในข้อที่ยาก

5.3 ค่าเฉลี่ยกำลังสอง ระหว่างกลุ่มจะบอกถึงความสอดคล้องระหว่างโค้งลักษณะข้อสอบแต่ละข้อกับโค้งลักษณะข้อสอบตามรูปแบบของราสส์ ซึ่งค่าเฉลี่ยกำลังสองมีค่าเท่ากับ 1 แสดงว่าโค้งทั้งสองมีความเหมาะสมกันพอดี

ขั้นที่ 6 โปรแกรมจะให้ค่าอำนาจจำแนก (biserial correlation) ถึงแม้รูปแบบราสส์จะมีข้อตกลงเบื้องต้นว่าค่าอำนาจจำแนกของข้อสอบแต่ละข้อมีค่าเท่ากันและไม่ได้นำมาใช้ในการวิเคราะห์ก็ตาม แต่ค่าอำนาจจำแนกของข้อสอบใดที่มีค่าใกล้ 1.00 จะแสดงว่าโครงสร้างของข้อสอบนั้นมีความใกล้ชิดกับโครงสร้างข้อสอบที่คาดหวังของรูปแบบราสส์

ขั้นที่ 7 โปรแกรมจะเรียงข้อสอบที่ประมาณค่าพารามิเตอร์แล้วใน 3 ลักษณะ คือ

1. เรียงตามลำดับของข้อสอบ
2. เรียงตามลำดับความยากจากง่ายที่สุดไปยังยากที่สุด
3. เรียงตามลำดับความเหมาะสมกับโมเดล

รูปแบบในการเก็บรวบรวมข้อมูล

ในการเทียบคะแนนแบบสอบต่างชุดซึ่งสอบกับผู้สอบต่างกลุ่มนั้น การประมาณค่าพารามิเตอร์แยกกันระหว่างแบบสอบแต่ละชุดจะไม่อยู่ในสเกลเดียวกัน ดังนั้นจึงจำเป็นต้องออกแบบในการเก็บรวบรวมข้อมูลภายใต้รูปแบบที่สามารถจะจัดกระทำทางสถิติ เพื่อหาค่าคงที่สำหรับการแปลงค่าพารามิเตอร์ที่ประมาณค่าได้จากแบบสอบต่างชุดให้อยู่ในสเกลร่วมกัน และสามารถเทียบหาคะแนนสมมูลระหว่างแบบสอบได้ สำหรับรูปแบบในการเก็บรวบรวมข้อมูลเพื่อเทียบคะแนนนั้นพอแบ่งได้เป็น 3 รูปแบบ ดังนี้ (Lord. 1982 : 145; Hambleton. 1985 : 198)

1. รูปแบบที่ใช้กลุ่มผู้สอบกลุ่มเดียว หรือใช้ผู้สอบร่วมโดยกลุ่มเดียวทำแบบสอบทั้งสองชุด การประมาณค่าความยาก จึงไม่เกี่ยวข้องกับความสามารถในระดับความสามารถของกลุ่มผู้สอบ แต่ลำดับการสอบชุดใดก่อนหลังอาจทำให้เกิดการเรียนรู้ การฝึกฝน และความเหนื่อยล้า ซึ่งจะมีผลกระทบต่อคะแนนผลการสอบได้ ดังนั้นเพื่อขจัดความลำเอียงเหล่านี้จึงจัดแปลงรูปแบบนี้โดยผู้สอบเป็นสองกลุ่ม แต่ละกลุ่มทำแบบสอบทั้งสองชุดในลักษณะของการจัดลำดับก่อนหลังต่างกันโดยวิธีสุ่ม
2. รูปแบบที่ใช้กลุ่มที่เท่าเทียมกัน รูปแบบนี้กลุ่มผู้สอบที่เท่าเทียมกันสองกลุ่ม ซึ่งอาจได้มาโดยการสุ่ม หรือเป็นกลุ่มที่ได้จากการจับคู่กัน (matching) ทางด้านความรู้และความสามารถ

ทั้งนี้เพื่อให้คะแนนผลการสอบจากแบบสอบแต่ละชุดที่จะนำมาเทียบคะแนนนั้นไม่มีผลของความแตกต่างในความสามารถระหว่างกลุ่มผู้สอบ อย่างไรก็ตามกลุ่มผู้สอบทั้งสอง อาจมีความแตกต่างในการแจกแจงความสามารถ ถึงแม้จะมีเพียงเล็กน้อยก็อาจทำให้เกิดความลำเอียงในการเทียบคะแนนได้

3. รูปแบบที่ใช้แบบสอบรวม รูปแบบนี้กลุ่มผู้สอบสองกลุ่มที่แตกต่างกัน (อาจเป็นกลุ่มที่ได้มาจากการสุ่มหรือไม่ได้มาจากการสุ่ม) แต่ละกลุ่มทำแบบสอบเพียง 1 ชุด ซึ่งแต่ละชุดจะประกอบด้วยข้อสอบรวมจำนวนหนึ่ง โดยอาจสอบรวมหรือสอบแยกต่างหากจากแบบสอบทั้งสองชุดก็ได้ แบบสอบรวมจะต้องคู่ขนานหรือเป็นตัวแทน แบบสอบทั้งสองชุดนั้นให้มากที่สุด แบบสอบรวมจะกำหนดสเกลรวมของคะแนนจากแบบสอบแต่ละชุดที่สอบโดยผู้สอบทั้งสองกลุ่ม เพื่อนำไปปรับค่าความยากของข้อสอบ หรือค่าความสามารถของผู้สอบจากแบบสอบต่างชุดให้อยู่ในสเกลร่วมกัน

การหาค่าคงที่ในการแปลงค่าประมาณพารามิเตอร์

ในกรณีที่เราทราบค่าพารามิเตอร์แต่ละคู่ของแบบสอบแต่ละชุด เช่น ค่าความยากของข้อสอบ (b_x, b_y) และค่าความสามารถของผู้สอบ (θ_x, θ_y) เมื่อนำมาพล็อตกราฟจะให้ความสัมพันธ์เชิงเส้นตรง ซึ่งสามารถคำนวณหาค่าความชัน และจุดตัดแกนของสมการเชิงเส้นตรง แต่หากทราบเฉพาะค่าประมาณพารามิเตอร์แล้ว ค่าต่าง ๆ เหล่านี้จะมีการกระจายโดยรอบเส้นความสัมพันธ์เชิงเส้นตรงนี้ ดังนั้นในการแปลงค่าประมาณพารามิเตอร์ของแบบสอบแต่ละชุดให้อยู่ในสเกลร่วมกันเพื่อให้สามารถเทียบหาคะแนนสมมูลได้ จึงใช้การแปลงเชิงเส้นตรง โดยเลือกวิธีการในการหาค่าคงที่ได้จากวิธีการต่อไปนี้ (Hambleton. 1985 : 205 - 207)

1. วิธีการถดถอย (regression method) วิธีนี้เป็นวิธีที่ใช้หาความสัมพันธ์เชิงเส้นตรงระหว่าง 2 ตัวแปรได้โดยตรงที่สุด โดยใช้สมการ

$$Y = \alpha x + \beta + e$$

เมื่อ $Y = \theta_y$ และ $X = \theta_x$ (เมื่อเทียบค่า θ)

และ $Y = b_y$ และ $X = b_x$ (เมื่อเทียบค่าความยาวของข้อสอบ)

e คือ ค่าความคลาดเคลื่อน

$$\alpha = r_{xy} S_y / S_x$$

$$\beta = \bar{Y} - \alpha \bar{X}$$

เมื่อ r_{xy} คือ ค่าความสัมพันธ์ระหว่าง X และ Y

Y, X คือ ค่าเฉลี่ยของคะแนนชุด Y และ X ตามลำดับ

S_y, S_x คือ ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนชุด Y และ X ตามลำดับ

วิธีนี้ให้ความสัมพันธ์ที่ไม่เป็นสมมาตร โดยค่าสัมประสิทธิ์ของสมการถดถอยจะขึ้น

กับแบบสอบที่เลือกเป็นตัวแปรต้น (ยกเว้นกรณีที่ $\alpha = 1$)

2. วิธีที่ใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน (mean and sigma method) วิธีนี้

กำหนดความสัมพันธ์เชิงเส้นตรงดังนี้

$$\text{ถ้า } Y = \alpha X + \beta$$

$$\text{แล้ว } \bar{Y} = \alpha \bar{X} + \beta$$

$$\text{และ } S_y = \alpha S_x$$

$$\text{จะได้ } \alpha = S_y / S_x$$

$$\text{และ } \beta = \bar{Y} - \alpha \bar{X}$$

ความสัมพันธ์มีคุณลักษณะสมมาตร (สำหรับรูปแบบราสส์ วิธีนี้จะให้ผลเหมือนกับวิธี

ที่ 1)

3. วิธีที่ใช้ค่าที่ถ่วงน้ำหนักแล้วของค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน (robust mean and sigma method) วิธีนี้กำหนดน้ำหนักของแต่ละคู่ของค่าประมาณพารามิเตอร์ ที่ได้จากแบบสอบต่างชุดตามความแปรปรวน ในการประมาณค่าของแต่ละชุด โดยมีวิธีการในการหาค่าคงที่ของการเทียบคะแนนโดยสรุปดังนี้

3.1 กำหนด w_j สำหรับแต่ละคู่ของ (X_j, Y_j) ด้วย

$$w_j = \max[V(X_j), V(Y_j)], j = 1, \dots, k$$

เมื่อ $V(\theta)$ คือ ความแปรปรวนของค่าประมาณที่ j

3.2 ให้สเกลน้ำหนักคือ

$$w_j = w_j / (\sum_{j=1}^n w_j)$$

3.3 คำนวณค่า $X'_j = w'_j X_j, j = 1, \dots, k$

และ $Y'_j = w'_j Y_j, j = 1, \dots, k$

3.4 หาค่า X', Y', S'_x, S'_y

เมื่อ X', Y', S'_x, S'_y คือ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ

คะแนนที่ถ่วงน้ำหนัก (weighted scores)

3.5 คำนวณค่า α และ β จากสมการ $\beta = Y - \alpha X$ โดยใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานที่ถ่วงน้ำหนัก

นอกจากนี้อาจใช้การกำหนดน้ำหนักซ้ำ ๆ (reweight) เพื่อให้ได้ค่า α และ β ที่ให้น้ำหนักค่าต่อค่าประมาณพารามิเตอร์ที่ไม่ดี (Stocking and Lord. 1983; Hambleton. 1985 : 208 - 209)

4. วิธีที่ใช้โคลงลักษณะข้อสอบ วิธีนี้กำหนดความสัมพันธ์ระหว่างค่าความยากของข้อสอบจากสมการ

$$b_y = \alpha b_x + \beta$$

คะแนนจริง ของผู้สอบซึ่งมีความสามารถ θ_a ในแบบสอบชุด X คือ

$$\xi_{xa} = \sum_{i=1}^n P_i(\theta_a, a_{xi}, b_{xi}, c_{xi})$$

คะแนนจริงของผู้สอบ a ซึ่งมีความสามารถ θ_a ในแบบสอบชุด Y คือ

$$\xi_{ya} = \sum_{i=1}^n P_i(\theta_a, a_{yi}, b_{yi}, c_{yi})$$

$$\text{เมื่อ } b_{yi} = \alpha b_{xi} + \beta$$

$$a_{yi} = a_{xi} / \alpha$$

$$c_{yi} = c_{xi}$$

ค่าคงที่ α และ β ควรเลือกให้มีความแตกต่างระหว่าง ϵ_{xa} และ ϵ_{ya} น้อย

ดั่งสมการ (Stocking and Lord. 1983 : 203 - 204)

$$F = \frac{1}{N} \sum_{a=1}^n (\epsilon_{xa} - \epsilon_{ya})^2$$

สต็อก และลอร์ด พบว่าวิธีที่ใช้โด่งลักษณะข้อสอบนี้ ให้ผลดีเทียบได้กับวิธีที่ใช้ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน อย่างไรก็ตามในกรณีที่ต้องการแปลงค่าอำนาจจำแนกพบว่า วิธีที่ใช้โด่งลักษณะข้อสอบให้ผลดีกว่าเพราะใช้ข้อมูลที่ประมาณค่าได้ทุกค่า

การเทียบค่าประมาณความสามารถของผู้สอบ

การเทียบค่าประมาณความสามารถ จะแตกต่างกันไปตามรูปแบบในการเก็บรวบรวมข้อมูล ซึ่งมีรายละเอียดในการเทียบดังนี้

1. รูปแบบที่มีผู้สอบกลุ่มเดียว เมื่อแบบสอบชุด X และชุด Y สอบกับผู้สอบกลุ่มเดียวกัน มีวิธีการที่เหมาะสม 2 วิธี คือ

วิธีที่ 1 รวมข้อมูลทั้งสองชุดและประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถ ซึ่งจะได้ค่าพารามิเตอร์ที่อยู่ในสเกลของแบบสอบชุด X และแบบสอบชุด Y

วิธีที่ 2 แยกข้อมูลแต่ละชุด (ถ้าจำเป็นต้องประมาณค่าพารามิเตอร์แยกกันสำหรับแบบสอบแต่ละชุด) แล้วกำหนดค่า θ ให้เหมือนกันในการประมาณค่าพารามิเตอร์แต่ละชุด ด้วยวิธีนี้ ค่าประมาณความสามารถจะอยู่บนสเกลร่วมกัน

2. รูปแบบที่ใช้กลุ่มที่เท่าเทียมกัน กรณีนี้ไม่มีข้อสอบร่วมหรือผู้สอบร่วม จึงจำเป็นต้องประมาณค่าพารามิเตอร์แยกกันสำหรับแบบสอบแต่ละชุด โดยมีขั้นตอนดังนี้

2.1 กลุ่มผู้สอบที่ได้จากการสุ่ม (จำนวนผู้สอบเท่ากัน) โดยแต่ละกลุ่มทำแบบสอบเพียง 1 ชุด คือ ชุด X หรือชุด Y

2.2 ประมาณค่าพารามิเตอร์แยกกันสำหรับกลุ่มผู้สอบแต่ละกลุ่ม โดยกำหนดค่า θ ให้เหมือนกันในการประมาณค่าทั้งสองกลุ่ม

2.3 เรียงลำดับค่าประมาณ θ และจับคู่ระหว่างค่าประมาณที่ต่ำสุดของ θ_x กับ θ_y ที่ต่ำสุดที่ละคู่เรื่อยไป

2.4 พล็อตกราฟระหว่างค่า θ_x กับค่า θ_y ซึ่งถ้าเป็นไปตามข้อตกลงของทฤษฎีการตอบข้อสอบแล้ว ผลที่ได้จะเป็นเส้นตรง อย่างไรก็ตามในทางปฏิบัติใช้ค่าประมาณพารามิเตอร์ ดังนั้นผลที่ได้อาจไม่เป็นเส้นตรง โดยเฉพาะที่ส่วนปลายของค่า θ_x และ θ_y จะมีความคลาดเคลื่อนในการประมาณค่ามากกว่าส่วนอื่น

3. รูปแบบที่ใช้แบบสอบร่วม ในรูปแบบนี้ผู้สอบ N_x ทำแบบสอบชุด X ซึ่งมี n_x ข้อ และแบบสอบร่วม n_a ข้อ เช่นเดียวกับผู้สอบ N_y ทำแบบสอบชุด Y ซึ่งมี n_y ข้อ และข้อสอบร่วมอีก n_a ข้อ การเทียบค่าประมาณความสามารถมีขั้นตอนดังนี้

3.1 ประมาณค่าแยกกันสำหรับแบบสอบแต่ละชุด คือ ชุด X มีผู้สอบ N_x คน และมีข้อสอบ $(n_x + n_a)$ ข้อ ส่วนชุด Y มีผู้สอบ N_y และมีข้อสอบ $(n_y + n_a)$ ข้อ กำหนดค่า θ ให้เหมือนกัน ในการประมาณค่าทั้งสองชุดแล้วหาเส้นความสัมพันธ์ระหว่างแบบสอบ โดยผ่านข้อสอบร่วม n_a ข้อ ที่สอบกับกลุ่มผู้สอบทั้งสองกลุ่ม $(N_x + N_y)$ คน ตามวิธีที่ได้อธิบายในข้อ 1 ดังนั้น ความสามารถของผู้สอบจะเทียบกันได้โดยใช้เส้นความสัมพันธ์

3.2 ใช้โปรแกรม LOGIST ประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบพร้อมกันทั้งสองชุดดังนี้

3.2.1 จัดข้อมูลรวบกับว่าผู้สอบ $N_x + N_y$ คน ทำแบบสอบ $(n_x + n_y + n_a)$ ข้อ

3.2.2 จัดข้อสอบ n_y ข้อ ซึ่งผู้สอบ N_x คน ไม่ได้ทำแบบสอบให้เป็นข้อสอบที่ผู้สอบทำไม่ถึง (not reached) ส่วนข้อสอบ n_x ข้อ ซึ่งผู้สอบ N_y คน ไม่ได้ทำข้อสอบ ก็จะจัดให้เป็นข้อสอบที่ผู้สอบทำไม่ถึงเช่นเดียวกัน

3.2.3 ประมวลค่าพารามิเตอร์ความสามารถ สำหรับผู้สอบ ($N_x + N_y$ คน) และพารามิเตอร์ของข้อสอบ สำหรับข้อสอบ ($n_x + n_y + n_a$ ข้อ) พร้อมกันครั้งเดียว ด้วยวิธีนี้ค่าประมวลพารามิเตอร์ที่ได้จากแบบสอบต่างชุดจะอยู่ในสเกลเดียวกันทั้งหมด และทำการเทียบค่าประมวลความสามารถระหว่างผู้สอบแบบสอบต่างชุดได้

การเทียบคะแนนโดยใช้คะแนนจริง

ในบางกรณีการเทียบคะแนนโดยใช้สเกลความสามารถดังกล่าวอาจไม่เหมาะสม ดังนั้นค่า θ อาจแปลงไปสู่ค่าคะแนนจริงที่สอดคล้องกันได้ ทั้งนี้โดยอาศัยสมการความสัมพันธ์เชิงคณิตศาสตร์ระหว่างค่าความสามารถ θ และคะแนนจริงที่ได้จากแบบสอบแต่ละชุด ชุด X หรือชุด Y ดังนี้ (Hambleton. 1985 : 212; Lord. 1980 : 199 - 203)

$$\xi_x = \sum_{i=1}^n p_i(\theta_x)$$

เมื่อ $p_i(\theta)$ คือ ค่าความน่าจะเป็นในการตอบข้อสอบที่ i ได้ถูกต้อง เมื่อผู้สอบมีความสามารถ θ

θ_x คือ ระดับความสามารถของผู้สอบในแบบสอบชุด X

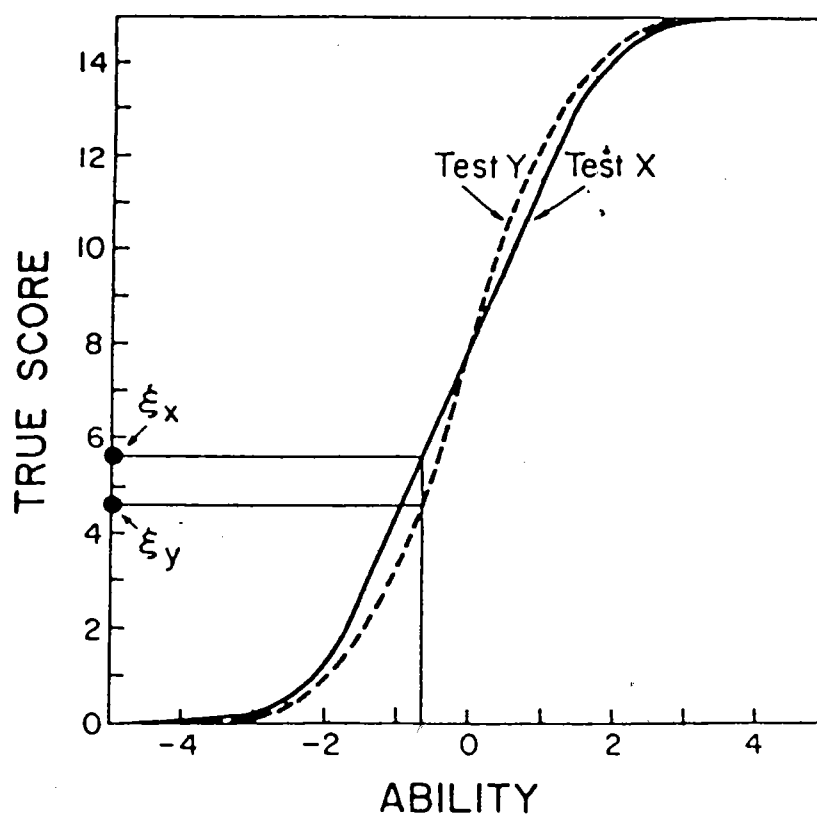
ξ_x คือ คะแนนจริงที่ได้จากแบบสอบชุด X

ในทำนองเดียวกัน ถ้า θ_y คือ ระดับความสามารถของผู้สอบในแบบสอบชุด Y แล้วคะแนนจริงที่สมมูลกันในแบบสอบชุด Y คือ

$$\xi_y = \sum_{j=1}^m p_j(\theta_y) = \sum_{j=1}^m p_j(\alpha\theta_x + \beta)$$

เมื่อ $\theta_y = \alpha\theta_x + \beta$ คือ เส้นความสัมพันธ์ระหว่าง θ_y และ θ_x สำหรับให้ค่าจริงที่
 สมมูลกันแต่ละคู่ (ξ_x, ξ_y) ดังนั้นจึงมีความเป็นไปได้ที่จะเทียบคะแนนจริงจากแบบสอบต่างชุด
 โดยคำนวณค่าคงที่ของการเทียบคะแนน α และ β ตามวิธีการที่ได้กล่าวมาข้างต้น หลังจากนั้น
 จะสามารถพล็อตกราฟเพื่อเทียบคะแนนจริงระหว่างแบบสอบได้ ดังแสดงในภาพประกอบที่ 2

(Hambleton. 1985 : 213; Lord. 1980 : 199 - 200)



ภาพประกอบ 2 การเทียบคะแนนจริงตามทฤษฎีการตอบข้อสอบ ระหว่างแบบสอบสองชุด คือ
 ชุด X และชุด Y

การประเมินความเพียงพอของวิธีการเทียบคะแนน

การเทียบคะแนนวิธีการใดก็ตามจะให้ผลดีที่สุด เมื่อคะแนนที่ได้จากแบบสอบที่ต้องการเทียบ มีคุณสมบัติเป็นไปตามเงื่อนไขต่าง ๆ ที่ได้กำหนดไว้ของวิธีการเทียบแต่ละวิธีนั้นอย่างเคร่งครัด แต่ในสถานการณ์จริงอาจมีข้อจำกัดบางประการซึ่งทำให้ไม่สามารถได้ข้อมูลตรงตามเงื่อนไขดังกล่าว เช่น ข้อมูลไม่เป็นไปตามข้อตกลงของวิธีการเทียบนั้น ๆ หรือไม่สามารถควบคุมให้สถานการณ์ในการสอบเป็นไปตามรูปแบบในการเก็บรวบรวมข้อมูลได้อย่างครบถ้วน ดังนั้นจึงมีความจำเป็นที่จะต้องทำการตรวจสอบความเพียงพอของการเทียบคะแนน หรือประเมินประสิทธิภาพของวิธีการเทียบคะแนนที่แตกต่างกัน (Petersen, Marco and Stewart. 1982 : 72, 91) ซึ่งมีผู้เสนอวิธีการในการตรวจสอบความเพียงพอของการเทียบคะแนนไว้ดังนี้

1. ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ (percentile comparison index)

โคเลน และวิทนี (Kolen and Whitney. 1982 : 284) ได้แนะนำให้ใช้ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์เพื่อประเมินความเพียงพอของการเทียบคะแนน ทั้งนี้โดยอาศัยกลุ่มตัวอย่างตรวจสอบผล ซึ่งเป็นกลุ่มอิสระอีกกลุ่มหนึ่งที่สุ่มมาจากประชากรเดียวกันกับกลุ่มผู้สอบที่ใช้สร้างตารางการเทียบคะแนน วิธีการหาค่าดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์คือการวัดความไม่สอดคล้องระหว่างการแจกแจงคะแนนที่ได้จากกลุ่มตรวจสอบผล คือ คะแนนจากแบบสอบชุด X กับคะแนนซึ่งได้มาจากการแปลงคะแนนจากแบบสอบชุด Y ไปยังสเกลของแบบสอบชุด X ตามตารางการเทียบคะแนนที่ได้จากวิธีการที่ใช้เทียบโดยหาค่าสมมูล ณ ที่ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน ดัชนีดังกล่าวคำนวณได้จาก

$$C = \frac{\sum_{i=1}^n (x_i - x_i^*)^2}{n \cdot k}$$

เมื่อ n คือ จำนวนคะแนนดิบในกลุ่มตรวจสอบผล

k คือ จำนวนข้อในแบบสอบชุด X

ถ้าค่า C น้อยกว่าแสดงว่าวิธีการเทียบคะแนนที่นำมาใช้สร้างตารางการแปลงคะแนนนั้น มีความเพียงพอที่จะนำมาใช้ให้ได้ผลการเทียบที่คงที่มากกว่า

2. ดัชนีความแตกต่าง (discrepancy indices)

ปีเตอร์เซน มาร์โค และสจวร์ต (Petersen, Marco and Stewart. 1982 : 91) ได้เสนอวิธีการประเมินความเพียงพอของการเทียบคะแนนด้วยการใช้ดัชนีความแตกต่าง ซึ่งมีแนวคิดว่าจะแนบคิแต่ละตัวจากแบบสอบชุด X จะมีคะแนนเกณฑ์ t และคะแนนแปลง t' ที่ได้จาก การประมาณค่า โดยใช้วิธีการเทียบคะแนนเฉพาะอันหนึ่ง ค่าความแตกต่าง d ระหว่าง t' และ t เป็นความคลาดเคลื่อนในการเทียบคะแนน ซึ่งค่าที่น้อยกว่าแสดงว่ารูปแบบการเทียบคะแนนนั้น มีความเหมาะสมมากกว่า

ค่าดัชนีความแตกต่าง คือ ค่าความคลาดเคลื่อนรวมในการเทียบคะแนนของวิธีการเทียบ คะแนนที่ใช้นั้น ซึ่งคำนวณจากค่าเฉลี่ยยกกำลังสองของความแตกต่างระหว่างคะแนนแปลงกับคะแนน เกณฑ์ และถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ ซึ่งเท่ากับค่าความแปรปรวนของความ แตกต่างรวมกับค่ากำลังสองของความลำเอียง

$$\sum_j f_j d_j^2 / ns_t^2 = \sum_j f_j (d_j - d)^2 / ns_t^2 + d^2 / s_t^2$$

(total error) = (variance of difference) + (squared bias)

เมื่อ $d_j = (t_j - t'_j)$
 $t'_j =$ คะแนนแปลงสำหรับคะแนนคิ x_j
 $t_j =$ คะแนนเกณฑ์สำหรับคะแนนคิ x_j
 $s_j =$ ความเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ t_j
 $f_j =$ ความถี่ของ x_j
 $n = \sum_j f_j =$ จำนวนคะแนนที่ใช้
 $d = \sum_j f_j d_j / n$

ดัชนีเปรียบเทียบความแตกต่างนี้เป็นค่ามาตรฐาน (แสดงในรูปสัดส่วนของความเบี่ยงเบน มาตรฐานของคะแนนเกณฑ์ ค่าความคลาดเคลื่อนรวมที่ได้นี้จึงสามารถเปรียบเทียบกันได้โดยตรง ระหว่างค่าที่ได้จากการใช้รูปแบบและสถานการณ์การเทียบที่แตกต่างกัน (Petersen, Marco and Stewart. 1982 : 91)

งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการเทียบคะแนนเท่าที่ผ่านมานั้น ส่วนใหญ่ทำการเปรียบเทียบผลการเทียบคะแนนโดยใช้วิธีการเทียบคะแนนแบบต่าง ๆ โดยอาจพิจารณาจากตัวแปรที่แตกต่างกัน เช่น กลุ่มผู้สอบมาจากการสอบหรือไม่ได้มาจากการสอบ ระดับความสามารถของกลุ่มผู้สอบคล้ายคลึงกันหรือไม่แตกต่างกัน ลักษณะการกระจายของคะแนนคล้ายคลึงกันหรือไม่แตกต่างกัน เป็นต้น นอกจากนี้ ได้มีการศึกษาที่เน้นเชิงสถิติ ได้แก่ การศึกษาความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนแบบต่าง ๆ ความคงเส้นคงวา (stability) ของผลการเทียบคะแนน ความแรงต่อการฝ่าฝืนข้อตกลงเบื้องต้นของวิธีการเทียบคะแนนแต่ละวิธี ตลอดจนการประเมินความเพียงพอของวิธีการเทียบคะแนนที่นำมาใช้เป็นต้น ซึ่งผู้วิจัยได้นำมาเสนอตามลำดับดังนี้

คาแกน และสตอค (Kagan and Stock. 1980 : 34 - 37) ได้ศึกษาคะแนนที่เท่าเทียมกันของแบบสอบ MAT (Miller Analogies Test) กับแบบสอบ GRE (Graduate Record Examination) โดยทำการเทียบคะแนนด้วยวิธีเชิงเส้นตรง และวิธีกระดกถอย ซึ่งทั้งสองวิธีมีสมการในการแปลงคะแนนต่างกัน ดังนี้

$$\text{วิธีการเชิงเส้นตรง ใช้สมการ} \quad Y = \frac{S_y}{S_x} x + y - \frac{S_y}{S_x} x$$

$$\text{วิธีการกระดกถอย ใช้สมการ} \quad Y = r \frac{S_y}{S_x} x + y - r \frac{S_y}{S_x} x$$

ซึ่งถ้า r_{xy} เท่ากับ 1 แล้วสมการทั้งสองคือสมการเดียวกัน ดังนั้นวิธีการทั้งสองวิธีจะให้ผลเหมือนกัน กลุ่มตัวอย่างที่ใช้ได้จากผลการสอบแบบสอบมาตรฐานปี 1970 - 1976 ของมหาวิทยาลัยแอริโซนา (Arizona State University) ซึ่งมี 18,017 คน โดยเลือกกลุ่มตัวอย่างตามลักษณะของประชากรมาจำนวน 154 คน ซึ่งสอบแบบสอบทั้งสองชุด ผลการเทียบ

คะแนนทั้งสองวิธีพบว่า ผลการแปลงด้วยวิธีการเชิงเส้นตรงแตกต่างจากวิธีการกระดกน้อย โดยคะแนนที่ห่างจากกึ่งกลางจะมีความแตกต่างกันเพิ่มขึ้น ทั้งนี้เป็นเพราะวิธีกระดกน้อยจะได้คะแนนแปลงเข้าใกล้ค่าเฉลี่ยมากกว่าวิธีการเชิงเส้นตรง ด้วยเหตุนี้ผู้ที่สอบได้คะแนนต่ำกว่าค่าเฉลี่ยจะได้ประโยชน์มากขึ้นขณะที่ผู้ได้คะแนนสูงกว่าค่าเฉลี่ยจะได้คะแนนแปลงที่ไม่ได้ประโยชน์

ฮัทเทน (Hutten. 1982 : 4799-A) ได้ศึกษาความเหมาะสมของข้อมูลจริงกับรูปแบบของทฤษฎีการตอบข้อสอบ 2 รูปแบบ คือ รูปแบบของราสส์ กับรูปแบบ 3 พารามิเตอร์ ในการประมาณค่าพารามิเตอร์ความสามารถและความยากของข้อสอบ สำหรับกลุ่มตัวอย่างขนาดเล็ก (250 คน) และแบบสอบสั้น (20 ข้อ) ผลการศึกษาพบว่ารูปแบบราสส์และรูปแบบ 3 พารามิเตอร์ มีความเหมาะสมกับข้อมูลประมาณ ร้อยละ 80 ทั้งสองรูปแบบ เมื่อเปรียบเทียบคะแนนที่ทำนายกับคะแนนที่วัดได้โดยใช้ค่าสถิติ Kolmogorov-Smirnov แล้ว รูปแบบของราสส์เหมาะสมกับข้อมูลทั้งหมดดีกว่ารูปแบบ 3 พารามิเตอร์ ผู้วิจัยได้สนับสนุนให้ใช้รูปแบบ 3 พารามิเตอร์ กับกลุ่มตัวอย่างที่มีขนาดตั้งแต่ 1,000 คนขึ้นไป เพื่อจะได้ความแม่นยำในการประมาณค่าพารามิเตอร์ต่าง ๆ ผลการวิจัยนี้สอดคล้องกับที่โดแกลสส์ (Douglass. 1981 : 4000-A) ได้ศึกษาเปรียบเทียบการใช้รูปแบบของทฤษฎีการตอบข้อสอบกับผลการสอบในระดับห้องเรียนโดยใช้รูปแบบของทฤษฎีการตอบข้อสอบประมาณค่าพารามิเตอร์ของข้อสอบ 100 ข้อ ที่เป็นข้อสอบปลายปีของ 4 ปีการศึกษา จากกลุ่มตัวอย่างระดับวิทยาลัย จำนวน 594 - 1,082 คน จัดข้อสอบแบ่งเป็น 4 ชุด จากนั้นทำการสุ่มข้อสอบชุดละ 43 - 53 ข้อ สุ่มผู้สอบจำนวน 200, 600 คน และ 800 คน มาวิเคราะห์ปรากฏว่า การวิเคราะห์ข้อสอบด้วยรูปแบบของราสส์มีความคงที่ของค่าพารามิเตอร์ดีกว่ารูปแบบ 2 พารามิเตอร์ และเมื่อมีการเทียบคะแนนของแบบสอบข้ามกลุ่มตัวอย่าง รูปแบบของราสส์ก็มีความคงที่ดีกว่าด้วย

โคเลน (Kolen. 1981 : 1 - 11) ได้เปรียบเทียบผลการเทียบคะแนนระหว่างรูปแบบดั้งเดิมสองวิธี กับรูปแบบ TRT เจ็ดวิธี โดยใช้ข้อมูลจากผลการสอบนักเรียนในรัฐไอโอวา ปี 1978 ตามโครงการชื่อ The IOWA Tests of Educational Development แบบสอบที่ใช้ คือ แบบสอบที่พิมพ์ใหม่ครั้งที่ 7 เทียบไปสู่แบบสอบชุดเก่าที่พิมพ์ครั้งที่ 6 แบบสอบที่พิมพ์ครั้งที่ 6 มีหนึ่งชุด

คือ X6 ใช้สอนนักเรียนในทั้งสองระดับ คือ ระดับ 1 และระดับ 2 ซึ่งระดับ 1 เป็นนักเรียนเกรดที่ 9 และ 10 ระดับ 2 เป็นนักเรียนเกรด 11 และ 12 ส่วนแบบสอบที่พิมพ์ครั้งที่ 7 มีสองชุดที่คู่ขนานกัน คือ X7 และ Y7 แต่ละชุดแบ่งแบบสอบเป็น 2 ระดับ ซึ่งมีความยากแตกต่างกันเพื่อแยกสอนนักเรียนแต่ละคน โดยมีแผนการรวบรวมข้อมูลดังนี้ นักเรียนระดับ 1 สอบแบบสอบ X6, X7 ระดับ 1 และ Y7 ระดับ 1 ส่วนนักเรียนระดับ 2 สอบแบบสอบ X6, X7 ระดับ 2 และ Y7 ระดับ 2 ทั้งนี้ นักเรียนแต่ละคนจะได้รับแบบสอบเพียง 1 ชุด คือ X6, X7 หรือ Y7 โดยการสุ่ม ดังนั้นจำนวนนักเรียนที่สอบแบบสอบแต่ละชุดจะมี 1 ใน 3 ของจำนวนนักเรียนในแต่ละระดับ ส่วนแบบสอบ X6 จะมีนักเรียนจำนวน 1 ใน 3 ของจำนวนนักเรียนรวมสองระดับ สำหรับเกณฑ์ในการเปรียบเทียบผลใช้ค่าดัชนีจากกลุ่มตรวจสอบผล ซึ่งเป็นนักเรียนทุก ๆ คนที่ 3 จากการสอบแบบสอบแต่ละชุดและแต่ละระดับ

การเทียบคะแนนแบบดั้งเดิมใช้วิธีการเชิงเส้นตรงและอิกวิเปอร์เซนไทล์ซึ่งจัดกระทำตามวิธีของแองกอฟฟ์ ส่วนรูปแบบของทฤษฎีการตอบข้อสอบ (IRT) ใช้ชนิดหนึ่ง, สอง และสามพารามิเตอร์โลจิสติกโมเดล โดยแต่ละโมเดลเทียบ 2 แบบ คือ เทียบด้วยค่าประมาณคะแนนจริง และเทียบด้วยค่าประมาณคะแนนสังเกต ส่วนอีกวิธีหนึ่ง คือ ราสส์โมเดล

ดัชนีสำหรับตรวจสอบผล โคลเนใช้วิธีคำนวณจากค่าเฉลี่ยกำลังสองของความแตกต่าง (สำหรับผู้สอบแบบสอบ X6 ในกลุ่มตรวจสอบผล) ระหว่างคะแนนจากแบบสอบ X6 กับคะแนนจากแบบสอบ X7 และ Y7 ที่แปลงแล้ว ณ ที่ตำแหน่งเปอร์เซนไทล์เดียวกัน ทั้งนี้โดยใช้วิธีการเทียบคะแนนแบบต่าง ๆ ทั้งเก้าวิธี หากค่าดัชนีจากกลุ่มตรวจสอบผลที่ได้จากวิธีการเทียบใดมีค่าน้อยกว่า แสดงว่าวิธีนั้นมีความสอดคล้องระหว่างการแจกแจงคะแนนของแบบสอบชุดเก่ากับคะแนนของแบบสอบชุดใหม่ที่แปลงไปยังสเกลของชุดเก่ามากกว่า นั่นคือ วิธีนั้นมีความคงเส้นคงวามากกว่า ทั้งนี้โดยโคลเนทดสอบความแตกต่างของดัชนีที่ได้จากวิธีการเทียบคะแนนแบบต่าง ๆ เพื่อเปรียบเทียบผลของแต่ละวิธี โดยใช้ค่าสถิติทดสอบฟรายด์แมน (Friedman test)

ผลการวิจัยพบว่า วิธีการเทียบคะแนนแบบต่าง ๆ ให้ผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 เมื่อใช้แบบสอบระดับ 1 และ 2 ตามลำดับ ถ้าพิจารณาแบบสอบระดับที่ 1 พบว่า วิธีที่ให้ผลคะแนนสูงที่สุด คือ วิธีไออาร์ที (IRT) ชนิดสามพารามิเตอร์ ที่ประมาณค่าสังเกต ส่วนวิธีที่ให้ผลคะแนนน้อยที่สุด คือ วิธีการเชิงเส้นตรง ถัดขึ้นมาคือวิธีไออาร์ทีชนิดหนึ่งพารามิเตอร์ สำหรับแบบสอบระดับที่ 2 พบว่า วิธีการที่ให้ผลคะแนนสูงที่สุด คือวิธีไออาร์ทีชนิดสามพารามิเตอร์ที่ประมาณค่าคะแนนจริง วิธีการของราสซีย์ให้ผลคะแนนมากกว่าวิธีไออาร์ทีชนิดหนึ่งพารามิเตอร์ที่ประมาณค่าคะแนนจริง

โคเลนอภิปราชผลโดยส่วนรวมว่า วิธีไออาร์ทีชนิดหนึ่งพารามิเตอร์ให้ผลการเทียบที่ไม่เพียงพอ ทั้งนี้อาจเนื่องมาจากการเดาของผู้สอบมากเกินไปเมื่อจัดการเดาของผู้สอบโดยใช้วิธีการของไรท์ และสโตน (Wright and Stone, 1979) ซึ่งตัดข้อสอบที่พบว่าไม่เหมาะสมกับราสซีย์โมเดลออกจากผลการสอบ ผลการวิจัยสนับสนุนงานวิจัยของสไลด์ และลินน์ (Slinde and Linn, 1978 : 23 - 34; Slinde and linn, 1979 : 159 - 165) ที่พบว่าการเทียบคะแนนในแนวตั้ง โดยใช้แบบสอบที่มีความยากต่างกันสอบกับนักเรียนต่างระดับชั้นตามวิธีราสซีย์โมเดลให้ผลที่ไม่เพียงพอ ซึ่งกัสตาฟสัน (Gustafson, 1979 : 153 - 158) ได้ทำการศึกษาซ้ำกับกลุ่มตัวอย่างใหม่ และได้เสนอแนะว่าวิธีเทียบคะแนนโดยใช้ราสซีย์โมเดลกับแบบสอบที่มีความยากแตกต่างกันควรใช้กับกลุ่มตัวอย่างที่มีการเดาคำ

สำหรับวิธีเทียบคะแนนแบบไออาร์ทีชนิดสามพารามิเตอร์ให้ผลที่เพียงพอมากที่สุดยกเว้นการเทียบคะแนนที่อยู่ต่ำกว่าระดับการเดาซึ่งไม่สามารถอธิบายผลการเทียบคะแนนได้ ดังนั้นการเทียบคะแนนในส่วนนี้จึงต้องอาศัยวิธีการอื่น เช่น วิธีการเชิงเส้นตรง

ส่วนวิธีเทียบแบบอิกวิเปอร์เซนไทล์ พบว่าให้ผลที่เพียงพอที่จะนำมาใช้เทียบคะแนนแบบสอบที่มีความยากต่างกัน แต่วิธีเทียบเชิงเส้นตรงพบว่า ให้ผลที่ไม่น่าพอใจในกรณีนี้ ซึ่งจากผลการวิจัยของสไลด์ และลินน์ (Slinde and Linn, 1977 : 23 - 31) ก็ได้พบว่าการเทียบคะแนนแบบอิกวิเปอร์เซนไทล์ให้ผลดีกว่าการเทียบเชิงเส้นตรงในกรณีที่แบบสอบมีค่าความยากต่างกัน

คุก ดันบาร์ และไอเนอร์ (Cook, Dunbar and Eignor. 1981 : 1 - 52) ได้ศึกษาผลการเทียบคะแนนแบบไออาร์ที่เปรียบเทียบกับ การเทียบคะแนนแบบดั้งเดิม 2 วิธี คือ การเทียบคะแนนเชิงเส้นตรง และการเทียบคะแนนแบบอควิเปอร์เซนไทล์ การเทียบคะแนนแบบ ไออาร์ที่ใช้โมเดลแบบสามพารามิเตอร์ ซึ่งประมาณค่าพารามิเตอร์โดยใช้โปรแกรมโลจิส (LOGIST) ด้วยวิธีความเป็นไปได้สูงสุด (maximum likelihood procedure) สำหรับการ ประมาณค่าที่อยู่ต่ำกว่าระดับการเดาใช้วิธีการเชิงเส้นตรง (linear interpolation) ใน การเทียบคะแนนศึกษาทั้งสองกรณีคือมีแบบสอบรวม และไม่มีแบบสอบรวม โดยใช้กลุ่มตัวอย่างที่ไม่ได้ สุ่มจากประชากรเดียวกัน ทั้งนี้เพื่อศึกษาความเป็นไปได้ (feasibility) ของวิธีการไออาร์ที่ ในการนำไปแก้ปัญหาทางปฏิบัติ การเปรียบเทียบพิจารณา 2 ประการ คือ ความสอดคล้องสัมพัทธ์ (relative agreement) ระหว่างวิธีไออาร์ที่กับวิธีดั้งเดิมแต่ละวิธี โดยพิจารณาจากการพล็อต กราฟ อีกประการหนึ่งคือ คำนึงความแตกต่างสำหรับการแจกแจงคะแนนรวม และแต่ละส่วนของ การแจกแจง (สูงกว่า 20 เปอร์เซนต์, ตรงกลาง 60 เปอร์เซนต์ และต่ำกว่า 20 เปอร์เซนต์ ของการแจกแจง) ซึ่งคำนวณจากค่าที่ถ่วงน้ำหนักของค่าเฉลี่ยกำลังสองของความแตกต่างระหว่าง ค่าประมาณจากการเทียบคะแนนแต่ละวิธีที่ศึกษากับคะแนนเกณฑ์ เกณฑ์ในที่นี้คือคะแนนแปลงด้วย การเทียบคะแนนแบบไออาร์ที่ชนิดสามพารามิเตอร์ ซึ่งงานวิจัยที่มีมาก่อน (Kolen. 1981 : 1 - 11; Slind and Linn. 1977 : 23 - 31) แนะนำว่ามีความเหมาะสมในการเทียบ คะแนนในกรณีที่มีแบบสอบมีความยากต่างกัน สอดคล้องกับกลุ่มตัวอย่างที่มีความสามารถต่างกัน

ข้อมูลที่ใช้ในการศึกษาครั้งนี้ เป็นผลการสอบ 2 ครั้งที่ดำเนินการแล้ว โดย The Collage Board Admissions Testing Program ซึ่งใช้แบบสอบ 2 ชุด ที่มีความแตกต่างกันในความยาว และความยาก แต่ละชุดมีข้อสอบ 2 ตอน คือ ข้อสอบภาษาและคณิตศาสตร์ใช้สอบกับกลุ่มตัวอย่างที่ไม่ได้เป็นกลุ่มเท่าเทียมกันโดยการสุ่ม

ผลการศึกษาพบว่าวิธีการเทียบคะแนนแบบดั้งเดิม กับวิธีเทียบแบบไออาร์ที่มีความสอดคล้อง กันมาก เนื่องจากโดยความจริงแล้วการแจกแจงคะแนนดิบจากแบบสอบทั้งสองชุดมีรูปร่างคล้ายคลึง

กันมาก จึงทำให้รูปแบบเชิงเส้นตรงและเชิงเส้นโค้งหรืออควิเปอร์เซนไทล์มีผลใกล้เคียงกัน โดยการเทียบเชิงเส้นโค้งมีผลใกล้เคียงกับการเทียบแบบไออาร์ที่มากกว่าการเทียบเชิงเส้นตรง อย่างไรก็ตามจากความแตกต่างที่เกิดขึ้นที่ส่วนปลายของวิธีการเทียบแบบดั้งเดิมทุกวิธีที่ต่างไปจากการเทียบแบบไออาร์ที่เป็นผลจากข้อมูลส่วนนี้มีน้อยและหายากมาก ในทางทฤษฎีจึงแนะนำให้ใช้การเทียบแบบไออาร์ที่เพราะไม่มีผลเนื่องจากขาดแคลนข้อมูลในส่วนปลายการแจกแจง อย่างไรก็ตาม การเทียบแบบไออาร์ที่ก็ไม่สามารถเตรียมการเทียบคะแนนในส่วนที่อยู่ทางปลายด้านล่างที่ต่ำกว่าระดับการเคาได้ ซึ่งจำเป็นต้องใช้วิธีอื่นประมาณค่าเพิ่มเติม นอกจากนี้ผลการวิจัยพบว่าการเทียบคะแนนด้วยวิธีไออาร์ที่โดยตรงเมื่อไม่มีแบบสอบร่วม และเมื่อกลุ่มตัวอย่างไม่ได้เป็นกลุ่มที่เท่าเทียมกัน โดยการสุ่มนั้น ให้ผลแตกต่างจากการเทียบเมื่อมีแบบสอบร่วมเพียงเล็กน้อย (Cook, Dunbar and Eignor. 1981 : 50)

ปีเตอร์เสน มาร์โก และสจ๊วต (Petersen, Marco and Stewart. 1982 : 71 - 185) ได้ศึกษาความเพียงพอของการเทียบคะแนนเชิงเส้นตรงหลาย ๆ วิธี และวิธีอควิเปอร์เซนไทล์ ในสถานการณ์ต่าง ๆ โดยใช้ข้อมูลจากผลการสอบระดับชาติสองครั้ง คือ ครั้งที่สองเมื่อเดือนเมษายน 1975 เป็นแบบสอบชุด SAT (Scholastic Aptitude Test) ซึ่งมี 3 ฉบับ คือ ee, fe และ fm อีกครั้งหนึ่งสอบเมื่อเดือนพฤศจิกายน 1975 เป็นแบบสอบชุด TSWE (Test of Standard Written English) ซึ่งมี 3 ฉบับ คือ eg, fg และ fo ซึ่งบันทึกผลการตอบข้อสอบรายชื่อของผู้สอบทั้งหมดไว้ในคอมพิวเตอร์ กลุ่มตัวอย่างสร้างขึ้นตามลักษณะเฉพาะที่ได้กำหนดไว้ โดยทำการสุ่มมาจำนวน 6 กลุ่ม แต่ละกลุ่มมีจำนวนผู้สอบ 4,731 คน แบ่งแต่ละกลุ่มออกเป็น 3 กลุ่มย่อยโดยไม่ให้ซ้ำกัน ทั้งนี้ให้มีลักษณะต่างกันดังนี้ กลุ่มสุ่มอย่างง่าย กลุ่มที่มีความสามารถคล้ายคลึงกัน (ใช้ค่าเฉลี่ยความสามารถทางภาษาเป็นเกณฑ์) และกลุ่มที่มีความสามารถไม่คล้ายคลึงกัน ดังนั้นจึงมีกลุ่มย่อยทั้งหมด 54 กลุ่ม การออกแบบการเทียบคะแนนใช้แบบสอบร่วมลักษณะต่าง ๆ กัน กับแบบสอบที่ต้องการเทียบ คือ ต่างกันในตำแหน่งของแบบสอบร่วม (ชนิดภายนอกและภายใน) เนื้อหา

(คล้ายคลึงกันและไม่คล้ายคลึงกัน) และความยาก (คล้ายคลึงกันและไม่คล้ายคลึงกัน) โดยทำการเทียบแบบสอบชุด SAT ส่วนที่เป็นภาษากับตัวเองเหมือนกับว่าเป็นการเทียบแบบสอบสองชุด โดยใช้แบบสอบร่วมลักษณะต่าง ๆ เพื่อศึกษาระดับความคล้ายคลึงระหว่างแบบสอบที่ต้องการเทียบกับแบบสอบร่วม และระดับความคล้ายคลึงระหว่างกลุ่มตัวอย่างสองกลุ่มที่ใช้ในการเทียบที่มีต่อผลการเทียบคะแนน ทั้งนี้การจักรกระทำตัวแปรต่าง ๆ ดำเนินการอย่างเป็นระบบ นอกจากนี้ได้มีการเทียบคะแนนในระหว่างแบบสอบ (SAT กับ TSWE) โดยใช้แบบสอบร่วมภายใน เพื่อศึกษาระดับความคล้ายคลึงระหว่างแบบสอบที่ต้องการเทียบทั้งสองชุด และระดับของความคล้ายคลึงระหว่างกลุ่มตัวอย่างที่ใช้เทียบสองกลุ่มที่มีผลต่อการเทียบคะแนนโดยแบบสอบที่เทียบมีลักษณะแตกต่างกันด้านต่าง ๆ คือ เนื้อหา ความยากและความยาวของแบบสอบ การวิเคราะห์ข้อสอบทำการเลือกคะแนนตามจุดมุ่งหมายแล้วคำนวณค่าสถิติรายข้อและข้อมูลอื่น ๆ ที่ต้องการใช้เพื่อเทียบแบบสอบในสถานการณ์ต่าง ๆ

เกณฑ์ที่ใช้ในการประเมินความเพียงพอของวิธีการเทียบคะแนนใช้ค่าดัชนีความแตกต่างซึ่งเป็นค่าที่ถ่วงน้ำหนักแล้วของค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างค่าประมาณหรือคะแนนแปลงกับคะแนนเกณฑ์ คะแนนเกณฑ์ที่ใช้คือคะแนนแปลงจากคะแนนดิบชุดเดียวกันด้วยรูปแบบการเทียบคะแนนแบบไออาร์ที ดัชนีนี้เป็นค่ามาตรฐานโดยอยู่ในรูปสัดส่วนของความเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ ดังนั้นจึงสามารถเปรียบเทียบผลการเทียบคะแนนระหว่างสถานการณ์ต่าง ๆ และระหว่างวิธีการเทียบคะแนนแบบต่าง ๆ ได้ ทั้งนี้ค่าดัชนีความแตกต่างหรือความคลาดเคลื่อนซึ่งคำนวณได้จากสมการแสดงอยู่ในสเกลที่มีค่าเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์เท่ากับ 100 และจำแนกตามระดับคุณภาพสำหรับประเมินความเพียงพอของวิธีการเทียบคะแนนดังนี้

$TE < 25$	คือ	น่าพอใจอย่างมาก	(Zone A)
$25 < TE < 100$	คือ	น่าพอใจ	(Zone B)
$100 < TE < 225$	คือ	ปานกลาง	(Zone C)
$225 < TE < 400$	คือ	ไม่น่าพอใจ	(Zone D)
$400 < TE$	คือ	ไม่น่าพอใจอย่างมาก	(Zone E)

ระดับต่าง ๆ ของความคลาดเคลื่อนรวมนี้กำหนดขึ้นจากเหตุผลที่ว่า ถ้าความคลาดเคลื่อนรวมเนื่องมาจากความลำเอียงเพียงอย่างเดียวแล้ว ผลการเทียบคะแนนจะอยู่ในระดับที่น่าพอใจอย่างมาก นั่นคือค่าความคลาดเคลื่อนรวมเท่ากับความลำเอียง แต่ถ้าค่าความคลาดเคลื่อนรวมเป็นผลมาจากความแตกต่างระหว่างค่าประมาณหรือคะแนนแปลงกับเกณฑ์หรือค่าจริงด้วยแล้ว โดยทั่วไปเป็นที่ยอมรับกันว่าค่าเฉลี่ยของความแตกต่างไม่เกินร้อยละ 5 ของส่วนเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์จึงจะถือว่าผลของการเทียบคะแนนน่าพอใจอย่างมาก และจะไม่เป็นที่ยอมรับถ้าค่าเฉลี่ยของความแตกต่างถึงร้อยละ 20 ของส่วนเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ ซึ่งผู้วิจัยกล่าวว่าจุดคิดนี้เป็นวิธีการที่มีเหตุผลทำให้สามารถทำการเปรียบเทียบผลการเทียบคะแนนจากวิธีการที่ต่างกันและในสถานการณ์ต่าง ๆ กันได้

ผลการวิจัยสรุปได้ดังนี้

1. รูปแบบการเทียบคะแนนเชิงเส้นตรงโดยใช้เทคนิคของพอททอฟ (Potthoff) ให้ผลที่น่าพอใจในเกือบทุกสถานการณ์ที่ศึกษา
2. กรณีที่กลุ่มตัวอย่างขนาดเท่ากัน รูปแบบอควิเปอร์เซนไทล์ให้ผลที่มีความคลาดเคลื่อนรวมมากกว่ารูปแบบเชิงเส้นตรง
3. รูปแบบการเทียบคะแนนเชิงเส้นตรงโดยเทคนิคของทักเกอร์ 1 (Tucker 1) ให้ผลการเทียบที่มีความคลาดเคลื่อนรวมมากกว่าวิธีเชิงเส้นตรงอื่น ๆ ในกรณีที่กลุ่มตัวอย่างแตกต่างกัน ยกเว้นสำหรับรูปแบบเชิงเส้นตรงโดยใช้เทคนิคของพอททอฟ
4. รูปแบบอควิเปอร์เซนไทล์ให้ผลการเทียบดีกว่ารูปแบบเชิงเส้นตรง เมื่อแบบทดสอบสองชุดมีความสัมพันธ์เชิงเส้นโค้งอันเนื่องมาจากแบบทดสอบมีความยากแตกต่างกันและจากกลุ่มตัวอย่างที่ไม่คล้ายคลึงกัน
5. ค่าความยากระหว่างแบบสอบรวมและแบบสอบที่เทียบคะแนนแตกต่างกันจะทำให้ความคลาดเคลื่อนในการเทียบมีมากกว่าความแตกต่างในเนื้อหา
6. ค่าความยากที่ต่างกันระหว่างแบบสอบทั้งสองชุดที่จะเทียบคะแนนจะทำให้ความคลาดเคลื่อนในการเทียบมีมากกว่าความแตกต่างในเนื้อหา

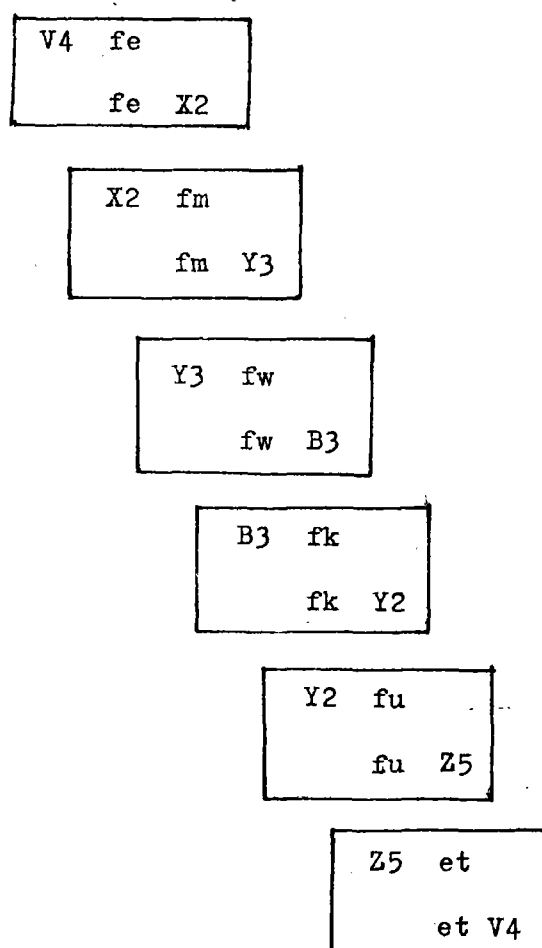
7. แบบสอบร่วมที่สร้างให้มีลักษณะคล้ายคลึงกับแบบสอบที่จะเทียบคะแนนอย่างมาก หรืออาจเรียกว่าเป็นฉบับย่อส่วนแล้ว การเทียบคะแนนจะให้ผลดีที่สุด

ผู้วิจัยได้ให้ข้อสังเกตว่า การสรุปผลการวิจัยเพื่อนำไปใช้ทั่วไปต้องทำอย่างรอบคอบโดยพิจารณาจากความเหมาะสมของรูปแบบ สถานการณ์ และลักษณะของกลุ่มตัวอย่าง

โคเลน และวิทนี (Kolen and Whitney. 1982 : 279 - 293) ได้เปรียบเทียบความเพียงพอของวิธีการเทียบคะแนน 4 วิธี คือ วิธีอีควิเปอร์เซนไทล์ วิธีเชิงเส้นตรง วิธีไออาร์ที ชนิดหนึ่งพารามิเตอร์โลจิสติกโมเดล หรือราสซ์โมเดล และวิธีไออาร์ที ชนิดสามพารามิเตอร์โลจิสติกโมเดล โดยใช้แบบสอบ GED (Tests of General Educational Development) ซึ่งเป็นแบบสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียนจำนวน 5 วิชา แต่ละวิชามีเนื้อหาแตกต่างกันดังนี้ ทักษะการเขียน (80 ข้อ) การศึกษาสังคม (60 ข้อ) วิทยาศาสตร์ (60 ข้อ) ทักษะการอ่าน (40 ข้อ) และคณิตศาสตร์ (50 ข้อ) ซึ่งใช้สำหรับตัดสินให้ประกาศนียบัตรแก่นักเรียนระดับอุดมศึกษาทั่วประเทศที่มีความรู้เทียบเท่าระดับที่แสดงว่าได้เรียนสำเร็จแล้ว แบบสอบแต่ละวิชามี 12 ชุด คำเนิการสอบผู้สอบกลุ่มละ 2 ชุด และทุกกลุ่มต้องสอบแบบสอบชุดที่ 12 ซึ่งแยกออกมาเป็นแบบสอบร่วม การจัดลำดับแบบสอบให้กับแต่ละกลุ่มใช้วิธีสมดุลย์ (counterbalance) ทำการเทียบคะแนนแต่ละคู่ของแบบสอบในแนวนอน ด้วยวิธีการเทียบ 4 วิธีดังกล่าว

กลุ่มตัวอย่างสุ่มจากการสอบในปี 1980 ซึ่งมีจำนวนมากกว่า 800,000 คน โดยสุ่มมาขนาดเล็ก กลุ่มละประมาณ 200 คน คิดเป็นสัดส่วนโรงเรียนละ 22 คนด้วยวิธีสุ่มแบบแบ่งชั้น (stratified random sampling) ตามประเภทของโรงเรียน สภาพทางภูมิศาสตร์และสถานภาพทางสังคม กลุ่มตรวจสอบผลก็ใช้วิธีสุ่มเช่นเดียวกัน โดยสุ่มมาจากผู้สอบแต่ละวิชา และแต่ละชุดให้ได้ผู้สอบ 20 คน (ประมาณร้อยละ 10 ของผู้เข้าสอบทั้งหมด) การเทียบคะแนนเชิงเส้นตรงและอีควิเปอร์เซนไทล์ กระทำตามวิธีของแองกอฟฟ์ ส่วนวิธีการไออาร์ทีประมาณค่าพารามิเตอร์โดยใช้โปรแกรมโลจิสเฉพาะแบบสอบร่วมก่อนจากนั้นใช้ค่าพารามิเตอร์ความสามารถของผู้สอบที่กำหนดให้คงที่ในการประมาณค่าพารามิเตอร์จากแบบสอบอีก 11 ชุด แล้วสร้างตารางคะแนนสมมูลระหว่าง

ลอร์ด และวิงเกอร์สกี (Lord and Wingersky. 1984 : 453 - 461) ได้
 เปรียบเทียบการเทียบคะแนนระหว่างรูปแบบไออาร์ทีที่เทียบโดยใช้คะแนนจริง และรูปแบบไออาร์ที
 ที่เทียบคะแนนสังเกตโดยใช้ไอคิวเปอร์เซ็นต์ไทล์ จากแบบสอบ SAT ด้านภาษา จำนวน 6 ชุด เทียบ
 เข้าสู่กันในลักษณะใช้โดยใช้แบบสอบร่วม ทั้งนี้แบบสอบชุดแรกและชุดสุดท้ายเป็นชุดเดียวกัน
 การออกแบบรวบรวมข้อมูลดำเนินการดังภาพ



จากภาพแบบสอบ V4 เทียบไปสู่แบบสอบ X2 โดยใช้แบบสอบร่วม fe ส่วนแบบสอบ X2 เทียบไปสู่ Y3 โดยใช้แบบสอบร่วม fm ซึ่งจะได้ผลการเทียบจาก V4 ไป Y3 กระทำการเทียบคะแนนในลักษณะเดียวกันไปยังแบบสอบชุดอื่น สุดท้ายจะได้ผลการเทียบ V4 กับ V4 ซึ่งเป็นการเทียบแบบสอบกับตัวเอง และใช้เป็นเกณฑ์ในการเปรียบเทียบความคงเส้นคงวาของวิธีการเทียบคะแนนทั้ง 2 วิธี ทั้งนี้เกณฑ์นี้แตกต่างจากที่โคเลน (Kolen. 1981 : 1 - 11) ได้เคยใช้ศึกษาความคงเส้นคงวาของวิธีการเทียบคะแนนแบบต่าง ๆ โคเลนใช้เกณฑ์จากกลุ่มตรวจสอบผลซึ่งเป็นกลุ่มอิสระอีกกลุ่มหนึ่ง โดยผู้วิจัยกล่าวไว้ว่าเกณฑ์ที่โคเลนใช้ไม่ใช่เกณฑ์ที่มีคุณสมบัติสำหรับการเลือกวิธีการเทียบคะแนนที่ดีที่สุด เพราะวิธีการเทียบคะแนนที่ไม่ถูกต้องอาจให้ผลการเทียบคงเส้นคงวามากกว่าวิธีการที่ถูกต้อง เกณฑ์การเทียบแบบสอบกับตัวเองที่ใช้ในการศึกษานี้อาจไม่เที่ยงตรงในเชิงทฤษฎี และไม่น่าพอใจในการปฏิบัติสำหรับการเทียบแบบสอบสองฉบับที่แตกต่างกัน แต่เป็นเกณฑ์ที่มีประโยชน์ต่อการเปลี่ยนวิธีการเทียบคะแนนใหม่หากพบว่าคะแนนที่แปลงแล้วในขั้นสุดท้ายไม่เป็นอย่างเดียวกันกับคะแนนเดิมของตัวเอง

ผลการวิจัยยังไม่ได้มีการตัดสินใจทางทฤษฎีที่ชัดเจนในการเลือกใช้วิธีการเทียบคะแนนแบบไฮอาร์ที่ที่ใช้คะแนนจริงกับแบบไฮอาร์ที่ที่ใช้คะแนนสังเกตโดยอควิเปอร์เซนไทล์ เนื่องจากทั้งสองวิธีมีความสอดคล้องกัน

โดรานส์ และคิงสตัน (Dorans and Kingston. 1985 : 249 - 262) ได้ศึกษาผลของการฝ่าฝืนเกี่ยวกับความมีมิติเดียว ในการประมาณค่าพารามิเตอร์ข้อสอบ และพารามิเตอร์ความสามารถ โดยใช้การเทียบคะแนนแบบไฮอาร์ที่ชนิดสามพารามิเตอร์กับแบบสอบวัดความถนัดชุด GRE ด้านภาษาจำนวน 4 ชุด แต่ละชุดวัด 2 ด้าน คือ ความเข้าใจในการอ่าน และหลักภาษา การออกแบบเทียบคะแนนมี 2 แบบ คือ เทียบคะแนนรวมทั้งชุดซึ่งมีความเป็นวิวิธพันธ์ (heterogeneous) และเทียบคะแนนรวมแต่ละด้านซึ่งมีความเป็นเอกพันธ์ (homogeneous) โดยออกแบบในการรวบรวมข้อมูล เมื่อใช้กลุ่มที่เท่าเทียมกัน และเมื่อใช้แบบสอบร่วม จำนวนกลุ่มตัวอย่างมีตั้งแต่ 2,579 ถึง 4,351 คน

ผลการวิจัยพบว่า การเทียบคะแนนแบบไออาร์ที่เมื่อผ่านข้อตกลงเกี่ยวกับแบบสอบต้องมีมิติเดียวนั้นอาจมีผลกระทบ (impact) ต่อการเทียบ แต่ไม่อาจยืนยันได้ เนื่องจากมีความคล้ายคลึงกันระหว่างผลการเทียบเมื่อข้อสอบเป็นเอกพันธ์ และวิวิธพันธ์ ดังนั้นผู้วิจัยจึงแนะนำว่าการเทียบโดยใช้ไออาร์ที่อาจมีความแม่นยำเพียงพอต่อการผ่านการมีมิติเดียว

ฟิลลิปส์ (Phillips . 1986 : 107 - 118) ได้ศึกษาผลการเทียบคะแนนในแนวตั้งโดยใช้รหัสโมเดล เพื่อหาคำตอบว่า

1. การตัดคนที่ไม่เหมาะสมในการวิเคราะห์ออกไปแล้วจะช่วยปรับปรุงข้อมูลให้เหมาะสมกับโมเดลขึ้นหรือไม่ โดยเปรียบเทียบจากค่าอำนาจจำแนกที่เหมาะสมกับโมเดล เมื่อก่อนและหลังการตัดคนที่ไม่เหมาะสม โดยใช้โปรแกรมไออาร์ที่ประมาณค่าพารามิเตอร์
2. การตัดคนที่ไม่เหมาะสมในการวิเคราะห์ จะมีผลที่แตกต่างในการเทียบคะแนนแบบสอบ 2 ชุดหรือไม่ โดยเปรียบเทียบจากค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าประมาณความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบที่มีค่าสถิติที่เหมาะสม ค่าความเชื่อมั่นที่พิจารณาจากความสอดคล้องภายใน (internal consistency reliability) และจำนวนข้อสอบที่เหมาะสมกับโมเดลระหว่างก่อนและหลังจากตัดคนที่ไม่เหมาะสมในการวิเคราะห์
3. การตัดคนที่ไม่เหมาะสมในการวิเคราะห์ จะเปลี่ยนผลการเทียบคะแนนโดยใช้รหัสโมเดลหรือไม่ โดยเปรียบเทียบผลการเทียบคะแนนโดยใช้รหัสโมเดลก่อนและหลังการตัดคนที่ไม่เหมาะสมกับผลการเทียบคะแนนแบบอีควิเปอร์เซนไทล์

ข้อมูลที่ใช้ได้มาจากการสอบแบบสอบมาตรฐานระดับชาติเมื่อปี 1980 ซึ่งใช้วัดหลายระดับชั้น เฉพาะแบบสอบย่อยต่างกัน 2 ด้าน คือ ด้านการอ่านและด้านคณิตศาสตร์ กลุ่มตัวอย่างคือนักเรียนระดับ 4 และระดับ 8 จำนวนระดับละ 300 - 500 คน การเทียบคะแนนจึงดำเนินการเมื่อแบบสอบมีพิสัยความยากและเนื้อหาที่กว้างกับนักเรียนกลุ่มตัวอย่างทั้งสองระดับ

ผลการวิจัยพบว่า

1. การตัดคนที่ไม่เหมาะสมในการวิเคราะห์ให้ผลน้อยในการปรับปรุงข้อมูล

2. การใช้วิธีการของราสส์ทั้งก่อนและหลังการตัดคนที่ไม่เหมาะสมกับโมเดล ให้คะแนนสมมูลคล้ายกัน เมื่อจำนวนร้อยละของคนที่ไม่เหมาะสมถูกตัดออกน้อย นอกจากนี้ข้อมูลของราสส์ชี้ว่าการตัดคนที่ไม่เหมาะสมในการวิเคราะห์ ให้สัดส่วนของการแจกแจงคะแนนผลสัมฤทธิ์ที่ต่ำกว่าเมื่อเกิดการเคามากขึ้น ค่าพารามิเตอร์ความยากของข้อสอบทั้งก่อนและหลังการตัดคนที่ไม่เหมาะสมมีแนวโน้มต่างกัน น้อยกว่าค่าเบี่ยงเบนมาตรฐานทั้งก่อนและหลังการตัดคนที่ไม่เหมาะสม

3. การใช้วิธีการของราสส์ทั้งก่อนและหลังจากตัดคนที่ไม่เหมาะสมในการวิเคราะห์ให้ผลการเทียบที่แตกต่างจากผลการเทียบแบบอควิเปอร์เช่นไพล์

ผู้วิจัยกล่าวว่า การตัดคนที่มีความสามารถต่ำอาจลดการเคา แต่บางครั้งอาจทำให้ค่าประมาณของโค้งลักษณะข้อสอบมีสัดส่วนของสเกลที่มีความถูกต้องน้อยลง

โคป (Cope. 1987 : 143 - 149) ได้เปรียบเทียบคะแนนเชิงเส้นตรง 5 วิธี ระหว่างวิธีของแองกอฟฟ์ (Angoff. 1971) ตามรูปแบบที่ 5 ซึ่งมี 3 วิธี กับวิธีของทักเกอร์ (Tucker) และเลวิน (Levine) ในกรณีที่แบบสอบมีความเชื่อมั่นเท่ากัน โดยใช้แบบสอบรวมแบบสอบที่ใช้เทียบคะแนนเป็นแบบสอบเพื่อให้ประกาศนียบัตรในวิชาชีพ จำนวน 5 ชุด สอบกับกลุ่มประชากรที่ไม่เท่าเทียมกัน 5 กลุ่ม กลุ่มละ 1 ชุด ปีละ 2 ครั้ง คือ แบบสอบชุด B, D และ E สอบในฤดูใบไม้ผลิ ส่วนแบบสอบชุด A กับชุด C สอบในฤดูฝน การเปรียบเทียบวิธีการเทียบคะแนนมี 2 แบบ คือ การเปรียบเทียบผลการเทียบคะแนนระหว่างแบบสอบแต่ละคู่ภายใต้วิธีการเชิงเส้นตรงที่แตกต่างกัน 5 วิธี ทั้งนี้โดยไม่เปรียบเทียบความถูกต้องของวิธีการต่าง ๆ เพราะว่สเกลคะแนนดิบของแบบสอบสองชุดถูกคาดว่าจะต่างกัน อีกแบบหนึ่งคือ การเปรียบเทียบผลการเทียบคะแนนแบบสายโซ่ ซึ่งเป็นการเปรียบเทียบคะแนนแบบสอบแต่ละชุดกับตัวเองโดยผ่านแบบสอบชุดอื่นที่เลือก ทั้งนี้เพื่อทำการเปรียบเทียบความถูกต้องของวิธีการเหล่านั้นด้วย

การวัดความคลาดเคลื่อนของการเทียบคะแนนใช้สถิติ 2 ตัว คือ ค่าประมาณความลำเอียง (estimated bias) และค่าประมาณที่ถ่วงน้ำหนักแล้วของรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนในการเทียบคะแนน (estimated weighted root squared error : RMSE) ซึ่งใช้เปรียบเทียบระหว่างวิธีที่แตกต่างกัน ซึ่งคำนวณจากสมการต่อไปนี้

$$\text{BIAS} = \bar{X}' - \bar{X}$$

เมื่อ $\bar{X}$ คือ ค่าเฉลี่ยของคะแนนดิบของแบบสอบชุดหนึ่ง

$\bar{X}'$ คือ ค่าเฉลี่ยของคะแนนดิบที่เท่าเทียมกันโดยการเทียบแบบสอบชุดนั้นไปยัง

ตัวเองแบบสายโซ่

$$\text{และ RMSE} = \frac{\sum n_i (\bar{x}'_i - \bar{x}_i)^2}{\sum n_i}$$

เมื่อ x_i คือ คะแนนดิบของคนที่ i

$\bar{x}'_i$ คือ คะแนนดิบที่เท่าเทียมกัน

n_i คือ จำนวนผู้สอบ

ในกรณีที่ผู้สอบจำนวนมาก การคำนวณ RMSE จะง่ายขึ้นโดยใช้สมการ

$$\text{RMSE} = [(m-1)^2 \text{Var}(X) + (\text{BIAS})^2]^{1/2}$$

เมื่อ m คือ ความชัน (slope) ของฟังก์ชันการเทียบ $x' = mx + b$ และ $\text{Var}(x)$

คือ ความแปรปรวนของ X

ผลการวิจัยสรุปได้ว่า ในสถานการณ์ที่ศึกษาภายใต้รูปแบบที่ 5 ตามวิธีของแองกอฟฟ์ ให้คะแนนดิบที่สมมูลกันใกล้เคียงกับวิธีการของหัตคเคอร์และเลวิน โดยการเทียบคะแนนแบบสายโซ่ทั้ง 5 วิธีให้ค่าประมาณความลำเอียงและรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนที่ใกล้เคียงกัน ส่วนการเทียบคะแนนแต่ละคู่ของแบบสอบพบว่าคะแนนดิบที่เท่าเทียมกันจากรูปแบบที่ 5 มีค่าอยู่ระหว่างคะแนนจากวิธีของหัตคเคอร์และเลวิน ผู้วิจัยกล่าวว่าผลการวิจัยนี้สนับสนุนวิธีการเทียบคะแนนเชิงเส้นตรงตามรูปแบบที่ 5 ซึ่งมีข้อตกลงที่มีข้อจำกัดน้อยกว่าวิธีของหัตคเคอร์และเลวิน ดังนั้นในทางปฏิบัติควรใช้วิธีการตามรูปแบบที่ 5 วิธีการใดวิธีการหนึ่งใน 3 วิธี

เขาวดี ราชชัยกุล (2518 : 87 - 94) ได้ทำการเทียบคะแนนผลการสอบในประเทศ
ไทย ที่กระทรวงศึกษาธิการได้จัดดำเนินการสอบให้กับนักเรียนในระบบโรงเรียนเมื่อเดือนมีนาคม

พ.ศ. 2516 และ พ.ศ. 2517 และการสอบที่จัดให้กับบุคคลทั่วไปที่อยู่นอกระบบโรงเรียน เมื่อเดือนสิงหาคม พ.ศ. 2516 และ พ.ศ. 2517 โดยมีจุดมุ่งหมายเพื่อสำรวจขอบเขตแบบสอบทั้ง 4 ครั้ง ซึ่งใช้แบบสอบต่างชุดกันนี้ว่ามีความเท่าเทียมกันเพียงใดที่จะนำมาใช้ตัดสินว่านักเรียนสมควรได้ผ่านชั้นมัธยมศึกษาปีที่ 5 หรือไม่ และเพื่อเสนอแนะวิธีที่ดีที่สุดสำหรับการเทียบคะแนนแบบสอบต่างชุด กลุ่มตัวอย่างที่ใช้คือคะแนนผลการสอบของนักเรียนแต่ละครั้ง ครั้งละ 500 คน เทียบคะแนนในวิชาภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ โดยใช้วิธีเทียบคะแนนเชิงเส้นตรง รูปแบบที่ 1 ที่เสนอโดยแองกอฟฟ์ (Angoff) ทำการเทียบคะแนนแบบสอบแต่ละชุดไปสู่สเกลร่วมที่มีช่วงคะแนน 1 - 175 ค่าเฉลี่ย 100 ค่าเบี่ยงเบนมาตรฐาน 25

ผลการศึกษาพบว่าคะแนนจากแบบสอบทั้ง 4 ชุด ไม่เท่าเทียมกันทุกวิชา กล่าวคือ จุดตัดของแบบสอบแต่ละชุดไม่สามารถเปรียบเทียบกันได้ เช่น การสอบเมื่อเดือนมีนาคม พ.ศ. 2517 พบว่าคะแนนจุดตัดสำหรับการผ่านในสเกลร่วมเท่ากับ 88 แต่การสอบเมื่อเดือนมีนาคม พ.ศ. 2516 มีคะแนนจุดตัดสำหรับการผ่านในสเกลร่วมเท่ากับ 78 และภายหลังจากการตรวจสอบสัดส่วนการผ่านของผู้สอบแล้ว พบว่านักเรียนที่ควรสอบได้แต่ได้รับการตัดสินให้ตกมีจำนวนน่าสนใจ ตัวอย่างเช่น การเทียบคะแนนโดยใช้แบบสอบเมื่อเดือนมีนาคม พ.ศ. 2516 เป็นหลัก พบว่ามีจำนวนนักเรียนถึงร้อยละ 37.6 ที่ตกเมื่อเดือนสิงหาคม พ.ศ. 2516 และร้อยละ 9 ที่ตกเมื่อเดือนมีนาคม พ.ศ. 2517 ซึ่งควรเป็นผู้สอบได้ ผู้วิจัยได้เสนอแนะวิธีการซึ่งควรใช้ในการเทียบคะแนนในสถานการณ์การสอบนี้ว่าควรใช้รูปแบบที่ 4 ตามที่แองกอฟฟ์ได้เสนอไว้ นั่นคือรูปแบบที่มีข้อสอบร่วม

ซูซีพ พงษ์สมบูรณ์ (2528 : 112 - 118) ได้ทำการเปรียบเทียบประสิทธิภาพและความคงที่ของการเทียบคะแนน 3 วิธี คือ การเทียบคะแนนเชิงเส้นตรง การเทียบคะแนนแบบอีควิเปอร์เซนไทล์ และการเทียบคะแนนโดยใช้โด่งลักษณะข้อสอบ เมื่อใช้รูปแบบที่ใช้แบบสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วม การเปรียบเทียบประสิทธิภาพของการเทียบคะแนนพิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน ส่วนการเปรียบเทียบความคงที่ของวิธีการเทียบคะแนน (ความไม่แปรผันไปตามกลุ่มตัวอย่าง ความสมมาตร และความเสมอภาค) นั้นทำการทดสอบนัยสำคัญ

ของความแตกต่างโดยใช้สถิติทดสอบโคลโมโกรอฟ-สมิรโนฟ (Kolmogorov-Smirnov two sample test) กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2516 ทั่วประเทศ ที่เข้าสอบตามโครงการตรวจสอบคุณภาพการศึกษาของกรมวิชาการ กระทรวงศึกษาธิการ ซึ่งสุ่มมาจำนวน 3,721 คน แบ่งศึกษาเป็นกลุ่มย่อย 6 กลุ่ม กลุ่มละ 620 คน แต่ละกลุ่มมีความสามารถใกล้เคียงกัน แบบสอบที่ใช้มีข้อสอบทั้งหมด 60 ข้อ จัดแบ่งเป็น 2 ชุด ชุดละ 38 ข้อ ที่มีข้อสอบรวมภายในจำนวน 20 ข้อ แบบสอบทั้งสองชุดนี้มีความยากใกล้เคียงกัน

ผลการวิจัยพบว่าประสิทธิภาพของการเทียบคะแนนระหว่างรูปแบบที่ใช้แบบสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วมด้วยวิธีการที่ต่างกัน 3 วิธีไม่แตกต่างกัน ส่วนผลการเปรียบเทียบความคงที่ของวิธีการเทียบคะแนนโดยรูปแบบที่ใช้ผู้สอบร่วมทั้ง 3 วิธี พบว่าการเทียบโดยใช้คิวเปอร์เซนไทล์มีความคงที่มากกว่าการเทียบเชิงเส้นตรง และมีความคงที่พอ ๆ กับการเทียบโดยใช้คั้งลักษณะข้อสอบซึ่งความคงที่ของวิธีการเทียบคะแนนโดยใช้รูปแบบที่ใช้ผู้สอบร่วมนี้ให้ผลเหมือนกับการใช้รูปแบบที่ใช้แบบสอบร่วมทั้ง 3 วิธี

ภาวณี ศรีสุขวัฒนานันท์ (2529 : 155 - 170) ได้เปรียบเทียบผลของการใช้รูปแบบการเทียบคะแนนที่ต่างกัน 3 รูปแบบ คือ รูปแบบคิวเปอร์เซนไทล์ รูปแบบเชิงเส้นตรง และรูปแบบอิงทฤษฎีการตอบข้อสอบแบบสามพารามิเตอร์ โดยใช้แบบสอบร่วมภายในที่ต่างกัน 3 ขนาด คือ ขนาดร้อยละ 20 (7 ข้อ), ขนาดร้อยละ 40 (14 ข้อ) และขนาดร้อยละ 60 (21 ข้อ) โดยมีข้อสอบทั้งหมด 100 ข้อ จัดแบ่งเป็น 2 ชุด ชุดละ 35 ข้อ ซึ่งทั้งสองมีความยากง่ายใกล้เคียงกัน ที่เหลือจัดเป็นแบบสอบร่วมขนาดต่าง ๆ ดังกล่าว สำหรับกลุ่มตัวอย่างแยกเป็น 2 กรณี คือ กรณีแบบสอบคัดเลือก 2 กลุ่ม กลุ่มละ 1,500 คน และกลุ่มตรวจสอบผลอีก 1 กลุ่ม จำนวน 1,500 คน อีกกรณีหนึ่งคือกรณีแบบวัดผลสัมฤทธิ์ ซึ่งมีจำนวนกลุ่มตัวอย่างเช่นเดียวกับกรณีแรก ในการประเมินผลการเทียบคะแนนมี 2 ลักษณะ คือ

1. เปรียบเทียบปริมาณความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนด้วยวิธีการที่ต่างกันของแบบสอบแต่ละกรณี
2. ตรวจสอบความเพียงพอของวิธีเทียบคะแนนแต่ละวิธี โดยใช้ค่าดัชนีเปรียบเทียบความแตกต่าง (index C) ที่ได้จากกลุ่มตรวจสอบผล

ผลการวิจัยพบว่า การใช้แบบสอบร่วมที่ยาวกว่าให้ประสิทธิภาพของการเทียบคะแนนที่สูงกว่า (หรือมีความคลาดเคลื่อนมาตรฐานในการเทียบคะแนนน้อยกว่า) และให้ผลในระดับที่น่าพอใจมากกว่า ทั้งสองกรณีของแบบสอบ

การศึกษารูปแบบของการเทียบคะแนน พบว่าในกรณีของแบบสอบคัดเลือกนั้น รูปแบบที่ให้ผลที่มีความเพียงพอมากที่สุด คือ รูปแบบอีควิเปอร์เซนไทล์ รองลงมาคือ รูปแบบทฤษฎีการตอบข้อสอบ และรูปแบบเชิงเส้นตรงตามลำดับ ซึ่งให้ผลที่แตกต่างกับในกรณีของแบบสอบวัดผลสัมฤทธิ์ที่พบว่า รูปแบบที่ให้ผลที่มีความเพียงพอมากที่สุด คือ รูปแบบเชิงเส้นตรง รองลงมาคือรูปแบบอีควิเปอร์เซนไทล์ และรูปแบบทฤษฎีการตอบข้อสอบตามลำดับ และเมื่อทดสอบความแตกต่างของค่าดัชนี C เป็นรายคู่ พบว่ากรณีแบบสอบวัดผลสัมฤทธิ์วิธีการที่ใช้รูปแบบเชิงเส้นตรงมีความเพียงพอมากกว่าวิธีที่ใช้รูปแบบทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ .05 ส่วนการเปรียบเทียบรูปแบบคู่อื่นไม่สามารถสรุปความแตกต่างอย่างมีนัยสำคัญ

ผู้วิจัยอภิปรายผลว่ารูปแบบของการเทียบคะแนนที่ให้ความเพียงพอต่างกันกรณีของแบบสอบคัดเลือก และกรณีของแบบสอบวัดผลสัมฤทธิ์ อาจเนื่องมาจากปัจจัยด้านแบบทดสอบและปัจจัยด้านประชากรผู้สอบ ในด้านแบบทดสอบนั้นกรณีแบบสอบคัดเลือกเป็นแบบสอบวัดความรู้ ความสามารถทั่วไปในเชิงความถนัด มีพิสัยระดับความยากพอสมควร หลังจากจัดแบ่งเป็นสองชุดแล้วมีความเชื่อมั่นใกล้เคียงกัน ด้านประชากรผู้สอบนั้นเป็นกลุ่มที่มีการกระจายความสามารถกว้าง ดังนั้นการเทียบคะแนนในกรณีรูปแบบอีควิเปอร์เซนไทล์จึงให้ผลดีที่สุด เมื่อใช้แบบสอบร่วมขนาดเดียวกัน ส่วนในกรณีของแบบสอบวัดผลสัมฤทธิ์นั้น แบบสอบมีขอบเขตเนื้อหาที่จำกัดภายในกรอบของหลักสูตรและประชากรของผู้สอบได้ผ่านการเรียนการสอนในวิชาที่สอบมาแล้ว จึงทำให้คะแนนมีพิสัยที่แคบกว่ากรณีแบบสอบคัดเลือก แต่เนื่องจากกลุ่มตัวอย่างสุ่มทั้งกลุ่มมีการแจกแจงของคะแนนคล้ายคลึงกัน ดังนั้นจึงพบว่าการเทียบรูปแบบเชิงเส้นตรงให้ผลดีกว่ารูปแบบอีควิเปอร์เซนไทล์

เรวดี อินทสระ (2530 : 39, 64 - 69) ได้เปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน และความเพียงพอของวิธีการเทียบคะแนน 2 วิธี คือ วิธีเทียบคะแนนโดยใช้

รูปแบบอิงทฤษฎีการตอบข้อสอบ รูปแบบที่ใช้พารามิเตอร์สามตัวกับรูปแบบที่ใช้เทคนิคการวิเคราะห์องค์ประกอบ กลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2529 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดพัทลุง โดยนำแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ เรื่องอัตราส่วนร้อยละ ซึ่งสร้างขึ้นจำนวน 150 ข้อ ไปทดลองกับกลุ่มตัวอย่างจำนวน 1,266 คน แล้วนำข้อสอบที่มีคุณภาพจากการวิเคราะห์ผลการทดลองมาจัดแบ่งเป็นแบบสอบ 2 ชุด ชุดละ 30 ข้อ และแต่ละชุดมีแบบสอบรวมอีก 15 ข้อ นำไปดำเนินการสอบกลุ่มตัวอย่างอื่น จำนวน 1,557 คน โดยทุกคนต้องทำแบบสอบทั้ง 2 ชุด หลังจากนั้นจัดสุ่มใช้คะแนนผลการสอบจากแบบสอบชุดที่ 1 จำนวน 779 คน และอีก 778 คน ใช้คะแนนจากแบบสอบชุดที่ 2 ส่วนคะแนนที่เหลือใช้เป็นคะแนนตรวจสอบผลของการเทียบคะแนนแต่ละรูปแบบ ประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้โปรแกรม LOGIST 5 และวิเคราะห์องค์ประกอบโดยใช้โปรแกรมสำเร็จรูป SPSSX

ผลการวิจัยพบว่า การเทียบคะแนนรูปแบบอิงทฤษฎีการตอบข้อสอบให้ความเพียงพอในระดับที่น่าพอใจ ส่วนรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบให้ความเพียงพอในระดับปานกลาง สำหรับการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนเมื่อพิจารณาแบบสอบทั้งสองชุดพบว่า ความคลาดเคลื่อนของรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบมีความคลาดเคลื่อนสูงกว่ารูปแบบอิงทฤษฎีการตอบข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (t-test) แต่เมื่อพิจารณาในแต่ละช่วงคะแนนพบว่าแบบทั้งสองให้คะแนนสมมูลใกล้เคียงกันในช่วงตรงกลางของการแจกแจงคะแนน (คะแนน 19 - 23) ส่วนช่วงคะแนนต่ำและคะแนนสูง พบว่าวิธีทั้งสองให้ผลแตกต่างในทางตรงกันข้าม กล่าวคือในช่วงคะแนนต่ำ (คะแนน 1 - 18) นั้นรูปแบบอิงทฤษฎีการตอบข้อสอบมีคะแนนสมมูลสูงกว่ารูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ ส่วนช่วงคะแนนสูง (คะแนน 24 - 35) รูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบมีคะแนนสมมูลสูงกว่ารูปแบบการใช้ทฤษฎีการตอบข้อสอบ

อาภรณ์ กาญจนกิจโสภณ (2531 : 66 - 67) ได้ศึกษาการสร้างแบบทดสอบและตารางเทียบระดับคะแนนตามแนวคิดในวิชาคณิตศาสตร์ เรื่องสมการและอสมการโดยใช้วิธีเทียบคะแนน

แบบราชส์สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น กลุ่มตัวอย่างที่ใช้เทียบคะแนนเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 จำนวนระดับชั้นละ 490, 482 และ 415 คนตามลำดับ ผู้วิจัยได้นำแบบสอบที่สร้างขึ้นระดับชั้นละ 90 ข้อไปทดลองกับกลุ่มตัวอย่างอื่นแล้วคัดเลือกข้อสอบที่มีคุณภาพเรียงตามค่าความยากที่ต้องการระดับชั้นละ 40, 30 และ 30 ข้อ โดยแต่ละชั้นจะมีข้อสอบร่วมระหว่างชั้นอยู่ 10 ข้อ หลังจากนั้นนำไปทดสอบกับกลุ่มตัวอย่างที่ใช้เทียบคะแนน และสร้างตารางเทียบระดับคะแนนเพื่อใช้ศึกษาพัฒนาการของนักเรียนกลุ่มเดิมที่ผ่านขึ้นไปเรียนในชั้นสูงขึ้นว่า นักเรียนมีความก้าวหน้าเป็นไปตามปกติหรือเบี่ยงเบนไปจากปกติ

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังกล่าวมา พอสรุปได้ดังนี้

1. ผลการศึกษาเท่าที่ผ่านมา พบว่าไม่มีวิธีการเทียบคะแนนวิธีใดเพียงวิธีเดียวที่ให้ผลดีที่สุดในทุกสถานการณ์ ดังนั้นในทางปฏิบัติจำเป็นต้องศึกษาข้อกำหนดต่าง ๆ ของวิธีการเทียบคะแนนแต่ละวิธี ได้แก่ ลักษณะการแจกแจงของคะแนนที่ได้จากแบบสอบต่างชุด ลักษณะของแบบสอบ และลักษณะของกลุ่มตัวอย่าง เป็นต้น ทั้งนี้เพื่อเป็นแนวทางในการตัดสินใจเลือกวิธีการในการเทียบคะแนนที่เหมาะสมในแต่ละสถานการณ์การเทียบ ซึ่งจากผลการวิจัยดังกล่าวมาพอสรุปได้ว่า

1.1 เมื่อแบบสอบต่างชุดที่นำมาเทียบคะแนนมีระดับความยากคล้ายคลึงกันและกลุ่มตัวอย่างเป็นกลุ่มที่เท่าเทียมกันแล้ว ควรใช้วิธีการเชิงเส้นตรงซึ่งสะดวกในการแปลงคะแนนและมีความเป็นปรนัยมากกว่า แต่ผลการเทียบคะแนนจะขึ้นอยู่กับกลุ่มตัวอย่างที่ใช้

1.2 เมื่อแบบสอบมีความยากต่างกัน กลุ่มตัวอย่างเป็นกลุ่มที่เท่าเทียมกันแล้ว การใช้วิธีการเทียบแบบอควิเปอร์-เซนไทล์ หรือการเทียบเชิงเส้นโค้งจะเหมาะสมกว่า แต่ถ้ากลุ่มตัวอย่างเป็นกลุ่มที่ไม่เท่าเทียมกันแล้ว รูปร่างของการแจกแจงคะแนนที่ได้จากแบบสอบต่างชุดจะแตกต่างกันมาก ในกรณีนี้ผลการเทียบคะแนนจะมีความคลาดเคลื่อนมาก

1.3 เมื่อใช้แบบสอบที่มีความยากต่างกัน และกลุ่มตัวอย่างไม่ได้เป็นกลุ่มที่เท่าเทียมกัน ควรใช้วิธีการแบบไออาร์ทีชนิดสามพารามิเตอร์ซึ่งให้ผลที่น่าพอใจกว่าวิธีอื่น

1.4 เมื่อกลุ่มตัวอย่างค่อนข้างน้อย และแบบสอวัดในสิ่งเดียวกันแล้ว วิธีการเชิงเส้นตรง และวิธีการโออาร์ทีชนิดหนึ่งพารามิเตอร์ให้ผลที่น่าพอใจกว่าวิธีอีควิเปอร์เซนไทล์ และโออาร์ทีชนิดสามพารามิเตอร์

2. การประเมินความเพียงพอของผลที่ได้จากการเทียบคะแนน ว่าวิธีการเทียบคะแนนวิธีใดมีความแน่นอนหรือคงเส้นคงวามากกว่ากันนั้น ในงานวิจัยโดยทั่วไปใช้ค่าดัชนีความแตกต่างซึ่งคำนวณจากค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างคะแนนที่แปลงแล้วด้วยวิธีการที่ศึกษากับคะแนนเกณฑ์ ค่าที่ได้นี้ถ่วงน้ำหนักด้วยความเบี่ยงเบนมาตรฐานของการแจกแจงคะแนนเกณฑ์ ดังนั้นจึงเป็นค่ามาตรฐานที่สามารถเปรียบเทียบระหว่างวิธีการที่ต่างกันได้ คำนี้นี้มีค่าน้อยแสดงว่าวิธีการเทียบคะแนนวิธีนั้นมีความคงเส้นคงวามากกว่า สำหรับคะแนนเกณฑ์ที่ใช้อาจได้มาจากการเทียบด้วยวิธีการอื่นที่เชื่อถือได้ ดังเช่น ปีเตอร์เสน มาร์โค และสจวร์ต (Peterson, Marco and STewart. 1982 : 71 - 135) ซึ่งใช้คะแนนที่ได้จากการเทียบด้วยวิธีการโออาร์ทีในข้อมูลเดียวกันกับการเทียบวิธีที่ศึกษาเป็นคะแนนเกณฑ์ เกณฑ์อีกอย่างหนึ่งคือ คะแนนจากกลุ่มตรวจสอบผลซึ่งเป็นกลุ่มตัวอย่างที่เท่าเทียมกัน ดังที่โคเลน (Kolen. 1981 : 1 - 11) และ โคเลน และ วิทนี (Kolen and Whitney. 1982 : 284) ได้ศึกษาโดยหาความแตกต่างระหว่างคะแนนที่แปลงแล้วกับคะแนนก่อนแปลง ณ ที่ตำแหน่งเปอร์เซนไทล์เดียวกันในกลุ่มตรวจสอบผล นอกจากนี้ อาจใช้คะแนนดิบจากแบบสอบชุดแรกเป็นเกณฑ์โดยทำการเทียบคะแนนแบบสอบครั้งละ 2 ชุด ต่อเนื่องกันมากกว่า 2 ชุด ในลักษณะสายโซ่ เมื่อเทียบได้ 2 ถึง 3 ครั้ง ก็จะนำคะแนนที่ได้จากการเทียบครั้งสุดท้ายซึ่งเป็นการเทียบกับชุดแรกมาเปรียบเทียบกับคะแนนดิบจากแบบสอบชุดแรกก่อนที่จะแปลงคะแนน (Lord and Wingersgy. 1984 : 453 - 461) ความแตกต่างนี้จะเป็นดัชนีที่แสดงความเพียงพอในการเทียบคะแนนเช่นเดียวกัน

สำหรับในการวิจัยครั้งนี้ ผู้วิจัยจะศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในช่วง 3 ปีการศึกษาที่ผ่านมา โดยจะทำการเทียบคะแนนแบบสอบต่างชุดในกลุ่มทักษะที่ใช้ดำเนินการสอบ ตั้งแต่ปีการศึกษา 2529 ถึง 2531 ปีการศึกษาละ 1 ชุด ตามโครงการประเมินคุณภาพนักเรียน

ชั้นประถมศึกษาปีที่ 6 ระดับประเทศ ซึ่งได้สร้างแบบชุดใหม่เพื่อใช้ประเมินในทุกปีการศึกษา ตั้งแต่ปีการศึกษา 2527 เป็นต้นมา แต่ในปัจจุบันยังไม่ได้ทำการเทียบคะแนนเพื่อใช้ในการรายงานคะแนนผลการสอบดังกล่าว การศึกษาครั้งนี้ผู้วิจัยจะทำการเทียบคะแนนโดยใช้วิธีเทียบคะแนนรูปแบบราสส์ ซึ่งค่าพารามิเตอร์ของความสามารถและค่าความยากของข้อสอบไม่แปรผันไปตามกลุ่มตัวอย่างที่ใช้เทียบคะแนน ในขณะที่การเทียบคะแนนตามทฤษฎีดั้งเดิมไม่มีคุณสมบัติเช่นนี้ และวิธีการของราสส์นี้มีความซับซ้อนน้อยที่สุดในการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบข้อสอบ แต่จากงานวิจัยพบว่ามีความแกร่งที่จะนำมาใช้ให้ได้ผลในระดับที่น่าพอใจ

วิธีดำเนินการวิจัย

เนื่องจากสถานการณ์จริงในการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ตั้งแต่ปีการศึกษา 2529 ถึงปีการศึกษา 2531 ในครั้งนี้ ข้อมูลคะแนนได้มาจากการสอบที่ได้ดำเนินการสอบไปแล้ว ซึ่งใช้แบบสอบต่างชุดสอบกับนักเรียนต่างกลุ่มโดยไม่มีแบบสอบร่วมและไม่มีกลุ่มผู้สอบร่วม แต่จากแนวคิดของการเทียบคะแนนนั้นการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนดังกล่าวจำเป็นต้องมีข้อมูลที่น่ามาใช้เชื่อมค่าประมาณพารามิเตอร์จากแบบสอบแต่ละชุดให้อยู่ในสเกลเดียวกัน (Lord. 1982 : 145; Hambleton. 1985 : 198) ด้วยเหตุนี้ผู้วิจัยจึงอาศัยคุณสมบัติของการเทียบคะแนนตามทฤษฎีการตอบข้อสอบ ซึ่งจะให้ผลการเทียบคะแนนที่ไม่แปรผันไปตามคุณลักษณะของกลุ่มตัวอย่างที่นำมาวิเคราะห์ (Lord. 1980 : 199) โดยจะทำการรวบรวมข้อมูลจากกลุ่มตัวอย่างอื่นซึ่งมีข้อมูลในส่วนที่จะนำมาใช้เชื่อมค่าประมาณพารามิเตอร์จากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน ทั้งนี้โดยศึกษาผลการสอบในกรณีที่ใช้ผู้สอบร่วม หลังจากนั้นจะหาค่าคงที่ เพื่อใช้ในการสมการสำหรับแปลงค่าประมาณพารามิเตอร์จากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน สุดท้ายจะทำการเทียบคะแนนเพื่อให้ได้ตารางคะแนนสมมูลสำหรับการรายงานผลความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนของนักเรียนต่อไป ส่วนรายละเอียดของวิธีการวิจัยจะนำเสนอตามลำดับดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากรในการศึกษาครั้งนี้ คือ นักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 เมื่อปีการศึกษา 2529, 2530 และ 2531 ในสังกัดสำนักงานการประถมศึกษาจังหวัดปทุมธานี เฉพาะอำเภอที่นำมาศึกษา 4 อำเภอ จากอำเภอทั้งหมด 7 อำเภอ ซึ่งแต่ละปีการศึกษามีจำนวนประชากรจำนวน 3,875 คน 4,306 คน และ 3,973 คน ตามลำดับ

กลุ่มตัวอย่างที่ใช้ในการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในครั้งนี้ คือ กลุ่มตัวอย่างที่สุ่มไว้แล้วตามโครงการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศ เมื่อปีการศึกษา 2529, 2530 และ 2531 โดยผู้วิจัยจะนำมาศึกษาเฉพาะอำเภอที่มีข้อมูลผลการสอบรายชื่อนักเรียนรายบุคคลครบทั้ง 3 ปีการศึกษา ซึ่งมี 4 อำเภอ จากอำเภอทั้งหมด 7 อำเภอ จำนวนโรงเรียนและนักเรียนกลุ่มตัวอย่างในแต่ละปีการศึกษา ดังตาราง 2

ตาราง 2 จำนวนโรงเรียนและนักเรียนกลุ่มตัวอย่างในแต่ละปีการศึกษา จำแนกเป็นรายอำเภอ

ปีการศึกษา	อำเภอ	จำนวนโรงเรียน	จำนวนนักเรียน	ปีการศึกษา	อำเภอ	จำนวนโรงเรียน	จำนวนนักเรียน
2529	1. เมือง	6	184	2531	1. เมือง	5	109
	2. คลองหลวง	5	136		2. คลองหลวง	5	125
	3. ลาดหลุมแก้ว	4	99		3. ลาดหลุมแก้ว	5	103
	4. หนองเสือ	6	115		4. หนองเสือ	5	89
	รวม	21	534		รวม	20	467
2530	1. เมือง	5	119		รวมทั้งสิ้น	61	1454
	2. คลองหลวง	6	138	2532	1. ัญบุรี*	4	581
	3. ลาดหลุมแก้ว	5	87				
	4. หนองเสือ	4	109				
	รวม	20	453				

*คือ กลุ่มตัวอย่างที่ใช้ในการหาค่าคงที่ในการแปลงค่าพารามิเตอร์ของแบบสอบแต่ละชุดให้อยู่ในสเกลร่วมกัน ซึ่งสุ่มจากนักเรียนชั้นประถมศึกษาปีที่ 6 ที่กำลังเรียนภาคปลายปีการศึกษา 2532 ในอำเภอัญบุรี จังหวัดปทุมธานี จำนวน 1 กลุ่มโรงเรียน และทำการทดสอบเฉพาะโรงเรียนที่มีนักเรียนชั้นประถมศึกษาปีที่ 6 อย่างน้อย 2 ห้องเรียน ทั้งนี้แต่ละห้องจะต้องสอบแบบสอบ 2 ชุด โดยจัดลำดับก่อนหลังในการสอบต่างกันโดยวิธีสุ่ม

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการศึกษาครั้งนี้ คือ แบบสอบวัดผลสัมฤทธิ์ทางการเรียนที่สร้างขึ้นตามโครงการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศของสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ซึ่งใช้สอบนักเรียนกลุ่มตัวอย่างเมื่อปีการศึกษา 2529, 2530 และ 2531 ผู้วิจัยศึกษาในกลุ่มทักษะ (วิชาคณิตศาสตร์ และภาษาไทย) เฉพาะภาคความรู้ กลุ่มวิชาละ 3 ชุด ซึ่งแต่ละชุดวัดสมรรถภาพต่อไปนี้

1. คณิตศาสตร์ เป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 38 ข้อ วัด 3 สมรรถภาพ คือ

สมรรถภาพที่ 1 ความรู้ความเข้าใจเกี่ยวกับคณิตศาสตร์ 12 ข้อ

สมรรถภาพที่ 2 ทักษะการคิดคำนวณ 10 ข้อ

สมรรถภาพที่ 3 ทักษะการแก้ปัญหาโจทย์ 16 ข้อ

2. ภาษาไทย เป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก จำนวน 30 ข้อ วัด 3 สมรรถภาพ คือ

สมรรถภาพที่ 1 ความเข้าใจในการฟัง 8 ข้อ

สมรรถภาพที่ 2 ความเข้าใจในการอ่านจับใจความ 15 ข้อ

สมรรถภาพที่ 3 ความรู้เกี่ยวกับหลักเกณฑ์ทางภาษา 7 ข้อ

คุณภาพของแบบสอบ

คุณภาพของแบบสอบ แต่ละชุดเน้นความเที่ยงตรงเชิงเนื้อหา นั่นคือแบบสอบต้องวัดได้ตรงตามสมรรถภาพที่ต้องการวัด โดยสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ได้กำหนดขอบเขตของแต่ละสมรรถภาพไว้เหมือนกันทุกปีเพื่อให้เป็นแนวทางในการประเมินและสามารถเปรียบเทียบผลการสอบในต่างปีการศึกษาได้ สำหรับผู้สร้างแบบสอบประกอบด้วยนักวิชาการ และ

นักวัดผลการศึกษาจากหน่วยงานต่าง ๆ ได้แก่ วิทยาลัยครู กรมวิชาการ สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ตลอดจนศึกษานิเทศก์และครูผู้สอนจากจังหวัดต่าง ๆ ซึ่งจะทำการสร้างแบบสอบขึ้นทุกปี โดยการสร้างแบบสอบชุดใหม่ในปีการศึกษาต่อ ๆ มา ก็จะสร้างข้อคำถามใหม่ให้มีลักษณะคล้ายคลึงกับแบบสอบชุดเดิมให้มากที่สุดทั้งในด้านเนื้อหา และโครงสร้างของแบบสอบ เพื่อให้แบบสอบชุดใหม่ยังคงทำหน้าที่วัดความรู้ ความสามารถ และทักษะภายในขอบเขตอย่างเดียวกันกับแบบสอบชุดเดิม และจะดำเนินการสอบให้เป็นมาตรฐานเดียวกันทุกปีการศึกษา

การเก็บรวบรวมข้อมูล

การศึกษาค้นคว้าครั้งนี้ ใช้ข้อมูลจากการสอบในสถานการณ์จริงที่ได้ดำเนินการไปแล้วเมื่อปลายปีการศึกษา 2529, 2530 และ 2531 ถือเป็นผลการตอบข้อสอบรายข้อของนักเรียนรายบุคคลที่สอบ โดยสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ โดยมอบให้ศึกษานิเทศก์อำเภอทุกอำเภอเป็นผู้เก็บรวบรวมข้อมูลและบันทึกเฉพาะคะแนนรวมของนักเรียนรายบุคคล จำแนกตามกลุ่มวิชาและสมรรถภาพที่ต้องการวัดส่งไปให้เท่านั้น ดังนั้นผู้วิจัยจึงขออนุญาตใช้ข้อมูลกระดาษคำตอบของนักเรียนที่มีอยู่ที่สำนักงานการประถมศึกษาจังหวัดพุมธานี และบันทึกผลการตอบข้อสอบรายข้อของผู้สอบรายบุคคลโดยใช้รหัสดังนี้ ถ้าตอบถูกต้องให้เป็น 1 ถ้าตอบไม่ถูกต้องให้เป็น 0 และถ้าเว้นไม่ตอบให้เป็น 0

ส่วนการเก็บรวบรวมข้อมูลในกลุ่มตัวอย่างที่ใช้ในการหาค่าคงที่ในการแปลงค่าประมาณพารามิเตอร์นั้น เนื่องจากในกรณีที่แบบสอบที่ใช้ประกอบด้วยสมรรถภาพย่อย ๆ เช่นนี้ การใช้ข้อสอบรวม อาจไม่เป็นตัวแทนแบบสอบแต่ละชุด ดังนั้นผู้วิจัยจึงศึกษาจากผู้สอบรวมโดยนำแบบสอบในแต่ละกลุ่มวิชาไปทำการสอบนักเรียนกลุ่มตัวอย่าง โดยแต่ละกลุ่มจะสอบแบบสอบชุด X (แบบสอบที่ใช้ในปีการศึกษา 2529) และสอบแบบสอบอีกหนึ่งชุด คือ ชุด Y (แบบสอบที่ใช้ในปีการศึกษา 2530) หรือแบบสอบชุด Z (แบบสอบที่ใช้ในปีการศึกษา 2531) ทั้งนี้เพื่อให้ได้ข้อมูลผลการสอบเมื่อใช้ผู้สอบรวมสำหรับแบบสอบแต่ละชุดโดยผู้วิจัยจะทำการรวบรวมข้อมูลโรงเรียนละ 2 วัน ดังตาราง 3

ตาราง 3 แบบสอบที่ใช้สอบในโรงเรียนกลุ่มตัวอย่าง จำแนกตามกลุ่มวิชาและห้องเรียน

โรงเรียน	ห้องที่	จำนวน นักเรียน	แบบสอบ			
			คณิตศาสตร์		ภาษาไทย	
			วันที่ 1	วันที่ 2	วันที่ 1	วันที่ 2
1. ชุมชนประชาธิปไตยวิทยาการ	1	40	X	Y	X	Y
	2	40	Y	X	Y	X
	3	40	X	Z		
	4	40			Z	X
	5	40	X	Y		
	6	40			Y	X
2. ทองพูลอุทิศ	1	31	X	Z	X	Z
	2	31	Z	X	Z	X
3. รัษฎาสีทิธิศิลป์	1	37	X	Y	X	Y
	2	38	Y	X		
	3	37	X	Z		
	4	38			Z	X
4. วัดเขียนเขต	1	37			X	Y
	2	32	Y	X	Y	X
	3	31	X	Z	X	Z
	4	29	Z	X	Z	X

เนื่องจากผู้วิจัยทำการสอบนักเรียนเป็นรายห้อง ห้องละ 2 วัน ดังนั้นในกรณีที่นักเรียนขาดสอบหรือมีข้อมูลผลการสอบเพียงชุดเดียว ผู้วิจัยจะไม่นำมาวิเคราะห์ในครั้งนี้ และสำหรับข้อมูลผลการสอบของนักเรียนที่มีสมบูรณ์ทั้ง 2 ชุด ผู้วิจัยจะนำมาบันทึกลงในเทปในลักษณะของผลการตอบข้อสอบรายข้อรายบุคคล ซึ่งตอบข้อสอบทั้ง 2 ชุด โดยใช้รหัสเช่นเดียวกับการใช้รหัสผลการตอบข้อสอบในกลุ่มตัวอย่างที่ศึกษาความก้าวหน้า คือ ตอบถูกให้ 1 ตอบผิดให้ 0 และเว้นไม่ตอบให้ 0

การวิเคราะห์ข้อมูล

ผู้วิจัยจะดำเนินการวิเคราะห์ข้อมูล โดยแบ่งการวิเคราะห์ออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ในช่วงระยะเวลา 3 ปีการศึกษา โดยใช้การเทียบคะแนนรูปแบบราสส์ ผู้วิจัยจะใช้แบบสอบในปีการศึกษา 2529 (ชุด X) เป็นหลัก และทำการเทียบคะแนนจากแบบสอบในปีการศึกษา 2530 (ชุด Y) และแบบสอบในปีการศึกษา 2531 (ชุด Z) ไปสู่สเกลเดียวกับแบบสอบในปีการศึกษา 2529 ในลักษณะของการเทียบคะแนนจากแบบสอบที่สร้างขึ้นใหม่ทุกชุดไปยังสเกลเดียวกับแบบสอบชุดเก่า หลังจากนั้นผู้วิจัยจะรายงานผลสัมฤทธิ์ และความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนโดยใช้คะแนนจริง ซึ่งสะดวกต่อการนำไปใช้มากกว่าคะแนนความสามารถ (Hambleton. 1985 : 205) ทั้งนี้ผู้วิจัยจะทำการวิเคราะห์ข้อมูลเหมือนกันทั้งสองกลุ่มวิชา รวมทั้งจะศึกษาเป็นรายอำเภอและรวมทุกอำเภอด้วย โดยมีลำดับขั้นตอนของการวิเคราะห์ดังนี้

ขั้นที่ 1 หาค่าสถิติพื้นฐาน เช่น ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ (skewness) ค่าความโด่ง (kurtosis) และทำการแจกแจงความถี่ของคะแนนสอบเป็นรายกลุ่มวิชา ทั้งเมื่อจำแนกเป็นรายอำเภอและรวมทุกอำเภอ โดยใช้โปรแกรมคอมพิวเตอร์สำเร็จรูป SPSS ด้วยคำสั่ง frequency

ขั้นที่ 2 ประมวลค่าพารามิเตอร์ของแบบสอบแต่ละชุด ที่ได้จากผลการสอบในกลุ่มตัวอย่างที่ใช้ศึกษาค่าคงที่ในการเทียบคะแนน โดยใช้โปรแกรมคอมพิวเตอร์ BICAL ที่ศูนย์สารสนเทศ สำนักงานปลัดกระทรวงศึกษาธิการ ทั้งนี้เพื่อให้ได้ข้อมูลต่อไปนี้

2.1 การแจกแจงความถี่ของผู้ที่ได้คะแนนรวมแต่ละค่า (distribution of ability)

2.2 การแจกแจงความถี่และสัดส่วนของการตอบข้อสอบแต่ละข้อได้ถูก (item easiness distribution)

2.3 ค่าประมาณความยากของข้อสอบแต่ละข้อ (item difficulty, b) และค่าความคลาดเคลื่อนมาตรฐาน (standard error)

2.4 ค่าประมาณความสามารถ (log ability, θ) ของผู้สอบที่ได้คะแนนรวมแตกต่างกัน พร้อมทั้งค่าความคลาดเคลื่อนมาตรฐาน ค่าเฉลี่ยของค่าประมาณความสามารถและค่าความเบี่ยงเบนมาตรฐานของค่าประมาณความสามารถ

2.5 ค่าอำนาจจำแนกของข้อสอบ (biserial correlation) และค่าสถิติความเหมาะสมของโครงสร้างข้อสอบตามรูปแบบราสช์กับข้อมูลจริง

ขั้นที่ 3 คำนวณค่าคงที่ (α และ β) เพื่อใช้ในการแปลงค่าประมาณพารามิเตอร์จากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน โดยใช้วิธีหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน (mean and sigma method) ของค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแต่ละชุด ซึ่งคำนวณจากผลการสอบโดยกลุ่มผู้สอบร่วมหรือกลุ่มผู้สอบที่ทำแบบสอบทั้งสองชุดในกลุ่มตัวอย่างที่ใช้ในการหาค่าคงที่ของการเทียบคะแนน ทั้งนี้โดยใช้สมการเชิงเส้นตรง $y = \alpha x + \beta$ ในการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบชุด Y ไปยังสเกลเดียวกันกับค่าประมาณพารามิเตอร์ของผู้สอบแบบสอบชุด X ดังนี้ (Hambleton. 1985 : 207)

$$\theta_x = \alpha \theta_y + \beta$$

โดยที่ $\alpha = s_x / s_y$

$$\text{และ } \beta = \bar{X} - \bar{Y}$$

ในที่นี้ θ_x และ θ_y คือ ค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ

แบบสอบชุด X และชุด Y ตามลำดับ

s_x และ s_y คือ ค่าส่วนเบี่ยงเบนมาตรฐานของค่าประมาณพารามิเตอร์

ความสามารถของผู้สอบชุด X และชุด Y ตามลำดับ

$\bar{X}$ และ $\bar{Y}$ คือ ค่าเฉลี่ยของค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ

แบบสอบชุด X และชุด Y ตามลำดับ

ขั้นที่ 4 คำนวณค่าคงที่ในการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ

แบบสอบชุด Z ไปยังสเกลเดียวกันกับค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบชุด X ทั้งนี้โดยใช้วิธีการคำนวณในทำนองเดียวกันกับการคำนวณในขั้นที่ 3

ขั้นที่ 5 ทำการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบชุด Y ให้อยู่ในสเกลเดียวกันกับค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบชุด X โดยแทนค่าในสมการ $\theta_x = \alpha\theta_y + \beta$

ขั้นที่ 6 ทำการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบชุด Z ให้อยู่ในสเกลเดียวกันกับค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบชุด X ทั้งนี้โดยใช้วิธีการในทำนองเดียวกันกับขั้นที่ 5

ขั้นที่ 7 ในทำนองเดียวกันทำการแปลงค่าประมาณพารามิเตอร์ความยากของข้อสอบชุด Y และชุด Z ให้อยู่ในสเกลเดียวกันกับค่าประมาณความยากของข้อสอบชุด X

ขั้นที่ 8 คำนวณค่าความน่าจะเป็นที่ผู้สอบแต่ละระดับความสามารถจะตอบข้อสอบแต่ละข้อในแบบแต่ละชุดได้ถูกต้องโดยใช้สูตร

$$P_i(\theta) = \frac{\ell \cdot Da_i(\theta - b_i)}{1 + \ell \cdot Da_i(\theta - b_i)} \quad ; \quad i = 1, 2, 3, \dots, n$$

เมื่อ $P_i(\theta)$ คือ ความน่าจะเป็นที่ผู้สอบซึ่งมีความสามารถ θ จะตอบ
ข้อสอบข้อที่ i ในแบบสอบแต่ละชุดได้ถูกต้อง

b_i คือ ค่าความยากของข้อสอบข้อที่ i

a คือ ค่าเฉลี่ยของค่าอำนาจจำแนกของข้อสอบให้เท่ากับ 1

D คือ scaling factor มีค่า 1.7

θ คือ ระดับความสามารถของผู้สอบ

ℓ คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

การกำหนดค่า θ นี้ได้มาจากค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ
แบบสอบชุด X กับของผู้สอบชุด Y ที่แปลงไปสู่สเกลของแบบสอบชุด X แล้ว และค่าประมาณความ
สามารถของผู้สอบแบบสอบชุด X กับของผู้สอบแบบสอบชุด Z ที่แปลงไปสู่สเกลของแบบสอบชุด X
แล้ว สำหรับการคำนวณครั้งนี้ผู้วิจัยจะกำหนดค่า θ ให้ครอบคลุมพิสัยของค่าความสามารถตั้งแต่ค่า
 -4 ถึง $+4$ โดยแจกแจงให้ค่า θ แต่ละค่าแตกต่างกันเท่ากับ 0.2 เพื่อให้สะดวกในการเทียบค่า
ความสามารถ θ ระหว่างแบบสอบแต่ละชุด

ส่วนค่าความยากของข้อสอบนั้น คือ ค่าที่ประมาณได้จากแบบสอบแต่ละชุด ซึ่ง
ได้ทำการแปลงให้อยู่ในสเกลเดียวกันทั้งหมดแล้ว

ขั้นที่ 9 ประมวลค่าคะแนนจริงของผู้สอบแบบสอบแต่ละชุด ที่แต่ละระดับความสามารถ
โดยใช้สูตร

$$\xi_x = \sum_{i=1}^{\ell} P_i(\theta_x)$$

$$\xi_y = \sum_{j=1}^m P_j(\alpha\theta_x + \beta)$$

$$\xi_z = \sum_{k=1}^n P_k(\alpha\theta_x + \beta)$$

- เมื่อ ξ_x คือ คะแนนจริงของผู้สอบแบบสอบชุด X
 ξ_y คือ คะแนนจริงของผู้สอบแบบสอบชุด Y ซึ่งมีระดับความสามารถเท่าเทียมกันกับผู้สอบในแบบสอบชุด X
 ξ_z คือ คะแนนจริงของผู้สอบแบบสอบชุด Z ซึ่งมีระดับความสามารถเท่าเทียมกันกับผู้สอบในแบบสอบชุด X
 α, β คือ ค่าคงที่ (ค่าความชัน และจุดตัดแกนของสมการที่ใช้ในการแปลงค่าประมาณพารามิเตอร์จากแบบสอบแต่ละชุดให้อยู่ในสเกลเดียวกัน)

ขั้นที่ 10 พล็อตกราฟแสดงความสัมพันธ์ระหว่างคะแนนจริงจากแบบสอบแต่ละคู่ คือ ชุด X กับ Y และชุด X กับชุด Z ณ ตำแหน่งค่า 0 ที่เท่าเทียมกัน

ขั้นที่ 11 สร้างตารางเทียบคะแนนจริงระหว่างแบบสอบทั้ง 3 ชุด ในแต่ละกลุ่มวิชา

ขั้นที่ 12 คำนวณหาค่าเฉลี่ย และร้อยละนักเรียนที่มีผลน่าพอใจ จากการแจกแจงคะแนนจริงในแบบสอบแต่ละชุด

ขั้นที่ 13 ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในช่วง 3 ปีการศึกษา โดยพิจารณาจากปริมาณที่เพิ่มขึ้น หรือลดลงของค่าเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษา โดยใช้คะแนนจุดตัดการผ่านเกณฑ์ขั้นต่ำที่น่าพอใจที่ค่าเฉลี่ยร้อยละ 50 ของคะแนนเต็มเหมือนกันทุกปีการศึกษา และจะทำการศึกษาความก้าวหน้าของนักเรียนแต่ละอำเภอและรวมทุกอำเภอ

ตอนที่ 2 การเปรียบเทียบความแตกต่างของผลสัมฤทธิ์ทางการเรียนระหว่างปีการศึกษาต่าง ๆ เมื่อได้ทำการเทียบคะแนนแล้ว

ผู้วิจัยจะทดสอบนัยสำคัญของความแตกต่างระหว่างคะแนนเฉลี่ยผลการสอบระหว่าง 3 ปีการศึกษาเมื่อได้เทียบคะแนนจริงแล้ว โดยจะทำการวิเคราะห์ความแปรปรวน (Analysis of Variance : AVOVA) ทั้งนี้โดยคำนึงการเหมือนกันทั้งสองกลุ่มวิชา

ตอนที่ 3 การเปรียบเทียบความแตกต่างของสัดส่วนจำนวนนักเรียนที่ผ่านคะแนนจุดตัดของการผ่านเกณฑ์ที่น่าพอใจ ระหว่างสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์เมื่อเทียบคะแนนแล้วกับสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์เมื่อยังไม่ได้เทียบคะแนน

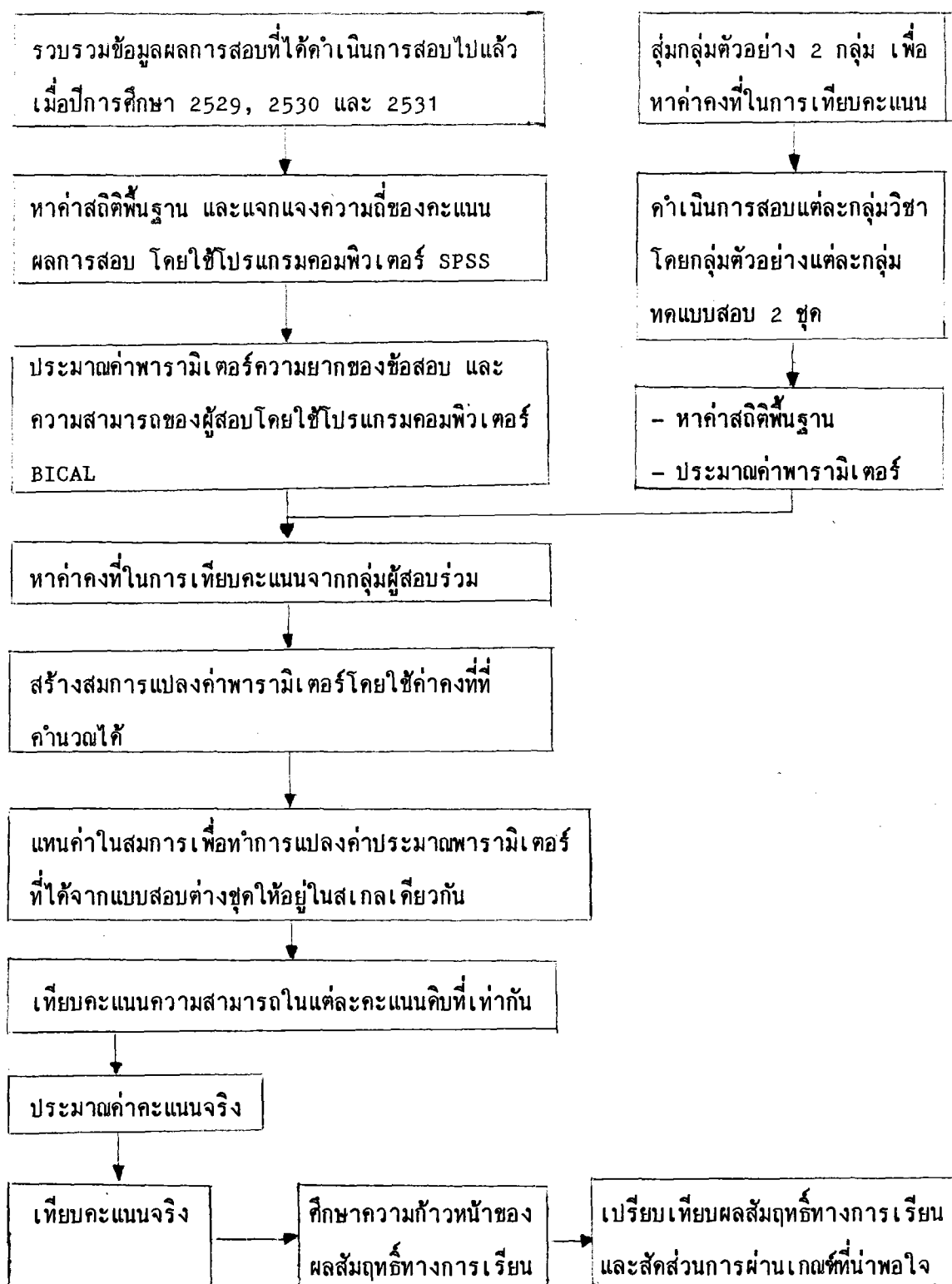
ผู้วิจัยใช้สถิติทดสอบซี (Z-test) ทดสอบในกรณีที่มีสัดส่วนทั้งสองเป็นอิสระจากกัน (uncorrelated data) หลังจากนั้นจะแจกแจงค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละกลุ่มวิชา ระหว่างผลที่ได้จากการเทียบคะแนน กับผลที่ได้จากการไม่เทียบคะแนนตามตารางสองทางดังนี้

		การเทียบคะแนน	
		ผ่าน	ไม่ผ่าน
การไม่เทียบคะแนน	ผ่าน		(2)
	ไม่ผ่าน	(1)	

ผู้วิจัยจะพิจารณาความคลาดเคลื่อน 2 ชนิด คือ

1. ร้อยละนักเรียนที่ไม่ผ่านคะแนนจุดตัดเมื่อยังไม่ได้เทียบคะแนน แต่ผ่านจุดตัดเมื่อได้เทียบคะแนนแล้ว
2. ร้อยละนักเรียนที่ผ่านคะแนนจุดตัดเมื่อยังไม่ได้เทียบคะแนน แต่ไม่ผ่านคะแนนจุดตัดเมื่อได้เทียบคะแนนแล้ว

จากที่กล่าวมา ผู้วิจัยสรุปเป็นแบบแผนในการวิจัย เพื่อศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงสามปีการศึกษาที่ผ่านมา (ปีการศึกษา 2529 ถึง 2531) โดยใช้วิธีเทียบคะแนนรูปแบบราสส์สำหรับการวิจัยในครั้งนี้ ดังภาพประกอบ 3



ภาพประกอบ 3 สรุปแบบแผนในการดำเนินการวิจัยครั้งนี้

ผลการวิเคราะห์ข้อมูล

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

ในการเสนอผลการวิเคราะห์ข้อมูล เพื่อความสะดวกและความเข้าใจตรงกัน ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูลดังนี้

N	แทน	จำนวนนักเรียนในกลุ่มตัวอย่าง
K	แทน	จำนวนข้อสอบของแบบสอบ
X	แทน	ค่าคะแนนเฉลี่ยของแบบสอบ
S.D.	แทน	ค่าเบี่ยงเบนมาตรฐาน
S.E.	แทน	ค่าความคลาดเคลื่อนมาตรฐาน
θ	แทน	ค่าประมาณความสามารถของผู้สอบ
θ_X	แทน	ค่าประมาณความสามารถของผู้สอบแบบสอบชุด X (ปี 2529)
θ_Y	แทน	ค่าประมาณความสามารถของผู้สอบแบบสอบชุด Y (ปี 2530)
θ_Z	แทน	ค่าประมาณความสามารถของผู้สอบแบบสอบชุด Z (ปี 2531)
θ_{X1}	แทน	ค่าประมาณความสามารถของผู้สอบแบบสอบชุด X ในกลุ่มที่ 1
θ_{X2}	แทน	ค่าประมาณความสามารถของผู้สอบแบบสอบชุด X ในกลุ่มที่ 2
θ_{Y1}	แทน	ค่าประมาณความสามารถของผู้สอบแบบสอบชุด Y ในกลุ่มที่ 1
θ_{Z2}	แทน	ค่าประมาณความสามารถของผู้สอบแบบสอบชุด Z ในกลุ่มที่ 2
θ'_{X1}	แทน	ค่าประมาณความสามารถของผู้สอบแบบสอบชุด Y ในกลุ่มที่ 1 ที่แปลงให้อยู่ในสเกลเดียวกับชุด X ในกลุ่มที่ 1

- θ_X^2 แทน ค่าประมาณความสามารถของผู้สอบแบบสอบชุด Z ในกลุ่มที่ 2
 ที่แปลงให้อยู่ในสเกลเดียวกับชุด X ในกลุ่มที่ 2
 θ_X^1 แทน ค่าประมาณความสามารถของผู้สอบแบบสอบชุด Z ในกลุ่มที่ 2
 ที่แปลงให้อยู่ในสเกลเดียวกับชุด X ในกลุ่มที่ 2 และ
 แปลงให้อยู่ในสเกลเดียวกับชุด X ในกลุ่มที่ 1
 ξ แทน ค่าคะแนนจริงของแบบสอบ

การนำเสนอผลการวิเคราะห์ข้อมูล

ผู้วิจัยวิเคราะห์ข้อมูลตามแบบแผนการวิจัยที่กล่าวไว้ในบทที่ 3 และจะรายงานผลสุดท้าย
 ของการเทียบคะแนนเพื่อศึกษาผลสัมฤทธิ์ทางการเรียนในแต่ละปีการศึกษา และความก้าวหน้าของ
 ผลสัมฤทธิ์ทางการเรียนตั้งแต่ปีการศึกษา 2529 ถึง 2531 พร้อมทั้งสัดส่วนจำนวนนักเรียนที่มีผล
 การสอบผ่านเกณฑ์ที่น่าพอใจในแต่ละปีการศึกษา หลังจากนั้นจะเปรียบเทียบคะแนนผลสัมฤทธิ์เมื่อ
 เทียบคะแนนแล้ว และเปรียบเทียบสัดส่วนจำนวนนักเรียนที่มีผลน่าพอใจระหว่างเมื่อเทียบคะแนนกับ
 เมื่อยังไม่ได้เทียบคะแนน ดังเสนอตามลำดับดังนี้

1. ค่าสถิติพื้นฐานของคะแนนผลการสอบ (คะแนนดิบ) ที่ได้จากการสอบในแต่ละปี
การศึกษา ก่อนการเทียบคะแนน
2. ค่าสถิติพื้นฐานของคะแนนผลการสอบที่ได้จากกลุ่มผู้สอบร่วม และค่าคงที่ที่ใช้ในการ
ปรับสเกลความสามารถของผู้สอบแบบสอบแต่ละชุดให้อยู่ในสเกลเดียวกัน
3. ผลการปรับสเกลความสามารถของผู้สอบในปีการศึกษา 2529 ถึง 2531 ให้อยู่ใน
สเกลเดียวกัน
4. ผลการเทียบคะแนน (คะแนนจริง) ระหว่างแบบสอบในแต่ละกลุ่มวิชา
5. ผลสัมฤทธิ์ทางการเรียนในแต่ละปีการศึกษา และความก้าวหน้าของผลสัมฤทธิ์ทาง
การเรียน พร้อมทั้งสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจเมื่อเทียบคะแนนแล้ว

6. การเปรียบเทียบความแตกต่างของผลสัมฤทธิ์ทางการเรียนระหว่างปีการศึกษา 2529 ถึง 2531 เมื่อเทียบคะแนนแล้ว

7. การเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ ในแต่ละปีการศึกษา ระหว่างการเทียบคะแนนกับการไม่เทียบคะแนน

ผู้วิจัยนำคะแนนผลการสอบที่ได้ดำเนินการสอบกับกลุ่มตัวอย่างไปแล้ว เมื่อปีการศึกษา 2529 ถึง 2531 ตามโครงการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศ ดังกล่าวมาทำการวิเคราะห์หาค่าสถิติพื้นฐาน โดยใช้โปรแกรมคอมพิวเตอร์ SPSS-X ซึ่งปรากฏผล ดังตาราง 4.

ตาราง 4 ค่าสถิติพื้นฐานของคะแนนผลการสอบในแต่ละปีการศึกษา ก่อนการเทียบคะแนน จำแนก เป็นรายกลุ่มวิชา

ค่าสถิติ กลุ่มวิชา	จำนวน ข้อ (k)	จำนวน ผู้สอบ (n)	ค่า ต่ำสุด (min)	ค่า สูงสุด (max)	ค่าเฉลี่ย ($\bar{X}$)	ค่าเบี่ยงเบน มาตรฐาน (S.D.)	ค่าความ คลาดเคลื่อน (S.E.)
<u>ภาษาไทย</u>							
ปีการศึกษา 2529	30	534	5	24	15.168	3.427	.148
ปีการศึกษา 2530	30	453	5	24	14.731	3.558	.167
ปีการศึกษา 2531	30	467	5	26	17.097	3.673	.170

ตาราง 4 (ต่อ)

ค่าสถิติ	จำนวน ข้อ (k)	จำนวน ผู้สอบ (n)	ค่า ต่ำสุด (min)	ค่า สูงสุด (max)	ค่าเฉลี่ย ค่าเฉลี่ย ($\bar{X}$)	ค่าเบี่ยงเบน มาตรฐาน (S.D.)	ค่าความ คลาดเคลื่อน (S.E.)
กลุ่มวิชา							
<u>คณิตศาสตร์</u>							
ปีการศึกษา 2529	38	534	3	25	13.828	4.117	.178
ปีการศึกษา 2530	38	453	3	30	15.575	5.028	.236
ปีการศึกษา 2531	38	467	3	32	15.127	4.945	.229

จากตาราง 4 ก่อนที่จะทำการเทียบคะแนนผลการสอบ (คะแนนดิบ) ในแต่ละปีการศึกษา ตั้งแต่ปีการศึกษา 2529 ถึง 2531 ซึ่งได้มาจากการดำเนินการสอบนักเรียนที่เป็นกลุ่มตัวอย่างที่เรียนจบชั้นประถมศึกษาปีที่ 6 ในแต่ละปีการศึกษา ด้วยแบบสอบต่างชุด โดยไม่มีข้อสอบร่วมกันปรากฏผล ดังนี้ ในกลุ่มทักษะวิชาภาษาไทยซึ่งมีคะแนนเต็ม 30 คะแนน พบว่าคะแนนต่ำสุดของผลการสอบแบบสอบแต่ละชุด มีค่าเท่ากันคือ 5 ส่วนคะแนนสูงสุดพบว่ามีค่าใกล้เคียงกันคือมีค่าตั้งแต่ 24 ถึง 26 คะแนน โดยในทุกปีการศึกษาไม่มีผู้สอบคนใดได้คะแนนศูนย์หรือได้คะแนนเต็ม สำหรับคะแนนเฉลี่ยผลการสอบในสามปีการศึกษา มีค่าต่ำสุด 14.731 คะแนนเฉลี่ยสูงสุด 17.097 พิสัยของคะแนนเฉลี่ยเท่ากับ 2.366 คะแนนเฉลี่ยต่ำสุดในปีการศึกษา 2530 คะแนนเฉลี่ยสูงสุดในปีการศึกษา 2531 โดยมีเพียงปีการศึกษา 2530 เท่านั้นที่มีคะแนนเฉลี่ยต่ำกว่าครึ่งหนึ่งของคะแนนเต็มเล็กน้อย เมื่อพิจารณาค่าเบี่ยงเบนมาตรฐานของการแจกแจงคะแนนในแต่ละปีการศึกษา พบว่ามีค่าใกล้เคียงกัน คือมีค่าต่ำสุดเท่ากับ 3.427 และมีค่าสูงสุดเท่ากับ 3.673

ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์ ซึ่งมีคะแนนเต็ม 38 คะแนน พบว่าคะแนนต่ำสุดจากผลการสอบแบบสอบแต่ละชุดมีค่าเท่ากันคือ มีค่าเท่ากับ 3 และมีคะแนนสูงสุดเท่ากับ 32 คะแนน (ปีการศึกษา 2531) โดยในทุกปีการศึกษาไม่มีผู้สอบคนใดได้คะแนนศูนย์หรือได้คะแนนเต็มเลย สำหรับคะแนนเฉลี่ยผลการสอบในสามปีการศึกษามีค่าต่ำสุดเท่ากับ 13.828 และคะแนนเฉลี่ยสูงสุด 15.575 พิสัยของคะแนนเฉลี่ยเท่ากับ 1.747 คะแนนเฉลี่ยต่ำสุดในปีการศึกษา 2529 คะแนนเฉลี่ยสูงสุดในปีการศึกษา 2530 ทั้งนี้คะแนนเฉลี่ยในทุกปีการศึกษา มีค่าต่ำกว่าครึ่งหนึ่งของคะแนนเต็ม เมื่อพิจารณาค่าเบี่ยงเบนมาตรฐานของการแจกแจงคะแนนในแต่ละปีการศึกษา พบว่ามีค่าต่ำสุดเท่ากับ 4.117 และมีค่าสูงสุดเท่ากับ 5.028

ในทำนองเดียวกัน ผู้วิจัยหาค่าสถิติพื้นฐานของคะแนนผลการสอบในแต่ละปีการศึกษาจำแนกเป็นรายอำเภอ ซึ่งปรากฏผลดังตาราง 5

ตาราง 5 ค่าสถิติพื้นฐานของคะแนนผลการสอบ (คะแนนดิบ) ในกลุ่มทักษะวิชาภาษาไทยในแต่ละปีการศึกษา ก่อนการเทียบคะแนน จำแนกเป็นรายอำเภอ

ค่าสถิติ อำเภอ	จำนวนข้อ (k)	จำนวน ผู้สอบ (n)	ค่าต่ำสุด (min)	ค่าสูงสุด (max)	ค่าเฉลี่ย ($\bar{X}$)	ค่าเบี่ยงเบน มาตรฐาน (S.D.)	ค่าความ คลาดเคลื่อน (S.E.)
<u>อำเภอเมือง</u>							
ปีการศึกษา 2529	30	184	5	23	15.589	3.550	.261
ปีการศึกษา 2530	30	119	7	22	15.866	3.165	.290
ปีการศึกษา 2531	30	109	8	26	17.797	3.332	.274

ตาราง 5 (ต่อ)

ค่าสถิติ อำเภอ	จำนวนข้อ	จำนวน ผู้สอบ	ค่าต่ำสุด	ค่าสูงสุด	ค่าเฉลี่ย	ค่าเบี่ยงเบน มาตรฐาน	ค่าความ คลาดเคลื่อน
	(k)	(n)	(min)	(max)	($\bar{X}$)	(S.D.)	(S.E.)
<u>อำเภอคลองหลวง</u>							
ปีการศึกษา 2529	30	136	6	22	15.184	2.847	.244
ปีการศึกษา 2530	30	138	7	24	15.406	3.543	.302
ปีการศึกษา 2531	30	125	5	26	16.760	3.987	.357
<u>อำเภอลาดหลุมแก้ว</u>							
ปีการศึกษา 2529	30	99	7	24	15.848	3.453	.347
ปีการศึกษา 2530	30	87	8	21	13.989	3.138	.336
ปีการศึกษา 2531	30	103	7	25	17.505	3.319	.327
<u>อำเภอหนองเสือ</u>							
ปีการศึกษา 2529	30	115	6	22	13.887	3.534	.330
ปีการศึกษา 2530	30	109	5	23	13.229	3.983	.353
ปีการศึกษา 2531	30	89	7	26	15.933	3.858	.409

จากตาราง 5 ก่อนทำการเทียบคะแนนพบว่าในกลุ่มทักษะวิชาภาษาไทยนั้น อำเภอมืองและอำเภอกลองหลวง มีคะแนนเฉลี่ยผลการสอบของทุกปีการศึกษาเกินครึ่งหนึ่งของคะแนนเต็ม และมีคะแนนเฉลี่ยที่สูงขึ้นกว่าเดิมทุกปีการศึกษา โดยอำเภอมืองมีคะแนนเฉลี่ยตั้งแต่ 15.589 ถึง 17.797 พิสัยของคะแนนเฉลี่ยเท่ากับ 2.208 คะแนน อำเภอกลองหลวงมีคะแนนเฉลี่ยตั้งแต่ 15.184 ถึง 16.760 พิสัยของคะแนนเฉลี่ยเท่ากับ 1.276 คะแนน ส่วนอำเภอลาดหลุมแก้วและอำเภอนองเสือพบว่ามีคะแนนเฉลี่ยลดลงในปีการศึกษา 2530 แต่มีคะแนนเฉลี่ยสูงขึ้นในปีการศึกษา 2531 โดยอำเภอลาดหลุมแก้วมีคะแนนเฉลี่ยตั้งแต่ 13.989 ถึง 17.505 คะแนนเฉลี่ยต่ำสุดในปีการศึกษา 2530 ซึ่งเป็นปีการศึกษาเดียวที่มีคะแนนเฉลี่ยต่ำกว่าครึ่งหนึ่งของคะแนนเต็มและคะแนนเฉลี่ยสูงสุดในปีการศึกษา 2531 พิสัยของคะแนนเฉลี่ยเท่ากับ 3.516 ส่วนอำเภอนองเสือมีคะแนนเฉลี่ยตั้งแต่ 13.229 ถึง 15.933 คะแนน เฉลี่ยต่ำสุดในปีการศึกษา 2530 และคะแนนเฉลี่ยสูงสุดในปีการศึกษา 2531 ซึ่งเป็นปีการศึกษาเดียวที่มีคะแนนเฉลี่ยเกินครึ่งหนึ่งของคะแนนเต็ม พิสัยของคะแนนเฉลี่ยเท่ากับ 2.704 เมื่อพิจารณาค่าเบี่ยงเบนมาตรฐานของคะแนนทุกปีการศึกษาของทุกอำเภอพบว่ามีค่าต่ำสุด 2.847 และมีค่าสูงสุด 3.987

ส่วนผลการหาค่าสถิติพื้นฐานของคะแนนผลการสอบในกลุ่มทักษะวิชาคณิตศาสตร์ ปรากฏผลดังตาราง 6

ตาราง 6 ค่าสถิติพื้นฐานของคะแนนผลการสอบ (คะแนนดิบ) ในกลุ่มวิชาคณิตศาสตร์ในแต่ละ
ปีการศึกษา ก่อนการเทียบคะแนน จำแนกเป็นรายอำเภอ

อำเภอ	ค่าสถิติ	จำนวนข้อ (k)	จำนวน ผู้สอบ (n)	ค่าต่ำสุด (min)	ค่าสูงสุด (max)	ค่าเฉลี่ย ($\bar{X}$)	ค่าเบี่ยงเบน มาตรฐาน (S.D.)	ค่าความ คลาดเคลื่อน (S.E.)
<u>อำเภอเมือง</u>								
ปีการศึกษา 2529		30	185	3	34	14.286	4.201	.309
ปีการศึกษา 2530		30	119	6	27	17.420	4.884	.448
ปีการศึกษา 2531		30	148	6	28	14.162	4.354	.358
<u>อำเภอดงหลวง</u>								
ปีการศึกษา 2529		38	136	5	25	13.949	4.389	.351
ปีการศึกษา 2530		38	138	8	29	16.246	4.664	.397
ปีการศึกษา 2531		38	125	3	32	14.024	4.611	.412
<u>อำเภอลาดหลุมแก้ว</u>								
ปีการศึกษา 2529		38	99	8	25	15.313	3.638	.366
ปีการศึกษา 2530		38	87	7	30	15.954	5.060	.542
ปีการศึกษา 2531		38	103	8	28	17.835	5.128	.505

ตาราง 6 (ต่อ)

ค่าสถิติ อำเภอ	จำนวนข้อ (k)	จำนวน ผู้สอบ (n)	ค่าต่ำสุด (min)	ค่าสูงสุด (max)	ค่าเฉลี่ย ($\bar{X}$)	ค่าเบี่ยงเบน มาตรฐาน (S.D.)	ค่าความ คลาดเคลื่อน (S.E.)
<u>อำเภอหนองเสือ</u>							
ปีการศึกษา 2529	38	115	6	23	11.670	3.573	.333
ปีการศึกษา 2530	38	109	3	26	13.073	4.596	.440
ปีการศึกษา 2531	38	89	6	25	15.146	4.994	.529

จากตาราง 6 ก่อนทำการเทียบคะแนนพบว่าในกลุ่มทักษะวิชาคณิตศาสตร์ ทุกอำเภอมีคะแนนเฉลี่ยผลการสอบของทุกปีการศึกษาคือว่าครั้งหนึ่งของคะแนนเต็ม โดยอำเภอลาดหลุมแก้วและอำเภอหนองเสือมีคะแนนเฉลี่ยสูงขึ้นกว่าเดิมทุกปีการศึกษา สำหรับอำเภอลาดหลุมแก้วมีคะแนนเฉลี่ยตั้งแต่ 15.313 ถึง 17.835 พิสัยของคะแนนเฉลี่ยเท่ากับ 2.522 อำเภอหนองเสือมีคะแนนเฉลี่ยตั้งแต่ 11.670 ถึง 15.146 พิสัยของคะแนนเฉลี่ยเท่ากับ 3.467 ส่วนอำเภอเมืองและอำเภอดงหลวงมีคะแนนเฉลี่ยสูงขึ้นในปีการศึกษา 2530 แต่มีคะแนนเฉลี่ยลดลงในปีการศึกษา 2531 ซึ่งค่าเฉลี่ยในปีการศึกษา 2531 เกือบเท่ากับค่าเฉลี่ยเดิมในปีการศึกษา 2529 โดยอำเภอเมืองมีคะแนนเฉลี่ยตั้งแต่ 14.162 ถึง 17.420 ค่าเฉลี่ยต่ำสุดในปีการศึกษา 2531 ค่าเฉลี่ยสูงสุดในปีการศึกษา 2530 พิสัยของคะแนนเฉลี่ยเท่ากับ 3.258 ส่วนอำเภอดงหลวงมีคะแนนเฉลี่ยตั้งแต่ 13.949 ถึง 16.246 ค่าเฉลี่ยต่ำสุดในปีการศึกษา 2529 และค่าเฉลี่ยสูงสุดในปีการศึกษา 2530 พิสัยของคะแนนเฉลี่ยเท่ากับ 2.297 และเมื่อพิจารณาค่าเบี่ยงเบนมาตรฐานของคะแนนทุกปีการศึกษาของทุกอำเภอพบว่าค่าต่ำสุด 3.573 ค่าสูงสุด 5.128

หลังจากนั้นผู้วิจัยทำการประมาณค่าพารามิเตอร์ความยากของข้อสอบแต่ละข้อและความสามารถของผู้สอบที่ได้คะแนนรวมต่าง ๆ กันในแบบสอบแต่ละชุด ซึ่งสอบกับผู้สอบต่างกลุ่ม โดยใช้โปรแกรมคอมพิวเตอร์ BICAL ซึ่งจากการหาค่าสถิติพื้นฐานของคะแนนผลการสอบในแบบสอบแต่ละชุด ดังกล่าวพบว่า ไม่มีผู้ใดได้คะแนนศูนย์หรือได้คะแนนเต็มเลย ดังนั้นจึงเหมาะที่จะวิเคราะห์ผลการสอบตามทฤษฎีการตอบข้อสอบซึ่งมีข้อกำหนดว่า การประมาณค่าพารามิเตอร์ตามทฤษฎีนี้จะตัดผู้สอบที่ได้คะแนนศูนย์หรือได้คะแนนเต็มออกจากการวิเคราะห์เพราะถือว่าข้อสอบดังกล่าวไม่มีลักษณะจำเป็นที่จะนำมาวิเคราะห์ (Wright and Stone. 1977 : 32 - 65) ดังนั้น การหาค่าประมาณพารามิเตอร์ในครั้งนี้จึงใช้ข้อมูลคะแนนผลการสอบที่เก็บรวบรวมข้อมูลได้ทุกค่า และปรากฏผลการประมาณค่าพารามิเตอร์ของแบบสอบแต่ละชุดดังตาราง 7

ตาราง 7 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าประมาณพารามิเตอร์ความยากของข้อสอบและของค่าประมาณพารามิเตอร์ความสามารถของผู้สอบที่สอบแบบสอบในปีการศึกษา 2529 ถึง 2531 ก่อนการปรับให้อยู่ในสเกลเดียวกัน จำแนกเป็นรายวิชา

ค่าประมาณ	ปีการศึกษา 2529				ปีการศึกษา 2530				ปีการศึกษา 2531			
	min	max	$\bar{X}$	S.D.	min	max	$\bar{X}$	S.D.	min	max	$\bar{X}$	S.D.
<u>ภาษาไทย</u>												
ค่าความยาก	-2.33	1.79	.00	1.10	-2.38	2.31	.00	1.09	-1.59	2.17	.00	.83
ค่าความสามารถ	-3.79	3.68	.06	.58	-3.73	3.77	-.06	.59	-3.53	3.56	.32	.59
<u>คณิตศาสตร์</u>												
ค่าความยาก	-1.90	2.09	.00	.83	-1.72	2.48	.00	.87	-1.73	1.88	.00	.67
ค่าความสามารถ	-3.80	3.83	-.66	.54	-3.82	3.88	-.43	.65	-3.71	3.72	-.47	.61

จากตาราง 7 โปรแกรมคอมพิวเตอร์ BICAL จะให้ค่าประมาณพารามิเตอร์ความยากของข้อสอบ และความสามารถของผู้สอบ โดยแยกประมาณค่าจากคะแนนผลการสอบแบบสอบที่ละชุด และทำการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบที่ได้คะแนนรวมต่าง ๆ กันทีละคนที่ละข้อ โดยใช้วิธีการทำซ้ำ ๆ (iterative) จนได้ค่าประมาณพารามิเตอร์คงที่ เพื่อกำหนดจุดเริ่มต้นของสเกลความสามารถและประมาณค่าพารามิเตอร์ความสามารถของผู้สอบที่ได้คะแนนดิบต่าง ๆ

ที่เป็นไปได้ในแต่ละชุด ทั้งนี้ค่าประมาณพารามิเตอร์เหล่านี้จะมีหน่วยการวัดแบบล็อก (log) เหมือนกัน แต่จุดเริ่มต้นของสเกลอาจแตกต่างกัน (Rentz and Bashaw. 1977 : 162) ส่วนการปรับสเกลความสามารถระหว่างแบบสอบต่างชุดซึ่งประมาณค่าแยกกันเพื่อให้อยู่ในสเกลที่มีจุดเริ่มต้นที่เดียวกันนั้นต้องใช้ข้อมูลจากแบบสอบร่วมหรือผู้สอบร่วมในการกำหนดค่าคงที่ของการเทียบคะแนน ซึ่งจะใช้ค่าคงที่นี้แทนในสมการเชิงเส้นตรงสำหรับการปรับสเกลความสามารถให้อยู่ในสเกลเดียวกันทั้งหมด (Hambleton. 1985 : 205 - 207).

แต่เนื่องจากการดำเนินการสอบในแต่ละปีการศึกษาที่ผ่านมาผู้ใช้แบบสอบต่างชุดสอบนักเรียนต่างกลุ่ม โดยไม่มีข้อมูลจากแบบสอบร่วมหรือผู้สอบร่วม ด้วยเหตุนี้ผู้วิจัยจึงจะหาค่าคงที่ในการเทียบคะแนนจากกลุ่มตัวอย่างอื่นที่ได้มาจากการสุ่มดังกล่าวไว้ในบทที่ 3 ทั้งนี้จะศึกษาในกรณีที่ใช้ผู้สอบร่วม (กลุ่มตัวอย่างเดียวสอบแบบสอบสองชุด) ซึ่งเริ่มด้วยการแบ่งกลุ่มตัวอย่างที่สุ่มมาออกเป็นสองกลุ่มย่อย แล้วจัดให้แต่ละกลุ่มสอบแบบสอบสองชุด โดยต้องสอบแบบสอบชุดที่ใช้ในปีการศึกษา 2529 หนึ่งชุดเหมือนกัน และต้องสอบแบบสอบที่ใช้ในปีการศึกษา 2530 หรือในปีการศึกษา 2531 อีกหนึ่งชุด และเพื่อควบคุมผลที่จะเกิดขึ้นจากการสอบแบบสอบชุดแรกในแบบสอบชุดหลัง ได้แก้ความคล่องในการตอบข้อสอบ และความเหนื่อยล้าของผู้สอบ ผู้วิจัยจึงจัดลำดับก่อนหลังในการสอบแบบสอบสองชุดให้กับผู้สอบต่างกันโดยวิธีสุ่ม ผลการวิเคราะห์ข้อมูลจากผู้สอบร่วมนี้จะสามารถหาความแตกต่างระหว่างค่าความสามารถเฉลี่ยของแบบสอบแต่ละชุดที่สอบโดยผู้สอบแต่ละกลุ่ม และจะสามารถประมาณค่าจุดเริ่มต้นของสเกลความสามารถร่วมระหว่างผู้สอบแต่ละกลุ่ม เพื่อนำมาหาค่าคงที่และสร้างสมการในการปรับสเกลความสามารถระหว่างผู้สอบแบบสอบต่างชุดได้ ซึ่งปรากฏผลการวิเคราะห์ข้อมูลจากกลุ่มผู้สอบร่วมดังต่อไปนี้

ตาราง 8 ค่าสถิติพื้นฐานของคะแนนผลการสอบของผู้สอบกลุ่มที่ 1 ที่สอบแบบสอบชุดที่ใช้ในปีการศึกษา 2529 (ชุด X) และชุดที่ใช้ในปีการศึกษา 2530 (ชุด Y) และจากผู้สอบกลุ่มที่ 2 ที่สอบแบบสอบชุดที่ใช้ในปีการศึกษา 2529 และชุดที่ใช้ในปีการศึกษา 2531 (ชุด Z) จำแนกเป็นรายวิชา

ค่าสถิติ กลุ่มผู้สอบ	จำนวน ผู้สอบ	ชุด X				ชุด Y				ชุด Z			
		$\bar{X}$	S.D.	min	max	$\bar{X}$	S.D.	min	max	$\bar{X}$	S.D.	min	max
กลุ่มที่ 1													
ภาษาไทย	218	17.440	3.779	8	25	15.564	3.221	7	24				
คณิตศาสตร์	224	13.955	3.856	6	25	11.786	3.180	4	21				
กลุ่มที่ 2													
ภาษาไทย	196	16.209	3.271	7	23					17.857	3.684	6	26
คณิตศาสตร์	197	12.817	3.901	4	25					13.868	4.336	4	29

จากตาราง 8 ในกลุ่มทักษะวิชาภาษาไทย กรณีแบบสอบชุด X และชุด Y ที่สอบโดยผู้สอบกลุ่มเดียวกัน (กลุ่มที่ 1) พบว่าชุด X มีคะแนนเฉลี่ย 17.440 และชุด Y มีคะแนนเฉลี่ย 15.564 ซึ่งชุด X มีคะแนนเฉลี่ยสูงกว่าชุด Y เท่ากับ 1.876 คะแนน ส่วนในกรณีแบบสอบชุด X และชุด Z ที่สอบโดยผู้สอบกลุ่มเดียวกัน (กลุ่มที่ 2) พบว่าชุด X มีคะแนนเฉลี่ย 16.209 ส่วนชุด Z มีคะแนนเฉลี่ย 17.857 ซึ่งชุด X มีคะแนนเฉลี่ยต่ำกว่าชุด Z เท่ากับ 1.648 คะแนน ทั้งนี้คะแนนเฉลี่ยชุด X ที่ได้จากผู้สอบกลุ่มที่ 1 มีค่าสูงกว่าคะแนนเฉลี่ยของผู้สอบกลุ่มที่ 2 เท่ากับ 1.231 คะแนน

ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์ กรณีแบบสอบชุด X และชุด Y ที่สอบโดยผู้สอบกลุ่มเดียวกัน (กลุ่มที่ 1) พบว่าชุด X มีคะแนนเฉลี่ย 13.955 ชุด Y มีคะแนนเฉลี่ย 11.786 ซึ่งชุด X มีคะแนนเฉลี่ยสูงกว่าชุด Y เท่ากับ 2.169 คะแนน ส่วนกรณีแบบสอบชุด X และชุด Z ที่สอบโดยผู้สอบกลุ่มเดียวกัน (กลุ่มที่ 2) พบว่าชุด X มีคะแนนเฉลี่ย 12.817 ชุด Z มีคะแนนเฉลี่ย 13.868 ซึ่งชุด X มีคะแนนเฉลี่ยต่ำกว่าชุด Z เท่ากับ 1.051 คะแนน ทั้งนี้คะแนนเฉลี่ยชุด X ที่ได้จากผู้สอบกลุ่มที่ 1 มีค่าสูงกว่าคะแนนเฉลี่ยจากผู้สอบกลุ่มที่ 2 เท่ากับ 1.138 คะแนน อย่างไรก็ตามผลการวิเคราะห์พบว่าไม่มีผู้สอบคนใดได้คะแนนศูนย์หรือได้คะแนนเต็มในแบบสอบทุกชุดในทั้งสองกลุ่มวิชา

การวิเคราะห์คะแนนผลการสอบที่ได้จากแบบสอบแต่ละชุด ซึ่งสอบโดยผู้สอบร่วมดังกล่าว พอสรุปได้ว่า หากกลุ่มผู้สอบกลุ่มเดียวสอบแบบสอบชุดใดชุดหนึ่งในสามชุดของแต่ละกลุ่มวิชาจะได้คะแนนเฉลี่ยแตกต่างกัน กล่าวคือหากสอบแบบสอบชุด Z จะได้คะแนนเฉลี่ยสูงสุด รองลงมาคือ การสอบแบบสอบชุด X และหากสอบแบบสอบชุด Y จะได้คะแนนเฉลี่ยต่ำสุดซึ่งแสดงให้เห็นถึงความจำเป็นในการเทียบคะแนนเพื่อหาคะแนนสมมูลระหว่างแบบสอบ

หลังจากนั้นผู้วิจัยทำการประมาณค่าพารามิเตอร์ของแบบสอบแต่ละชุดที่สอบโดยผู้สอบแต่ละกลุ่ม โดยใช้โปรแกรมคอมพิวเตอร์ BICAL ทั้งนี้เพื่อนำผลมาใช้ในการคำนวณค่าคงที่ในการสร้างสมการเชิงเส้นตรงเพื่อปรับสเกลความสามารถของผู้สอบแบบสอบต่างชุดให้อยู่ในสเกลเดียวกันต่อไป สรุปได้ค่าประมาณพารามิเตอร์ของแบบสอบแต่ละชุดที่สอบโดยผู้สอบแต่ละกลุ่ม ดังตาราง 9

ตาราง 9 สรุปค่าประมาณพารามิเตอร์ความยากของข้อสอบ (b) และความสามารถของผู้สอบ (θ) ในแบบสอบแต่ละชุดที่ผู้สอบโดยผู้สอบรวม

ค่า ประมาณ พารามิเตอร์	กลุ่มที่		กลุ่มที่ 1								กลุ่มที่ 2							
			ชุด X				ชุด Y				ชุด X				ชุด Z			
	MIN	MAX	X	S.D	MIN	MAX	X	S.D	MIN	MAX	X	S.D	MIN	MAX	X	S.D		
ภาษาไทย																		
b	-2.31	2.30	.00	.42	-2.32	2.31	.00	.93	-2.02	1.81	.00	.87	-1.69	2.22	.00	.74		
θ	-3.83	3.91	.42	.69	-3.77	3.83	.08	.53	-3.81	3.78	.22	.57	-3.54	3.57	.44	.53		
คณิตศาสตร์																		
b	-2.23	2.55	.00	1.26	-1.36	2.01	.00	1.14	-1.77	2.07	.00	1.19	-2.06	1.80	.00	.84		
θ	-3.88	3.94	-.66	.51	-3.85	3.91	-.96	.46	-3.82	3.88	-.80	.53	-3.77	3.75	-.63	.57		

จากตาราง 9 เมื่อพิจารณาค่าประมาณพารามิเตอร์ความสามารถของผู้สอบกลุ่มเดียวกันที่ทำแบบสอบต่างชุดปรากฏผลดังนี้ ในกลุ่มทักษะวิชาภาษาไทยพบว่า ผู้สอบในกลุ่มที่ 1 มีค่าเฉลี่ยความสามารถในการทำแบบสอบชุด X เท่ากับ .42 และมีค่าเฉลี่ยความสามารถในการทำแบบสอบชุด Y เท่ากับ .08 ความแตกต่างของค่าเฉลี่ยมีค่าเท่ากับ .34 ส่วนในกลุ่มที่ 2 มีค่าเฉลี่ยความสามารถในการทำแบบสอบชุด X เท่ากับ .22 และมีค่าเฉลี่ยความสามารถในการทำแบบสอบชุด Z เท่ากับ .44 ความแตกต่างของค่าเฉลี่ยเท่ากับ .22 นอกจากนี้ความแตกต่างของค่าเฉลี่ยในการทำแบบสอบชุด X เหมือนกันระหว่างกลุ่มที่ 1 และ กลุ่มที่ 2 มีค่าเท่ากับ .20

สำหรับในกลุ่มทักษะวิชาคณิตศาสตร์พบว่า ผู้สอบกลุ่มที่ 1 มีค่าเฉลี่ยความสามารถในการทำแบบสอบชุด X เท่ากับ -0.66 และมีค่าเฉลี่ยความสามารถในการทำแบบสอบชุด Y เท่ากับ -0.96 ความแตกต่างของค่าเฉลี่ยมีค่าเท่ากับ $.30$ ส่วนในกลุ่มที่ 2 มีค่าเฉลี่ยความสามารถในการทำแบบสอบชุด X เท่ากับ -0.80 และมีค่าเฉลี่ยความสามารถในการทำแบบสอบชุด Z เท่ากับ -0.63 ความแตกต่างของค่าเฉลี่ยมีค่าเท่ากับ $.17$ นอกจากนี้ได้พบว่า ความแตกต่างของค่าเฉลี่ยในการทำแบบสอบชุด X เหมือนกันระหว่างกลุ่มที่ 1 และกลุ่มที่ 2 มีค่าเท่ากับ $.14$

จากผลการประมาณค่าเฉลี่ยความสามารถโดยใช้ผู้สอบร่วมของทั้งสองกลุ่มวิชาแสดงให้เห็นว่าในการสอบแบบสอบ ทั้งสามชุดของแต่ละกลุ่มวิชาผู้สอบต้องใช้ความสามารถต่างกัน จึงทำให้ได้คะแนนเฉลี่ยในการสอบแต่ละชุดต่างกัน

ดังนั้นผู้วิจัยจึงทำการปรับค่าประมาณพารามิเตอร์ความสามารถของผู้สอบร่วมให้อยู่ในสเกลเดียวกันทั้งหมด คือในสเกลเดียวกับค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบชุด X ในกลุ่มที่ 1 สเกลความสามารถที่ได้ใหม่นี้จึงเหมือนกับการใช้ผู้สอบเพียงกลุ่มเดียวคือกลุ่มที่ 1 แต่ทำแบบสอบทั้งสามชุด ซึ่งเป็นวิธีหนึ่งที่สามารถเทียบหาคะแนนดิบหรือคะแนนจริงระหว่างแบบสอบสามชุดในแต่ละกลุ่มวิชาได้ ทั้งนี้ในการปรับหรือแปลงค่าประมาณความสามารถให้อยู่ในสเกลเดียวกันดังกล่าวนี้จำเป็นต้องกำหนดค่าคงที่ เพื่อนำมาสร้างสมการเชิงเส้นตรงสำหรับแปลงค่าประมาณพารามิเตอร์ความสามารถจากแบบสอบต่างชุดในสเกลเดียวกันเสียก่อน ซึ่งปรากฏผลดังตาราง 10

ตาราง 10 ค่าคงที่ α, β และสมการเชิงเส้นตรงในการแปลงค่าประมาณพารามิเตอร์ ความสามารถ (๑) ของผู้สอบที่ได้จากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน

ค่าคงที่/สมการ กลุ่มวิชา	ค่าคงที่		สมการเชิงเส้นตรง
	α	β	
ภาษาไทย			
1. แปลงจากชุด Y_1 ไปยังชุด X_1	1.302	.316	$\theta X_1 = 1.302 \theta Y_1 + .316$
2. แปลงจากชุด Z_2 ไปยังชุด X_2	.966	-.205	$\theta X_2 = .966 \theta Z_2 - .205$
และแปลงจากชุด X_2 ไปยังชุด X_1	1.211	.154	$\theta X_1 = 1.211 \theta X_2 + .154$
คณิตศาสตร์			
1. แปลงจากชุด Y_1 ไปยังชุด X_1	1.109	.405	$\theta X_1 = 1.109 \theta Y_1 + .405$
2. แปลงจากชุด Z_2 ไปยังชุด X_2	.930	-.214	$\theta X_2 = .930 \theta Z_2 - .214$
และแปลงจากชุด X_2 ไปยังชุด X_1	.962	.110	$\theta X_1 = .962 \theta X_2 + .110$

จากตาราง ค่าคงที่ α และ β ที่ได้รับเหล่านั้นนำไปแทนลงในสมการเชิงเส้นตรง $\theta X = \alpha \theta Y + \beta$ ซึ่งอยู่ในรูป $Y = \alpha X + \beta$ โดยที่ $\alpha = S_X/S_Y$ และ $\beta = \bar{X} - \bar{Y}$ เพื่อใช้ในการแปลงค่าประมาณพารามิเตอร์ความสามารถที่ได้จากแบบสอบแต่ละคู่จากกลุ่มผู้สอบพร้อมดังกล่าวโดยมีขั้นตอนการแปลงในแต่ละกลุ่มวิชาเหมือนกันดังนี้

1. แปลงค่าประมาณพารามิเตอร์ของผู้สอบกลุ่มที่ 1 จากแบบสอบชุด Y ไปยังสเกลเดียวกับชุด X
2. แปลงค่าประมาณพารามิเตอร์ของผู้สอบกลุ่มที่ 2 จากแบบสอบชุด Z ไปยังสเกลเดียวกับชุด X แล้วแปลงจากค่าที่แปลงให้อยู่ในสเกลของแบบสอบชุด X ในกลุ่มที่ 2 แล้วนั้นไปยังสเกลเดียวกันกับค่าประมาณพารามิเตอร์ของชุด X ในกลุ่มที่ 1 ดังนั้นจะได้ค่าประมาณของแบบสอบทั้งสามชุดที่แปลงไปยังสเกลเดียวกันทั้งหมด คือสเกลของแบบสอบชุด X ในกลุ่มที่ 1 ซึ่งปรากฏผลดังตาราง 11 และ 12

ตาราง 11 ผลการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบร่วมที่สอบแบบสอบสามชุด
ในกลุ่มทักษะวิชาภาษาไทยให้อยู่ในสเกลเดียวกัน

คะแนนดิบ	ชุด X		ชุด Y		ชุด Z		
	θX_1	θX_2	θY_1	$\theta'X_1$	θZ_2	$\theta'X_2$	$\theta''X_1$
29	3.91	3.78	3.83	5.30	3.57	3.24	4.08
28	3.16	3.05	3.07	4.31	2.83	2.53	3.22
27	2.68	2.59	2.58	3.68	2.38	2.09	2.68
26	2.31	2.24	2.22	3.21	2.03	1.76	2.29
25	2.01	1.95	1.91	2.80	1.75	1.49	1.96
24	1.74	1.70	1.65	2.46	1.51	1.25	1.67
23	1.50	1.48	1.41	2.15	1.30	1.05	1.43
22	1.28	1.27	1.20	1.88	1.11	.87	1.21
21	1.07	1.07	1.00	1.62	.93	.69	.99
20	.87	.88	.82	1.38	.76	.53	.80
19	.68	.70	.64	1.15	.60	.37	.60
18	.50	.53	.47	.93	.44	.22	.42
17	.32	.35	.30	.71	.29	.08	.25
16	.14	.18	.14	.50	.14	-.07	.07
15	-.03	.01	-.02	.29	-.01	-.21	-.10
14	-.20	-.16	-.18	.08	-.15	-.35	-.27
13	-.37	-.33	-.34	-.13	-.30	-.49	-.44

ตาราง 11 (ต่อ)

คะแนนดิบ	ชุด X		ชุด Y		ชุด Z		
	θX_1	θX_2	θY_1	$\theta' X_1$	$\theta \cdot Z_2$	$\theta' X_2$	$\theta'' X_1$
12	-.55	-.51	-.50	-.34	-.45	-.64	-.62
11	-.72	-.69	-.67	-.56	-.61	-.79	-.80
10	-.91	-.87	-.84	-.78	-.77	-.95	-1.00
9	-1.09	-1.06	-1.02	-1.01	-.94	-1.11	-1.19
8	-1.29	-1.26	-1.21	-1.26	-1.12	-1.29	-1.41
7	-1.50	-1.48	-1.42	-1.53	-1.31	-1.47	-1.63
6	-1.73	-1.71	-1.64	-1.82	-1.52	-1.67	-1.87
5	-1.98	-1.96	-1.90	-2.16	-1.75	-1.90	-2.15
4	-2.27	-2.25	-2.19	-2.54	-2.03	-2.17	-2.47
3	-2.63	-2.61	-2.55	-3.00	-2.36	-2.48	-2.85
2	-3.09	-3.07	-3.02	-3.62	-2.81	-2.92	-3.38
1	-3.83	-3.81	-3.77	-4.59	-3.54	-3.62	-4.23

จากตาราง 11 ผลการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบรวมให้อยู่ในสเกลเดียวกัน พบว่าในกลุ่มทักษะวิชาภาษาไทย แบบสอบชุด X ที่สอบโดยผู้สอบสองกลุ่มมีค่าประมาณแตกต่างกันเล็กน้อย กล่าวคือ ในกลุ่มที่ 1 มีค่าต่ำสุด -3.83 ค่าสูงสุด 3.91 ส่วนในกลุ่มที่ 2 มีค่าต่ำสุด -3.81 ค่าสูงสุด 3.78 และเมื่อพิจารณาในแต่ละคู่ของค่าประมาณพารามิเตอร์ความสามารถซึ่งตรงกับคะแนนดิบเดียวกัน พบว่า ความแตกต่างในแต่ละคู่มิมีค่าไม่เกิน 0.13 และมีค่าเฉลี่ยของความแตกต่างเท่ากับ $.003$

สำหรับค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบชุด Y และชุด Z ซึ่งต่างก็แปลงให้อยู่ในสเกลเดียวกันกับค่าประมาณความสามารถของผู้สอบแบบสอบชุด X ในกลุ่มที่ 1 แล้วนั้น พบว่า แบบสอบทั้งสองชุดต่างมีพิสัยของความสามารถที่กว้างขึ้นกว่าเดิม ดังนี้ ค่าประมาณที่แปลงแล้วของชุด Y มีค่าต่ำสุด -4.59 ค่าสูงสุด 5.30 ส่วนชุด Z มีค่าต่ำสุด -4.23 และค่าสูงสุด 4.08

ตาราง 12 ผลการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบร่วม ที่สอบแบบสอบสามชุด
ในกลุ่มทักษะวิชาคณิตศาสตร์ ให้อยู่ในสเกลเดียวกัน

คะแนนดิบ	ชุด X		ชุด Y		ชุด Z		
	θx_1	θx_2	θy_1	$\theta' x_1$	θz_2	$\theta' x_2$	$\theta'' x_1$
37	3.94	3.88	3.91	4.74	3.75	3.27	3.26
36	3.20	3.15	3.18	3.93	3.03	2.60	2.61
35	2.74	2.70	2.74	3.44	2.59	2.19	2.22
34	2.40	2.36	2.40	3.07	2.27	1.90	1.94
33	2.12	2.09	2.13	2.77	2.01	1.66	1.71
32	1.89	1.86	1.90	2.51	1.79	1.45	1.50
31	1.68	1.65	1.69	2.28	1.59	1.26	1.32
30	1.49	1.47	1.51	2.08	1.42	1.11	1.18
29	1.32	1.30	1.34	1.89	1.26	.96	1.03
28	1.16	1.14	1.18	1.71	1.11	.82	.90
27	1.01	.99	1.03	1.55	.97	.69	.77
26	.87	.85	.88	1.38	.83	.56	.65
25	.73	.72	.74	1.23	.71	.45	.54
24	.60	.59	.61	1.08	.58	.33	.43
23	.47	.46	.48	.94	.46	.21	.31
22	.35	.34	.35	.79	.35	.11	.22
21	.23	.22	.23	.66	.23	.01	.12
20	.11	.10	.10	.52	.12	-.10	.01

ตาราง 12 (ต่อ)

คะแนนดิบ	ชุด X		ชุด Y		ชุด Z		
	θX_1	θX_2	θY_1	$\theta' X_1$	θZ_2	$\theta' X_2$	$\theta'' X_1$
19	-.01	-.02	-.02	.38	.00	-.21	-.09
18	-.13	-.13	-.14	.25	-.11	-.32	-.20
17	-.25	-.25	-.26	.12	-.22	-.42	-.29
16	-.38	-.37	-.39	-.03	-.34	-.53	-.40
15	-.50	-.49	-.51	-.16	-.46	-.64	-.51
14	-.62	-.62	-.64	-.30	-.58	-.75	-.61
13	-.75	-.74	-.77	-.45	-.70	-.87	-.73
12	-.89	-.88	-.90	-.59	-.83	-.99	-.84
11	-1.03	-1.01	-1.04	-.75	-.96	-1.11	-.96
10	-1.17	-1.16	-1.19	-.91	-1.10	-1.24	-1.08
9	-1.33	-1.31	-1.34	-1.08	-1.25	-1.38	-1.22
8	-1.49	-1.47	-1.51	-1.27	-1.42	-1.53	-1.36
7	-1.68	-1.65	-1.69	-1.47	-1.59	-1.68	-1.51
6	-1.88	-1.85	-1.88	-1.68	-1.79	-1.87	-1.69
5	-2.10	-2.07	-2.11	-1.94	-2.01	-2.08	-1.89
4	-2.37	-2.33	-2.37	-2.22	-2.28	-2.33	-2.13
3	-2.70	-2.66	-2.69	-2.58	-2.60	-2.63	-2.42
2	-3.15	-3.10	-3.13	-3.07	-3.05	-3.05	-2.82
1	-3.88	-3.82	-3.85	-3.86	-3.77	-3.72	-3.47

จากตาราง 12 ผลการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบร่วมที่ทำแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน พบว่าในกลุ่มทักษะวิชาคณิตศาสตร์ แบบสอบชุด X ที่สอบโดยผู้สอบสองกลุ่มมีค่าประมาณที่แตกต่างกันเล็กน้อย กล่าวคือกลุ่มที่ 1 มีค่าต่ำสุด -3.88 ค่าสูงสุด 3.94 ส่วนกลุ่มที่ 2 มีค่าต่ำสุด -3.82 ค่าสูงสุด 3.88 และเมื่อพิจารณาในแต่ละคู่ของค่าประมาณพารามิเตอร์ความสามารถที่ตรงกับคะแนนดิบค่าเดียวกัน พบว่าความแตกต่างกันในแต่ละคู่มีค่าไม่เกิน $.06$ และค่าเฉลี่ยของความแตกต่างเท่ากับ $.002$

สำหรับค่าประมาณพารามิเตอร์ความสามารถของผู้สอบชุด Y และชุด Z ซึ่งต่างก็แปลงให้อยู่ในสเกลเดียวกันกับค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบชุด X ในกลุ่มที่ 1 แล้วนั้นพบว่า ค่าที่แปลงแล้วของชุด Y มีค่าต่ำสุด -3.86 ค่าสูงสุด 4.74 ซึ่งมีพิสัยที่กว้างขึ้นกว่าเดิม ในช่วงปลายด้านคะแนนสูง ส่วนชุด Z มีค่าต่ำสุด -3.47 ค่าสูงสุด 3.26 ซึ่งมีพิสัยที่แคบลงกว่าเดิม

เมื่อพิจารณาผลการวิเคราะห์ในแต่ละกลุ่มวิชา หลังจากที่ได้แปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแต่ละชุดให้อยู่ในสเกลเดียวกันแล้ว พบว่าหากใช้ผู้สอบกลุ่มเดียวทำแบบสอบทั้งสามชุด ผู้สอบแต่ละระดับความสามารถจะได้คะแนนหรือคะแนนรวมจากการสอบแบบสอบแต่ละชุดแตกต่างกัน ดังนั้นจึงต้องทำการเทียบคะแนนระหว่างแบบสอบต่างชุดที่ประกอบไปด้วยข้อสอบต่างข้อกัน เพื่อให้ได้คะแนนที่สมมูลกันระหว่างแบบสอบซึ่งมีคุณสมบัติที่จะนำคะแนนมาเปรียบเทียบกันได้โดยตรง

สำหรับในการเทียบคะแนนเพื่อศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนตั้งแต่ปีการศึกษา 2529 ถึง 2531 ในการวิจัยครั้งนี้เดิมไม่มีข้อมูลจากแบบสอบร่วมหรือผู้สอบร่วมดังกล่าวมาแล้วนั้น ผู้วิจัยได้ใช้ค่าคงที่และสมการเชิงเส้นตรงที่ศึกษาจากกลุ่มผู้สอบร่วมนี้ เพื่อเป็นตัวเชื่อมข้อมูลที่ได้จากแบบสอบต่างชุดนั้นเข้าด้วยกัน (Rentz and Beshaw. 1977 : 162 - 163) โดยจะดำเนินการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบต่างกลุ่มที่ได้รับการสอบด้วยแบบสอบในแต่ละปีการศึกษา ทั้งนี้จะทำการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบในปีการศึกษา 2530 และในปีการศึกษา 2531 ไปสู่สเกลความสามารถเดียวกันกับสเกลของค่าประมาณพารามิเตอร์ความสามารถของผู้สอบในปีการศึกษา 2529 ปราบผลการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบในแต่ละปีการศึกษาที่ได้คะแนนดิบต่าง ๆ กัน

ดังตาราง 13

ตาราง 13 ผลการใช้ค่าคงที่และสมการในการแปลงค่าประมาณพารามิเตอร์ความสามารถของ
ผู้สอบแบบสอบสามชุดในกลุ่มทักษะวิชาภาษาไทยที่ดำเนินการสอบเมื่อปีการศึกษา 2529 ถึง 2531
ให้อยู่ในสเกลเดียวกัน

คะแนนดิบ	ปีการศึกษา 2529	ปีการศึกษา 2530		ปีการศึกษา 2531		
	θ_{x_1}	θ_{y_1}	θ'_{x_1}	θ_{z_2}	θ'_{x_2}	θ'_{x_1}
29	3.68	3.77	5.22	3.56	3.23	4.07
28	2.95	3.02	4.25	2.82	2.52	3.21
27	2.50	2.54	3.62	2.37	2.08	2.67
26	2.16	2.18	3.15	2.03	1.76	2.29
25	1.88	1.88	2.76	1.75	1.49	1.96
24	1.64	1.62	2.43	1.51	1.25	1.67
23	1.42	1.39	2.13	1.30	1.05	1.43
22	1.22	1.18	1.85	1.11	.87	1.21
21	1.04	.99	1.60	.93	.69	.99
20	.86	.80	1.36	.76	.53	.80
19	.69	.63	1.14	.60	.37	.60
18	.53	.46	.91	.44	.22	.42
17	.36	.30	.71	.29	.08	.25
16	.20	.14	.50	.14	-.07	.07
15	.04	-.02	.29	-.01	-.21	-.10
14	-.12	-.17	.09	-.15	-.35	-.27
13	-.29	-.33	-.11	-.30	-.49	-.44

ตาราง 13 (ต่อ)

คะแนนดิบ	ปีการศึกษา 2529	ปีการศึกษา 2530		ปีการศึกษา 2531		
	θx_1	θy_1	$\theta' x_1$	θz_2	$\theta' x_2$	$\theta'' x_1$
12	-.46	-.49	-.32	-.45	-.64	-.62
11	-.63	-.65	-.53	-.61	-.79	-.80
10	-.81	-.82	-.75	-.77	-.95	-1.00
9	-1.00	-1.00	-.99	-.93	-1.10	-1.18
8	-1.20	-1.19	-1.23	-1.11	-1.28	-1.40
7	-1.42	-1.39	-1.49	-1.30	-1.46	-1.61
6	-1.65	-1.62	-1.79	-1.51	-1.66	-1.86
5	-1.91	-1.87	-2.12	-1.75	-1.90	-2.15
4	-2.21	-2.16	-2.50	-2.02	-2.16	-2.46
3	-2.57	-2.51	-2.95	-2.36	-2.48	-2.85
2	-3.04	-2.98	-3.56	-2.80	-2.91	-3.37
1	-3.79	-3.73	-4.54	-3.53	-3.61	-4.22

จากตาราง 13 ผลการใช้ค่าคงที่และสมการในการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน สำหรับในกลุ่มทักษะวิชาภาษาไทย พบว่าโดยทั่วไปผู้สอบที่มีระดับความสามารถเดียวกัน หากได้รับการสอบด้วยแบบสอบต่างชุดกันจะได้คะแนนดิบต่างกัน

ส่วนผลการใช้ค่าคงที่และสมการในการแปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบต่างชุดให้อยู่ในสเกลเดียวกันสำหรับกลุ่มทักษะวิชาคณิตศาสตร์นั้นปรากฏผลดังตาราง 14

ตาราง 14 ผลการใช้ค่าคงที่และสมการในการแปลงค่าประมาณพารามิเตอร์ความสามารถของ
ผู้สอบแบบสอบสามชุดในกลุ่มทักษะวิชาคณิตศาสตร์ ที่ดำเนินการสอบเมื่อปีการศึกษา 2529 ถึง
2531 ให้อยู่ในสเกลเดียวกัน

คะแนนดิบ	ปีการศึกษา 2529	ปีการศึกษา 2530		ปีการศึกษา 2531		
	θX_1	θY_1	$\theta' X_1$	θZ_2	$\theta' X_2$	$\theta'' X_1$
37	3.83	3.88	4.71	3.72	3.25	3.24
36	3.11	3.15	3.90	3.00	2.58	2.59
35	2.66	2.69	3.39	2.57	2.18	2.21
34	2.33	2.36	3.02	2.24	1.87	1.91
33	2.06	2.08	2.71	1.98	1.63	1.67
32	1.83	1.85	2.46	1.76	1.42	1.48
31	1.63	1.65	2.23	1.57	1.25	1.31
30	1.45	1.47	2.04	1.39	1.08	1.15
29	1.29	1.30	1.85	1.23	.93	1.00
28	1.13	1.14	1.67	1.09	.80	.88
27	.99	.99	1.50	.95	.67	.75
26	.85	.86	1.36	.82	.55	.64
25	.72	.72	1.20	.69	.43	.52
24	.59	.59	1.06	.57	.32	.42
23	.47	.47	.93	.45	.20	.30
22	.35	.35	.79	.34	.10	.21
21	.23	.23	.66	.22	-.01	.10
20	.11	.11	.53	.11	-.11	.00

ตาราง 14 (ต่อ)

คะแนนดิบ	ปีการศึกษา 2529	ปีการศึกษา 2530		ปีการศึกษา 2531		
	θx_1	θy_1	$\theta' x_1$	θz_2	$\theta' x_2$	$\theta'' x_1$
19	-.01	-.01	.39	.00	-.21	-.09
18	.12	-.13	.26	-.11	-.32	-.20
17	-.24	-.25	.13	-.23	-.43	-.30
16	-.36	-.37	-.01	-.34	-.53	-.40
15	-.48	-.49	-.14	-.46	-.64	-.51
14	-.60	-.61	-.27	-.57	-.74	-.60
13	-.73	-.74	-.42	-.69	-.86	-.72
12	-.86	-.87	-.56	-.82	-.98	-.83
11	-1.00	-1.01	-.72	-.95	-1.10	-.95
10	-1.14	-1.15	-.87	-1.09	-1.23	-1.07
9	-1.29	-1.31	-1.05	-1.24	-1.37	-1.21
8	-1.45	-1.47	-1.23	-1.40	-1.52	-1.35
7	-1.63	-1.65	-1.42	-1.57	-1.67	-1.50
6	-1.83	-1.85	-1.65	-1.76	-1.85	-1.67
5	-2.05	-2.07	-1.89	-1.98	-2.06	-1.87
4	-2.31	-2.34	-2.19	-2.24	-2.30	-2.10
3	-2.64	-2.66	-2.54	-2.56	-2.59	-2.38
2	-3.08	-3.10	-3.03	-3.00	-3.00	-2.78
1	-3.80	-3.82	-3.83	-3.71	-3.66	-3.41

จากตาราง 14 ผลการใช้ค่าคงที่และสมการในการแปลงค่าประมาณพารามิเตอร์
ความสามารถของผู้สอบในแต่ละปีการศึกษาให้อยู่ในสเกลเดียวกัน สำหรับในกลุ่มทักษะวิชาคณิตศาสตร์
พบว่าผู้สอบที่มีระดับความสามารถเดียวกันหากได้รับการสอบด้วยแบบสอบต่างชุดกันจะได้คะแนนดิบ
ต่างกัน

ผลการวิเคราะห์ระดับความสามารถของผู้สอบที่สอบแบบสอบแต่ละชุด ซึ่งแปลงให้อยู่ในสเกล
เดียวกันแล้วนี้ จะนำมาใช้สำหรับการเทียบหาคะแนนที่สมมูลกันระหว่างแบบสอบ 3 ชุด ในแต่ละกลุ่ม
วิชา เพื่อศึกษาผลสัมฤทธิ์ทางการเรียนในแต่ละปีการศึกษา และความก้าวหน้าของผลสัมฤทธิ์ทางการ
เรียนในช่วง 3 ปีการศึกษาต่อไป ซึ่งในการเทียบหาคะแนนสมมูลอาจทำได้โดยการเทียบคะแนนดิบ
กันระหว่างแบบสอบที่ตรงกับระดับความสามารถเดียวกัน หรือการเทียบคะแนนจริงระหว่างแบบสอบ
ที่ตรงกับคะแนนค่าเดียวกันของแบบสอบแต่ละชุด สำหรับการเทียบคะแนนที่สมมูลกันระหว่าง
แบบสอบนั้นอาศัยการเขียนเส้นกราฟแสดงค่าคะแนนดิบในแต่ละระดับความสามารถของแบบสอบ
แต่ละชุด และเทียบหาคะแนนที่สมมูลกันที่ระดับความสามารถเดียวกัน จึงอาจมีการคลาดเคลื่อน
เนื่องมาจากการเขียนและการแปลเส้นกราฟให้เรียบร้อยเหตุนี้ผู้วิจัยจึงทำการเทียบคะแนนโดยใช้
คะแนนจริง ซึ่งจะไม่มีความคลาดเคลื่อนนี้ ทั้งนี้โดยนำค่าประมาณพารามิเตอร์ที่ได้ผ่านการ
แปลงให้อยู่ในสเกลเดียวกันดังกล่าวแล้วนี้มาทำการคำนวณหาค่าคะแนนจริงของผู้สอบแบบสอบ
แต่ละชุดในแต่ละกลุ่มวิชาที่คะแนนดิบต่าง ๆ โดยใช้สูตรในการคำนวณดังนี้ (Hambleton. 1985 :
212 ; Lord. 1980 : 199 - 203)

$$\xi = \sum_{i=1}^l P_i(\theta_x)$$

$$\text{โดยที่ } P_i(\theta) = \frac{\frac{Da_i(\theta - b_i)}{l}}{1 + \frac{Da_i(\theta - b_i)}{l}} \quad ; \quad i = 1, 2, 3, \dots, n$$

เมื่อ ξ	คือคะแนนจริงของผู้สอบแบบสอบแต่ละชุด
$P_i(\theta)$	คือความน่าจะเป็นที่ผู้สอบซึ่งมีความสามารถ θ จะตอบข้อสอบข้อที่ i ในแบบสอบแต่ละชุดได้ถูกต้อง
b_i	คือค่าความยากของข้อสอบข้อที่ i
a_i	คือค่าเฉลี่ยของค่าอำนาจจำแนกของข้อสอบให้มีค่าเท่ากับ i
D	คือ Scaling factor มีค่าเท่ากับ 1.702
θ	คือระดับความสามารถของผู้สอบ
e	คือค่าคงที่มีค่าเท่ากับ 2.7182818

หลังจากนั้นผู้วิจัยนำค่าประมาณคะแนนจริงของผู้สอบแบบสอบแต่ละชุดมาเทียบคะแนนกัน ณ ที่คะแนนดิบค่าเดียวกัน ซึ่งปรากฏดังตาราง 15 และ 16

ตาราง 15 ผลการเทียบคะแนนจริงในกลุ่มทักษะวิชาภาษาไทย ตั้งแต่ปีการศึกษา 2529 ถึง 2531

คะแนนดิบ	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	θx_1	คะแนนเทียบ (คะแนนจริง)	θx_1	คะแนนเทียบ (คะแนนจริง)	θx_1	คะแนนเทียบ (คะแนนจริง)
29	3.68	29.8	5.22	29.9	4.07	29.9
28	2.95	29.4	4.25	29.9	3.21	29.6
27	2.50	28.8	3.62	29.6	2.67	29.2
26	2.16	28.0	3.15	29.3	2.29	28.6
25	1.88	27.0	2.76	28.8	1.96	27.9
24	1.64	26.0	2.43	28.2	1.67	26.9
23	1.42	24.8	2.13	27.5	1.43	25.9

ตาราง 15 (ต่อ)

คะแนนดิบ	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	θX_1	คะแนนเทียบ (คะแนนจริง)	$\theta' X_1$	คะแนนเทียบ (คะแนนจริง)	$\theta'' X_1$	คะแนนเทียบ (คะแนนจริง)
22	1.22	23.6	1.85	26.6	1.21	24.8
21	1.04	22.4	1.60	25.7	.99	23.4
20	.86	21.1	1.36	24.6	.80	22.0
19	.69	19.8	1.14	23.5	.60	20.5
18	.53	18.5	.91	22.2	.42	18.9
17	.36	17.1	.71	20.9	.25	17.3
16	.20	15.8	.50	19.4	.07	15.7
15	-.04	14.5	.29	17.7	-.10	14.1
14	-.12	13.3	.09	16.0	-.27	12.5
13	-.29	12.1	-.11	14.3	-.44	11.0
12	-.46	10.9	-.32	12.5	-.62	9.4
11	-.63	9.8	-.53	10.7	-.80	8.0
10	-.81	8.7	-.75	9.0	-1.00	6.5
9	-1.00	7.6	-.99	7.3	-1.18	5.4
8	-1.20	6.5	-1.23	5.8	-1.40	4.2
7	-1.42	5.4	-1.49	4.5	-1.61	3.2
6	-1.65	4.4	-1.79	3.3	-1.86	2.3
5	-1.91	3.3	-2.12	2.2	-2.15	1.5

ตาราง 15 (ต่อ)

คะแนนดิบ	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	θx_1	คะแนนเทียบ (คะแนนจริง)	$\theta' x_1$	คะแนนเทียบ (คะแนนจริง)	$\theta'' x_1$	คะแนนเทียบ (คะแนนจริง)
4	-2.21	2.3	-2.50	1.4	-2.46	0.9
3	-2.57	1.4	-2.95	0.7	-2.85	0.5
2	-3.04	0.7	-3.56	0.3	-3.37	0.2
1	-3.79	0.2	-4.54	0.0	-4.22	0.0

จากตาราง 15 ผลการเทียบคะแนนจริงในกลุ่มทักษะวิชาภาษาไทยตั้งแต่ปีการศึกษา 2529 ถึง 2531 พบว่าคะแนนที่เทียบแล้วของแบบสอบแต่ละชุดที่ตรงกับคะแนนดิบค่าเดียวกันมีค่าแตกต่างกัน โดยทั่วไปคะแนนที่เทียบแล้วของปีการศึกษา 2530 จะมีค่าสูงสุด รองลงมาคือคะแนนเทียบของปีการศึกษา 2531 ส่วนคะแนนเทียบของปีการศึกษา 2529 พบว่ามีค่าต่ำกว่าคะแนนเทียบของปีการศึกษาอื่น ทั้งนี้คะแนนเทียบระหว่างสามปีการศึกษาจะมีค่าใกล้เคียงกันมากในช่วงที่ตรงกับคะแนนดิบทางปลายด้านคะแนนสูงมากและต่ำมาก ส่วนในช่วงที่ตรงกับคะแนนดิบที่อยู่ตรงกลางจะมีค่าแตกต่างกันมากขึ้น กล่าวคือในช่วงคะแนนดิบตั้งแต่ 1 ถึง 9 คะแนนและตั้งแต่ 14 ถึง 29 คะแนน จะมีความแตกต่างของคะแนนเทียบระหว่างสามปีการศึกษาไม่เกิน 2.3 คะแนน ส่วนในช่วงคะแนนดิบตั้งแต่ 10 ถึง 23 คะแนน พบว่ามีความแตกต่างของคะแนนเทียบตั้งแต่ 2.5 ถึง 3.8 คะแนน

ตาราง 16 ผลการเทียบคะแนนจริงในกลุ่มทักษะวิชาคณิตศาสตร์ ตั้งแต่ปีการศึกษา 2529 ถึง 2531

คะแนนดิบ	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	θx_1	คะแนนเทียบ (คะแนนจริง)	$\theta' x_1$	คะแนนเทียบ (คะแนนจริง)	$\theta'' x_1$	คะแนนเทียบ (คะแนนจริง)
37	3.83	37.8	4.71	37.9	3.24	37.6
36	3.11	37.4	3.90	37.8	2.59	37.8
35	2.66	36.9	3.39	37.5	2.21	36.4
34	2.33	36.3	3.02	37.3	1.91	35.6
33	2.06	35.6	2.71	36.9	1.67	34.7
32	1.83	34.7	2.46	36.4	1.48	33.8
31	1.63	33.8	2.23	35.9	1.31	32.8
30	1.45	32.9	2.04	35.3	1.15	31.7
29	1.29	31.9	1.85	34.7	1.00	30.5
28	1.13	30.8	1.67	33.9	.88	29.5
27	.99	29.7	1.50	33.1	.75	28.2
26	.85	28.5	1.36	32.2	.64	27.0
25	.72	27.3	1.20	31.2	.52	25.7
24	.59	26.0	1.06	30.1	.42	24.5
23	.47	24.7	.93	29.1	.30	23.0
22	.35	23.4	.79	27.8	.21	21.8
21	.23	22.0	.66	26.6	.10	20.3
20	.11	20.5	.53	25.2	.00	19.0
19	-.01	19.1	.39	23.7	-.09	17.8

ตาราง 16 (ต่อ)

คะแนนดิบ	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	θx_1	คะแนนเทียบ (คะแนนจริง)	$\theta' x_1$	คะแนนเทียบ (คะแนนจริง)	$\theta'' x_1$	คะแนนเทียบ (คะแนนจริง)
18	-.12	17.7	.26	22.2	-.20	16.3
17	-.24	16.3	.13	20.7	-.30	15.0
16	-.36	14.8	-.01	19.1	-.40	13.8
15	-.48	13.4	-.14	17.5	-.51	12.4
14	-.60	12.1	-.27	16.0	-.60	11.4
13	-.73	10.7	-.42	14.3	-.72	10.1
12	-.86	9.4	-.56	12.7	-.83	9.0
11	-1.00	8.2	-.72	11.0	-.95	7.9
10	-1.14	7.0	-.87	9.6	-1.07	6.8
9	-1.29	5.9	-1.05	8.0	-1.21	5.8
8	-1.45	4.9	-1.23	6.6	-1.35	4.9
7	-1.63	3.9	-1.42	5.3	-1.50	4.0
6	-1.83	3.0	-1.65	4.0	-1.67	3.2
5	-2.05	2.2	-1.89	2.9	-1.87	2.4
4	-2.31	1.5	-2.19	1.9	-2.10	1.7
3	-2.64	0.9	-2.54	1.1	-2.38	1.1
2	-3.08	0.4	-3.03	0.5	-2.78	0.6
1	-3.80	0.1	-3.83	0.1	-3.41	0.2

จากตาราง 16 ผลการเทียบคะแนนจริงในกลุ่มทักษะวิชาคณิตศาสตร์ตั้งแต่ปีการศึกษา 2529 ถึง 2531 พบว่าคะแนนที่เทียบแล้วของแบบสอบแต่ละชุดที่ตรงกับคะแนนดิบค่าเดียวกันมีค่าแตกต่างกัน ซึ่งโดยทั่วไปคะแนนที่เทียบแล้วของปีการศึกษา 2530 จะมีค่าสูงสุด รองลงมาคือคะแนนเทียบของปีการศึกษา 2529 ส่วนคะแนนเทียบของปีการศึกษา 2531 พบว่ามีค่าต่ำกว่าคะแนนเทียบของปีการศึกษาอื่น ทั้งนี้คะแนนเทียบระหว่างสามปีการศึกษาจะมีค่าใกล้เคียงกันมากในช่วงที่ตรงกับคะแนนดิบทางปลายด้านคะแนนสูงมากและต่ำมาก ส่วนในช่วงที่ตรงกับคะแนนดิบที่อยู่ตรงกลางจะมีค่าแตกต่างกันมากขึ้น กล่าวคือ ในช่วงคะแนนดิบตั้งแต่ 1 ถึง 9 คะแนน และตั้งแต่ 33 ถึง 37 คะแนน จะมีความแตกต่างของคะแนนเทียบไม่เกิน 2.2 คะแนน ส่วนในช่วงคะแนนดิบตั้งแต่ 10 ถึง 32 คะแนน พบว่ามีความแตกต่างของคะแนนเทียบตั้งแต่ 2.6 ถึง 6.3 คะแนน

ผู้วิจัยแจกแจงความถี่และทำการเทียบคะแนนจริงของผลการสอบที่ได้ดำเนินการไปแล้วเมื่อปีการศึกษา 2529 ถึง 2531 ตามโครงการประเมินคุณภาพนักเรียนชั้นประถมปีที่ 6 ระดับประเทศ ดังกล่าว ทั้งนี้โดยคิดคะแนนเทียบเป็นค่าจำนวนเต็มทีใกล้ที่สุด แล้วคำนวณหาค่าคะแนนเฉลี่ยของผลการสอบในแต่ละปีการศึกษา รวมทั้งสัดส่วนจำนวนนักเรียนที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจ (คะแนนเฉลี่ยตั้งแต่ร้อยละ 50 ขึ้นไปของคะแนนเต็มในแต่ละกลุ่มวิชา) นอกจากนี้ผู้วิจัยจะศึกษาถึงความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในแต่ละปีการศึกษาที่ผ่านมาจากคะแนนที่เทียบให้อยู่ในสเกลเดียวกันทั้งหมดแล้ว ซึ่งมีคุณสมบัติที่จะนำมาเปรียบเทียบกันได้โดยตรง ดังผลการแจกแจงความถี่ของคะแนนในแต่ละปีการศึกษา และคะแนนเทียบ (จำนวนเต็ม) ดังตาราง 17 และ 18 และแสดงค่าเฉลี่ยค่าเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษาเมื่อได้เทียบคะแนนแล้ว ตลอดจนความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในช่วงสามปีการศึกษา ดังตาราง 19

ตาราง 17 ผลการแจกแจงความถี่ของคะแนนที่ได้จากการสอบเมื่อปีการศึกษา 2529 ถึง 2531
และคะแนนเทียบในแต่ละปีการศึกษา ในกลุ่มทักษะวิชาภาษาไทย

คะแนนดิบ	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	ความถี่	คะแนนเทียบ (คะแนนจริง)	ความถี่	คะแนนเทียบ (คะแนนจริง)	ความถี่	คะแนนเทียบ (คะแนนจริง)
26					4	29
25					3	28
24	1	26	1	28	4	27
23	5	25	3	28	18	26
22	5	24	9	27	29	25
21	11	22	11	26	26	23
20	29	21	15	25	34	22
19	34	20	23	24	45	21
18	51	19	34	22	50	19
17	68	17	56	21	67	17
16	60	16	48	19	42	16
15	59	15	45	18	33	14
14	52	13	43	16	31	13
13	41	12	39	14	32	11
12	45	11	39	13	18	9
11	17	10	27	11	12	8
10	22	9	28	9	5	7

ตาราง 17 (ต่อ)

คะแนนดิบ	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	ความถี่	คะแนนเทียบ (คะแนนจริง)	ความถี่	คะแนนเทียบ (คะแนนจริง)	ความถี่	คะแนนเทียบ (คะแนนจริง)
9	20	8	10	7	2	5
8	4	7	13	6	3	4
7	6	5	6	5	5	3
6	3	4	2	3	1	2
5	2	3	1	2	1	2

จากตาราง 17 ในกลุ่มทักษะวิชาภาษาไทย เมื่อเทียบคะแนนจริงจากแบบสอบชุด Y และชุด Z ไปสู่สเกลเดียวกับสเกลคะแนนจริงของแบบสอบชุด Z แล้วพบว่า คะแนนจุดตัดของแบบสอบแต่ละชุดมีค่าแตกต่างกัน กล่าวคือ คะแนนจุดตัดของชุด Y มีค่าสูงสุด คือมีค่าเท่ากับ 18 คะแนน รองลงมาคือชุด X ซึ่งมีคะแนนจุดตัดเท่ากับ 15 คะแนน และชุด Z มีคะแนนจุดตัดต่ำสุด คือเท่ากับ 14 คะแนน

ตาราง 18 ผลการแจกแจงความถี่ของคะแนนที่ได้จากการสอบเมื่อปีการศึกษา 2529 ถึง 2531 และคะแนนเทียบในแต่ละปีการศึกษา ในกลุ่มทักษะวิชาคณิตศาสตร์

คะแนนดิบ	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	ความถี่	คะแนนเทียบ (คะแนนจริง)	ความถี่	คะแนนเทียบ (คะแนนจริง)	ความถี่	คะแนนเทียบ (คะแนนจริง)
32					1	34
31						
30			1	35		
29			1	35		
28			1	34	5	30
27			5	33	3	28
26			5	32	8	27
25	2	27	6	31	6	26
24	4	26	17	30	9	25
23	8	25	11	29	12	23
22	9	23	19	28	8	22
21	11	22	15	27	18	20
20	14	21	27	25	21	19
19	26	19	28	24	26	18
18	37	18	24	22	19	16

ตาราง 18 (ต่อ)

คะแนนดิบ	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	ความถี่	คะแนนเทียบ (คะแนนจริง)	ความถี่	คะแนนเทียบ (คะแนนจริง)	ความถี่	คะแนนเทียบ (คะแนนจริง)
17	26	16	35	21	25	15
16	43	15	30	19	31	14
15	44	13	36	18	33	12
14	42	12	30	16	34	11
13	50	11	33	14	51	10
12	50	9	27	13	37	9
11	49	8	28	11	40	8
10	41	7	24	10	34	7
9	36	6	20	3	18	6
8	18	5	12	7	9	5
7	14	4	12	5	10	4
6	5	3	4	4	5	3
5	4	2	1	3	1	2
3	2	1	1	1	1	1

จากตาราง 18 ในกลุ่มทักษะวิชาคณิตศาสตร์ เมื่อเทียบคะแนนจริงจากแบบสอบชุด Y และชุด Z ไปอยู่ในสเกลเดียวกับสเกลคะแนนจริงของแบบสอบชุด X แล้ว พบว่าคะแนนจุดตัดของแต่ละชุดมีค่าแตกต่างกัน กล่าวคือคะแนนจุดตัดของชุด Y มีค่าสูงสุด คือมีค่าเท่ากับ 24 คะแนน รองลงมาคือ X ซึ่งมีคะแนนจุดตัดเท่ากับ 19 คะแนน และชุด Z มีคะแนนจุดตัดต่ำสุดคือเท่ากับ 18 คะแนน

จากตาราง 17 และ 18 แสดงให้เห็นว่าโดยทั่วไปหากผู้สอบได้รับการสอบด้วยแบบสอบชุด Y จะได้รับคะแนนสูงกว่าได้รับการสอบด้วยแบบสอบชุดอื่น รองลงมาคือได้รับการสอบด้วยแบบสอบชุด X และชุด Z ตามลำดับ ซึ่งข้อค้นพบนี้มีลักษณะเหมือนกันทั้งสองกลุ่มวิชา

ตาราง 19 ค่าเฉลี่ย ค่าเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจในแต่ละกลุ่มวิชา และในแต่ละปีการศึกษา เมื่อเทียบคะแนนแล้ว (ตัวเลขในวงเล็บบรรทัดแรกคืออัตราความก้าวหน้าในแต่ละปีการศึกษา ตัวเลขในวงเล็บบรรทัดที่สองคืออัตราความก้าวหน้าในช่วงสามปีการศึกษา)

ปีการศึกษา กลุ่มวิชา	ปีการศึกษา 2529			ปีการศึกษา 2530			ปีการศึกษา 2531		
	$\bar{X}$	$\bar{X}\%$	%น.ร	$\bar{X}$	$\bar{X}\%$	%น.ร	$\bar{X}$	$\bar{X}\%$	%น.ร
ภาษาไทย	15.026	50.087	60	16.976	56.587	64	17.428	58.093	69
					(6.500)	(4)		(1.506)	(5)
								(8.006)	(9)
คณิตศาสตร์	12.082	31.795	14	18.486	48.647	50	12.987	34.176	20
					(16.852)	(36)		(-14.471)	(-30)
								(2.381)	(6)

จากตาราง 19 เมื่อเทียบคะแนนแล้ว พบว่า กลุ่มทักษะวิชาภาษาไทยมีคะแนนเฉลี่ยในแต่ละปีการศึกษา ตั้งแต่ 15.026 ถึง 17.428 คะแนน จากคะแนนเต็ม 30 คะแนน คิดเป็นร้อยละ 50.087 ถึง 58.093 คะแนนเฉลี่ยร้อยละต่ำสุดในปีการศึกษา 2529 คะแนนเฉลี่ยร้อยละสูงสุดในปีการศึกษา 2531 โดยมีอัตราความก้าวหน้าของคะแนนเฉลี่ยร้อยละเพิ่มขึ้นทุกปีการศึกษาค้างนี้ ในปีการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากปีการศึกษา 2529 เท่ากับร้อยละ 6.50 สำหรับในปีการศึกษา 2531 มีคะแนนเฉลี่ยเพิ่มขึ้นจากในปีการศึกษา 2530 เท่ากับร้อยละ 1.506 รวมมีค่าเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 8.006 ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษาพบว่า มีค่าตั้งแต่ร้อยละ 60 ถึง 69 ร้อยละต่ำสุดในปีการศึกษา 2529 ร้อยละสูงสุดในปีการศึกษา 2531 เมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจพบว่า ในปีการศึกษา 2530 มีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 4 ในปีการศึกษา 2531 มีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2530 อีกร้อยละ 5 ดังนั้นจึงมีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 9

สำหรับในกลุ่มทักษะวิชาคณิตศาสตร์พบว่า มีคะแนนเฉลี่ยผลการสอบในแต่ละปีตั้งแต่ 12.082 ถึง 18.486 จากคะแนนเต็ม 38 คะแนน เมื่อคิดเป็นค่าเฉลี่ยร้อยละมีค่าตั้งแต่ร้อยละ 31.795 ถึง 48.647 คะแนนเฉลี่ยร้อยละต่ำสุดในปีการศึกษา 2529 และคะแนนเฉลี่ยร้อยละสูงสุดในปีการศึกษา 2530 โดยทุกปีการศึกษามีคะแนนเฉลี่ยร้อยละต่ำกว่าร้อยละ 50 เมื่อพิจารณาอัตราความก้าวหน้าของคะแนนเฉลี่ยร้อยละในแต่ละปีการศึกษาพบว่าในปีการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นมากจากในปีการศึกษา 2529 คือเพิ่มขึ้นเท่ากับร้อยละ 16.852 แต่ในปีการศึกษา 2531 มีคะแนนเฉลี่ยร้อยละลดลงมากเช่นกัน คือลดลง เท่ากับร้อยละ 14.471 ซึ่งยังมีค่าเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 2.381 ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 14 ถึง 50 ซึ่งเมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจพบว่าในปีการศึกษา 2530 มีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 36 แต่ในปีการศึกษา 2531 มีจำนวนลดลงจากในปีการศึกษา 2530 เท่ากับร้อยละ 20 ซึ่งยังมีจำนวนสูงกว่าในปีการศึกษา 2529 เท่ากับร้อยละ 6

ในทำนองเดียวกัน ผู้วิจัยเทียบคะแนนผลการสอบในแต่ละกลุ่มวิชาจำแนกเป็นรายอำเภอ ซึ่งปรากฏผลการเทียบคะแนนในแต่ละปีการศึกษา ดังตาราง 20 และ 21

ตาราง 20 คะแนนเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจในกลุ่มทักษะวิชาภาษาไทยเมื่อเทียบคะแนนจำแนกเป็นรายอำเภอ (ตัวเลขในวงเล็บบรรทัดแรกคืออัตราความก้าวหน้าในแต่ละปีการศึกษา ตัวเลขในวงเล็บบรรทัดที่สองคืออัตราความก้าวหน้าในช่วงสามปีการศึกษา)

ปีการศึกษา กลุ่มวิชา	ปีการศึกษา 2529			ปีการศึกษา 2530			ปีการศึกษา 2531		
	N	$\bar{X}\%$	%น.ร	N	$\bar{X}\%$	%น.ร	N	$\bar{X}\%$	%น.ร
1. เมือง	185	51.910	66	119	62.689 (10.779)	77 (11)	148	61.689 (-1.000) (9.779)	78 (1) (12)
2. คลองหลวง	136	49.903	60	138	60.024 (10.121)	70 (10)	125	56.187 (-3.837) (6.284)	63 (-7) (3)
3. ลาดหลุมแก้ว	99	52.963	68	87	52.835 (-0.128)	54 (-14)	103	60.324 (7.489) (7.361)	74 (20) (6)
4. หนองเสือ	115	44.900	45	109	48.563 (3.663)	49 (4)	89	52.210 (3.647) (7.310)	58 (9) (13)

จากตาราง 20 เมื่อเทียบคะแนนผลการสอบในกลุ่มทักษะวิชาภาษาไทย ซึ่งมีคะแนนเต็ม 30 คะแนน จำแนกเป็นรายอำเภอแล้วพบว่า อำเภอเมืองมีคะแนนเฉลี่ยร้อยละในสามปีการศึกษา มากกว่าร้อยละ 50 ทุกปีการศึกษาคือมีค่าตั้งแต่ร้อยละ 51.910 ถึง 62.689 คะแนนเฉลี่ยร้อยละต่ำสุดในปีการศึกษา 2529 คะแนนเฉลี่ยร้อยละสูงสุดในปีการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของค่าคะแนนเฉลี่ยร้อยละพบว่า ในปีการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากปีการศึกษา 2529 เท่ากับร้อยละ 10.779 และในปีการศึกษา 2531 มีคะแนนเฉลี่ยร้อยละลดลงจากในปีการศึกษา 2530 เท่ากับร้อยละ 1 แต่ยังมีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 9.779 ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 66 ถึง 78 ร้อยละต่ำสุดในปีการศึกษา 2529 ร้อยละสูงสุดในปีการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจพบว่า ในปีการศึกษา 2530 มีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 11 และในปีการศึกษา 2531 มีจำนวนลดลงจากในปีการศึกษา 2530 เท่ากับร้อยละ 1 แต่ยังคงมีจำนวนที่เพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 12

อำเภอคลองหลวง มีคะแนนเฉลี่ยผลการสอบในแต่ละปีการศึกษาตั้งแต่ร้อยละ 49.903 ถึง 60.024 คะแนนเฉลี่ยร้อยละต่ำสุดในปีการศึกษา 2529 คะแนนเฉลี่ยร้อยละสูงสุดในปีการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของคะแนนเฉลี่ยร้อยละพบว่าในปีการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 10.121 และในปีการศึกษา 2531 มีคะแนนเฉลี่ยร้อยละลดลงจากในปีการศึกษา 2530 เท่ากับร้อยละ 3.837 แต่ยังมีอัตราที่เพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 6.284 ส่วนค่าร้อยละนักเรียนที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจ พบว่ามีค่าตั้งแต่ร้อยละ 60 ถึง 70 ร้อยละต่ำสุดในปีการศึกษา 2529 ร้อยละสูงสุดในปีการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจพบว่า ในปีการศึกษา 2530 มีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 10 ในปีการศึกษา 2531 มีจำนวนลดลงจากในปีการศึกษา 2530 เท่ากับร้อยละ 7 แต่ยังมีจำนวนที่เพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 3

อำเภอลาดหลุมแก้ว มีคะแนนเฉลี่ยร้อยละของผลการสอบในแต่ละปีการศึกษามากกว่าร้อยละ 50 ทุกปีการศึกษา คือมีคะแนนเฉลี่ยร้อยละตั้งแต่ร้อยละ 52.835 ถึง 60.324 คะแนนเฉลี่ยร้อยละต่ำสุดในปีการศึกษา 2530 คะแนนเฉลี่ยร้อยละสูงสุดในปีการศึกษา 2531 เมื่อพิจารณาอัตราความก้าวหน้าของคะแนนเฉลี่ยร้อยละพบว่าในปีการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละลดลงจากในปีการศึกษา 2529 เพียงเล็กน้อย คิดเป็นร้อยละ 0.128 ในปีการศึกษา 2531 มีค่าร้อยละเพิ่มขึ้นจากในปีการศึกษา 2530 เท่ากับร้อยละ 7.489 และมีค่าร้อยละเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 7.361 ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจพบว่า มีค่าตั้งแต่ร้อยละ 54 ถึง 74 ร้อยละต่ำสุดในปีการศึกษา 2530 ร้อยละสูงสุดในปีการศึกษา 2531 เมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจ พบว่าในปีการศึกษา 2530 มีจำนวนลดลงจากในปีการศึกษา 2529 เท่ากับร้อยละ 14 ในปีการศึกษา 2531 มีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2530 เท่ากับร้อยละ 20 และมีจำนวนที่เพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 6

ส่วนอำเภอนองเสือ มีคะแนนเฉลี่ยร้อยละของผลการสอบในแต่ละปีการศึกษาตั้งแต่ร้อยละ 44.900 ถึง 52.210 คะแนนเฉลี่ยร้อยละต่ำสุดในปีการศึกษา 2529 และร้อยละสูงสุดในปีการศึกษา 2531 โดยมีอัตราความก้าวหน้าเพิ่มขึ้นทุกปีการศึกษาดังนี้ ในปีการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 3.663 ในปีการศึกษา 2531 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากในปีการศึกษา 2530 เท่ากับร้อยละ 3.647 และเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 7.310 ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 45 ถึง 58 ร้อยละต่ำสุดในปีการศึกษา 2529 ร้อยละสูงสุดในปีการศึกษา 2531 เมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจ พบว่า ในปีการศึกษา 2530 มีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 4 ในปีการศึกษา 2531 มีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2530 อีกร้อยละ 9 ดังนั้นจึงมีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 13

ตาราง 21 คะแนนเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจในกลุ่มทักษะวิชาคณิตศาสตร์ เมื่อเทียบคะแนนเป็นรายอำเภอ (ตัวเลขในวงเล็บบรรทัดแรก คืออัตราความก้าวหน้าในแต่ละปีการศึกษาตัวเลขในวงเล็บบรรทัดที่สองคืออัตราความก้าวหน้าในช่วงสามปีการศึกษา)

ปีการศึกษา อำเภอ	ปีการศึกษา 2529			ปีการศึกษา 2530			ปีการศึกษา 2531		
	N	$\bar{X}\%$	%น.ร	N	$\bar{X}\%$	%น.ร	N	$\bar{X}\%$	%น.ร
1. เมือง	185	33.429	15	119	55.242	65	148	31.046	14
				(21.813)	(50)		(-24.196)	(-51)	
							(-2.383)	(-1)	
2. คลองหลวง	136	32.161	15	138	50.745	55	125	30.505	10
				(18.584)	(40)		(-20.240)	(-45)	
							(-1.656)	(-5)	
3. ลาดหลุมแก้ว	99	36.789	20	87	49.214	48	103	43.051	37
				(12.425)	(28)		(-6.163)	(-11)	
							(6.262)	(17)	
4. ท้องเสือ	115	24.395	6	109	38.339	28	89	34.270	24
				(13.944)	(44)		(-4.096)	(-4)	
							(9.875)	(18)	

จากตาราง 21 เมื่อเทียบคะแนนในกลุ่มทักษะวิชาคณิตศาสตร์ ซึ่งมีคะแนนเต็ม 38 คะแนน จำแนกเป็นรายอำเภอแล้ว ปรากฏผลดังนี้ อำเภอเมืองมีคะแนนเฉลี่ยร้อยละตั้งแต่ร้อยละ 31.046 ถึง 55.242 คะแนนเฉลี่ยร้อยละต่ำสุดในการศึกษา 2531 คะแนนเฉลี่ยร้อยละสูงสุดในการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของคะแนนเฉลี่ยร้อยละพบว่าในการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากในการศึกษา 2529 เท่ากับร้อยละ 21.813 ในการศึกษา 2531 มีค่าลดลงจากในการศึกษา 2530 เท่ากับร้อยละ 24.196 และมีค่าลดลงจากในการศึกษา 2529 เท่ากับร้อยละ 2.383 ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษาพบว่ามีค่าตั้งแต่ร้อยละ 14 ถึง 65 ค่าร้อยละต่ำสุดในการศึกษา 2531 ค่าร้อยละสูงสุดในการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจพบว่าในการศึกษา 2530 มีจำนวนร้อยละนักเรียนที่มีผลน่าพอใจลดลงจากในการศึกษา 2530 เท่ากับร้อยละ 51 ซึ่งลดลงจากในการศึกษา 2529 เท่ากับ ร้อยละ 1

อำเภอคลองหลวง มีคะแนนเฉลี่ยร้อยละตั้งแต่ร้อยละ 30.505 ถึงร้อยละ 50.745 คะแนนเฉลี่ยร้อยละต่ำสุดในการศึกษา 2531 คะแนนเฉลี่ยร้อยละสูงสุดในการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของคะแนนเฉลี่ยร้อยละ พบว่าในการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากในการศึกษา 2529 เท่ากับร้อยละ 18.584 ในการศึกษา 2531 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากในการศึกษา 2530 เท่ากับร้อยละ 20.240 และลดลงจากในการศึกษา 2529 เท่ากับ ร้อยละ 1.656 ส่วนค่าร้อยละนักเรียนที่มีผลที่น่าพอใจในแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 10 ถึง 55 ค่าร้อยละต่ำสุดในการศึกษา 2531 ค่าร้อยละสูงสุดในการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจพบว่าในการศึกษา 2530 มีจำนวนเพิ่มขึ้นจากในการศึกษา 2529 เท่ากับร้อยละ 40 ในการศึกษา 2531 มีจำนวนลดลงจากในการศึกษา 2530 เท่ากับร้อยละ 45 และลดลงจากในการศึกษา 2529 เท่ากับร้อยละ 5

อำเภอลาดหลุมแก้ว มีคะแนนเฉลี่ยร้อยละตั้งแต่ร้อยละ 36.789 ถึง 49.241 คะแนนเฉลี่ยร้อยละต่ำสุดในการศึกษา 2529 คะแนนเฉลี่ยร้อยละสูงสุดในการศึกษา 2530 เมื่อพิจารณา

อัตราความก้าวหน้าของคะแนนเฉลี่ยร้อยละ พบว่าในปีการศึกษา 2530 มีค่าเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 12.425 ในปีการศึกษา 2531 มีค่าลดลงจากในปีการศึกษา 2530 เท่ากับร้อยละ 6.163 แต่ยังคงเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 6.262 ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 20 ถึง 48 ร้อยละต่ำสุดในปีการศึกษา 2529 ร้อยละสูงสุดในปีการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจ พบว่าในปีการศึกษา 2530 มีจำนวนเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 28 ในปีการศึกษา 2531 มีจำนวนลดลงจากในปีการศึกษา 2530 เท่ากับร้อยละ 11 แต่เพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 17

อำเภอหนองเสือ มีคะแนนเฉลี่ยร้อยละของคะแนนผลการสอบตั้งแต่ร้อยละ 24.395 ถึง 38.339 คะแนนเฉลี่ยร้อยละต่ำสุดในปีการศึกษา 2529 คะแนนเฉลี่ยสูงสุดในปีการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของค่าเฉลี่ยร้อยละ พบว่าในปีการศึกษา 2530 มีค่าเฉลี่ยร้อยละเพิ่มขึ้นจากปีการศึกษา 2529 เท่ากับร้อยละ 13.944 ในปีการศึกษา 2531 มีค่าเฉลี่ยร้อยละลดลงจากในปีการศึกษา 2530 เท่ากับร้อยละ 4.069 แต่เพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 9.875 ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 6 ถึง 28 ร้อยละต่ำสุดในปีการศึกษา 2529 ร้อยละสูงสุดในปีการศึกษา 2530 เมื่อพิจารณาอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจ พบว่าในปีการศึกษา 2530 มีค่าร้อยละนักเรียนที่เพิ่มขึ้นกว่าในปีการศึกษา 2529 เท่ากับร้อยละ 22 ในปีการศึกษา 2531 มีค่าร้อยละนักเรียนที่มีผลน่าพอใจลดลงจากในปีการศึกษา 2530 เท่ากับร้อยละ 4 แต่ยังคงมีค่าเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 18

หลังจากนั้นผู้วิจัยทำการวิเคราะห์ความแปรปรวน (ANOVA) เมื่อเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยระหว่างสามปีการศึกษาในแต่ละกลุ่มวิชา ซึ่งปรากฏผลดังตาราง 22 และ 23

ตาราง 22 ผลการวิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบความแตกต่างของคะแนนเฉลี่ย
ระหว่างสามปีการศึกษา ในแต่ละกลุ่มวิชา

แหล่งความแปรปรวน	SS	df	MS	F
ภาษาไทย				
ระหว่างกลุ่ม	1651.270	2	825.635	31.589**
ภายในกลุ่ม	37898.200	1450	26.137	
รวม	39549.470	1452		
คณิตศาสตร์				
ระหว่างกลุ่ม	11362.550	2	5681.275	142.149**
ภายในกลุ่ม	57951.460	1450	39.967	
รวม	69314.011	1452		

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 22 พบว่าคะแนนเฉลี่ยระหว่าง 3 ปีการศึกษาของทั้งสองกลุ่มวิชา มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ตาราง 23 ผลการวิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบความแตกต่างของคะแนนเฉลี่ย
ระหว่างสามปีการศึกษา ในกลุ่มทักษะวิชาภาษาไทย จำแนกเป็นรายอำเภอ

แหล่งความแปรปรวน	SS	df	MS	F
อำเภอเมือง				
ระหว่างกลุ่ม	1034.207	2	517.104	22.931**
ภายในกลุ่ม	10124.811	449	22.550	
รวม	11159.018	451		
อำเภอลองหลวง				
ระหว่างกลุ่ม	642.507	2	321.254	12.543**
ภายในกลุ่ม	10142.280	396	25.612	
รวม	10784.787	398		
อำเภอลาดหลุมแก้ว				
ระหว่างกลุ่ม	328.579	2	164.289	7.024**
ภายในกลุ่ม	6689.844	286	23.391	
รวม	7018.443	288		

ตาราง 23 (ต่อ)

แหล่งความแปรปรวน	SS	df	MS	F
อำเภอหนองเสือ				
ระหว่างกลุ่ม	242.805	2	121.403	4.321**
ภายในกลุ่ม	8709.265	310	28.094	
รวม	8952.070	312		

* มีนัยสำคัญทางสถิติที่ระดับ .01

ตาราง 24 ผลการวิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบความแตกต่างของคะแนนเฉลี่ย
ระหว่างสามปีการศึกษา ในกลุ่มทักษะวิชาคณิตศาสตร์ จำแนกเป็นรายอำเภอ

แหล่งความแปรปรวน	SS	df	MS	F
อำเภอเมือง				
ระหว่างกลุ่ม	6689.900	2	3344.950	95.995**
ภายในกลุ่ม	15645.560	449	34.845	
รวม	22335.460	451		
อำเภอลองหลวง				
ระหว่างกลุ่ม	4919.740	2	2459.870	68.284**
ภายในกลุ่ม	14265.553	396	36.024	
รวม	19185.293	398		
อำเภอลาดหลุมแก้ว				
ระหว่างกลุ่ม	1034.164	2	517.082	12.696**
ภายในกลุ่ม	11647.898	286	40.727	
รวม	12682.062	288		

ตาราง 24 (ต่อ)

แหล่งความแปรปรวน	SS	df	MS	F
อำเภอนองเสือ				
ระหว่างกลุ่ม	1659.288	2	829.644	23.662**
ภายในกลุ่ม	11869.332	310	35.062	
รวม	12528.620	312		

* มีนัยสำคัญทางสถิติที่ระดับ .01

จากตารางพบว่าคะแนนเฉลี่ยระหว่าง 3 ปีการศึกษามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติระดับ .01 ทุกอำเภอนองเสือทั้งสองกลุ่มวิชา หลังจากนั้นผู้วิจัยจึงทำการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ต่อไป โดยใช้วิธีการของ sheffe ซึ่งปรากฏผลดังตาราง 25 และ 26

ตาราง 25 ผลการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ ในกลุ่มทักษะ
วิชาภาษาไทย

ปีการศึกษา	ปีการศึกษา 2529	ปีการศึกษา 2530	ปีการศึกษา 2531
ปีการศึกษา 2529	-	6.500**	8.006**
ปีการศึกษา 2530	-	-	1.506**
ปีการศึกษา 2531	-	-	-

* มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 25 พบว่าความแตกต่างของคะแนนเฉลี่ยรายคู่ทั้งสามคู่คือ ระหว่างค่าเฉลี่ย
ของปีการศึกษา 2529 กับ 2530 ระหว่างค่าเฉลี่ยของปีการศึกษา 2529 กับ 2531 และระหว่าง
ค่าเฉลี่ยของปีการศึกษา 2530 กับ 2531 นั้นมีนัยสำคัญทางสถิติที่ระดับ .01

ตาราง 26 ผลการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ ในกลุ่มทักษะ
วิชาคณิตศาสตร์

ปีการศึกษา	ปีการศึกษา 2529	ปีการศึกษา 2530	ปีการศึกษา 2531
ปีการศึกษา 2529	-	16.852**	2.381**
ปีการศึกษา 2530	-	-	- 14.471**
ปีการศึกษา 2531	-	-	-

* มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 26 พบว่าค่าเฉลี่ยรายคู่ในแต่ละปีการศึกษาทั้งสามคู่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ในทำนองเดียวกันผู้วิจัยทำการวิเคราะห์ความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ของแต่ละกลุ่มวิชา จำแนกเป็นรายอำเภอ ซึ่งปรากฏผลดังตาราง 27

ตาราง 27 ผลการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ของแต่ละกลุ่มวิชาจำแนกเป็นรายอำเภอ

อำเภอ	ปีการศึกษา	ภาษาไทย			คณิตศาสตร์		
		ปี 2529	ปี 2530	ปี 2531	ปี 2529	ปี 2530	ปี 2531
เมือง	ปี 2529	-	10.779 ^{**}	9.779 ^{**}	-	21.813 ^{**}	-2.383 ^{**}
	ปี 2530	-	-	-1.000 [*]	-	-	-24.196 ^{**}
	ปี 2531	-	-	-	-	-	-
คลองหลวง	ปี 2529	-	10.121 ^{**}	6.284 ^{**}	-	18.584 ^{**}	-1.656 [*]
	ปี 2530	-	-	-3.837 ^{**}	-	-	-20.240 ^{**}
	ปี 2531	-	-	-	-	-	-
ลาดหลุมแก้ว	ปี 2529	-	-0.128 [*]	7.361 ^{**}	-	12.425 ^{**}	6.262 ^{**}
	ปี 2530	-	-	7.489 ^{**}	-	-	-6.163 ^{**}
	ปี 2531	-	-	-	-	-	-
หนองเสือ	ปี 2529	-	3.663 [*]	7.310 ^{**}	-	13.944 ^{**}	9.875 ^{**}
	ปี 2530	-	-	3.647 ^{**}	-	-	-4.069 ^{**}
	ปี 2531	-	-	-	-	-	-

* มีนัยสำคัญทางสถิติที่ระดับ .01

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง พบว่าคะแนนเฉลี่ยร้อยละรายคู่ของการเปรียบเทียบมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ดังนั้นการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนจึงสรุปผลได้อย่างมั่นใจมากยิ่งขึ้น

หลังจากนั้นผู้วิจัยทำการทดสอบความแตกต่างของสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจระหว่างสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์เมื่อเทียบคะแนนแล้ว กับเมื่อยังไม่ได้ทำการเทียบคะแนน โดยใช้สถิติทดสอบ Z-Test for Uncorrelated Data ซึ่งปรากฏผลการเปรียบเทียบสัดส่วนจำนวนนักเรียนเมื่อเทียบคะแนนระหว่างในปีการศึกษา 2530 และ 2531 ที่เทียบคะแนนไปสู่สเกลคะแนนสเกลเดียวกับสเกลคะแนนในปีการศึกษา 2529 กับเมื่อยังไม่ได้เทียบคะแนนดังตาราง 28

ตาราง 28 ผลการเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ ระหว่างเมื่อเทียบคะแนนแล้ว กับเมื่อยังไม่ได้เทียบคะแนน จำแนกเป็นรายกลุ่มวิชา และรายอำเภอ

อำเภอ / กลุ่มวิชา	ภาษาไทย		คณิตศาสตร์	
	ปี 2530	ปี 2531	ปี 2530	ปี 2531
เมือง	1.175	1.326	3.460**	1.119
คลองหลวง	2.035*	1.931*	4.000**	1.171
ลาดหลุมแก้ว	1.227	0.983	2.521**	1.000
หนองเสือ	1.464	0.616	2.537**	1.015
รวม	3.030*	-2.448**	6.156**	-2.074*

* มีนัยสำคัญทางสถิติที่ระดับ .01

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 28 พบว่าสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ ระหว่างเมื่อเทียบคะแนนแล้ว กับเมื่อยังไม่ได้เทียบคะแนน มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทั้งสองกลุ่มวิชา ยกเว้นในกลุ่มทักษะวิชาคณิตศาสตร์เมื่อปีการศึกษา 2531 ที่พบว่าความแตกต่างมีนัยสำคัญทางสถิติที่ระดับ .05

ส่วนผลการเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ ระหว่างเมื่อเทียบคะแนนแล้ว กับเมื่อยังไม่ได้เทียบคะแนน จำแนกเป็นรายอำเภอ พบว่า อำเภอคลองหลวงมีสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจแตกต่างกันอย่างมีนัยสำคัญทางสถิติทั้งสองกลุ่มวิชา ยกเว้นในกลุ่มทักษะวิชาคณิตศาสตร์เมื่อปีการศึกษา 2531 ซึ่งพบว่าความแตกต่างไม่มีนัยสำคัญทางสถิติ นอกจากนี้ ได้พบว่าสัดส่วนจำนวนนักเรียนที่มีผลน่าพอใจในกลุ่มทักษะวิชาคณิตศาสตร์ เมื่อปีการศึกษา 2530 มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกอำเภอนอกนั้นพบว่าความแตกต่างไม่มีนัยสำคัญ

หลังจากนั้นผู้วิจัยแจกแจงสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจทั้งเมื่อได้ทำการเทียบคะแนนแล้วและเมื่อยังไม่ได้เทียบคะแนนของทั้งสองกลุ่มวิชา เพื่อศึกษาความคลาดเคลื่อนในการตัดสินใจผลให้ผ่านเกณฑ์ที่น่าพอใจและไม่ผ่านเกณฑ์ที่น่าพอใจ ซึ่งปรากฏผลดังตาราง 29

ตาราง 29 สัดส่วนจำนวนนักเรียนที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจในแต่ละปีการศึกษา เมื่อเทียบคะแนนแล้ว และเมื่อไม่เทียบคะแนน

ปีการศึกษา กลุ่มวิชา	ปีการศึกษา 2529		ปีการศึกษา 2530		ปีการศึกษา 2531	
	เทียบ	ไม่เทียบ	เทียบ	ไม่เทียบ	เทียบ	ไม่เทียบ
ภาษาไทย	60	60	64	54	69	76
คณิตศาสตร์	14	14	50	30	20	25

จากตาราง 29 ในกลุ่มทักษะวิชาภาษาไทย เมื่อเทียบคะแนนผลการสอบในปีการศึกษา 2530 และ 2531 ไปสู่สเกลเดียวกับสเกลคะแนนในปีการศึกษา 2529 แล้วพบว่า ในปีการศึกษา 2530 มีสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจเท่ากับร้อยละ 64 ซึ่งมีความมากกว่าสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจเมื่อไม่ได้เทียบคะแนนเท่ากับร้อยละ 10 แสดงว่าเมื่อไม่เทียบคะแนนจะมีนักเรียนจำนวนร้อยละ 10 ที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจ แต่ได้รับการตัดสินผลให้ไม่ผ่านเกณฑ์ที่น่าพอใจ ส่วนปีการศึกษา 2531 พบว่ามีสัดส่วนการผ่านเกณฑ์ที่น่าพอใจ ต่ำกว่าเมื่อไม่เทียบคะแนนเท่ากับร้อยละ 7 แสดงว่าเมื่อไม่เทียบคะแนนจะมีนักเรียนร้อยละ 7 ที่มีผลการสอบไม่ผ่านเกณฑ์ที่น่าพอใจแต่ได้รับการตัดสินผลให้ผ่านเกณฑ์ที่น่าพอใจ

ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์ เมื่อเทียบคะแนนแล้วพบว่าในปีการศึกษา 2530 มีสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจสูงกว่าเมื่อไม่เทียบคะแนนเท่ากับร้อยละ 20 แสดงว่าเมื่อไม่เทียบคะแนนจะมีนักเรียนจำนวนร้อยละ 20 ที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจ แต่ได้รับการตัดสินให้ไม่ผ่านเกณฑ์ที่น่าพอใจ ส่วนในปีการศึกษา 2531 มีสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ ต่ำกว่าเมื่อไม่เทียบเท่ากับร้อยละ 5 แสดงว่าเมื่อไม่เทียบคะแนนจะมีนักเรียนร้อยละ 5 ที่มีคะแนนผลการสอบไม่ผ่านเกณฑ์ที่น่าพอใจ แต่ได้รับการตัดสินผลให้ผ่านเกณฑ์ที่น่าพอใจ

จากตาราง 30 พบว่าค่าเฉลี่ย ค่าเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลการสอบผ่าน เกณฑ์ที่น่าพอใจในแต่ละกลุ่มวิชา และในแต่ละปีการศึกษา ระหว่างเมื่อเทียบคะแนน (คะแนนจริง) และเมื่อไม่ได้เทียบคะแนน (คะแนนจริง) มีค่าแตกต่างกัน โดยค่าคะแนนเฉลี่ยร้อยละ ในปีการศึกษา 2530 เมื่อเทียบคะแนนแล้วมีค่าแตกต่างจากเมื่อไม่ได้เทียบคะแนนมากกว่าในปีการศึกษา 2531 ทั้งสองกลุ่มวิชา ส่วนอัตราความก้าวหน้าของค่าคะแนนเฉลี่ยร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจ พบว่ามีการเพิ่มขึ้นหรือลดลงในตัวเองเดียวกัน แต่แตกต่างกันที่ปริมาณมากน้อยของการเพิ่มขึ้นหรือลดลง ยกเว้นในปีการศึกษา 2530 ที่พบว่ากลุ่มทักษะภาษาไทยเมื่อเทียบคะแนนแล้วมีค่าคะแนนเฉลี่ย ร้อยละ และร้อยละนักเรียนที่มีผลน่าพอใจในอัตราที่เพิ่มขึ้น แต่เมื่อไม่ได้เทียบคะแนนจะมีอัตราที่ลดลง ทั้งค่าคะแนนเฉลี่ยร้อยละและร้อยละนักเรียนที่มีผลน่าพอใจ อย่างไรก็ตาม เมื่อตั้งเกณฑ์การผ่านที่ ร้อยละ 50 ของคะแนนเต็มในแต่ละกลุ่มวิชา พบว่า การตัดสินผลการผ่านเกณฑ์ที่น่าพอใจของค่าเฉลี่ย ร้อยละ และสัดส่วนจำนวนนักเรียนที่มีผลผ่านเกณฑ์ที่น่าพอใจ ระหว่างเมื่อเทียบคะแนนแล้วกับเมื่อ ไม่ได้เทียบคะแนนให้ผลแตกต่างกัน

สรุปผล อภิปรายผลและข้อเสนอแนะ

จุดมุ่งหมายของการวิจัย

1. เพื่อศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบคะแนนรูปแบบราสช์
2. เพื่อเปรียบเทียบความแตกต่างของผลสัมฤทธิ์ทางการเรียน ระหว่างปีการศึกษาต่าง ๆ หลังจากที่ได้เทียบคะแนนแล้ว
3. เพื่อเปรียบเทียบสัดส่วนจำนวนนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนผ่านเกณฑ์ที่น่าพอใจ ระหว่างการเทียบคะแนนกับการไม่เทียบคะแนน

สมมติฐานของการวิจัย

1. ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรในแต่ละปีการศึกษา มีความก้าวหน้าขึ้นตามลำดับ
2. สัดส่วนจำนวนนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนผ่านเกณฑ์ที่น่าพอใจ ระหว่างการเทียบคะแนนกับการไม่เทียบคะแนนแตกต่างกัน

วิธีดำเนินการวิจัย

ประชากรที่ใช้ในการศึกษารั้งนี้ คือนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 เมื่อปีการศึกษา 2529, 2530 และ 2531 ในสังกัดสำนักงาน การประถมศึกษาจังหวัดปทุมธานี ผู้วิจัยศึกษา 4 อำเภอ ซึ่งแต่ละปีมีจำนวนนักเรียน 3,875 คน 4,036 คน และ 3,973 คน ตามลำดับ

กลุ่มตัวอย่างที่ใช้ในการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในครั้งนี้ คือ กลุ่มตัวอย่างที่สุ่มไว้แล้วตามโครงการประเมินคุณภาพนักเรียนชั้นประถมปีที่ 6 ระดับประเทศเมื่อ ปีการศึกษา 2529, 2530 และ 2531 โดยผู้วิจัยจะนำมาศึกษาเฉพาะอำเภอที่มีข้อมูลผลการสอบ รายชื่อของนักเรียนรายบุคคล ครบทั้ง 3 ปีการศึกษา จำนวน 4 อำเภอ (จากอำเภอทั้งหมด 7 อำเภอ) คือ อำเภอเมือง อำเภอคลองหลวง อำเภอลาดหลุมแก้วและอำเภอหนองเสือ รวม จำนวนนักเรียนกลุ่มตัวอย่างในแต่ละปีการศึกษาเท่ากับ 534 คน, 453 คน และ 467 คน ตามลำดับ

สำหรับกลุ่มตัวอย่างที่ใช้ในการหาค่าคงที่ในการแปลงค่าประมาณพารามิเตอร์ของแบบสอบ แต่ละชุดให้อยู่ในสเกลเดียวกัน ผู้วิจัยสุ่มจากนักเรียนชั้นประถมศึกษาปีที่ 6 ที่กำลังเรียนภาคปลาย ปีการศึกษา 2532 ในอำเภอธัญบุรี สังกัดสำนักงานการประถมศึกษาจังหวัดปทุมธานี จำนวน 4 โรงเรียน รวมมีนักเรียนทั้งสิ้น 581 คน

เครื่องมือที่ใช้ในการวิจัย คือแบบสอบวัดผลสัมฤทธิ์ที่สร้างขึ้นตามโครงการประเมินคุณภาพ นักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศของสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ซึ่งใช้สอบนักเรียนกลุ่มตัวอย่างเมื่อปีการศึกษา 2529, 2530 และ 2531 ผู้วิจัยศึกษาในกลุ่มทักษะ (วิชาคณิตศาสตร์และวิชาภาษาไทย) เฉพาะภาคความรู้ กลุ่มวิชาละ 3 ชุด โดยคุณภาพของแบบสอบ เน้นความเที่ยงตรงเชิงเนื้อหา กล่าวคือแบบสอบสามารถวัดได้ตรงตามสมรรถภาพที่ต้องการวัดใน แต่ละกลุ่มวิชา

การเก็บรวบรวมข้อมูล

ผู้วิจัยขออนุญาตใช้ข้อมูลกระดาษคำตอบในกลุ่มทักษะของนักเรียนกลุ่มตัวอย่าง จากสำนักงาน การประถมศึกษาจังหวัดปทุมธานี ที่ดำเนินการสอบไปแล้วเมื่อปีการศึกษา 2529 ถึง 2531 โดย สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติและมีข้อมูลนักเรียนรายบุคคล เก็บไว้ที่สำนักงาน การประถมศึกษาจังหวัดปทุมธานี ผู้วิจัยได้นำมาจำนวน 4 อำเภอ เฉพาะอำเภอที่มีข้อมูลสมบูรณ์ ทั้ง 3 ปีการศึกษา และได้้นำผลการตอบข้อสอบรายข้อรายบุคคลมาบันทึกลงเทป โดยแยกตาม ปีการศึกษาและกลุ่มวิชา ชนิดตอบถูกต้องให้ 1 คะแนน ตอบไม่ถูกต้องให้ 0 คะแนน และไม่ตอบให้ 0 คะแนน

ส่วนการเก็บรวบรวมข้อมูลในกลุ่มตัวอย่างที่ใช้หาค่าคงที่ในการแปลงค่าประมาณพารามิเตอร์ ให้อยู่ในสเกลเดียวกันนั้น ผู้วิจัยแบ่งกลุ่มตัวอย่างที่สุ่มมาได้เป็น 2 กลุ่ม ๆ ละ 8 ห้องเรียน ทำการ สอบนักเรียนเป็นรายห้อง ๆ ละ 2 วัน โดยแต่ละห้องจะต้องสอบแบบสอบ 2 ชุด ในแต่ละกลุ่มวิชา คือสอบชุดที่ใช้ในปีการศึกษา 2529 (ชุด X) 1 ชุด และสอบชุดที่ใช้ในปีการศึกษา 2530 (ชุด Y) หรือชุดที่ใช้ในปีการศึกษา 2531 (ชุด Z) อีก 1 ชุด ซึ่งทำการสอบวันละ 1 ชุด ชุดละ 2 ชั่วโมง การจัดลำดับก่อนหลังของแบบสอบใช้วิธีสุ่มและจะใช้ข้อมูลผลการสอบจากแบบสอบชุด X ซึ่งกลุ่ม ตัวอย่างทุกกลุ่มสอบเหมือนกันนั้น เป็นข้อมูลจากกลุ่มผู้สอบร่วมสำหรับใช้เชื่อมข้อมูลผลการสอบระหว่าง แบบสอบชุด Y และชุด Z ให้อยู่ในสเกลเดียวกับแบบสอบชุด X และสามารถหาค่าคงที่ในการแปลง ค่าประมาณพารามิเตอร์ที่ได้จากการสอบแบบสอบแต่ละชุดจากผลการสอบในสถานการณ์จริงเมื่อปี การศึกษา 2529 ถึง 2531 ให้อยู่ในสเกลเดียวกันทั้งหมดและสามารถเทียบคะแนนกันได้

การวิเคราะห์ข้อมูล

ผู้วิจัยวิเคราะห์ข้อมูลในแต่ละกลุ่มวิชา โดยมีขั้นตอนเหมือนกันดังนี้

1. หาค่าสถิติพื้นฐานและแจกแจงความถี่ของคะแนนที่ได้จากการสอบ แบบสอบแต่ละชุด โดยใช้โปรแกรมคอมพิวเตอร์ SPSS^X ด้วยคำสั่ง Frequency
2. ประมวลค่าพารามิเตอร์ความสามารถของผู้สอบและความยากของแบบสอบแต่ละชุด ที่ได้จากการสอบในกลุ่มตัวอย่างที่ใช้ในการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนและกลุ่มตัวอย่างที่ใช้ศึกษาค่าคงที่ในการเทียบคะแนน โดยการใช้โปรแกรมคอมพิวเตอร์ BICAL ที่ศูนย์สารสนเทศ สำนักงานปลัดกระทรวงศึกษาธิการ
3. คำนวณค่าคงที่ (α และ β) จากกลุ่มผู้สอบร่วมที่ใช้หาค่าคงที่ เพื่อใช้ในสมการเชิงเส้นตรงสำหรับการแปลงค่าประมาณพารามิเตอร์จากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกัน
4. แปลงค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบชุด Y (ปีการศึกษา 2530) ไปยังชุด X (ปีการศึกษา 2529) และแปลงจากชุด Z (ปีการศึกษา 2531) ไปยังชุด X ดังนั้น ค่าประมาณพารามิเตอร์ความสามารถของผู้สอบแบบสอบแต่ละชุดจะอยู่ในสเกลเดียวกันทั้งหมด
5. แปลงค่าประมาณพารามิเตอร์ความยากของข้อสอบแบบสอบชุด Y และชุด Z ไปยังชุด X โดยใช้วิธีการในทำนองเดียวกันกับข้อ 4
6. คำนวณค่าความน่าจะเป็นที่ผู้สอบแต่ละระดับความสามารถจะตอบข้อสอบแต่ละข้อในแบบสอบแต่ละชุดได้ถูกต้อง
7. ประมวลค่าคะแนนจริงของผู้สอบแบบสอบแต่ละชุด ที่แต่ละระดับความสามารถ θ
8. แจกแจงความถี่ของคะแนนจริงในแบบสอบแต่ละกลุ่ม
9. เทียบคะแนนจริงของผู้สอบแบบสอบต่างชุดที่ตรงกับคะแนนดิบค่าเดียวกัน
10. ศึกษาคะแนนจุดตัดในแต่ละปีการศึกษาซึ่งตรงกับคะแนนดิบที่ร้อยละ 50 ของคะแนนเต็มในแต่ละวิชา

11. คำนวณหาค่าเฉลี่ยร้อยละและร้อยละนักเรียนที่มีผลน่าพอใจจากการแจกแจงคะแนนจริงในแบบสอบแต่ละชุด

12. ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในช่วง 3 ปีการศึกษา ของนักเรียนแต่ละอำเภอและรวมทุกอำเภอ โดยพิจารณาจากปริมาณที่เพิ่มขึ้นหรือลดลงของค่าเฉลี่ยร้อยละ ($X\%$) และร้อยละนักเรียนที่มีผลน่าพอใจแต่ละปีการศึกษา โดยใช้คะแนนจุดตัดการผ่านเกณฑ์ขั้นต่ำที่น่าพอใจที่ค่าเฉลี่ยร้อยละ 50 ของคะแนนเต็มเหมือนกันทุกปีการศึกษา

13. เปรียบเทียบความแตกต่างของผลสัมฤทธิ์ทางการเรียนระหว่างปีการศึกษาต่าง ๆ เมื่อได้ทำการเทียบคะแนนแล้ว โดยทำการวิเคราะห์ความแปรปรวน (ANOVA)

14. เปรียบเทียบความแตกต่างของสัดส่วนจำนวนนักเรียนที่ผ่านคะแนนจุดตัดของการผ่านเกณฑ์ที่น่าพอใจเมื่อเทียบคะแนนแล้ว กับสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์เมื่อยังไม่ได้เทียบคะแนน โดยใช้สถิติทดสอบซี (Z-Test) ทดสอบในกรณีที่สองเป็นอิสระจากกัน หลังจากนั้นแจกแจงค่าร้อยละนักเรียนที่มีผลน่าพอใจระหว่างผลที่ได้จากการเทียบคะแนน กับผลที่ได้จากการไม่เทียบคะแนน เพื่อพิจารณาความคลาดเคลื่อน 2 ชนิด คือ

14.1 ร้อยละนักเรียนที่ไม่ผ่านคะแนนจุดตัดเมื่อยังไม่ได้เทียบคะแนนแต่ผ่านจุดตัดเมื่อได้เทียบคะแนนแล้ว

14.2 ร้อยละนักเรียนที่ผ่านคะแนนจุดตัดเมื่อยังไม่ได้เทียบคะแนนแต่ไม่ผ่านคะแนนจุดตัดเมื่อได้เทียบคะแนนแล้ว

สรุปผลการวิจัย

1. ผลการเทียบคะแนนระหว่างแบบสอบสามชุด ปรากฏผลดังนี้

1.1 ในกลุ่มทักษะวิชาภาษาไทย คะแนนที่เทียบแล้วของแบบสอบสามชุดซึ่งตรงกับคะแนนดิบค่าเดียวกันมีค่าแตกต่างกัน กล่าวคือโดยทั่วไปคะแนนที่เทียบแล้วของปีการศึกษา 2530 จะมีค่าสูงสุด รองลงมาคือคะแนนเทียบของปีการศึกษา 2531 และคะแนนเทียบของปีการศึกษา 2529

มีค่าต่ำกว่าคะแนนเทียบของปีการศึกษาอื่น ความแตกต่างของคะแนนเทียบระหว่างสามปีการศึกษา มีค่าสูงสุดในช่วงตรงกลางของการแจกแจงคะแนนเทียบ คือช่วงที่ตรงกับคะแนนดิบตั้งแต่ 10 ถึง 23 คะแนน ซึ่งพบว่ามีค่าความแตกต่างของคะแนนเทียบตั้งแต่ 2.5 ถึง 3.8 คะแนน

1.2 ในกลุ่มทักษะวิชาคณิตศาสตร์ คะแนนที่เทียบแล้วของแบบสอบสามชุดที่ตรงกับคะแนนดิบค่าเดียวกันมีค่าแตกต่างกัน กล่าวคือโดยทั่วไปคะแนนที่เทียบแล้วของปีการศึกษา 2530 จะมีค่าสูงสุด รองลงมาคือคะแนนเทียบของปีการศึกษา 2529 และคะแนนเทียบของปีการศึกษา 2531 มีค่าต่ำกว่าคะแนนเทียบของปีการศึกษาอื่น ความแตกต่างของคะแนนเทียบระหว่างสามปีการศึกษามีค่าสูงสุดในช่วงตรงกลางของการแจกแจงคะแนน คือในช่วงที่ตรงกับคะแนนดิบตั้งแต่ 10 ถึง 32 คะแนน ซึ่งพบว่ามีค่าความแตกต่างของคะแนนเทียบตั้งแต่ 2.6 ถึง 6.3 คะแนน

2. ผลการศึกษาผลสัมฤทธิ์ทางการเรียนและความก้าวหน้าของผลสัมฤทธิ์ทางการเรียน ตั้งแต่ปีการศึกษา 2529 ถึงปีการศึกษา 2531 ปรากฏดังนี้

2.1 ผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะวิชาภาษาไทยในแต่ละปีการศึกษามีค่าคะแนนเฉลี่ยร้อยละตั้งแต่ร้อยละ 50.087 ถึง 58.093 คะแนนเฉลี่ยร้อยละต่ำสุดในปีการศึกษา 2529 คะแนนเฉลี่ยร้อยละสูงสุดในปีการศึกษา 2531 โดยอัตราความก้าวหน้าของคะแนนเฉลี่ยร้อยละเพิ่มขึ้นในทุกปีการศึกษาตามลำดับดังนี้ ในปีการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 6.500 และในปีการศึกษา 2531 มีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากปีการศึกษา 2530 อีกร้อยละ 1.506 รวมมีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นจากในปีการศึกษา 2529 เท่ากับร้อยละ 8.006

ส่วนค่าเฉลี่ยร้อยละนักเรียนที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจในแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 60 ถึง 69 ร้อยละต่ำสุดในปีการศึกษา 2529 ร้อยละสูงสุดในปีการศึกษา 2531 โดยมีอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจเพิ่มขึ้นทุกปีการศึกษาตามลำดับดังนี้ ในปีการศึกษา 2530 มีร้อยละเพิ่มขึ้นจากปีการศึกษา 2529 เท่ากับร้อยละ 4 ในปีการศึกษา 2531 มีร้อยละเพิ่มขึ้นจากปีการศึกษา 2530 อีกร้อยละ 5 รวมมีจำนวนร้อยละนักเรียนที่มีผลน่าพอใจเพิ่มขึ้นจากปีการศึกษา 2529 เท่ากับร้อยละ 9

เมื่อพิจารณาผลสัมฤทธิ์ทางการเรียนในช่วง 3 ปีการศึกษา จำแนกเป็นรายอำเภอพบว่า มีคะแนนเฉลี่ยร้อยละตั้งแต่ร้อยละ 44.000 ถึง 62.689 โดยในปีการศึกษา 2531 ทุกอำเภอมีคะแนนเฉลี่ยร้อยละสูงกว่าร้อยละ 50 อำเภอเมืองและอำเภอลาดหลุมแก้วมีคะแนนเฉลี่ยร้อยละสูงกว่าร้อยละ 50 ทุกปีการศึกษา อำเภอกลองหลวงมีคะแนนเฉลี่ยร้อยละต่ำกว่าร้อยละ 50 ในปีการศึกษา 2529 ส่วนอำเภอหนองเสือมีคะแนนเฉลี่ยร้อยละต่ำกว่าร้อยละ 50 ในปีการศึกษา 2529 และปีการศึกษา 2530 สำหรับอัตราความก้าวหน้าของผลสัมฤทธิ์ทางการเรียน พบว่า อำเภอหนองเสือมีอัตราความก้าวหน้าขึ้นตามลำดับในทุกปีการศึกษา โดยในปีการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละสูงขึ้นจากปีการศึกษา 2529 เท่ากับร้อยละ 3.663 และในปีการศึกษา 2531 มีคะแนนเฉลี่ยร้อยละสูงขึ้นจากปีการศึกษา 2530 อีกร้อยละ 3.647 อำเภอเมืองและอำเภอกลองหลวงมีคะแนนเฉลี่ยร้อยละสูงขึ้นในปีการศึกษา 2530 คิดเป็นร้อยละ 10.779 และ 10.121 ตามลำดับ แต่มีค่าร้อยละลดลงในปีการศึกษา 2531 คิดเป็นร้อยละ 1.000 และ 3.837 ตามลำดับ ส่วนอำเภอลาดหลุมแก้วมีคะแนนเฉลี่ยร้อยละลดลงเล็กน้อยในปีการศึกษา 2530 คือลดลงคิดเป็นร้อยละ 0.128 แต่ในปีการศึกษา 2531 มีคะแนนเฉลี่ยร้อยละสูงขึ้นจากปีการศึกษา 2530 คิดเป็นร้อยละ 7.489

ส่วนค่าเฉลี่ยร้อยละนักเรียนที่มีผลน่าพอใจแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 45 ถึง 78 โดยทุกอำเภอมีร้อยละนักเรียนที่มีผลน่าพอใจสูงกว่าร้อยละ 50 ยกเว้นอำเภอหนองเสือที่มีร้อยละนักเรียนที่มีผลน่าพอใจต่ำกว่าร้อยละ 50 ในปีการศึกษา 2529 และปีการศึกษา 2530 ซึ่งคิดเป็นร้อยละ 45 และร้อยละ 49 ตามลำดับ สำหรับอัตราความก้าวหน้าของร้อยละนักเรียนที่มีผลน่าพอใจของแต่ละอำเภอ พบว่า อำเภอเมืองและอำเภอหนองเสือมีร้อยละนักเรียนที่มีผลน่าพอใจสูงขึ้นกว่าเดิมทุกปีการศึกษา โดยในปีการศึกษา 2530 พบว่าแต่ละอำเภอมีร้อยละนักเรียนที่มีผลน่าพอใจสูงขึ้นคิดเป็นร้อยละ 11 และร้อยละ 4 ตามลำดับ ส่วนในปีการศึกษา 2531 แต่ละอำเภอมีร้อยละนักเรียนที่มีผลน่าพอใจสูงขึ้น คิดเป็นร้อยละ 1 และร้อยละ 9 ตามลำดับ อำเภอกลองหลวงมีร้อยละนักเรียนที่มีผลน่าพอใจสูงขึ้นในปีการศึกษา 2530 คิดเป็นร้อยละ 10 แต่ในปีการศึกษา 2531

มีร้อยละนักเรียนที่มีผลน่าพอใจลดลงจากปีการศึกษา 2530 คิดเป็นร้อยละ 7 อำเภอลาดหลุมแก้วมีร้อยละนักเรียนที่มีผลน่าพอใจลดลงในปีการศึกษา 2530 คิดเป็นร้อยละ 14 แต่ในปีการศึกษา 2531 มีร้อยละนักเรียนที่มีผลน่าพอใจสูงขึ้นจากในปีการศึกษา 2530 คิดเป็นร้อยละ 20

2.2 ผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะวิชาคณิตศาสตร์ในแต่ละปีการศึกษาพบว่า มีค่าคะแนนเฉลี่ยร้อยละตั้งแต่ร้อยละ 31.795 ถึงร้อยละ 34.176 คะแนนเฉลี่ยร้อยละต่ำสุดในปีการศึกษา 2529 คะแนนเฉลี่ยร้อยละสูงสุดในปีการศึกษา 2530 โดยในปีการศึกษา 2530 มีคะแนนเฉลี่ยร้อยละสูงขึ้นจากในปีการศึกษา 2529 คิดเป็นร้อยละ 16.852 แต่ในปีการศึกษา 2531 มีคะแนนเฉลี่ยร้อยละลดลงจากในปีการศึกษา 2530 คิดเป็นร้อยละ 14.471 ซึ่งยังมีค่าสูงขึ้นจากปีการศึกษา 2529 เท่ากับร้อยละ 2.381

ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 14 ถึง 50 ร้อยละต่ำสุดในปีการศึกษา 2529 ร้อยละสูงสุดในปีการศึกษา 2530 โดยในปีการศึกษา 2530 มีค่าร้อยละนักเรียนที่มีผลน่าพอใจสูงขึ้นในปีการศึกษา 2529 คิดเป็นร้อยละ 36 แต่ในปีการศึกษา 2531 มีค่าร้อยละนักเรียนที่มีผลน่าพอใจลดลงจากปีการศึกษา 2530 คิดเป็นร้อยละ 30 ซึ่งยังมีค่าสูงกว่าร้อยละนักเรียนที่มีผลน่าพอใจในปีการศึกษา 2529 คิดเป็นร้อยละ 6

เมื่อพิจารณาผลสัมฤทธิ์ทางการเรียนในช่วง 3 ปีการศึกษา จำแนกเป็นรายอำเภอพบว่า มีคะแนนเฉลี่ยร้อยละตั้งแต่ร้อยละ 24.395 ถึงร้อยละ 55.242 โดยมีเพียงอำเภอเมืองและอำเภอคลองหลวงที่มีคะแนนเฉลี่ยร้อยละสูงกว่าร้อยละ 50 ในปีการศึกษา 2530 เพียงปีการศึกษาเดียว ส่วนอำเภอลาดหลุมแก้วและอำเภอหนองเสือมีคะแนนเฉลี่ยร้อยละต่ำกว่าร้อยละ 50 ในทุกปีการศึกษา สำหรับอัตราความก้าวหน้าของผลสัมฤทธิ์ทางการเรียน พบว่า ทุกอำเภอมียกคะแนนเฉลี่ยร้อยละสูงขึ้นในปีการศึกษา 2530 ซึ่งมีค่าที่สูงขึ้นจากปีการศึกษา 2529 ตั้งแต่ร้อยละ 12.425 ถึง 21.813 อำเภอเมืองมีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นสูงสุด ส่วนอำเภอลาดหลุมแก้วมีคะแนนเฉลี่ยร้อยละเพิ่มขึ้นน้อยกว่าอำเภออื่น และในปีการศึกษา 2531 ทุกอำเภอมียกคะแนนเฉลี่ยร้อยละลดลงจากปีการศึกษา 2530 โดยมีค่าที่ลดลงตั้งแต่ ร้อยละ 4.069 ถึง ร้อยละ 24.196 อำเภอเมืองมีคะแนนเฉลี่ยร้อยละลดลงสูงสุด ส่วนอำเภอหนองเสือมีคะแนนเฉลี่ยร้อยละลดลงน้อยกว่าอำเภออื่น

ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษา พบว่ามีค่าตั้งแต่ร้อยละ 6 ถึง ร้อยละ 65 อำเภอเมืองมีร้อยละนักเรียนที่มีผลน่าพอใจสูงสุดในปีการศึกษา 2530 และอำเภอหนองเสือมีร้อยละนักเรียนที่มีผลน่าพอใจต่ำสุดในปีการศึกษา 2529 นอกจากนี้ พบว่ามีอำเภอเมือง และอำเภอคลองหลวงเท่านั้น ที่มีร้อยละนักเรียนที่มีผลน่าพอใจมากกว่าร้อยละ 50 ในปีการศึกษา 2530 นอกนั้นทุกอำเภอมีร้อยละนักเรียนที่มีผลน่าพอใจต่ำกว่าร้อยละ 50 สำหรับอัตราความก้าวหน้าของร้อยละนักเรียนที่มีผลน่าพอใจของแต่ละอำเภอ พบว่า ทุกอำเภอมีร้อยละนักเรียนที่มีผลน่าพอใจสูงขึ้นในปีการศึกษา 2530 โดยสูงขึ้นคิดเป็นร้อยละตั้งแต่ร้อยละ 22 ถึงร้อยละ 50 อำเภอเมืองมีความก้าวหน้าสูงสุดและอำเภอหนองเสือมีความก้าวหน้าน้อยกว่าอำเภออื่น ส่วนในปีการศึกษา 2531 ทุกอำเภอมีค่าร้อยละนักเรียนที่มีผลน่าพอใจลดลงจากปีการศึกษา 2530 คิดเป็นร้อยละตั้งแต่ร้อยละ 4 ถึงร้อยละ 51 โดยอำเภอเมืองมีร้อยละนักเรียนที่มีผลน่าพอใจลดลงต่ำสุด

3. ผลการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละระหว่าง 3 ปีการศึกษา พบว่าความแตกต่างมีนัยสำคัญทางสถิติที่ระดับ .01 เหมือนกันทั้งสองกลุ่มวิชา และหลังจากนั้น ผู้วิจัยทำการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ต่อไป โดยใช้วิธีการของ Sheffe พบว่าคะแนนเฉลี่ยร้อยละรายคู่ทั้ง 3 คู่ คือ ระหว่างคะแนนเฉลี่ยร้อยละของปีการศึกษา 2529 กับปีการศึกษา 2530 คะแนนเฉลี่ยร้อยละของปีการศึกษา 2529 กับปีการศึกษา 2531 และระหว่างคะแนนเฉลี่ยร้อยละของปีการศึกษา 2530 กับปีการศึกษา 2531 ต่างมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งเมื่อพิจารณาประกอบกับสรุปผลการวิจัยข้อ 2.1 และ 2.2 พบว่าผลสัมฤทธิ์ในกลุ่มทักษะวิชาภาษาไทยมีความก้าวหน้าขึ้นทุกปีการศึกษาจริงซึ่งสอดคล้องกับสมมติฐานข้อ 1 ที่ได้ตั้งไว้ว่าผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามสูตรในแต่ละปีการศึกษามีความก้าวหน้าขึ้นตามลำดับ ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์พบว่าไม่เป็นไปตามสมมติฐานข้อ 1 กล่าวคือในปีการศึกษา 2530 มีผลสัมฤทธิ์เพิ่มขึ้นจากเดิม แต่ในปีการศึกษา 2531 มีผลสัมฤทธิ์ลดลง อัตราการเพิ่มขึ้นและลดลงนี้มีความแตกต่างจากปีการศึกษาก่อนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

สำหรับผลการวิเคราะห์ความแปรปรวน เพื่อเปรียบเทียบความแตกต่างของคะแนนเฉลี่ย ร้อยละ ระหว่าง 3 ปีการศึกษาของทั้งสองกลุ่มวิชา จำแนกเป็นรายอำเภอ พบว่า ในกลุ่มทักษะวิชา ภาษาไทยมีคะแนนเฉลี่ยร้อยละรายคู่ของผลการสอบระหว่างปีการศึกษา 2530 กับปีการศึกษา 2531 ของอำเภอเมือง และคะแนนเฉลี่ยร้อยละรายคู่ของผลการสอบระหว่างปีการศึกษา 2529 กับปี การศึกษา 2530 ของอำเภอลาดหลุมแก้ว ซึ่งมีความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ นอกนั้น ทุกคู่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์ พบว่า มีคะแนนเฉลี่ยร้อยละของปีการศึกษา 2529 กับปีการศึกษา 2531 ของอำเภอกลองหลวงเท่านั้นที่มี ความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ นอกนั้นคะแนนเฉลี่ยร้อยละรายคู่มีความแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

4. เมื่อเทียบคะแนนแบบสอบในปีการศึกษา 2530 และ 2531 ไปสู่สเกลคะแนนเดียว กับสเกลคะแนนในปีการศึกษา 2529 แล้ว พบว่า คะแนนจุดตัดการผ่านเกณฑ์ที่น่าพอใจของแบบสอบ ทั้งสามชุดแตกต่างกัน โดยคะแนนจุดตัดของปีการศึกษา 2530 จะมีค่าสูงสุด รองลงมาคือคะแนน จุดตัดของปีการศึกษา 2529 ส่วนคะแนนจุดตัดของปีการศึกษา 2531 จะมีค่าต่ำสุด การค้นพบ ดังกล่าวนี้นับว่าเหมือนกันทั้งสองกลุ่มวิชา กล่าวคือในกลุ่มทักษะวิชาภาษาไทยมีคะแนนจุดตัดใน แต่ละปีการศึกษาเท่ากับ 15 18 และ 14 ตามลำดับ พิสัยของคะแนนจุดตัดเท่ากับ 4 คะแนน ส่วน ในกลุ่มทักษะวิชาคณิตศาสตร์มีคะแนนจุดตัดในแต่ละปีการศึกษาเท่ากับ 19, 24 และ 18 ตามลำดับ พิสัยของคะแนนจุดตัดเท่ากับ 6 คะแนน

5. ผลการเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ ระหว่างสัดส่วนเมื่อ เทียบคะแนนแล้วกับสัดส่วนเมื่อยังไม่ได้เทียบคะแนน จำแนกเป็นรายกลุ่มวิชาและรายอำเภอ โดยใช้ สถิติทดสอบ Z-Test พบว่า สัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจเมื่อเทียบคะแนนแล้วกับเมื่อ ยังไม่ได้เทียบคะแนนของทั้งปีการศึกษา 2530 และปีการศึกษา 2531 มีความแตกต่างกันอย่างมี นัยสำคัญทางสถิติที่ระดับ .01 ทั้งสองกลุ่มวิชา ยกเว้นสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ ในกลุ่มทักษะวิชาคณิตศาสตร์เมื่อปีการศึกษา 2531 ที่พบว่าเมื่อเทียบคะแนนแล้วกับเมื่อยังไม่ได้เทียบ

คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ข้อค้นพบนี้สอดคล้องกับสมมุติฐาน ข้อ 2 ที่ได้ตั้งไว้ว่าสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์น่าพอใจระหว่างการเทียบคะแนนกับการไม่เทียบคะแนนแตกต่างกัน

เมื่อพิจารณาเป็นรายอำเภอ พบว่า อำเภอคลองหลวงมีสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจแตกต่างกันอย่างมีนัยสำคัญทางสถิติในทั้งสองกลุ่มวิชา ยกเว้นในกลุ่มทักษะวิชาคณิตศาสตร์ เมื่อปีการศึกษา 2531 ซึ่งพบว่าความแตกต่างไม่มีนัยสำคัญ และพบว่าสัดส่วนจำนวนนักเรียนที่มีผลน่าพอใจในกลุ่มทักษะวิชาคณิตศาสตร์ เมื่อปีการศึกษา 2530 มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกอำเภอ นอกนั้นพบว่าความแตกต่างไม่มีนัยสำคัญทางสถิติ

เนื่องจากผลการเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจเมื่อปีการศึกษา 2530 และปีการศึกษา 2531 ระหว่างเมื่อเทียบคะแนนแล้วกับเมื่อยังไม่เทียบคะแนน ในแต่ละกลุ่มวิชากรณีรวมทุกอำเภอนั้น ได้พบว่าความแตกต่างมีนัยสำคัญทางสถิติทั้งสองกลุ่มวิชาดังกล่าว ดังนั้น เมื่อผู้วิจัยได้แจกแจงร้อยละนักเรียนที่ผ่านเกณฑ์และไม่ผ่านเกณฑ์ ทั้งกรณีเทียบและไม่เทียบคะแนนแล้ว สรุปผลได้ดังนี้

1. กลุ่มทักษะวิชาภาษาไทยปีการศึกษา 2530 พบว่า เมื่อไม่เทียบคะแนนมีนักเรียนจำนวนร้อยละ 10 ที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจ แต่ได้รับการตัดสินผลให้ไม่ผ่านเกณฑ์ที่น่าพอใจ ส่วนปีการศึกษา 2531 พบว่าเมื่อไม่เทียบคะแนนมีนักเรียนร้อยละ 7 ที่มีคะแนนผลการสอบไม่ผ่านเกณฑ์ที่น่าพอใจ แต่ได้รับการตัดสินให้ผ่านเกณฑ์ที่น่าพอใจ

2. กลุ่มทักษะวิชาคณิตศาสตร์ ในปีการศึกษา 2530 พบว่า เมื่อไม่เทียบคะแนนมีนักเรียนจำนวนร้อยละ 20 ที่มีผลการสอบผ่านเกณฑ์ที่น่าพอใจ แต่ได้รับการตัดสินผลให้ไม่ผ่านเกณฑ์ที่น่าพอใจ ส่วนปีการศึกษา 2531 พบว่าเมื่อไม่เทียบคะแนนมีนักเรียนร้อยละ 5 ที่มีคะแนนผลการสอบไม่ผ่านเกณฑ์ที่น่าพอใจ แต่ได้รับการตัดสินผลให้ผ่านเกณฑ์ที่น่าพอใจ

อภิปรายผล

การตรวจสอบคุณภาพการศึกษาโดยใช้แบบสอบวัดผลสัมฤทธิ์ทางการเรียนสอบวัดผู้เรียนที่เรียนจบหลักสูตรในทุกปีการศึกษาอย่างต่อเนื่องหรือสอบวัดในแต่ละช่วงเวลาต่าง ๆ กัน เช่น ในช่วงสองปีการศึกษา หรือในช่วงสามปีการศึกษา เป็นต้น เพื่อจะศึกษาผลสัมฤทธิ์ทางการเรียนและความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนของนักเรียนที่เรียนจบหลักสูตรแต่ละรุ่นนั้น หากใช้แบบสอบต่างชุดกันแล้ว คะแนนที่ได้จากแบบสอบแต่ละชุดจะไม่สามารถนำมาเปรียบเทียบกันโดยตรง ทั้งนี้เพราะคะแนนจากแบบสอบต่างชุดจะอยู่ในสเกลที่ต่างกัน ซึ่งขึ้นอยู่กับความแตกต่างในความยากระหว่างแบบสอบเป็นสำคัญ ดังนั้นในการวิจัยครั้งนี้ ผู้วิจัยจึงไม่ได้ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนที่ได้จากการสอบแบบสอบต่างชุดในกรณีที่ยังไม่ได้เทียบคะแนน

หลังจากผู้วิจัยทำการเทียบคะแนน เพื่อปรับคะแนนที่ได้จากแบบสอบต่างชุดในแต่ละกลุ่มวิชาให้อยู่ในสเกลเดียวกันทั้งหมด โดยใช้สเกลคะแนนผลการสอบเมื่อปีการศึกษา 2529 เป็นหลัก และทำการเทียบคะแนนที่ได้จากการสอบแบบสอบที่ใช้ในปีการศึกษา 2530 และ 2531 ไปสู่สเกลเดียวกับสเกลคะแนนผลการสอบเมื่อปีการศึกษา 2529 แล้วจึงสามารถนำคะแนนที่สมมูลกันระหว่างแบบสอบต่างชุดมาเปรียบเทียบกันได้อย่างมีความหมาย ซึ่งจากผลการวิจัยครั้งนี้ พบว่าผู้สอบที่สอบแบบสอบต่างชุดกันแต่ได้คะแนนเท่ากันนั้น ภายหลังจากทำการเทียบคะแนนแล้วจะได้คะแนนต่างกัน หรือกล่าวได้ว่าจะคะแนนที่เท่ากันระหว่างแบบสอบต่างชุดมีค่าแตกต่างกัน โดยในกลุ่มทักษะวิชาภาษาไทย มีความแตกต่างกันสูงสุด 3.8 คะแนน ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์มีความแตกต่างกันสูงสุด 6.3 คะแนน ดังนั้นจึงมีผลให้คะแนนของผู้สอบแต่ละคนและคะแนนเฉลี่ยของผลการสอบในปีการศึกษา 2530 และ 2531 ระหว่างเมื่อเทียบคะแนนแล้วกับเมื่อยังไม่ได้เทียบคะแนนมีค่าแตกต่างกันทั้งสองกลุ่มวิชา ซึ่งจะเห็นว่าการเทียบคะแนนได้ให้ข้อสรุปใหม่ที่มีความคมชัด (precise) และเชื่อถือได้มากยิ่งขึ้น ทั้งนี้เพราะว่าการเปรียบเทียบคะแนนระหว่างแบบสอบต่างชุดไม่มีผลอันเนื่องมาจากความแตกต่างในความยากระหว่างแบบสอบต่างชุดนั้น

สำหรับการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) จากการวิจัยครั้งนี้พบว่าในกลุ่มทักษะวิชาภาษาไทยนั้นนักเรียนมีผลสัมฤทธิ์ทางการเรียนที่ก้าวหน้าขึ้นกว่าเดิมทุกปีการศึกษา ซึ่งสอดคล้องกับสมมติฐานข้อ 1 ที่ตั้งไว้ว่าผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบหลักสูตรในแต่ละปีการศึกษามีความก้าวหน้าขึ้นตามลำดับ ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์พบว่าในปีการศึกษา 2530 นักเรียนมีผลสัมฤทธิ์ทางการเรียนสูงขึ้นจากในปีการศึกษา 2529 แต่ในปีการศึกษา 2531 นักเรียนมีผลสัมฤทธิ์ทางการเรียนลดลงจากในปีการศึกษา 2530 ซึ่งไม่เป็นไปตามสมมติฐานข้อ 1 และเมื่อผู้วิจัยทำการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในแต่ละปีการศึกษาจำแนกเป็นรายอำเภอ แล้วพบว่า ในกลุ่มทักษะวิชาภาษาไทยมีเพียง 1 อำเภอ จากอำเภอที่ศึกษา 4 อำเภอ ที่มีคะแนนผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิมทุกปีการศึกษา ส่วนอำเภออื่นมีผลสัมฤทธิ์ในบางปีการศึกษาลดลงจากปีการศึกษา ก่อนเพียงเล็กน้อย เมื่อพิจารณาผลสัมฤทธิ์ทางการเรียนระหว่างปีการศึกษา 2529 กับปีการศึกษา 2531 พบว่าทุกอำเภอยังคงมีผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิม ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์พบว่าในปีการศึกษา 2530 ทุกอำเภอมียกคะแนนเฉลี่ยร้อยละสูงขึ้นมากแต่ในปีการศึกษา 2531 ทุกอำเภอกลับมียกคะแนนเฉลี่ยร้อยละลดลงมากเช่นกัน โดยมี 2 อำเภอที่ยังคงมียกคะแนนเฉลี่ยร้อยละในปีการศึกษา 2531 สูงกว่าในปีการศึกษา 2529 ส่วนอีก 2 อำเภอมีคะแนนเฉลี่ยร้อยละลดลงจากในปีการศึกษา 2529 เพียงเล็กน้อย เมื่อผู้วิจัยทำการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละของผลการสอบระหว่าง 3 ปีการศึกษาในแต่ละกลุ่มวิชา ทั้งเมื่อรวมทุกอำเภอและจำแนกเป็นรายอำเภอ โดยใช้สถิติวิเคราะห์ความแปรปรวน (ANOVA) แล้วพบว่าคะแนนเฉลี่ยระหว่าง 3 ปีการศึกษามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติทั้งสองกลุ่มวิชา หลังจากนั้นผู้วิจัยจึงทำการเปรียบเทียบความแตกต่างของคะแนนเฉลี่ยร้อยละรายคู่ต่อไป ซึ่งก็ได้พบว่าความแตกต่างมีนัยสำคัญทางสถิติทุกการเปรียบเทียบ แสดงว่าความแตกต่างที่เพิ่มขึ้นหรือลดลงของคะแนนเฉลี่ยร้อยละระหว่างปีการศึกษาต่าง ๆ นั้น มีความแตกต่างกันจริง สำหรับการลดลงของคะแนนเฉลี่ยร้อยละ

ในบางปีการศึกษา ซึ่งไม่สอดคล้องกับสมมติฐานข้อ 1 นั้น อาจเป็นเพราะระบบการเร่งรัดคุณภาพ การศึกษาทั้งทางด้านการจัดการเรียนการสอน การบริหารและการนิเทศการศึกษา ยังทำได้ไม่ทั่วถึง และจริงจังเท่าที่ควร

ส่วนค่าร้อยละนักเรียนที่มีผลน่าพอใจในแต่ละปีการศึกษานั้นในกลุ่มทักษะวิชาภาษาไทย พบว่าส่วนใหญ่ักเรียนที่มีผลน่าพอใจสูงกว่าร้อยละ 50 ทุกปีการศึกษายกเว้นในปีการศึกษา 2529 และ 2530 ซึ่งมีนักเรียนที่มีผลน่าพอใจต่ำกว่า 50 เพียงหนึ่งอำเภอจากทั้งหมดสี่อำเภอ ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์พบว่าส่วนใหญ่ักเรียนมีผลน่าพอใจต่ำกว่าร้อยละ 50 ยกเว้นในปีการศึกษา 2530 ซึ่งพบว่ามีสองอำเภอจากสี่อำเภอที่มีนักเรียนที่มีผลน่าพอใจสูงกว่าร้อยละ 50 การที่นักเรียน มีผลน่าพอใจในแต่ละกลุ่มวิชาไม่ถึงร้อยละ 50 นั้น เนื่องมาจากนักเรียนมีผลสัมฤทธิ์ทางการเรียน ต่ำกว่าเกณฑ์ที่น่าพอใจคือได้คะแนนเฉลี่ยต่ำกว่าร้อยละ 50 ของคะแนนเต็มในแต่ละกลุ่มวิชาและ อาจเป็นเพราะความยากในเนื้อหาวิชาคณิตศาสตร์ จึงทำให้นักเรียนที่มีผลน่าพอใจในกลุ่มทักษะ วิชาคณิตศาสตร์มีจำนวนร้อยละน้อยกว่าในกลุ่มทักษะวิชาภาษาไทยทุกปีการศึกษา ทั้งเมื่อรวมทุกอำเภอ และจำแนกเป็นรายอำเภอ

ผลการพิจารณาจุดตัดการผ่านเกณฑ์ที่น่าพอใจในแต่ละกลุ่มวิชา เมื่อกำหนดจุดตัดการผ่าน เกณฑ์ที่ร้อยละ 50 ของคะแนนเต็มเหมือนกันทุกปีการศึกษา พบว่าหลังจากทำการเทียบคะแนนแล้ว จุดตัดของทั้งสามปีการศึกษามีค่าแตกต่างกันทั้ง 2 กลุ่มวิชา กล่าวคือจุดตัดในกลุ่มทักษะวิชาภาษาไทย ตั้งแต่ปีการศึกษา 2529 ถึง 2531 มีค่าเท่ากับ 15, 18 และ 14 ตามลำดับ ส่วนจุดตัดในกลุ่ม ทักษะวิชาคณิตศาสตร์มีค่าเท่ากับ 19, 24 และ 18 ตามลำดับ ดังนั้นการตัดสินผลการสอบเมื่อ ยังไม่ได้เทียบคะแนน ซึ่งจะมีจุดตัดการผ่านในกลุ่มทักษะวิชาภาษาไทยที่ 15 คะแนนทุกปีการศึกษา และในกลุ่มทักษะวิชาคณิตศาสตร์มีจุดตัด 19 คะแนนทุกปีการศึกษานั้น จึงให้ผลแตกต่างจากการตัดสิน ผลการสอบเมื่อเทียบคะแนนแล้ว

สำหรับผลการเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจในแต่ละกลุ่มวิชา ระหว่างสัดส่วนเมื่อทำการเทียบคะแนนแล้วกับเมื่อยังไม่ได้ทำการเทียบคะแนนโดยใช้สถิติ *Z-test*

พบว่าให้ผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติในทุกปีการศึกษาที่ทำการเทียบคะแนนในทั้งสองกลุ่มวิชา ซึ่งสอดคล้องกับสมมุติฐานข้อ 2 ที่ตั้งไว้ว่าสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจระหว่างเมื่อเทียบคะแนนกับเมื่อไม่ได้เทียบคะแนนแตกต่างกัน แต่เมื่อจำแนกเป็นรายอำเภอ พบว่าผลการเปรียบเทียบของบางอำเภอและบางปีการศึกษานั้นมีสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ทั้งนี้อาจเป็นเพราะเมื่อจำแนกเป็นรายอำเภอแล้ว กลุ่มตัวอย่างมีจำนวนน้อยลง ดังนั้นเมื่อกำหนดจุดตัดการผ่านเกณฑ์ที่น่าพอใจทั้งก่อนการเทียบคะแนนและหลังการเทียบคะแนนซึ่งมีค่าแตกต่างกันนั้น หากการแจกแจงคะแนนผลการสอบของแต่ละอำเภอไม่ได้เกาะกลุ่มกันอยู่ในช่วงใกล้กับคะแนนจุดตัดและมีจำนวนนักเรียนที่ผ่านจุดตัดเมื่อเทียบคะแนนแล้วกับเมื่อยังไม่ได้เทียบคะแนนแตกต่างกันน้อยแล้วในการใช้สถิติทดสอบ z ซึ่งขึ้นอยู่กับจำนวนกลุ่มตัวอย่าง จึงปรากฏแล้วว่าความแตกต่างดังกล่าวไม่มีนัยสำคัญทางสถิติ ข้อค้นพบนี้แสดงว่าหากกลุ่มตัวอย่างยังมีขนาดใหญ่มากขึ้นแล้ว การตัดสินใจผ่านเกณฑ์ที่น่าพอใจระหว่างเมื่อเทียบคะแนนแล้วกับเมื่อไม่ได้ทำการเทียบคะแนนจะยิ่งแตกต่างกันมากขึ้น ซึ่งสำหรับในการวิจัยครั้งนี้พบว่าในปีการศึกษา 2530 กลุ่มทักษะวิชาภาษาไทย มีนักเรียนร้อยละ 10 ที่ควรเป็นผู้สอบผ่านแต่ได้รับการตัดสินใจให้ไม่ผ่าน และในกลุ่มทักษะวิชาคณิตศาสตร์พบว่านักเรียนร้อยละ 20 ที่ควรเป็นผู้สอบผ่านแต่ได้รับการตัดสินใจให้ไม่ผ่าน ส่วนในปีการศึกษา 2531 พบว่าในกลุ่มทักษะวิชาภาษาไทยมีนักเรียนร้อยละ 7 และกลุ่มทักษะวิชาคณิตศาสตร์มีนักเรียนร้อยละ 5 ที่ควรเป็นผู้สอบไม่ผ่าน แต่ได้รับการตัดสินใจให้ผ่านเมื่อไม่เทียบคะแนน ดังนั้นการเทียบคะแนนจึงช่วยให้การตัดสินใจมีความถูกต้องมากยิ่งขึ้น และช่วยให้ผู้เกี่ยวข้องกับการจัดการศึกษาทุกระดับสามารถนำไปประกอบการตัดสินใจในการปรับปรุงคุณภาพทางการศึกษาในส่วนที่พบว่านักเรียนยังมีผลสัมฤทธิ์ทางการเรียนไม่ผ่านเกณฑ์ที่น่าพอใจได้อย่างมีประสิทธิภาพต่อไป

ข้อเสนอแนะ

1. ในการศึกษาครั้งนี้ ผู้วิจัยศึกษาในสถานการณ์จริงซึ่งการดำเนินการสอบนักเรียนที่เรียนจบหลักสูตรแต่ละรุ่น เมื่อปีการศึกษา 2529 ถึง 2531 ทำโดยไม่มีข้อสอบร่วมหรือกลุ่มผู้สอบร่วม ดังนั้นในการศึกษาต่อไปควรมีการวางแผนดำเนินการในการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนของนักเรียนแต่ละรุ่น โดยใช้วิธีใดวิธีหนึ่ง ดังนี้

1.1 ควรใช้แบบสอบต่างชุดดำเนินการสอบนักเรียนที่เรียนจบหลักสูตรแต่ละรุ่นและควรมีแบบสอบร่วมซึ่งวัดในสิ่งเดียวกันกับแบบสอบที่จะทำการเทียบคะแนนนั้น ซึ่งแบบสอบร่วมจะประกอบด้วยข้อสอบที่วัดสมรรถภาพย่อย เช่นเดียวกันกับแบบสอบที่จะทำการเทียบคะแนนแต่อาจมีจำนวนข้อน้อยกว่าได้ ทั้งนี้ควรมีจำนวนข้ออย่างน้อยร้อยละ 20 ของข้อสอบชุดที่ใช้ทำการเทียบคะแนน ในชุดที่มีข้อสอบจำนวนมากกว่า โดยแบบสอบร่วมจะมีลักษณะคล้ายกับว่าย่อมาจากแบบสอบชุดที่ใช้เทียบคะแนนนั้นหรือเป็นตัวแทน (represent) แบบสอบชุดที่ใช้เทียบคะแนนนั้นได้ซึ่งหากแบบสอบร่วมมีความสัมพันธ์กับแบบสอบชุดที่ใช้ทำการเทียบคะแนนมากเท่าใด ผลการเทียบคะแนนจะยิ่งมีความคมชัดมากยิ่งขึ้น ทั้งนี้อาจสอบรวมหรือสอบต่างเวลากับแบบสอบต่างชุดที่ใช้เทียบคะแนนในแต่ละปีการศึกษานั้นก็ได้ แต่กลุ่มผู้สอบต้องทำข้อสอบร่วมเหมือนกันทุกรุ่น ข้อมูลจากข้อสอบร่วมจะใช้เชื่อมโยงคะแนนผลการสอบในสถานการณ์จริงที่ดำเนินการสอบนักเรียนแต่ละรุ่นโดยใช้แบบสอบต่างชุดนั้นเข้าด้วยกัน เพื่อให้สามารถแปลงคะแนนที่ได้จากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกันทั้งหมดและสามารถเทียบคะแนนระหว่างแบบสอบต่างชุดได้ อนึ่งการรักษาลักษณะของข้อสอบร่วมอาจทำได้โดยการใช้ข้อสอบร่วมชุดอื่น และทำการเทียบคะแนนระหว่างแบบสอบร่วมก่อนทำการเทียบคะแนนไปสู่คะแนนผลการสอบในแต่ละปีการศึกษานั้น หรืออาจใช้แบบสอบร่วมกับกลุ่มตัวอย่างย่อยที่สุ่มมากลุ่มหนึ่งเท่านั้น โดยไม่ต้องสอบกับนักเรียนกลุ่มตัวอย่างทุกคน ทั้งนี้เพื่อกำหนดค่าคงที่และสมการเชิงเส้นตรงสำหรับการเทียบคะแนนผลการสอบในแต่ละปีการศึกษาต่อไป

1.2 หากไม่สะดวกที่จะสร้างข้อสอบร่วมให้เป็นตัวแทนในแต่ละสมรรถภาพ หรือ ไม่ต้องการพบกับปัญหาในเรื่องการรักษาความลับของข้อสอบร่วม อาจดำเนินการโดยใช้ผู้สอบร่วม กล่าวคือในแต่ละปีการศึกษาจะสร้างแบบสอบชุดใหม่ขึ้นโดยไม่มีข้อสอบข้อใดซ้ำกับข้อสอบชุดเดิม การดำเนินการสอบในแต่ละปีการศึกษาจะทำการสุ่มสอบนักเรียนกลุ่มตัวอย่างย่อยกลุ่มหนึ่ง โดยให้ สอบแบบสอบทั้ง 2 ชุด ได้แก่ แบบสอบชุดที่จะใช้เป็นหลักในการเทียบคะแนน และแบบสอบที่สร้างขึ้น ในแต่ละปีการศึกษานั้น การจัดลำดับก่อนหลังของการทำแบบสอบสองชุดใช้วิธีสุ่ม ส่วนกลุ่มตัวอย่าง ที่เหลือทั้งหมดจะสอบแบบสอบที่ใช้ในแต่ละปีการศึกษาเพียง 1 ชุด ผลจากการใช้ผู้สอบร่วมจะนำไป กำหนดค่าคงที่และสร้างสมการในการแปลงคะแนนจากแบบสอบต่างชุดให้อยู่ในสเกลเดียวกันใน สถานการณ์การดำเนินการสอบจริงในแต่ละปีการศึกษา ซึ่งผู้สอบแต่ละรุ่นจะสอบแบบสอบเพียงชุดเดียว เท่านั้นก็จะสามารถเทียบคะแนนได้

1.3 ควรสร้างคลังข้อสอบที่เคยผ่านการทดลองใช้และได้อาศัยทฤษฎีการตอบข้อสอบ ในการหาประมาณพารามิเตอร์ของผู้สอบและของข้อสอบรายข้อไว้แล้ว เมื่อจะดำเนินการสอบ นักเรียนแต่ละรุ่นผู้สร้างแบบสอบจะจัดชุดของข้อสอบแต่ละครั้งแตกต่างกัน มาประกอบในแบบสอบ แต่ละชุด วิธีนี้นักวิจัยสามารถสร้างตารางคะแนนสมมูลจากแบบสอบชุดหนึ่งไปยังอีกชุดหนึ่งไว้ก่อน เมื่อดำเนินการสอบนักเรียนในแต่ละรุ่นจึงสามารถทำการเทียบคะแนนได้ทันทีโดยใช้ตารางคะแนน สมมูลซึ่งสร้างไว้แล้วนั้น ดังนั้นจึงสะดวกสำหรับการเสนอผลการสอบได้อย่างรวดเร็ว วิธีนี้เป็นวิธี การเทียบคะแนนซึ่งเรียกว่า Pre-equating ซึ่งจะต้องมีข้อสอบในคลังข้อสอบเป็นจำนวนมาก เพื่อให้สามารถสร้างชุดของข้อสอบได้โดยไม่ซ้ำกัน

2. สำหรับการศึกษาต่อไปควรนำรูปแบบการเทียบคะแนนรูปแบบต่าง ๆ ซึ่งในปัจจุบันมี ผู้คิดค้นไว้หลายรูปแบบมาทำการศึกษา ซึ่งได้แก่การเทียบคะแนนโดยใช้วิธีวิเคราะห์องค์ประกอบ และการเทียบคะแนนโดยใช้วิธี Pre-equating เป็นต้น ทั้งนี้โดยศึกษาถึงความเพียงพอของวิธีการ เทียบคะแนนแบบต่าง ๆ และเลือกใช้วิธีการที่เหมาะสมกับการสอบในแต่ละสถานการณ์ด้วย

3. ควรได้นำวิธีการในการเทียบคะแนนแบบสอบไปใช้ในจุดมุ่งหมายอื่น ๆ เช่น โครงการคัดเลือกบุคคลเข้าทำงาน โครงการทดสอบเพื่อปรับตำแหน่งในการปฏิบัติงาน เป็นต้น ทั้งนี้เพื่อให้มีมาตรฐานในการตัดสินผลการสอบและให้ความยุติธรรมแก่ผู้สอบมากขึ้น

4. ในการทดสอบเพื่อวัดผลสัมฤทธิ์ทางการเรียนตามหลักสูตรในชั้นเรียนซึ่งมีจำนวนนักเรียนไม่มากนัก หากครูผู้สอนต้องการจะเปรียบเทียบคะแนนผลสัมฤทธิ์ทางการเรียนซึ่งใช้แบบสอบต่างชุด อาจใช้วิธีการเทียบคะแนนเชิงเส้นตรงหรือการเทียบคะแนนแบบอิกวิเปออร์เซนไทล์ หรือการเทียบคะแนนรูปแบบราสซซึ่งผู้วิจัยใช้ครั้งนี้ก็ได้ เพราะวิธีต่าง ๆ เหล่านี้เข้าใจง่ายและการคำนวณไม่ซับซ้อนมาก

5. ควรทำการเทียบคะแนนในแนวตั้ง ได้แก่ ศึกษาความก้าวหน้าของนักเรียนที่เรียนต่างระดับชั้นกัน เพื่อให้ได้ข้อมูลในการวางแผนปรับปรุงให้ผลสัมฤทธิ์ทางการเรียนของนักเรียนถึงเกณฑ์มาตรฐานที่น่าพอใจที่กำหนดไว้สำหรับแต่ละชั้นต่อไป

บรรณานุกรม

บรรณานุกรม

- ชูชีพ พงษ์สมบูรณ์. การเปรียบเทียบการเทียบมาตราระหว่างรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบร่วม. วิทยานิพนธ์ ก.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2530. 118 หน้า (130).
- ภาวิณี ศรีสุขวัฒนานนท์. การเปรียบเทียบผลจากการใช้รูปแบบการเทียบมาตราที่ต่างกัน เมื่อแบบสอบร่วมมีความยาวต่างกัน. วิทยานิพนธ์ ก.ค. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2530. 197 หน้า.
- เรวดี อินทสระ. การเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตราระหว่างรูปแบบอิงทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ. ปริญญานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2530. 69 หน้า (112).
- สงบล ลักษณะ. "การเทียบระดับคะแนนระหว่างแบบสอบ," วารสารการวัดผลการศึกษา. 2525, 4(2) ; 21 - 32.
- สุพัตน์ สุกมลสันต์. การวิเคราะห์ข้อทดสอบแนวใหม่ด้วยเครื่องคอมพิวเตอร์. เอกสารลำดับที่ 1 สถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย, 2530. 150 หน้า.
- สุรัตน์ ขวัญบุญจันทร์. การสร้างแบบทดสอบและตารางเทียบคะแนนในแนวนอนวิชาวิทยาศาสตร์ที่วิเคราะห์ด้วยราสซีโมเดล. ปริญญานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2531. 99 หน้า.
- สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ. รายงานการประเมินความก้าวหน้า : คุณภาพนักเรียนชั้นประถมปีที่ 6 ระดับประเทศ ปีการศึกษา 2530. โรงพิมพ์คุรุสภาลาดพร้าว กรุงเทพมหานคร, 2531. 266 หน้า.
- อาภรณ์ กาญจนกิจโสภณ. การสร้างแบบทดสอบและตารางเทียบคะแนนตามแนวตั้งในวิชาคณิตศาสตร์. ปริญญานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2531. 89 หน้า.

- Angoff, W.H. "Scale, Norm and Equivalent Scores," In R.L. Thorndike (Ed.). Educational Measurement. 2nd ed. Washington : D.C. American Council on Educational. 1971 : 508 - 600.
- Angoff, W.H. "Summary and Derivation of Equating Methods Used at ETS," In P.W. Holland and D.B. Rubin (Eds), Test Equating New York : Academic Press, 1972 : 55 - 70.
- ✓ Angoff, William H. "Technical and Practical Issues in Equating : A Discussion of Four Papers," Applied Psychological Measurement. 11(1987) : 291 - 300.
- ✓ Angoff, William H. and Cowell, William R. "An Examination of the Assumption that the Equating of Parallel Forms is Population - Independent," Journal of Education Measurement. 23(1986) : 327 - 345.
- Bianchini, J.C. and Loret, P.C. "Anchor test study : the long and the short of it," Journal of Education Measurment. 12(3) : 201 - 213 ; 1975.
- Braun, Henry I. and Holland, P.W. "Observed-Score Test Equating : A mathematical Analysis of Some ETS Equating Procedures," In P.W. Holland and D.W. Holland and D.B. Rubin (Eds), Test Equating. New York : Academic Press. 1982 : 9 - 50.
- Brennan Robert L. and Kolen Michael. "Some Practical Issues in Equating," Applied Psychological Measurement. 11(1987) : 279 - 290.
- Budescu David V. "Selecting on Equating Method Linear or Equipercentile?," Journal of Educational Statistic. 12(1987) : 33 - 43.
- Budescu David. "Efficiency of Linear Equating as a Function of the Length of the Anchor Test," Journal of Educational Measurement. 12(1985) : 13 - 20.
- Carsrud, Karen Banks Ligon, Glynn. "Equating Studies : A Manual of Issues, Options and Decissions for Public Evaluators," A paper presented at the annual meeting of the American Educational Research Associaltion Los Angeles, 1981.
- Cook Linda L; Dunbar; S.B. and Eignor; D.R. "IRT Equating : A Flexible Alternative to Conventional Methods for Solving Practical Testing Problems," A paper presented at the annual meeting of the American Educational Research Association Los Angeles, 1981.

- Cook Linda L. and Petersen Nancy S. "Problems Related to the Use of Conventional and Item Response Theory Equating Methods in Less Than Optimal Circumstances," Applied Psychological Measurement. 11(1987) : 224 - 225.
- Cook Linda L. Dorans Neil J. and Eignor Daniel R. "An Assessment of the Dimensionality of three SAT-Verbal Test Editions," Journal of Educational Statistics. 13(1988) : 19 - 43.
- Cope, Ronald T. "How Well Do the Angoff Design V Linear Equating Methods Compaere With the Tucker and Levine Methods," Applied Psychological Measurement. 11(1987) : 143 - 149.
- Dorans, Neil J. and Kingston, Neal M. "The Effects of Vioations of Unidimensionality on the Estimation of Item and Ability Parameters and on Item Response Theory Equating of the GTE Verbal Scale," Journal of Educational Measurement. 22(1985) : 249 - 262.
- Douglass, J.B. "A Comparison of Item Characteristic Curve Models for classroom examination system," Dissertation Abstracts International. 41(March. 1981) : 4000-A.
- Faibank, Jr. Benjamin A. "The Use of Presmoothing and Postsmoothing to Increase the Precision of Equipercetile Equating," Applied Psychological Measurement 11. (1987) : 245 - 262.
- Flanagan, T.C. "Unit, Scores and Norms," In E.F. Lindquist (Ed.) Educational Measurement. Washington D.C. : American Council on Education, 1951 : 695 - 763.
- Fligmer Michael A. "Small Sample Asymptotics," Journal of Educational Statistics. 13(1988) : 53 - 61.
- Gulliksen, Harold. Theory of Mental Tests. New York : John Wiley and Sons, Inc., 1950.
- Holland, Paul W. and Rubin Donald B. (Ed). Test Equating. Educational Testing Service, Princetion, New Jersey : Academic Press, 1982 : 365 p.
- Hembleton, R.K. and others. "Development in Latent Trait Theory Model Technical Issues and Applications," Review of Educational Research. 48(1978) : 467 - 510.

- Hambleton R.K. and Swaminathan H. Item Response Theory. Kluwer. Nijhoff, Boston. 1985 : 312 p.
- Hutten, Leah R. "The fit of Empirical Data to two Latent Trait Models. Dissertation Abstracts International. 42(1982) : 4799-A.
- Jaeger, R.M. "Some Explorator Indices for Selection of a Test Equating Method," Journal of Educational Measurement. 18(1981) : 23 - 38.
- Jansen Paul G.W. and other. "Correcting Unconditional Parameter Estimates in the Rasch Model for Inconsistence," Applies Psychological Measurement. 12(1988) : 297 - 306.
- Kelder man Henk. "Common Item Equating Using the Loglinear Rasch Model," Journal of Educational Statistics. 13(1988) : 319 - 336.
- Klein, Lawrence W. and Jarjoura, David. "The Importance of Content Representation for Common-Item Equating with Nonrandom Group," Journal of Educational Measurement. 22(1985) : 197 - 206.
- Kolen, M.J. "Comparison of Traditional and Item Respose theory Methods for Equating Tests," Journal of Educational Measurement. 18(1981) : 1 - 11.
- Kolen, M.J. and Whitney, D.R. "Comparison of Four Procedures for Equating the Tests of General Educational Development," " Journal of Educational Measurement. 19(1982) : 279 - 294.
- Kagan, D.M. and Stock, W.A. "Equivalencing MAT and GRE Scores using simple Linear transformation and regression methods," Journal of Experimetal Education. 1980 : 34 - 37.
- Lord, F.M. Applicational of Item Response Theory to Practical Testting Problems. Hillsdale, N.J. Erbaum, 1980 : 273 p.
- Lord, F.M. "Standard Error of and Equating by Item Response Theory," Educational Testing Service, Princeton, New Jersey : ERIC Document Reproduction Service, November, 1981.
- Lord, F.M. (a). "The Standard Error of Equipercentile Equating," Journal of Education Statistics. 7(1982) : 165 - 174.
- Lord, F.M. (b). "Item Response Theory and Equating-A Technical Summary," In P.W. Holland and D.B. Ruin (Eds), Test Equating. New York : Academic Press, 1982 : 141 - 148.
- Lord, F.M. and Wingersky, M.S. "Comparison of IRT true score and equipercentile observed-score equating," Journal of Applied psychological measurement. 1984 : 453 - 461.

- Marascuilo, L.A. and Mcsweeney, M. Nonparametric and Distribution free Methods for the Social Sciences, Monterey, C.A. :
- Marco, Gary L. "Item Characteristic Curve Solutions to three Intractable Testting Problems," Journal of Educational Measurement. 14(1977) : 139 - 160.
- Masters Geofferey N. "Common - Person Equating with the Rasch Model," Applied Psychological Measurement. 9(1985) : 73 - 82.
- Petersen, N.S., Marco, G.L. and Stewart, E.E. "A test of the Adequacy of linear score equating Methods," In P.W. Holland and D.B. Rubin (Eds.) Test Equating. New York : Academic Press. 1982 : 71 - 135.
- Potthoff, R.F. "Some Issues in Test Equating," In P.W. Holland and D.B. Rubin (Eds.) Test Equating. New York : Academic press. 1982 : 201 - 242.
- Rangchaikul, Yavadee. "Equating and Calibration Technique for Test Program in Thailand," A Dissertation Submitted to the Area of Educational Management System, The Florida State University. December. 1975.
- Rentz and Bashaw. "The National Reference Scale for Reading : An Application of the Rasch Model," Journal of Educational Measurement. 14(1977) : 161 - 178.
- Rubin, D.B. "Discussion of Observed-Score Equating : A Mathematical Analysis of Some ETS Equating Procedures," In P.W. Holland and D.B. Rubin (Eds.) Test Equating. New York. Academic Press. 1982 : 51 - 54.
- Skaggs, Gary. and Lissitz, Robert W. "IRT Test Equating : Relevent Issues and a Review of Recent Research," Review of Educational Research. 56(1986) : 495 - 529.
- Skaggs Gary. Lissitz, Robert W. "An Exploration of the Robustness of Four Test Equating Models," Applied Psychological Measurement. 10(1986) : 303 - 317.
- Stroud, T.W. "Discussion of A test of the Adequacy of Linear Score Equating Models," In P.W. Holland and D.B. Rubin (Eds.) Test Equating. New York. Academic Press. 1982 : 137 - 140.

- Van der Wollenberg Arnold L. Wierda Folkert W. and Jansen Paul G.W.
 "Consisttency of Rasch Model Rarmeter Estimation : A simulation
 Study," Applied Psychological Measurement. 12(1988) : 307 - 313.
- Vijver Fons J.R. Van de. "The Robustness of Rasch Estimateates," Applied
 Pyschological Measurement. 10(1986) : 45 - 57.
- Wingersky M.S., Barton and Lord, F.M. Logist 5 - A compater program for
 estimating examinee ability and item characteristic curve parameter.
 Princeton NI : Educational Testing Service, 1982.
- Wright, B.D. "Solving Measurement Problems with the rasch Model,"
Journal of Education Measurement. 14(1977) : 97 - 116.
- Wright, B.D. "Misunderstanding the Rasch Model," Journal of Education
 Measurement. 14(1977).
- Wright, B.D. and Panchapakesan, N. "A procedure for sample-item analysis,"
Educational and Psychological Measurement. 29(1967).
- Wright, B.D. and Stone, M.H. Best Test Design. Chicago : Mesa Press,
 1979 : 222 p.
- Yen W.M. "Obtaining Some Degree of Correspondence Between Unequatable
 Scores ; A Comparison of Item Response Theory and Equipercentile
 Equating Methods," A Paper Presented at the American Educational
 Resarch Association Meetings, New York, 1982.

BICAL "THAI" 2529 30 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW		LOG	STANDARD	RAW		LOG	STANDARD
SCORE	COUNT	ABILITY	ERRORS	SCORE	COUNT	ABILITY	ERRORS
29	0	3.68	1.04	14	52	-.12	.41
28	0	2.95	.76	13	41	-.29	.42
27	0	2.50	.63	12	45	-.46	.42
26	0	2.16	.56	11	17	-.63	.43
25	0	1.88	.52	10	22	-.81	.44
24	1	1.64	.49	9	20	-1.00	.45
23	5	1.42	.46	8	4	-1.20	.46
22	5	1.22	.45	7	6	-1.42	.48
21	11	1.04	.43	6	3	-1.65	.51
20	29	.86	.42	5	2	-1.91	.54
19	34	.69	.42	4	0	-2.21	.58
18	51	.53	.41	3	0	-2.57	.65
17	68	.36	.41	2	0	-3.04	.77
16	60	.20	.41	1	0	-3.79	1.05
15	59	.04	.41				

PERSON SEPARABILITY INDEX .48 (EQUIVALENT TO KR20)

38 ITEMS CALIBRATED ON 535 PERSONS

535 MEASURABLE PERSONS WITH MEAN ABILITY = .06 AND STD. DEV. = .58

BICAL "THAI" 2529 30 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.06 ABILITY SCALE FACTOR 1.20

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	.640	.093	.005	.648	.635
2	2	-1.531	.116	-.012	-1.557	-1.520
3	3	-1.741	.124	-.013	-1.774	-1.729
4	4	.725	.094	.006	.735	.719
5	5	.189	.090	.002	.191	.187
6	6	1.311	.105	.009	1.338	1.303
7	7	-1.598	.119	-.012	-1.626	-1.586
8	8	1.290	.105	.009	1.316	1.281
9	9	-.939	.100	-.007	-.951	-.932
10	10	.459	.091	.004	.465	.455
11	11	1.787	.120	.011	1.834	1.777
12	12	-2.328	.153	-.017	-2.387	-2.311
13	13	.379	.091	.004	.383	.376
14	14	.300	.090	.003	.303	.297
15	15	-.078	.090	.000	-.079	-.078
16	16	-.221	.091	-.001	-.223	-.219
17	17	-1.392	.112	-.011	-1.414	-1.382
18	18	-1.571	.118	-.012	-1.598	-1.559
19	19	.024	.090	.001	.024	.024
20	20	.331	.091	.003	.335	.328
21	21	.811	.095	.006	.823	.805
22	22	1.107	.100	.008	1.126	1.099
23	23	1.422	.108	.009	1.452	1.413
24	24	.873	.096	.007	.887	.867
25	25	-.779	.097	-.006	-.788	-.773

26	26	-.205	.091	-.001	-.207	-.204
27	27	.451	.091	.004	.456	.447
28	28	.395	.091	.004	.399	.392
29	29	-1.143	.105	-.009	-1.158	-1.134
30	30	1.030	.099	.007	1.047	1.023

ROOT MEAN SQUARE = .008

30 ITEMS CALIBRATED ON 535 PERSONS

535 MEASURABLE PERSONS WITH MEAN ABILITY = .06 AND STD. DEV. = .42

BICAL "THAI" 2530 30 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.06 ABILITY SCALE FACTOR 1.20

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-2.385	.163	-.006	-2.464	-2.379
2	2	-.694	.102	-.003	-.707	-.692
3	3	.292	.099	.001	.294	.291
4	4	-.614	.101	-.002	-.625	-.612
5	5	-.382	.099	-.001	-.389	-.380
6	6	-.755	.103	-.003	-.769	-.753
7	7	-.166	.098	-.001	-.171	-.165
8	8	-.297	.099	-.001	-.303	-.296
9	9	-.297	.099	-.001	.303	-.296
10	10	.244	.099	.001	.245	.244
11	11	.235	.099	.001	.236	.234
12	12	-.372	.099	-.001	-.380	-.371
13	13	1.152	.114	.004	1.174	1.148
14	14	.417	.100	.001	.421	.415
15	15	-1.151	.111	-.004	-1.173	-1.147
16	16	-.535	.100	-.002	-.545	-.534
17	17	2.139	.155	.007	2.209	2.132
18	18	.397	.100	.001	.401	.396
19	19	2.313	.166	.007	2.395	2.307
20	20	-.054	.098	.000	-.058	-.054
21	21	1.420	.122	.005	1.452	1.415
22	22	.881	.107	.003	.895	.879
23	23	-.027	.098	.000	-.029	-.026
24	24	-.231	.098	-.001	-.237	-.230
25	25	-.468	.100	-.002	-.477	-.466

26	26	.495	.101	.002	.508	.494
27	27	.011	.098	.000	.008	.011
28	28	-1.485	.120	-.004	-1.518	-1.480
29	29	1.923	.143	.006	1.980	1.917
30	30	-2.007	.142	-.005	-2.064	-2.002

ROOT MEAN SQUARE = .003

30 ITEMS CALIBRATED ON 453 PERSONS

453 MEASURABLE PERSONS WITH MEAN ABILITY = -.06 AND STD. DEV. = .42

BICAL "THAI" 2531 30 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW		LOG	STANDARD	RAW		LOG	STANDARD
SCORE	COUNT	ABILITY	ERRORS	SCORE	COUNT	ABILITY	ERRORS
29	0	3.56	1.04	14	31	-.15	.40
28	0	2.82	.76	13	32	-.30	.39
27	0	2.37	.63	12	18	-.45	.40
26	4	2.03	.56	11	12	-.61	.40
25	3	1.75	.52	10	5	-.77	.41
24	4	1.51	.48	9	2	-.93	.42
23	18	1.30	.46	8	3	-1.11	.44
22	29	1.11	.44	7	5	-1.30	.46
21	26	.93	.42	6	1	-1.51	.48
20	34	.76	.41	5	1	-1.75	.51
19	45	.60	.40	4	0	-2.02	.56
18	50	.44	.40	3	0	-2.36	.63
17	67	.29	.39	2	0	-2.80	.75
16	42	.14	.39	1	0	-3.53	1.04
15	33	-.01	.39				

PERSON SEPARABILITY INDEX .51 (EQUIVALENT TO KR20)

30 ITEMS CALIBRATED ON 465 PERSONS

465 MEASURABLE PERSONS WITH MEAN ABILITY = .32 AND STD. DEV. = .59

BICAL "THAI" 2531 30 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.06 ABILITY SCALE FACTOR 1.12

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-1.034	.117	-.003	-1.058	-1.031
2	2	-.060	.098	.000	-.060	-.060
3	3	-1.592	.139	-.005	-1.638	-1.587
4	4	.151	.097	.001	.155	.151
5	5	-.398	.102	-.001	-.404	-.397
6	6	.540	.097	.002	.550	.538
7	7	-.145	.099	-.001	-.145	-.144
8	8	-.932	.114	-.003	-.951	-.929
9	9	.377	.097	.001	.384	.376
10	10	.827	.099	.003	.842	.824
11	11	.106	.097	.000	.109	.105
12	12	-.521	.104	-.002	-.530	-.520
13	13	.780	.099	.003	.794	.777
14	14	1.062	.102	.004	1.082	1.058
15	15	2.169	.135	.006	2.230	2.163
16	16	-.564	.105	-.002	-.573	-.562
17	17	-.752	.109	-.003	-.766	-.749
18	18	.124	.097	.000	.127	.123
19	19	1.196	.104	.004	1.219	1.192
20	20	.440	.097	.001	.448	.439
21	21	-.378	.102	-.001	-.383	-.377
22	22	.413	.097	.001	.421	.412
23	23	-1.573	.138	-.005	-1.619	-1.569
24	24	.485	.097	.002	.494	.484
25	25	.594	.097	.002	.605	.592

26	26	.014	.098	.000	.016	.014
27	27	-.661	.107	-.002	-.673	-.659
28	28	-.650	.107	-.002	-.662	-.648
29	29	-.511	.104	-.002	-.519	-.509
30	30	.494	.097	.002	.504	.493

ROOT MEAN SQUARE = .003

30 ITEMS CALIBRATED ON 465 PERSONS

465 MEASURABLE PERSONS WITH MEAN ABILITY = .32 AND STD. DEV. = .42

BICAL "MATH" 2529 38 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.06 ABILITY SCALE FACTOR 1.12

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-.358	.090	-.001	-.358	-.357
2	2	-.766	.090	-.001	-.780	-.764
3	3	.235	.097	.000	.235	.235
4	4	-.213	.091	.000	-.221	-.212
5	5	1.075	.121	.002	1.100	1.073
6	6	.235	.097	.000	.235	.235
7	7	-.196	.092	.001	-.204	-.196
8	8	-1.904	.105	-.002	-1.938	-1.902
9	9	1.148	.124	.002	1.177	1.146
10	10	2.086	.179	.003	2.166	2.084
11	11	.616	.106	.001	.625	.615
12	12	-.310	.091	.000	-.319	-.309
13	13	.717	.109	.001	.729	.716
14	14	.226	.097	.000	.225	.226
15	15	.263	.098	.000	.263	.263
16	16	1.670	.150	.003	1.724	1.667
17	17	-.366	.090	-.001	-.376	-.365
18	18	1.194	.126	.002	1.224	1.192
19	19	-.358	.090	-.001	-.368	-.357
20	20	.263	.098	.000	.263	.263
21	21	.030	.094	.000	.026	.030
22	22	1.061	.121	.002	1.085	1.059
23	23	-1.556	.097	-.003	-1.582	-1.554
24	24	-.711	.090	-.001	-.725	-.710
25	25	-.633	.090	-.001	-.646	-.632

BICAL "MATH" 2529 38 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW		LOG	STANDARD	RAW		LOG	STANDARD
SCORE	COUNT	ABILITY	ERRORS	SCORE	COUNT	ABILITY	ERRORS
37	0	3.83	1.03	18	37	-.12	.35
36	0	3.11	.75	17	26	-.24	.35
35	0	2.66	.62	16	43	-.36	.35
34	0	2.33	.55	15	44	-.48	.35
33	0	2.06	.50	14	42	-.60	.36
32	0	1.83	.47	13	50	-.73	.36
31	0	1.63	.44	12	50	-.86	.37
30	0	1.45	.42	11	49	-1.00	.38
29	0	1.29	.41	10	41	-1.14	.39
28	0	1.13	.39	9	36	-1.29	.40
27	0	.99	.38	8	18	-1.45	.42
26	0	.85	.37	7	14	-1.65	.44
25	2	.72	.37	6	5	-1.83	.46
24	4	.59	.36	5	4	-2.05	.50
23	8	.47	.35	4	0	-2.31	.55
22	9	.35	.35	3	2	-2.64	.62
21	11	.23	.35	2	0	-3.08	.74
20	14	.11	.35	1	0	-3.80	1.03
19	26	-.01	.35				

PERSON SEPARABILITY INDEX .53 (EQUIVALENT TO KR20)

38 ITEMS CALIBRATED ON 535 PERSONS

535 MEASURABLE PERSONS WITH MEAN ABILITY = -.66 AND STD. DEV. = .54

26	26	-.147	.092	.000	-.154	-.147
27	27	-.221	.091	.000	-.229	-.221
28	28	-.326	.091	-.001	-.335	-.325
29	29	-.625	.090	-.001	-.638	-.624
30	30	-.413	.090	-.001	-.424	-.413
31	31	.021	.094	.000	.017	.021
32	32	-.445	.090	-.001	-.456	-.444
33	33	-.891	.090	-.001	-.907	-.890
34	34	-1.316	.094	-.002	-1.338	-1.314
35	35	.398	.101	.001	.401	.397
36	36	.190	.097	.000	.188	.189
37	37	.226	.097	.000	.225	.226
38	38	.100	.095	.000	.097	.100

ROOT MEAN SQUARE = .001

38 ITEMS CALIBRATED ON 535 PERSONS

535 MEASURABLE PERSONS WITH MEAN ABILITY = -.66 AND STD. DEV. = .39

BICAL "MATH" 2530 38 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW		LOG	STANDARD	RAW		LOG	STANDARD
SCORE	COUNT	ABILITY	ERRORS	SCORE	COUNT	ABILITY	ERRORS
37	0	3.88	1.04	18	24	-.13	.35
36	0	3.15	.75	17	35	-.25	.35
35	0	2.69	.63	16	30	-.37	.35
34	0	2.36	.56	15	36	-.49	.36
33	0	2.08	.51	14	30	-.61	.36
32	0	1.85	.47	13	33	-.74	.36
31	0	1.65	.45	12	27	-.87	.37
30	1	1.47	.42	11	28	-1.01	.38
29	1	1.30	.41	10	24	-1.15	.39
28	1	1.14	.39	9	20	-1.31	.40
27	5	.99	.38	8	12	-1.47	.42
26	5	.86	.37	7	12	-1.65	.44
25	6	.72	.37	6	4	-1.85	.47
24	17	.59	.36	5	1	-2.07	.50
23	11	.47	.36	4	0	-2.34	.55
22	19	.35	.35	3	1	-2.66	.62
21	15	.23	.35	2	0	-3.10	.74
20	27	.11	.35	1	0	-3.82	1.03
19	28	-.01	.35				

PERSON SEPARABILITY INDEX .68 (EQUIVALENT TO KR20)

38 ITEMS CALIBRATED ON 453 PERSONS

453 MEASURABLE PERSONS WITH MEAN ABILITY = -.43 AND STD. DEV. = .65

BICAL "MATH" 2530 38 ITEMS 4 CHOICES. (2530)

PROCEDURE IS UCON .

DIFFICULTY SCALE FACTOR 1.08 ABILITY SCALE FACTOR 1.13

NUMBER OF ITERATIONS = 2

SEQUENCE	ITEM	ITEM	STANDARD	LAST DIFF	PROX	FIRST
NUMBER	NAME	DIFFICULTY	ERROR	CHANGE	DIFF	CYCLE
1	1	.249	.103	.001	.246	.248
2	2	-.780	.100	-.002	-.789	-.777
3	3	.333	.105	.001	.331	.332
4	4	-.521	.099	-.002	-.528	-.519
5	5	1.553	.142	.005	1.590	1.548
6	6	.820	.115	.002	.828	.818
7	7	-.255	.099	-.001	-.261	-.254
8	8	-.111	.100	.000	-.116	-.110
9	9	.354	.105	.001	.353	.353
10	10	-.043	.100	.000	-.048	-.042
11	11	-.312	.099	-.001	-.319	-.311
12	12	-.226	.099	-.001	-.232	-.225
13	13	.475	.107	.001	.475	.473
14	14	.322	.104	.001	.321	.321
15	15	-.502	.099	-.002	-.509	-.500
16	16	1.234	.128	.004	1.256	1.230
17	17	.744	.113	.002	.751	.742
18	18	2.483	.204	.006	2.582	2.477
19	19	1.385	.134	.004	1.414	1.381
20	20	-1.324	.107	-.004	-1.340	-1.320
21	21	-.197	.099	-.001	-.204	-.197
22	22	.408	.106	.001	.408	.407
23	23	-1.719	.116	-.005	-1.744	-1.715
24	24	-1.719	.116	-.005	-1.744	-1.715
25	25	-.693	.099	-.002	-.701	-.691

26	26	-.378	.099	-.001	-.385	-.377
27	27	-.967	.102	-.003	-.978	-.964
28	28	-.274	.099	-.001	-.280	-.273
29	29	-1.079	.103	-.003	-1.091	-1.076
30	30	.259	.103	.001	.257	.258
31	31	-.217	.099	-.001	-.223	-.216
32	32	.565	.109	.002	.567	.563
33	33	.744	.113	.002	.751	.742
34	34	.486	.107	.002	.487	.484
35	35	-.003	.101	.000	-.008	-.003
36	36	.270	.104	.001	.267	.269
37	37	-.664	.099	-.002	-.672	-.662
38	38	-.702	.099	-.002	-.711	-.700

ROOT MEAN SQUARE = .002

38 ITEMS CALIBRATED ON 453 PERSONS

453 MEASURABLE PERSONS WITH MEAN ABILITY = -.43 AND STD. DEV. = .53

BICAL "MATH" 2531 38 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW		LOG	STANDARD	RAW		LOG	STANDARD
SCORE	COUNT	ABILITY	ERRORS	SCORE	COUNT	ABILITY	ERRORS
37	0	3.72	1.03	18	19	-.11	.34
36	0	3.00	.74	17	25	-.23	.34
35	0	2.57	.62	16	31	-.34	.34
34	0	2.24	.54	15	33	-.46	.35
33	0	1.98	.50	14	34	-.57	.35
32	0	1.76	.46	13	51	-.69	.36
31	1	1.57	.43	12	37	-.82	.36
30	0	1.39	.41	11	40	-.95	.37
29	0	1.23	.40	10	34	-1.09	.38
28	5	1.09	.38	9	18	-1.24	.40
27	3	.95	.37	8	9	-1.40	.41
26	8	.82	.36	7	10	-1.57	.43
25	6	.69	.36	6	5	-1.76	.46
24	9	.57	.35	5	1	-1.98	.49
23	12	.45	.35	4	0	-2.24	.54
22	8	.34	.34	3	1	-2.56	.62
21	18	.22	.34	2	0	-3.00	.74
20	21	.11	.34	1	0	-3.71	1.03
19	26	.00	.34				

PERSON SEPARABILITY INDEX .64 (EQUIVALENT TO KR20)

38 ITEMS CALIBRATED ON 465 PERSONS

465 MEASURABLE PERSONS WITH MEAN ABILITY = -.47 AND STD. DEV. = .61

BICAL "MATH" 2531 30 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.07 ABILITY SCALE FACTOR 1.08

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-.380	.097	-.001	-.384	-.379
2	2	-1.087	.100	-.002	-1.100	-1.085
3	3	.051	.100	.000	.050	.051
4	4	.229	.102	.000	.230	.228
5	5	.099	.100	.000	.099	.099
6	6	-.260	.097	-.001	-.264	-.260
7	7	.041	.100	.000	.040	.041
8	8	-.260	.097	-.001	-.264	-.260
9	9	-1.730	.113	-.003	-1.760	-1.727
10	10	-.839	.098	-.002	-.848	-.837
11	11	.714	.112	.001	.722	.713
12	12	.060	.100	.000	.059	.060
13	13	-.223	.098	.000	-.227	-.223
14	14	-.074	.099	.000	-.076	-.074
15	15	-.168	.098	.000	-.170	-.167
16	16	.595	.109	.001	.601	.594
17	17	.249	.102	.000	.250	.249
18	18	.666	.111	.001	.673	.664
19	19	-.526	.097	-.001	-.531	-.525
20	20	1.489	.140	.003	1.521	1.486
21	21	-.205	.098	.000	-.208	-.204
22	22	-.635	.097	-.001	-.642	-.634
23	23	-1.331	.104	-.002	-1.349	-1.329
24	24	-.196	.098	.000	-.199	-.195
25	25	-.036	.099	.000	-.038	-.036

26	26	.012	.099	.000	.011	.012
27	27	-.214	.098	.000	-.217	-.214
28	28	.290	.103	.001	.292	.290
29	29	-.416	.097	-.001	-.421	-.415
30	30	.239	.102	.000	.240	.239
31	31	-.102	.098	.000	-.104	-.102
32	32	.702	.112	.001	.710	.700
33	33	1.884	.161	.004	1.935	1.880
34	34	.099	.100	.000	.099	.099
35	35	.726	.112	.001	.734	.725
36	36	.239	.102	.000	.204	.239
37	37	.109	.100	.000	.109	.109
38	38	.188	.101	.000	.189	.188

ROOT MEAN SQUARE = .001

30 ITEMS CALIBRATED ON 465 PERSONS

465 MEASURABLE PERSONS WITH MEAN ABILITY = -.47 AND STD. DEV. = .49

THAI 2529 (Form 1) 30 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	LOG COUNT	ABILITY	STANDARD ERRORS	RAW SCORE	LOG COUNT	ABILITY	STANDARD ERRORS
29	0	3.91	1.06	14	14	-.20	.42
28	0	3.16	.78	13	14	-.37	.42
27	0	2.68	.66	12	12	-.55	.43
26	0	2.31	.59	11	6	-.72	.43
25	1	2.01	.54	10	3	-.91	.44
24	1	1.74	.51	9	4	-1.09	.45
23	18	1.50	.49	8	1	-1.29	.46
22	14	1.28	.47	7	0	-1.50	.48
21	21	1.07	.46	6	0	-1.73	.50
20	21	.87	.45	5	0	-1.98	.53
19	17	.68	.44	4	0	-2.27	.58
18	20	.50	.43	3	0	-2.63	.64
17	22	.32	.43	2	0	-3.09	.76
16	16	.14	.42	1	0	-3.83	1.05
15	13	-.03	.42				

PERSON SEPARABILITY INDEX .59 (EQUIVALENT TO KR20)

30 ITEMS CALIBRATED ON 218 PERSONS

218 MEASURABLE PERSONS WITH MEAN ABILITY = .42 AND STD. DEV. = .69

THAI 2529 (กลุ่มที่ 1) 30 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.09 ABILITY SCALE FACTOR 1.25

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	2.156	.187	.025	2.204	2.131
2	2	-1.408	.194	-.018	-1.436	-1.391
3	3	-1.408	.194	-.018	-1.436	-1.391
4	4	.274	.144	.003	.280	.271
5	5	-.540	.157	-.008	-.540	-.533
6	6	2.299	.195	.026	2.356	2.274
7	7	-1.523	.202	-.018	-1.557	-1.506
8	8	1.024	.148	.014	1.034	1.010
9	9	-.713	.162	-.010	-.716	-.704
10	10	1.219	.152	.016	1.232	1.203
11	11	1.705	.166	.021	1.731	1.684
12	12	-2.305	.270	-.023	-2.395	-2.283
13	13	.374	.144	.005	.380	.369
14	14	-.011	.146	-.001	-.005	-.010
15	15	-.540	.157	-.008	-.540	-.533
16	16	-.791	.165	-.011	-.796	-.781
17	17	-1.445	.197	-.018	-1.475	-1.428
18	18	-1.408	.194	-.018	-1.436	-1.391
19	19	-.179	.149	-.003	-.175	-.176
20	20	-.244	.150	-.004	-.240	-.240
21	21	1.045	.148	.014	1.055	1.031
22	22	1.002	.148	.014	1.012	.989
23	23	2.024	.180	.024	2.065	2.001

24	24	1.678	.165	.022	1.703	1.657
25	25	-1.301	.188	-.016	-1.323	-1.285
26	26	-.613	.159	-.009	-.614	-.604
27	27	.051	.146	.000	.057	.051
28	28	.153	.145	.002	.159	.152
29	29	-1.168	.181	-.015	-1.185	-1.153
30	30	.593	.144	.008	.600	.586

ROOT MEAN SQUARE = .015

30 ITEMS CALIBRATED ON 218 PERSONS

218 MEASURABLE PERSONS WITH MEAN ABILITY = .42 AND STD. DEV. = .53

THAI 2529 (กลุ่ม 2) 30 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	LOG COUNT	ABILITY	STANDARD ERRORS	RAW SCORE	LOG COUNT	ABILITY	STANDARD ERRORS
29	0	3.78	1.04	14	11	-.16	.42
28	0	3.05	.76	13	17	-.33	.42
27	0	2.59	.64	12	9	-.51	.43
26	0	2.24	.57	11	12	-.69	.43
25	0	1.95	.53	10	5	-.87	.44
24	0	1.70	.50	9	3	-1.06	.45
23	2	1.48	.47	8	0	-1.26	.46
22	3	1.27	.46	7	1	-1.48	.48
21	8	1.07	.45	6	0	-1.71	.50
20	17	.88	.44	5	0	-1.96	.53
19	28	.70	.43	4	0	-2.25	.58
18	19	.53	.43	3	0	-2.63	.64
17	24	.35	.42	2	0	-3.09	.76
16	21	.18	.42	1	0	-3.83	1.05
15	16	.01	.42				

PERSON SEPARABILITY INDEX .44 (EQUIVALENT TO KR20)

30 ITEMS CALIBRATED ON 196 PERSONS

196 MEASURABLE PERSONS WITH MEAN ABILITY = .22 AND STD. DEV. = .57

THAI 2529 (กล่มที่ 2) 30 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.06 ABILITY SCALE FACTOR 1.23

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	1.108	.161	.011	1.129	1.097
2	2	-1.289	.187	-.013	-1.314	-1.276
3	3	-1.469	.198	-.015	-1.501	-1.455
4	4	.532	.150	.006	.540	.526
5	5	-.480	.157	-.005	-.486	-.476
6	6	1.671	.183	.015	1.713	1.657
7	7	-1.809	.221	-.018	-1.857	-1.792
8	8	1.317	.167	.013	1.344	1.304
9	9	-1.222	.184	-.012	-1.245	-1.210
10	10	1.059	.159	.011	1.078	1.048
11	11	1.807	.190	.016	1.855	1.791
12	12	-2.016	.239	-.018	-2.075	-1.998
13	13	.381	.149	.005	.386	.376
14	14	.619	.151	.007	.629	.613
15	15	-.947	.171	-.010	-.962	-.937
16	16	-.703	.163	-.007	-.712	-.696
17	17	-1.588	.205	-.016	-1.625	-1.573
18	18	-1.189	.182	-.012	-1.211	-1.178
19	19	-.204	.152	-.002	-.205	-.202
20	20	.080	.149	.001	.082	.079
21	21	1.034	.159	.011	1.053	1.024
22	22	1.372	.169	.013	1.401	1.359
23	23	1.807	.190	.016	1.855	1.791

24	24	.686	.152	.008	.696	.679
25	25	-1.359	.191	-.014	-1.386	-1.345
26	26	-.204	.152	-.002	-.205	-.202
27	27	.686	.152	.008	.696	.679
28	28	.252	.149	.003	.256	.249
29	29	-1.064	.176	-.011	-1.082	-1.054
30	30	1.133	.161	.011	1.155	1.122

ROOT MEAN SQUARE = .012

30 ITEMS CALIBRATED ON 196 PERSONS

196 MEASURABLE PERSONS WITH MEAN ABILITY = .22 AND STD. DEV. = .38

THAI 2530 (กลุ่มที่ 1) 30 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS	RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS
29	0	3.83	1.06	14	24	-.18	.41
28	0	3.07	.78	13	19	-.34	.41
27	0	2.58	.66	12	10	-.50	.41
26	0	2.22	.59	11	8	-.67	.42
25	0	1.91	.54	10	13	-.84	.43
24	1	1.65	.51	9	4	-1.02	.44
23	0	1.41	.48	8	1	-1.21	.45
22	3	1.20	.46	7	1	-1.42	.47
21	5	1.00	.45	6	0	-1.64	.50
20	13	.82	.43	5	0	-1.90	.53
19	23	.64	.42	4	0	-2.19	.58
18	22	.47	.42	3	0	-2.55	.65
17	20	.30	.41	2	0	-3.02	.77
16	28	.14	.41	1	0	-3.77	1.06
15	23	-.02	.41				

PERSON SEPARABILITY INDEX .39 (EQUIVALENT TO KR20)

30 ITEMS CALIBRATED ON 218 PERSONS

218 MEASURABLE PERSONS WITH MEAN ABILITY = .08 AND STD. DEV. = .53

THAI 2530 (๒๕๓๐) 30 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.06 ABILITY SCALE FACTOR 1.22

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-2.318	.243	-.006	-2.407	-2.312
2	2	-.972	.158	-.004	-.995	-.968
3	3	.135	.140	.000	.136	.134
4	4	-.529	.146	-.002	-.541	-.527
5	5	-.249	.142	-.001	-.255	-.248
6	6	-.347	.143	-.001	-.355	-.346
7	7	-.347	.143	-.001	-.355	-.346
8	8	-.190	.141	-.001	-.195	-.190
9	9	-.786	.152	-.003	-.805	-.783
10	10	-.268	.142	-.001	-.275	-.267
11	11	.307	.141	.001	.312	.306
12	12	.173	.140	.001	.175	.172
13	13	1.352	.166	.005	1.387	1.347
14	14	.288	.141	.001	.292	.287
15	15	-1.070	.161	-.004	-1.097	-1.066
16	16	-.591	.147	-.002	-.605	-.589
17	17	2.256	.223	.006	2.339	2.249
18	18	.561	.144	.002	.572	.559
19	19	2.305	.227	.007	2.391	2.298
20	20	-.152	.141	-.001	-.156	-.151
21	21	1.248	.162	.004	1.279	1.243
22	22	.541	.144	.002	.551	.540
23	23	.207	.142	.001	.215	.205

24	24	-.268	.142	-.001	-.275	-.267
25	25	-.288	.142	-.001	-.295	-.287
26	26	1.173	.159	.004	1.202	1.169
27	27	.135	.140	.000	.136	.134
28	28	-1.844	.203	-.005	-1.905	-1.899
29	29	2.162	.215	.006	2.240	2.156
30	30	-2.107	.224	-.005	-2.183	-2.102

ROOT MEAN SQUARE = .003

30 ITEMS CALIBRATED ON 218 PERSONS

218 MEASURABLE PERSONS WITH MEAN ABILITY = .08 AND STD. DEV. = .33

THAI 2531 (ฉบับที่ 2) 30 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS	RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS
29	0	3.57	1.05	14	7	-.15	.39
28	0	2.83	.76	13	11	-.30	.39
27	0	2.38	.64	12	11	-.45	.40
26	1	2.03	.57	11	2	-.61	.40
25	1	1.75	.52	10	1	-.77	.41
24	4	1.51	.48	9	1	-.94	.42
23	12	1.30	.46	8	1	-1.12	.44
22	9	1.11	.44	7	0	-1.31	.46
21	23	.93	.42	6	2	-1.52	.48
20	24	.76	.41	5	0	-1.75	.51
19	23	.60	.40	4	0	-2.03	.56
18	19	.44	.40	3	0	-2.36	.63
17	14	.29	.39	2	0	-2.81	.75
16	15	.14	.39	1	0	-3.54	1.04
15	15	-.01	.39				

PERSON SEPARABILITY INDEX .51 (EQUIVALENT TO KR20)

30 ITEMS CALIBRATED ON 196 PERSONS

196 MEASURABLE PERSONS WITH MEAN ABILITY = .44 AND STD. DEV. = .59

THAI 2531 (กลุ่๑ 2) 30 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.07 ABILITY SCALE FACTOR 1.12

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-.333	.159	-.001	-.339	-.332
2	2	.193	.150	.001	.198	.192
3	3	-1.694	.231	-.005	-1.754	-1.689
4	4	.408	.149	.001	.418	.407
5	5	-.432	.162	-.001	-.441	-.431
6	6	.408	.149	.001	.418	.407
7	7	-.407	.161	-.001	-.415	-.406
8	8	-.407	.161	-.001	-.415	-.406
9	9	.149	.150	.001	.153	.149
10	10	1.136	.156	.004	1.164	1.132
11	11	.193	.150	.001	.198	.192
12	12	-.483	.163	-.002	-.494	-.482
13	13	.731	.150	.002	.748	.728
14	14	1.330	.161	.004	1.364	1.325
15	15	2.217	.202	.006	2.295	2.211
16	16	-.728	.172	-.002	-.745	-.725
17	17	-1.071	.188	-.004	-1.101	-1.067
18	18	-.120	.154	.000	-.121	-.119
19	19	.952	.153	.003	.974	.949
20	20	.494	.149	.002	.505	.492
21	21	-.097	.154	.000	-.097	-.096
22	22	.494	.149	.002	.505	.492
23	23	-1.291	.201	-.004	-1.330	-1.286

24	24	.430	.149	.001	.440	.428
25	25	.408	.149	.001	.418	.407
26	26	.061	.151	.000	.064	.061
27	27	-.876	.178	-.003	-.898	-.873
28	28	-1.291	.201	-.004	-1.330	-1.286
29	29	-.458	.163	-.001	-.467	-.456
30	30	.083	.151	.000	.086	.083

ROOT MEAN SQUARE = .003

30 ITEMS CALIBRATED ON 196 PERSONS

196 MEASURABLE PERSONS WITH MEAN ABILITY = .44 AND STD. DEV. = .42

MATH 2529 (FORM 1) 38 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS	RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS
37	0	3.94	1.05	18	13	-.13	.35
36	0	3.20	.76	17	16	-.25	.35
35	0	2.74	.63	16	16	-.38	.35
34	0	2.40	.56	15	17	-.50	.36
33	0	2.12	.51	14	26	-.62	.36
32	0	1.89	.48	13	27	-.75	.37
31	0	1.68	.45	12	17	-.89	.37
30	0	1.49	.43	11	17	-1.03	.38
29	0	1.32	.41	10	18	-1.17	.39
28	0	1.16	.40	9	13	-1.33	.41
27	0	1.01	.39	8	8	-1.49	.42
26	0	.87	.38	7	3	-1.68	.44
25	0	.73	.37	6	4	-1.88	.47
24	1	.60	.36	5	0	-2.10	.50
23	1	.47	.36	4	0	-2.37	.55
22	1	.35	.36	3	0	-2.70	.63
21	7	.23	.35	2	0	-3.15	.75
20	5	.11	.35	1	0	-3.88	1.04
19	10	-.01	.35				

PERSON SEPARABILITY INDEX .46 (EQUIVALENT TO KR20)

38 ITEMS CALIBRATED ON 224 PERSONS

224 MEASURED PERSONS WITH MEAN ABILITY = -.66 AND STD. DEV. = .51

MATH 2529 (กลุ่มที่ 1) 38 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.06 ABILITY SCALE FACTOR 1.15

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-.551	.138	-.001	-.568	-.551
2	2	.961	.139	-.001	-.984	-.960
3	3	.048	.146	.000	.044	.048
4	4	-.133	.142	.000	-.142	-.133
5	5	1.372	.208	.001	1.417	1.371
6	6	.704	.168	.001	.719	.703
7	7	-.961	.139	-.001	-.984	-.960
8	8	-2.232	.178	-.001	-2.292	-2.230
9	9	1.104	.189	.001	1.136	1.103
10	10	2.550	.342	.002	2.669	2.548
11	11	.153	.148	.000	.152	.153
12	12	-.477	.138	-.001	-.492	-.476
13	13	-.345	.139	.000	-.358	-.345
14	14	-.533	.138	-.001	-.549	-.532
15	15	.028	.145	.000	.023	.028
16	16	1.602	.227	.002	1.660	1.600
17	17	.028	.145	.000	.023	.028
18	18	1.763	.242	.002	1.830	1.761
19	19	-.402	.139	.000	-.415	-.401
20	20	.424	.156	.000	.429	.423
21	21	-.364	.139	.000	-.377	-.364
22	22	.907	.178	.001	.930	.906
23	23	-1.655	.153	-.001	-1.695	-1.654

24	24	-.551	.138	-.001	-.568	-.551
25	25	-.054	.144	.000	-.060	-.054
26	26	-.211	.141	.000	-.221	-.211
27	27	.623	.164	.001	.636	.623
28	28	.546	.161	.001	.556	.546
29	29	-.033	.144	.000	-.040	-.033
30	30	-.942	.139	-.001	-.965	-.941
31	31	-.153	.142	.000	-.162	-.153
32	32	-.211	.141	.000	-.221	-.211
33	33	-.755	.138	-.001	-.775	-.754
34	34	-1.373	.146	-.001	-1.405	-1.372
35	35	.448	.157	.000	.454	.447
36	36	-.231	.141	.000	-.241	-.230
37	37	.496	.159	.000	.504	.496
38	38	.330	.153	.000	.333	.330

ROOT MEAN SQUARE = .001

38 ITEMS CALIBRATED ON 224 PERSONS

224 MEASURABLE PERSONS WITH MEAN ABILITY = -.66 AND STD. DEV. = .35

MATH 2529 (PART 2) 38 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS	RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS
37	0	3.88	1.04	18	8	-.13	.35
36	0	3.15	.75	17	14	-.25	.35
35	0	2.70	.63	16	17	-.37	.35
34	0	2.36	.56	15	14	-.49	.35
33	0	2.09	.51	14	16	-.62	.36
32	0	1.86	.47	13	17	-.74	.36
31	0	1.65	.45	12	19	-.88	.37
30	0	1.47	.43	11	23	-1.01	.38
29	0	1.30	.41	10	16	-1.16	.39
28	0	1.14	.40	9	11	-1.31	.40
27	0	.99	.38	8	15	-1.47	.42
26	0	.85	.38	7	6	-1.65	.44
25	1	.72	.37	6	5	-1.85	.46
24	0	.59	.36	5	1	-2.07	.50
23	2	.46	.36	4	2	-2.33	.55
22	2	.34	.35	3	0	-2.66	.62
21	1	.22	.35	2	0	-3.10	.74
20	2	.10	.35	1	0	-3.82	1.03
19	5	-.02	.35				

PERSON SEPARABILITY INDEX .47 (EQUIVALENT TO KR20)

38 ITEMS CALIBRATED ON 197 PERSONS

197 MEASURABLE PERSONS WITH MEAN ABILITY = -.80 AND STD. DEV. = .53

MATH 2529 (กลุ่มที่ 2) 38 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.06 ABILITY SCALE FACTOR 1.13

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-.258	.152	.000	-.272	-.258
2	2	-.714	.148	-.001	-.738	-.714
3	3	-.121	.155	.000	-.130	-.120
4	4	.277	.165	.000	.288	.227
5	5	2.069	.313	.001	2.171	2.068
6	6	.450	.174	.000	.459	.450
7	7	-.969	.148	-.001	-.997	-.968
8	8	-1.769	.163	.000	-1.816	-1.768
9	9	.510	.177	.000	.521	.510
10	10	1.978	.300	.001	2.073	1.977
11	11	.636	.183	.000	.652	.636
12	12	-.144	.155	.000	-.154	-.144
13	13	.604	.181	.000	.619	.603
14	14	.421	.173	.000	.429	.421
15	15	-.167	.154	.000	-.178	-.167
16	16	1.604	.257	.002	1.673	1.602
17	17	-.213	.153	.000	-.225	-.213
18	18	1.482	.244	.001	1.543	1.480
19	19	-.370	.150	.000	-.386	-.370
20	20	.572	.180	.000	.585	.572
21	21	.281	.167	.000	.283	.280
22	22	.807	.192	.000	.831	.807
23	23	-1.668	.160	.000	-1.712	-1.667

24	24	-.629	.148	.000	-.651	-.629
25	25	-.544	.149	.000	-.564	-.543
26	26	.098	.161	.000	-.094	.098
27	27	-.281	.152	.000	-.295	-.281
28	28	.023	.159	.000	.017	.023
29	29	-.565	.148	.000	-.585	-.565
30	30	-.735	.147	-.001	-.759	-.735
31	31	-.370	.150	.000	-.386	-.370
32	32	-.693	.148	.000	-.716	-.693
33	33	-.799	.147	-.001	-.824	-.798
34	34	-1.249	.151	-.001	-1.284	-1.249
35	35	.572	.180	.000	.585	.572
36	36	.048	.159	.000	.043	.048
37	37	.227	.165	.000	.228	.227
38	38	-.348	.151	.000	-.363	-.348

ROOT MEAN SQUARE = .001

38 ITEMS CALIBRATED ON 197 PERSONS

197 MEASURABLE PERSONS WITH MEAN ABILITY = -.80 AND STD. DEV. = .37

MATH 2530 (กล้วยน้ำ 1) 38 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS	RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS
37	0	3.91	1.04	18	4	-.14	.35
36	0	3.18	.75	17	9	-.26	.36
35	0	2.74	.63	16	11	-.39	.36
34	0	2.40	.56	15	16	-.51	.36
33	0	2.13	.51	14	28	-.64	.36
32	0	1.90	.47	13	22	-.77	.37
31	0	1.69	.45	12	30	-.90	.38
30	0	1.51	.43	11	23	-1.04	.38
29	0	1.34	.41	10	20	-1.19	.39
28	0	1.18	.40	9	23	-1.34	.41
27	0	1.03	.39	8	15	-1.51	.42
26	0	.88	.38	7	9	-1.69	.44
25	1	.74	.37	6	5	-1.88	.47
24	0	.61	.37	5	5	-2.11	.50
23	0	.48	.36	4	1	-2.37	.55
22	0	.35	.36	3	0	-2.69	.62
21	1	.23	.36	2	0	-3.13	.74
20	0	.10	.36	1	0	-3.85	1.03
19	2	-.02	.35				

PERSON SEPARABILITY INDEX .29 (EQUIVALENT TO KR20)

38 ITEMS CALIBRATED ON 224 PERSONS

224 MEASURABLE PERSONS WITH MEAN ABILITY = -.96 AND STD. DEV. = .46

MATH 2530 (PART 1) 38 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.05 ABILITY SCALE FACTOR 1.15

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	-.227	.146	.000	-.241	-.226
2	2	-1.135	.138	.000	-1.170	-1.135
3	3	-.543	.140	.000	-.565	-.543
4	4	-.164	.147	.000	-.176	-.164
5	5	1.430	.242	.001	1.488	1.429
6	6	.380	.167	.000	.386	.380
7	7	-.842	.137	.000	-.870	-.841
8	8	-1.135	.138	.000	-1.170	-1.135
9	9	-.268	.145	.000	-.283	-.267
10	10	-.768	.138	.000	-.795	-.768
11	11	-1.360	.140	.000	-1.399	-1.359
12	12	-.842	.137	.000	-.870	-.841
13	13	.750	.138	.000	-.776	-.749
14	14	-.034	.151	.000	-.042	-.034
15	15	-1.062	.137	.000	-1.095	-1.061
16	16	.463	.171	.000	.473	.463
17	17	1.374	.236	.001	1.429	1.374
18	18	-1.154	.138	.000	-1.189	-1.153
19	19	-.712	.138	.000	-.738	-.712
20	20	1.549	.254	.001	1.615	1.549
21	21	1.270	.226	.000	1.319	1.270
22	22	1.430	.242	.001	1.488	1.429
23	23	.056	.154	.000	.051	.056
24	24	.249	.161	.000	.251	.249
25	25	.643	.181	.000	.661	.643

24	24	.249	.161	.000	.251	.249
25	25	.643	.181	.000	.661	.643
26	26	-.933	.137	.000	-.964	-.933
27	27	2.007	.311	.001	2.100	2.006
28	28	-.308	.144	.000	-.325	-.308
29	29	.776	.189	.000	.800	.776
30	30	.353	.166	.000	.359	.353
31	31	1.221	.222	.000	1.268	1.221
32	32	.811	.191	.000	.836	.811
33	33	.080	.155	.000	.075	.079
34	34	-1.117	.138	.000	-1.151	-1.117
35	35	.151	.157	.000	.148	.150
36	36	.249	.161	.000	.251	.249
37	37	.127	.156	.000	.124	.126
38	38	-1.265	.139	.000	-1.302	-1.265

ROOT MEAN SQUARE = .000

38 ITEMS CALIBRATED ON 224 PERSONS

224 MEASURABLE PERSONS WITH MEAN ABILITY = -.96 AND STD. DEV. = .25

MATH 2531 (กลุ่มที่ 2) 38 ITEMS 4 CHOICES.

COMPLETE SCORE EQUIVALENCE TABLE

RAW		LOG	STANDARD	RAW		LOG	STANDARD
SCORE	COUNT	ABILITY	ERRORS	SCORE	COUNT	ABILITY	ERRORS
37	0	3.75	1.03	18	10	-.11	.34
36	0	3.03	.74	17	7	-.22	.34
35	0	2.59	.62	16	17	-.34	.35
34	0	2.27	.55	15	14	-.46	.35
33	0	2.01	.50	14	24	-.58	.35
32	0	1.79	.46	13	21	-.70	.36
31	0	1.59	.44	12	20	-.83	.37
30	0	1.42	.42	11	17	-.96	.38
29	1	1.26	.40	10	11	-1.10	.39
28	0	1.11	.39	9	5	-1.25	.40
27	0	.97	.37	8	6	-1.42	.42
26	1	.83	.37	7	7	-1.59	.44
25	0	.71	.36	6	7	-1.79	.46
24	1	.58	.35	5	0	-2.01	.50
23	2	.46	.35	4	2	-2.28	.55
22	3	.35	.35	3	0	-2.60	.62
21	10	.23	.34	2	0	-3.05	.75
20	3	.12	.34	1	0	-3.77	1.03
19	8	.00	.34				

PERSON SEPARABILITY INDEX .57 (EQUIVALENT TO KR20)

38 ITEMS CALIBRATED ON 197 PERSONS

197 MEASURABLE PERSONS WITH MEAN ABILITY = -.63 AND STD. DEV. = .57

MATH 2531 (กลุ่^๑ม 2) 38 ITEMS 4 CHOICES.

PROCEDURE IS UCON

DIFFICULTY SCALE FACTOR 1.07 ABILITY SCALE FACTOR 1.09

NUMBER OF ITERATIONS = 2

SEQUENCE NUMBER	ITEM NAME	ITEM DIFFICULTY	STANDARD ERROR	LAST DIFF CHANGE	PROX DIFF	FIRST CYCLE
1	1	.038	.155	.000	.035	.038
2	2	-1.187	.153	-.002	-1.216	-1.185
3	3	-.321	.149	.000	-.332	-.321
4	4	-.055	.153	.000	-.059	-.055
5	5	.061	.156	.000	.060	.061
6	6	.109	.157	.000	.108	.109
7	7	-.429	.148	-.001	-.442	-.429
8	8	-.278	.150	.000	-.288	-.277
9	9	-2.057	.182	-.003	-2.111	-2.054
10	10	-1.030	.150	-.002	-1.056	-1.029
11	11	.924	.189	.002	.952	.922
12	12	-.190	.151	.000	-.198	-.189
13	13	-.408	.149	-.001	-.420	-.407
14	14	.038	.155	.000	.035	.038
15	15	-.429	.148	-.001	-.442	-.429
16	16	.231	.160	.000	.234	.230
17	17	.606	.173	.001	.622	.605
18	18	.758	.180	.001	.780	.757
19	19	-.643	.148	-.001	-.660	-.642
20	20	1.796	.257	.002	1.874	1.794
21	21	.157	.158	.000	.158	.157
22	22	-.472	.148	-.001	-.486	-.471
23	23	-1.446	.159	-.002	-1.481	-1.444

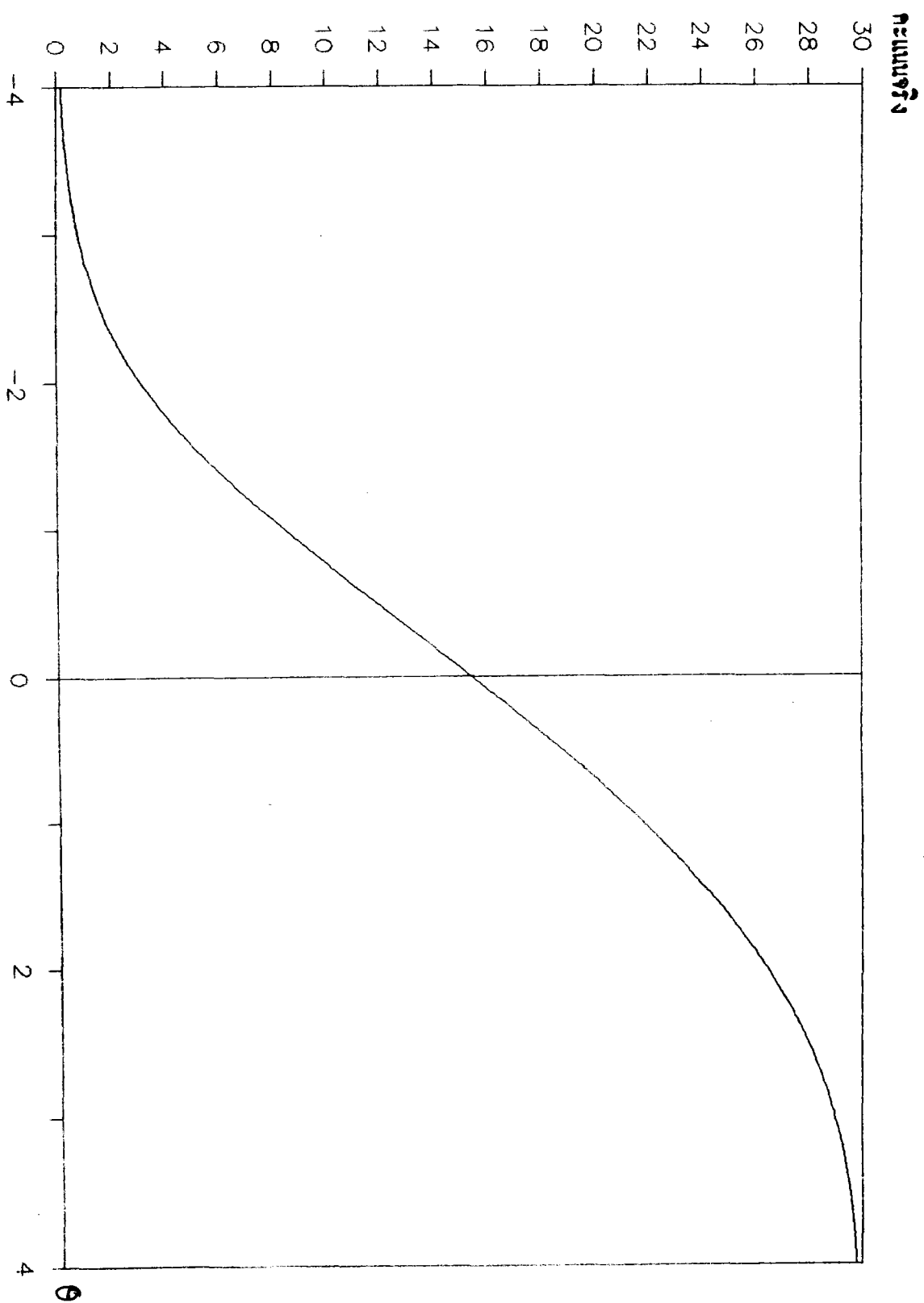
24	24	.306	.162	.000	.312	.306
25	25	.181	.159	.000	.183	.181
26	26	-.145	.152	.000	-.152	-.145
27	27	.411	.166	.001	.420	.410
28	28	.157	.158	.000	.158	.157
29	29	-.408	.149	-.001	-.420	.407
30	30	.332	.163	.001	.338	.332
31	31	-.167	.151	.000	-.175	-.167
32	32	.855	.185	.001	.881	.854
33	33	1.673	.245	.002	1.743	1.671
34	34	.384	.165	.001	.392	.384
35	35	.520	.170	.001	.533	.520
36	36	.281	.161	.000	.285	.280
37	37	-.145	.152	.000	-.152	-.145
38	38	-.009	.154	.000	-.012	-.009

ROOT MEAN SQUARE = .001

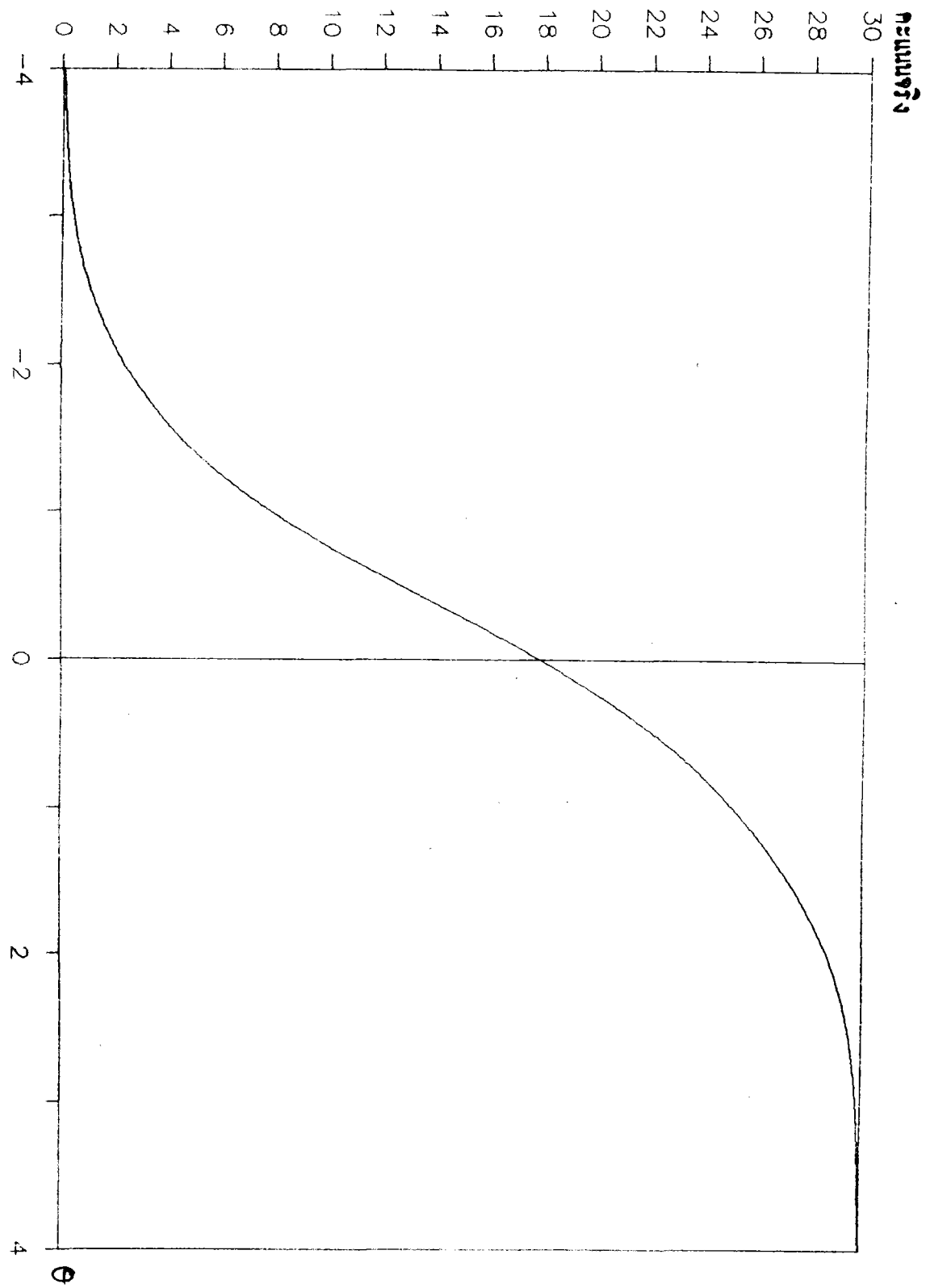
38 ITEMS CALIBRATED ON 197 PERSONS

197 MEASURABLE PERSONS WITH MEAN ABILITY = -.63 AND STD. DEV. = .43

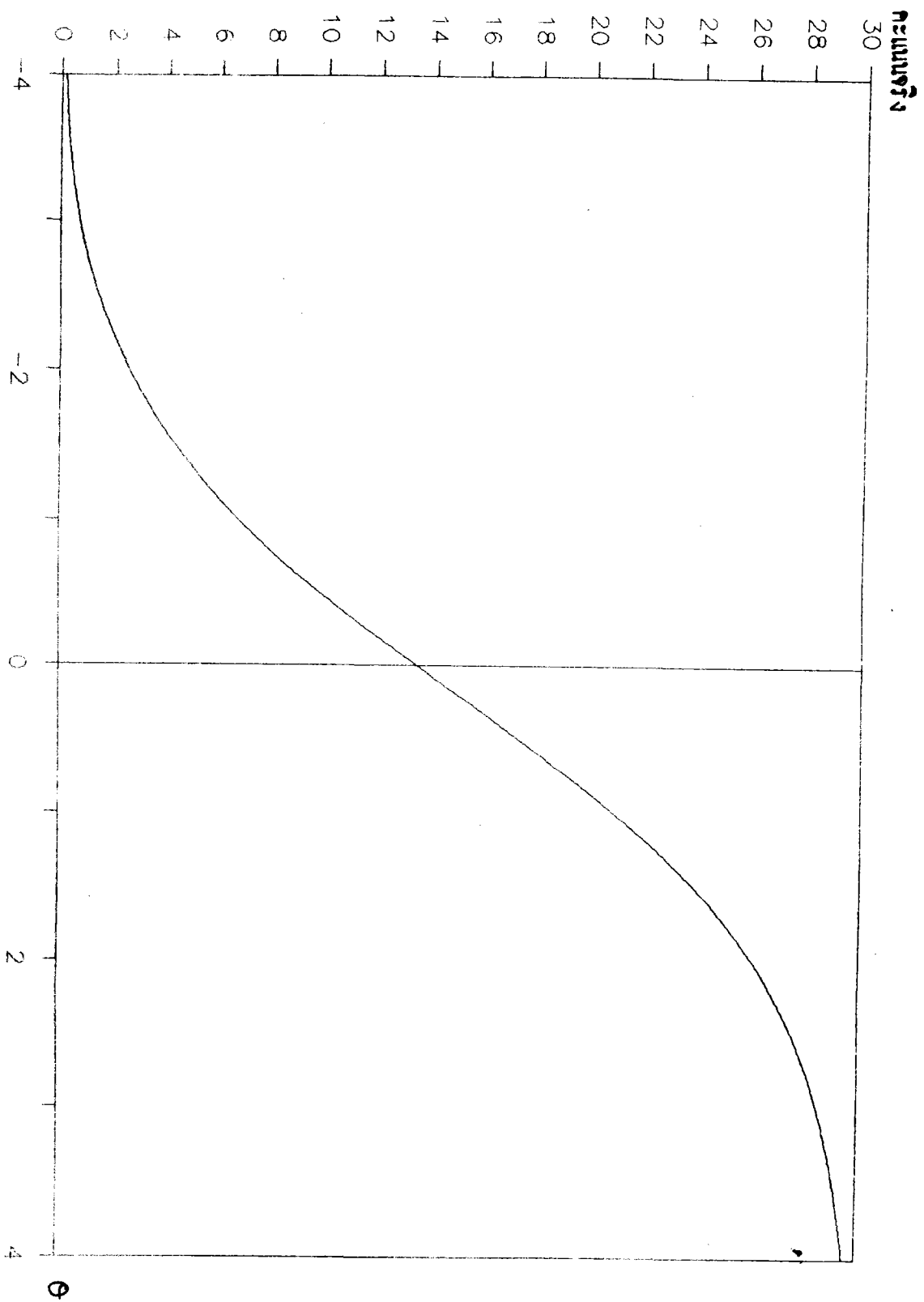
ภาพที่ 4 แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ ในกลุ่มทักษะวิชาภาษาไทย
เมื่อปีการศึกษา 2529



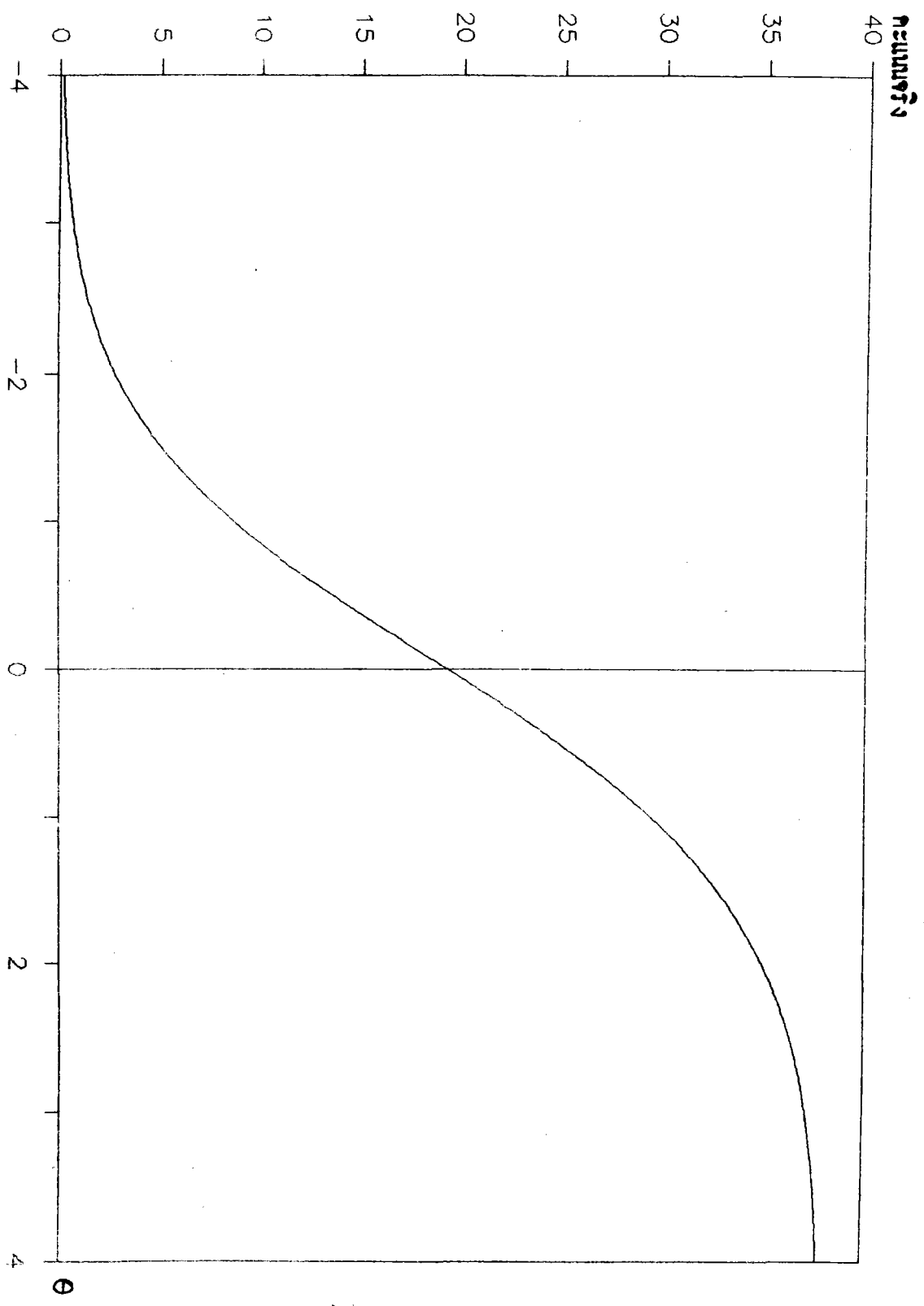
ภาพที่ 5 แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ
เมื่อปีการศึกษา 2530



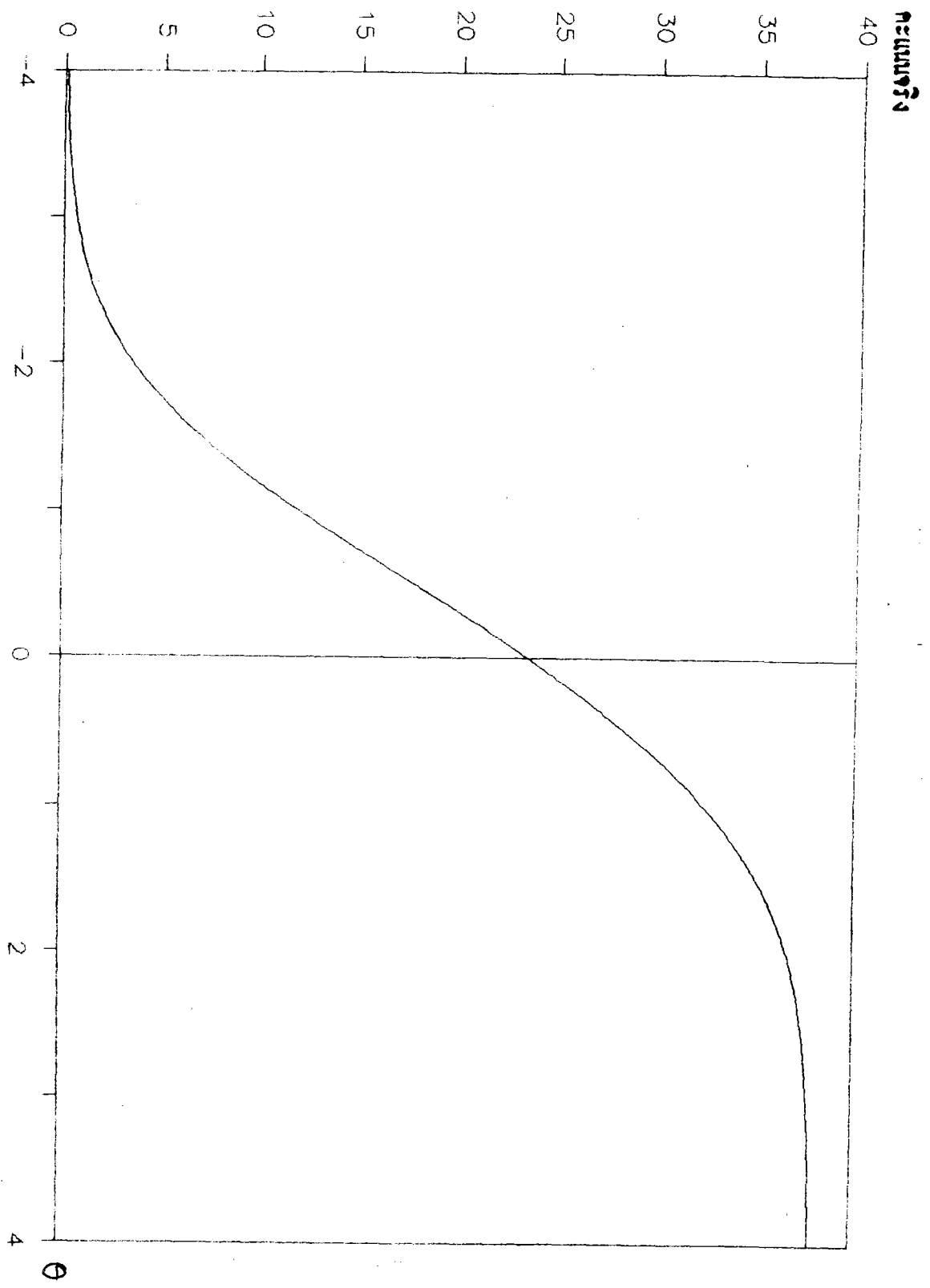
ภาพที่ 6 แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ
เมื่อปีการศึกษา 2531



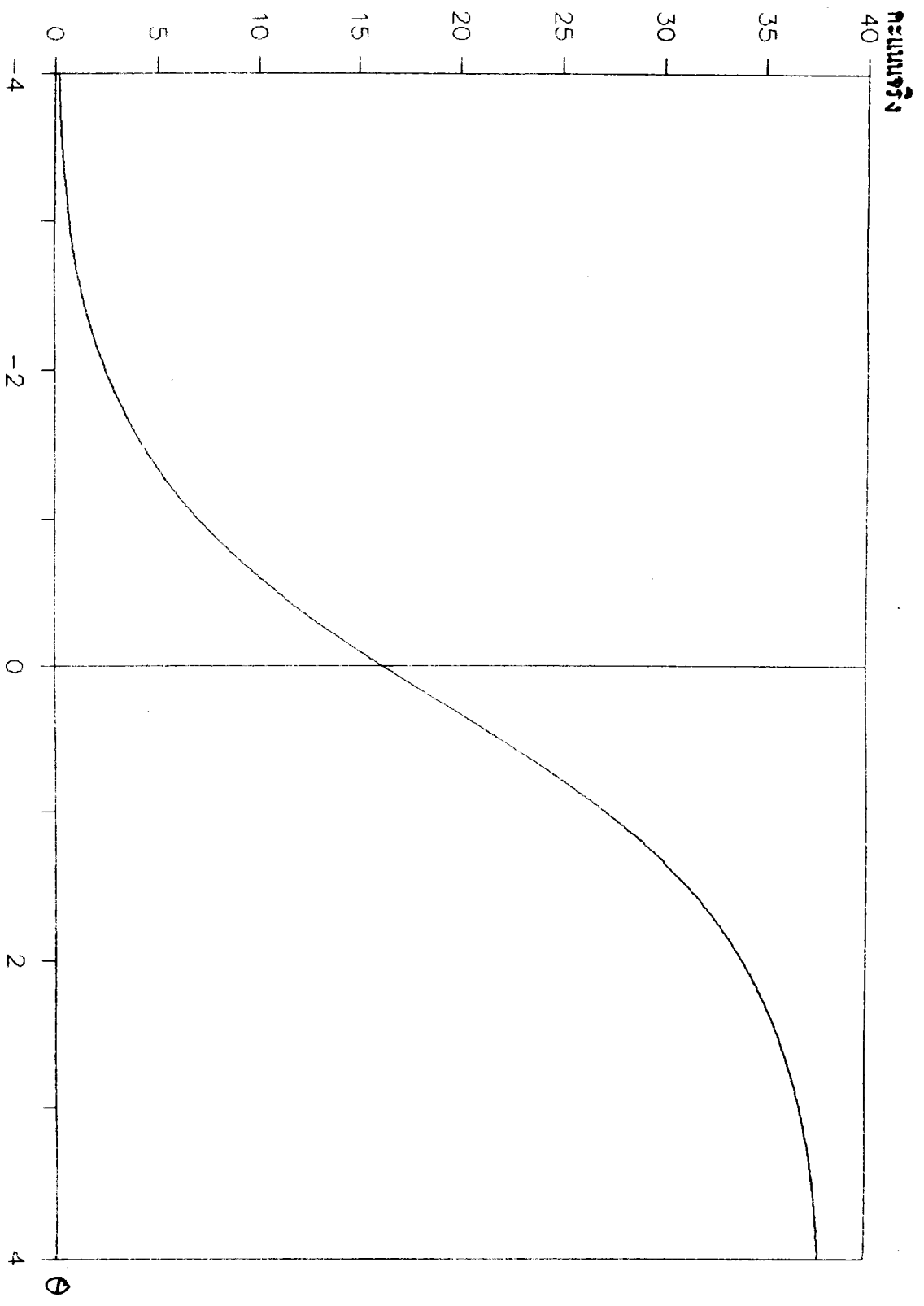
ภาพที่ 7 แสดงการกะแผนจริงของผู้สอบแต่ละระดับกับความสามารถที่ได้จากการสอบแบบสอบ ในกลุ่มทักษะวิชาการคณิตศาสตร์
เมื่อปีการศึกษา 2529



ภาพที่ 8 แสดงค่าคะแนนจริงของข้อสอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ
เมื่อปีการศึกษา 2530



ภาพที่ 9 แสดงค่าคะแนนจริงของผู้สอบแต่ละระดับความสามารถที่ได้จากการสอบแบบสอบ
เมื่อปีการศึกษา 2531



ตาราง 31 การเทียบคะแนนจริงในกลุ่มทักษะวิชาภาษาไทย ตั้งแต่ปีการศึกษา 2529 ถึง 2531

คะแนนดิบ	ปีการศึกษา 2529	ปีการศึกษา 2530	ปีการศึกษา 2531
29	30	30	30
28	29	30	30
27	29	30	29
26	28	29	29
25	27	29	28
24	26	28	27
23	25	28	26
22	24	27	25
21	22	26	23
20	21	25	22
19	20	24	21
18	19	22	19
17	17	21	17
16	16	19	16
15	15	18	14
14	13	16	13
13	12	14	11
12	11	13	9
11	10	11	8
10	9	9	7

ตาราง 31 (ต่อ)

คะแนนดิบ	ปีการศึกษา 2529	ปีการศึกษา 2530	ปีการศึกษา 2531
9	8	7	5
8	7	6	4
7	5	5	3
6	4	3	2
5	3	2	2
4	2	1	1
3	1	1	1
2	1	0	0
1	0	0	0

ตาราง 32 เทียบคะแนนจริงในกลุ่มทักษะวิชาคณิตศาสตร์ ตั้งแต่ปีการศึกษา 2529 ถึง 2531

คะแนนดิบ	ปีการศึกษา 2529	ปีการศึกษา 2530	ปีการศึกษา 2531
37	38	38	38
36	37	38	38
35	37	38	36
34	36	37	36
33	36	37	35
32	35	36	34
31	34	36	33
30	33	35	32
29	32	35	31
28	31	34	30
27	30	33	28
26	29	32	27
25	27	31	26
24	26	30	25
23	25	29	23
22	23	28	22
21	22	27	20
20	21	25	19

ตาราง 32 (ต่อ)

คะแนนดิบ	ปีการศึกษา 2529	ปีการศึกษา 2530	ปีการศึกษา 2531
19	19	24	18
18	18	22	16
17	16	21	15
16	15	19	14
15	13	18	12
14	12	16	11
13	11	14	10
12	9	13	9
11	8	11	8
10	7	10	7
9	6	8	6
8	5	7	5
7	4	5	4
6	3	4	3
5	2	3	2
4	2	2	2
3	1	1	1
2	0	1	1
1	0	0	0

ประวัติย่อของผู้วิจัย

ชื่อ สุจินดา ชื่อสกุล ผ่องอักษร

เกิดปี พ.ศ. 2497

สถานที่เกิด จังหวัดพระนครศรีอยุธยา

สถานที่อยู่ปัจจุบัน 512/139 หมู่ 2 ตำบลประชาธิปไตย อำเภอดอนบุรี จังหวัดพทุมธานี

ตำแหน่งหน้าที่

พ.ศ. 2517 ครูตรีโรงเรียนวัดไม้ตราสมาธิาราม อำเภอบางไทร
จังหวัดพระนครศรีอยุธยา

พ.ศ. 2525 อาจารย์ 1 โรงเรียนวัดไม้ตราสมาธิาราม

พ.ศ. 2526 อาจารย์ 1 โรงเรียนวัดไม้ตราสมาธิาราม ช่วยราชการกองวิชาการ
สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ กระทรวงศึกษาธิการ

พ.ศ. 2528 นักวิชาการศึกษา กองวิชาการ สำนักงานคณะกรรมการการประถมศึกษา
แห่งชาติ กระทรวงศึกษาธิการ

ประวัติการศึกษา

พ.ศ. 2517 ป.กศ. สูง จากวิทยาลัยครูพระนครศรีอยุธยา

พ.ศ. 2518 พ.ม.

พ.ศ. 2522 กศ.บ. (คณิตศาสตร์) จากมหาวิทยาลัยศรีนครินทรวิโรฒ บางเขน

พ.ศ. 2526 กศ.ม. (การวัดผลการศึกษา) จาก มหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร

พ.ศ. 2534 กศ.ด. (การวิจัยและพัฒนาหลักสูตร) จากมหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร

การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6
ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา
(ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบคะแนนรูปแบบราสส์

บทคัดย่อ

ของ

สุจินดา ผ่องอักษร

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาคุณวุฒิปดพิท สาขาการวิจัยและพัฒนาหลักสูตร

ธันวาคม 2533

การศึกษาค้นคว้าครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะ (วิชาคณิตศาสตร์และภาษาไทย) ของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบคะแนนรูปแบบราสซัส กลุ่มตัวอย่างที่ใช้ศึกษาความก้าวหน้าคือนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบหลักสูตรประถมศึกษา พุทธศักราช 2521 เมื่อปีการศึกษา 2529 ถึง 2531 ในจังหวัดปทุมธานี จำนวน 4 อำเภอ จากอำเภอทั้งหมด 7 อำเภอ รวมทั้งสิ้น 1,454 คน ส่วนกลุ่มตัวอย่างที่ใช้ศึกษาค่าคงที่ในการเทียบคะแนนในกรณีที่ใช้ผู้สอบร่วมคือนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบหลักสูตรประถมศึกษา พุทธศักราช 2521 เมื่อปีการศึกษา 2532 ที่สุ่มอย่างง่ายจากนักเรียนในอำเภอธัญบุรี จังหวัดปทุมธานี จำนวน 581 คน ผู้วิจัยทำการเทียบคะแนนโดยหาค่าคงที่เพื่อใช้ในสมการเชิงเส้นตรง สำหรับปรับคะแนนที่ได้จากแบบสอบต่างชุดในแต่ละกลุ่มวิชาให้อยู่ในสเกลเดียวกันทั้งหมด โดยใช้สเกลคะแนนผลการสอบเมื่อปีการศึกษา 2529 เป็นหลัก และทำการเทียบคะแนนในปีการศึกษา 2530 และ 2531 ไปสู่สเกลเดียวกับสเกลคะแนนในปีการศึกษา 2529 โดยใช้วิธีเทียบคะแนนรูปแบบราสซัส หลังจากนั้นจึงนำคะแนนที่สมมูลกันระหว่างแบบสอบต่างชุดมาเปรียบเทียบกันโดยตรง

ผลการศึกษาพบว่าหลังจากการเทียบคะแนนแล้ว คะแนนดิบที่เท่ากันระหว่างแบบสอบต่างชุดมีค่าแตกต่างกันโดยในกลุ่มทักษะวิชาภาษาไทยมีความแตกต่างกันสูงสุด 3.8 คะแนน ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์มีความแตกต่างกันสูงสุด 6.3 คะแนน จุดตัดในแต่ละปีการศึกษาของทั้งสามปีการศึกษามีค่าแตกต่างกัน กล่าวคือจุดตัดในกลุ่มทักษะวิชาภาษาไทยตั้งแต่ปีการศึกษา 2529 ถึง 2531 มีค่าเท่ากับ 15, 18 และ 14 ตามลำดับ ส่วนจุดตัดในกลุ่มทักษะวิชาคณิตศาสตร์มีค่าเท่ากับ 19, 24 และ 18 ตามลำดับ ดังนั้นการศึกษาค้นคว้าความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในแต่ละปีการศึกษาและการตัดสินผลการสอบในแต่ละปีการศึกษา เมื่อก่อนเทียบคะแนนและหลังการเทียบคะแนนจึงให้ผลแตกต่างกัน

ผลการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนพบว่า ในกลุ่มทักษะวิชาภาษาไทย ส่วนใหญ่นักเรียนมีผลสัมฤทธิ์ทางการเรียนสูงกว่าระดับที่น่าพอใจและมีผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิม

ตามลำดับ ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์พบว่าส่วนใหญ่มีผลสัมฤทธิ์ทางการเรียนต่ำกว่าระดับที่น่าพอใจ และมีผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิมเฉพาะในปีการศึกษา 2530 แต่ในปีการศึกษา 2531 พบว่ามีผลสัมฤทธิ์ลดลงจากเดิม อย่างไรก็ตามผลสัมฤทธิ์ที่ลดลงนี้ยังมีค่าสูงกว่าผลสัมฤทธิ์ทางการเรียนเมื่อปีการศึกษา 2529 เพียงเล็กน้อย ซึ่งความแตกต่างที่เพิ่มขึ้นและลดลงของผลสัมฤทธิ์ทางการเรียนนี้ พบว่ามีความแตกต่างจากปีการศึกษาก่อนอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ผลการเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจระหว่างก่อนเทียบคะแนน และหลังจากการเทียบคะแนน พบว่ามีความแตกต่างกัน กล่าวคือในปีการศึกษา 2530 พบว่า กลุ่มทักษะวิชาภาษาไทยมีนักเรียนร้อยละ 10 และในกลุ่มทักษะวิชาคณิตศาสตร์พบว่ามีนักเรียนร้อยละ 20 ที่ควรเป็นผู้สอบผ่านแต่ได้รับการตัดสินให้เป็นผู้สอบไม่ผ่าน ส่วนในปีการศึกษา 2531 พบว่าในกลุ่มทักษะวิชาภาษาไทยมีนักเรียน ร้อยละ 7 และในกลุ่มทักษะวิชาคณิตศาสตร์พบว่ามีนักเรียน ร้อยละ 5 ที่ควรเป็นผู้สอบไม่ผ่านแต่ได้รับการตัดสินให้ผ่าน

The Progressive Study of Achievment in Basic Skills of Prathom VI Students
Who Finished Primary Curriculum 2521 B.E. During the Three Academic
Years (2529 - 2531) by Using Rasch Model Equating

BY

SUJINDA PHONGAKSORN

Presented in partial fulfillment of the requirements for the Doctor
of Education degree in Curriculum research and development
at Srinakharinwirot University

January 1991

The objective of the research was to study the progress of academic achievement in the skill subject area (mathematics and Thai language) of Prathom VI pupils who completed the Primary Curriculum B.E. 2521 (1978) during the three academic years (1986 - 1988) by using the Rasch model equating. The samples for studying the progress of academic achievement in the skill subject area were 1454 Prathom VI pupils who completed the Primary Curriculum B.E. 2521, from 4 out of the 7 districts in Pratoomthani Province. The two common person groups for studying the constant value of true score equating were 581 pupils simply choosed at random from other Prathom VI pupils who completed primary education in the 1989 academic year. Each group took the two tests ; the 1986 test and either 1987 or the 1988 test.

The researcher used the constant value as the linear equation to transform the estimated parameter ability of the 1987 and 1988 tests to be the same scale based on that of the 1986 tests. The true scores of various tests were estimated by using Rasch model. After that, the true scores of the tests in the three academic years were directly compared.

The result of the study on the same raw-score from different tests was that they had true scores differed, the most difference of the true score in Thai subject was 3.8, and the most difference of the true score in mathematics was 6.3. The cut-off true score in each of the three academic year is different. The cut-off true score in Thai language during the 1986 - 1988 academic year was 15, 18 and 14 respectively,

and that of mathematics was 19, 24 and 18 respectively. Therefore, the study on the academic achievement, the progress of academic achievement and the percentage of pupils who had reached the satisfactory level in each academic true-score equating were different.

From the study on the progress of academic achievement in the skill subject area, it was found that the majority of pupils' academic achievement in Thai Subject had reached the satisfactory level and it was progressed in each academic year. Regarding to academic achievement in mathematics, the majority of pupils' achievement had not reached the satisfactory level. That in the two academic years was however than that of the 1986 academic year. Although the mathematics academic achievement of the 1988 academic year was decreased from that of the 1987 academic year, it was still a little higher than that of the 1986 academic year. It was found that increasing and decreasing of academic achievement in the skill subject area was significantly differed from the previous academic year at .01 level.

In comparing percentage of pupils who had reached the satisfactory level before and after equating the true scores, it was found that there was statistically significant at .01 level, e.g., in the 1987 academic year, 10 percent of pupils in the Thai language and 20 percent of pupils in the mathematics who should have passed the test had not passed the test ; in the 1988 academic year 7 percent of pupils in the Thai language and 5 percent of pupils in the mathematics subject who should not have passed the test has passed the test.