



การปรับเทียบคะแนน O-NET ระหว่างปี โดยวิธีการเทียบเป็นมาตราเดียวกัน  
และโดยรูปแบบ IRT

นักวิจัย

สำราญ มีแจ่ม

ประภัสสร วงษ์ดี

ยุพิน โถงทา

งานวิจัยนี้ได้รับทุนสนับสนุนจากสถาบันทดสอบทางการศึกษาแห่งชาติ  
(องค์การมหาชน)

กันยายน 2552

## สารบัญ

### บทที่ 1 บทนำ

|                            |   |
|----------------------------|---|
| ความเป็นมา                 | 1 |
| คำถามการวิจัย              | 5 |
| วัตถุประสงค์ของการวิจัย    | 5 |
| สมมติฐานการวิจัย           | 5 |
| ขอบเขตการวิจัย             | 6 |
| นิยามศัพท์ที่ใช้ในการวิจัย | 6 |
| ประโยชน์ที่คาดว่าจะได้รับ  | 7 |
| ข้อจำกัดของการวิจัย        | 8 |

### บทที่ 2 เอกสารและงานวิจัยที่เกี่ยวข้อง

|  |    |
|--|----|
| ตอนที่ 1 แนวคิดเกี่ยวกับการเปรียบเทียบคะแนน          |    |
| 1.1 ความหมายของการเปรียบเทียบคะแนน                   | 9  |
| 1.2 เงื่อนไขของการเปรียบเทียบคะแนน                   | 10 |
| 1.3 ประเภทของการเปรียบเทียบคะแนน                     | 11 |
| 1.4 รูปแบบวิธีการเก็บข้อมูล                          | 13 |
| 1.5 รูปแบบของการเปรียบเทียบคะแนน                     | 17 |
| ตอนที่ 2 การหาคุณภาพของการเปรียบเทียบคะแนน           |    |
| 2.1 ความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบคะแนน     | 35 |
| 2.2 การประเมินความเพียงพอของการเปรียบเทียบคะแนน      | 37 |
| ตอนที่ 3 งานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบคะแนน |    |
| 3.1 งานวิจัยต่างประเทศ                               | 40 |
| 3.2 งานวิจัยในประเทศ                                 | 46 |

### บทที่ 3 วิธีดำเนินการวิจัย

|                         |    |
|-------------------------|----|
| ประชากร                 | 62 |
| กลุ่มตัวอย่าง           | 62 |
| วิธีการเปรียบเทียบคะแนน | 63 |
| การวิเคราะห์ข้อมูล      | 65 |

## สารบัญ (ต่อ)

|   | หน้า |
|---|------|
| <b>บทที่ 4 ผลการวิเคราะห์ข้อมูล</b>   |      |
| ตอนที่ 1 ลักษณะของคะแนน O-NET ในปี 2550 และ ปี 2551 ในแต่ละรายวิชา  | 70   |
| ตอนที่ 2 ลักษณะของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เป็นคะแนนเดิมและคะแนนแปลงโดยวิธีการเทียบให้เป็นมาตราเดียวกัน  | 73   |
| ตอนที่ 3 การเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ระหว่างคะแนนเดิมกับคะแนนแปลงโดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน   | 76   |
| ตอนที่ 4 การเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในปีเดียวกันที่เป็นคะแนนเดิมกับคะแนนแปลงที่ได้จากการปรับเทียบคะแนนมาจากคนละปีโดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน  | 77   |
| ตอนที่ 5 ผลการเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ตามรูปแบบการปรับเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ (Equating Through Item Response Theory หรือ IRT) โดยใช้โมเดล 3 พารามิเตอร์ (รูปแบบ IRT)   | 78   |
| ตอนที่ 6 ผลการเปรียบเทียบคุณภาพของวิธีการแปลงคะแนน O-NET โดยการปรับเทียบให้เป็นมาตราเดียวกัน และปรับเทียบตามรูปแบบ IRT โดยการศึกษาความสัมพันธ์ระหว่างคะแนนสอบ O-NET เดิมกับคะแนน O-NET ที่ได้จากการปรับเทียบให้เป็นมาตราเดียวกัน และที่ได้จากการปรับเทียบด้วยรูปแบบ IRT | 84   |
| <b>บทที่ 5 สรุป อภิปรายผล และข้อเสนอแนะ</b>   |      |
| สรุปผลการวิจัย  | 89   |
| อภิปรายผล   | 93   |
| ข้อเสนอแนะ  | 94   |
| บรรณานุกรม  | 96   |
| ภาคผนวก   |      |
| ประวัติผู้วิจัย   | 103  |

## สารบัญตาราง

| ตารางที่ |   | หน้า |
|----------|---|------|
| 4.1      | แสดงการแจกแจงผลของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2550 และ 2551 จำแนกตามรายวิชา และเป็นกลุ่มตัวอย่างในการนำคะแนนมาเปรียบเทียบ                                       | 71   |
| 4.2      | แสดงการแจกแจงผลของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2550 ที่เป็นคะแนนเดิมและคะแนนแปลงโดยการปรับเทียบให้เป็นมาตรฐานเดียวกัน จำแนกตามรายวิชา                            | 73   |
| 4.3      | แสดงการแจกแจงผลของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2551 ก่อนการปรับเทียบและคะแนนแปลงโดยการปรับเทียบให้เป็นมาตรฐานเดียวกัน จำแนกตามรายวิชา                            | 74   |
| 4.4      | แสดงผลการเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2550 เดิมและคะแนนแปลงโดยการปรับเทียบให้เป็นมาตรฐานเดียวกัน จำแนกตามรายวิชา                                      | 76   |
| 4.5      | แสดงผลการเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2550 ที่เป็นคะแนนเดิมและคะแนนแปลงมาจากคะแนน O-NET ปี 2551 โดยการปรับเทียบให้เป็นมาตรฐานเดียวกัน จำแนกตามรายวิชา | 77   |
| 4.6      | แสดงผลการเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2551 ที่เป็นคะแนนเดิมและคะแนนแปลงมาจากคะแนน O-NET ปี 2550 โดยการปรับเทียบให้เป็นมาตรฐานเดียวกัน จำแนกตามรายวิชา | 78   |
| 4.7      | การเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ที่เป็นคะแนนจริงและร้อยละของคะแนนจริง ตามระดับความสามารถ สำหรับวิชาภาษาไทย   | 78   |
| 4.8      | 8 การเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ที่เป็นคะแนนจริงและร้อยละของคะแนนจริง ตามระดับความสามารถ สำหรับวิชาสังคมศึกษา ศาสนา และวัฒนธรรม  | 79   |
| 4.9      | การเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ที่เป็นคะแนนจริงและร้อยละของคะแนนจริง ตามระดับความสามารถ สำหรับวิชาภาษาอังกฤษ  | 80   |

## สารบัญตาราง (ต่อ)

| ตารางที่ |  | หน้า |
|----------|--|------|
| 4.10     | การเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ที่เป็นคะแนนจริงและร้อยละของคะแนนจริง ตามระดับความสามารถ สำหรับวิชาวิทยาศาสตร์  | 81   |
| 4.11     | การเปรียบเทียบค่าคะแนนจริงที่ได้จากการเทียบคะแนน O-NET วิชาภาษาไทย ปีการศึกษา 2550 กับ ปีการศึกษา 2551 ณ ความสามารถของผู้สอบระดับเดียวกัน  | 82   |
| 4.12     | การเปรียบเทียบค่าคะแนนจริงที่ได้จากการเทียบคะแนน O-NET วิชาสังคมศึกษา ศาสนา และวัฒนธรรม ปีการศึกษา 2550 กับ ปีการศึกษา 2551 ณ ความสามารถของผู้สอบระดับเดียวกัน   | 82   |
| 4.13     | การเปรียบเทียบค่าคะแนนจริงที่ได้จากการเทียบคะแนน O-NET วิชาภาษาอังกฤษ ปีการศึกษา 2550 กับ ปีการศึกษา 2551 ณ ความสามารถของผู้สอบระดับเดียวกัน   | 83   |
| 4.14     | การเปรียบเทียบค่าคะแนนจริงที่ได้จากการเทียบคะแนน O-NET วิชาวิทยาศาสตร์ ปีการศึกษา 2550 กับ ปีการศึกษา 2551 ณ ความสามารถของผู้สอบระดับเดียวกัน  | 84   |
| 4.15     | ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบในปี 2550 กับคะแนนแปลงที่ได้จากการปรับเทียบให้เป็นมาตรฐานเดียวกัน และ ได้จากการปรับเทียบโดยรูปแบบ IRT และระหว่างคะแนนแปลงด้วยตัวเอง                              | 84   |
| 4.16     | ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบในปี 2551 กับคะแนนแปลงที่ได้จากการปรับเทียบให้เป็นมาตรฐานเดียวกัน และ ได้จากการปรับเทียบโดยรูปแบบ IRT และระหว่างคะแนนแปลงด้วยตัวเอง                              | 85   |
| 4.17     | ความแตกต่างระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบในปี 2550 กับคะแนนแปลงที่ปรับเทียบโดยรูปแบบการเทียบให้เป็นมาตรฐานเดียวกัน และ ที่ปรับเทียบโดยรูปแบบ IRT และระหว่างคะแนนแปลงด้วยตัวเอง          | 85   |
| 4.18     | ความแตกต่างระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบในปี 2551 กับคะแนนแปลงที่ได้จากการปรับเทียบโดยรูปแบบการเทียบให้เป็นมาตรฐานเดียวกัน และ ที่ปรับเทียบโดยรูปแบบ IRT และระหว่างคะแนนแปลงด้วยตัวเอง | 86   |



# บทที่ 1

## บทนำ

### ความเป็นมา

ในการพัฒนาประเทศให้ทัดเทียมกับอารยประเทศนั้น ปัจจัยหนึ่งที่สำคัญยิ่ง คือ ประสิทธิภาพด้านความรู้ความสามารถ และควมมีคุณธรรมจริยธรรมของคนในชาติ ดังนั้น การจัดการศึกษาให้ทันต่อสถานการณ์ที่แปรเปลี่ยนไปอยู่ตลอดเวลา เพื่อให้คนมีความรู้ ความสามารถ มีไหวพริบที่จะนำประเทศให้อยู่รอด โดยไม่เบียดเบียนใคร แต่กลับมีความเอื้ออาทร ช่วยเหลือ เกื้อกูลทั้งผู้อยู่ใกล้ชิด บุคคลในสังคมของตน และในสังคมอื่น ๆ ด้วย การพัฒนาคนให้มีศักยภาพ ดังกล่าวนี้ต้องอาศัยการจัดการศึกษาของหน่วยงานที่รับผิดชอบทุกระดับ เพราะการให้การศึกษา เป็นหนทางแห่งการพัฒนาคนในชาติให้มีความพร้อมสำหรับการเปลี่ยนแปลงใหม่ ๆ ในสังคมที่ พัฒนาแล้วและที่กำลังพัฒนาแล้วให้ความสำคัญต่อการพัฒนาการศึกษา เพราะสามารถนำไปสู่ การพัฒนาคุณภาพชีวิตที่ดีของคนในชาติได้ และการศึกษายังเป็นปัจจัยสำคัญในการพัฒนาคน ให้รู้เท่าทันและพัฒนาตนเองให้เหมาะสมกับการเปลี่ยนแปลง สามารถดำรงชีพและประกอบอาชีพ ได้อย่างมีความสุข นอกจากนี้ยังเป็นพลังสร้างสรรค์ในการพัฒนาประเทศอีกด้วย

ด้วยความสำคัญของการจัดการศึกษาเพื่อให้เกิดการพัฒนาคนดังกล่าวมาแล้วนั้น จึงมี การกำหนดบทบัญญัติเพื่อแนวทางการจัดการศึกษาลงในพระราชบัญญัติการศึกษาแห่งชาติ ซึ่ง พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 และที่แก้ไขเพิ่มเติม(ฉบับที่ 2) พ.ศ. 2545 ได้ กำหนดความสำคัญของการศึกษาต่อการพัฒนาคนในชาติ โดยในมาตรา 6 ได้กำหนดว่า “ให้ การจัดการศึกษาต้องเป็นไปเพื่อพัฒนาคนไทยให้เป็นมนุษย์ที่สมบูรณ์ทั้งร่างกาย จิตใจ สติปัญญา ความรู้ และคุณธรรม มีจริยธรรมและวัฒนธรรมในการดำรงชีวิต สามารถอยู่ร่วมกับผู้อื่นได้อย่างมี ความสุข” โดยมีบุคลากรที่ต้องทำการอบรมให้ความรู้ โดยเริ่มตั้งแต่การจัดเตรียมความพร้อมเพื่อ สนับสนุนให้บุตรหลานตั้งแต่แรกเกิดจนถึงก่อนวัยเรียนที่ผู้ปกครองมีความสำคัญยิ่งในการดูแล จนก้าวเข้าสู่การเข้าเรียนตามระบบการศึกษาภาคบังคับ ซึ่งสำนักงานคณะกรรมการการศึกษาขั้น พื้นฐาน กระทรวงศึกษาธิการเป็นผู้รับผิดชอบหลักในการจัดการศึกษา โดยรับผิดชอบการจัด การศึกษาจนถึงระดับชั้นมัธยมศึกษาตอนปลายสำหรับสายสามัญ และยังรับผิดชอบการศึกษา สายอาชีพ และอื่น ๆ อีก ต่อจากนั้นสำนักงานคณะกรรมการการอุดมศึกษามีหน้าที่หลักในการ รับผิดชอบจัดการศึกษาตั้งแต่ระดับปริญญาตรี ไปจนถึงระดับสูงสุดคือปริญญาเอก ซึ่งจะเห็นว่า มีผู้ที่เกี่ยวข้องในการจัดการศึกษาหลากหลายกลุ่ม เพื่อให้การจัดการศึกษาของทุกกลุ่มสอดคล้อง เป็นไปในทิศทางเดียวกัน เป็นไปตามมาตรฐาน และมุ่งสู่เป้าหมายเดียวกันของการจัดการศึกษา ของชาติ จึงได้มีการกำหนดให้มีหมวด 6 ว่าด้วย มาตรฐานและการประกันคุณภาพการศึกษาและ

ภายใต้หมวดดังกล่าวนี้ยังได้ระบุ มาตรา 48 ที่ให้มีสำนักงานรับรองมาตรฐานและประเมินคุณภาพการศึกษา มีฐานะเป็นองค์กรมมหาชน ทำหน้าที่พัฒนาเกณฑ์ วิธีการประเมินคุณภาพภายนอก และทำการประเมินผลการจัดการศึกษาเพื่อตรวจสอบคุณภาพและแนวทางการจัดการศึกษา (สำนักงานคณะกรรมการการศึกษาแห่งชาติ. 2545)

จากที่กล่าวมาแล้ว นอกจากสำนักงานรับรองมาตรฐานและประเมินคุณภาพ จะต้องทำหน้าที่ในการตรวจสอบคุณภาพการจัดการศึกษาแล้ว หน่วยงานที่รับผิดชอบในการจัดการศึกษาทุกระดับจำเป็นต้องมีแนวทางการตรวจสอบคุณภาพการจัดการศึกษาของตนเองด้วย โดยควรมีระบบการวัดผลและการประเมินผลการศึกษาที่ได้มาตรฐาน ในสภาพที่เห็นทั่วไปมักพบว่า การวัดและประเมินผลการศึกษาส่วนใหญ่นิยมใช้แบบทดสอบเป็นเครื่องมือที่ทำให้ได้ข้อมูลหรือสารสนเทศเกี่ยวกับการจัดกิจกรรมทางการศึกษา ผลที่ได้จากการทดสอบมีความสำคัญต่อการพัฒนาคุณภาพของการศึกษา ตลอดจนใช้เป็นข้อมูลพื้นฐานในการตัดสินใจของครู ผู้บริหารและผู้ที่เกี่ยวข้องกับการจัดการศึกษาทุกระดับ นอกจากนี้ ยังมีการนำแบบทดสอบไปใช้อย่างแพร่หลายในปัจจุบัน ตามจุดมุ่งหมายที่ต่างกัน เช่น การสอบแข่งขัน การสอบคัดเลือกบุคคลเข้าเรียน เข้าทำงาน จัดตำแหน่ง เลื่อนระดับ ซึ่งในการสอบแต่ละครั้งนั้นอาจมีการใช้แบบทดสอบในลักษณะที่แตกต่างกันไป เพื่อให้สามารถวัดได้ตรงกับความสามารถที่แท้จริงของแต่ละบุคคลและเพื่อให้เกิดความยุติธรรมแก่ทุกคนด้วย

นอกจากสถานศึกษาที่จะต้องพัฒนาแนวทางการตรวจสอบคุณภาพการจัดการศึกษาของตนเองแล้ว ในปัจจุบันยังมีอีกหนึ่งหน่วยงานที่มีส่วนเกี่ยวข้องในการตรวจสอบคุณภาพการจัดการศึกษาของสถานศึกษาที่มีแนวทางการจัดการศึกษาแตกต่างกันไป ตามศักยภาพ และวิสัยทัศน์ของผู้นำและบุคลากรทางการศึกษา ในช่วงที่ผ่านมาได้ทำหน้าที่สะท้อนผลการจัดการศึกษาของสถานศึกษาในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน หน่วยงานนี้คือ สถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) หรือ สทศ. ซึ่งเป็นหน่วยงานของรัฐ จัดตั้งขึ้นตามพระราชกฤษฎีกาจัดตั้งสถาบันทดสอบทางการศึกษาแห่งชาติ (องค์การมหาชน) พ.ศ. 2548 (สถาบันทดสอบทางการศึกษาแห่งชาติ , <http://www.niets.or.th/pdf/history.pdf> สืบค้นวันที่ 7 พฤศจิกายน 2551) โดยมีเป้าหมายเพื่อบริหารจัดการและดำเนินการเกี่ยวกับการศึกษาวิจัย พัฒนา และให้บริการทางการประเมินผลทางการศึกษาและทดสอบทางการศึกษาอย่างต่อเนื่อง รวมทั้งเป็นศูนย์กลางความร่วมมือด้านการทดสอบทางการศึกษาระดับชาติและนานาชาติ (สถาบันทดสอบทางการศึกษาแห่งชาติ , <http://www.niets.or.th/pdf/obligation.pdf> สืบค้นวันที่ 7 พฤศจิกายน 2551) นอกจากนี้ยังมีอำนาจหน้าที่หลัก คือ ดำเนินการเกี่ยวกับ 1) การจัดทำระบบ วิธีการทดสอบ และพัฒนาเครื่องมือวัด และประเมินผลตามมาตรฐานการศึกษา 2) การประเมินผลการจัดการศึกษา และการทดสอบทางการศึกษาระดับชาติ ตลอดจนให้ความ

ร่วมมือและสนับสนุนการทดสอบทั้งระดับเขตพื้นที่การศึกษา และสถานศึกษา 3) การทดสอบทางการศึกษา บริการสอบวัดความรู้ความสามารถและการสอบวัดมาตรฐานวิชาการและวิชาชีพ เพื่อนำผลไปใช้เป็นส่วนหนึ่งในการเทียบระดับ และเทียบโอนผลการเรียนที่มาจากการศึกษาในระบบเดียวกันหรือการศึกษาผ่านระบบ 4) การศึกษาวิจัย และเผยแพร่นวัตกรรมเกี่ยวกับการทดสอบทางการศึกษา ตลอดจนเผยแพร่เทคนิคการวัดและประเมินผลการศึกษา 5) การเป็นศูนย์กลางข้อมูลการทดสอบและประเมินผลทางการศึกษา ตลอดจนสนับสนุนและให้บริการผลการทดสอบแก่หน่วยงานต่าง ๆ ได้ทั้งในประเทศและต่างประเทศ 6) การพัฒนาและส่งเสริมวิชาการด้านการทดสอบและประเมินผลทางการศึกษา รวมถึงการพัฒนาบุคลากรด้านการทดสอบและประเมินผลด้านการติดตามและประเมินผลคุณภาพบัณฑิต รวมทั้งการให้การรับรองมาตรฐานของระบบวิธีการ เครื่องมือวัดของหน่วยงานการประเมินผลและทดสอบทางการศึกษา และ 7) การเป็นศูนย์กลางความร่วมมือด้านการทดสอบทางการศึกษาทั้งในระดับประเทศและระดับนานาชาติ (สถาบันทดสอบทางการศึกษาแห่งชาติ , <http://www.niets.or.th/pdf/authority.pdf> สืบค้นวันที่ 7 พฤศจิกายน 2551)

จากที่กล่าวมาแล้วข้างต้น จะเห็นว่าหน้าที่หลัก ๆ ของ สทศ. นั้น ส่วนใหญ่เกี่ยวข้องกับการวัดผลและการประเมินผลการศึกษา ซึ่งในการดำเนินการทดสอบที่ สทศ. รับผิดชอบนั้นต้องเป็นกระบวนการที่ได้มาตรฐาน โดยเฉพาะเครื่องมือที่ใช้ในการวัดผลและประเมินผลนั้นจะต้องมีความเชื่อมั่นและมีความเที่ยงตรงในการวัดสูง สำหรับการดำเนินงานหนึ่งที่สถานศึกษาทุกแห่งต้องทำการทดสอบผู้เรียน โดยการบริหารจัดการด้านการสอบจาก สทศ. คือ การทดสอบ O-NET ซึ่งคะแนนจากการทดสอบ O-NET นี้นอกจากจะสามารถใช้เป็นเกณฑ์ร่วมกันในการวัดคุณภาพของทุกหน่วยงานว่าสามารถจัดการศึกษาขั้นพื้นฐานอยู่ในระดับที่ใกล้เคียงกัน หรือแตกต่างกันมากน้อยเพียงใดแล้ว สำหรับคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 นั้น ยังมีความสำคัญอีกชั้นหนึ่งต่อตัวนักเรียนเอง นั่นคือ คะแนน O-NET นี้ได้ถูกใช้เป็นส่วนหนึ่งในการยื่นเข้าไปพิจารณาเพื่อคัดเลือกเข้าศึกษาต่อในระดับอุดมศึกษาต่อไป แต่ เนื่องจากจำนวนในการรับนักเรียนเข้าไปเรียนในระดับอุดมศึกษามีจำกัด จึงทำให้มีนักเรียนบางส่วนที่อาจพลาดในการผ่านเกณฑ์การเข้าไปเรียนในระดับอุดมศึกษาได้ทันทีที่จบการศึกษาระดับมัธยมศึกษาตอนปลาย ดังนั้น จึงต้องเก็บคะแนนที่ได้จากการสอบ O-NET ไว้เพื่อนำไปเป็นส่วนหนึ่งเพื่อร่วมพิจารณาให้ผ่านเกณฑ์การเข้าไปศึกษาต่อในระดับอุดมศึกษาต่อไป โดยจะมีผลต่อคะแนนสอบเข้า ร้อยละ 35-70 ทั้งนี้ขึ้นอยู่กับข้อกำหนดของแต่ละคณะ แต่ในการสอบนักเรียนสามารถทดสอบเพื่อเก็บคะแนนไว้พิจารณาสอบเข้าเรียนระดับอุดมศึกษาได้เพียงครั้งเดียว คือช่วงที่กำลังศึกษาในระดับมัธยมศึกษาปีที่ 6 จึงอาจเกิดเหตุการณ์ที่มีการนำคะแนน O-NET ที่ทดสอบต่างปีกันมารวม

พิจารณาสอบเข้าเรียนในคณะเดียวกันได้ ดังนั้นแบบทดสอบที่ใช้ทดสอบแต่ละปีต้องได้มาตรฐาน ไม่แตกต่างกัน

จากพื้นฐานความเชื่อเกี่ยวกับการเทียบผลจากการทดสอบที่ผู้สอบทำแบบทดสอบต่างฉบับกัน และสามารถเปรียบเทียบความสามารถของผู้สอบได้อย่างไม่ซับซ้อน นั่นคือ การวัดผล การทดสอบจากแบบทดสอบคู่ขนาน ซึ่งสามารถนำคะแนนดิบที่ได้จากการทดสอบฉบับต่าง ๆ มาเปรียบเทียบกันได้โดยตรง (สุรรัตน์ ขวัญบุญจันทร์. 2531 : 1-2) แต่การสร้างแบบทดสอบให้เป็นคู่ขนานกันอย่างแท้จริงนั้นทำได้ยาก เพราะ จะต้องสร้างให้คู่ขนานกันทั้งในด้านโครงสร้างของ ข้อสอบ(Structure) เวลาที่ใช้ในการสอบ(Timing) ชนิดของข้อสอบ(Item Types) รูปแบบข้อสอบ และเนื้อหาในข้อสอบ (ชูศักดิ์ ชัมภลชิต. 2529 :144) นอกจากนี้ในการพยายามสร้างข้อสอบที่เป็น คู่ขนานจริงอาจต้องสูญเสียเวลาและไม่คุ้มค่างับการลงทุน จึงมีผู้ คิดค้นรูปแบบที่อาศัยวิธีการ ทางสถิติมาทำการปรับคะแนนให้เกิดความเท่าเทียมกันได้ โดยเรียก ว่าวิธี “การปรับเทียบ คะแนน” (Test Equating) ซึ่งวิธีการนี้จะสามารถแปลงคะแนนของแบบทดสอบชุดหนึ่งไปสู่มาตรา เดียวกันกับคะแนนของแบบทดสอบอีกชุดหนึ่งคะแนนของแบบทดสอบต่างชุดกันจะอยู่บนสเกล การวัดเดียวกัน จึงสามารถหาคะแนนสมมูลระหว่างแบบทดสอบต่างชุดนั้นและนำคะแนนมา เปรียบเทียบกันได้(Angoff. 1971 : 562 ; Rentz and Bashaw. 1977 :163) ทั้งนี้ไม่ ว่า แบบทดสอบต่างชุดนั้นจะคู่ขนานกันหรือไม่ก็ตาม ถ้ามุ่งวัดในสิ่งเดียวกันก็จะสามารถแปลง คะแนนจากแบบทดสอบต่างชุดให้มีคุณสมบัติที่เทียบเท่ากันได้เสมอ (Lord. 1977 :128 ; Kolen. 1981 :1-2)

ประเภทของการปรับเทียบคะแนนสามารถแบ่งได้เป็น 2 ประเภท (Hambleton and Swaminathan1985:197) คือ 1) การปรับเทียบคะแนนตามแนวนอน (Horizontal Equating) เป็นการปรับเทียบคะแนนระหว่างแบบทดสอบที่มีความยากใกล้เคียงกันและกลุ่มผู้สอบมีความสามารถใกล้เคียงกันและ 2) การปรับเทียบคะแนนตามแนวตั้ง (Vertical Equating) เป็นการปรับเทียบคะแนนระหว่างแบบทดสอบที่มีความยากต่างกัน และกลุ่มผู้สอบมีความสามารถแตกต่างกัน อาจเป็นผู้สอบที่เรียนต่างระดับกัน สำหรับวิธีการปรับเทียบคะแนนมีหลายรูปแบบ ถ้าแบ่งตามทฤษฎีที่สนับสนุนจะแบ่งได้สองกลุ่มใหญ่ ๆ (Flanagan. 1951 : 750 - 760) คือ 1) กลุ่มทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) ประกอบด้วยรูปแบบเชิงเส้นตรง(Linear Equating) รูปแบบอิกวิเปอร์เซ็นไทล์ (Equipercetile Equating) และรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ (Equating Using the Confirmatory Factor Analysis Model) และ 2) กลุ่มทฤษฎีการวัดความสามารถแฝง (Latent Trait Theory) ประกอบด้วยรูปแบบอิงทฤษฎีการตอบข้อสอบ (Item Response Theory Equating Model)ซึ่งได้จำแนกการปรับเทียบตามโมเดล 1 พารามิเตอร์ หรือ

บางครั้งเรียกว่าราชสีโมเดล (Rasch Model) การปรับเทียบตามโมเดล 2 พารามิเตอร์ (Two Parameter Model) และการปรับเทียบตามโมเดล 3 พารามิเตอร์ (Three Parameter Model)

จากวิธีการที่กล่าวมานี้ มีกฎในการแปลงคะแนนต่างกันและขึ้นอยู่กับวิธีแต่ละอย่าง จึงเกิดปัญหาว่าวิธีที่ให้ค่าการแปลงคะแนนได้ถูกต้องที่สุดหรือให้โอกาสที่จะคลาดเคลื่อนได้น้อยที่สุด เพื่อให้เกิดความยุติธรรมแก่ผู้สอบมากที่สุด และจากการที่ สทศ. ได้เห็นความสำคัญต่อในการนำคะแนนสอบที่อาจมาจากการสอบต่างปีกันมาพิจารณาเปรียบเทียบกันอย่างไม่ให้เกิดความลำเอียง และสนใจต่อการดำเนินการปรับเทียบคะแนน O-NET ที่ทำการทดสอบต่างปีกันให้สามารถปรับเทียบกันได้ ดังกล่าวไปแล้วในข้างต้นนั้น จึงเห็นควรทำการศึกษาค้นคว้าหาวิธีการที่เหมาะสมกับการปรับเทียบคะแนน O-NET ระหว่างปีขึ้น

### คำถามการวิจัย

วิธีการปรับเทียบ คะแนน O-NET ระหว่างปีการศึกษา 2550 และปีการศึกษา 2551 ที่ดำเนินการทดสอบโดย สำนักทดสอบแห่งชาติ(สทศ.) ที่เหมาะสม ควรเป็นวิธีการใด ระหว่างวิธีการเทียบด้วยมาตราเดียวกัน และ โดยการปรับเทียบตามรูปแบบการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ (หรือ รูปแบบIRT)

### วัตถุประสงค์ของการวิจัย

เพื่อเปรียบเทียบ วิธีการปรับเทียบคะแนน O-NET ระหว่างปี 2550 และปี 2551 โดยใช้วิธีการเทียบเป็นมาตราเดียวกัน และ โดยการปรับเทียบตามรูปแบบการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ (หรือ รูปแบบIRT)

### สมมติฐานการวิจัย

ในการศึกษาคั้งนี้มุ่งศึกษาวิธีการเปรียบเทียบวิธีการปรับเทียบคะแนน O-NET ระหว่างปี 2550 และ ปี 2551 โดยวิธีการเทียบเป็นมาตราเดียวกัน และ ปรับเทียบตามรูปแบบการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อจะดูว่าวิธีใดมีคุณภาพ และเหมาะสมที่จะนำมาใช้ในการปรับเทียบได้ดีกว่ากัน ในแต่ละวิชา ผู้วิจัยได้ตั้งสมมติฐานการวิจัยไว้ดังนี้

1. คะแนน O-NET เดิมกับคะแนนแปลง O-NET ที่ปรับเทียบแล้วไม่แตกต่างกัน
2. คะแนน O-NET ที่ได้จากการปรับเทียบในแต่ละวิธีมีความสัมพันธ์กับคะแนน O-NET เดิม ทางบวก
3. คะแนน O-NET ที่ได้จากการปรับเทียบต่างวิธีกันมีความสัมพันธ์กันทางบวก

## ขอบเขตของการวิจัย

1. การวิจัยครั้งนี้มีขอบเขตการศึกษาวิธีการปรับเทียบคะแนนในแนวนอน (Horizontal Equating) โดยใช้วิธีการปรับเทียบคะแนน 2 วิธี คือ ใช้วิธีการปรับเทียบเป็นมาตราเดียวกัน และรูปแบบการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ (IRT)

### 2. ประชากรและกลุ่มตัวอย่าง

2.1 ในการวิจัยครั้งนี้ ทำการศึกษาจากคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปี 2550 และ ปี 2551 ซึ่งดำเนินการสร้างและบริหารจัดการสอบ โดย สทศ.

2.2 กลุ่มตัวอย่างในการศึกษา เป็นคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปี 2550 และ ปี 2551 รายวิชาละ 1998 – 1999 คน

3. คะแนน O-NET ที่นำมาเปรียบเทียบในครั้งนี้ มี 4 รายวิชา คือ 1) วิชาภาษาไทย 2) สังคมศึกษา ศาสนาและวัฒนธรรม 3) ภาษาอังกฤษ และ 4) วิทยาศาสตร์

### 4. ตัวแปรที่ทำการศึกษวิจัย

ตัวแปรต้น ได้แก่

- วิธีการปรับเทียบคะแนน
- รายวิชาของคะแนน O-NET
- ปีที่ทำการทดสอบ คะแนน O-NET

ตัวแปรตาม ได้แก่ คะแนนแปลง O-NET ที่ได้จากการปรับเทียบคะแนน O-NET ระหว่างปีการศึกษา 2550 และ ปีการศึกษา 2551

5. เกณฑ์ที่ใช้ในการพิจารณาคุณภาพของวิธีการปรับเทียบคะแนน O-NET คือ วิธีการปรับเทียบคะแนน O-NET ที่มีคุณภาพ จะเป็นวิธีที่ทำให้ได้คะแนนแปลง O-NET ที่มีความสัมพันธ์กับคะแนน O-NET เดิมสูงกว่า ความสัมพันธ์ระหว่างคะแนนแปลง O-NET โดยวิธีอื่นกับคะแนน O-NET เดิม

## นิยามศัพท์ที่ใช้ในการวิจัย

1. คะแนน O-NET หมายถึงคะแนนที่จัดสอบขึ้นโดย สทศ. โดยในการศึกษาครั้งนี้เป็นคะแนนสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบในปี 2550 และ ปี 2551 สำหรับ O-NET เป็นคำที่ย่อมาจาก Ordinary National Education Test หรือ เรียกเป็นภาษาไทยว่า การทดสอบทางการศึกษาแห่งชาติขั้นพื้นฐาน

2. การปรับเทียบคะแนน (Test Equating) หมายถึง กระบวนการที่ใช้วิธีการทางสถิติเพื่อแปลงคะแนนที่ได้จากแบบทดสอบต่างฉบับกัน ให้มาอยู่ในคะแนนมาตรฐานหรือหน่วยวัด หรือ อยู่ในบสเกลเดียวกัน เพื่อให้คะแนนจากแบบทดสอบทั้งสองฉบับนั้นสามารถเปรียบเทียบกันได้ โดยใช้รูปแบบการปรับเทียบมาตรา การใช้รูปแบบทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ และการพิจารณาความเป็นข้อสอบคู่ขนาน

3. วิธีการปรับเทียบคะแนนโดยวิธีการเทียบเป็นมาตราเดียวกัน หมายถึง วิธีการปรับเทียบคะแนน O-NET ที่ผู้สอบต่างกัน ทดสอบข้อสอบคนละชุด หรือคนละปีกัน โดยใช้วิธีการเทียบเป็นมาตราเดียวกันจากสูตรในการเทียบคะแนน O-NET ในแต่ละปีดังนี้

$$X'_{O-NET51} = X_{O-NET 50} \left( \frac{SD_{O-NET51}}{SD_{O-NET50}} \right) - \bar{X}_{O-NET 50} \left( \frac{SD_{O-NET51}}{SD_{O-NET50}} \right) + \bar{X}_{O-NET 51}$$

$$X'_{O-NET 50} = X_{O-NET 51} \left( \frac{SD_{O-NET50}}{SD_{O-NET51}} \right) - \bar{X}_{O-NET 51} \left( \frac{SD_{O-NET50}}{SD_{O-NET51}} \right) + \bar{X}_{O-NET 50}$$

โดยที่

$X'_{O-NET}$  หมายถึง คะแนน O-NET ที่ได้จากการปรับเทียบ

$X_{O-NET}$  หมายถึง คะแนน O-NET รายบุคคล

$\bar{X}_{O-NET}$  หมายถึง ค่าเฉลี่ยคะแนน O-NET รายบุคคล

$SD_{O-NET}$  หมายถึง ค่าเบี่ยงเบนมาตรฐานคะแนน O-NET รายบุคคล

4. วิธีการปรับเทียบคะแนนตามรูปแบบ IRT หมายถึง วิธีการปรับเทียบคะแนน O-NET โดยยึดตามทฤษฎีตอบสนองข้อสอบที่มีสามพารามิเตอร์

### ประโยชน์หรือความรู้ที่คาดว่าจะได้รับ

1. ได้วิธีการปรับเทียบคะแนน O-NET ของนักเรียนที่มาจากต่างกลุ่มกันและทำการทดสอบแบบทดสอบต่างปี ซึ่งสามารถเลือกใช้วิธีการปรับเทียบได้หลายวิธี ตามความเหมาะสม
2. ทำให้ได้องค์ความรู้เกี่ยวกับวิธีการปรับเทียบคะแนนที่ต่างวิธีกัน และคะแนนแปลงที่ได้จากการปรับเทียบของวิธีการปรับเทียบที่แตกต่างกัน
3. นำวิธีการปรับเทียบคะแนนที่เหมาะสมไปใช้ประโยชน์ในการปรับเทียบคะแนน O-NET ที่สอบต่างปีกันนักเรียน ซึ่งจะสามารถสร้างความมั่นใจในความยุติธรรมต่อการยื่นคะแนนสอบเข้าเรียนระดับอุดมศึกษาให้กับผู้ปกครอง และนักวิชาการ ซึ่งสามารถช่วยลดปัญหาความไม่เสมอภาคของการนำคะแนนในปีที่แตกต่างกันมาเทียบคะแนนกัน

## ข้อจำกัดของการวิจัย

ข้อจำกัดของการวิจัยครั้งนี้ มีดังนี้

1. การที่ไม่สามารถนำข้อมูลคะแนน O-NET ทุกค่ามาทำการ ศึกษา เพราะการปรับเทียบคะแนน O-NET ตามรูปแบบ IRT นั้นผู้วิจัยใช้โปรแกรมการวิเคราะห์ ของอาจารย์ คณิต ไช่มุกด์ ซึ่งสามารถวิเคราะห์ข้อมูลจากคะแนนรายข้อของผู้สอบ จากแบบสอบที่มีความยาวไม่เกิน 100 ข้อ ได้ประมาณ 2,000 ชุด ผู้วิจัยจึงทำการสุ่มตัวอย่างคะแนนนักเรียนจากคะแนนทั้งหมดที่ สทศ. สามารถจัดสอบนักเรียน มาเป็นกลุ่มตัวอย่างในการศึกษาครั้งนี้ และในการศึกษาครั้งนี้
2. ข้อมูลคะแนน O-NET รายวิชาที่ 4 คือวิชาคณิตศาสตร์นั้น เมื่อใช้โปรแกรมของ อ.คณิต ไช่มุกด์ ทำการวิเคราะห์ข้อมูลตามรูปแบบ IRT พบว่าข้อมูล ปีการศึกษา 2551 ไม่สามารถทำการวิเคราะห์หาค่าพารามิเตอร์ ซึ่งผู้วิจัยได้ทำการสุ่มตัวอย่างหลายครั้งและพบว่าทุกครั้งไม่สามารถทำการวิเคราะห์หาค่าพารามิเตอร์ได้เหมือนกัน ในการวิจัยครั้งนี้จึงทำการศึกษาเพียง 4 รายวิชา คือ ภาษาไทย สังคมศึกษา ศาสนา และวัฒนธรรม ภาษาอังกฤษ และวิทยาศาสตร์

## บทที่ 2

### เอกสารและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาเอกสาร บทความ และงานวิจัยที่เกี่ยวข้องกับการปรับเทียบคะแนนครั้งนี้ ผู้วิจัยขอเสนอผลการศึกษา ตามลำดับดังนี้

ตอนที่ 1 แนวคิดเกี่ยวกับการปรับเทียบคะแนน

- 1.1 ความหมายของการปรับเทียบคะแนน
- 1.2 เงื่อนไขของการปรับเทียบคะแนน
- 1.3 ประเภทของการปรับเทียบคะแนน
- 1.4 รูปแบบวิธีการเก็บข้อมูล
- 1.5 รูปแบบของการปรับเทียบคะแนน

ตอนที่ 2 การหาคุณภาพของการปรับเทียบคะแนน

- 2.1 ความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน
- 2.2 การประเมินความเพียงพอของการปรับเทียบคะแนน

ตอนที่ 3 งานวิจัยที่เกี่ยวข้องกับการปรับเทียบคะแนน

- 3.1 งานวิจัยต่างประเทศ
- 3.2 งานวิจัยในประเทศ

**ตอนที่ 1 แนวคิดเกี่ยวกับการปรับเทียบคะแนน**

**1.1 ความหมายของการปรับเทียบคะแนน**

ความหมายและนิยามของการปรับเทียบคะแนน ที่นักการศึกษาหลายท่านได้กล่าวไว้ เช่น กัลลิกเสน (ภาวิณี ศรีสุขวัฒนานนท์. 2529 : 17 ; อ้างอิงจาก Gulliksen. 1950 : 298) ได้ให้ความหมายของการปรับเทียบคะแนนระหว่างแบบทดสอบว่าเป็นวิธีการทำคะแนนจากแบบทดสอบสองชุดที่วัดในวิชาเดียวกันให้เป็นคะแนนสมมูล (Equivalent scores) ที่เปรียบเทียบกันโดยตรง โดยเสนอวิธีการให้ผู้สอบ โดยเสนอวิธีการให้ผู้สอบกลุ่มเดียวทำแบบทดสอบสองชุด และใช้วิธีง่าย ๆ คือ แปลงคะแนนแต่ 'ละชุดให้เป็นคะแนนมาตรฐานแล้วนำคะแนนที่แปลงแล้วมาเทียบกันโดยตรง แต่ต้องตรวจสอบให้แน่ใจก่อนว่า คะแนนสองชุดนั้น มีการกระจายของคะแนน

เป็นโค้งปกติ ซึ่งสอดคล้องกับ ปีเตอร์เซนและคณะ (Petersen & other. 1982 : 242) ที่ให้นิยามของการปรับเทียบคะแนนระหว่างแบบทดสอบว่าเป็นกระบวนการเชิงประจักษ์ที่ใช้ในการแปลงคะแนนจากแบบทดสอบฉบับหนึ่งไปยังแบบทดสอบอีกฉบับหนึ่ง โดยที่แบบทดสอบทั้งสองนั้นต้องวัดคุณลักษณะเดียวกัน และสอดคล้องกับศิริชัย กาญจนวาสี (2541 : 2) ซึ่งได้ให้นิยามว่า การปรับเทียบคะแนนระหว่างแบบทดสอบเป็นกระบวนการแปลงคะแนนของทดสอบต่างฉบับที่มุ่งวัดคุณลักษณะเดียวกันให้เป็นคะแนนที่สมมูลกันเพื่อให้คะแนนสามารถเปรียบเทียบกันได้โดยตรง ส 'วนลอร์ด ( Lord. 1980 : 195) ได้กล่าวถึงการปรับเทียบคะแนนระหว่างแบบทดสอบว่าเป็นการแปลงคะแนนจากแบบทดสอบต่างฉบับ ให้มีความหมายสลับเปลี่ยนกันได้ และเพื่อความเสมอภาคของบุคคล ผู้รับการทดสอบ ในขณะที่แองกอฟ (Angoff. 1984 : 85) ให้ความหมายของการปรับเทียบคะแนนระหว่างแบบทดสอบว่า หมายถึงกระบวนการแปลงระบบคะแนนของแบบทดสอบฉบับหนึ่งไปสู่ระบบคะแนนของแบบทดสอบอีกฉบับหนึ่งซึ่งวัดสิ่งเดียวกัน เพื่อให้คะแนนแปลงจากแบบทดสอบทั้งสองฉบับนั้น สามารถเทียบเท่ากัน และใช้แทนกันได้

จากแนวคิดและการให้ความหมายดังกล่าว ข้างต้นพอจะสรุปได้ว่า การปรับเทียบคะแนนเป็นวิธีการทางสถิติในการนำคะแนนจากแบบสอบต่างชุดกันแต่วัด สิ่งเดียวกันไปแปลงให้มีหน่วยวัดเดียวกัน เพื่อให้คะแนนจากแบบสอบต่างชุดกันสามารถเปรียบเทียบกันได้โดยตรง

## 1.2 เงื่อนไขของการปรับเทียบคะแนน

เนื่องจากวัตถุประสงค์ของการปรับเทียบคะแนนระหว่างแบบทดสอบ คือการทำให้คะแนนที่ได้ จากแบบทดสอบ 2 ฉบับใด ๆ มีความเท่าเทียมกันมากที่สุด ซึ่งต้องเกี่ยวข้องกับแบบแผนในการเก็บรวบรวมข้อมูล และกฎเกณฑ์การแปลงคะแนนจากแบบทดสอบฉบับหนึ่งไปยังแบบทดสอบอีกฉบับหนึ่ง จึงต้องกำหนดเงื่อนไขบางประการในการปรับเทียบคะแนน ซึ่งลอร์ด (Petersen and others. 1982 : 242 ; citing Lord. 1980) ได้กล่าวถึงเงื่อนไข 4 ข้อที่จะต้องนำมาใช้ในการปรับเทียบคะแนน ดังนี้

1. แบบทดสอบทั้งสองฉบับจะต้องวัดความสามารถเดียวกัน คือ แบบทดสอบทั้งสองต้องวัดคุณลักษณะเดียวกัน คุณลักษณะนี้อาจเป็นคุณลักษณะแฝง หรือความสามารถ หรือทักษะอย่างใดอย่างหนึ่งก็ได้

2. ความเสมอภาค (Equity) คือ เมื่อทุกกลุ่มมีความสามารถเดียวกัน การแจกแจงคะแนนของแบบทดสอบที่ได้ หลังจากที่มีการแปลงคะแนนแล้ว จะมีการแจกแจงเหมือนกับการแจกแจงของคะแนนจากแบบทดสอบที่ใช้เป็นฉบับเทียบคะแนน

3. ความไม่แปรผันตามกลุ่ม (Invariance Across Groups) คือคะแนนที่ได้จากการ

แปลงคะแนนไม่ว่าจะมาจากกลุ่มใด ๆ ก็ตามจะมีค่าเท่าเทียมกันหรือมีค่าไม่เปลี่ยนแปลงไปตามกลุ่มผู้เข้าสอบ

4. มีความสมมาตร (Symmetry) คือผลของการปรับเทียบคะแนนต้องเหมือนกันไม่ว่าจะเทียบจากแบบทดสอบฉบับ X ไปยังฉบับ Y หรือเทียบจากแบบทดสอบฉบับ Y ไปยังฉบับ X

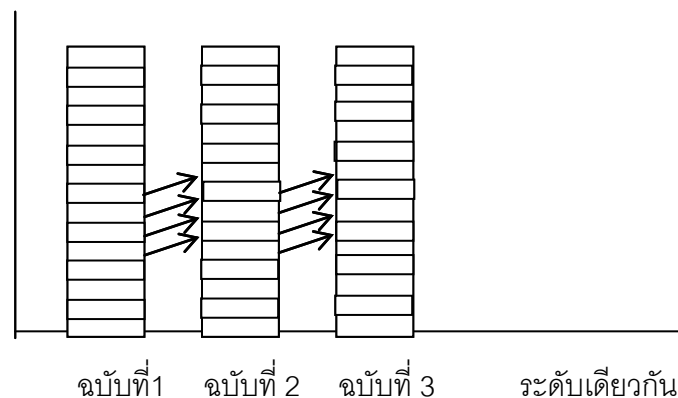
### 1.3 ประเภทของการปรับเทียบคะแนน

การทดสอบที่มีความจำเป็นต้องใช้เทคนิคการปรับเทียบคะแนนระหว่างแบบทดสอบสามารถจำแนกได้เป็น 2 สถานการณ์ (ศิริชัย กาญจนวาสี. 2541 : 5 อ้างอิงมาจาก Hambleton and Swaminathan, 1985) ซึ่งต้องใช้รูปแบบการปรับเทียบคะแนน (Equating Forms) 2 รูปแบบ ดังนี้

#### 3.1 การปรับเทียบคะแนนตามแนวนอน (Horizontal Equating)

การปรับเทียบคะแนนตามแนวนอนเป็นการปรับเทียบคะแนนระหว่างแบบทดสอบต่างฉบับกันเมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกันมีระดับความยากใกล้เคียงกัน และกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ในประชากรเดียวกันและมีความสามารถใกล้เคียงกัน การปรับเทียบคะแนนตามแนวนอนเป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบทดสอบเนื้อหาเดียวกันขึ้นมาหลาย ๆ ฉบับ ด้วยเหตุผลของการนำไปใช้ในการทดสอบให้เกิดความยุติธรรมและป้องกันความลับของข้อสอบเมื่อใช้ในเวลาที่ต่างกัน สำหรับผู้สอบกลุ่มขนาดใหญ่เพื่อเทียบว่าคะแนนที่ได้จากฉบับหนึ่งเทียบเป็นเท่าไรของอีกฉบับหนึ่ง ซึ่งวัดในระดับเดียวกันจึงเป็นการปรับเทียบคะแนนระหว่างแบบทดสอบต่างฉบับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบระดับชั้นเดียวกัน เช่น การปรับเทียบคะแนนระหว่างแบบทดสอบวิชาคณิตศาสตร์ 2 ฉบับ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 เป็นต้น

คะแนน

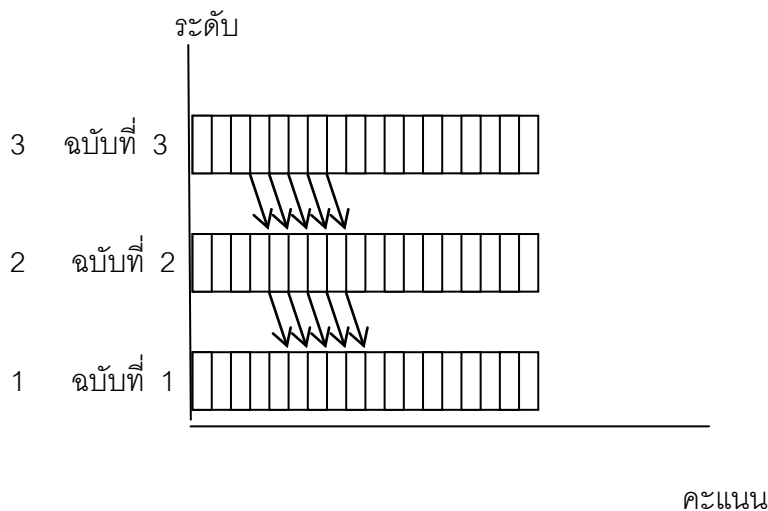


### ภาพประกอบ 1 การปรับเทียบคะแนนตามแนวนอน (Horizontal Equating)

แบบทดสอบต่างฉบับที่ต้องการนำมาเทียบมาตรฐานตามแนวนอนนั้น ถึงแม้ว่าแบบทดสอบเหล่านั้นยากที่จะมีความเท่าเทียมกันในทุกด้าน แต่ก็ควรเป็นแบบทดสอบที่พยายามสร้างหรือมุ่งให้คู่ขนานกัน (Alternate Forms) นอกจากนี้การแจกแจงความสามารถของผู้สอบที่ทำการสอบด้วยแบบทดสอบแต่ละฉบับควรมีการแจกแจงที่ประมาณได้ว่ามีระดับความสามารถที่ใกล้เคียงกัน

### 3.2 การปรับเทียบคะแนนตามแนวตั้ง (Vertical Equating)

การปรับเทียบคะแนนแนวตั้งเป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบทดสอบเนื้อหาเดียวกัน แต่ต่างฉบับต่างมุ่งวัดความสามารถของผู้สอบที่ต่างระดับกัน เพื่อเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่งเทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน จึงเป็นการปรับเทียบคะแนนระหว่างแบบทดสอบต่างระดับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบต่างระดับชั้นกันเช่น การปรับเทียบคะแนนระหว่างแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น เป็นต้น



### ภาพประกอบ 2 การปรับเทียบคะแนนตามแนวตั้ง (Vertical Equating)

แบบทดสอบต่างฉบับที่วัดเนื้อหาเรื่องเดียวกัน แต่ต่างระดับกันที่ต้องการนำมาเทียบคะแนนกันนั้น แบบทดสอบแต่ละฉบับจึงมีความยากแตกต่างกันไป รวมทั้งกลุ่มผู้สอบแต่ละฉบับมีการแจกแจงความสามารถอยู่ต่างประชากรหรือมีการแจกแจงความสามารถที่อยู่ระดับต่างประชากรหรือมีการแจกแจงความสามารถที่อยู่ในระดับที่ต่างกัน ดังนั้นการปรับเทียบคะแนนตาม

แนวตั้ง จึงมีความสลับซับซ้อนกว่าการเปรียบเทียบคะแนนตามแนวนอนทั้งในแง่ทฤษฎีและการปฏิบัติ

รูปแบบของการเปรียบเทียบคะแนนสามารถแบ่งได้ 2 รูปแบบ คือ การเทียบมาตรฐานตามแนวนอนเป็นการเปรียบเทียบคะแนนระหว่างแบบทดสอบที่มีความยากใกล้เคียงกัน และกลุ่มผู้สอบมีความสามารถใกล้เคียงกันด้วย การเปรียบเทียบคะแนนตามแนวตั้งเป็นการเปรียบเทียบคะแนนของแบบทดสอบที่มีความยากต่างกัน และกลุ่มผู้สอบก็มีความสามารถต่างกันด้วย

#### 1.4 รูปแบบวิธีการเก็บข้อมูล

ข้อมูลจากการใช้แบบทดสอบหลายฉบับและเก็บรวบรวมข้อมูลสำหรับนำมาใช้ในการเทียบมาตรฐานสามารถใช้จำแนกเป็นแบบแผนการเก็บรวบรวมข้อมูลได้ดังนี้ (ศิริชัย กาญจนวาสี 2541 : 8-11 อ้างอิงมาจาก Petersen and others. 1989 : 244 - 246)

##### 1.4.1 รูปแบบผู้สอบกลุ่มเดียว (Single Group Design)

ก. รูปแบบผู้สอบกลุ่มเดียวที่ไม่ได้รับการจัดให้สมดุล (Unconterbalanced Design)

รูปแบบนี้เป็นรูปแบบที่ง่ายที่สุด ใช้ผู้สอบกลุ่มเดียวทำหน้าที่เป็นผู้สอบร่วมวิธีการเก็บรวบรวมข้อมูลกระทำโดยใช้กลุ่มตัวอย่างผู้สอบกลุ่มเดียว แต่ละคนทำแบบทดสอบทั้ง 2 ฉบับ ให้ผู้สอบทำแบบทดสอบฉบับหนึ่งแล้วตามด้วยแบบทดสอบอีกฉบับหนึ่ง การให้ผู้สอบกลุ่มเดียวกันทำแบบทดสอบทั้ง 2 ฉบับ ระดับความสามารถของผู้สอบทั้ง 2 ฉบับ เท่ากัน จึงไม่出现问题แทรกซ้อนรวมทั้งไม่เป็นปัญหาในเรื่องความยากของแบบทดสอบต่างฉบับ คะแนนจากแบบทดสอบจึงสามารถเปรียบเทียบกันได้บนพื้นฐานความสามารถที่เท่ากันจากคนกลุ่มเดียวกัน แต่ในทางปฏิบัติ การทดสอบฉบับหลังผู้สอบอาจได้รับปัจจัยที่ส่งผลกระทบต่อผลการทดสอบฉบับแรก เช่น การเรียนรู้ การฝึกฝน ความเมื่อยล้าจากการทำแบบทดสอบฉบับแรก เป็นต้น ซึ่งอาจมีอิทธิพลต่อการเปรียบเทียบคะแนน

ข. รูปแบบกลุ่มเดียวที่ได้รับการจัดให้สมดุล (Counterbalanced Design)

รูปแบบนี้ปรับจากแบบแรก เพื่อขจัดผลของลำดับการทดสอบ ก่อน-หลัง โดยการสุ่มผู้สอบแยกเป็น 2 กลุ่มย่อย แต่ละกลุ่มได้รับการทดสอบด้วยแบบทดสอบทั้ง 2 ฉบับ ในลักษณะที่กลุ่มย่อยหนึ่งให้ทำแบบทดสอบฉบับ 1 แล้วตามด้วยแบบทดสอบฉบับ 2 ส่วนอีกกลุ่มย่อยหนึ่งทำให้แบบทดสอบฉบับ 2 แล้วตามด้วยฉบับ 1 สลับกันเพื่อให้เกิดความสมดุล การใช้รูปแบบนี้กลุ่มผู้สอบจึงได้รับอิทธิพลจากลำดับการทดสอบ การเรียนรู้ การฝึกฝนและความเมื่อยล้าที่สมดุลกันในทั้งสองกลุ่มย่อย

| การเก็บรวบรวมข้อมูล                                 | กลุ่มตัวอย่าง | แบบทดสอบ  |           |              |
|---|---------------|-----------|-----------|--------------|
|   |               | ฉบับที่ 1 | ฉบับที่ 2 | แบบทดสอบร่วม |
| 1.4.1 รูปแบบผู้สอบกลุ่มเดียว                        |               |           |           |              |
| ก. รูปแบบผู้สอบกลุ่มเดียวที่ไม่ได้รับการจัดให้สมดุล | $P_1$         | X         | Y         |              |
| ข. รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล    | $P_1$         | X         | Y         |              |
|   | $P_2$         | Y         | X         |              |

ภาพประกอบ 3 แผนการเก็บข้อมูลรูปแบบผู้สอบกลุ่มเดียว

#### 1.4.2 รูปแบบผู้สอบกลุ่มเท่าเทียมกัน (Equivalent-group Design)

รูปแบบนี้เป็นการใช้กลุ่มเท่าเทียมกัน โดยการจัดกลุ่มผู้สอบให้มีความคล้ายคลึงกันมากที่สุดเท่าที่จะทำได้ แล้วให้ผู้สอบในแต่ละกลุ่มทำแบบทดสอบฉบับเดียวกัน ถึงแม้การจัดกลุ่มจะไม่เหมือนกันทุกประการ (Identical groups) แต่ก็ควรมีลักษณะทั่วไปคล้ายกันมากที่สุดอาจจะได้มาโดยวิธีการสุ่ม (Random) ข้อดีของรูปแบบนี้คือ แต่ละกลุ่มทำแบบทดสอบฉบับเดียว จึงสามารถหลีกเลี่ยงปัญหา การเรียนรู้ การฝึกฝนและความเมื่อยล้าจากการทำแบบทดสอบ แต่ปัญหาที่ตามมาคือ กลุ่มที่ใช้ไม่เหมือนกันอาจมีการแจกแจงความสามารถที่แตกต่างกัน และไม่มีข้อมูลที่น่ามาปรับความแตกต่างของกลุ่ม ความแตกต่างที่เกิดขึ้นแม้เล็กน้อยย่อมส่งผลกระทบต่อความลำเอียงในการเทียบมาตราได้ วิธีการที่จะช่วยลดความแตกต่างระหว่างกลุ่มได้โดยการใช้กลุ่มตัวอย่างสุ่มขนาดใหญ่

| การเก็บรวบรวมข้อมูล                 | กลุ่มตัวอย่าง | แบบทดสอบ  |           |              |
|-------------------------------------|---------------|-----------|-----------|--------------|
|                                     |               | ฉบับที่ 1 | ฉบับที่ 2 | แบบทดสอบร่วม |
| 1.4.2 รูปแบบผู้สอบกลุ่มเท่าเทียมกัน | $P_1$         | X         |           |              |
|                                     | $P_2$         |           | Y         |              |

ภาพประกอบ 4 แผนการเก็บข้อมูลรูปแบบผู้สอบกลุ่มเท่าเทียมกัน

### 1.4.3 รูปแบบผู้สอบต่างกลุ่มโดยใช้แบบทดสอบร่วม (Anchor-test Design)

ก. รูปแบบผู้สอบกลุ่มสุ่มโดยใช้แบบทดสอบร่วม (Anchor-test Random group Design)

รูปแบบนี้เป็นการใช้กลุ่มผู้สอบซึ่งได้จากการสุ่มแล้วให้ผู้สอบแต่ละกลุ่มสอบแบบทดสอบฉบับเดียวกัน โดยแบบทดสอบแต่ละฉบับมีข้อสอบร่วมกันจำนวนหนึ่ง ซึ่งเรียกว่าแบบทดสอบร่วม (Anchor Test) การใช้แบบทดสอบร่วมนี้สามารถกระทำได้ 2 ลักษณะดังนี้

1. แบบทดสอบร่วมภายใน (Internal Anchor Test) เป็นแบบทดสอบที่มีข้อสอบร่วมที่จัดไว้อยู่ในแบบสอบที่ต้องการนำมาเปรียบเทียบคะแนน

2. แบบทดสอบร่วมภายนอก (External Anchor Test) เป็นแบบที่มีข้อสอบร่วมที่ถูกจัดแยกออกมาต่างหากจากแบบทดสอบที่ต้องการนำมาเปรียบเทียบคะแนน ซึ่งกลุ่มผู้สอบจะต้องทำและจับเวลาแยกออกจากตัวแบบทดสอบที่ต้องการนำมาเปรียบเทียบคะแนน

รูปแบบนี้จะช่วยลดความลำเอียงในการเทียบคะแนนมีสาเหตุมาจากการเรียนรู้ ผีปกฝน ความเมื่อยล้าและความแตกต่างระดับความสามารถของกลุ่มผู้สอบรูปแบบนี้จึงได้เปรียบเหนือกว่า 2 รูปแบบแรก แต่อย่างไรก็ตามประสิทธิภาพของการเทียบคะแนนขึ้นอยู่กับปัจจัยที่สำคัญบางประการ เช่น คุณภาพของแบบทดสอบร่วม ซึ่งจะต้องคำนึงถึง ความสัมพันธ์ระหว่างข้อสอบร่วมที่ต้องมีเนื้อหาและระดับความยากคล้ายกับแบบทดสอบที่ต้องการเทียบคะแนนและความยากของแบบทดสอบร่วมเป็นสำคัญ

ข. รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบทดสอบร่วม (Anchor-test Nonequivalent groups Design)

รูปแบบนี้เป็นการใช้กลุ่มผู้สอบซึ่งมีลักษณะไม่เท่าเทียมกัน แล้วให้ผู้สอบแต่ละกลุ่มทำแบบทดสอบฉบับเดียวกันกลุ่มละฉบับ โดยแบบทดสอบแต่ละฉบับมีข้อสอบร่วมกันจำนวนหนึ่ง ซึ่งเรียกว่า ข้อสอบร่วม (Anchor Test) ซึ่งอาจใช้แบบทดสอบร่วมภายในหรือแบบทดสอบร่วมภายนอกก็ได้ในการทดสอบบางสถานการณ์ อาจมีความจำเป็นต้องใช้แบบทดสอบร่วมกับกลุ่มผู้สอบต่างประชากรกัน เช่น กลุ่มผู้สอบต่างโปรแกรม เวลา หรือระดับ เป็นต้น ซึ่งคะแนนจากแบบทดสอบร่วมจะเป็นกลไกสำคัญสำหรับการเทียบคะแนนระหว่างแบบทดสอบ

| การเก็บรวบรวมข้อมูล                                   | กลุ่ม<br>ตัวอย่าง | แบบทดสอบ  |           |              |
|---|-------------------|-----------|-----------|--------------|
|   |                   | ฉบับที่ 1 | ฉบับที่ 2 | แบบทดสอบร่วม |
| 1.4.3 รูปแบบผู้สอบต่างกลุ่มโดยใช้แบบทดสอบร่วม         |                   |           |           |              |
| ก. รูปแบบผู้สอบกลุ่มสุ่มโดยใช้แบบทดสอบร่วม            |                   |           |           |              |
| (1) แบบทดสอบร่วมภายใน                                 | $P_1$             | XV        |           |              |
|   | $P_2$             |           | YV        |              |
| (2) แบบทดสอบร่วมภายนอก                                | $P_1$             | X         |           | V            |
|   | $P_2$             |           | Y         | V            |
| ข. รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบทดสอบร่วม |                   |           |           |              |
| (1) แบบทดสอบร่วมภายใน                                 | $P_1$             | XV        |           |              |
|   | $Q_2$             |           | YV        | V            |
| (2) แบบทดสอบร่วมภายนอก                                | $P_1$             | X         |           | V            |
|   | $Q_2$             |           | Y         | V            |

ภาพประกอบ 5 แผนการเก็บข้อมูลรูปแบบผู้สอบต่างกลุ่มโดยใช้แบบทดสอบร่วม

#### 4.4 รูปแบบผู้สอบต่างกลุ่มโดยใช้ผู้สอบร่วม (Common-person Design)

รูปแบบนี้เป็นการใช้กลุ่มผู้สอบต่างกลุ่มกันทำแบบทดสอบต่างฉบับแต่มีจำนวนผู้สอบส่วนหนึ่งจากทั้งสองกลุ่มทำแบบสอบทั้ง 2 ฉบับ ซึ่งเป็นรูปแบบของการใช้ผู้สอบร่วม

| การเก็บรวบรวมข้อมูล                        | กลุ่ม<br>ตัวอย่าง | แบบทดสอบ  |           |              |
|--|-------------------|-----------|-----------|--------------|
|  |                   | ฉบับที่ 1 | ฉบับที่ 2 | แบบทดสอบร่วม |
| 1.4.4 รูปแบบผู้สอบต่างกลุ่มโดยมีผู้สอบร่วม |                   |           |           |              |
|  | $P_{1c}$          | X         |           |              |
|  | $P$               | X         | Y         |              |
|  | $Q_1$             |           | Y         |              |
|  | $Q_{1c}$          | X         | Y         |              |

ภาพประกอบ 6 แผนการเก็บข้อมูลรูปแบบผู้สอบต่างกลุ่มโดยมีผู้สอบร่วม

ในการเก็บรวบรวมข้อมูลที่ใช้ในแบบทดสอบหลายฉบับและการเก็บข้อมูลเพื่อการเปรียบเทียบคะแนนแบบทดสอบ สามารถจำแนกการเก็บรวบรวมข้อมูลได้ 4 รูปแบบ คือ รูปแบบผู้สอบกลุ่มเดียว รูปแบบผู้สอบกลุ่มเท่าเทียมกัน รูปแบบผู้สอบต่างกลุ่มโดยใช้แบบทดสอบร่วม และ รูปแบบผู้สอบต่างกลุ่มโดยมีผู้สอบร่วม

### 1.5 รูปแบบวิธีการปรับเทียบคะแนน (Designs for Equating Method)

วิธีการที่ใช้ ในการปรับเทียบคะแนนมีหลายวิธี แต่สามารถ จำแนกตามแนวคิดพื้นฐานของการปรับเทียบได้ เป็นรูปแบบหลัก 2 รูปแบบ (ศิริชัย กาญจนวาสี. 2541 : 11-42) ได้แก่

1.5.1 รูปแบบการปรับเทียบคะแนนโดยใช้ทฤษฎีดั้งเดิม(Classical Models of Equating)

1.5.2 รูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Equating through Item Response Theory)

ซึ่งรูปแบบการปรับเทียบคะแนนแต่ละรูปแบบมีรายละเอียดดังต่อไปนี้

#### 1.5.1 รูปแบบการปรับเทียบโดยใช้ทฤษฎีดั้งเดิม(Classical Models of Equating)

วิธีที่นิยมใช้กันทั่วไปมี 3 วิธีดังนี้

##### 1.5.1.1 วิธีการปรับเทียบคะแนนแบบอิกวิเปอร์ เซ็นไทล์(Equipercetile Equating)

การปรับเทียบคะแนนแบบอิกวิเปอร์ เซ็นไทล์ เป็นวิธีการปรับเทียบคะแนนโดยยึดหลักการพื้นฐานที่ว่า การแจกแจงของคะแนนจากแบบทดสอบ X และแบบทดสอบ Y มีลักษณะคล้ายกัน หรือถ้ามีความแตกต่างเกิดขึ้นบ้างก็มีเพียงเล็กน้อยเช่น ต่างกันที่ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน การเทียบหาคะแนนสมมูลทำได้โดยใช้คะแนน ณ ตำแหน่งเปอร์เซ็นต์เดียวกันของคะแนน 2 ชุดนั้น ผลการปรับเทียบคะแนนแสดงได้ ด้วยกราฟ (ภาวิณี ศรีสุขวัฒนานันท์. 2529 : 30) ภาพที่ได้จากการแปลงคะแนนของวิธีการนี้ ทำให้มองเห็นภาพของระดับความยากของแบบทดสอบทั้งสองชุด นั่นคือ ถ้าหากความยากของแบบทดสอบทั้งสองชุดใกล้เคียงกันเส้นกราฟที่เกิดขึ้นจะใกล้เคียงเส้นตรงและหากว่าความยากต่างกันเส้นกราฟจะเป็นเส้นโค้ง (Curvilinear) ดังนี้

**นิยาม** คะแนนจากแบบทดสอบ 2 ฉบับ X และ Y (เมื่อแบบทดสอบ X และ Y วัดสิ่งเดียวกันและมีความเชื่อมั่นสูงพอ) จะถือว่าเป็นคะแนนสมมูลกัน (Equivalent Scores) เมื่อคะแนนของแต่ละฉบับนั้น มีตำแหน่งเปอร์เซ็นต์ไทล์ (Percentile Ranks) เท่ากัน (ศิริชัย กาญจนวาสี. 2541 : 12 อ้างอิงจาก Flanagan, 1951)

|         |  |
|---------|--|
| นั่นคือ | $X_i$ และ $Y_i$ เป็นคะแนนสมมูล   |
| เมื่อ   | $X_{PRK} = Y_{PRK}$  |
| โดย     | $X_{PRK}$ แทน คะแนนจากแบบทดสอบฉบับ X ตรงตำแหน่งเปอร์เซ็นต์ไทล์ k<br>$Y_{PRK}$ แทน คะแนนจากแบบทดสอบฉบับ Y ตรงตำแหน่งเปอร์เซ็นต์ไทล์ k |

และ

$$X_{PRK} = L_0 + \left[ \frac{\frac{kn}{100} - F_b}{f} \right] (c)$$

|       |       |     |   |
|-------|-------|-----|---|
| เมื่อ | $L_0$ | แทน | คะแนนขีดจำกัดล่างที่แท้จริงของคะแนนในชั้นเปอร์เซ็นต์ไทล์ที่ k |
|       | n     | แทน | จำนวนสมาชิกในกลุ่มตัวอย่าง                                    |
|       | f     | แทน | ความถี่ของคะแนนในชั้นเปอร์เซ็นต์ไทล์ที่ k                     |
|       | $F_b$ | แทน | ความถี่สะสมก่อนถึงคะแนนในชั้นเปอร์เซ็นต์ไทล์ที่ k             |
|       | c     | แทน | อันตรภาคชั้น  |

การเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์ มีนัยามว่า คะแนนจากแบบทดสอบสองฉบับจะถือว่าเป็นคะแนนที่สมมูลกัน เมื่อคะแนนของแต่ละฉบับมีตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน การเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์ประกอบด้วย 2 ขั้นตอน (Petersen, Kolen and Hoover, 1983 : 247-248) ดังนี้

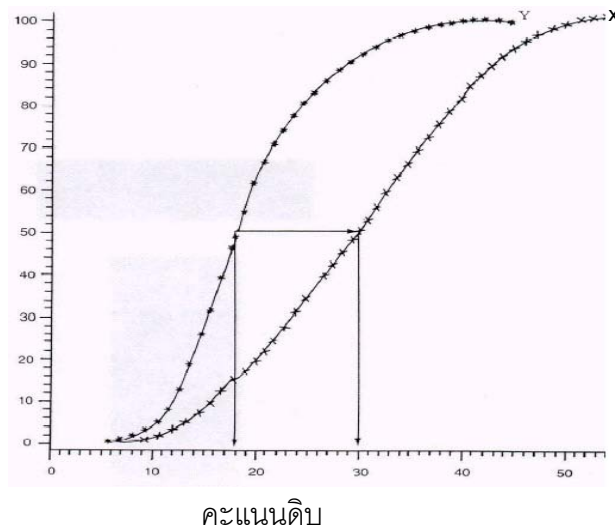
ขั้นตอนที่หนึ่ง การแจกแจงความถี่สะสมสัมพัทธ์สำหรับแบบสอบที่จะนำมาเทียบคะแนน โดยมีขั้นตอนย่อยๆ คือ

1. นำคะแนนของกลุ่มผู้สอบที่มีความสามารถทั้งเก่ง ปานกลาง และอ่อน ซึ่งสุ่มเป็นสองกลุ่มย่อย ให้กลุ่มหนึ่งทำแบบสอบ X และอีกกลุ่มทำแบบสอบ Y มาทำการแจกแจงคะแนน X และ Y

2. คำนวณหาจุดกลางเปอร์เซ็นต์ไทล์ของแต่ละการแจกแจง

3. อ่านและทำเครื่องหมายค่าคะแนนของแบบสอบฉบับ X และฉบับ Y ของการแจกแจงที่สมมูลกันบนกระดาษกราฟ โดยแกนนอนเป็นคะแนนดิบและแกนตั้งเป็นตำแหน่งเปอร์เซ็นต์ไทล์ประมาณ 30 จุด แล้วลากเส้นเชื่อมต่อกันเป็นกราฟ

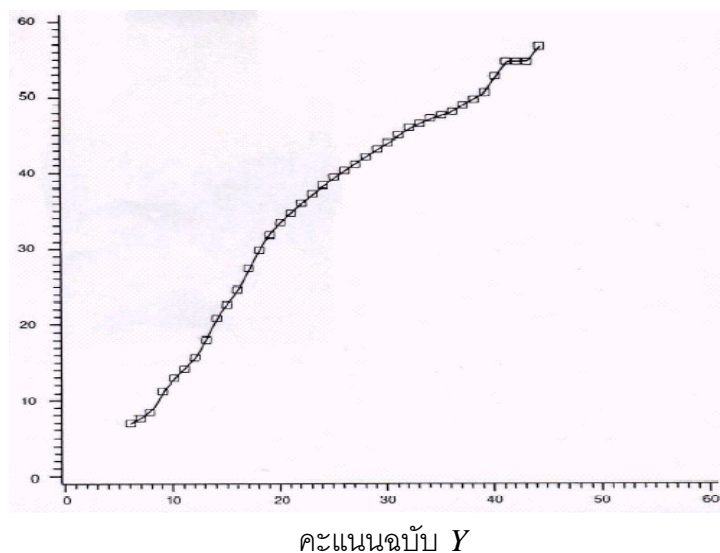
## เปอร์เซ็นต์ไทล์



ภาพประกอบ 7 กระบวนการเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์

ขั้นตอนที่สอง เทียบคะแนนโดยนำคะแนนจากกราฟในขั้นตอนที่หนึ่งมาเขียนกราฟใหม่ลงบนกระดาษกราฟ โดยแกนนอนเป็นคะแนนฉบับ  $Y$  แกนตั้งเป็นคะแนนฉบับ  $X$  ทำการปรับเส้นกราฟให้เรียบ เส้นกราฟที่ได้จะอ่านค่า  $X$  ที่สมมูลกับ  $Y$  จากนั้นสร้างตารางสำเร็จเพื่ออ่านค่าคะแนนแปลงจากกราฟ ดังภาพ

คะแนนฉบับ  $X$



ภาพประกอบ 8 การแปลงคะแนนของการเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์  
การเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์โดยใช้แบบสอบร่วม

ลอร์ดและเลวิน (Lord and Levine อ้างถึงใน วรเนตร พิษิตกริกพล , 2537 :17) ได้เสนอขั้นตอนการเทียบคะแนนวิธีคิวเปอร์เซนไทล์โดยใช้แบบทดสอบร่วมสำหรับกลุ่มที่ไม่ได้สุ่มซึ่งมีความสามารถไม่แตกต่างกันมากนัก โดยที่กลุ่ม  $\alpha$  สอบแบบสอบฉบับ  $X$  และฉบับ  $U$  ส่วนกลุ่ม  $\beta$  สอบแบบสอบฉบับ  $Y$  และฉบับ  $U$  ขั้นตอนในการเทียบคะแนนคือประมาณค่าความถี่ในแบบสอบฉบับ  $X$  และ  $Y$  สำหรับกลุ่ม  $t (\alpha + \beta)$  มีขั้นตอนดังนี้

1. รวมคะแนนของฉบับ  $U$  ที่สอบโดยกลุ่ม  $\alpha, \beta$  และกลุ่ม  $t$
2. หาสัดส่วนความถี่  $f_{it} / f_{i\alpha}, f_{it} / f_{i\beta}$  ที่ทุกช่วงคะแนน  $i$
3. กระจายความถี่ของฉบับ  $U$  ที่ทุกช่วงคะแนนของฉบับ  $X$  และ  $Y$
4. คูณความถี่ของฉบับ  $U$  ในแต่ละช่วงคะแนนของฉบับ  $X$  ด้วยสัดส่วน  $f_{i\alpha} / f_{i\alpha}$
5. คูณความถี่ของฉบับ  $U$  ในแต่ละช่วงคะแนนของฉบับ  $Y$  ด้วยสัดส่วน  $f_{i\beta} / f_{i\beta}$
6. หาตำแหน่งเปอร์เซนไทล์ของความถี่ที่ประมาณค่าใหม่ของทั้งสองฉบับ
7. ดำเนินการเทียบตามขั้นตอนที่สองของการเทียบคะแนนโดยวิธีคิวเปอร์เซนไทล์

ดังกล่าวแล้วข้างต้น

ในการเทียบคะแนนรูปแบบคิวเปอร์เซนไทล์มีข้อจำกัดที่ควรคำนึงอยู่หลายประการดังนี้ (Angoff, 1971 ; Potthoff, 1982 อ้างถึงใน พรพิมล นาคเวช, 2537 : 14 – 15)

1. การสร้างกราฟเปอร์เซนไทล์เพื่อเทียบคะแนนระหว่างแบบสอบ ควรสร้างด้วยความระมัดระวังและไม่ลำเอียง
2. รูปแบบคิวเปอร์เซนไทล์มีความไวต่อความแปรปรวนเชิงสุ่มมาก โดยเฉพาะกลุ่มตัวอย่างที่มีขนาดเล็กจึงควรใช้กลุ่มผู้สอบที่มีขนาดใหญ่และมีความสามารถค่อนข้างกระจาย
3. แบบสอบที่นำมาเทียบคะแนนควรมีความเที่ยงสูงพอกัน ถ้าความเที่ยงต่างกันมากผลการเทียบคะแนนจะขาดความคงที่
4. การเทียบคะแนนระหว่างแบบสอบสองชุดทำได้เฉพาะในช่วงพิสัยของคะแนนที่มีความถี่ของคะแนนสังเกตเพียงพอ ส่วนช่วงที่มีความถี่ของคะแนนน้อยจะมีความคลาดเคลื่อนสูงมาก

### การปรับเทียบคะแนนรูปแบบคิวเปอร์เซนไทล์โดยใช้แบบทดสอบร่วม

ลอร์ด และเลวิน (Angoff. 1984 : 113 – 115 ;citing Lord. 1957 and Levine. 1958) ได้เสนอขั้นตอนการปรับเทียบคะแนนรูปแบบคิวเปอร์เซนไทล์โดยใช้แบบทดสอบร่วมสำหรับกลุ่มที่ได้สุ่มซึ่งมีความสามารถไม่แตกต่างกันมากนัก โดยที่กลุ่ม  $\alpha$  สอบฉบับ  $X$  และฉบับ  $U$  กลุ่ม  $\beta$  สอบฉบับ  $Y$  และฉบับ  $U$  ขั้นตอนในการปรับเทียบคะแนน คือประมาณค่าความถี่ในแบบทดสอบฉบับ  $X$  และ  $Y$  สำหรับกลุ่ม  $t (\alpha + \beta)$  มีขั้นตอนดังนี้

1. รวมคะแนนของฉบับ U ที่สอบโดยกลุ่ม  $\alpha$  ,  $\beta$  และ t
  2. หาสัดส่วนความถี่  $f_{it} / f_{i\alpha}$  และ  $f_{it} / f_{i\beta}$  ที่ทุกช่วงคะแนน i
  3. กระจายความถี่ของฉบับ U ที่ทุกช่วงคะแนนของฉบับ X และ Y
  4. คูณความถี่ของฉบับ U ในแต่ละช่วงคะแนนของฉบับ X ด้วยสัดส่วน  $f_{it} / f_{i\alpha}$
  5. คูณความถี่ของฉบับ U ในแต่ละช่วงคะแนนของฉบับ Y ด้วยสัดส่วน  $f_{it} / f_{i\beta}$
  6. หาค่าแห่งเปอร์เซ็นต์ไทล์ของความถี่ที่ประมาณค่าใหม่ของทั้งสองชุด
  7. ดำเนินการเทียบตามขั้นตอนที่สองของการเปรียบเทียบโดยวิธีอิกวิเปอร์เซ็นต์ไทล์
- ดังกล่าวข้างต้น

### ข้อจำกัดของการเปรียบเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์ มีดังนี้

(ภาวิณี ศรีสุขวัฒน์นันท์, 2529 : 34)

1. ในการเขียนเส้นกราฟที่แสดงคะแนนสมมูลที่ยังไม่ได้ปรับหรือเกลามาให้เรียบจะมีความคลาดเคลื่อนมาก แต่เทคนิคการปรับเส้นเท่าที่ทำกันมานั้นยังไม่มีวิธีใดที่ให้หลักประกันผลได้อย่างน่าเชื่อถือ ลักษณะเช่นนี้ยังก่อให้เกิดความลำเอียง
2. การเปรียบเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นต์ไทล์มีความไวต่อความแปรปรวนเชิงสุ่มมาก โดยเฉพาะกลุ่มตัวอย่างที่มีขนาดเล็ก
3. ในกรณีที่ แบบทดสอบสองชุดมีความเชื่อมั่นต่างกันมากผลของการเปรียบเทียบคะแนนจะขาดความคงที่
4. การเปรียบเทียบคะแนนสมมูลของแบบทดสอบสองชุด ทำได้เฉพาะในช่วงพิสัยของคะแนนที่มีความถี่ของคะแนนสังเกตเท่านั้น ส่วนที่อยู่นอกพิสัยดังกล่าวจะมีความคลาดเคลื่อนของการเปรียบเทียบคะแนนสูงมาก

#### 1.5.1.2 วิธีเปรียบเทียบคะแนนเชิงเส้นตรง (Linear Equating)

การเทียบคะแนนเชิงเส้นตรงมีนิยามว่า สำหรับกลุ่มผู้สอบที่กำหนดคะแนนจากแบบทดสอบฉบับจะถือว่าเท่าเทียมกันถ้าต่างก็ตรงกับคะแนนมาตรฐานเดียวกัน (Angoff, 1971 ; Petersen and Other, 1982 อ้างถึงใน สุนิสา จัยม่วงศรี , 2537 : 13) การเทียบคะแนนเชิงเส้นตรงเป็นวิธีการเทียบคะแนนที่ง่ายและสะดวกที่สุดในการนำไปใช้ (Hills, Subhiyah and Hirsch, 1988 : 229)

แองกอฟฟ์ (Angoff, 1984 อ้างถึงในสุนิสา จัยม่วงศรี , 2537 : 13) ได้เสนอแบบแผนการรวบรวมและการจัดกระทำข้อมูลในทางสถิติของการเทียบคะแนนเชิงเส้นตรงไว้ 6

รูปแบบ ซึ่งแต่ละรูปแบบจะมีวิธีการประมาณค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแบบทดสอบแตกต่างกันออกไป ตามเงื่อนไขของรูปแบบการรวบรวมข้อมูล แต่ทุกรูปแบบจะตัดสินคะแนนสมมูลจากค่าคะแนนมาตรฐานเดียวกันคือ

$$\frac{Y - M_y}{S_y} = \frac{X - M_x}{S_x} \quad \dots\dots\dots(1)$$

จากสมการ (1) สามารถเขียนใหม่ได้ดังนี้

$$(Y - M_y) / S_y = (X - M_x) / S_x$$

เมื่อ  $x, y$  คือ คะแนนจากแบบทดสอบฉบับ X และฉบับ Y

$M_x, M_y$  คือ ค่าเฉลี่ยจากคะแนนแบบทดสอบฉบับ X และฉบับ Y

$S_x, S_y$  คือ ค่าเบี่ยงเบนมาตรฐานของคะแนนจากแบบทดสอบฉบับ X และฉบับ Y

โดยที่คะแนนจากแบบทดสอบฉบับ X และแบบทดสอบฉบับ Y สามารถนำมาเปรียบเทียบเชิงเส้นตรง ได้ดังสมการ

$$Y = A_{yx}X + B_{yx}$$

$$\text{เมื่อ } A_{yx} = \frac{S_y}{S_x} \quad \text{และ } B_{yx} = M_y - A_{yx} M_x$$

ดังนั้น  $Y = A_{yx}X + B_{yx}$  อาจเขียนในรูป

$$Y = \left( \frac{S_y}{S_x} \right) X + \left( M_y - \frac{S_y}{S_x} M_x \right) \dots\dots\dots[2]$$

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

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

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

$$= \frac{S_x}{S_y} Y + \left[ M_x - M_y \frac{S_x}{S_y} \right] \dots\dots\dots[3]$$

### 1.5.1.3 วิธีเปรียบเทียบคะแนนโดยใช้สมการถดถอย(Regression Method)

วิธีเปรียบเทียบคะแนนโดยใช้สมการถดถอย สามารถทำได้ 2 วิธี

(ศิริชัย กาญจนวาสี. 2541: 16-18) ดังนี้

- (1) สร้างสมการทำนายคะแนนของแบบทดสอบฉบับหนึ่งจากคะแนนของแบบทดสอบอีกฉบับหนึ่ง
- (2) สร้างความสัมพันธ์ระหว่างคะแนนจากแบบทดสอบทั้งสองฉบับ โดยใช้เกณฑ์ภายนอก

1. สร้างสมการทำนายคะแนนของแบบทดสอบฉบับหนึ่งจากคะแนนของแบบทดสอบอีกฉบับหนึ่ง

วิธีการนี้ถือว่า คะแนนจากแบบทดสอบ 2 ฉบับ คะแนน  $x_i$  และ  $y_i$  เป็นคะแนนสมมูลกันเมื่อคะแนน  $y_i$  เป็นคะแนนที่ทำนายได้จากคะแนน  $x_i$  ด้วยสมการทำนาย

$$Y_i = T_{(xi)} = B_{yx} + A_{yx}x_i$$

เมื่อ  $T_{(xi)}$  แทน คะแนนแปลงของ  $x_i$  ที่สมมูลกับ  $y_i$

$$A_{yx} \text{ แทน ความชัน (Slope) } = r_{xy} \frac{S_y}{S_x}$$

$$B_{yx} \text{ แทน ค่าคงที่ (Constant) } = \bar{Y} - A_{yx} \bar{X}$$

จะเห็นว่า วิธีเปรียบเทียบคะแนนเชิงเส้นตรง (Linear Equating) มีลักษณะการคำนวณคล้ายกับวิธีการเปรียบเทียบคะแนนโดยสร้างสมการทำนายคะแนนด้วยสมการถดถอย แต่มีวิธีการเก็บรวบรวมข้อมูลแตกต่างกัน

วิธีการเปรียบเทียบคะแนนระหว่างแบบทดสอบ X และแบบทดสอบ Y โดยการสร้างสมการถดถอยทำนายคะแนนของแบบทดสอบ Y จากคะแนนของแบบทดสอบ X สามารถกระทำดังต่อไปนี้

- (1) สุ่มกลุ่มผู้สอบให้เป็นตัวแทนประชากรเป้าหมายที่ต้องการนำแบบทดสอบไปใช้ แล้วให้กลุ่มตัวอย่างทำแบบทดสอบทั้งฉบับ X และ Y
- (2) นำคะแนนของแบบทดสอบแต่ละฉบับมาคำนวณค่า  $\bar{X}$ ,  $S_x$ ,  $\bar{Y}$  และ  $S_y$  แล้วนำมาคำนวณหาค่า  $A_{yx}$  และ  $B_{yx}$  เพื่อสร้างสมการทำนาย

(3) สร้างตารางเทียบคะแนนระหว่างแบบทดสอบ X และ Y โดยแทนค่า  $x_i$  ลงในสมการ เพื่อหาคะแนนสมมูล  $y_i$

การปรับเทียบคะแนนโดยใช้ สมการถดถอยวิธี นี้เป็นการใช้สมการเส้นตรงในการ ทำนายตัวแปรตาม (Dependent Variable) จากตัวแปรอิสระ (Independent Variable) ซึ่งมี ลักษณะที่ไม่สมมาตร คือ สมการทำนาย  $y$  จาก  $x$  หรือสมการทำนาย  $x$  จาก  $y$  เป็นสมการที่ไม่ สมมาตร หรือให้ ผลการทำนายไปในทิศทางเดียวกัน นอกจากนี้คะแนนจากแบบทดสอบที่ใช้ เป็นตัวทำนายยังมีข้อตกลงเบื้องต้นว่าจะต้องมีค่าความเชื่อมั่นเป็น 1

2. สร้างความสัมพันธ์ระหว่างคะแนนจากแบบทดสอบทั้งสองฉบับ โดยใช้ เกณฑ์ภายนอก (External Criterion)

วิธีนี้ถือว่า คะแนนจากแบบทดสอบ 2 ฉบับ คะแนน  $x_i$  กับ  $y_i$  เป็นคะแนนสมมูลกัน เมื่อ คะแนน  $x_i$  และ  $y_i$  ต่างทำนายคะแนนเกณฑ์ภายนอก  $W_i$  เดียวกัน

นั่นคือ  $x_i$  กับ  $y_i$  เป็นคะแนนสมมูลกัน

เมื่อ  $R_x (W/x) = R_y (W/y)$

โดย  $R_x (W/x)$  แทน คะแนนเกณฑ์ภายนอก  $W$  ที่ทำนายได้จาก  $x$  โดยใช้สมการทำนาย

$R_y (W/y)$  แทน คะแนนเกณฑ์ภายนอก  $W$  ที่ทำนายได้จาก  $y$  โดยใช้สมการทำนาย

$$\therefore R_x (W/X) = T_{(yi)} = B_{wx} + A_{wx} X_i$$

$$\text{เมื่อ } A_{wx} = r_{xw} \frac{S_w}{S_x} \text{ และ } B_{wx} = \bar{w} - A_{wx} \bar{x}$$

$$\therefore R_y (W/Y) = T_{(yi)} = B_{wy} + A_{wy} Y_i$$

$$\text{เมื่อ } A_{wy} = r_{yw} \frac{S_w}{S_x} \text{ และ } B_{wy} = \bar{w} - A_{wy} \bar{y}$$

คะแนนเกณฑ์ภายนอก  $W$  อาจเป็นคะแนนของแบบทดสอบมาตรฐาน หรือ แบบทดสอบร่วม ซึ่งเป็นแบบทดสอบที่เป็นอิสระจากแบบทดสอบ X และแบบทดสอบ Y วิธีการปรับเทียบคะแนนระหว่างแบบทดสอบ X และแบบทดสอบ Y โดยใช้เกณฑ์ ภายนอก ด้วยวิธีสร้างสมการถดถอย สามารถกระทำได้ ดังนี้

(1) สุ่มกลุ่มตัวอย่างให้ เป็นตัวแทนประชากรเป้าหมายที่ต้องการนำแบบทดสอบไป ใช้ แล้วทำการสุ่มแยกผู้สอบออกเป็น 2 กลุ่ม โดยให้กลุ่มหนึ่งทำแบบทดสอบฉบับ X และ W ส่วนอีกกลุ่มหนึ่งทำแบบทดสอบฉบับ Y และ W

(2) ทำคะแนนจากแบบทดสอบ X และ W มาคำนวณหาค่า  $S_x$   $S_w$  และ  $r_{xw}$  เพื่อ สร้างสมการทำนายคะแนนเกณฑ์  $W_i$  จาก  $x_i$  และนำคะแนนจากแบบทดสอบ Y และ W มา

คำนวณค่า  $y$ ,  $S_y$ ,  $w$ ,  $S_w$  และ  $r_{yw}$  เพื่อสร้างสมการทำนายคะแนนเกณฑ์  $W_i$  จาก  $y_i$

(3) สร้างกราฟความสัมพันธ์ระหว่าง  $x_i$  และ  $y_i$  สำหรับแต่ละค่าของ  $W_i$  เดียวกัน เพื่อสร้างตารางเทียบคะแนนระหว่างแบบทดสอบ X และแบบทดสอบ Y

ซึ่ง ลอร์ด (ศิริชัย กาญจนวาสี . 2545 : 18 ; อ้างอิงจาก Lord. 1980 : 208 – 209) ได้เสนอแนะว่า การนำไปใช้ให้เกิดความเสมอภาคได้นั้น แบบทดสอบ X และ Y จะต้องคู่ขนานกันอย่างแท้จริง หรือแบบทดสอบทั้งสองฉบับ จะต้องมีความเชื่อมั่นอย่างสมบูรณ์ (Reliability = 1) เนื่องจากการทำนาย Y จาก X หรือทำนาย X จาก Y ในทางวิเคราะห์การถดถอยนั้น จะมีความถูกต้องก็ต่อเมื่อ การวัดค่า  $x$  (หรือ  $y$ ) ได้ถูกต้องสมบูรณ์ปราศจากความคลาดเคลื่อน นอกจากนี้แบบทดสอบ X และ Y จะต้องมีความสัมพันธ์กับคะแนนเกณฑ์เท่ากัน มิเช่นนั้น จะทำให้ความสัมพันธ์ของ X และ Y ที่ได้ ผันแปรไปตามกลุ่มที่ใช้ศึกษา

### ข้อเสนอแนะในการนำวิธีปรับเทียบคะแนนวิธีนี้ไปใช้

(1) แบบทดสอบที่ต้องการนำมาปรับเทียบคะแนนกันต้องเป็นแบบทดสอบที่วัดคุณลักษณะเดียวกัน มีความเป็นคู่ขนานกัน และมีความเชื่อมั่นสูง

(2) แบบทดสอบที่ ต้องการนำมาปรับเทียบคะแนนกัน ต้องมี ความสัมพันธ์กับคะแนนเกณฑ์เท่าเทียมกันมิเช่นนั้นแล้วจะทำให้คะแนนจากแบบทดสอบฉบับหนึ่งจะสามารถทำนายคะแนนเกณฑ์ได้ แม่นยำกว่าอีกฉบับหนึ่ง ทำให้การปรับเทียบคะแนนมีความผันแปรตามกลุ่มที่ใช้ศึกษา

### 1.5.2 รูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ(Equating Through Item Response Theory)

รูปแบบวิธีการปรับเทียบตามแนวทฤษฎีดั้งเดิมที่ใช้คะแนนดิบมาปรับเทียบคะแนนดิบ (raw score) ระหว่างแบบสอบ แต่มีจุดอ่อนทั้งในด้านความเสมอภาค (equity) ความสมมาตร (symmetry) และความไม่ผันแปรตามกลุ่ม (invariance) ซึ่งสามารถแก้ปัญหาดังกล่าวได้ด้วยวิธีการปรับเทียบคะแนนตามแนวทางทฤษฎีการตอบสนองข้อสอบ(IRT) ถ้ามีการเลือกใช้โมเดลการตอบสนองข้อสอบ (item response models) ที่มีความสอดคล้องกับข้อมูล (Kolen, 1981 อ้างถึงใน ศิริชัย กาญจนวาสี, 2541: 18-19)

IRT ได้เสนอโมเดลการตอบสนองข้อสอบที่แสดงความสัมพันธ์ระหว่างความสามารถของผู้สอบ ( $\theta$ ) กับโอกาสของการตอบข้อสอบถูกเป็นรายข้อ  $[P_i(\theta)]$  โดย  $P_i(\theta)$  เป็นฟังก์ชันของ  $\theta$

และคุณลักษณะของข้อสอบ (ค่าอำนาจจำแนก ความยาก และอำนาจการเดา หรือแทนด้วยค่าพารามิเตอร์  $a$   $b$  และ  $c$ ) ซึ่งค่าพารามิเตอร์ของข้อสอบไม่ผันแปรตามกลุ่ม จะประมาณความสามารถ  $\theta$  ของผู้สอบได้อย่างอิสระจากชุดของข้อสอบ เมื่อทราบค่าพารามิเตอร์ของข้อสอบ

ในการประมาณค่า  $\theta$  ค่าประมาณ  $\theta$  ของผู้สอบแต่ละคน จะไม่ได้รับผลกระทบจากความแตกต่างของชุดข้อสอบที่ใช้ ไม่ว่าผู้สอบจะถูกทดสอบด้วยแบบทดสอบที่ยาก ปานกลาง หรือง่าย ความสามารถ  $\theta$  ของผู้สอบที่ประมาณได้จากรูปแบบ IRT จึงสามารถเปรียบเทียบกันได้ ภายใต้ความคลาดเคลื่อนอย่างสุ่ม ดังนั้นการทราบค่าพารามิเตอร์ของข้อสอบ จึงเปรียบเสมือนว่าเมตริกซ์ความสามารถ  $\theta$  ของผู้สอบถูกกำหนดให้อยู่บนสเกลเดียวกัน จึงสามารถเปรียบเทียบกันได้โดยตรง จึงไม่จำเป็นในการปรับเทียบคะแนนระหว่างแบบสอบ (Hambleton and Swaminathan, 1985 อ้างถึงใน ศิริชัย กาญจนวาสี, 2541: 19)

กรณีที่ไม่ทราบค่าพารามิเตอร์ของข้อสอบมาก่อน เมื่อนำแบบสอบฉบับเดียวกันไปทดสอบกับผู้สอบ 2 กลุ่ม ฟังก์ชันการตอบสนองของข้อสอบจะมีคุณลักษณะไม่ผันแปรตามกลุ่ม ถ้าในการคำนวณได้มีการกำหนดหรือตรึงค่า (fix) ความสามารถของกลุ่มผู้สอบ ( $\theta$ ) หรือค่าพารามิเตอร์ความยากของข้อสอบ ( $b$ ) ให้อยู่บนสเกลที่มีค่าเฉลี่ย ( $\bar{X}$ ) และส่วนเบี่ยงเบนมาตรฐาน (SD) เท่ากัน

ในกรณีที่ไม่นำทั้งค่าพารามิเตอร์ของข้อสอบและพารามิเตอร์ของผู้สอบมาก่อน เมื่อนำแบบสอบฉบับเดียวกันไปทดสอบกับผู้สอบ 2 กลุ่ม (ข้อสอบร่วม) หรือผู้สอบกลุ่มเดียวกันทำแบบสอบ 2 ฉบับ (ผู้สอบร่วม) แล้วนำข้อมูลที่ได้มาวิเคราะห์แยกกัน ทฤษฎีการตอบสนองของข้อสอบจะประมาณค่าพารามิเตอร์ของข้อสอบ และความสามารถของผู้สอบ ผลที่ได้อาจได้ค่าที่แตกต่างกันระหว่างแบบสอบหรือระหว่างกลุ่มผู้สอบ ความแตกต่างของค่าพารามิเตอร์ที่เกิดขึ้นเนื่องมาจากการกำหนดค่าเฉลี่ยของการแจกแจงค่าความสามารถของผู้สอบ ( $\theta$ ) หรือ ค่าเฉลี่ยของการแจกแจงค่าพารามิเตอร์ความยากของข้อสอบ ( $b$ ) ในการวิเคราะห์แต่ละครั้งที่ไม่เหมือนกัน แต่ค่าพารามิเตอร์ระหว่างฉบับหรือระหว่างกลุ่มเหล่านั้น จะมีความสัมพันธ์เชิงเส้นตรง (linear relationship) ซึ่งสามารถสร้างสมการแปลงค่าพารามิเตอร์ระหว่างฉบับหรือระหว่างกลุ่มให้อยู่บนสเกลเดียวกันได้

การปรับเทียบคะแนนระหว่างแบบสอบโดยใช้ทฤษฎีการตอบสนองของข้อสอบ มีหลายวิธี โดยพิจารณาให้สอดคล้องกับรูปแบบของการเก็บรวบรวมข้อมูล และการใช้โปรแกรมคอมพิวเตอร์ ในการวิเคราะห์ข้อมูล เช่น LOGIST , BILOG หรือโปรแกรมของ คณิต ไช่มุกด์ เป็นต้น วิธีการปรับเทียบโดยใช้ทฤษฎีการตอบสนองของข้อสอบ สามารถจำแนกเป็นกลุ่มวิธีตามรูปแบบที่ใช้ในการเก็บรวบรวมข้อมูล ซึ่งมีรายละเอียดดังนี้

### 1) รูปแบบผู้สอบกลุ่มเดียว ( Single Group Design)

ผู้สอบกลุ่มเดียวกันทำแบบสอบทั้ง 2 ฉบับที่ต้องการนำคะแนนมาเปรียบเทียบกัน สามารถใช้วิธีการเปรียบเทียบวิธีใดวิธีหนึ่ง ดังนี้

- 1.1 การเปรียบเทียบโดยการวิเคราะห์ข้อมูลจากแบบสอบทั้ง 2 ฉบับร่วมกัน มีขั้นตอน ดังนี้
  - (1) นำแบบทดสอบทั้ง 2 ฉบับที่ต้องการเปรียบเทียบคะแนนความสามารถกัน ไปทดสอบกับกลุ่มผู้สอบกลุ่มเดียวกัน
  - (2) นำผลการตอบข้อสอบจากแบบสอบทั้ง 2 ฉบับมาต่อกันเป็นเมตริกซ์ข้อมูลเดียวกัน แล้วทำการวิเคราะห์ค่าพารามิเตอร์ของข้อสอบและค่าพารามิเตอร์ความสามารถของผู้สอบพร้อมกันไป
  - (3) ผลการวิเคราะห์ ทำให้ได้ค่าประมาณความสามารถของผู้สอบแต่ละคน และค่าพารามิเตอร์ของข้อสอบแต่ละข้อ ซึ่งอยู่บนสเกลร่วมกัน ( Common Scale) จึงไม่จำเป็นต้องปรับค่าระหว่างกลุ่มผู้สอบ ค่าที่ได้จากวิธีนี้จึงไม่ต้องนำมาสร้างสมการเปรียบเทียบ สามารถใช้เปรียบเทียบกันได้โดยตรง
  
- 1.2 การเปรียบเทียบโดยการวิเคราะห์ข้อมูลจากแบบสอบ 2 ฉบับแยกจากกัน มีวิธีการดังนี้
  - (1) นำแบบทดสอบทั้ง 2 ฉบับที่ต้องการเปรียบเทียบคะแนนความสามารถกัน ไปทดสอบกับกลุ่มผู้สอบกลุ่มเดียวกัน
  - (2) นำผลการตอบข้อสอบจากแบบสอบแต่ละฉบับ มาวิเคราะห์แยกสำหรับแต่ละฉบับ โดยกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยากของข้อสอบ ( b) ของแต่ละแบบสอบ ในการวิเคราะห์แต่ละครั้งให้เท่ากัน เช่น  $\bar{b}=0, S_b=1$  เป็นต้น
  - (3) สร้างสมการความสัมพันธ์เชิงเส้นตรง สำหรับปรับค่า ระหว่างแบบสอบทั้ง 2 ให้มีความสมมูลกัน

#### กรณีโมเดล 1-พารามิเตอร์ (1-Parameter Model)

ถ้ามีการกำหนดค่าเฉลี่ยของ  $b$  ให้เป็น 0 ( $\bar{b}=0$ ) ในการวิเคราะห์สำหรับแต่ละแบบสอบ ความสัมพันธ์ระหว่างคะแนนความสามารถจากแบบสอบ  $X(\theta_X)$  กับคะแนนความสามารถจากแบบสอบ  $Y(\theta_Y)$  เป็นดังนี้

$$\theta_X - \mu_{\theta_X} = \theta_Y - \mu_{\theta_Y}$$

$$\therefore \theta_Y = \theta_X + (\mu_{\theta_Y} - \mu_{\theta_X})$$

$$\text{นั่นคือ } \theta_Y = T_{(\theta_X)} = \theta_X + (\mu_{\theta_Y} - \mu_{\theta_X})$$

$$\text{หรือ } \theta_Y = T_{(\theta_X)} = \alpha\theta_X + \beta$$

$$\text{เมื่อ } \alpha = 1$$

$$\text{และ } \beta = \mu_{\theta_Y} - \mu_{\theta_X}$$

ในเมื่อ  $T_{\theta(X)}$  แทน คะแนนความสามารถที่แปลงแล้วของ  $\theta_X$  ซึ่งสมมูลกับ  $\theta_Y$

$\mu_{\theta_X}$  แทน ค่าเฉลี่ยความสามารถของผู้สอบฉบับ X

$\mu_{\theta_Y}$  แทน ค่าเฉลี่ยความสามารถของผู้สอบฉบับ Y

### กรณีโมเดล 2 และ 3-พารามิเตอร์ (2- and 3 -Parameter Model)

ถ้ามีการกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของ  $b$  ให้เป็น 0 และ 1 ( $\bar{b}=0, S_b = 1$ ) ความสัมพันธ์ระหว่างคะแนนความสามารถจากแบบสอบ X ( $\theta_X$ ) กับคะแนนความสามารถจากแบบสอบ Y ( $\theta_Y$ ) เป็นดังนี้

$$\theta_X - \mu_{\theta_X} / \sigma_{\theta_X} = \theta_Y - \mu_{\theta_Y} / \sigma_{\theta_Y}$$

$$\theta_X - \mu_{\theta_X} \frac{\sigma_{\theta_Y}}{\sigma_{\theta_X}} = \theta_Y - \mu_{\theta_Y}$$

$$\theta_X = \frac{\sigma_{\theta_Y}}{\sigma_{\theta_X}} \theta_X + \left[ \mu_{\theta_Y} - \frac{\sigma_{\theta_Y}}{\sigma_{\theta_X}} \mu_{\theta_X} \right]$$

$$\text{นั่นคือ } \theta_Y = T_{(\theta_X)} = \alpha\theta_X + \beta$$

$$\text{เมื่อ } \alpha = \frac{\sigma_{\theta_Y}}{\sigma_{\theta_X}}$$

$$\text{และ } \beta = \mu_{\theta_Y} - \frac{\sigma_{\theta_Y}}{\sigma_{\theta_X}} \mu_{\theta_X}$$

ในเมื่อ

$T_{\theta(X)}$  แทน คะแนนความสามารถที่แปลงแล้วของ  $\theta_X$  ซึ่งสมมูลกับ  $\theta_Y$

$\mu_{\theta_X}$  แทน ค่าเฉลี่ยความสามารถของผู้สอบฉบับ X

$\mu_{\theta_Y}$  แทน ค่าเฉลี่ยความสามารถของผู้สอบฉบับ Y

$\sigma_{\theta_X}$  แทน ค่าส่วนเบี่ยงเบนมาตรฐานความสามารถของผู้สอบฉบับ X

$\sigma_{\theta_Y}$  แทน ค่าส่วนเบี่ยงเบนมาตรฐานความสามารถของผู้สอบฉบับ Y

ความสัมพันธ์เชิงเส้นของ  $\theta_x$  และ  $\theta_y$  ที่เกิดขึ้นเป็นผลเนื่องมาจากการใช้ฟังก์ชันความสัมพันธ์ของทฤษฎีการตอบสนองข้อสอบในการวิเคราะห์ ซึ่งต่างจากการปรับเทียบคะแนนระหว่างแบบสอบด้วยวิธีเชิงเส้น (Linear Equating) ตามทฤษฎีการทดสอบแบบดั้งเดิม ซึ่งถือว่าความสัมพันธ์ระหว่าง X และ Y เป็นข้อตกลงเบื้องต้น

วิธีที่ 2 มีการกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความสามารถของผู้สอบ ( $\theta$ )

การปรับเทียบคะแนนระหว่างแบบสอบโดยวิธีนี้มีขั้นตอนของการปรับเทียบ ดังนี้

- (1) นำแบบสอบทั้ง 2 ฉบับที่ต้องการปรับเทียบคะแนนความสามารถกันไปทำการทดสอบกับกลุ่มผู้สอบกลุ่มเดียวกัน
- (2) นำผลการตอบข้อสอบจากแบบสอบแต่ละฉบับ มาวิเคราะห์แยกสำหรับแต่ละฉบับ โดยกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความสามารถผู้สอบ ( $\theta$ ) ในการวิเคราะห์แต่ละครั้งให้เท่ากัน เช่น  $\mu_{\theta_x} = \mu_{\theta_y} = 0$  และ  $\sigma_{\theta_x} = \sigma_{\theta_y} = 1$  เป็นต้น
- (3) จากสมการปรับเทียบข้างต้น เมื่อกำหนดเงื่อนไขตามข้อ 2) จึงได้  $\alpha=1$  และ  $\beta = 0$  ทำให้ค่าความสามารถที่วิเคราะห์ได้จากแบบสอบฉบับหนึ่งกับค่าความสามารถที่วิเคราะห์ได้จากแบบสอบอีกฉบับหนึ่งเป็นค่าที่สมมูลกัน ( $\theta_x = \theta_y$ ) ดังนั้นการปรับเทียบกันจึงไม่มีความจำเป็น

## 2) รูปแบบผู้สอบกลุ่มเท่าเทียมกัน (Equivalent Group Design)

ในการออกแบบการเก็บรวบรวมข้อมูลใช้กลุ่มเท่าเทียมกัน โดยให้กลุ่มหนึ่งทำแบบทดสอบฉบับ X และอีกกลุ่มหนึ่งทำแบบทดสอบฉบับ Y การออกแบบวิธีนี้จึงไม่มีทั้งข้อสอบร่วมและผู้สอบร่วม การปรับเทียบคะแนนความสามารถ จึงดำเนินการตามขั้นตอนต่อไปนี้

- (1) สุ่มกลุ่มผู้สอบ 2 กลุ่ม ที่มีจำนวนผู้สอบเท่ากัน ให้แต่ละกลุ่มทำแบบสอบแต่ละฉบับ โดยกลุ่มหนึ่งทำฉบับ X และอีกกลุ่มหนึ่งทำฉบับ Y
- (2) นำผลการตอบข้อสอบจากแบบสอบแต่ละฉบับมาวิเคราะห์แยกกัน โดยกำหนดค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของความสามารถผู้สอบ ( $\theta$ ) ของแต่ละกลุ่มให้เท่ากันสำหรับวิเคราะห์ในแต่ละครั้ง เช่น  $\bar{\theta} = 0$ ,  $S_{\theta} = 1$  เป็นต้น
- (3) เรียงลำดับค่าประมาณความสามารถที่ได้จากแต่ละฉบับ และจับคู่กัน ( $\theta_x$ ,  $\theta_y$ ) เป็นคู่ ๆ จากค่าต่ำสุดถึงสูงสุด
- (4) สร้างความสัมพันธ์ระหว่าง  $\theta_x$  และ  $\theta_y$  โดยการพล็อตค่าของ  $\theta_x$  และ  $\theta_y$  เป็นรายคู่

ถ้าข้อมูลที่ได้เป็นไปตามข้อตกลงเบื้องต้นตามทฤษฎีการตอบสนองรายข้อ การพล็อตค่า  $\theta_x$ , กับ  $\theta_y$  จะเป็นเส้นตรง แต่ในทางปฏิบัติอาจไม่เป็นเช่นนั้น เนื่องจากมีความคลาดเคลื่อนของการเปรียบเทียบเกิดขึ้น ความคลาดเคลื่อนจะเกิดมากขึ้น สำหรับค่า  $\theta$  ที่อยู่สุดขั้ว

### 3) รูปแบบผู้สอบต่างกลุ่มโดยใช้แบบสอบร่วม (Anchor Test Design)

เมื่อการออกแบบการเก็บรวบรวมข้อมูลมีการใช้ผู้สอบต่างกลุ่มทำแบบสอบต่างฉบับกัน กลุ่มหนึ่งทำฉบับ X และอีกกลุ่มหนึ่งทำฉบับ Y โดยแบบสอบแต่ละฉบับมีข้อสอบร่วมกันจำนวนหนึ่ง (Common Items) ซึ่งเรียกว่าแบบสอบร่วม (Anchor Test) การเปรียบเทียบคะแนนทำได้ 2 วิธี ดังนี้

3.1 การเปรียบเทียบโดยการวิเคราะห์ข้อมูลจากแบบสอบ 2 ฉบับแยกกัน มีขั้นตอนดังต่อไปนี้

(1) ใช้กลุ่มผู้สอบต่างกลุ่มกัน 2 กลุ่ม ซึ่งอาจมีการแจกแจงความสามารถที่แตกต่างกัน ให้กลุ่มหนึ่งทำแบบสอบฉบับ X และอีกกลุ่มหนึ่งทำแบบสอบฉบับ Y โดยแบบสอบฉบับ X และ Y มีข้อสอบร่วมกันจำนวนหนึ่ง

(2) นำผลการตอบข้อสอบจากแบบสอบแต่ละฉบับมาวิเคราะห์แยกกัน โดยกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความสามารถผู้สอบ ( $\theta$ ) ของแต่ละกลุ่มให้เท่ากัน สำหรับการวิเคราะห์แต่ละครั้ง

(3) นำค่าพารามิเตอร์ของข้อสอบร่วมเป็นรายข้อที่คำนวณได้จากแต่ละกลุ่มมาจับเป็นคู่ ๆ เช่น ค่าความยากของข้อสอบ ( $b_x, b_y$ ) ค่าอำนาจจำแนกของข้อสอบ ( $a_x, a_y$ )

เนื่องจากค่าพารามิเตอร์ของข้อสอบร่วมที่วิเคราะห์ได้จากผู้สอบ 2 กลุ่มจะมีความสัมพันธ์เชิงเส้นตรงต่อกัน ดังสมการ

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

$$a_y = \frac{a_x}{\alpha}$$

เมื่อ 
$$\alpha = \frac{\sigma_{by}}{\sigma_{bx}}$$

$$\beta = \mu_{by} - \alpha \mu_{bx}$$

ความสัมพันธ์ระหว่างค่าพารามิเตอร์คล้ายรูปแบบที่ 1 (รูปแบบผู้สอบกลุ่มเดียว) ที่กล่าวมาแล้ว แต่ต่างกันตรงที่ได้สารสนเทศเพิ่มเติมขึ้นมา คือ ค่าความชันของเส้นแสดงความสัมพันธ์

ระหว่างความยาก ( $\alpha$ ) เป็นส่วนกลับของค่าความชันของเส้นแสดงความสัมพันธ์ระหว่างค่าอำนาจจำแนก ( $1/\alpha$ )

(4) นำสมการที่ได้จากข้อ 3) มาใช้ปรับค่าพารามิเตอร์ของข้อสอบทั้ง 2 ฉบับให้อยู่บนสเกลร่วมกัน เมื่อทราบค่าพารามิเตอร์ของข้อสอบแต่ละฉบับ นำแบบสอบไปใช้ก็จะทำให้การประมาณค่าความสามารถของผู้สอบ ( $\theta$ ) ไม่ว่าจะใช้ฉบับใด ย่อมอยู่บนสเกลเดียวกัน

3.2 การเปรียบเทียบโดยการวิเคราะห์ข้อมูลจากแบบสอบ 2 ฉบับรวมกัน มีขั้นตอนการเปรียบเทียบ ดังนี้

(1) ใช้กลุ่มผู้สอบต่างกลุ่มกัน 2 กลุ่ม ทำแบบสอบต่างฉบับกัน โดยแบบสอบแต่ละฉบับมีข้อสอบร่วมกันจำนวนหนึ่ง

(2) นำผลการตอบข้อสอบจากแบบสอบแต่ละฉบับมาวิเคราะห์ร่วมกัน โดยจัดกระทำ ดังนี้

(2.1) ในการจัดข้อมูลให้ทำเสมือนว่ามีผู้สอบ  $N_x + N_y$  คน โดยทุกคนทำข้อสอบ  $n_x + n_y + n_a$  ข้อ

(2.2) จัดดำเนินการการลงรหัสข้อมูลในลักษณะที่ผู้สอบ  $N_x$  คน ตอบข้อสอบ  $n_x + n_a$  ข้อ แต่ไม่ได้ตอบข้อสอบ  $n_y$  ข้อและผู้สอบ  $N_y$  คน ตอบข้อสอบ  $n_y + n_a$  ข้อ แต่ไม่ได้ตอบข้อสอบ  $n_x$  ข้อ

(2.3) ใช้โปรแกรม LOGIST ประมาณค่าความสามารถของผู้สอบ  $N_x + N_y$  คน และค่าพารามิเตอร์ของข้อสอบจำนวน  $n_x + n_y + n_a$  ข้อ พร้อมกันไป

ผลการวิเคราะห์ที่ได้ จะทำให้ค่าพารามิเตอร์ของข้อสอบอยู่บนสเกลร่วมกันและมีความสมมูลกัน

#### 4. รูปแบบผู้สอบร่วม ( Common – Person Design)

เมื่อการออกแบบการเก็บรวบรวมข้อมูลมีการใช้ “ผู้สอบร่วม” โดยให้ผู้สอบสองกลุ่มทำแบบสอบต่างฉบับกัน กลุ่มหนึ่งทำฉบับ X และอีกกลุ่มหนึ่งทำฉบับ Y แต่มีผู้สอบส่วนหนึ่งของทั้งสองกลุ่มทำแบบสอบทั้งฉบับ X และ Y จากลักษณะของการใช้ “ผู้สอบร่วม” ซึ่งเป็นอีกลักษณะหนึ่งของรูปแบบที่ 1 การเปรียบเทียบคะแนนความสามารถสำหรับวิธีนี้ มีขั้นตอน ดังนี้

(1) ใช้กลุ่มผู้สอบต่างกลุ่มกัน 2 กลุ่ม ให้กลุ่มหนึ่งทำแบบสอบฉบับ X และอีกกลุ่มหนึ่งทำแบบสอบฉบับ Y โดยมีผู้สอบส่วนหนึ่งของทั้งสองกลุ่มทำแบบสอบทั้งสองฉบับ X และ Y

(2) จากกลุ่มผู้สอบร่วม ซึ่งทำแบบสอบทั้ง 2 ฉบับ เราสามารถนำข้อมูลมาวิเคราะห์ร่วมกันเพื่อคำนวณค่าพารามิเตอร์ของแบบสอบทั้ง 2 ฉบับให้อยู่บนสเกลร่วมกัน

ทั้งนี้ในการคำนวณจะต้องมีการกำหนดค่าการแจกแจงความสามารถของผู้สอบให้คงที่ เช่น  $\theta = 0$  และ  $S_\theta = 1$

- (3) จากค่าพารามิเตอร์ของข้อสอบทั้ง 2 ฉบับ ซึ่งอยู่บนสเกลร่วมกันแล้วจึงสามารถนำไปใช้ในการสอบเพื่อประมาณค่าความสามารถของผู้สอบ ไม่ว่าจะสอบฉบับใดก็ตามจะได้คะแนนความสามารถที่สมมูลกัน

### รูปแบบวิธีการเทียบคะแนนตามรูปแบบ IRT

การออกแบบวิธีการเทียบคะแนนระหว่างแบบทดสอบเป็นการกำหนดเงื่อนไขและกฎเกณฑ์การแปลงมาตราจากแบบทดสอบต่างชุดที่ต้องการนำมาเทียบกันให้สามารถเทียบกันได้ อย่างมีความหมาย วิธีการที่ใช้ในการเทียบคะแนนมีหลายวิธีสามารถจำแนกตามความคิดพื้นฐานของการเทียบได้เป็นแบบหลัก 2 รูปแบบ คือ การเทียบคะแนนรูปแบบดั้งเดิม และการเทียบคะแนนรูปแบบอิงทฤษฎีการตอบสนองข้อสอบ ซึ่งทั้ง 2 รูปแบบมีหลักการในการเทียบคะแนนที่เหมือนกัน เพียงแต่วิธีการในการเทียบคะแนนต่างกัน รูปแบบวิธีการเทียบคะแนนมีดังนี้

ลอร์ด (Lord, 1980 อ้างถึงใน สุนิสา จัยม่วงศรี, 2537: 22) ได้กล่าวถึงการเทียบคะแนนโดยใช้ทฤษฎีการตอบสนองข้อสอบไว้ว่า แบบสอบสองฉบับใด ๆ ที่ใช้ในการเทียบคะแนนจะต้องเป็นแบบสอบที่วัดคุณลักษณะเดียวกันและเทียบมาตราคะแนนที่สมมูลกันที่ระดับความสามารถเดียวกัน ซึ่งมีข้อกำหนด 3 ประการดังนี้

1. ความเสมอภาค (Equity) หมายถึง สำหรับทุกๆ ระดับความสามารถ ( $\theta$ ) การแจกแจงความถี่แบบมีเงื่อนไขของคะแนนแปลงต้องเหมือนกับการแจกแจงความถี่แบบมีเงื่อนไขของคะแนนแบบสอบที่ต้องการเทียบ
2. ความไม่ผันแปรตามกลุ่ม (Invariance across Groups) หมายถึง คะแนนที่แปลงแล้วจะคงที่โดยไม่แปรเปลี่ยนไปตามกลุ่มประชากรที่นำมาเทียบจึงสามารถนำผลไปใช้กับกลุ่มอื่นๆ ได้
3. ความสมมาตร (Symmetry) หมายถึง คะแนนแปลงที่ได้จากการเทียบคะแนนจะต้องเหมือนกัน ไม่ว่าจะเป็นการปรับเทียบคะแนนจากแบบสอบฉบับ ฉบับ  $X$  ไปยังแบบสอบฉบับ  $Y$  หรือจากแบบสอบฉบับ  $Y$  ไปยังแบบสอบฉบับ  $X$

การเทียบคะแนนรูปแบบทฤษฎีการตอบสนองข้อสอบสามารถแบ่งได้ 2 วิธีการใหญ่ๆ ดังนี้ (Lord, 1980 อ้างถึงใน สุนิสา จัยม่วงศรี, 2537: 23 – 24)

- 1) การเทียบคะแนนจากคะแนนจริง (True-Score Equating)

จากผลการเปรียบเทียบความสามารถ ( $\theta$ ) บนสเกลของคะแนนมาตรฐาน (Standard score) ซึ่งมีค่าเป็นไปได้อันทั้งค่าบวกและค่าลบ โดยมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของการแจกแจงเป็น 0 และ 1 ตามลำดับ นั้น อาจส่งผลให้เกิดความสับสนในการแปลความหมาย จึงเกิดเป็นแนวคิดของการแปลงคะแนนความสามารถ ( $\theta$ ) ให้อยู่ในรูปของคะแนนจริง (true score) เพื่อให้มีความสอดคล้องกับสภาพความเป็นจริง และเข้าใจได้ง่ายขึ้น

แบบสอบที่มีความยากต่างกันถึงแม้ว่าจะนำไปสอบกับประชากรที่มีความสามารถเท่ากันหรือคุณลักษณะเดียวกัน แต่คะแนนที่ได้จะมีการแจกแจงแตกต่างกัน ดังนั้น ถ้าสามารถหาคะแนนจริงได้ ก็จะสามารถเทียบคะแนนระหว่างความสามารถ ( $\theta$ ) กับคะแนนจริงได้จากความสัมพันธ์เชิงคณิตศาสตร์ดังนี้

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

$$\eta = \eta(\theta) = \sum_{i=1}^m P_i(\theta_y)$$

เมื่อ  $\xi$  คือ คะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับ X

$\eta$  คือ คะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับ Y

m คือ จำนวนข้อของแบบสอบฉบับ X

n คือ จำนวนข้อของแบบสอบฉบับ Y

$\theta_x$  คือ ค่าความสามารถของผู้สอบที่ได้จากแบบสอบฉบับ X

$\theta_y$  คือ ค่าความสามารถของผู้สอบที่ได้จากแบบสอบฉบับ Y

ดังนั้น ถ้าแทนค่าความสามารถ ( $\theta$ ) ใด ๆ ลงในสมการที่  $\theta$  เท่ากันจะหาคะแนนสมมูลของแบบสอบฉบับ X และฉบับ Y ได้โดยคำนวณจากค่า  $P_i(\theta)$  และ  $P_j(\theta)$  และในทางตรงกันข้าม ก็สามารถประมาณค่าความสามารถ ( $\theta$ ) เมื่อทราบคะแนนจริง

การหาค่า  $P_i(\theta)$  และ  $P_j(\theta)$  ได้จากการประมาณค่าพารามิเตอร์, สองพารามิเตอร์ หรือสามพารามิเตอร์จากนั้นนำคะแนนจริงของแบบสอบฉบับ X และฉบับ Y หาความสัมพันธ์กันโดยใช้ค่าความสามารถ ( $\theta$ ) ที่ระดับเดียวกัน

## 2) การเทียบคะแนนจากคะแนนที่สังเกต (Observed – Score Equating)

ปัญหาของการเทียบคะแนนโดยใช้คะแนนจริงคือไม่สามารถทราบคะแนนจริงของแต่ละคนได้นอกจากใช้วิธีประมาณจากผลการสอบ ซึ่งคะแนนที่ได้เป็นเพียงค่าประมาณเท่านั้น ยังไม่มีคุณสมบัติเป็นคะแนนจริงและไม่สามารถทำการเทียบคะแนนคะแนนที่ต่ำกว่าคะแนนการเดาได้ (ศิริชัย กาญจนวาสี, 2541 : 30) ลอร์ด (Lord, 1980 อ้างถึงใน ภาวิณี ศรีสุข

วัฒนานันท์, 2529 : 43) จึงเสนอการเทียบคะแนนโดยใช้คะแนนสังเกต โดยอาศัยข้อมูลจากการตอบแบบสอบถาม (Anchor test)

การแจกแจงทางทฤษฎีของคะแนนที่สังเกตได้ของผู้สอบที่มีความสามารถ  $\theta$  จากการทำแบบสอบฉบับหนึ่ง สามารถแสดงได้ดังนี้ (Hambleton & Swaminathan, 1985 : 215)

$$\sum_{r=0}^n f(r|\theta)r^r = \prod_{i=1}^n [Q_i(\theta) + tP_i(\theta)]$$

ขั้นตอนในการเทียบคะแนนคะแนนที่สังเกตได้โดยใช้ทฤษฎีการตอบสนองข้อสอบ สามารถสรุปเป็นขั้นตอนได้ดังนี้ (ศิริชัย กาญจนวาสี, 2545 : 164)

1. จากกลุ่มผู้สอบแบบสอบฉบับ  $X$  นำข้อมูลวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบ แล้วนำค่า  $\theta_x$  ของผู้สอบแต่ละคนมาคำนวณความถี่สัมพัทธ์อย่างมีเงื่อนไขทางทฤษฎีของการแจกแจงคะแนนดิบ  $f_x(r|\theta_x)$
2. แปลงค่า  $\theta_x$  ให้สมมูลกับ  $\theta_y$  ด้วยสูตร  $T(\theta_y) = \alpha\theta_x + \beta$  แล้วนำ  $T(\theta_x)$  ของกลุ่มผู้สอบมาคำนวณความถี่ของคะแนนดิบทางทฤษฎี  $f_x(r)$  จะได้ตารางการแจกแจงความถี่ของคะแนนดิบทางทฤษฎีของกลุ่มผู้สอบ ฉบับ  $X$
3. จากกลุ่มผู้สอบแบบสอบฉบับ  $Y$  นำข้อมูลมาวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ แล้วนำค่า  $\theta_y$  ของผู้สอบแต่ละคนมาคำนวณความถี่สัมพัทธ์อย่างมีเงื่อนไขทางทฤษฎีของการแจกแจงคะแนนดิบ  $f_y(r|\theta_y)$
4. นำค่า  $\theta_y$  ของกลุ่มผู้สอบมาคำนวณความถี่ของคะแนนดิบทางทฤษฎี  $f_y(r)$  จะได้ตารางการแจกแจงความถี่ของคะแนนดิบทางทฤษฎีของกลุ่มผู้สอบ ฉบับ  $Y$
5. เทียบคะแนนระหว่างคะแนนดิบของแบบสอบฉบับ  $X$  และ  $Y$  (จากข้อ 2 และ 4) โดยใช้วิธีอีควิเปอร์เซนไทล์

รูปแบบวิธีการเทียบคะแนนระหว่างแบบทดสอบให้สามารถนำมาเทียบกันได้ อย่างมีความหมาย การเทียบคะแนนสามารถจำแนกได้ตามความคิดพื้นฐานของการเทียบคะแนน ได้เป็น 2 รูปแบบ ดังต่อไปนี้

1. รูปแบบการเทียบคะแนนโดยใช้ทฤษฎีดั้งเดิม รูปแบบนี้มีวิธีที่นิยมใช้ คือวิธีเทียบคะแนนแบบอีควิเปอร์เซนไทล์ วิธีการเทียบคะแนนเชิงเส้น และวิธีการเทียบคะแนนโดยใช้สมการถดถอย

2. รูปแบบการเทียบคะแนนโดยใช้ทฤษฎีตอบสนองข้อสอบรูปแบบนี้จำแนกเป็นกลุ่มวิธีการ 2 กลุ่มใหญ่ๆคือ การเทียบคะแนนจากคะแนนจริง และการเทียบคะแนนจากคะแนนที่สังเกต

## ตอนที่ 2 การหาคุณภาพของการปรับเทียบคะแนน

ในการศึกษาคุณภาพของการปรับเทียบคะแนน นิยมใช้การประเมินคุณภาพโดยใช้ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน (Standard Error of Equating:SEE) และการตรวจสอบความเพียงพอของการปรับเทียบคะแนนซึ่งมีผู้เสนอแนวคิด และวิธีปฏิบัติ ดังนี้

### 2.1 ความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน

การประเมินความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนที่สมบูรณ์ คือการวิเคราะห์แหล่งความลำเอียง (Bias) และความแปรผัน (Variability) แล้วนำเสนอให้เห็นเป็นภาพเดียวกัน เพื่อช่วยในการตัดสินใจว่าการเทียบมาตรฐานด้วยรูปแบบนั้นๆ ไปใช้มีความสมเหมาะสมหรือไม่อย่างไร

#### 2.1.1 ความลำเอียงของการเทียบคะแนน (Bias Error)

ความลำเอียงเป็นองค์ประกอบหนึ่งของความคลาดเคลื่อนมาตรฐาน ซึ่งจะต้องพิจารณากันอย่างรอบคอบ เมื่อต้องการนำผลของการเทียบคะแนนไปใช้ บรูโนและฮอลแลนด์ (พรพิมล นาคเวช, 2537:25 อ้างอิงมาจาก Braun and Hooland. 1982) ได้เสนอบทการวิเคราะห์ไว้ว่าความลำเอียงทางสถิติ มีความหมาย 2 ประการคือ

ก. ความแตกต่างของค่าเฉลี่ยของค่าตัววัดประมาณตลอดจนการทำการสุ่มซ้ำๆจากประชากรเดียวกันกับค่าของประชากรที่ถูกประมาณ

ข. ความแตกต่างของฟังก์ชันของการเทียบคะแนนโดยประมาณ (Estimated Equating Function) และค่าที่แท้จริงของฟังก์ชันการเทียบคะแนน (Equating Function)

แหล่งของความลำเอียงของการเทียบคะแนนที่สำคัญมี 2 แหล่ง คือ

1. ความผันแปรของประชากร (Population Variability) เนื่องจากการเทียบคะแนนไม่ได้กระทำกับประชากรกลุ่มเดียวโดยตลอด ฟังก์ชันของการเทียบคะแนนที่สร้างขึ้นเพื่อใช้กับประชากรหนึ่งอาจนำไปใช้กับประชากรหนึ่งไม่ได้เช่น ถ้า  $X_p*(Y)$  เทียบคะแนนจาก Y ไปยังคะแนน X ในประชากร P และ  $Y_Q*(Z)$  เทียบคะแนนจาก Z ไปยังคะแนน Y ในประชากร Q

แล้วจะสรุปได้ว่าฟังก์ชันเชิงซ้อน  $x_p * (y_Q(Z))$  เทียบคะแนนจาก  $Z$  ไปยัง  $X$  ในประชากร  $P$  หรือ  $Q$  ย่อมไม่ได้

ในแบบแผนการเทียบคะแนนที่ใช้แบบทดสอบร่วมได้แก้ไขปัญหานี้ โดยการสร้างประชากรสังเคราะห์ (Synthesis Population) ซึ่งเป็นการรวมของประชากร  $P$  และ  $Q$  ตามสัดส่วนแต่ยังไม่รับประกันผลว่าจะไม่มีความลำเอียงในทุกๆ ประชากรที่แปลเปลี่ยนไป

2. ความคลาดเคลื่อนของรูปแบบ (Model Error) ความคลาดเคลื่อนนี้เกิดขึ้นจากการจัดกระทำ ข้อสมมติฐานที่ว่าด้วยรูปแบบการแจกแจงให้ง่ายขึ้นอย่างผิดพลาด ได้แบ่งออกเป็น 2 พวกคือ

1) ความคลาดเคลื่อนรูปแบบที่ทดสอบได้ (Testable Model Error) เช่น ความเป็นเส้นตรงของฟังก์ชันการถดถอย สามารถทดสอบได้จากข้อมูลจำนวนมาก ซึ่งจะเห็นผลได้ว่ามีความเพียงพอหรือไม่

2) ความคลาดเคลื่อนรูปแบบที่ทดสอบไม่ได้ (Non Testable Model Error) เป็นกรณีที่ไม่สามารถหาข้อมูลมาทดสอบเพื่อพิสูจน์ได้ว่า สมมติฐานที่ว่าไว้เพียงพอหรือไม่

2.1.2 ความแปรผันเชิงสุ่มและความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

วิธีการเทียบคะแนนทุกวิธีไม่ว่าจะอยู่ในรูปแบบการเทียบคะแนนใด เมื่อกลุ่มตัวอย่างผู้สอบเป็นกลุ่มที่สุ่มมาจากประชากรเดียวหรือหลายประชากรย่อมมีความผันแปรเชิงสุ่มเกิดขึ้นจึงนิยมใช้เทคนิคการประมาณค่าความคลาดเคลื่อนเชิงสุ่ม (Estimate Sampling Error) กับการเทียบคะแนนต่างๆ ความคลาดเคลื่อนเชิงสุ่มมีสมมติฐานเบื้องต้นว่าการสุ่มตัวอย่างมาจากการสุ่มและใช้ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน (Standard Error of Equating: SEE) เป็นการวัดความแปรผันประเภทนี้

ความคลาดเคลื่อนมาตรฐานโดยอิงรูปแบบทฤษฎีการตอบสนองข้อสอบ ค่าประมาณคะแนนจะนับข้อที่ตอบถูก (Number – right Score)  $\xi$  ของผู้สอบที่ทำแบบทดสอบ  $X$  และ  $\eta$  ของผู้สอบที่ทำแบบทดสอบ  $Y$  มีค่าเท่ากับฟังก์ชันคุณลักษณะที่ประเมินที่ระดับความสามารถเดียวกัน ให้ความหมายของการเท่ากันในเชิงคะแนนสมมูล ในทางปฏิบัติจึงนำความสัมพันธ์เชิงฟังก์ชัน  $\xi$  และ  $\eta$  มาใช้ในการเทียบคะแนน  $X$  และ  $Y$  จากคะแนนที่สอบได้จริงของคนทั้งสองกลุ่ม การประมาณเพื่อให้เทียบ  $\xi$  และ  $\eta$  ตามฟังก์ชัน จะต้องใช้ค่าประมาณของประชากรข้อสอบค่าประมาณเหล่านี้ได้ คือที่มาของความคลาดเคลื่อนเชิงสุ่ม (Sampling Error) ในการเทียบคะแนน

การประเมินความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนที่สมบูรณ์ คือ การวิเคราะห์แหล่งความลำเอียงของการเทียบคะแนน ความแปรผันเชิงสุ่มและความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน

## 2.2 การประเมินความเพียงพอของการปรับเทียบคะแนน

การเทียบคะแนนรูปแบบใดก็ตามจะมีคุณภาพดีที่สุดเมื่อคะแนนที่ได้จากแบบทดสอบที่ต้องการเป็นเงื่อนไขต่างๆ ที่กำหนดไว้ของแต่ละรูปแบบ แต่ในสถานการณ์จริงมีข้อจำกัดทำให้ไม่สามารถได้ข้อมูลตามเงื่อนไข ดังนั้นจึงจำเป็นต้องตรวจสอบความเพียงพอของการเทียบคะแนนหรือประเมินคุณภาพของรูปแบบที่แตกต่างกันซึ่งวิธีการประเมินความเพียงพอ มีผู้เสนอแนวความคิดและวิธีปฏิบัติดังนี้

### การประเมินก่อนดำเนินการเทียบคะแนน

การประเมินก่อนดำเนินการเทียบคะแนน เจเกอร์ (Jaeger, 1981 : 26) ได้เสนอดัชนี 5 ตัว เพื่อตรวจสอบดัชนีความเพียงพอของการใช้วิธีเชิงเส้นตรงว่าเทคนิค วิธีที่ได้นำมาใช้ในการเทียบเพียงพอกับการปรับความแตกต่างระหว่างค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน การทดสอบความเหมือนของการแจกแจงของคะแนนจากแบบทดสอบหรือไม่ หรือจำเป็นต้องมีวิธีอื่นที่มีความสลับซับซ้อนมากขึ้น ดัชนี 5 ตัวมีดังนี้

ตัวที่ 1 ดัชนีความคล้ายคลึงของการแจกแจงคะแนนสะสมของแบบทดสอบชุดเก่า และชุดใหม่โดยการปรับความแตกต่างระหว่างค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน การทดสอบความเหมือนของการแจกแจง

ตัวที่ 2 รูปแบบของการแจกแจงคะแนนดิบกับคะแนนแปลง (Shape of The Raw Score Transformation) เหตุของการใช้ดัชนีตัวนี้ คือ ถ้าการเทียบคะแนนด้วยวิธีเชิงเส้นตรงสามารถอธิบายความแตกต่างในการแจกแจงของคะแนนดิบทั้งสองชุดอย่างพอเพียง ก็เป็นเหตุผลอย่างพอเพียงเช่นกันที่จะยอมรับว่าการแปลงคะแนนดิบจากแบบทดสอบชุดใหม่ไปยังแบบทดสอบชุดเก่าเป็นเส้นตรงอย่างแน่นอน

ตัวที่ 3 ความคงเส้นคงวา ของผลลัพธ์ของการเทียบคะแนนตามวิธีเส้นตรงกับการเทียบที่ตำแหน่งเปอร์เซ็นต์ไทล์ (Consistency of Linear and Equipercentile Equating Results) การวิเคราะห์นี้อาศัยข้อตกลงที่เป็นสมมติฐานเบื้องต้นที่ว่าถ้าวิธีเชิงเส้นตรงมีความเพียงพอแล้ว ฟังก์ชันของวิธีเชิงเส้นตรงที่สมนัยกัน

ตัวที่ 4 ความคล้ายคลึงของการแจกแจงความยาวของข้อสอบ (Similarity of Item Difficulty Distributions) โดยอาศัยหลักการที่ว่าการใช้วิธีการเชิงเส้นตรงมีความพอเพียงอย่าง

แท้จริงกับแบบทดสอบคู่ขนานกัน ถ้ามีการเบี่ยงเบนจากความเป็นคู่ขนานมากเท่าไร แสดงว่าต้องการวิธีการเทียบมาตรฐานที่ซับซ้อนขึ้นเพราะการแจกแจงของแบบทดสอบที่ไม่ใช่คู่ขนานจะมีความแตกต่างเกิดขึ้นในรูปโมเมนต์ที่สูงขึ้น

ตัวที่ 5 ความคล้ายคลึงของค่าอำนาจจำแนกของข้อสอบ (Similarity of Item Discrimination Distributions) เหตุผลทำนองเดียวกับดัชนีตัวที่ 4

#### การประเมินผลการเทียบคะแนน

การประเมินผลการเทียบประเมินคะแนนเป็นการประเมินซึ่งอาศัยคะแนนเกณฑ์ที่ได้เลือกสรรไว้แล้วเป็นหลักในการเทียบหาความแตกต่างมีรายละเอียดดังนี้

##### 1. ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ (Percentile Comparison Index)

โคเลนและวิทนี (Kolen and Whitney. 1982 : 284) ได้แนะนำให้ใช้ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ เพื่อประเมินความเพียงพอของการเทียบคะแนน วิธีการหาค่าดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์นี้ คือ การวัดความไม่สอดคล้อง (Dissimilarity) ระหว่างการแจกแจงคะแนนที่ได้จากกลุ่มตรวจสอบผล คือ คะแนนจากแบบทดสอบชุด X กับคะแนนซึ่งได้จากการแปลงคะแนนจากแบบทดสอบชุด Y ไปยังสเกลของแบบทดสอบชุด X ตามตารางการเทียบคะแนนที่ได้จากวิธีการใช้เทียบโดยหาค่าสมมูล ณ ที่ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกันดัชนีดังกล่าวคำนวณได้จาก

$$C = \frac{\sum (X_i - X_i^*)^2}{nk}$$

|       |         |     |   |
|-------|---------|-----|---|
| เมื่อ | C       | คือ | ดัชนีความเพียงพอ                                    |
|       | $X_i$   | คือ | คะแนนที่ได้จากการสอบชุด X ของคน i                   |
|       | $X_i^*$ | คือ | คะแนนที่ได้เทียบด้วยตารางคะแนนที่สมนัยกันของคนที่ i |
|       | n       | คือ | จำนวนคะแนนดิบในกลุ่มตรวจสอบผล                       |
|       | k       | คือ | จำนวนข้อสอบในแบบทดสอบร่วมที่ใช้                     |

ถ้าค่า C น้อยกว่า แสดงว่าวิธีการเทียบคะแนนที่นำมาใช้สร้างตารางการแปลงคะแนนนั้นมีความเพียงพอที่จะนำมาใช้ให้ได้ผลการเทียบที่คงที่มากกว่า

## 2. ดัชนีความเพียงพอหรือดัชนีความแตกต่าง (Discrepancy Indices)

ปีเตอร์เซน มาร์โค และสตีเวอรัท (Petersen, Maco and Stewart. 1982 : 91) ได้เสนอวิธีการประเมินความเพียงพอของการเทียบมาตรฐานด้วยการใช้ดัชนีความแตกต่างซึ่งมีแนวคิดว่าจะแนบแต่ละตัวจากแบบทดสอบชุด X จะมีคะแนนเกณฑ์ t และคะแนน t' ที่ได้จากการประมาณค่าโดยใช้วิธีการเทียบคะแนนเฉพาะอันหนึ่งค่าความแตกต่าง d ระหว่าง t และ t' เป็นความคลาดเคลื่อนในการเทียบคะแนน ซึ่งค่าที่น้อยกว่าแสดงว่า วิธีการมาตรฐานนั้นมีความเหมาะสมมากกว่า

ค่าดัชนีความเพียงพอ คือ ค่าความคลาดเคลื่อนรวมในการเทียบคะแนนของวิธีการเทียบคะแนนที่ใช้นั้นซึ่งคำนวณจากค่าเฉลี่ยยกกำลังสองของความแตกต่างระหว่างคะแนนแปลงกับคะแนนเกณฑ์และถ่วงน้ำหนักด้วยความแปรปรวนของเกณฑ์ (Standardized Weighted Mean-square Difference) ซึ่งเท่ากับค่าความแปรปรวนของความแตกต่างรวมกับกำลังสองของความลำเอียง

$$\begin{aligned} \text{Total error} &= j \sum f_j d_j / ns_i^2 \\ \text{เมื่อ } d_j &= (t - t') \\ N &= \text{จำนวนคะแนนที่ใช้} \\ S_t^2 &= \text{ความแปรปรวนของคะแนน } t \end{aligned}$$

ดัชนีเปรียบเทียบความแตกต่างนี้เป็นค่ามาตรฐานที่แสดงในรูปสัดส่วนของความเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ ค่าความคลาดเคลื่อนรวมที่ได้นี้จึงสามารถเปรียบเทียบกันได้โดยตรงระหว่างค่าที่ได้จากการใช้วิธีและสถานการณ์การเทียบคะแนนต่างกัน

วิธีการประเมินความเพียงพอที่ได้กล่าวมาแล้วนี้ อาจจำแนกเป็นสองพวก พวกแรกเป็นการประเมินก่อนการดำเนินการ เช่น ดัชนีความคล้อยคลึงของการแจกแจงคะแนนสะสมของแบบทดสอบสองชุด ดัชนีความคล้อยคลึงของการแจกแจงของค่าความยากของข้อสอบ เป็นต้น พวกหลังเป็นการประเมินผลของการเทียบคะแนน ซึ่งอาศัยคะแนนเกณฑ์ที่เลือกสรรแล้วเป็นหลักในการเทียบหาความแตกต่างสำหรับดัชนีที่เสนอโดย ปีเตอร์เซนและคณะอื่น (Petersen and other. 1982 : 91) คะแนนเกณฑ์ที่ใช้ คือ ผลการแปลงคะแนนด้วยรูปแบบอิงทฤษฎีการตอบสนองข้อสอบ ค่าดัชนีที่คำนวณมีลักษณะเป็นหน่วยมาตรฐานแล้ว สามารถนำมาหาค่าเหล่านี้ที่ได้จากการใช้รูปแบบที่ต่างกันตลอดจนสถานการณ์ที่ได้ข้อมูลที่ต่างกันมาเปรียบเทียบกันโดยตรง ได้ส่วนดัชนีที่เสนอโดยโคเลนและวิทนี (Kolen and Whitney. 1982 : 284) ได้ใช้ชุดข้อมูลจากผู้สอบเองเป็นเกณฑ์ในการหาความแตกต่าง ข้อมูลเหล่านี้ได้จากการออกแบบด้วยการใช้กลุ่มตัวอย่างทานผล ซึ่งผู้สอบในกลุ่มตัวอย่างนี้ได้รับการทดสอบด้วยแบบทดสอบทั้งสองชุด ดังนั้น การใช้คะแนนของตนเองเป็นเกณฑ์จึงมีความเป็นอิสระไม่ขึ้นอยู่กับการแปลงคะแนนอื่นๆ

เช่น วิธีเสนอโดยปีเตอร์เซนและคณะ ด้วยเหตุผลดังกล่าว ผู้ศึกษาจึงได้เลือกวิธีการประเมินความเพียงพอจากการวิเคราะห์ผลกลุ่มตัวอย่างสอบทานผล แต่การหาค่าดัชนี ได้ดัดแปลงมาจากสูตรของโคเลนไปใช้ตามแนวความคิดของปีเตอร์เซนและคณะ คือ ใช้ค่าความแปรปรวนเป็นตัวถ่วงน้ำหนักเพื่อให้ค่าที่ได้มีหน่วยเป็นมาตรฐาน

สูตรที่ดัดแปลงและนำมาใช้ในการศึกษาครั้งนี้ คือ

$$C = \frac{\sum_{j=1}^n (X_{ij} - X_{kij}^*)^2}{n S_{ii}^2}$$

|       |             |  |
|-------|-------------|--|
| เมื่อ | C           | แทน ค่าความเพียงพอ   |
|       | $X_{ij}$    | แทน คะแนนจากแบบทดสอบชุดที่ i ของนักเรียนคนที่ j  |
|       | $X_{kij}^*$ | แทน คะแนนจากแบบทดสอบชุดที่ k เทียบไปเป็นคะแนนจากแบบทดสอบชุดที่ i โดยใช้ตารางเทียบคะแนนของนักเรียนคนที่ j |
|       | $S_{ii}^2$  | แทน ความแปรปรวนของการแจกแจงคะแนนจากแบบทดสอบชุดที่ i  |
|       | n           | แทน จำนวนคนในกลุ่มตัวอย่างสอบทานผล   |

อาจสรุปได้ว่าการประเมินความพอเพียงของการเทียบคะแนนใช้เกณฑ์ในการประเมิน 2 เกณฑ์ คือ ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ ซึ่งเป็นการวัดความไม่สอดคล้อง ระหว่างการแจกแจงคะแนนที่ได้จากกลุ่มตรวจสอบผล และดัชนีความเพียงพอหรือดัชนีความแตกต่าง

### ตอนที่ 3 งานวิจัยที่เกี่ยวข้องกับการปรับเทียบคะแนน

#### 3.1 งานวิจัยต่างประเทศ

การศึกษางานวิจัยของต่างประเทศที่น่าเสนอมีดังนี้

ฮิวลิน ลิสส์ค และ ดราสโกว์ (Hulin, lissak, and Drasgow. 1982 :153-158) ได้ใช้การจำลองข้อมูลเพื่อศึกษา ขนาดของกลุ่มตัวอย่าง ความยาวของแบบสอบ และการประมาณค่าพารามิเตอร์ด้วยโปรแกรม LOGIST โดยมีการเปลี่ยนความยาวแบบสอบและเปลี่ยนขนาดของกลุ่มตัวอย่าง กับ IRT โมเดล 3 พารามิเตอร์ และพารามิเตอร์ความสามารถมีการแจกแจงแบบโค้งปกติ ผลการวิจัยพบว่าความยาวของแบบสอบและขนาดของกลุ่มตัวอย่างมีผลต่อการประมาณค่าเมื่อประมาณค่าพารามิเตอร์ 3 พารามิเตอร์โดยใช้ข้อสอบ 60 ข้อ กับกลุ่มผู้สอบจำนวน 1,000 คน

พบว่าให้ผลออกมาในลักษณะคงที่ ถ้าลดข้อสอบลงครึ่งหนึ่งคือเหลือ 30 ข้อ และเพิ่มขนาดของกลุ่มตัวอย่างเป็นสองเท่าจะให้ผลใกล้เคียงกัน และนอกจากนี้ยังพบว่าโปรแกรม LOGIST ใช้กับกลุ่มตัวอย่างขนาด 15 คนไม่ได้

คาลด์เวลล์ (Caldwell. 1984 : 2847) ได้ทำการศึกษาเพื่อวัดประสิทธิผลสัมพัทธ์ในโมเดลการปรับเทียบความยากของแบบทดสอบ รูปแบบที่ใช้คือ รูปแบบเชิงเส้นตรง แบบแผนที่ 4 และราสซิมิเดล ในการปรับเทียบได้พิจารณาถึง การเพิ่มค่าเฉลี่ยและการลดส่วนเบี่ยงเบนมาตรฐานสำหรับการสอบครั้งที่ 2 ซึ่งคาดว่าโมเดลเชิงเส้นตรงจะมีประสิทธิผลน้อยกว่าราสซิมิเดล เนื่องจากโมเดลเชิงเส้นตรงมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานบรรจุอยู่ในสูตร ส่วนราสซิมิเดลไม่มี และคาดว่าประสิทธิผลสัมพัทธ์ของราสซิมิเดลจะเพิ่มขึ้นขณะที่ความแตกต่างของค่าเฉลี่ยเพิ่มขึ้น ได้มีการประเมินแบบทดสอบรวมทั้งสองแบบ แบบทดสอบร่วมชุดหนึ่งประกอบด้วยข้อสอบที่มีความยากในระดับปานกลางส่วนอีกชุดหนึ่งมีความยากสูงสุด แบบทดสอบที่มีค่าความยากปานกลางได้รับการพิจารณาว่าให้ประสิทธิผลต่ำสุด และแบบทดสอบร่วมที่มีความยากปานกลางได้รับการพิจารณาว่าให้ประสิทธิผลที่ดีกว่า ผลการวิจัยพบว่าเมื่อพิจารณาคะแนนทั้งหมด ราสซิมิเดลเหนือกว่ารูปแบบเชิงเส้นตรงทั้งในแบบทดสอบร่วมที่มีความยากระดับปานกลางและระดับสูง แต่เมื่อพิจารณาคะแนนจุดตัดที่แสดงถึงความสามารถต่ำสุด ราสซิมิเดลเหนือกว่าในลักษณะที่ให้ความลำเอียงเพียงเล็กน้อยแต่เชิงเส้นตรง ให้ค่าความคลาดเคลื่อนในระดับต่ำดัชนีความแตกต่างของราสซิมิเดลมีค่าเฉลี่ยเกือบเป็นศูนย์และมีการแปรเปลี่ยนในช่วงกว้าง รูปแบบเชิงเส้นตรงให้ค่าความลำเอียงด้านลบ แต่มีการแปรเปลี่ยนที่น้อยกว่า

สแก็กส์ (Skaggs. 1984 : 1259) ได้ทำการตรวจสอบวิธีการปรับเทียบคะแนน 4 วิธีภายใต้เงื่อนไขที่พารามิเตอร์ข้อสอบและพารามิเตอร์ความสามารถแปรเปลี่ยนอย่างเป็นระบบ วิธีการปรับเทียบที่ใช้ได้แก่ IRT 2 แบบ คือ ราสซิมิเดล กับ โมเดล 3-พารามิเตอร์ และแบบมาตรฐานเดิมอีก 2 แบบ คือการปรับเทียบแบบเส้นตรงและแบบอิกวิเปอร์เซ็นไทล์ ข้อมูลที่ใช้ในการศึกษาใช้การจำลองขึ้นมา ตัวแปรที่ใช้ในการศึกษาได้แก่ ค่าความยาก อำนาจจำแนก และระดับโอกาสการเดา และค่าเฉลี่ยของความสามารถ เกณฑ์ที่ใช้ในการปรับเทียบคือคะแนนจริงจากโมเดล 3-พารามิเตอร์ ข้อแตกต่างระหว่างวิธีการปรับเทียบทั้ง 4 แบบกับเกณฑ์นำมาวิเคราะห์หาค่าสถิติเชิงบรรยายและแสดงด้วยกราฟและจุดประสงค์หลักของการศึกษาคั้งนี้เพื่อทดสอบความแกร่งของราสซิมิเดลเมื่อมีการละเมิดข้อตกลง ผลการวิจัยพบว่าราสซิมิเดลยังไม่มีความแกร่งเมื่อมีการละเมิดในข้อตกลงกรณีนี้เป็นจริงเมื่อค่าเฉลี่ยของอำนาจจำแนกมีค่าไม่เท่ากัน และโมเดล 3-พารามิเตอร์ ก็ประสบปัญหาเช่นกันเมื่ออำนาจจำแนกไม่เท่ากัน โดยภาพรวมการ

ปรับเทียบคะแนนแบบอิกวิเปอร์เซ็นไทล์ให้ผลดีเกือบทั้งหมด การวิจัยครั้งนี้ได้ให้ข้อเสนอแนะว่า การใช้รหัสโมเดลและ โมเดล 3-พารามิเตอร์ ต้องใช้ด้วยความระมัดระวัง

วูดดรัฟ (Woodruff. 1986 : 245-257) ได้ศึกษาวิธีปรับเทียบคะแนนแบบเส้นตรงโดยใช้แบบแผนข้อสอบร่วมที่มีประชากรแตกต่างกัน (CINEPย่อมาจาก Common Item Non-Equivalent Population ) และแบบจำลองคะแนนจริงสัมพันธ์ โดยใช้กลุ่มตัวอย่างจำนวน 4,000 คน ทดสอบกับแบบทดสอบ 5 ฉบับ ที่เป็นฟอร์มใหม่ มีความยาวที่แตกต่างกันคือตั้งแต่จำนวน 25 ถึง 115 ข้อ และใช้แบบทดสอบที่เป็นฟอร์มเก่าร่วมกับแบบทดสอบเชื่อมโยงมาทดสอบกับกลุ่มตัวอย่างจำนวน 4,000 คน นำคะแนนที่ได้มาศึกษาวิธีปรับเทียบคะแนน 2 วิธี ผลการศึกษาพบว่ามีความแตกต่างของกลุ่มผู้สอบเล็กน้อย โดยดูจากความเป็นเบี่ยงเบนมาตรฐานจากแบบทดสอบ 5 ฉบับ ที่ใช้แบบทดสอบหลักและใช้วิธีการปรับเทียบคะแนน 2 วิธี ให้ความเพียงพอในการปฏิบัติ แม้ว่าแบบทดสอบทั้ง 5 ฉบับ จะวัดในลักษณะที่ต่างกัน

โคป (Cope. 1987 : 143-149) ได้เปรียบเทียบการปรับเทียบคะแนนเชิงเส้นตรง 5 วิธีระหว่างวิธีของแองกอฟ(Angoff. 1971)ตามวิธีที่ 5 ซึ่งมี 3 วิธี กับวิธีของทักเกอร์ (Tucker) และเลวิน (Levine) ในกรณีที่แบบทดสอบมีความเชื่อมั่นเท่ากันโดยใช้แบบทดสอบรวมแบบทดสอบที่ใช้ปรับเทียบคะแนนเป็นแบบทดสอบเพื่อให้ประกาศนียบัตรในวิชาชีพ จำนวน 5 ชุด สอบกับกลุ่มประชากรที่ไม่เท่าเทียมกัน 5 กลุ่ม ๆ ละ 1 ชุด ปีละ 2 ครั้ง การเปรียบเทียบผลการเทียบระหว่างแบบทดสอบแต่ละคู่ (Single link) ภายใต้วิธีการเชิงเส้นตรงที่แตกต่างกัน 5 วิธี ทั้งนี้ไม่มีการเปรียบเทียบความถูกต้องของวิธีการต่าง ๆ เพราะว่าสเกลคะแนนดิบของแบบทดสอบสองชุดถูกคาดว่าจะต่างกัน อีกแบบหนึ่งคือ การเปรียบเทียบผลการปรับเทียบคะแนนแบบโซ่ ซึ่งเป็นการเปรียบเทียบคะแนนแบบทดสอบแต่ละชุดกับตัวเอง โดยผ่านแบบทดสอบชุดอื่นที่เลือก ทั้งนี้เพื่อทำการเปรียบเทียบความถูกต้องของวิธีการเหล่านั้นด้วย

การวัดความคลาดเคลื่อนของการปรับเทียบคะแนน (Equating Error) ใช้สถิติ 2 ตัว คือ ค่าประมาณความลำเอียง (Estimate Bias) และค่าประมาณที่ถ่วงน้ำหนักแล้วของรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนในการปรับเทียบคะแนน(Estimated Weighted Root Squared Error : RMSE) ซึ่งใช้เปรียบเทียบระหว่างวิธีที่แตกต่างกัน คำนวณจากสมการ

$$\text{BIAS} = \frac{\bar{x}'}{\bar{x}} - \bar{x}$$

เมื่อ  $\bar{x}$  คือ ค่าเฉลี่ยของคะแนนดิบของแบบทดสอบชุดหนึ่ง  
 $\bar{x}'$  คือ ค่าเฉลี่ยของคะแนนดิบที่เท่าเทียมกันโดยการเทียบแบบทดสอบชุดนั้นไปยังตัวเอง

$$\text{และ RMSE} = \left[ \frac{\sum N_i (x'_i - x_i)^2}{\sum N_i} \right]^{\frac{1}{2}}$$

|             |     |                               |
|-------------|-----|-------------------------------|
| เมื่อ $x_i$ | คือ | คะแนนดิบของคนที่ $i$          |
| $x'_i$      | คือ | คะแนนดิบที่เท่าเทียมกัน       |
| $N_i$       | คือ | จำนวนผู้สอบซึ่งได้คะแนน $X_i$ |

ในกรณีที่มีผู้สอบจำนวนมาก การคำนวณ RMSE จะง่ายขึ้นโดยใช้สมการ

$$\text{RMSE} = \left[ (m-1)^2 \text{Var}(x) + (\text{BIAS})^2 \right]^{\frac{1}{2}}$$

|                 |     |  |
|-----------------|-----|--|
| เมื่อ $m$       | คือ | ความชันของฟังก์ชันการเทียบ $x' = mx + b$ |
| $\text{Var}(x)$ | คือ | ความแปรปรวนของ $x$                       |

ผลการวิจัยสรุปว่า ในสถานการณ์ที่ศึกษาภายใต้รูปแบบที่ 5 ตามวิธีของแองกอฟ ให้คะแนนดิบ (Raw Score Equivalent) ใกล้เคียงกับวิธีการของทักเกอร์และเลวินโดยการปรับเทียบคะแนนแบบสายโซ่ทั้ง 5 วิธี ให้ค่าประมาณความลำเอียงและรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนที่ใกล้เคียงกัน ส่วนการปรับเทียบคะแนนจากวิธีการของทักเกอร์และเลวิน ผู้วิจัยกล่าวว่าผลการวิจัยนี้สนับสนุนวิธีการปรับเทียบคะแนนเชิงเส้นตรงตามรูปแบบที่ 5 ซึ่งมีข้อตกลงที่มีข้อจำกัดน้อยกว่าวิธีของทักเกอร์และเลวิน ดังนั้นในทางปฏิบัติควรใช้วิธีการตามรูปแบบที่ 5 วิธีการใดวิธีหนึ่งใน 3 วิธี

แมคแคน (MacCann. 1989 : 263-276) ได้เปรียบเทียบวิธีปรับเทียบคะแนนสองวิธีที่ใช้ข้อตกลงว่ามีความเชื่อมั่นเท่าเทียมกันและใช้แบบทดสอบที่เป็นคะแนนจริงสัมพัทธ์ ทำการทดสอบกับกลุ่มตัวอย่างทั้งที่เป็นกลุ่มคล้ายคลึงกันและกลุ่มที่มีความแตกต่างกัน โดยมีขนาดกลุ่มตัวอย่างที่แตกต่างกันในแต่ละวิชาที่สอบ และแต่ละกลุ่มที่ใช้เป็นกลุ่มปรับเทียบคะแนน ตั้งแต่จำนวน 326 คน ถึง 5,610 คน ใช้แบบทดสอบ 5 วิชา แบ่งเป็นส่วนย่อยที่แตกต่างกันในแต่ละวิชา และใช้แบบทดสอบเชื่อมโยงที่มีความยาวคิดเป็นสัดส่วนแตกต่างกัน ทั้งที่มากกว่าและน้อยกว่าแบบทดสอบที่นำมาใช้ปรับเทียบคะแนน แล้วจึงนำมาเปรียบเทียบความคลาดเคลื่อนมาตรฐานของแต่ละวิธี (สี่วิธี) คือ วิธีของแมคแคน สองวิธี และวิธีของทักเกอร์กับวิธีของเลวิน พบว่า วิธีที่ 1 ให้ผลใกล้เคียงกับวิธีของทักเกอร์ ส่วนวิธีที่ 2 ให้ผลใกล้เคียงกับวิธีของเลวิน และทั้งสี่วิธี ให้ผลความคลาดเคลื่อนมาตรฐานในระดับที่น่าพอใจเป็นส่วนมาก ผู้วิจัยสรุปว่า วิธีที่ 1 และวิธีที่ 2 มี

ข้อดีคือไม่จำเป็นต้องใช้ข้อตกลงที่ว่าการกระจายของคะแนนจากแบบทดสอบสองฉบับ ที่นำมาเทียบคะแนนกัน จะต้องมีการแจกแจงเป็นโค้งปกติ

เพเทียน (Patience. 1990 : 486) ได้ศึกษาเปรียบเทียบผลของการปรับเทียบคะแนนระหว่างระดับชั้นโดยใช้วิธีปรับเทียบคะแนน 5 วิธี คือ วิธีการตามทฤษฎีการสอบแบบดั้งเดิม 2 วิธี ได้แก่ อิกวิเปอร์เซ็นไทล์ และวิธีการของเธอร์สโตน และวิธีการตามทฤษฎีการตอบสนองข้อสอบ 3 วิธี ได้แก่ วิธีใช้แบบจำลองโลจิสติก แบบหนึ่ง สอง และสามพารามิเตอร์ โดยศึกษาว่า การใช้วิธีใด ให้ผลการปรับเทียบคะแนนที่เหมาะสมที่สุดแบบทดสอบที่ใช้แบ่งเป็น 3 ระดับ คือ ง่าย ยากง่ายปานกลาง และยาก การปรับเทียบคะแนนแบบทดสอบแต่ละระดับ ใช้วิธีการปรับเทียบคะแนนทั้ง 5 วิธี แบบทดสอบฉบับง่าย ใช้กับนักเรียนระดับ 9 แบบทดสอบฉบับยากง่ายปานกลางใช้กับนักเรียนระดับ 10 และแบบทดสอบฉบับยากใช้กับนักเรียนระดับ 11 คะแนนที่ได้จากแบบทดสอบแต่ละระดับ จะนำไปเปรียบเทียบกับคะแนนโดยรวมของนักเรียน วิธีการปรับเทียบคะแนนที่เหมาะสมที่สุดกำหนดว่า เป็นวิธีการที่ให้ผลการปรับเทียบคะแนนใกล้เคียงกับผลที่ได้จากคะแนนของนักเรียนจากแบบทดสอบรวมทุกระดับ

ผลการวิจัยพบว่าวิธีการปรับเทียบคะแนนทั้ง 5 วิธี ให้ผลการเทียบคะแนนที่คล้ายคลึงกัน เมื่อใช้แบบทดสอบฉบับง่ายและฉบับยากง่ายปานกลาง เมื่อใช้แบบทดสอบฉบับยาก พบว่า ผลการปรับเทียบคะแนนโดยวิธีอิกวิเปอร์เซ็นไทล์ วิธีการของเธอร์สโตนและวิธีสามพารามิเตอร์ให้ผลคล้ายคลึงกับคะแนนที่ได้จากแบบทดสอบรวมมากกว่าวิธีหนึ่งพารามิเตอร์และวิธีสองพารามิเตอร์

ทริสคาริ (Triscari. 1990 : 1996) ได้ศึกษาเปรียบเทียบวิธีการปรับเทียบคะแนน 5 วิธีโดยใช้แบบแผนการเก็บรวบรวมข้อมูลแบบความไม่เท่าเทียมกันของกลุ่มประชากร (Nonequivalent Population) และใช้ข้อสอบรวมแบบภายใน (Internal Anchor Test) โดยมีการเชื่อมโยงข้อสอบแบบเลือกตอบและแบบเรียงความ วิธีการปรับเทียบที่ใช้คือตามแนวทฤษฎีการวัดแบบมาตรฐานเดิม แบบเชิงเส้นตรง แบบอิกวิเปอร์เซ็นไทล์ และแบบทฤษฎีตอบสนองข้อสอบชนิด 1 , 2 และ 3 พารามิเตอร์ ซึ่งมีวิธีการให้คะแนนตามแนวพาเชียลเครดิต โมเดล (Partial Credit Model) การเปรียบเทียบใช้วิธีการจำลองสถานการณ์เมื่อทราบค่าคะแนนจริงที่เท่าเทียมกันของแบบทดสอบที่เป็นฐาน การประเมินประสิทธิภาพใช้ค่าความแตกต่างระหว่างคะแนนจริงของแบบสอบถาม นอกจากนี้ยังได้เก็บข้อมูลจริงเพื่อหาความสอดคล้องกับข้อมูลที่จำลองขึ้นมาด้วย ผลการวิจัยจากการจำลองข้อมูลพบว่า การปรับเทียบตามแนวทฤษฎีการตอบสนองข้อสอบที่ใช้พาราเชียลเครดิตโมเดล (Partial Credit Model) และมีการวิเคราะห์รวมให้ความคลาดเคลื่อนน้อยที่สุด และการปรับเทียบแบบเส้นตรง ให้ความคลาดเคลื่อนมากที่สุดจากการเก็บข้อมูลจริงพบว่า มีความสอดคล้องกับการจำลองข้อมูลร้อยละ 70

กลอวากิ (Glowacki. 1991 : 1722) ได้ตรวจสอบโมเดลของการเปรียบเทียบคะแนนที่มีความเหมาะสมกับการสอบของบัณฑิตวิทยาลัย แห่งมหาวิทยาลัยอลาบามา ปัญหาในการวิจัย คือ โมเดลของการเปรียบเทียบที่ตรวจสอบมีการแจกแจงของคะแนนดิบหรือคะแนนที่ผ่านจากการสอบแบบทดสอบการอ่าน และคณิตศาสตร์ แตกต่างกันหรือไม่ โมเดลที่ใช้ตรวจสอบคือ โมเดลเชิงเส้นตรง อิกวิเปอร์เซ็นไทล์และทฤษฎีการตอบสนองข้อสอบ ชนิด 1 2 และ 3 พารามิเตอร์ ผลการวิจัยพบว่า วิธีการเปรียบเทียบคะแนนทั้ง 5 โมเดล ในการสอบการอ่านและคณิตศาสตร์ให้ผลที่คล้ายคลึงกัน แสดงว่าโมเดลทั้งหมดสามารถนำมาใช้กับการเปรียบเทียบคะแนนได้ โดยไม่มีโมเดลใดที่ดีที่สุด

แฮริส (Harris. 1991 : 221-235) ได้ศึกษาผลของแบบแผนการรวบรวมข้อมูล 2 วิธี คือ แบบแผนที่ 1 ของแองกอฟ(ผู้สอบกลุ่มเดียวจัดให้สมดุล) และแบบแผนที่ 2 ของแองกอฟ (ผู้สอบกลุ่มเท่าเทียมกัน) ที่มีต่อการเทียบคะแนนตามแนวตั้ง โดยใช้วิธีเปรียบเทียบคะแนน 2 วิธี คือ วิธีอิกวิเปอร์เซ็นไทล์ และวิธีตามทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ แบบทดสอบที่ใช้คือ แบบทดสอบวิชาคณิตศาสตร์ 2 ฉบับ คือ EACT มีจำนวน 60 ข้อ และฉบับ PACT มีจำนวน 40 ข้อ แบบทดสอบทั้ง 2 ฉบับ มีเนื้อหาคาบเกี่ยวกัน แต่มีเนื้อหาเฉพาะและระดับความยากต่างกัน กลุ่มตัวอย่างเป็นนักเรียนระดับ 11 และ 12 ซึ่งแบ่งเป็นกลุ่มย่อยหลายกลุ่ม การวิเคราะห์ข้อมูลใช้ ค่าเฉลี่ยกำลังสองของความแตกต่าง (RMSD)

ผลการวิจัยสรุปได้ว่า การเปรียบเทียบคะแนนตามแนวตั้งโดยใช้แบบแผนการเก็บข้อมูลทั้ง 2 แบบ มีความเพียงพอ และให้ผลคล้ายคลึงกัน โดยรูปแบบที่ใช้ผู้สอบกลุ่มเดียว ให้ผลคงที่กว่ารูปแบบที่ใช้ผู้สอบกลุ่มเท่าเทียมเล็กน้อย

อเยอร์เว (Ayerve. 1992:1841) ได้ศึกษาเปรียบเทียบประสิทธิภาพของการเปรียบเทียบคะแนนตามแนวตั้งด้วยวิธีอิกวิเปอร์เซ็นไทล์ และวิธี IRT โมเดล 3 พารามิเตอร์โดยใช้การจำลองข้อมูลภายใต้เงื่อนไขของขนาดกลุ่มตัวอย่าง ความยาวของแบบทดสอบและความยาวของแบบทดสอบรวม โดยเปรียบเทียบใน 3 กรณี คือ (1) เปรียบเทียบประสิทธิภาพของวิธีการเปรียบเทียบคะแนนทั้งสองภายใต้เงื่อนไขทุกเงื่อนไข (2) เปรียบเทียบประสิทธิภาพของวิธีการเปรียบเทียบคะแนนแต่ละวิธีภายใต้เงื่อนไขของตัวแปรอิสระแต่ละตัว (3) ตรวจสอบผลของตัวแปรอิสระแต่ละตัวในการเปรียบเทียบคะแนนแต่ละวิธี ขนาดกลุ่มตัวอย่างที่ใช้ มี 3 ขนาด คือ 200 500 และ 1,000 คน ความยาวของแบบทดสอบที่ใช้มี 2 ขนาด คือ 30 ข้อ และ 60 ข้อ ส่วนความยาวของแบบทดสอบรวมที่ใช้มี 2 ขนาด คือ 5 ข้อ และ 10 ข้อ การวิเคราะห์ข้อมูล ใช้ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนซึ่งถ่วงน้ำหนักแล้ว (Weighted Mean Square Error :WMSE) และค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนซึ่งไม่ถ่วงน้ำหนัก (Unweighted Mean Square Error :UMSE)

ผลการวิจัยพบว่า โดยส่วนรวมแล้ว วิธีการเปรียบเทียบแบบ วิธีอิกวิเปอร์เซ็นไทล์ และวิธี IRT ไม่แตกต่างกันอย่างมีนัยสำคัญ สำหรับการเปรียบเทียบคะแนนโดยวิธีอิกวิเปอร์เซ็นไทล์ พบว่าความยาวของแบบทดสอบกับความยาวของแบบทดสอบร่วมเป็นองค์ประกอบที่สำคัญ ขณะที่การเปรียบเทียบคะแนนตามวิธี IRT 3 พารามิเตอร์ ขนาดของกลุ่มตัวอย่างเป็นองค์ประกอบที่สำคัญ คือ กลุ่มตัวอย่าง 200 คน มีแนวโน้มที่จะให้ผลการเปรียบเทียบคะแนนที่ไม่ถูกต้อง ในขณะที่กลุ่มตัวอย่างขนาดใหญ่ คือ 500 และ 1,000 จะให้ผลการเปรียบเทียบคะแนนที่ถูกต้องมากกว่า

ยัง (Young, 1995 : 558-571) ได้ปรับคะแนนเฉลี่ยสะสมของนักเรียนศึกษาชั้นปีที่ 1 ของโรงเรียนกฎหมายด้วยวิธีปรับคะแนนเฉลี่ยสะสม 2 วิธี คือ วิธีปรับคะแนนเฉลี่ยสะสมโดยใช้ทฤษฎีการตอบสนองข้อสอบภายใต้ Graded Response Model กับวิธีปรับคะแนนเฉลี่ยสะสมด้วยวิธีเชิงเส้นตรง เพื่อเปรียบเทียบว่าคะแนนเฉลี่ยสะสมที่ปรับด้วยวิธีใด จะเป็นเกณฑ์ที่ดีสำหรับตรวจสอบความตรงเชิงทำนาย ของคะแนน LSAT และคะแนนเฉลี่ยสะสมระดับมหาวิทยาลัย (UGPA) ข้อมูลที่ใช้ได้จากโรงเรียนกฎหมาย 4 แห่งในสหรัฐอเมริกา แยกวิเคราะห์เป็นรายโรงเรียน ผลการวิเคราะห์พบว่า โดยทั่วไปเมื่อใช้คะแนนเฉลี่ยสะสมปี 1 ของโรงเรียนกฎหมายที่ปรับด้วยการวิเคราะห์แบบเชิงเส้นตรงเป็นเกณฑ์จะให้ความเที่ยงตรงเชิงทำนายดีกว่าวิธีปรับคะแนนเฉลี่ยสะสมโดยใช้ทฤษฎีการตอบสนองข้อสอบภายใต้ Graded Response Model

### 3.2 งานวิจัยในประเทศ

สำหรับงานวิจัยในประเทศไทยที่เกี่ยวกับการปรับเทียบคะแนน ผู้วิจัยขอเสนอเป็นกลุ่ม ๆ ตามลักษณะเรื่องที่ทำการศึกษา ดังนี้

#### 3.2.1 งานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบแบบแผนการรวบรวมข้อมูลมี 2 เรื่องคือ

ชูชีพ พงษ์สมบุญ (2528 :112 - 118) ได้ศึกษาเปรียบเทียบประสิทธิภาพและความคงที่ของการปรับเทียบคะแนนโดยใช้แบบแผนผู้สอบร่วม และแบบแผนข้อสอบร่วม โดยใช้วิธีการปรับเทียบ 3 วิธี คือ วิธีอิกวิเปอร์เซ็นไทล์ วิธีเชิงเส้นตรง และวิธี IRT แบบ สามพารามิเตอร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศ สุ่มมาจำนวน 3,721 คน ที่เข้าสอบโครงการตรวจสอบคุณภาพของกรมวิชาการ กระทรวงศึกษาธิการ ซึ่งแบบทดสอบในโครงการมีจำนวน 60 ข้อ นำมาแบ่งเป็น 2 ฉบับฉบับละ 38 ข้อ มีแบบทดสอบร่วมภายใน จำนวน 20 ข้อ ผลการเปรียบเทียบ พบว่าประสิทธิภาพของการปรับเทียบคะแนนระหว่างแบบแผนข้อสอบร่วมกับแบบแผนผู้สอบร่วมในแต่ละวิธีไม่แตกต่างกัน ความคงที่ของวิธีการปรับเทียบคะแนนโดยแบบแผนผู้สอบร่วมได้ผลว่า วิธีการปรับเทียบโดยใช้วิธีอิกวิเปอร์เซ็นไทล์มีความคงที่มากกว่าวิธี

เชิงเส้นตรงแต่มีความคงที่พอ ๆ กับวิธีใช้ IRT แบบ 3 พารามิเตอร์ ความคงที่ของวิธีการปรับเทียบคะแนนในแบบแผนข้อสอบร่วมได้ผลเหมือนกับแบบแผนผู้สอบร่วม

พิชัย ละแมนชัย (2538 :131-134) ได้ทำการศึกษาเพื่อหาขนาดของกลุ่มตัวอย่างขั้นต่ำ ที่ทำให้การปรับเทียบคะแนนในแนวระดับ มีประสิทธิภาพ โดยศึกษา ตามแนวทฤษฎีการตอบสนองรายข้อ โมเดล 1 และ 3 พารามิเตอร์ ในแบบแผนกลุ่มสมมูล และแบบแผนข้อสอบร่วมที่มีความยาวของแบบทดสอบต่าง ๆ กัน คือ 30 50 70 90 110 130 และ 150 ข้อ ข้อมูลที่ใช้ในการวิจัยครั้งนี้ เป็นการจำลองจากโปรแกรม IRTDATA เกณฑ์ที่ใช้ในการตัดสินคือ การหาค่าดัชนีความแตกต่างที่อยู่ในแนวระดับน่าพอใจอย่างมากผลการวิจัยพบว่า

(1) ขนาดของกลุ่มตัวอย่างขั้นต่ำ ตามโมเดล 1 พารามิเตอร์ ในแบบแผนกลุ่มสมมูลที่มีความยาวของแบบทดสอบ เป็น 30, 50, 70, 90, 110, 130 และ 150 ข้อ คือ 1000, 850, 700, 600, 550, 450, และ 350 ตามลำดับ

(2) ขนาดของกลุ่มตัวอย่างขั้นต่ำ ตามโมเดล 3 พารามิเตอร์ในแบบแผนกลุ่มสมมูลที่มีความยาวของแบบทดสอบเป็น 30, 50, 70, 90, 110, 130, และ 150 ข้อ คือ 2550, 2350, 2150, 1750, 1550, 1400 และ 1150 ตามลำดับ

(3) ขนาดกลุ่มตัวอย่างขั้นต่ำ ตามโมเดล 1 พารามิเตอร์ในแบบแผนข้อสอบร่วมที่มีความยาวแบบทดสอบเป็น 30, 50, 70, 90, 110, 130, และ 150 ข้อ คือ 750, 650, 500, 400, 350, 300 และ 200 ตามลำดับ

(4) ขนาดของกลุ่มตัวอย่างขั้นต่ำตามโมเดล 3 พารามิเตอร์ ในแบบแผนข้อสอบร่วมที่มีความยาวแบบทดสอบเป็น 30, 50, 70, 90, 110, 130, และ 150 ข้อ คือ 2000, 1800, 1600, 1350, 1200, 1000 และ 800 ตามลำดับ

(5) โมเดลการตอบสนองรายข้อ 1 และ 3 พารามิเตอร์ ทั้งแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วม ขนาดกลุ่มตัวอย่างขั้นต่ำในแต่ละความยาวแบบทดสอบแตกต่างกัน โดยโมเดล 1 พารามิเตอร์ ใช้ขนาดกลุ่มตัวอย่างขั้นต่ำที่น้อยกว่า

(6) โมเดลการตอบสนองรายข้อที่ใช้คือ 1 และ 3 พารามิเตอร์ ทั้งแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วม ขนาดกลุ่มตัวอย่างขั้นต่ำในแต่ละความยาวแบบทดสอบแตกต่างกัน โดยแบบแผนข้อสอบร่วมใช้ขนาดกลุ่มตัวอย่างขั้นต่ำที่น้อยกว่า

### 3.2.2 การศึกษาเปรียบเทียบผลการใช้รูปแบบการปรับเทียบคะแนน

ภาวิณี ศรีสุขวัฒนานันท์ (2529 :155-170) ได้ศึกษาเปรียบเทียบผลของการปรับเทียบคะแนน 3 วิธีคือ วิธีอิกวิเปอร์เทนไทล์ วิธีเชิงเส้นตรง และวิธีอิงทฤษฎีตอบสนองข้อสอบแบบ 3 พารามิเตอร์ โดยใช้แบบทดสอบร่วมภายในที่ต่างกัน 3 ขนาด คือ ขนาดร้อยละ 20 (70ข้อ)

ร้อยละ 40 (14 ข้อ) และขนาดร้อยละ 60 (21 ข้อ) แบบทดสอบมีทั้งหมด 100 ข้อ จัดแบ่งเป็น 2 ฉบับ ๆ ละ 35 ข้อ แบบทดสอบทั้ง 2 ฉบับมีความยากใกล้เคียงกัน การประเมินผลการปรับเทียบคะแนนมี 2 ลักษณะ คือ

1. ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน (SEE) ด้วยวิธีที่ต่างกันของแบบทดสอบแต่ละกรณี และ
2. ตรวจสอบความเพียงพอของวิธีการปรับเทียบคะแนน แต่ละวิธี โดยใช้ดัชนีเปรียบเทียบความแตกต่าง (Index C) ที่ได้วิเคราะห์จากกลุ่มตรวจสอบผลกลุ่มตัวอย่างแยกเป็น 2 กรณี คือ กรณีแบบสอบคัดเลือก 2 กลุ่ม ๆ ละ 1,500 คน และกลุ่มตรวจสอบผลอีก 1 กลุ่ม จำนวน 1,500 คน อีกกรณีหนึ่งคือ กรณีแบบสอบวัดผลสัมฤทธิ์ซึ่งมีกลุ่มตัวอย่างเช่นเดียวกับกรณีแรก ผลการวิจัยพบว่า การใช้แบบทดสอบร่วมที่ยาวกว่าให้ประสิทธิภาพของการปรับเทียบคะแนนสูงกว่า (คลาดเคลื่อนน้อยกว่า) และให้ผลในระดับที่น่าพอใจมากกว่า ทั้งสองกรณีของแบบทดสอบ การศึกษาวิธีการปรับเทียบคะแนน พบว่าในกรณีแบบสอบคัดเลือก นั้นวิธีที่ให้ความเพียงพอมากที่สุด คือ วิธีอิกวิเปอร์เซ็นไทล์ รองลงมาคือ วิธีทฤษฎีการตอบข้อสอบ ซึ่งให้ผลที่แตกต่างกันในกรณีแบบสอบวัดผลสัมฤทธิ์ ที่พบว่า วิธีที่ให้ความเพียงพอมากที่สุด คือ วิธีเชิงเส้นตรง รองลงมาคือ วิธีอิกวิเปอร์เซ็นไทล์และเมื่อทดสอบความแตกต่างของค่าดัชนี C เป็นรายคู่ พบว่า กรณีแบบสอบวัดผลสัมฤทธิ์ วิธีเชิงเส้นตรงมีความเพียงพอมากกว่าวิธีที่ใช้ทฤษฎีการตอบข้อสอบอย่างมีนัยสำคัญทางสถิติที่ .05 ส่วนการเปรียบเทียบวิธีคู่อื่นไม่สามารถสรุปความแตกต่างอย่างมีนัยสำคัญ

เวรดี อินทสระ (2530 :64-69) ได้เปรียบเทียบความคลาดเคลื่อนและความเพียงพอในการปรับเทียบคะแนน 2 วิธี คือ วิธีอิงทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ กับวิธีวิเคราะห์องค์ประกอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2529 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดพัทลุง โดยแบบทดสอบที่ใช้ปรับเทียบคะแนนเป็นแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ เรื่องอัตราส่วน ร้อยละ ซึ่งผู้วิจัยสร้างขึ้นแล้วนำข้อสอบที่วิเคราะห์หาคุณภาพแล้วมาจัดแบ่งเป็นแบบสอบ 2 ชุด ชุดละ 45 ข้อ และแต่ละชุดมีข้อสอบรวมภายใน 15 ข้อ นำไปดำเนินการสอบกลุ่มตัวอย่างอื่น จำนวน 1,557 คน โดยทุกคนต้องทำแบบสอบทั้ง 2 ชุด หลังจากนั้นจัดสุ่มใช้คะแนนผลการสอบจากแบบทดสอบชุดที่ 1 จำนวน 779 คน และอีก 778 คน ใช้คะแนนจากแบบทดสอบชุดที่ 2 ส่วนคะแนนที่เหลือใช้เป็นคะแนนตรวจทานผลของการปรับเทียบคะแนนแต่ละรูปแบบ

ผลการวิจัยพบว่า การปรับเทียบคะแนน รูปแบบอิงทฤษฎีการตอบข้อสอบให้ ความเพียงพอ ในระดับที่น่าพอใจ ส่วนรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบให้ ความเพียงพอในระดับปานกลาง สำหรับการเปรียบเทียบความคลาดเคลื่อนของการปรับเทียบคะแนน

รูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบมีความคลาดเคลื่อนสูงกว่ารูปแบบอิงทฤษฎีการตอบข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (t-test) แต่เมื่อพิจารณาในแต่ละช่วงคะแนนพบว่ารูปแบบทั้งสองให้คะแนนสมมูลใกล้เคียงกันในช่วงตรงกลางของการแจกแจงคะแนน (คะแนน 19-23) ส่วนช่วงคะแนนต่ำและคะแนนสูง พบว่าวิธีทั้งสองให้ผลแตกต่างในทางตรงกันข้าม กล่าวคือในช่วงคะแนนต่ำ คะแนน 1-18) นั้น รูปแบบอิงทฤษฎีการตอบข้อสอบมีคะแนนสมมูลสูงกว่ารูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ ส่วนช่วงคะแนนสูง (คะแนน 24-35) รูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบมีคะแนนสมมูลสูงกว่ารูปแบบการอิงทฤษฎีการตอบข้อสอบ

วรรณตร พิชิตเกริกพล (2537:94-98) ได้ศึกษาเปรียบเทียบคุณภาพของวิธีการเปรียบเทียบคะแนนระหว่างแบบอควิเปอร์เซ็นไทล์กับแบบเชิงเส้นตรง โดยศึกษากรณีของลักษณะแบบทดสอบที่มีความยากใกล้เคียงกัน และแตกต่างกัน สอบกับกลุ่มตัวอย่างที่มีระดับความสามารถใกล้เคียงกันและแตกต่างกัน การประเมินผลการเปรียบเทียบคะแนนใช้การวิเคราะห์กลุ่มตรวจสอบผล ซึ่งเป็นเกณฑ์ภายนอก เพื่อตรวจสอบความคลาดเคลื่อนของคะแนนที่ได้จากการเปรียบเทียบคะแนนโดยตรงกับผลสุดท้ายของการเปรียบเทียบคะแนน กลุ่มตัวอย่างเป็นนักเรียนชั้น มัธยมศึกษาปีที่ 2 ปีการศึกษา 2534 จังหวัดพระนครศรีอยุธยา โดยแยกเป็นกลุ่มตัวอย่างเปรียบเทียบคะแนน 900 คน และกลุ่มตัวอย่างตรวจสอบผล 240 คน เครื่องมือที่ใช้เป็นแบบทดสอบวิชาคณิตศาสตร์ จำนวน 3 ฉบับ ๆ ละ 40 ข้อ โดยมีข้อสอบรวมฉบับละ 20 ข้อ แล้วสร้างตารางแปลงคะแนนสมมูลเพื่อวิเคราะห์ดัชนีความแตกต่างมาตรฐาน (C) และความคลาดเคลื่อนของคะแนนที่ได้จากการเปรียบเทียบ จากกลุ่มตัวอย่างตรวจสอบผล

ผลการวิจัยพบว่า (1) กรณีแบบทดสอบมีความยากใกล้เคียงกัน และกลุ่มตัวอย่างมีระดับความสามารถใกล้เคียงกัน วิธีเชิงเส้นตรงมีคุณภาพดีกว่า วิธีอควิเปอร์เซ็นไทล์ (2) กรณีแบบทดสอบมีความยากใกล้เคียงกันและกลุ่มตัวอย่างมีระดับความสามารถแตกต่างกัน วิธีอควิเปอร์ เซ็นไทล์ให้คุณภาพดีกว่าวิธีเชิงเส้นตรง (3) ในกรณีแบบทดสอบมีความยากแตกต่างกัน ในกลุ่มตัวอย่างที่มีระดับความสามารถใกล้เคียงกัน วิธีอควิเปอร์เซ็นไทล์ให้คุณภาพดีกว่าวิธีเชิงเส้นตรง (4) ในกรณีแบบทดสอบมีความยากแตกต่างกัน ในกลุ่มตัวอย่างที่มีระดับความสามารถแตกต่างกัน การเปรียบเทียบทั้ง 2 วิธีให้ความเพียงพอในระดับไม่น่าพอใจ

พรพิมล นาคเวช (2537 : 99-101) ได้ศึกษาคุณภาพของการเปรียบเทียบคะแนนในแนวตั้งโดยใช้ทฤษฎีการตอบสนองข้อสอบ สำหรับผู้มีความสามารถต่างกัน 3 ระดับ ซึ่งได้แก่นักเรียนชั้น ม.1 ม.2 และ ม.3 โดยวิเคราะห์ความคลาดเคลื่อนมาตรฐานและความเพียงพอของการเปรียบเทียบคะแนน กลุ่มตัวอย่างเป็นนักเรียนชั้น ม.1 ม.2 และ ม.3 สังกัดกองการมัธยมศึกษา

กรมสามัญศึกษา จำนวน 6,429 คน ที่ได้จากการสุ่มกลุ่มตัวอย่างแบบแบ่งชั้นหลายชั้นตอน แบ่งเป็นกลุ่มตัวอย่างเปรียบเทียบคะแนนระหว่างแบบทดสอบ จำนวน 3,279 คน และกลุ่มตัวอย่าง สอบทานผลเพื่อนำผลไปตรวจสอบคุณภาพของการเปรียบเทียบคะแนน จำนวน 3,105 คน เครื่องมือที่ใช้เป็นแบบทดสอบคณิตศาสตร์ เรื่องสมการและอสมการ สำหรับระดับชั้น ม.1 ม. 2 และ ม.3 รวม 3 ฉบับ ๆ ละ 25 ข้อ โดยมีข้อสอบร่วมระหว่างระดับชั้นที่ติดกันระดับชั้นละ 5 ข้อ ผล การศึกษาพบว่าความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบคะแนนมีค่าอยู่ระหว่าง .031 - .051 ซึ่งเป็นความคลาดเคลื่อนมาตรฐานที่ค่อนข้างต่ำ และผลการวิเคราะห์ความเพียงพอของการ เปรียบเทียบคะแนนตามแนวตั้งพบว่า ค่าดัชนีความแตกต่างของการเปรียบเทียบคะแนนอยู่ในระดับที่ น่าพอใจ

นอกจากนี้ สุณิสา จุ้ยม่วงศรี (2537 : 88-92) ได้ทำการตรวจสอบคุณภาพของการ เปรียบเทียบคะแนนเชิงเส้นตรงจากการใช้แบบทดสอบร่วมภายใน ขนาดความยาวต่างกัน 4 ขนาด คือ 10, 15, 20 และ 25 ข้อ โดยการเปรียบเทียบความคลาดเคลื่อนมาตรฐาน ของการเปรียบเทียบ คะแนน (SEE) และดัชนีความแตกต่าง (C) จากการวิเคราะห์กลุ่มสอบทานผล เครื่องมือที่ใช้เป็น แบบทดสอบวัดผลสัมฤทธิ์ปลายภาค วิชาคณิตศาสตร์ระดับมัธยมศึกษาปีที่ 2 ภาคต้น ของกลุ่ม โรงเรียนมัธยมศึกษาตอนต้น จังหวัดขอนแก่นกลุ่มที่ 2 จำนวน 2 ฉบับ ๆ ละ 60 ข้อ สร้างในปี การศึกษา 2532 (ฉบับ x) และ 2533 (ฉบับ y) และแบบทดสอบร่วมที่ผู้วิจัยสร้างขึ้นเพื่อใช้ในการ เปรียบเทียบคะแนนจำนวน 25 ข้อ กลุ่มตัวอย่างสำหรับเปรียบเทียบคะแนนมีจำนวน 810 ข้อ สุ่มอย่าง มีระบบให้แต่ละคนสอบแบบทดสอบฉบับ X หรือฉบับ Y คนละฉบับสลับกันและกลุ่มตัวอย่างสอบ ทานผล จำนวน 117 คน โดยใช้สอบแบบทดสอบทั้งสองฉบับ สลับกันด้วยการสุ่มอย่างมีระบบ ผลการวิจัยพบว่า (1) คะแนนสมมูลของแบบทดสอบฉบับ Y น้อยกว่าคะแนนของแบบทดสอบ ฉบับ X (2) ค่าความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบคะแนนเชิงเส้นตรง ในแบบทดสอบ ร่วมที่ยาวกว่าให้ค่าน้อยกว่า (3) ค่าประสิทธิภาพสัมพัทธ์ที่ได้จากอัตราส่วนร้อยละของความ คลาดเคลื่อนมาตรฐาน เมื่อใช้แบบทดสอบร่วมขนาด 10 15 และ 20 ข้อ เทียบกับแบบทดสอบร่วม ขนาด 25 ข้อ ให้ค่า 86.26 89.08 และ 93.30 ตามลำดับ (4) ค่าดัชนีความแตกต่างของการ เปรียบเทียบคะแนน จากการใช้แบบทดสอบร่วม 4 ขนาด ให้ผลอยู่ในระดับที่ยอมรับได้ทั้งสิ้น

สุมนา ไสตถิผลอนันต์ (2539 : 99-101) ได้ศึกษาวิธีการเปรียบเทียบคะแนนแบบ เส้นตรง 4 วิธี โดยมีข้อตกลงว่าแบบทดสอบที่นำมาเปรียบเทียบคะแนนนั้น ต้องเป็นไปตาม แบบจำลองคะแนนจริงสัมพันธ์ และมีความเชื่อมั่นเท่าเทียมกัน กลุ่มตัวอย่างเป็นนักศึกษา พยาบาลชั้นปีที่ 1 และชั้นปีที่ 2 หลักสูตรพยาบาลศาสตร์ และพยาบาลศาสตร์ระดับต้น ปี การศึกษา 2538 ของวิทยาลัยพยาบาลในเขตภาคกลาง จำนวน 864 คน โดยวิธีการสุ่มอย่างง่าย มาจัดเป็นกลุ่มเทียบคะแนน 12 กลุ่ม ที่มีความสามารถทางสมองเท่าเทียมกัน เครื่องมือที่ใช้เป็น

แบบทดสอบวัดการคิดเอกลัทธิแบบสัมพันธ์ที่มีฟอร์มต่างกัน 3 ฟอร์ม คือ ฟอร์ม A ฟอร์ม B และ ฟอร์ม C โดยใช้ฟอร์ม C เป็นแบบทดสอบหลักที่มีลักษณะการใช้เป็นแบบทดสอบเชื่อมโยง และเป็นข้อสอบร่วม มีขนาดความยาว 3 ขนาด คือ เป็นสัดส่วน 0.2, 0.3 และ 0.4 ของแบบทดสอบ เทียบคะแนนฟอร์ม A การวิเคราะห์ข้อมูลใช้วิธีการคำนวณจากสูตร 4 วิธี แล้วจึงเปรียบเทียบ คะแนนแปลงและความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน ด้วยวิธีการหาความแปรปรวนแบบการวัดซ้ำผลการวิจัยพบว่า คะแนนแปลงที่ได้จากวิธีการคำนวณด้วยสูตร 4 วิธี ผลการใช้แบบทดสอบเชื่อมโยงกับข้อสอบร่วม ผลของการเปรียบเทียบขนาดความยาวของแบบทดสอบหลัก และผลของการเปรียบเทียบความคลาดเคลื่อนมาตรฐานที่เกิดจากการใช้วิธีการคำนวณด้วยสูตร 4 วิธี แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

3.2.3 การนำวิธีการปรับเทียบคะแนนไปประยุกต์ ใช้ในการศึกษา ได้มีผู้นำวิธีการปรับเทียบคะแนนตามแนวตั้งและแนวนอนไปใช้สร้างตารางเทียบคะแนนแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ใช้ในการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนและใช้ในการปรับคะแนนเฉลี่ยสะสมของนักเรียนชั้นมัธยมศึกษาตอนปลายดังนี้

สุรัตน์ ขวัญบุญจันทร์ (2531 : 77-80) ได้สร้างแบบทดสอบ และตารางปรับเทียบคะแนนในแนวนอน วิชาวิทยาศาสตร์ เรื่องพลังงานและการเปลี่ยนแปลง ที่วิเคราะห์ด้วยราล์ช โมเดล กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2531 ในจังหวัดชัยภูมิ จำนวน 1,391 คน โดยใช้วิธีการสุ่มแบบหลายขั้นตอน ใช้แบบทดสอบ 2 ฉบับ ฉบับละ 60 ข้อ มีข้อสอบร่วม 10 ข้อ ผลการศึกษาพบว่า แบบทดสอบ 2 ฉบับ มีค่าความยากตั้งแต่  $-1.38$  ถึง  $2.25$  และ  $-1.63$  ถึง  $1.17$  มีค่า  $t$ -test ตั้งแต่  $-1.70$  ถึง  $1.71$  และ  $-1.40$  ถึง  $2.09$  ความเชื่อมั่นของแบบทดสอบที่คำนวณจาก KR-20 เท่ากับ  $.76$  และ  $.78$  ตามลำดับ เมื่อนำคะแนนดิบจากแบบทดสอบ 2 ฉบับ มาสร้างตารางเทียบคะแนนภายใต้ความสามารถที่เท่ากัน ได้คะแนนดิบแต่ละค่าเท่ากันทุกคะแนน และได้ค่าดัชนีความเพียงพอของความคลาดเคลื่อนในการเทียบคะแนนเท่ากับ  $.1019$  ซึ่งอยู่ในระดับที่น่าพอใจมาก แบบทดสอบแต่ละฉบับมีข้อสอบร่วมระหว่างชั้นฉบับละ 10 ข้อ นอกจากนี้ผู้วิจัยได้สร้างตารางเทียบคะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 1 2 และ 3 จากความสามารถที่เท่ากัน ผลการวิจัยสามารถนำไปใช้ในการศึกษาความก้าวหน้าทางการเรียน เรื่องสมการและอสมการของนักเรียนได้ และได้มีผู้นำวิธีการปรับเทียบคะแนนไปประยุกต์ใช้ในการศึกษาความก้าวหน้าทางการเรียนคือ

สุจินดา ผ่องอักษร (2533 : 143-157) นำวิธีการปรับเทียบคะแนนโดยใช้แบบจำลองของราล์ชไปใช้ในการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วง

ระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) ในการวิจัยครั้งนี้นอกจากศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนแล้วยังมีการเปรียบเทียบความแตกต่างของผลสัมฤทธิ์ทางการเรียนระหว่างปีการศึกษาต่าง ๆ หลังจากที่ได้เทียบคะแนนแล้ว และเปรียบเทียบสัดส่วนจำนวนนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนผ่านเกณฑ์ที่น่าพอใจระหว่างการเทียบคะแนนกับไม่เทียบคะแนน กลุ่มตัวอย่างการวิจัยเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ตามโครงการประเมินคุณภาพนักเรียนชั้นประถมศึกษาปีที่ 6 ระดับประเทศ ในจังหวัดปทุมธานี แบบทดสอบที่ใช้เป็นแบบทดสอบวัดผลสัมฤทธิ์ในกลุ่มทักษะ (วิชาคณิตศาสตร์และภาษาไทย) ซึ่งสำนักงานคณะกรรมการประถมศึกษาแห่งชาติสร้างขึ้นและใช้ทดสอบนักเรียนเมื่อปีการศึกษา 2529 2530 และ 2531 แบบแผนการเก็บรวบรวมข้อมูลใช้รูปแบบผู้สอบร่วม เนื่องจากแบบทดสอบที่ใช้ทั้ง 3 ฉบับ ไม่มีข้อสอบร่วมกัน

ผลการวิจัยพบว่า จากการปรับเทียบคะแนนระหว่างแบบทดสอบ 3 ฉบับคะแนนที่เทียบแล้วของแบบทดสอบ 3 ฉบับซึ่งตรงกับคะแนนดิบค่าเดียวกันมีค่าแตกต่างกันทั้งแบบทดสอบวิชาคณิตศาสตร์และแบบทดสอบภาษาไทย เมื่อศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียน พบว่าผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ในปีการศึกษา 2530 สูงกว่าปีการศึกษา 2529 แต่ลดลงในปีการศึกษา 2531 สำหรับผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยพบว่าอัตราความก้าวหน้าของคะแนนเฉลี่ยเพิ่มขึ้นทุกปีการศึกษา และอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจเพิ่มขึ้นทุกปีการศึกษา

สำหรับ วรณดี แสงประทีปทอง (2538 :123-133) ได้ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทย ของนักเรียนชั้น ม.1 ม.2 และ ม.3 โดยใช้วิธีการปรับเทียบคะแนนตามแนวตั้ง ตามทฤษฎีตอบสนองของข้อสอบ แบบจำลองโลจิสติกสามพารามิเตอร์ ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนทั้งในกลุ่มรวมและกลุ่มย่อยตามระดับความสามารถของนักเรียน กลุ่มตัวอย่าง เป็นนักเรียนชั้น ม.1 ม.2 และ ม.3 ในจังหวัดสิงห์บุรี ภาคเรียนที่ 1 ปีการศึกษา 2537 โดยแบ่งเป็น 2 กลุ่ม คือ กลุ่มตัวอย่างปรับเทียบคะแนนเป็นนักเรียน ม.1 ม.2 และ ม.3 จำนวน 1245, 1280 และ 1171 คนตามลำดับ และกลุ่มสอบทานผลเป็นนักเรียนชั้น ม.1 ม.2 และ ม.3 จำนวน 373 384 และ 352 คนตามลำดับ เครื่องมือที่ใช้เป็นแบบทดสอบปรับเทียบคะแนน 3 ฉบับ คือแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทย ท101 (ม.1) ท203 (ม.2) และ ท305(ม.3) และแบบทดสอบร่วมเป็นแบบทดสอบความสามารถทั่วไปทางภาษาไทย 1 ฉบับ รวบรวมข้อมูลจากสถานการณ์จริง ผลการวิจัยพบว่า

(1) ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทย ของนักเรียนชั้น ม.2 ไป ม.3 ก้าวหน้ากว่าปกติ แต่ผลสัมฤทธิ์ของการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.1 ไป ม.2 ก้าวหน้าตามปกติ

(2) ในกลุ่มนักเรียนที่มีความสามารถสูง ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทย ของนักเรียนชั้น ม.2 ไป ม.3 และ ม.1 ไป ม.2 ก้าวหน้าตามปกติ

(3) ในกลุ่มนักเรียนที่มีความสามารถปานกลางผลสัมฤทธิ์ทางการเรียนวิชา ภาษาไทยของนักเรียน ชั้น ม.2 ไป ม.3 และ ม.1 ไป ม.2 ก้าวหน้ามากกว่าปกติ

(4) ในกลุ่มนักเรียนที่มีความสามารถต่ำผลสัมฤทธิ์ทางการเรียนวิชา ภาษาไทยของนักเรียนชั้น ม.2 ไป ม.3 ก้าวหน้ามากกว่า ปกติ แต่ ม.1 ไป ม.2 ก้าวหน้าตามปกติ

นอกจากนี้ได้มีผู้นำวิธีการปรับเทียบคะแนนไปใช้ปรับคะแนนเฉลี่ยสะสมของ นักเรียนชั้นมัธยมศึกษาตอนปลาย โดยศึกษาในเชิงเปรียบเทียบระหว่างรูปแบบต่าง ๆ ดังนี้

กนิษฐา แสนแก้ว (2541 : 110-114) ได้เปรียบเทียบคุณภาพของวิธีปรับคะแนน เฉลี่ยสะสม 3 วิธี ได้แก่ วิธีอิกวิเปอร์เซ็นไทล์ วิธีเชิงเส้นตรงโดยการแปลงเป็นคะแนนมาตรฐาน (Z-Score) และวิธีปรับโดยใช้ทฤษฎีการตอบสนองของข้อสอบภายใต้ Graded Response Model โดยใช้ แบบทดสอบความถนัดทางการเรียนเป็นแบบทดสอบร่วม กลุ่มตัวอย่างเป็นนิสิตชั้นปีที่ 2 คณะ ครุศาสตร์ ศึกษาศาสตร์ ที่เข้าศึกษาในปีการศึกษา 2540 ของจุฬาลงกรณ์มหาวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ มหาวิทยาลัยเกษตรศาสตร์ และมหาวิทยาลัยศิลปากร จำนวน 559 คน ผลการวิจัยพบว่า เมื่อใช้คะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปีที่ 1 เป็นเกณฑ์ วิธีปรับ โดยใช้ทฤษฎีการตอบสนองของข้อสอบมีคุณภาพสูงสุด รองลงมาคือ วิธีเชิงเส้นตรง และวิธีอิกวิ เเปอร์เซ็นไทล์ เมื่อใช้คะแนนสอบเข้ามหาวิทยาลัยเป็นเกณฑ์ วิธีอิกวิเปอร์เซ็นไทล์มีคุณภาพสูงสุด รองลงมาคือวิธีปรับโดยใช้ทฤษฎีการตอบสนองของข้อสอบ และวิธีเชิงเส้นตรง ตามลำดับ

ในปีเดียวกัน สุภาภรณ์ คงทวี (2541 : 205 – 213) ได้เปรียบเทียบประสิทธิภาพ ของวิธีปรับคะแนนเฉลี่ยสะสม 3 วิธี ได้แก่ วิธีอิกวิเปอร์เซ็นไทล์ วิธีเชิงเส้นตรงรูปแบบที่ 3 ของแอง กอฟ (Angoff, 1971) และวิธีปรับโดยใช้ทฤษฎีการตอบสนองของข้อสอบภายใต้ Graded Response Model ของ ยิง (Young, 1990) โดยใช้แบบทดสอบเข้ามหาวิทยาลัยเป็นแบบทดสอบ ร่วม กลุ่มตัวอย่างเป็นนิสิต นักศึกษาชั้นปีที่ 1 ของมหาวิทยาลัยต่าง ๆ 6 แห่ง ในปีการศึกษา 2540 ที่จบชั้นมัธยมศึกษาตอนปลายในปีการศึกษา 2539 โดยวิเคราะห์แยกเป็น 6 กลุ่ม ได้แก่กลุ่ม สาขาวิชาวิทยาศาสตร์วิศวกรรมศาสตร์ สถาปัตยกรรม พาณิชยศาสตร์ และการบัญชี มนุษยศาสตร์ – สังคมศาสตร์ (แผนการสอนศิลป์) และมนุษยศาสตร์ – สังคมศาสตร์ (แผนการสอน วิทยาศาสตร์) ผลการวิจัยพบว่าเมื่อใช้คะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปีที่ 1 เป็นเกณฑ์วิธี ปรับเทียบคะแนนเชิงเส้นตรงมีประสิทธิภาพในการพยากรณ์สูงกว่าวิธีอิกวิเปอร์เซ็นไทล์ในกลุ่ม สาขาวิชาวิทยาศาสตร์ วิธีอิกวิเปอร์เซ็นไทล์ และวิธีปรับโดยใช้ทฤษฎีการตอบสนองของข้อสอบมี ประสิทธิภาพในการพยากรณ์สูงกว่าวิธีปรับเชิงเส้นตรง สำหรับกลุ่มสาขาวิชาวิศวกรรมศาสตร์

ส่วนในสาขาวิชาอื่น ๆ วิธีปรับคะแนนเฉลี่ยสะสมทั้ง 3 วิธีมีประสิทธิภาพในการพยากรณ์ไม่ต่างกัน

สำราญ มีแจ้ง (2542: 19-36) ได้ทำการเปรียบเทียบคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ที่มีวิธีแปลงคะแนนผลการเรียนสะสมต่างกัน และพบว่า 1) คะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ในส่วนของที่ได้จากการแปลงผลการเรียนสะสมตามวิธีของทบวง วิธีการแปลงเป็นคะแนนมาตรฐาน และวิธีการเทียบมาตรา แตกต่างกันอย่างมีนัยสำคัญทางสถิติ 2.1) โรงเรียนที่มีขนาดต่างกันมีคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาในส่วนของที่ได้จากการแปลงคะแนนผลการเรียนสะสมตามวิธีของทบวงไม่แตกต่างกัน 2.2) โรงเรียนที่มีขนาดต่างกันมีคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษา ในส่วนของที่ได้จากการแปลงคะแนนผลการเรียนสะสมตามวิธีการแปลงเป็นคะแนนมาตรฐานแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 2.3) โรงเรียนที่มีขนาดต่างกันมีคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาในส่วนของที่ได้จากการแปลงคะแนนผลการเรียนสะสมตามวิธีการเทียบมาตราแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 3) คะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ ในส่วนของที่ได้จากการแปลงคะแนนผลการเรียนสะสมในแต่ละวิธี มีความสัมพันธ์กับคะแนนผลการเรียนสะสมสูงในทางบวก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 4) ความสัมพันธ์ระหว่างคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาในส่วนของที่ได้จากการแปลงคะแนนผลการเรียนสะสมตามวิธีของทบวงกับวิธีการแปลงเป็นคะแนนมาตรฐาน และวิธีการเทียบมาตรา มีความสัมพันธ์กันสูงในทางบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

สุภมาส อังศุโชติ (2543 : 183-187) ได้ทำการศึกษาเพื่อ (1)วิเคราะห์ผลการปรับคะแนนเฉลี่ยสะสมของนักเรียนในระดับมัธยมศึกษาตอนปลายด้วยวิธีปรับคะแนนเฉลี่ยสะสม 5 วิธี คือ วิธีปรับคะแนนเฉลี่ยสะสมเชิงเส้นตรงตาม Design IV C-2 ของ Angoff (ANGOFF-4C2) วิธีปรับคะแนนเฉลี่ยสะสมโดยใช้ทฤษฎีการตอบสนองข้อสอบภายใต้ Graded Response Model (IRT-GRM) วิธีปรับคะแนนเฉลี่ยสะสมด้วยการวิเคราะห์ด้วยแบบเชิงเส้นทั่วไป (GLM-MODEL) วิธีปรับคะแนนเฉลี่ยสะสมด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันโมเดลการวัดคอนเจนเนอริค 1 องค์ประกอบ (CON-CFA) และวิธีปรับคะแนนเฉลี่ยสะสมด้วยโมเดลหลายฟาเซทของราสซ (RASCH-FACET) และ (2) เพื่อเปรียบเทียบความตรงเชิงทำนายระหว่างวิธีปรับคะแนนเฉลี่ยสะสมทั้ง 5 วิธี โดยใช้คะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 1 และ ปี 2 เป็นเกณฑ์ กลุ่มตัวอย่างในการปรับคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย ได้แก่ นักเรียนจำนวน 5,919 คนที่สำเร็จการศึกษาในปีการศึกษา 2539 จากโรงเรียนสังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ที่มีคุณภาพต่างกัน 3 ระดับในเขตกรุงเทพมหานครและภาคกลาง จำนวน 28 โรงเรียน และใช้เฉพาะกลุ่มตัวอย่างที่ผ่านการสอบคัดเลือกเข้าศึกษาในมหาวิทยาลัยของรัฐในปีการศึกษา 2540 จำนวน

1,029 คน ในการตรวจสอบความตรงเชิงทำนาย ข้อมูลที่ใช้ในการวิจัยประกอบด้วย คะแนนเฉลี่ยระดับมัธยมศึกษาตอนปลาย คะแนนสอบเข้ามหาวิทยาลัยและคะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 1 และปี 2 ของผู้ที่สอบคัดเลือกผ่าน

ผลการวิจัยพบว่า

1. โรงเรียนที่มีระดับคุณภาพในกลุ่มเดียวกันหรือต่างกลุ่มกันมีมาตรฐานการใช้คะแนนเฉลี่ยสะสมต่างกันเป็น 4 ลักษณะ คือ (1) ให้คะแนนเฉลี่ยสะสมสูงกว่าที่ควรจะเป็นในทุกช่วงคะแนนเฉลี่ยสะสม (2) ให้คะแนนเฉลี่ยสะสมต่ำกว่าที่ควรจะเป็นในทุกช่วงคะแนนเฉลี่ยสะสม (3) ให้คะแนนเฉลี่ยสะสมสูงกว่าที่ควรจะเป็นในช่วงคะแนนเฉลี่ยสะสมต่ำแต่ ให้คะแนนเฉลี่ยสะสมต่ำกว่าที่ควรจะเป็นในช่วงคะแนนเฉลี่ยสะสมสูง และ (4) ให้คะแนนเฉลี่ยสะสมต่ำกว่าที่ควรจะเป็นในช่วงคะแนนเฉลี่ยสะสมต่ำแต่ให้คะแนนเฉลี่ยสะสมสูงกว่าที่ควรจะเป็นในช่วงคะแนนเฉลี่ยสะสมสูง

2. เมื่อใช้คะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 1 และปี 2 ที่ปรับโดยใช้สมการถดถอย และที่ปรับด้วยโมเดลหลายพหุของราล์ซเป็นเกณฑ์ให้ผลสอดคล้องกันคือ ที่ระดับนัยสำคัญ .05 คะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายที่ปรับด้วยวิธีปรับ 3 วิธี คือ วิธี RASCH-FACET วิธี ANGOFF-4C2 และวิธี IRT-GRM ให้ความตรงเชิงทำนายสูงกว่าคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายเดิม และวิธี Con-CFA ให้ความตรงเชิงทำนายต่ำกว่าคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายเดิม ผลที่ต่างกันคือเมื่อใช้คะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 1 และปี 2 ที่ปรับโดยใช้สมการถดถอยเป็นเกณฑ์ คือ วิธี GLM-MODEL ให้ความตรงเชิงทำนายสูงกว่าคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายเดิมแต่จะไม่แตกต่างกันเมื่อใช้คะแนนเฉลี่ยสะสมระดับมหาวิทยาลัยปี 1 และปี 2 ที่ปรับด้วยโมเดลหลายพหุของราล์ซเป็นเกณฑ์ ส่วนการเปรียบเทียบความตรงเชิงทำนายระหว่างวิธีพบว่า วิธี RASCH-FACET ให้ผลดีที่สุด รองลงมาคือวิธี ANGOFF-4C2 และวิธี IRT-GRM ซึ่งสองวิธีนี้ให้ความตรงเชิงทำนายไม่แตกต่างกัน อันดับที่ดีที่สุดคือวิธี GLM-MODEL และอันดับสุดท้ายคือวิธี CON-CFA

อุทัยวรรณ พงศ์อร่าม (2545 : บทคัดย่อ) ได้ทำการศึกษาขนาดกลุ่มตัวอย่างที่เหมาะสมสำหรับการปรับ เทียบคะแนนด้วยวิธีอิกวิเปอร์เซ็นไทล์และวิธีเชิงเส้นตรงตามแบบจำลองคะแนนจริงสัมพันธ์ ที่มีแบบแผนการปรับเทียบและความยาวของแบบทดสอบแตกต่างกัน โดยมีจุดมุ่งหมายเพื่อศึกษาขนาดของกลุ่มตัวอย่างที่เหมาะสมสำหรับการปรับเทียบคะแนนด้วยวิธีอิกวิเปอร์เซ็นไทล์และวิธีเชิงเส้นตรงตามแบบจำลองคะแนนจริงสัมพันธ์ในแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วมที่มีความยาวของแบบทดสอบต่างกันคือ 60,50,40 และ 30 ข้อกลุ่มตัวอย่างที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ของโรงเรียนมัธยมศึกษา สังกัดกรมสามัญศึกษา จังหวัดชุมพร ภาคเรียนที่ 1 ปีการศึกษา 2544 จำนวน 3,000 คน ซึ่งได้มาโดยการ

กลุ่มสองชั้นตอน แบ่งเป็นสองกลุ่มคือกลุ่มตัวอย่างเปรียบเทียบคะแนน และกลุ่มตัวอย่างสอบทานผล เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ ค101 ชั้นมัธยมศึกษาปีที่1 จำนวนสองฉบับ การวิเคราะห์ข้อมูลใช้วิธีการเปรียบเทียบคะแนนด้วยวิธีอิกิเปอร์เซ็นไทล์ และวิธีเชิงเส้นตรงตามแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วมเพื่อสร้างตารางแปลงคะแนนสมมูล และวิเคราะห์ดัชนีความแตกต่าง(C) หาขนาดกลุ่มตัวอย่างที่เหมาะสมโดยใช้เกณฑ์ตัดสินจากค่าดัชนีความแตกต่าง (C) ที่อยู่ในระดับน่าพอใจและ น่าพอใจมาก

ผลการวิจัยพบว่า

1. ขนาดของกลุ่มตัวอย่างที่เหมาะสมในการเปรียบเทียบคะแนนวิธีเชิงเส้นตรง ในแบบแผน กลุ่มสมมูล ที่มีความยาวของแบบทดสอบเป็น 60, 50, และ 40 ข้อ เป็น 100 คน ส่วน 30 ข้อ เป็น 200 คน สำหรับการเปรียบเทียบคะแนนวิธีอิกิเปอร์เซ็นไทล์ในแบบแผนกลุ่มสมมูลที่มีความยาวของแบบทดสอบเป็น 60 ข้อมีขนาดกลุ่มตัวอย่างที่เหมาะสมเป็น 200 คน และ 50 กับ 40 ข้อ เป็น 300 คน ส่วน 30 ข้อ เป็น 400 คน
2. ขนาดของกลุ่มตัวอย่างที่เหมาะสมในการเปรียบเทียบคะแนนวิธีเชิงเส้นตรง ในแบบแผนข้อสอบร่วม ที่มีความยาวของแบบทดสอบเป็น 60, 50 และ 40 ข้อ เป็น 100 คน ส่วน 30 ข้อ เป็น 300 คน สำหรับการเปรียบเทียบคะแนนวิธีอิกิเปอร์เซ็นไทล์ในแบบแผนข้อสอบร่วม ที่มีความยาวของแบบทดสอบเป็น 60 และ 50 ข้อเป็น 200 คน ส่วน 40 และ 30 ข้อ เป็น 500 คน
3. การเปรียบเทียบคะแนนด้วยวิธีเชิงเส้นตรงและวิธีอิกิเปอร์เซ็นไทล์ทั้งแบบแผนกลุ่มสมมูล และแบบแผนข้อสอบร่วมขนาดของกลุ่มตัวอย่างที่เหมาะสมแตกต่างกันในแต่ละความยาวของแบบทดสอบโดยการเปรียบเทียบคะแนนด้วยวิธีเชิงเส้นตรงใช้กลุ่มตัวอย่างที่น้อยกว่า
4. ตามแบบแผนการเปรียบเทียบคะแนนแบบกลุ่มสมมูล และแบบแผนข้อสอบร่วมทั้งวิธีเชิงเส้นตรงและวิธีอิกิเปอร์เซ็นไทล์ ขนาดของกลุ่มตัวอย่างที่เหมาะสมในแต่ละความยาวของแบบทดสอบ ไม่แตกต่างกัน

ภัทราพร เกษสังข์ (2546 : บทคัดย่อ) ได้ทำการศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาตอนต้น โดยการเทียบคะแนนในแนวตั้งที่ใช้วิธีการที่เหมาะสมโดยมีจุดมุ่งหมายเพื่อศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียน และศึกษาประสิทธิภาพของการเทียบคะแนนในแนวตั้งตามทฤษฎีการตอบข้อสอบแบบโลจิสติกสองและสามพารามิเตอร์ ที่ใช้แบบทดสอบร่วมภายในและแบบทดสอบเชื่อมโยงภายนอก ที่หลากหลายลักษณะระดับความยาก (ยากและปานกลาง) และความยาว (15 10 และ 5 ข้อ) ของแบบทดสอบร่วม จากกลุ่มผู้สอบที่มีความสามารถแตกต่างกันสามระดับชั้น คือมัธยมศึกษาปีที่ 1, 2 และ 3 โดยการวิเคราะห์ความคลาดเคลื่อนมาตรฐานและความเพียงพอของการเทียบคะแนนแต่ละรูปแบบที่ศึกษา แล้วนำรูปแบบการเทียบคะแนนที่เหมาะสม มาใช้ในการศึกษาพัฒนาการ

ความสามารถทางคณิตศาสตร์ของนักเรียนกลุ่มตัวอย่างของการวิจัย เป็นนักเรียนมัธยมศึกษาปีที่ 1, 2 และ 3 สังกัดกองการมัธยมศึกษา กรมสามัญศึกษาในจังหวัดชัยนาท จำนวน 4,041 คน ที่ได้จากการสุ่มกลุ่มตัวอย่างแบบแบ่งชั้น แบ่งกลุ่มตัวอย่างเป็นสองกลุ่ม กลุ่มแรก จำนวน 3,081 คน ใช้สำหรับศึกษาการเทียบคะแนน เพื่อนำผลไปเทียบคะแนนระหว่างแบบทดสอบ และศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ กลุ่มที่สอง จำนวน 960 คน ใช้สำหรับสอบทานผล เพื่อนำผลไปตรวจสอบประสิทธิภาพของการเทียบคะแนน เครื่องมือวัดที่ใช้เป็น แบบทดสอบคณิตศาสตร์ เรื่องสมการและอสมการ ที่มีความเชื่อมั่นของระดับชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 เท่ากับ 0.745, 0.945 และ 0.951 ตามลำดับ และแบบทดสอบวัดความถนัดด้านตัวเลข ที่มีความเชื่อมั่นของระดับชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 เท่ากับ 0.792, 0.847 และ 0.708 ตามลำดับ

ผลการวิจัยสรุปได้ดังนี้

1. การวิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนน ตามรูปแบบที่ศึกษาสำหรับผู้ที่มีความสามารถต่างกันสามระดับ พบว่า มีค่าระหว่าง 0.1011 ถึง 0.1830 เป็นค่าระดับต่ำ ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนต่ำที่สุด เป็นการเทียบคะแนนในแนวตั้ง ตามทฤษฎีการตอบข้อสอบรูปแบบโลจิสติกสองพารามิเตอร์ ที่แบบทดสอบรวมภายใน 15 ข้อ มีความยากของข้อสอบรวมเฉลี่ยอยู่ในระดับยากมาก ( $\bar{x} = 2.2$ )

2. การวิเคราะห์ความเพียงพอของการเทียบคะแนน พบว่า ค่าดัชนีความแตกต่างของการเทียบคะแนน (C) อยู่ในระดับน่าพอใจมาก และระดับน่าพอใจ มีความเพียงพอของการเทียบคะแนนดีที่สุด เป็นการเทียบคะแนนในแนวตั้ง ตามทฤษฎีการตอบข้อสอบรูปแบบโลจิสติกสองพารามิเตอร์ ที่แบบทดสอบรวมภายใน 15 ข้อ ที่ลักษณะความยากของข้อสอบรวมเฉลี่ยระดับยากมาก

3. การศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ จากรูปแบบการเทียบคะแนนที่มีประสิทธิภาพ พบว่า ความสามารถทางคณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ไปชั้น

มัธยมศึกษาปีที่ 2 กับนักเรียนชั้นมัธยมศึกษาปีที่ 2 ไปชั้นมัธยมศึกษาปีที่ 3 มีพัฒนาการสูงขึ้น และนักเรียนชั้นมัธยมศึกษาปีที่ 1 ไปชั้นมัธยมศึกษาปีที่ 3 มีพัฒนาการสูงขึ้นมาก

สุนิสา จุ้ยม่วงศรี (2546) ได้ทำการศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลายค่า โดยมีจุดมุ่งหมายเพื่อศึกษาผลของการเปรียบเทียบคะแนนที่ข้อสอบตรวจให้คะแนนแบบหลายค่าตามวิธีการเทียบคะแนนในทฤษฎีการทดสอบแบบมาตรฐานเดิม ได้แก่ การเทียบคะแนนเชิงเส้นตรงและการเทียบคะแนนอีควิเปอร์เซ็นไทล์ และตามวิธีในทฤษฎีการตอบข้อสอบโมเดลเกรด เรสพอนส์ ได้แก่ การเทียบคะแนนวิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก วิธีโค้งลักษณะแบบทดสอบ และวิธีปรับค่าพารามิเตอร์พร้อม

กัน ซึ่งเทียบคะแนนทั้งคะแนนจริงและคะแนนสังเกต โดยพิจารณาคุณภาพของการเทียบคะแนน แต่ละวิธีจาก ความคล้ายคลึงกันของการแจกแจงคะแนนสะสม ดัชนีความแตกต่าง และดัชนีความแตกต่าง (C) จากการวิเคราะห์กลุ่มสอบทานผล เครื่องมือที่ใช้ในการวิจัย ประกอบด้วย แบบทดสอบเทียบคะแนน 2 ฉบับ เป็นแบบทดสอบเลือกตอบวัดผลสัมฤทธิ์ทางการเรียน คณิตศาสตร์ด้านเหตุผล ได้แก่ ฉบับ X และฉบับ Y ซึ่งผู้วิจัยสร้างขึ้นสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 แต่ละฉบับประกอบด้วยข้อสอบที่มี 5 ตัวเลือก จำนวน 30 ข้อ รวมข้อสอบรวมจำนวน 10 ข้อ แต่ละข้อตรวจให้คะแนน 6 ค่า คือ 0 1 2 3 4 และ 5 คะแนน กลุ่มตัวอย่างเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 สังกัดกรมสามัญศึกษา ในเขตกรุงเทพมหานคร ซึ่งเลือกมาด้วยวิธีการสุ่มแบบแบ่งชั้น จำนวน 2,901 คน แบ่งเป็น กลุ่มตัวอย่างเทียบคะแนนเพื่อนำผลการสอบไปวิเคราะห์เทียบคะแนนแบบทดสอบจำนวน 2,312 คน โดยการสุ่มอย่างมีระบบให้แต่ละคนทำแบบทดสอบฉบับ X หรือ ฉบับ Y เพียงหนึ่งฉบับ ได้ผู้สอบฉบับ X จำนวน 1,155 คน และ ฉบับ Y จำนวน 1,157 คน อีกกลุ่มใช้กลุ่มตัวอย่างสอบทานผลเพื่อนำผลการสอบไปตรวจสอบคุณภาพของการเทียบคะแนนในแต่ละวิธี จำนวน 589 คน แต่ละคนทำแบบทดสอบทั้งสองฉบับ โดยทำแบบทดสอบแต่ละฉบับสลับกัน

ผลการวิจัยสรุปได้ดังนี้

1. คะแนนสมมูลของแบบทดสอบฉบับ Y ส่วนใหญ่สูงกว่าคะแนนของแบบทดสอบฉบับ X จากการเทียบคะแนนทั้ง 5 วิธี 8 รูปแบบ โดยเฉพาะการเทียบคะแนนเชิงเส้นตรงให้ค่าสูงกว่าตลอดช่วงมาตรฐานของ X
2. การแจกแจงคะแนนสะสมระหว่างคะแนนแปลงจากการเทียบคะแนนแต่ละวิธีกับคะแนนจากแบบทดสอบเป้าหมาย (ฉบับ X) มีความคล้ายคลึงกัน
3. ค่าดัชนีคะแนนความแตกต่าง RMSD (Root Mean Square Difference) ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนัก MAD (Mean Absolute Difference) ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนักและ MSD (Mean Signed Difference) ไม่ถ่วงน้ำหนัก ให้ผลคล้ายคลึงกันคือฟังก์ชันการเทียบคะแนนจริงมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความยาก มีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากที่สุด โดยเฉพาะฟังก์ชันการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ สำหรับผลจากดัชนี MSD ถ่วงน้ำหนัก สรุปได้ว่าฟังก์ชันการเทียบคะแนนสังเกต วิธีปรับค่าพารามิเตอร์พร้อมกัน มีความคล้ายคลึงกับฟังก์ชันการเทียบคะแนนพื้นฐานมากที่สุด โดยเฉพาะฟังก์ชันการเทียบคะแนนอีคิวเปอร์เซ็นต์ไทล์ และดัชนี RMSD MAD และ MSD ทั้งถ่วงน้ำหนักและไม่ถ่วงน้ำหนักรวม 6 ดัชนี ให้ผลที่สอดคล้องกัน คือ ฟังก์ชันการเทียบคะแนนสังเกตวิธีโค้งลักษณะแบบทดสอบแตกต่างจากฟังก์ชันการเทียบคะแนนเชิงเส้นตรงมากที่สุด

4. ดัชนีความแตกต่างจากการเทียบคะแนนแต่ละวิธีให้ผลอยู่ในระดับที่ยอมรับได้ทั้งสิ้น คือ มีค่าต่ำกว่ากำลังสองของ 0.05 ของค่าส่วนเบี่ยงเบนมาตรฐานของผลสอบจากฉบับ X และพบว่าความคลาดเคลื่อนของการเทียบคะแนนในแต่ละวิธีจากกลุ่มสอบทานผลไม่แตกต่างกัน

เอกชัย สุขเคหา (2550) ได้ทำการศึกษาผลการเปรียบเทียบคะแนนแนวนอนที่วิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบเมื่อใช้รูปแบบการเก็บรวบรวมข้อมูลที่แตกต่างกัน โดยมีจุดมุ่งหมายเพื่อสร้างตารางเทียบคะแนนและเปรียบเทียบค่าร้อยละของคะแนนจริงที่ได้จากการเทียบคะแนนตามแนวนอน โดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ เมื่อใช้รูปแบบการเก็บรวบรวมข้อมูลที่แตกต่างกัน ในแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน กลุ่มสาระการเรียนรู้คณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เรื่องอัตราส่วนและร้อยละ กลุ่มตัวอย่างคือนักเรียนชั้นมัธยมศึกษาปีที่ 2 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาสิงห์บุรี ปีการศึกษา 2549 จำนวน 2,080 คน ได้มาจากการสุ่มแบบแบ่งกลุ่ม เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เรื่องอัตราส่วนและร้อยละ จำนวน 50 ข้อ มีค่าอำนาจจำแนก 0.464 ถึง 3.855 ค่าความยาก -2.299 ถึง 3.137 และค่าสัมประสิทธิ์การเดา 0.028 ถึง 0.445

ผลการวิจัยพบว่า

1. ผลการสร้างตารางเทียบคะแนนแบบทดสอบจากการใช้วิธีการเทียบคะแนนตามแนวนอนโดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ เมื่อใช้รูปแบบการเก็บรวบรวมข้อมูลที่แตกต่างกัน ได้ผล ดังนี้ รูปแบบการเก็บรวบรวมข้อมูลจากการสุ่มกลุ่มนักเรียนให้สอบโดยใช้แบบทดสอบทั้งสองฉบับที่เรียงอันดับการสอบต่างกัน คะแนนจริงของแบบทดสอบจากกลุ่มตัวอย่างที่ 1 ที่ระดับความสามารถ -3.0 ถึง 3.0 มีค่าคะแนนจริง 9.079 ถึง 29.834 คะแนนเฉลี่ยของคะแนนจริง 17.968 ค่าเบี่ยงเบนมาตรฐาน 7.171 คะแนนจริงของแบบทดสอบจากกลุ่มตัวอย่างที่ 2 ที่ระดับความสามารถ -3.0 ถึง 3.0 มีค่าคะแนนจริง 9.811 ถึง 30.869 คะแนนเฉลี่ยของคะแนนจริง 19.740 ค่าเบี่ยงเบนมาตรฐาน 7.275 คะแนนจริงของรูปแบบการเก็บรวบรวมข้อมูลจากการสุ่มนักเรียนให้สอบ โดยใช้แบบทดสอบทั้งสองฉบับที่เรียงอันดับการสอบต่างกัน (กลุ่มตัวอย่างที่ 1 และกลุ่มตัวอย่างที่ 2 รวมกัน) มีคะแนนเฉลี่ย 18.854 ค่าเบี่ยงเบนมาตรฐาน 7.135

ส่วนรูปแบบการเก็บรวบรวมข้อมูลจากการสุ่มนักเรียนให้สอบ โดยใช้แบบทดสอบกลุ่มละฉบับ และเพิ่มแบบทดสอบร่วมที่ต้องสอบทั้งสองกลุ่ม คะแนนจริงของแบบทดสอบจากกลุ่มตัวอย่างที่ 3 ที่ระดับความสามารถ -3.0 ถึง 3.0 มีค่าคะแนนจริง 7.726 ถึง 20.295 คะแนนเฉลี่ยของคะแนนจริง 12.509 ค่าเบี่ยงเบนมาตรฐาน 4.121 คะแนนจริงของแบบทดสอบจากกลุ่มตัวอย่างที่ 4 ที่ระดับความสามารถ -3.0 ถึง 3.0 มีค่าคะแนนจริง 9.811 ถึง

20.996 คะแนนเฉลี่ยของคะแนนจริง 13.351 ค่าเบี่ยงเบนมาตรฐาน 3.930 คะแนนจริงของรูปแบบการเก็บรวบรวมข้อมูลจากการสุ่มกลุ่มนักเรียนให้สอบ โดยใช้แบบทดสอบกลุ่มละฉบับ และเพิ่มแบบทดสอบร่วมที่ต้องสอบทั้งสองกลุ่ม (กลุ่มตัวอย่างที่ 3 และกลุ่มตัวอย่างที่ 4 รวมกัน) มีคะแนนเฉลี่ย 12.930 ค่าเบี่ยงเบนมาตรฐาน 3.969

2. ผลการทดสอบความแตกต่างของค่าร้อยละของคะแนนจริงที่ได้จากการเทียบคะแนนตามแนวนอนโดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ เมื่อใช้รูปแบบการเก็บรวบรวมข้อมูลที่แตกต่างกัน พบว่า มีค่าร้อยละของคะแนนจริงที่ได้จากการเทียบคะแนนแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ นั่นคือ รูปแบบการเก็บรวบรวมข้อมูลจากการสุ่มนักเรียนให้สอบโดยใช้แบบทดสอบทั้งสองฉบับที่เรียงอันดับการสอบต่างกับรูปแบบการเก็บรวบรวมข้อมูลจากการสุ่มนักเรียนให้สอบ โดยใช้แบบทดสอบกลุ่มละฉบับ และเพิ่มแบบทดสอบร่วมที่ต้องสอบทั้งสองกลุ่ม มีค่าเฉลี่ยของค่าร้อยละของคะแนนจริงไม่แตกต่างกัน

จากงานวิจัยต่างประเทศที่กล่าวมานี้ แสดงให้เห็นว่า ได้มีการศึกษาเกี่ยวกับวิธีการปรับเทียบคะแนนระหว่างแบบทดสอบไว้หลายแนวด้วยกัน ทั้งการเปรียบเทียบวิธีการปรับเทียบคะแนน การหารูปแบบที่เหมาะสม ขนาดของกลุ่มตัวอย่าง ความยาวของแบบทดสอบและความยาวของแบบทดสอบร่วม ซึ่งงานวิจัยเกี่ยวกับขนาดของกลุ่มตัวอย่างให้ผลไปในทำนองเดียวกันคือ กลุ่มตัวอย่างขนาดใหญ่ ตั้งแต่ 500 คน หรือ 1000 คนขึ้นไป จะให้ผลการปรับเทียบที่ดีกว่า และในรูปแบบการปรับเทียบคะแนนแบบ IRT การปรับเทียบตามโมเดล 3 พารามิเตอร์ ต้องใช้กลุ่มตัวอย่างที่มากกว่าโมเดล 1 และ 2 พารามิเตอร์ ในบางครั้งการปรับเทียบโดย โมเดล 1 หรือ 2 หรือ 3 พารามิเตอร์ นั้นต้องระมัดระวังที่จะไม่ให้ละเมิดข้อยกเว้น เช่นในกรณีที่ค่าเฉลี่ยของอำนาจจำแนกไม่เท่ากัน เป็นต้น เพราะจะทำให้ผลการปรับเทียบไม่แก่ง สำหรับเรื่องความยาวของแบบทดสอบร่วม พบว่าข้อสอบร่วมที่มากกว่าให้ผลการปรับเทียบที่เที่ยงตรงกว่า ส่วนในเรื่องวิธีการปรับเทียบคะแนนที่เหมาะสมที่สุดในการปรับเทียบกับข้อมูลที่เก็บมาในรูปแบบใดนั้น ยังหาแนวทางของข้อสรุปที่ชัดเจนได้ยาก เพราะผลจากการวิจัยยังมีการขัดแย้งกันอยู่เพราะในงานวิจัยบางเรื่อง สรุปได้ว่าทุกวิธีที่ปรับเทียบคะแนนที่มีการแจกแจงของคะแนนดิบแตกต่างกัน ได้ผลไม่แตกต่างกัน เป็นต้น

สำหรับผลการศึกษาวิจัยในประเทศ พบว่า วิธีการปรับเทียบคะแนนจากแบบสอบคัดเลือกที่ให้ความพอเพียงมากที่สุด คือ วิธีอีควิเปอร์เซ็นไทล์ รองลงมาคือวิธีตามรูปแบบ IRT และวิธีเชิงเส้นตรงตามลำดับ ส่วนวิธีการปรับเทียบคะแนนผลสัมฤทธิ์เชิงเส้นตรง ให้ความพอเพียงดีกว่าวิธีตามรูปแบบ IRT นอกจากนี้ยังพบว่า มีการใช้รูปแบบการปรับเทียบคะแนนตามแนวตั้งตามทฤษฎีการตอบสนองข้อสอบแบบ 3 พารามิเตอร์ ไปประยุกต์เพื่อศึกษาความก้าวหน้าของผู้เรียนในกลุ่ม

ต่าง ๆ อีกด้วย ส่วนผลการศึกษารูปแบบการเก็บรวบรวมข้อมูลจากการสุ่มนักเรียนให้สอบโดยใช้แบบสอบถามกลุ่มละฉบับและเพิ่มแบบสอบถามที่ต้องสอบทั้งสองกลุ่ม พบว่า ผู้สอบมีค่าเฉลี่ยคะแนนจริงไม่แตกต่างกัน

แต่อย่างไรก็ตาม สำหรับข้อสรุปที่เกี่ยวกับวิธีการเปรียบเทียบคะแนนว่าวิธีการใด เหมาะสมกับสถานการณ์ใดมากที่สุดนั้น ยังหาข้อสรุปที่เป็นฉันทามติยากพอสมควร เพราะ ผลจากการศึกษายังมีความขัดแย้งกันอยู่ ซึ่งขึ้นอยู่กับสถานการณ์ที่ทำการศึกษา

### บทที่ 3 วิธีการดำเนินการ

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อเพื่อเปรียบเทียบวิธีการปรับเทียบคะแนน O-NET ระหว่างปี 2550 และปี 2551 โดยใช้ วิธีการเทียบเป็นมาตราเดียวกัน และ โดยการปรับเทียบตามรูปแบบการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) 4 รายวิชา คือ 1) วิชาภาษาไทย 2) สังคมศึกษา ศาสนาและวัฒนธรรม 3) ภาษาอังกฤษ และ 4) วิทยาศาสตร์ เพื่อให้เป็นไปตามจุดมุ่งหมายที่กำหนดไว้ ผู้วิจัยได้เสนอขั้นตอนการดำเนินการวิจัย โดยนำเสนอ ประชากร การเลือกกลุ่มตัวอย่าง วิธีการเทียบมาตรา และการวิเคราะห์ข้อมูล ตามลำดับ โดยมีรายละเอียด ดังนี้

#### ประชากร

ในการวิจัยครั้งนี้ ทำการศึกษาจากคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปี 2550 และ ปี 2551 จำนวน 4 รายวิชา ซึ่งดำเนินการสร้างและบริหารจัดการสอบ โดย สทศ.

#### กลุ่มตัวอย่าง

กลุ่มตัวอย่าง เป็นคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปี 2550 และ ปี 2551 ใน 4 รายวิชา ทำการสุ่มตัวอย่างโดยใช้โปรแกรม SPSS จากคะแนนนักเรียนที่สอบวิชาที่ 1 ในปี 2550 และ ปี 2551 339586 คน และ 338224 คน คะแนนนักเรียนที่สอบวิชาที่ 2 ในปี 2550 และ ปี 2551 339268 คน และ 326691 คน คะแนนนักเรียนที่สอบวิชาที่ 3 ในปี 2550 และ ปี 2551 326691 คน และ 339268 คน และคะแนนนักเรียนที่สอบวิชาที่ 5 ในปี 2550 และ ปี 2551 326691 คน และ 337996 คน ได้รายวิชาละ 1998 – 1999 คน เพราะโปรแกรมในการวิเคราะห์ตามรูปแบบ IRT ของคณิต ไซมูกซ์ สามารถวิเคราะห์ได้ประมาณ 2000 คน โดยถือว่าเป็นขนาดของกลุ่มตัวอย่างที่ให้ผลในการปรับเทียบคะแนนได้อย่างถูกต้อง ดังผลจากการศึกษาของ Ayer (1992:1841) ที่พบว่า การปรับเทียบคะแนนตามวิธี IRT 3 พารามิเตอร์ ขนาดของกลุ่มตัวอย่างเป็นองค์ประกอบที่สำคัญ คือ กลุ่มตัวอย่าง 200 คน มีแนวโน้มที่จะให้ผลการปรับเทียบคะแนนที่ไม่ถูกต้อง ในขณะที่กลุ่มตัวอย่างขนาดใหญ่ คือ 500 และ 1,000 จะให้ผลการปรับเทียบคะแนนที่ถูกต้องมากกว่า ซึ่งสอดคล้องกับผลการศึกษาของ พิชัย ละแมนชัย (2538 :131-134) ที่ได้ทำการศึกษาเพื่อหาขนาดของกลุ่มตัวอย่างขั้นต่ำ ที่ทำให้การปรับเทียบคะแนนในแนวระดับ มีประสิทธิภาพ โดยศึกษา ตามแนวทฤษฎีการตอบสนองราย

ข้อ โมเดล 1 และ 3 พารามิเตอร์ ในแบบแผนกลุ่มสมมูล และแบบแผนข้อสอบร่วมที่มีความยาวของแบบทดสอบต่าง ๆ กัน คือ 30 ข้อ 50 ข้อ 70 ข้อ 90 ข้อ 110 ข้อ 130 ข้อ และ 150 ข้อ ข้อมูลที่ใช้ในการวิจัยครั้งนี้ เป็นการจำลองจากโปรแกรม IRTDATA เกณฑ์ที่ใช้ในการตัดสินคือการหาค่าดัชนีความแตกต่างที่อยู่ในแนวระดับน่าพอใจอย่างมากผลการวิจัยและพบว่า ขนาดของกลุ่มตัวอย่างขั้นต่ำ ตามโมเดล 3 พารามิเตอร์ในแบบแผนกลุ่มสมมูลที่มีความยาวของแบบทดสอบเป็น 70, 90, 110, 130, และ 150 ข้อ คือ 2150, 1750, 1550, 1400 และ 1150 ตามลำดับ ซึ่งจากผลการวิจัยที่ผ่านมา พอจะสรุปได้ว่าขนาดของกลุ่มตัวอย่างตั้งแต่ 1000 คนขึ้นไป ถือเป็นขนาดกลุ่มตัวอย่างที่เหมาะสมที่จะนำมาเปรียบเทียบคะแนนได้อย่างถูกต้อง

### วิธีการปรับเทียบคะแนน

ในการปรับเทียบคะแนนมีการเสนอวิธีการปรับเทียบไว้ 2 วิธี ซึ่งแต่ละวิธี มีขั้นตอนการดำเนินการ และวิธีการวิเคราะห์ข้อมูล ดังนี้

#### 1 . วิธีการปรับเทียบให้เป็นมาตราเดียวกัน

ในการปรับเทียบคะแนนให้เป็นมาตราเดียวกัน ของข้อสอบ O-NET ได้พิจารณา รูปแบบการเก็บรวบรวมข้อมูล ซึ่งพบว่า การจัดสอบข้อสอบ O-NET เป็นรูปแบบ Random group-one test administered to each group ดังที่ แองกอฟฟ์ ( Angoff.1971 : 569 – 586) ได้เสนอ โมเดลการปรับเทียบคะแนนในการออกแบบเพื่อรวบรวมข้อมูล และการจัดกระทำในทางสถิติเพื่อ ระบุการแปลงคะแนนไว้ ว่าเป็นรูปแบบที่ผู้สอบมาจากการสุ่มจากประชากรกลุ่มเดียวกัน แล้ว แบ่งเป็น 2 กลุ่มคือ กลุ่มนักเรียนชั้นม.6 ปีการศึกษา 2550 สอบแบบทดสอบ O-NET ปีการศึกษา 2550 และ กลุ่มนักเรียนชั้นม.6 ปีการศึกษา 2551 สอบแบบทดสอบ O-NET ปีการศึกษา 2551 ซึ่งนักเรียนเป็นคนละกลุ่มกัน และทำการทดสอบแบบทดสอบ O-NET คนละฉบับกัน ซึ่งในการปรับเทียบคะแนนให้เป็นมาตราเดียวกันหรือการปรับเทียบคะแนนเชิงเส้นตรง นี้ ใช้วิธีการปรับเทียบคะแนนเชิงเส้นตรงสำหรับแบบทดสอบ ดังนี้

1.1 คำนวณค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของคะแนนแบบทดสอบ O-NET ปีการศึกษา 2550 ที่สอบโดยกลุ่มนักเรียนชั้นม.6 ปีการศึกษา 2550 และคำนวณค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของคะแนนแบบทดสอบ O-NET ปีการศึกษา 2551 ที่สอบโดยกลุ่มนักเรียนชั้นม.6 ปีการศึกษา 2551 โดยดำเนินการที่ละรายวิชา แล้วคำนวณคะแนนมาตรฐานของทั้งสองกลุ่ม ดังนี้

$$\text{คะแนนมาตรฐาน O-NET 50} = \frac{(X_{O-NET50} - \bar{X}_{O-NET50})}{SD_{O-NET50}} \text{-----}(1)$$

$$\text{คะแนนมาตรฐาน}_{O\text{-NET}51} = \frac{(X_{O\text{-NET}51} - \bar{X}_{O\text{-NET}51})}{SD_{O\text{-NET}51}} \text{-----}(2)$$

โดยที่  $X_{O\text{-NET}}$  หมายถึง คะแนน O-NET รายบุคคล  
 $\bar{X}_{O\text{-NET}}$  หมายถึง คะแนน O-NET รายบุคคลเฉลี่ย

1.2 ในการจัดสอบข้อสอบ O-NET เนื่องจาก สทศ.เป็นผู้พัฒนาข้อสอบและดำเนินการสอบ O-NET เองทั้ง 2 ปี จึงคาดว่าข้อสอบทั้ง 2 ปีต้องเป็นข้อสอบคู่ขนานที่มีความคล้ายกันทั้งในเชิงโครงสร้าง เนื้อหา ความยาก และอำนาจจำแนก ดังนั้นคะแนนมาตรฐานทั้งสองปีควรมีค่าเท่ากัน นั่นคือจัดให้สมการที่(1) = (2) ดังนี้

$$\frac{(X_{O\text{-NET}50} - \bar{X}_{O\text{-NET}50})}{SD_{O\text{-NET}50}} = \frac{(X_{O\text{-NET}51} - \bar{X}_{O\text{-NET}51})}{SD_{O\text{-NET}51}} \text{-----}(3)$$

$$X_{O\text{-NET}51} = X_{O\text{-NET}50} \left( \frac{SD_{O\text{-NET}51}}{SD_{O\text{-NET}50}} \right) - \bar{X}_{O\text{-NET}50} \left( \frac{SD_{O\text{-NET}51}}{SD_{O\text{-NET}50}} \right) + \bar{X}_{O\text{-NET}51} \text{--}(4)$$

$$X_{O\text{-NET}50} = X_{O\text{-NET}51} \left( \frac{SD_{O\text{-NET}50}}{SD_{O\text{-NET}51}} \right) - \bar{X}_{O\text{-NET}51} \left( \frac{SD_{O\text{-NET}50}}{SD_{O\text{-NET}51}} \right) + \bar{X}_{O\text{-NET}50} \text{--}(5)$$

1.3 กรณีที่ทราบคะแนนนักเรียนในปีใดปีหนึ่ง เช่น ทราบคะแนน O-NET ปี 2550 แล้ว สมมุติว่านักเรียนที่สอบ O-NET ในปี 2550 ได้ยื่นคะแนนเพื่อให้พิจารณาเปรียบเทียบกับคะแนนสอบของในปี 2551 สามารถปรับเทียบคะแนนส่วนใหญ่ของนักเรียนที่สอบ O-NET ปี 2550 ให้สามารถนำมาเทียบกับคะแนน O-NET ปี 2551 จากสูตร (4) แล้วค่อยนำคะแนนปี 2550 ที่แปลงเป็นคะแนนปี 2551 ไปร่วมเปรียบเทียบกับคะแนนของนักเรียนส่วนใหญ่

## 2. วิธีการปรับเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ(Equating Through Item Response Theory)โดยใช้โมเดล 3 พารามิเตอร์ (รูปแบบ IRT)

มีขั้นตอนการปรับเทียบดังนี้

2.1 นำคะแนนสอบรายข้อของนักเรียนที่ได้จากการทดสอบด้วยแบบสอบทั้ง 2 ฉบับมาวิเคราะห์หาค่าความยาก ค่าอำนาจจำแนก และค่าสัมประสิทธิ์การเดาของข้อสอบเป็นรายข้อ โดยใช้โปรแกรมการวิเคราะห์ IRT ของคณิต ไข่มุกด์ ทำการวิเคราะห์ แล้วจะได้ค่าความยาก ค่าอำนาจจำแนก ค่าสัมประสิทธิ์การเดา ของข้อสอบทุกข้อ และยังได้ ค่าความสามารถ (ability) ของนักเรียนทุกคน

2.2 หาค่าความน่าจะเป็นในการตอบถูกของข้อสอบแต่ละข้อที่ระดับความสามารถต่าง ๆ

2.3 ปรับเทียบคะแนนจริงที่วิเคราะห์ได้ในแต่ละปีเทียบกับระดับความสามารถโดยสร้างตารางเทียบคะแนนตามแนวนอน

กรณีที่คะแนนเต็มเท่ากันทั้ง 2 ปี สามารถใช้คะแนนจริง ณ ระดับความสามารถเดียวกันเปรียบเทียบกันได้เลย แต่ถ้าคะแนนเต็มไม่เท่ากันให้ปรับคะแนนจริงของทั้งปี 2550 และ 2551 เป็นค่าร้อยละของคะแนนจริงแทนคะแนนจริง

### การวิเคราะห์ข้อมูล

1. นำคะแนนก่อนปรับเทียบ และคะแนนที่ได้จากการปรับเทียบโดยวิธีการเทียบให้เป็นมาตราเดียวกัน มาวิเคราะห์หาค่าสถิติต่าง ๆ ตามลำดับ ดังนี้

1.1 หาค่าสถิติพื้นฐานของคะแนนที่ปรับเทียบแล้ว ได้แก่ การวัดแนวโน้มเข้าสู่ศูนย์กลาง การวัดการกระจาย และลักษณะการแจกแจงความถี่ของข้อมูล

1.2 เปรียบเทียบคะแนน O-NET เดิม ทั้ง 2 ปี จำแนกตามรายวิชา โดยใช้สถิติ t-test independent

1.3 เปรียบเทียบคะแนน O-NET เดิม และคะแนนที่ปรับเทียบจากวิธีการเทียบให้เป็นมาตราเดียวกัน จำแนกตามรายวิชา โดยใช้สถิติ t-test dependent

2. นำค่าความสามารถของนักเรียนที่ได้จากการวิเคราะห์คะแนนรายข้อทั้ง 2 ปี ทุกรายวิชา มาทำการนำเสนอคะแนนแบบจับคู่ ณ ระดับค่าความสามารถเดียวกัน

2.1 วิเคราะห์เปรียบเทียบความแตกต่างของคะแนน โดยใช้สถิติ t-test dependent เพราะถือว่า ณ ระดับความสามารถเดียวกัน เสมือนกับเป็นบุคคลเดียวกันแต่ทำการทดสอบข้อสอบ O-NET ทั้ง 2 ปี

3. วิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิม กับคะแนนที่ปรับเทียบแล้วทั้ง 2 วิธี และวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ปรับเทียบแล้วทั้ง 2 วิธี โดยใช้การหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

4. ทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง คะแนน O-NET เดิม กับคะแนน O-NET ที่แปลงแล้วด้วยวิธีการเทียบให้เป็นมาตราเดียวกัน และค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนคะแนน O-NET เดิม กับคะแนน O-NET ที่แปลงแล้วตามรูปแบบ IRT โดยใช้ t-Hotelling

### สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. หาค่าเฉลี่ยของคะแนน O-NET ทั้งคะแนน O-NET เดิม และคะแนนที่ปรับเทียบแล้ว โดยใช้สูตร ดังนี้

$$\bar{X} = \frac{\sum X_{O-NET}}{n}$$

เมื่อ  $\bar{X}$  แทน ค่าเฉลี่ย  
 $\sum X_{O-NET}$  แทน ผลรวมของคะแนน O-NET ทั้งหมด  
 $n$  แทน จำนวนนักเรียนในกลุ่มตัวอย่าง

2. หาค่าส่วนเบี่ยงเบนมาตรฐานของคะแนน O-NET ทั้งคะแนน O-NET เดิม และคะแนนที่ปรับเทียบแล้ว โดยใช้สูตร ดังนี้

$$S.D. = \sqrt{\frac{n \sum X_{O-NET}^2 - (\sum X_{O-NET})^2}{n(n-1)}}$$

เมื่อ S.D. แทน ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนน  
 $X_{O-NET}$  แทน คะแนน O-NET ของแต่ละคน  
 $\sum X_{O-NET}^2$  แทน ผลรวมของคะแนน O-NET ที่แต่ละตัวยกกำลังสองแล้ว  
 $n$  แทน จำนวนนักเรียนในกลุ่มตัวอย่าง

3. เปรียบเทียบความแตกต่างของค่าเฉลี่ยขงคะแนน O-NET ทั้ง 2 ปี เช่น คะแนน O-NET เดิมทั้ง 2 ปี และ O-NET เดิมกับคะแนน O-NET ที่ปรับเทียบแล้ว

3.1 t-test dependent หรือ t-test pair sample โดยใช้สูตร ดังนี้

$$t = \frac{\sum D}{\sqrt{\frac{n \sum D^2 - (\sum D)^2}{n-1}}}, df = n-1$$

เมื่อ  $t$  แทน ค่าสถิติที่  
 $D$  แทน ผลต่างของคะแนน O-NET แต่ละคู่  
 $\sum D^2$  แทน ผลรวมของผลต่างของคะแนน O-NET ที่แต่ละตัวยกกำลังสองแล้ว  
 $n$  แทน จำนวนคู่ของข้อมูล หรือจำนวนนักเรียนในกลุ่มตัวอย่าง

df แทน องศาหรือชั้นความเป็นอิสระ

### 3.1 t-test independent หรือ t-test pair sample

#### 3.1.1 กรณีที่ความแปรปรวนเท่ากัน โดยใช้สูตร ดังนี้

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_p^2}{n_1} + \frac{s_p^2}{n_2}}}, \quad df = n_1 + n_2 - 2$$

เมื่อ  $s_p^2$  แทน ความแปรปรวนรวมซึ่งหาได้จาก

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

การคำนวณความแปรปรวน ได้จาก S.D.<sup>2</sup>

#### 3.1.2 กรณีที่ความแปรปรวนไม่เท่ากัน โดยใช้สูตร ดังนี้

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

เมื่อหา df ได้จาก

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}$$

เมื่อ  $s^2$  แทน ความแปรปรวนรวมซึ่งคำนวณได้จาก S.D.<sup>2</sup>

4. หาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิม กับคะแนนที่ปรับเทียบแล้ว โดยใช้สูตร ดังนี้

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

เมื่อ  $r_{xy}$  แทน ค่าสัมประสิทธิ์สหสัมพันธ์  
 X แทน คะแนน O-NET เดิม  
 Y แทน คะแนนที่ปรับเทียบแล้ว

ทดสอบความมีนัยสำคัญของค่าสหสัมพันธ์ โดยใช้สูตร ดังนี้

$$t = r \sqrt{\frac{n-2}{1-r^2}}, \text{ df} = n-2$$

เมื่อ  $r$  แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิม  
 กับคะแนนที่ปรับเทียบแล้ว  
 $n$  แทน จำนวนนักเรียนของกลุ่มตัวอย่าง  
 $t$  แทน สถิติทดสอบที่

5. ทดสอบความแตกต่างระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนน O-NET เดิม กับ  
 คะแนนที่แปลงแล้วโดยการเทียบเป็นมาตราเดียวกัน กับค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนน O-  
 NET เดิม กับคะแนนที่แปลงแล้วตามรูปแบบ IRT โดยใช้สูตรของ  $t$  – Hotelling (Guildford, 1978  
 :190;citing Hotelling:1940 ) ดังนี้

$$t = (r_{12} - r_{13}) \sqrt{\frac{(n-3)(1+r_{23})}{2(1-r_{23}^2 - r_{12}^2 - r_{13}^2 + 2r_{23}r_{12}r_{13})}}$$

เมื่อ  $r_{12}$  แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิม กับ  
 คะแนนที่ปรับเทียบโดยการเทียบให้เป็นมาตราเดียวกัน  
 $r_{13}$  แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิม กับ  
 คะแนนที่ปรับเทียบโดยรูปแบบ IRT  
 $r_{23}$  แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET ที่ปรับเทียบ  
 โดยการเทียบให้เป็นมาตราเดียวกัน กับคะแนนที่ปรับเทียบโดย  
 รูปแบบ IRT  
 $t$  แทน ค่าสถิติ  $t$ -Hotelling  
 $n$  แทน จำนวนนักเรียนของกลุ่มตัวอย่าง

## บทที่ 4

### ผลการวิเคราะห์ข้อมูล

#### สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

ในการนำเสนอผลการวิเคราะห์ข้อมูล ได้อย่างสะดวกและให้เกิดความเข้าใจตรงกัน ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล ดังนี้

|                     |   |   |
|---------------------|---|---|
| $X_{O-NET}$         | แทน คะแนน   | O-NET รายบุคคล  |
| $\bar{X}_{O-NET}$   | แทน คะแนน   | O-NET รายบุคคลเฉลี่ย  |
| $\bar{X}$           | แทน ค่าเฉลี่ย                                     |   |
| Min                 | แทน ค่าต่ำสุด                                     |   |
| Max                 | แทน ค่าสูงสุด                                     |   |
| SK                  | แทน ค่าความเบ้                                    |   |
| Ku                  | แทน ค่าความโด่ง                                   |   |
| S.D.                | แทน ค่าเบี่ยงเบนมาตรฐาน                           |   |
| $R_{O(50)_T1(50)}$  | แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน         | O-NET ปี 2550 กับคะแนนแปลงเป็นคะแนน O-NET ปี 2551 โดยวิธีการเทียบด้วยมาตราเดียวกัน  |
| $R_{O(50)_T2(50)}$  | แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน         | O-NET ปี 2550 กับค่าความสามารถของผู้เรียนที่ได้จากการปรับเทียบคะแนนตามรูปแบบ IRT  |
| $R_{T1(50)_T2(50)}$ | แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนแปลงเป็น | คะแนน O-NET ปี 2551(จากคะแนน O-NET ปี 2550) โดยวิธีการเทียบเป็นมาตราเดียวกัน กับค่าความสามารถของผู้เรียนที่ได้มาจากการปรับเทียบคะแนน O-NET ปี 2550 โดยการปรับเทียบตามรูปแบบ IRT |
| $R_{O(51)_T1(51)}$  | แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน         | O-NET ปี 2551 กับคะแนนแปลงเป็นคะแนน O-NET ปี 2550 โดยวิธีการเทียบด้วยมาตราเดียวกัน  |
| $R_{O(51)_T2(51)}$  | แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน         | O-NET ปี 2551 กับค่าความสามารถของผู้เรียนโดยการปรับเทียบ  |

ตามรูปแบบ IRT

$R_{T1(51)_{T2(51)}}$  แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนแปลงเป็น  
คะแนน O-NET ปี 2550(จากคะแนน O-NET ปี 2551)  
โดยวิธีการเทียบเป็นมาตราเดียวกัน กับค่า  
ความสามารถของผู้เรียนที่ได้มาจากการปรับเทียบ  
คะแนน O-NET ปี 2551 โดยการปรับเทียบตามรูปแบบ  
IRT

### ผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูลในการศึกษาครั้งนี้ จะนำเสนอเป็นตอน ๆ ดังนี้

1. ลักษณะของคะแนน O-NET ในปี 2550 และ ปี 2551 ในแต่ละรายวิชา
2. ลักษณะของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เป็นคะแนนเดิมและคะแนนแปลงโดยวิธีการเทียบให้เป็นมาตราเดียวกัน
3. การเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ระหว่างคะแนนเดิมกับคะแนนแปลงโดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน
4. การเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในปีเดียวกันระหว่างคะแนนเดิมกับคะแนนที่ได้จากการปรับเทียบมาจากคนละปีโดยวิธีการเทียบเป็นมาตราเดียวกัน
5. ผลการเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ตามรูปแบบการปรับเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ (Equating Through Item Response Theory )โดยใช้โมเดล 3 พารามิเตอร์ (รูปแบบ IRT)
6. ผลการเปรียบเทียบคุณภาพของวิธีการแปลงคะแนน O-NET โดยการปรับเทียบให้เป็นมาตราเดียวกัน และปรับเทียบตามรูปแบบ IRT โดยการหาความสัมพันธ์ระหว่างคะแนนสอบ O-NET เดิมกับคะแนน O-NET ที่ได้จากการปรับเทียบให้เป็นมาตราเดียวกัน และที่ได้จากการปรับเทียบด้วยรูปแบบ IRT

### ตอนที่ 1 ลักษณะของคะแนน O-NET ในปี 2550 และ ปี 2551 ในแต่ละรายวิชา

คะแนน O-NET ในปี 2550 และ 2551 ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จำแนกตามรายวิชา แสดงดังตารางที่ 4.1 และ 4.2

ตารางที่ 4.1 แสดงการแจกแจงผลของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6  
ที่ทดสอบปี 2550 และ 2551 จำแนกตามรายวิชา และเป็นกลุ่มตัวอย่างในการนำ  
คะแนนมาเปรียบเทียบ

| ค่าสถิติ          | วิชาภาษาไทย |         | วิชาสังคมศึกษา ฯ |         | วิชาภาษาอังกฤษ |         | วิชาวิทยาศาสตร์ |         |
|-------------------|-------------|---------|------------------|---------|----------------|---------|-----------------|---------|
|                   | ปี 2550     | ปี 2551 | ปี 2550          | ปี 2551 | ปี 2550        | ปี 2551 | ปี 2550         | ปี 2551 |
| N                 | 1999        | 1999    | 1999             | 1998    | 1999           | 1999    | 1999            | 1999    |
| $\bar{X}$ *       | 50.64*      | 36.22   | 37.7*            | 33.75   | 31.02          | 30.35   | 33.85*          | 31.39*  |
| S.D.              | 11.19       | 6.09    | 7.61             | 5.77    | 10.98          | 9.50    | 10.13           | 4.63    |
| Ku                | -.62        | -.02    | .83              | .00     | 6.73           | 1.57    | 3.31            | -.05    |
| Sk                | .23         | -.42    | .78              | .20     | 2.28           | 1.33    | 1.73            | -.02    |
| Min*              | 13.75*      | 14      | 15*              | 18      | 9              | 13      | 11.25*          | 11.25*  |
| Max*              | 88.75*      | 54      | 78.75*           | 56      | 92             | 68      | 93.75*          | 50*     |
| จำนวน<br>ข้อสอบ** | 80          | 100     | 80               | 100     | 100            | 100     | 80              | 80      |

หมายเหตุ \* ค่าเฉลี่ยของคะแนน ค่าต่ำสุด และค่าสูงสุดของคะแนน ถูกปรับให้มีคะแนนเต็มเท่ากันคือ 100  
คะแนน เพื่อให้สะดวกต่อการนำเสนอเปรียบเทียบค่าสถิติ

\*\* กำหนดให้คะแนนข้อสอบเป็นข้อละ 1 คะแนน

จากตารางที่ 4.1 ผลการพิจารณาคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่  
ทดสอบในปี 2550 ใน 4 รายวิชา มีดังนี้

ค่าเฉลี่ยในรายวิชาที่นักเรียนได้ค่าเฉลี่ยคะแนนสูงสุด คือวิชาภาษาไทย โดยมีค่าเฉลี่ย  
50.64 รองลงมาคือวิชาสังคมศึกษา ศาสนา และวัฒนธรรม และวิชาวิทยาศาสตร์ โดยมีค่าเฉลี่ย  
37.7 และ 33.85 ตามลำดับ ส่วนวิชาภาษาอังกฤษ มีค่าเฉลี่ยน้อยที่สุด คือ 31.02

เมื่อพิจารณาส่วนเบี่ยงเบนมาตรฐานของทั้ง 4 รายวิชา พบว่ามีค่าอยู่ระหว่าง 7.61 –  
11.19 โดย วิชาสังคมศึกษา ศาสนา และวัฒนธรรม มีค่าส่วนเบี่ยงเบนมาตรฐานน้อยกว่ารายวิชา  
อื่น ๆ แสดงว่าคะแนน O-NET ในวิชาสังคมศึกษา ศาสนา และวัฒนธรรม มีการกระจายน้อยกว่า  
รายวิชาอื่น ๆ

เมื่อพิจารณาที่ค่าความโค้ง พบว่า ค่าความโค้งของคะแนน O-NET ใน 4 รายวิชา มีค่าอยู่  
ระหว่าง -.62 – 6.73 แสดงว่าคะแนนในแต่ละรายวิชา มีการกระจายแตกต่างกัน

ในส่วนของค่าความเบ้ พบว่า ค่าความเบ้ของคะแนน O-NET ของทั้ง 4 รายวิชา มีค่าอยู่  
ระหว่าง .23 – 2.28 ซึ่งมีค่าเป็นบวก นั่นคือคะแนนส่วนใหญ่มีค่าน้อย

ในการพิจารณาค่าต่ำสุดและสูงสุด พบว่า ค่าต่ำสุดสูงสุดของทั้ง 4 รายวิชา มีค่าอยู่ระหว่าง 9- 93.75

การพิจารณาลักษณะของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของนักเรียนที่ทดสอบในปี 2551 ใน 4 รายวิชา มีดังนี้

ค่าเฉลี่ยในรายวิชาที่นักเรียนได้ค่าเฉลี่ยคะแนนสูงสุด คือวิชาภาษาไทย โดยมีค่าเฉลี่ย 36.22 ส่วนวิชาภาษาอังกฤษ มีค่าเฉลี่ยน้อยที่สุด คือ 30.35

เมื่อพิจารณาส่วนเบี่ยงเบนมาตรฐานของทั้ง 4 รายวิชา พบว่ามีค่าอยู่ระหว่าง 4.63 – 9.50 โดย วิชาวิทยาศาสตร์ มีค่าส่วนเบี่ยงเบนมาตรฐานน้อยกว่ารายวิชาอื่น ๆ แสดงว่าคะแนน O-NET ในวิชาวิทยาศาสตร์ มีการกระจายน้อยกว่ารายวิชาอื่น ๆ

เมื่อพิจารณาที่ค่าความโด่ง พบว่า ค่าความโด่งของคะแนน O-NET ใน 4 รายวิชา มีค่าอยู่ระหว่าง -0.05 – 1.57 แสดงว่าคะแนนในแต่ละรายวิชา มีการกระจายแตกต่างกัน

ในส่วนของค่าความเบ้ พบว่า ค่าความเบ้ของคะแนน O-NET ของทั้ง 4 รายวิชา มีค่าอยู่ระหว่าง -0.42 – 1.33 ซึ่งคะแนนในวิชาภาษาไทย และ 5 มีค่าความเบ้เป็นลบ แสดงว่าคะแนนส่วนใหญ่มีค่ามาก สำหรับคะแนนในวิชาสังคมศึกษา ศาสนา และวัฒนธรรม และ 3 มีค่าความเบ้เป็นบวก นั่นคือคะแนนส่วนใหญ่มีค่าน้อย

ในการพิจารณาค่าต่ำสุดและสูงสุด พบว่า ค่าต่ำสุดสูงสุดของทั้ง 4 รายวิชา มีค่าอยู่ระหว่าง 13- 68

**ตอนที่ 2 ลักษณะของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เป็นคะแนนเดิมและคะแนนแปลงโดยวิธีการเทียบให้เป็นมาตราเดียวกัน**

ตารางที่ 4.2 แสดงการแจกแจงผลของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2550 ที่เป็นคะแนนเดิมและคะแนนแปลงโดยการปรับเทียบให้เป็นมาตราเดียวกัน จำแนกตามรายวิชา

| ค่าสถิติ  | วิชาภาษาไทย   |               | วิชาสังคมศึกษา ฯ |               | วิชาภาษาอังกฤษ |               | วิชาวิทยาศาสตร์ |               |
|-----------|---------------|---------------|------------------|---------------|----------------|---------------|-----------------|---------------|
|           | ก่อนปรับเทียบ | หลังปรับเทียบ | ก่อนปรับเทียบ    | หลังปรับเทียบ | ก่อนปรับเทียบ  | หลังปรับเทียบ | ก่อนปรับเทียบ   | หลังปรับเทียบ |
| N         | 1999          | 1999          | 1999             | 1999          | 1999           | 1999          | 1999            | 1999          |
| $\bar{X}$ | 40.54         | 36.22         | 30.16            | 33.95         | 31.02          | 30.35         | 27.08           | 25.11         |
| S.D.      | 11.19         | 6.09          | 7.61             | 5.77          | 10.98          | 10.98         | 10.13           | 4.63          |
| Ku        | -.62          | -.62          | .83              | .83           | 6.73           | 6.73          | 3.31            | 3.31          |
| Sk        | .23           | .23           | .78              | .87           | 2.28           | 2.28          | 1.73            | 1.73          |
| Min       | 11            | 20.14         | 12               | 20.18         | 9              | 11.13         | 9               | 16.85         |
| Max       | 71            | 52.80         | 63               | 58.85         | 92             | 83.10         | 75              | 47.01         |

จากตารางที่ 4.2 ผลการพิจารณาคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของนักเรียนที่ทดสอบในปี 2550 ที่เป็นคะแนนเดิมและคะแนนแปลง โดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน ใน 4 รายวิชา มีดังนี้

วิชาภาษาไทย พบว่ามีค่าเฉลี่ยคะแนนเดิมสูงกว่าค่าเฉลี่ยคะแนนแปลง และมีค่าการกระจายของคะแนนเดิมสูงกว่าค่าการกระจายคะแนนแปลงแล้ว ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดแคบลง โดยเดิมมีช่วงคะแนนระหว่าง 11- 71 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 20.14-52.80

วิชาสังคมศึกษา ศาสนา และวัฒนธรรม พบว่ามีค่าเฉลี่ยคะแนนเดิมมีค่าน้อยกว่าค่าเฉลี่ยของคะแนนแปลงแล้ว และมีค่าการกระจายของคะแนนเดิมสูงกว่าค่าการกระจายคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดแคบลง โดยเดิมมีช่วงคะแนนระหว่าง 12-63 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 20.18-58.85

วิชาภาษาอังกฤษ พบว่ามีค่าเฉลี่ยคะแนนเดิมมีค่ามากกว่าค่าเฉลี่ยของคะแนนแปลงแล้ว แต่เป็นค่าที่แตกต่างกันไม่มาก โดยค่าเฉลี่ยคะแนนเดิมมีค่า 31.02 และหลังปรับเทียบมีค่า

30.35 มีค่าการกระจายของคะแนนเดิมก่อนและหลังปรับเทียบเท่ากัน ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดแคบลง โดยเดิมมีช่วงคะแนนระหว่าง 9 - 92 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 11.13 – 83.10

วิชาวิทยาศาสตร์ พบว่ามีค่าเฉลี่ยคะแนนเดิมมีค่ามากกว่าค่าเฉลี่ยของคะแนนแปลง และมีค่าการกระจายของคะแนนเดิมสูงกว่าค่าการกระจายคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดแคบลง โดยเดิมมีช่วงคะแนนระหว่าง 9 - 75 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 16.85 – 47.01

ตารางที่ 4.3 แสดงการแจกแจงผลของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2551 ก่อนการปรับเทียบและคะแนนแปลงโดยการปรับเทียบให้เป็นมาตรฐานเดียวกัน จำแนกตามรายวิชา

| ค่าสถิติ  | วิชาภาษาไทย   |               | วิชาสังคมศึกษา ฯ |               | วิชาภาษาอังกฤษ |               | วิชาวิทยาศาสตร์ |               |
|-----------|---------------|---------------|------------------|---------------|----------------|---------------|-----------------|---------------|
|           | ก่อนปรับเทียบ | หลังปรับเทียบ | ก่อนปรับเทียบ    | หลังปรับเทียบ | ก่อนปรับเทียบ  | หลังปรับเทียบ | ก่อนปรับเทียบ   | หลังปรับเทียบ |
| N         | 1999          | 1999          | 1998             | 1998          | 1999           | 1999          | 1999            | 1999          |
| $\bar{X}$ | 36.22         | 40.54         | 33.75            | 30.15         | 30.35          | 31.02         | 25.11           | 27.08         |
| S.D.      | 6.09          | 11.19         | 5.77             | 7.68          | 9.50           | 10.98         | 4.63            | 10.13         |
| Ku        | -.02          | -.02          | .00              | .00           | 1.57           | 1.57          | -.05            | -.05          |
| Sk        | -.42          | -.42          | .20              | .20           | 1.33           | 1.67          | -.02            | -.02          |
| Min       | 14            | -.29          | 18               | 9.21          | 13             | 10.96         | 9               | -8.17         |
| Max       | 54            | 73.21         | 56               | 59.76         | 68             | 74.54         | 40              | 59.66         |

จากตารางที่ 4.3 ผลการพิจารณาคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของนักเรียนที่ทดสอบในปี 2551 ที่เป็นคะแนนเดิมและคะแนนแปลง โดยวิธีการปรับเทียบให้เป็นมาตรฐานเดียวกัน ใน 4 รายวิชา มีดังนี้

วิชาภาษาไทย พบว่ามีค่าเฉลี่ยคะแนนเดิมน้อยกว่าค่าเฉลี่ยคะแนนแปลง และมีค่าการกระจายของคะแนนเดิมน้อยกว่าค่าการกระจายคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดกว้างขึ้น โดยเดิมมีช่วงคะแนนระหว่าง 14-54 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง -.29 – 73.21

วิชาสังคมศึกษา ศาสนา และวัฒนธรรม พบว่ามีค่าเฉลี่ยคะแนนเดิมมีค่ามากกว่าค่าเฉลี่ยของคะแนนแปลง และมีค่าการกระจายของคะแนนเดิมต่ำกว่าค่าการกระจายคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดกว้างขึ้น โดยเดิมมีช่วงคะแนนระหว่าง 18-56 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 9.21 – 59.76

วิชาภาษาอังกฤษ พบว่ามีค่าเฉลี่ยคะแนนเดิมมีค่าน้อยกว่าค่าเฉลี่ยของคะแนนแปลงแล้ว แต่เป็นค่าที่แตกต่างกันไม่มาก โดยค่าเฉลี่ยคะแนนเดิมมีค่า 30.35 และหลังปรับเทียบมีค่า 31.02 มีค่าการกระจายของคะแนนเดิมต่ำกว่าค่าการกระจายของคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดกว้างขึ้น โดยเดิมมีช่วงคะแนนระหว่าง 13 - 68 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 10.96 – 74.54

วิชาวิทยาศาสตร์ พบว่ามีค่าเฉลี่ยคะแนนเดิมมีค่าน้อยกว่าค่าเฉลี่ยของคะแนนแปลงแล้ว และมีค่าการกระจายของคะแนนเดิมต่ำกว่าค่าการกระจายคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดแคบลง โดยเดิมมีช่วงคะแนนระหว่าง 9 - 40 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง -8.17 – 59.66

จากตาราง 4.2 และ 4.3 พบว่า ค่าสถิติ ของคะแนนแปลงจากคะแนน O-NET ปี 2550 ที่มีค่าเท่ากับค่าสถิติของคะแนน O-NET ปี 2551 ได้แก่ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง สำหรับค่าที่มีการเปลี่ยนแปลงคือ คะแนนต่ำสุด และคะแนนสูงสุด และในทำนองเดียวกัน ค่าสถิติ ของคะแนนแปลงจากคะแนน O-NET ปี 2551 ที่มีค่าเท่ากับค่าสถิติของคะแนน O-NET ปี 2550 ได้แก่ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง สำหรับค่าที่มีการเปลี่ยนแปลงคือ คะแนนต่ำสุด และคะแนนสูงสุด

**ตอนที่ 3 การเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ระหว่างคะแนนเดิมกับคะแนนแปลงโดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน**

ตารางที่ 4.4 แสดงผลการเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2550 เดิมและคะแนนแปลงโดยการปรับเทียบให้เป็นมาตราเดียวกัน จำแนกตามรายวิชา

| วิชา        | N    | ผลต่างของคะแนนเดิม(ปี 2550 เดิม) และคะแนนแปลงแล้ว (เสมือนคะแนนปี 2551 เดิม) |      | t-test   | Sig. |
|-------------|------|---|------|----------|------|
|             |      | $\bar{X}$   | S.D. |          |      |
| ภาษาไทย     | 1999 | 4.32  | 5.10 | 37.84**  | .000 |
| สังคมศึกษา  | 1999 | -3.79   | 1.84 | -92.08** | .000 |
| ภาษาอังกฤษ  | 1999 | .67   | 1.48 | 20.21**  | .000 |
| วิทยาศาสตร์ | 1999 | 1.97  | 5.50 | 16.01**  | .000 |

\*\* p < .01

จากตาราง 4.4 พบว่าค่าเฉลี่ยคะแนน O-NET ปี 2550 เดิม และค่าเฉลี่ยคะแนน O-NET ปี 2550 ที่เป็นคะแนนแปลง (ซึ่งมีค่าเท่ากับค่าเฉลี่ยของคะแนน O-NET ปี 2551 เดิม) มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้ง 4 รายวิชา โดยในรายวิชาที่ 1 3 และ 5 มีค่าเฉลี่ยคะแนนเดิมสูงกว่าค่าเฉลี่ยของคะแนนแปลงที่ได้จากการปรับเทียบมาจากคะแนน O-NET ปี 2551 สำหรับในวิชาสังคมศึกษา ศาสนา และวัฒนธรรม นั้นมีค่าเฉลี่ยคะแนนเดิมต่ำกว่าค่าเฉลี่ยของคะแนนแปลงที่ได้จากการปรับเทียบมาจากคะแนน O-NET ปี 2551 นั่นคือค่าเฉลี่ยของคะแนน O-NET เดิมทั้ง 2 ปีมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ตอนที่ 4 การเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในปีเดียวกันที่เป็นคะแนนเดิมกับคะแนนแปลงที่ได้จากการปรับเทียบคะแนนมาจากคนละปีโดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน

ตารางที่ 4.5 แสดงผลการเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2550 ที่เป็นคะแนนเดิมและคะแนนแปลงมาจากคะแนน O-NET ปี 2551 โดยการปรับเทียบให้เป็นมาตราเดียวกัน จำแนกตามรายวิชา

| วิชา        | คะแนนปี 2550 |           |       | คะแนนแปลงปี 2550 ที่ปรับเทียบมาจากปี 2551 |           |       | t-test | Sig. |
|-------------|--------------|-----------|-------|---|-----------|-------|--------|------|
|             | N            | $\bar{X}$ | S.D.  | N   | $\bar{X}$ | S.D.  |        |      |
| ภาษาไทย     | 1999         | 40.54     | 11.19 | 1999                                      | 40.54     | 11.19 | -.009  | .993 |
| สังคมศึกษาฯ | 1999         | 30.15     | 7.61  | 1998                                      | 30.15     | 7.68  | .036   | .971 |
| ภาษาอังกฤษ  | 1999         | 31.02     | 10.98 | 1999                                      | 31.02     | 10.98 | -.015  | .988 |
| วิทยาศาสตร์ | 1999         | 27.08     | 10.13 | 1999                                      | 27.08     | 10.13 | -.005  | .996 |

ตารางที่ 4.6 แสดงผลการเปรียบเทียบคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบปี 2551 ที่เป็นคะแนนเดิมและคะแนนแปลงมาจากคะแนน O-NET ปี 2550 โดยการปรับเทียบให้เป็นมาตราเดียวกัน จำแนกตามรายวิชา

| วิชา        | คะแนนปี 2551 |           |      | คะแนนแปลงปี 2551 ที่ปรับเทียบมาจากปี 2550 |           |      | t-test | Sig. |
|-------------|--------------|-----------|------|---|-----------|------|--------|------|
|             | N            | $\bar{X}$ | S.D. | N   | $\bar{X}$ | S.D. |        |      |
| ภาษาไทย     | 1999         | 36.22     | 6.09 | 1999                                      | 36.02     | 6.09 | .009   | .993 |
| สังคมศึกษาฯ | 1998         | 33.75     | 5.77 | 1999                                      | 33.95     | 5.77 | -1.121 | .262 |
| ภาษาอังกฤษ  | 1999         | 30.35     | 9.50 | 1999                                      | 30.50     | 9.50 | .015   | .988 |
| วิทยาศาสตร์ | 1999         | 25.11     | 4.63 | 1999                                      | 25.11     | 4.63 | .005   | .996 |

จากตาราง 4.5 และ 4.6 พบว่าค่าเฉลี่ยคะแนน O-NET ปี 2550 เดิม และค่าเฉลี่ยคะแนน O-NET ปี 2550 ที่ได้จากการปรับเทียบมาจากคะแนน O-NET ปี 2551 ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทั้ง 4 รายวิชา และในทำนองเดียวกัน ค่าเฉลี่ยคะแนน O-NET ปี 2551 เดิม และค่าเฉลี่ยคะแนน O-NET ปี 2551 ที่ได้จากการปรับเทียบ มาจากคะแนน O-NET ปี 2550 ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทั้ง 4 รายวิชา เช่นเดียวกัน

ตอนที่ 5 ผลการเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ตามรูปแบบการปรับเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ (Equating Through Item Response Theory หรือ IRT) โดยใช้โมเดล 3 พารามิเตอร์ (รูปแบบ IRT)

ผลการเปรียบเทียบคะแนนที่ได้จากการปรับเทียบตามรูปแบบ IRT นำเสนอดังนี้

ตาราง 4.7 การเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ที่เป็นคะแนนจริงและร้อยละของคะแนนจริง ตามระดับความสามารถ สำหรับวิชาภาษาไทย

| ระดับ<br>ความสามารถ | ร้อยละของคะแนนจริงของ<br>นักเรียนที่สอบ<br>O-NET ปี 2550<br>(คะแนนเต็ม 80 คะแนน) | คะแนนจริงของนักเรียนที่สอบ<br>O-NET ปี 2551<br>(คะแนนเต็ม 100 คะแนน) |
|---------------------|--|--|
| -2.5                | 30.83  | 24.33  |
| -2.0                | 38.83  | 27.43  |
| -1.5                | 42.03  | 30.67  |
| -1.0                | 43.32  | 33.44  |
| -.5                 | 47.91  | 36.18  |
| .0                  | 48.48  | 39.48  |
| .5                  | 50.23  | 40.10  |
| 1.0                 | 45.38  | 41.52  |
| 1.5                 | 61.21  | 44.28  |
| 2.0                 | 62.95  | 45.50  |
| 2.4                 | 62.06  | 48.00  |
| $\bar{X}$           | 40.53<br>(คะแนนเต็ม 80 คะแนน)  | 36.22<br>(คะแนนเต็ม 100 คะแนน)                                       |
| S.D                 | 11.19  | 6.09   |

จากตาราง 4.7 พบว่า คะแนนจริงของแบบทดสอบ O-NET วิชาภาษาไทย ปี 2550 ที่ระดับความสามารถ -2.5 ถึง 2.4 มีค่าร้อยละของคะแนนจริง 30.83 ถึง 62.06 โดยมีค่าเฉลี่ยของร้อยละของคะแนนจริง 40.53 ค่าเบี่ยงเบนมาตรฐาน 11.19 คะแนนจริงของแบบทดสอบ O-NET วิชาภาษาไทย ปี 2551 ที่ระดับความสามารถ -2.9 ถึง 2.4 มีค่าร้อยละคะแนนจริงอยู่ระหว่าง

24.33 ถึง 48.00 โดยมีค่าเฉลี่ยของร้อยละของคะแนนจริง 36.22 และค่าเบี่ยงเบนมาตรฐาน 6.09

ตาราง 4.8 การเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ที่เป็นคะแนนจริงและร้อยละของคะแนนจริง ตามระดับความสามารถ สำหรับวิชาสังคมศึกษา ศาสนา และวัฒนธรรม

| ระดับ<br>ความสามารถ | ร้อยละของคะแนนจริงของ<br>นักเรียนที่สอบ<br>O-NET ปี 2550<br>(คะแนนเต็ม 80 คะแนน) | คะแนนจริงของนักเรียนที่สอบ<br>O-NET ปี 2551<br>(คะแนนเต็ม 100 คะแนน) |
|---------------------|--|--|
| -3.0                | 23.75  | 21.71  |
| -2.5                | 27.44  | 26.67  |
| -2.0                | 30.79  | 29.68  |
| -1.5                | 35.51  | 33.15  |
| -1.0                | 39.86  | 35.10  |
| -.5                 | 44.69  | 39.23  |
| .0                  | 48.00  | 41.27  |
| .5                  | 54.25  | 45.44  |
| 1.0                 | 59.25  | .7   |
| 2.0                 | 67.50  | 56.00  |
| $\bar{X}$           | 37.70  | 33.75  |
| S.D                 | 7.61   | 5.77   |

จากตาราง 4.8 พบว่า คะแนนจริงของแบบทดสอบ O-NET วิชาสังคมศึกษา ศาสนา และวัฒนธรรม ปี 2550 ที่ระดับความสามารถ -3.0 ถึง 2.0 มีค่าร้อยละของคะแนนจริงระหว่าง 23.75 ถึง 67.50 โดยมีค่าเฉลี่ยของร้อยละของคะแนนจริง 37.70 ค่าเบี่ยงเบนมาตรฐาน 7.61 ส่วนคะแนนจริงของแบบทดสอบ O-NET วิชาสังคมศึกษา ศาสนา และวัฒนธรรม ปี 2551 ที่ระดับความสามารถ -3.0 ถึง 2.0 มีค่าร้อยละของคะแนนจริงระหว่าง 21.71 ถึง 56.00 โดยมีค่าเฉลี่ยของร้อยละของคะแนนจริง 33.75 และมีค่าเบี่ยงเบนมาตรฐาน 5.77

ตาราง 4.9 การเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ที่เป็นคะแนนจริงและร้อยละของ  
คะแนนจริง ตามระดับความสามารถ สำหรับวิชาภาษาอังกฤษ

| ระดับ<br>ความสามารถ | คะแนนจริงของนักเรียนที่สอบ<br>O-NET ปี 2550<br>(คะแนนเต็ม 100 คะแนน) | คะแนนจริงของนักเรียนที่สอบ<br>O-NET ปี 2551<br>(คะแนนเต็ม 100 คะแนน) |
|---------------------|--|--|
| -3.0                | 16.89  | 18.38  |
| -2.5                | 21.85  | 22.11  |
| -2.0                | 26.68  | 25.68  |
| -1.5                | 30.00  | 29.52  |
| -1.0                | 33.88  | 32.74  |
| -.5                 | 36.90  | 35.21  |
| .0                  | 39.88  | 40.00  |
| .5                  | 42.44  | 42.30  |
| 1.0                 | 46.60  | 47.33  |
| 1.5                 | 50.50  | 48.83  |
| 2.0                 | 51.00  | 52.90  |
| 2.5                 | 58.17  | 54.71  |
| 3.0                 | 63.50  | 59.29  |
| $\bar{X}$           | 31.02  | 30.35  |
| S.D                 | 10.98  | 9.50   |

จากตาราง 4.9 พบว่า คะแนนจริงของแบบทดสอบ O-NET วิชาภาษาอังกฤษ ปี 2550 ที่ระดับความสามารถ -3.0 ถึง 3.0 มีค่าคะแนนจริงระหว่าง 16.86 ถึง 63.50 มีค่าเฉลี่ยของคะแนนจริง 31.02 และค่าเบี่ยงเบนมาตรฐาน 10.98 ส่วนคะแนนจริงของแบบทดสอบ O-NET วิชาภาษาอังกฤษ ปี 2551 ที่ระดับความสามารถ -3.0 ถึง 3.0 มีค่าคะแนนจริงระหว่าง 18.38 ถึง 59.29 มีค่าเฉลี่ยของคะแนนจริง 33.35 และมีค่าเบี่ยงเบนมาตรฐาน 9.50

ตาราง 4.10 การเปรียบเทียบคะแนน O-NET ปี 2550 และ 2551 ที่เป็นคะแนนจริงและร้อยละของคะแนนจริง ตามระดับความสามารถ สำหรับวิชาวิทยาศาสตร์

| ระดับ<br>ความสามารถ | คะแนนจริงของนักเรียนที่สอบ<br>O-NET ปี 2550<br>(คะแนนเต็ม 80 คะแนน) | คะแนนจริงของนักเรียนที่สอบ<br>O-NET ปี 2551<br>(คะแนนเต็ม 80 คะแนน) |
|---------------------|---|---|
| -2.5                | 18.94   | 16.67   |
| -2.0                | 22.86   | 20.32   |
| -1.5                | 26.49   | 22.48   |
| -1.0                | 29.75   | 24.65   |
| -.5                 | 33.70   | 26.96   |
| .0                  | 37.43   | 32.40   |
| .3                  | 38.81   | 35.00   |
| $\bar{X}$           | 27.08   | 23.45   |
| S.D                 | 10.13   | 4.07  |

จากตาราง 4.10 พบว่า คะแนนจริงของแบบทดสอบ O-NET วิชาวิทยาศาสตร์ ปี 2550 ที่ระดับความสามารถ -2.5 ถึง .3 มีค่าคะแนนจริง 18.94 ถึง 38.81 ค่าเฉลี่ยของคะแนนจริง 27.08 ค่าเบี่ยงเบนมาตรฐาน 10.13 คะแนนจริงของแบบทดสอบ O-NET วิชาวิทยาศาสตร์ ปี 2551 ที่ระดับความสามารถ -2.5 ถึง .3 มีค่าคะแนนจริง 16.67 ถึง 35.00 คะแนนเฉลี่ยของคะแนนจริง 23.45 ค่าเบี่ยงเบนมาตรฐาน 4.07

จากตารางเทียบคะแนนจริงในตาราง 4.7 – 4.10 สามารถนำมาเปรียบเทียบความแตกต่างของค่าเฉลี่ยคะแนนจริง ของคะแนน O-NET ทั้ง 2 ปี โดยสถิติ t-test Dependent samples เนื่องจากได้รายงานคะแนนจริงของผู้สอบ ณ ระดับความสามารถเท่ากันเป็นคู่ ๆ ทั้ง 2 ปี จึงเสมือนเป็นการจัดผู้สอบกลุ่มเดียวกันเพราะมีความสามารถเท่ากันมาทำการทดสอบข้อสอบ O-NET 2 ครั้ง คือสอบในปี 2550 และ ปี 2551 โดยใช้แบบทดสอบปีละฉบับ ผลการทดสอบนำเสนอ ดังตารางต่อไปนี้

ตาราง 4.11 การเปรียบเทียบค่าคะแนนจริงที่ได้จากการเทียบคะแนน O-NET วิชาภาษาไทย  
ปีการศึกษา 2550 กับ ปีการศึกษา 2551 ณ ความสามารถของผู้สอบระดับเดียวกัน

| ผลต่างของคะแนนสอบ             | จำนวนคู่ของ<br>คะแนนจริง | Mean Difference<br>( $\bar{D}$ ) | S.D. <sub>D</sub> | t-value |
|-------------------------------|--------------------------|----------------------------------|-------------------|---------|
| O-NET ปี 2550 – O-NET ปี 2551 | 51                       | 11.30                            | 3.53              | 22.84** |

\*\*  $p < .01$

จากตาราง 4.11 พบว่า แบบทดสอบ O-NET วิชาภาษาไทย ที่ใช้ทดสอบนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในปีการศึกษา 2550 และปีการศึกษา 2551 มีค่าเฉลี่ยของความแตกต่างระหว่างคะแนนจริง ณ ความสามารถของผู้สอบระดับเดียวกัน 11.30 มีค่าเบี่ยงเบนมาตรฐานของค่าความแตกต่างระหว่างคะแนนจริง 3.53 และมีค่าสถิติ t เท่ากับ 22.84 แสดงว่า ผลการเทียบคะแนน O-NET วิชาภาษาไทย ที่ทำการทดสอบในปีการศึกษา 2550 และ ปีการศึกษา 2551 โดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ มีค่าคะแนนจริงที่ได้จากการเทียบคะแนนแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยผู้สอบที่มีระดับความสามารถเดียวกัน ได้ค่าเฉลี่ยของคะแนนจริง(หรือค่าเฉลี่ยร้อยละของคะแนนจริง) จากการทดสอบข้อสอบ O-NET ปีการศึกษา 2550 สูงกว่าค่าเฉลี่ยของคะแนนจริง(หรือค่าเฉลี่ยร้อยละของคะแนนจริง) จากการทดสอบข้อสอบ O-NET ปีการศึกษา 2551

ตาราง 4.12 การเปรียบเทียบค่าคะแนนจริงที่ได้จากการเทียบคะแนน O-NET วิชาสังคมศึกษา  
ศาสนา และวัฒนธรรม ปีการศึกษา 2550 กับ ปีการศึกษา 2551 ณ ความสามารถ  
ของผู้สอบระดับเดียวกัน

| ผลต่างของคะแนนสอบ             | จำนวนคู่ของ<br>คะแนนจริง | Mean Difference<br>( $\bar{D}$ ) | S.D. <sub>D</sub> | t-value |
|-------------------------------|--------------------------|----------------------------------|-------------------|---------|
| O-NET ปี 2550 – O-NET ปี 2551 | 44                       | 4.90                             | 3.53              | 9.23**  |

\*\*  $p < .01$

จากตาราง 4.12 พบว่า แบบทดสอบ O-NET วิชาสังคมศึกษา ศาสนา และวัฒนธรรม ที่ใช้ทดสอบนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในปีการศึกษา 2550 และปีการศึกษา 2551 มีค่าเฉลี่ยของค่าความแตกต่างระหว่างคะแนนจริง ณ ความสามารถของผู้สอบระดับเดียวกัน 4.90 มีค่าเบี่ยงเบนมาตรฐานของค่าความแตกต่างของคะแนนจริง 3.53 และมีค่าสถิติ t เท่ากับ 9.23 แสดง

ว่า ผลการเทียบคะแนน O-NET วิชาสังคมศึกษา ศาสนา และวัฒนธรรม ที่ทำการทดสอบในปี การศึกษา 2550 และ ปีการศึกษา 2551 โดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ มีค่าคะแนนจริงที่ได้จากการเทียบคะแนนแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดย ผู้สอบที่มีระดับความสามารถเดียวกัน ได้ค่าเฉลี่ยของคะแนนจริง(หรือค่าเฉลี่ยร้อยละของคะแนน จริง) จากการทดสอบข้อสอบ O-NET ปีการศึกษา 2550 สูงกว่าค่าเฉลี่ยของคะแนนจริง(หรือ ค่าเฉลี่ยร้อยละของคะแนนจริง) จากการทดสอบข้อสอบ O-NET ปีการศึกษา 2551

ตาราง 4.13 การเปรียบเทียบค่าคะแนนจริงที่ได้จากการเทียบคะแนน O-NET วิชาภาษาอังกฤษ ปีการศึกษา 2550 กับ ปีการศึกษา 2551 ณ ความสามารถของผู้สอบระดับเดียวกัน

| ผลต่างของคะแนนสอบ             | จำนวนคู่ของ<br>คะแนนจริง | Mean Difference<br>( $\bar{D}$ ) | S.D. <sub>D</sub> | t-value |
|-------------------------------|--------------------------|----------------------------------|-------------------|---------|
| O-NET ปี 2550 – O-NET ปี 2551 | 70                       | 0.90                             | 1.93              | 3.85**  |

\*\*  $p < .01$

จากตาราง 4.13 พบว่า แบบทดสอบ O-NET วิชาภาษาอังกฤษ ที่ใช้ทดสอบนักเรียนชั้น มัธยมศึกษาปีที่ 6 ในปีการศึกษา 2550 และปีการศึกษา 2551 มีค่าเฉลี่ยของค่าความแตกต่าง ระหว่างคะแนนจริง ณ ความสามารถของผู้สอบระดับเดียวกัน 0.90 มีค่าเบี่ยงเบนมาตรฐานของ ค่าความแตกต่างของคะแนนจริง 1.93 และมีค่าสถิติ t เท่ากับ 3.85 แสดงว่า ผลการเทียบ คะแนน O-NET วิชาภาษาอังกฤษ ที่ทำการทดสอบในปีการศึกษา 2550 และ ปีการศึกษา 2551 โดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ มีค่าคะแนนจริงที่ได้จากการเทียบคะแนน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยผู้สอบที่มีระดับความสามารถเดียวกัน ได้ ค่าเฉลี่ยของคะแนนจริง(หรือค่าเฉลี่ยร้อยละของคะแนนจริง) จากการทดสอบข้อสอบ O-NET ปี การศึกษา 2550 สูงกว่าค่าเฉลี่ยของคะแนนจริง(หรือค่าเฉลี่ยร้อยละของคะแนนจริง) จากการ ทดสอบข้อสอบ O-NET ปีการศึกษา 2551

ตาราง 4.14 การเปรียบเทียบค่าคะแนนจริงที่ได้จากการเทียบคะแนน O-NET วิชาวิทยาศาสตร์  
ปีการศึกษา 2550 กับ ปีการศึกษา 2551 ณ ความสามารถของผู้สอบระดับเดียวกัน

| ผลต่างของคะแนนสอบ             | จำนวนคู่ของ<br>คะแนนจริง | Mean Difference<br>( $\bar{D}$ ) | S.D. <sub>D</sub> | t-value |
|-------------------------------|--------------------------|----------------------------------|-------------------|---------|
| O-NET ปี 2550 – O-NET ปี 2551 | 33                       | 3.98                             | 1.30              | 17.60** |

\*\*  $p < .01$

จากตาราง 4.14 พบว่า แบบทดสอบ O-NET วิชาวิทยาศาสตร์ ที่ใช้ทดสอบนักเรียนชั้นมัธยมศึกษาปีที่ 6 ในปีการศึกษา 2550 และปีการศึกษา 2551 มีค่าเฉลี่ยของค่าความแตกต่างระหว่างคะแนนจริง ณ ความสามารถของผู้สอบระดับเดียวกัน 3.98 มีค่าเบี่ยงเบนมาตรฐานของค่าความแตกต่างของคะแนนจริง 1.30 และมีค่าสถิติ t เท่ากับ 17.60 แสดงว่า ผลการเทียบคะแนน O-NET วิชาวิทยาศาสตร์ ที่ทำการทดสอบในปีการศึกษา 2550 และ ปีการศึกษา 2551 โดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ มีค่าคะแนนจริงที่ได้จากการเทียบคะแนนแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยผู้สอบที่มีระดับความสามารถเดียวกัน ได้ค่าเฉลี่ยของคะแนนจริง(หรือค่าเฉลี่ยร้อยละของคะแนนจริง) จากการทดสอบข้อสอบ O-NET ปีการศึกษา 2550 สูงกว่าค่าเฉลี่ยของคะแนนจริง(หรือค่าเฉลี่ยร้อยละของคะแนนจริง) จากการทดสอบข้อสอบ O-NET ปีการศึกษา 2551

**ตอนที่ 6 ผลการเปรียบเทียบคุณภาพของวิธีการแปลงคะแนน O-NET โดยการปรับเทียบให้เป็นมาตราเดียวกัน และปรับเทียบตามรูปแบบ IRT โดยการหาความสัมพันธ์ระหว่างคะแนนสอบ O-NET เดิมกับคะแนน O-NET ที่ได้จากการปรับเทียบให้เป็นมาตราเดียวกัน และที่ได้จากการปรับเทียบด้วยรูปแบบ IRT**

ตารางที่ 4.15 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบในปี 2550 กับคะแนนแปลงที่ได้จากการปรับเทียบให้เป็นมาตราเดียวกัน และ ได้จากการปรับเทียบโดยรูปแบบ IRT และระหว่างคะแนนแปลงด้วยตนเอง

| รายวิชา     | N    | $R_{O50_T151}$ | $R_{O50_T2}$ | $R_{T150_T250}$ |
|-------------|------|----------------|--------------|-----------------|
| ภาษาไทย     | 1999 | 1.00**         | .709**       | .709**          |
| สังคมศึกษาฯ | 1999 | 1.00**         | .951**       | .951**          |
| ภาษาอังกฤษ  | 1999 | 1.00**         | .972**       | .972**          |
| วิทยาศาสตร์ | 1999 | 1.00**         | .977**       | .977**          |

\*\*  $p < .01$

ตารางที่ 4.16 ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบในปี 2551 กับคะแนนแปลงที่ได้จากการปรับเทียบให้เป็นมาตรฐานเดียวกัน และ ได้จากการปรับเทียบโดยรูปแบบ IRT และระหว่างคะแนนแปลงด้วยตัวเอง

| รายวิชา     | N    | $R_{O51,T1}$ | $R_{O51,T2}$ | $R_{T151,T251}$ |
|-------------|------|--------------|--------------|-----------------|
| ภาษาไทย     | 1999 | 1.00**       | .837**       | .837**          |
| สังคมศึกษาฯ | 1998 | 1.00**       | .871**       | .871**          |
| ภาษาอังกฤษ  | 1999 | 1.00**       | .957**       | .957**          |
| วิทยาศาสตร์ | 1999 | 1.00**       | .793**       | .793**          |

\*\*  $p < .01$

จากตารางที่ 4.15 และ 4.16 พบว่า คะแนน O-NET ที่ได้จากการปรับเทียบโดยวิธีการเทียบเป็นมาตรฐานเดียวกัน และที่ได้จากการปรับเทียบโดยรูปแบบ IRT มีความสัมพันธ์กับคะแนน O-NET เดิมสูงในทางบวก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกรายวิชา ทั้ง 2 ปี และคะแนน O-NET ที่ได้จากการปรับเทียบโดยวิธีการเทียบเป็นมาตรฐานเดียวกัน มีความสัมพันธ์กับคะแนน O-NET ที่ได้จากการปรับเทียบโดยรูปแบบ IRT สูงในทางบวก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกรายวิชา ทั้ง 2 ปี

ตารางที่ 4.17 ความแตกต่างระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบในปี 2550 กับคะแนนแปลงที่ปรับเทียบโดยรูปแบบการเทียบให้เป็นมาตรฐานเดียวกัน และ ที่ปรับเทียบโดยรูปแบบ IRT และระหว่างคะแนนแปลงด้วยตัวเอง

| รายวิชา     | N    | $R_{O50,T151}$ | $R_{O50,T2}$ | $R_{T150,T250}$ | T<br>(จาก Hotelling's $T^2$ ) |
|-------------|------|----------------|--------------|-----------------|-------------------------------|
| ภาษาไทย     | 1999 | 1.00           | .709         | .709            | 0                             |
| สังคมศึกษาฯ | 1999 | 1.00           | .951         | .951            | 0                             |
| ภาษาอังกฤษ  | 1999 | 1.00           | .972         | .972            | 0                             |
| วิทยาศาสตร์ | 1999 | 1.00           | .977         | .977            | 0                             |

ตารางที่ 4.18 ความแตกต่างระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบในปี 2551 กับคะแนนแปลงที่ได้จากการปรับเทียบโดยรูปแบบการเทียบให้เป็นมาตราเดียวกัน และ ที่ปรับเทียบโดยรูปแบบ IRT และระหว่างคะแนนแปลงด้วยกันเอง

| รายวิชา     | N    | $R_{O51_{T1}}$ | $R_{O51_{T2}}$ | $R_{T151_{T251}}$ | T<br>(จาก Hotelling's $T^2$ ) |
|-------------|------|----------------|----------------|-------------------|-------------------------------|
| ภาษาไทย     | 1999 | 1.00           | .837           | .837              | 0                             |
| สังคมศึกษาฯ | 1998 | 1.00           | .871           | .871              | 0                             |
| ภาษาอังกฤษ  | 1999 | 1.00           | .957           | .957              | 0                             |
| วิทยาศาสตร์ | 1999 | 1.00           | .793           | .793              | 0                             |

จากตารางที่ 4.17 และ 4.18 พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนน O-NET ทั้ง 2 ปี ทั้ง 4 รายวิชา ที่ได้จากการปรับเทียบโดยวิธีการเทียบเป็นมาตราเดียวกัน กับคะแนน O-NET เดิม ไม่แตกต่างจากค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนน O-NET เดิม กับค่าความสามารถที่ได้จากการปรับเทียบตามรูปแบบ IRT

## บทที่ 5

### สรุป อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อเพื่อเปรียบเทียบวิธีการปรับเทียบคะแนน O-NET ระหว่างปี 2550 และปี 2551 โดยใช้ วิธีการเทียบเป็นมาตราเดียวกัน และ โดยการปรับเทียบตามรูปแบบการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบและผู้สอบตามทฤษฎีการตอบสนองข้อสอบ (IRT) 4 รายวิชา คือ 1) วิชาภาษาไทย 2) สังคมศึกษา ศาสนาและวัฒนธรรม 3) ภาษาอังกฤษ และ 4) วิทยาศาสตร์ โดยได้กำหนดสมมติฐานของการวิจัยไว้ดังนี้

1. คะแนน O-NET ต่างปีกัน ทำให้ได้คะแนน O-NET ที่ได้จากการปรับเทียบต่างกัน
2. คะแนน O-NET ต่างวิชากัน ในปีเดียวกัน ทำให้ได้คะแนน O-NET ที่ได้จากการปรับเทียบต่างกัน
3. คะแนน O-NET วิชาเดียวกัน ในต่างปีกัน ทำให้ได้คะแนน O-NET ที่ได้จากการปรับเทียบไม่แตกต่างกัน
4. คะแนน O-NET ที่ได้จากการปรับเทียบในแต่ละวิธีมีความสัมพันธ์กับคะแนน O-NET เดิม ทางบวก
5. คะแนน O-NET ที่ได้จากการปรับเทียบต่างวิธีกันมีความสัมพันธ์กันทางบวก

#### วิธีดำเนินการวิจัย

##### ประชากร

ในการวิจัยครั้งนี้ ทำการศึกษาจากคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปี 2550 และ ปี 2551 ซึ่งดำเนินการสร้างและบริหารจัดการสอบ โดย สทศ.

##### กลุ่มตัวอย่าง

กลุ่มตัวอย่าง เป็นคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทำการทดสอบข้อสอบ O-NET ปี 2550 และ ปี 2551 ใน 4 รายวิชา ทำการสุ่มตัวอย่างโดยใช้โปรแกรม SPSS และทำการสุ่มตัวอย่างให้ได้วิชาละ 1998 - 1999 คน

##### เกณฑ์ที่ใช้ในการพิจารณาคุณภาพของวิธีการปรับเทียบคะแนน O-NET

เกณฑ์ที่ใช้ในการพิจารณาคุณภาพของวิธีการปรับเทียบคะแนน O-NET คือ วิธีการปรับเทียบคะแนน O-NET ที่มีคุณภาพ จะเป็นวิธีที่ทำให้ได้คะแนนแปลง O-NET ที่มีความสัมพันธ์กับคะแนน O-NET เดิมสูงกว่า ความสัมพันธ์ระหว่างคะแนนแปลง O-NET โดยวิธีอื่นกับคะแนน O-NET เดิม

ในการวิจัยครั้งนี้ ใช้ข้อมูลทุติยภูมิของคะแนนสอบ O-NET ชั้นมัธยมศึกษาปีที่ 6 ที่สอบระหว่างปีการศึกษา 2550 – 2551 4 รายวิชา คือ 1)วิชาภาษาไทย 2)สังคมศึกษา ศาสนา และวัฒนธรรม 3)ภาษาอังกฤษ และ4)วิทยาศาสตร์

### การวิเคราะห์ข้อมูล

1. นำคะแนนก่อนปรับเทียบ และคะแนนที่ได้จากการปรับเทียบโดยวิธีการเทียบให้เป็นมาตรฐานเดียวกัน มาวิเคราะห์หาค่าสถิติต่าง ๆ ตามลำดับ ดังนี้

1.1 หาค่าสถิติพื้นฐานของคะแนนที่ปรับเทียบแล้ว ได้แก่ การวัดแนวโน้มเข้าสู่ศูนย์กลาง การวัดการกระจาย และลักษณะการแจกแจงความถี่ของข้อมูล

1.2 เปรียบเทียบคะแนน O-NET เดิม ทั้ง 2 ปี จำแนกตามรายวิชา โดยใช้สถิติ t-test independent

1.3 เปรียบเทียบคะแนน O-NET เดิม และคะแนนที่ปรับเทียบจากวิธีการเทียบให้เป็นมาตรฐานเดียวกัน จำแนกตามรายวิชา โดยใช้สถิติ t-test dependent

2. นำค่าความสามารถของนักเรียนที่ได้จากการวิเคราะห์คะแนนรายข้อทั้ง 2 ปี ทุกรายวิชา มาทำการนำเสนอคะแนนแบบจับคู่ ณ ระดับค่าความสามารถเดียวกัน

2.1 วิเคราะห์เปรียบเทียบความแตกต่างของคะแนน โดยใช้สถิติ t-test dependent เพราะถือว่า ณ ระดับความสามารถเดียวกัน เสมือนกับเป็นบุคคลเดียวกันแต่ทำการทดสอบข้อสอบ O-NET ทั้ง 2 ปี

3. วิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิม กับคะแนนที่ปรับเทียบแล้วทั้ง 2 วิธี และวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ปรับเทียบแล้วทั้ง 2 วิธี โดยใช้การหาค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน

4. ทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง คะแนน O-NET เดิม กับคะแนน O-NET ที่แปลงแล้วด้วยวิธีการเทียบให้เป็นมาตรฐานเดียวกัน และค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนคะแนน O-NET เดิม กับคะแนน O-NET ที่แปลงแล้วตามรูปแบบ IRT โดยใช้ t- Hotelling

### สรุปผลการวิจัย

#### 1. ผลสรุปลักษณะของคะแนน O-NET เดิม และที่ปรับเทียบแล้ว

1.1 ผลการพิจารณาคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่ทดสอบในปี 2550 ใน 4 รายวิชา พบว่าค่าเฉลี่ยในรายวิชาที่นักเรียนได้ค่าเฉลี่ยคะแนนสูงสุด คือวิชาภาษาไทย โดยมีค่าเฉลี่ย 50.64 รองลงมาคือรายวิชาที่ 2 และวิชาวิทยาศาสตร์ โดยมีค่าเฉลี่ย 37.7 และ

33.85 ตามลำดับ ส่วนวิชาภาษาอังกฤษ มีค่าเฉลี่ยน้อยที่สุด คือ 31.02 เมื่อพิจารณาส่วน เบี่ยงเบนมาตรฐานของทั้ง 4 รายวิชา พบว่ามีค่าอยู่ระหว่าง 7.61 – 11.19 โดย วิชาสังคมศึกษา ศาสนา และวัฒนธรรม มีค่าส่วนเบี่ยงเบนมาตรฐานน้อยกว่ารายวิชาอื่น ๆ แสดงว่าคะแนน O-NET ในวิชาสังคมศึกษา ศาสนา และวัฒนธรรม มีการกระจายน้อยกว่ารายวิชาอื่น ๆ สำหรับค่า ความโด่งของคะแนน O-NET ใน 4 รายวิชา พบว่ามีค่าอยู่ระหว่าง -.62 – 6.73 แสดงว่าคะแนนใน แต่ละรายวิชา มีการกระจายแตกต่างกัน ในส่วนของค่าความเบ้ของคะแนน O-NET ของทั้ง 4 รายวิชาพบว่ามีค่าอยู่ระหว่าง .23 – 2.28 ซึ่งมีค่าเป็นบวก นั่นคือคะแนนส่วนใหญ่มีค่าน้อย ค่า ต่ำสุดและสูงสุดของคะแนน ทั้ง 4 รายวิชา พบว่า มีค่าอยู่ระหว่าง 9- 93.75

1.2 ลักษณะของคะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของนักเรียนที่ทดสอบ ในปี 2551 ใน 4 รายวิชา พบว่า ค่าเฉลี่ยในรายวิชาที่นักเรียนได้ค่าเฉลี่ยคะแนนสูงสุด คือวิชา ภาษาไทย โดยมีค่าเฉลี่ย 36.22 ส่วนวิชาภาษาอังกฤษ มีค่าเฉลี่ยน้อยที่สุด คือ 30.35 ค่าส่วน เบี่ยงเบนมาตรฐานของทั้ง 4 รายวิชา พบว่ามีค่าอยู่ระหว่าง 4.63 – 9.50 โดย วิชาวิทยาศาสตร์ มี ค่าส่วนเบี่ยงเบนมาตรฐานน้อยกว่ารายวิชาอื่น ๆ แสดงว่าคะแนน O-NET ในวิชาวิทยาศาสตร์ มี การกระจายน้อยกว่ารายวิชาอื่น ๆ เมื่อพิจารณาที่ค่าความโด่งของคะแนน O-NET ใน 4 รายวิชา พบว่า มีค่าอยู่ระหว่าง -.05 – 1.57 แสดงว่าคะแนนในแต่ละรายวิชา มีการกระจายแตกต่างกัน ใน ส่วนของค่าความเบ้ของคะแนน O-NET ของทั้ง 4 รายวิชา พบว่า มีค่าอยู่ระหว่าง -.42 – 1.33 ซึ่ง คะแนนในวิชาภาษาไทย และ 5 มีค่าความเบ้เป็นลบ แสดงว่าคะแนนส่วนใหญ่มีค่ามาก สำหรับ คะแนนในวิชาสังคมศึกษา ศาสนา และวัฒนธรรม และ 3 มีค่าความเบ้เป็นบวก นั่นคือคะแนน ส่วนใหญ่มีค่าน้อย สำหรับค่าต่ำสุดสูงสุดของทั้ง 4 รายวิชา พบว่า มีค่าอยู่ระหว่าง 13- 68

1.3 คะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของนักเรียนที่ทดสอบในปี 2550 ที่ เป็นคะแนนเดิมและคะแนนแปลง โดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน ใน 4 รายวิชา มี ดังนี้

วิชาภาษาไทย พบว่ามีค่าเฉลี่ยคะแนนเดิมเดิมสูงกว่าค่าเฉลี่ยคะแนนแปลง และมีค่าการ กระจายของคะแนนเดิมสูงกว่าค่าการกระจายคะแนนแปลงแล้ว ค่าความโด่งและค่าความเบ้เท่า เดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดแคบลง โดยเดิมมีช่วงคะแนนระหว่าง 11- 71 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 20.14-52.80

วิชาสังคมศึกษา ศาสนา และวัฒนธรรม พบว่ามีค่าเฉลี่ยคะแนนเดิมเดิมมีค่าน้อยกว่า ค่าเฉลี่ยของคะแนนแปลงแล้ว และมีค่าการกระจายของคะแนนเดิมเดิมสูงกว่าค่าการกระจายคะแนน แปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุด แคบลง โดยเดิมมีช่วงคะแนนระหว่าง 12-63 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 20.18- 58.85

วิชาภาษาอังกฤษ พบว่ามีค่าเฉลี่ยคะแนนเดิมมีค่ามากกว่าค่าเฉลี่ยของคะแนนแปลงแล้ว แต่เป็นค่าที่แตกต่างกันไม่มาก โดยค่าเฉลี่ยคะแนนเดิมเดิมมีค่า 31.02 และหลังปรับเทียบมีค่า 30.35 มีค่าการกระจายของคะแนนเดิมก่อนและหลังปรับเทียบเท่ากัน ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดแคบลง โดยเดิมมีช่วงคะแนนระหว่าง 9 - 92 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 11.13 – 83.10

วิชาวิทยาศาสตร์ พบว่ามีค่าเฉลี่ยคะแนนเดิมเดิมมีค่ามากกว่าค่าเฉลี่ยของคะแนนแปลงและมีค่าการกระจายของคะแนนเดิมสูงกว่าค่าการกระจายคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดแคบลง โดยเดิมมีช่วงคะแนนระหว่าง 9 - 75 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 16.85 – 47.01

1.4 คะแนน O-NET ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ของนักเรียนที่ทดสอบในปี 2551 ที่เป็นคะแนนเดิมและคะแนนแปลง โดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน ใน 4 รายวิชา มีดังนี้

วิชาภาษาไทย พบว่ามีค่าเฉลี่ยคะแนนเดิมน้อยกว่าค่าเฉลี่ยคะแนนแปลง และมีค่าการกระจายของคะแนนเดิมน้อยกว่าค่าการกระจายคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดกว้างขึ้น โดยเดิมมีช่วงคะแนนระหว่าง 14-54 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง -29 – 73.21

วิชาสังคมศึกษา ศาสนา และวัฒนธรรม พบว่ามีค่าเฉลี่ยคะแนนเดิมมีค่ามากกว่าค่าเฉลี่ยของคะแนนแปลง และมีค่าการกระจายของคะแนนเดิมต่ำกว่าค่าการกระจายคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดกว้างขึ้น โดยเดิมมีช่วงคะแนนระหว่าง 18-56 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 9.21 – 59.76

วิชาภาษาอังกฤษ พบว่ามีค่าเฉลี่ยคะแนนเดิมมีค่าน้อยกว่าค่าเฉลี่ยของคะแนนแปลงแล้ว แต่เป็นค่าที่แตกต่างกันไม่มาก โดยค่าเฉลี่ยคะแนนเดิมมีค่า 30.35 และหลังปรับเทียบมีค่า 31.02 มีค่าการกระจายของคะแนนเดิมต่ำกว่าค่าการกระจายของคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดกว้างขึ้น โดยเดิมมีช่วงคะแนนระหว่าง 13 - 68 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง 10.96 – 74.54

วิชาวิทยาศาสตร์ พบว่ามีค่าเฉลี่ยคะแนนเดิมน้อยกว่าค่าเฉลี่ยของคะแนนแปลงแล้วและมีค่าการกระจายของคะแนนเดิมต่ำกว่าค่าการกระจายคะแนนแปลง ค่าความโด่งและค่าความเบ้เท่าเดิม เมื่อปรับเทียบแล้วมีช่วงพิสัยคะแนนสูงสุดและต่ำสุดแคบลง โดยเดิมมีช่วงคะแนนระหว่าง 9 - 40 เมื่อปรับเทียบแล้วมีช่วงคะแนนระหว่าง -8.17 – 59.66

1.5 ลักษณะค่าสถิติ ของคะแนนแปลงจากคะแนน O-NET ปี 2550 ที่มีค่าเท่ากับ ค่าสถิติของคะแนน O-NET ปี 2551 ได้แก่ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง สำหรับค่าที่มีการเปลี่ยนแปลงคือ คะแนนต่ำสุด และคะแนนสูงสุด และในทำนองเดียวกัน ค่าสถิติ ของคะแนนแปลงจากคะแนน O-NET ปี 2551 ที่มีค่าเท่ากับค่าสถิติของคะแนน O-NET ปี 2550 ได้แก่ ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเบ้ และค่าความโด่ง สำหรับค่าที่มีการเปลี่ยนแปลงคือ คะแนนต่ำสุด และคะแนนสูงสุด

1.6 ค่าเฉลี่ยคะแนน O-NET ปี 2550 เดิม และค่าเฉลี่ยคะแนน O-NET ปี 2550 ที่เป็นคะแนนแปลง (ซึ่งมีค่าเท่ากับค่าเฉลี่ยของคะแนน O-NET ปี 2551 เดิม) มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้ง 4 รายวิชา โดยในรายวิชาภาษาไทย ภาษาอังกฤษ และวิทยาศาสตร์ มีค่าเฉลี่ยคะแนนเดิมสูงกว่าค่าเฉลี่ยของคะแนนแปลงที่ได้จากการปรับเทียบมาจากคะแนน O-NET ปี 2551 สำหรับในวิชาสังคมศึกษา ศาสนา และวัฒนธรรม นั้นมีค่าเฉลี่ยคะแนนเดิมต่ำกว่า ค่าเฉลี่ยของคะแนนแปลงที่ได้จากการปรับเทียบมาจากคะแนน O-NET ปี 2551 โดยสรุป ค่าเฉลี่ยของคะแนน O-NET เดิมทั้ง 2 ปีมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

1.7 ค่าเฉลี่ยคะแนน O-NET ปี 2550 เดิม และค่าเฉลี่ยคะแนน O-NET ปี 2550 ที่เป็นคะแนนแปลง (ซึ่งมีค่าเท่ากับค่าเฉลี่ยของคะแนน O-NET ปี 2551 เดิม) มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้ง 4 รายวิชา โดยในรายวิชาภาษาไทย ภาษาอังกฤษ และวิทยาศาสตร์ มีค่าเฉลี่ยคะแนนเดิมสูงกว่าค่าเฉลี่ยของคะแนนแปลงที่ได้จากการปรับเทียบมาจากคะแนน O-NET ปี 2551 สำหรับในวิชาสังคมศึกษา ศาสนา และวัฒนธรรม นั้นมีค่าเฉลี่ยคะแนนเดิมต่ำกว่า ค่าเฉลี่ยของคะแนนแปลงที่ได้จากการปรับเทียบมาจากคะแนน O-NET ปี 2551 โดยสรุป ค่าเฉลี่ยของคะแนน O-NET เดิมทั้ง 2 ปีมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

## 2. สรุปผลการวิจัยตามสมมติฐานของการวิจัย

สมมติฐานข้อที่ 1: คะแนน O-NET เดิมกับคะแนนแปลง O-NET ที่ปรับเทียบแล้วไม่แตกต่างกัน

1) คะแนน O-NET เดิมกับคะแนนแปลง O-NET ที่ปรับเทียบโดยวิธีการเทียบเป็นมาตราเดียวกัน จากการศึกษพบว่า ค่าเฉลี่ยคะแนน O-NET ปี 2550 เดิม และค่าเฉลี่ยคะแนน O-NET ปี 2550 ที่ได้จากการปรับเทียบมาจากคะแนน O-NET ปี 2551 ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทั้ง 4 รายวิชา และในทำนองเดียวกัน ค่าเฉลี่ยคะแนน O-NET ปี 2551 เดิม และค่าเฉลี่ยคะแนน O-NET ปี 2551 ที่ได้จากการปรับเทียบ มาจากคะแนน O-NET ปี 2550 ไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ทั้ง 4 รายวิชา เช่นเดียวกัน

2) คะแนน O-NET เดิมกับคะแนนแปลง O-NET ที่ปรับเทียบตามรูปแบบ IRT จากการศึกษาพบว่า ผลการเทียบคะแนน O-NET ทั้ง 4 รายวิชา ที่ทำการทดสอบในปีการศึกษา 2550 และ ปีการศึกษา 2551 โดยใช้ทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ มีค่าคะแนนจริงที่ได้จากการเทียบคะแนนแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยผู้สอบที่มีระดับความสามารถเดียวกัน ได้ค่าเฉลี่ยของคะแนนจริง(หรือค่าเฉลี่ยร้อยละของคะแนนจริง) จากการทดสอบข้อสอบ O-NET ปีการศึกษา 2550 สูงกว่าค่าเฉลี่ยของคะแนนจริง(หรือค่าเฉลี่ยร้อยละของคะแนนจริง) จากการทดสอบข้อสอบ O-NET ปีการศึกษา 2551

สมมติฐานข้อที่ 2 : คะแนน O-NET ที่ได้จากการปรับเทียบในแต่ละวิธีมีความสัมพันธ์กับคะแนน O-NET เดิม ทางบวก

คะแนน O-NET ที่ได้จากการปรับเทียบโดยวิธีการเทียบเป็นมาตราเดียวกัน และที่ได้จากการปรับเทียบโดยรูปแบบ IRT มีความสัมพันธ์กับคะแนน O-NET เดิมสูงในทางบวก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกรายวิชา ทั้ง 2 ปี

สมมติฐานข้อที่ 3: คะแนน O-NET ที่ได้จากการปรับเทียบต่างวิธีกันมีความสัมพันธ์กันทางบวก

คะแนน O-NET ที่ได้จากการปรับเทียบโดยวิธีการเทียบเป็นมาตราเดียวกัน มีความสัมพันธ์กับคะแนน O-NET ที่ได้จากการปรับเทียบโดยรูปแบบ IRT สูงในทางบวก อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกรายวิชา ทั้ง 2 ปี

### 3. สรุปผลการเปรียบเทียบคุณภาพวิธีการปรับเทียบคะแนน O-NET

ผลการเปรียบเทียบคุณภาพวิธีการปรับเทียบคะแนน O-NET ระหว่างวิธีการเทียบเป็นมาตราเดียวกัน กับ การปรับเทียบตามรูปแบบ IRT พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนน O-NET ทั้ง 2 ปี ทั้ง 4 รายวิชา ที่ได้จากการปรับเทียบโดยวิธีการเทียบเป็นมาตราเดียวกัน กับคะแนน O-NET เดิม ไม่แตกต่างจากค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนน O-NET เดิม กับค่าความสามารถที่ได้จากการปรับเทียบตามรูปแบบ IRT

#### อภิปรายผล

จากผลการศึกษาครั้งนี้ซึ่งมีวัตถุประสงค์ที่สำคัญคือ เพื่อเปรียบเทียบการปรับเทียบคะแนน O-NET ที่จัดสอบระหว่างปีการศึกษา 2550 และ ปีการศึกษา 2551 โดยใช้ รูปแบบการปรับเทียบให้เป็นมาตราเดียวกัน กับ การปรับเทียบตามรูปแบบ IRT ซึ่งอภิปรายผลได้ ดังนี้

1. จากผลการศึกษาลักษณะของคะแนน O-NET เดิม ทั้ง 2 ปี พบว่า ค่าเฉลี่ยของคะแนนปี 2550 และ ปี 2551 มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และผนวกกับผลการศึกษาความแตกต่างของคะแนนจริง ณ ค่าความสามารถเดียวกัน ที่พบว่า หากบุคคลที่มีความสามารถเท่ากันทำการทดสอบข้อสอบ O-NET ทั้ง 2 ปี จะได้คะแนนต่างกัน ซึ่งข้อค้นพบดังกล่าวยืนยันว่าในการเทียบคะแนนทั้งสองปีนั้นไม่ควรนำมาเทียบกันโดยตรง เพราะจะทำให้เกิดความลำเอียง และไม่เป็นธรรมได้ เพราะโดยหลักการนำคะแนนของผู้สอบซึ่งเป็นคนละกลุ่ม และทำการทดสอบคนละฉบับมาทำการปรับเทียบกันได้โดยตรงนั้น ลอร์ด (ศิริชัย กาญจนวาสี . 2543 : 18 ; อ้างอิงจาก Lord. 1980 : 208 – 209) ได้เสนอแนะว่า การเทียบคะแนนให้เกิดความเสมอภาคได้นั้น แบบทดสอบ X และ Y จะต้องคู่ขนานกันอย่างแท้จริง หรือแบบทดสอบทั้งสองฉบับจะต้องมีความเชื่อมั่นอย่างสมบูรณ์ (Reliability = 1) เนื่องจากการทำนาย Y จาก X หรือทำนาย X จาก Y ในทางวิเคราะห์การถดถอยนั้น จะมีความถูกต้องก็ต่อเมื่อ การวัดค่า x (หรือ y) ได้อย่างถูกต้องสมบูรณ์ปราศจากความคลาดเคลื่อน นอกจากนี้แบบทดสอบ X และ Y จะต้องมีความสัมพันธ์กับคะแนนเกณฑ์เท่ากัน มิเช่นนั้น จะทำให้ความสัมพันธ์ของ X และ Y ที่ได้ ผันแปรไปตามกลุ่มที่ใช้ศึกษา แต่อย่างไรก็ตามในการสร้างข้อสอบให้เป็นข้อสอบที่คู่ขนานกันอย่างแท้จริงนั้นทำได้ยาก เพื่อให้เกิดความเสมอภาคในการนำคะแนนมาเทียบกันไม่ว่าจะเพื่อวัตถุประสงค์ใดก็ตามควรมีการใช้วิธีการปรับเทียบคะแนนเสียก่อนที่จะนำคะแนนมาเทียบเคียงกัน

2. ผลจากการศึกษาที่พบว่า คะแนน O-NET ปี 2550 เดิม กับคะแนนแปลง O-NET ที่ได้จากการปรับเทียบมาจากคะแนน O-NET ปี 2551 โดยวิธีการปรับเทียบให้เป็นมาตราเดียวกันมีค่าเฉลี่ยของคะแนนไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และผลการศึกษากับคะแนน O-NET ปี 2551 เดิมก็ได้ผลเช่นเดียวกัน ซึ่งสอดคล้องกับสมมติฐาน

ขณะเดียวกันผลการปรับเทียบคะแนนโดยรูปแบบ IRT ได้ผลที่ไม่สอดคล้องกับสมมติฐาน เพราะคะแนนจริงของผู้เรียนที่ค่าความสามารถเดียวกันมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

จากผลดังกล่าวนี้ ผู้ดำเนินการจัดสอบคะแนน O-NET ควรต้องพิจารณาถึงความเหมาะสมในการเลือกวิธีการปรับเทียบในแต่ละวิธีจากข้อค้นพบดังกล่าวถ้าต้องการใช้วิธีการปรับเทียบแบบแรก ต้องมั่นใจข้อมูลคะแนนของข้อสอบ 2 ชุดนี้มีการแจกแจงเหมือนกัน แล้วจึงเลือกใช้วิธีการปรับเทียบให้เป็นมาตราเดียวกันนี้ ซึ่งถือว่าเป็นวิธีการที่ไม่ซับซ้อน สามารถดำเนินการได้ง่าย แต่ถ้าไม่มั่นใจในการแจกแจงข้อมูลคะแนน 2 ชุด และเป็นการสอบจากผู้สอบคนละกลุ่มและใช้แบบสอบคนละฉบับ ซึ่งเป็นลักษณะของการจัดสอบ O-NET อยู่ในขณะนี้ วิธีการปรับเทียบตามรูปแบบ IRT จะเป็นวิธีการที่เหมาะสมมากกว่า

3. จากผลการศึกษาที่พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมกับคะแนนแปลงที่ได้จากการปรับเทียบทั้งโดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน และวิธีการปรับเทียบตามรูปแบบ IRT มีค่าเป็นบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งสอดคล้องกับสมมติฐานที่การวิจัยข้อที่ 2 ที่ตั้งไว้ ทั้งนี้อาจเป็นเพราะว่าวิธีการปรับเทียบคะแนนต่างวิธีกันมีความตรงในการปรับเทียบคะแนน และสามารถเลือกใช้วิธีการปรับเทียบคะแนนทั้ง 2 วิธีนี้แทนกันได้

4. จากข้อค้นพบที่ว่าค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิมกับคะแนนแปลงที่ปรับเทียบโดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน มีค่าสูงกว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนน O-NET เดิม กับค่าความสามารถที่ปรับเทียบตามรูปแบบ IRT แต่เมื่อทำการทดสอบความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ทั้งสองค่าโดยใช้สถิติ t-Hotelling แล้วพบว่า ค่าสัมประสิทธิ์ทั้ง 2 ค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งถ้าพิจารณาเกณฑ์ในการพิจารณาคุณภาพวิธีการปรับเทียบคะแนนแล้ว น่าจะพอสรุปได้ว่าวิธีการปรับเทียบคะแนน O-NET ทั้ง 2 วิธีให้คุณภาพสูง เพราะทั้ง 2 วิธีต่างก็ให้คะแนนแปลงที่มีค่าสัมประสิทธิ์สหสัมพันธ์สูง และสัมพันธ์ทางบวกกับคะแนน O-NET เดิมทั้งคู่ และให้คุณภาพที่ไม่แตกต่างกันด้วย

### ข้อเสนอแนะสำหรับสถาบันทดสอบทางการศึกษาแห่งชาติ

จากผลการศึกษาพบว่าวิธีการปรับเทียบคะแนนทั้ง 2 วิธีนี้มีประสิทธิภาพไม่แตกต่างกัน แต่ความยากง่ายในการปรับเทียบแตกต่างกัน จึงเสนอแนะเพื่อให้เป็นทางเลือกในการพิจารณาดัดลึกลงไปใช้ ดังนี้

1. ในกรณีที่มีการพิจารณาเปรียบเทียบคะแนน O-NET ต่างปีกัน ควรพิจารณาเปรียบเทียบความเท่ากันของค่าเฉลี่ยของคะแนนในแต่ละปีก่อน ถ้าไม่เท่ากันควรมีการปรับเทียบคะแนนก่อนแล้วจึงนำมาเปรียบเทียบกัน

2. สำหรับเกณฑ์การเลือกใช้วิธีการปรับเทียบคะแนน O-NET ที่ต่างปีกัน โดยวิธีการปรับเทียบให้เป็นมาตราเดียวกัน คือการแจกแจงของคะแนนจากแบบสอบทั้ง 2 ชุด ควรมีลักษณะคล้ายกัน หรือหากมีความแตกต่างกันก็มีเพียงเล็กน้อย เช่นอาจมีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานแตกต่างกัน หรือถ้าผู้พัฒนาข้อสอบได้ทำการตรวจสอบแล้วว่าพบว่าแบบทดสอบทั้งสองฉบับมีความเป็นคู่ขนานกัน ก็สามารถเลือกใช้วิธีการปรับเทียบคะแนนโดยวิธีนี้ได้ ซึ่งถ้าผลการศึกษาลักษณะของข้อมูลคะแนนทั้ง / ชุดเป็นไปตามเกณฑ์ ขอเสนอแนะให้ใช้วิธีนี้เป็นทางเลือกในการใช้ปรับเทียบคะแนน เนื่องจากคะแนนที่ได้จากการปรับเทียบมีค่าสัมประสิทธิ์

สหสัมพันธ์สูงสุดทางบวก และเป็นวิธีการที่ง่าย มีขั้นตอนที่ไม่ยุ่งยากซับซ้อน และสามารถคำนวณกับข้อมูลคะแนนจำนวนมาก ๆ ได้ เพราะไม่มีข้อจำกัดเกี่ยวกับโปรแกรมในการวิเคราะห์ข้อมูล

3. ในการเลือกใช้วิธีการเปรียบเทียบคะแนน O-NET ที่ต่างปีกัน และผู้สอบมาจากต่างกลุ่มกัน ถ้าไม่มั่นใจว่าแบบทดสอบ 2 ฉบับนั้นมีความเป็นคู่ขนานกันจริงหรือไม่ วิธีการที่ปลอดภัยที่สุดตามคำแนะนำของ Lord (1977 อ้างถึงใน ภาวณิ ศรีสวัสดิ์วัฒนานันท์. 2529:41) คือ ให้จัดกระทำในฐานะที่ข้อสอบนั้นไม่เป็นคู่ขนานกัน และใช้วิธีการประเภทที่ไม่ใช่โมเดลเส้นตรง นั่นคือใช้การออกแบบเพื่อศึกษาการเทียบมาตรฐานโดยใช้ทฤษฎีการตอบสนองข้อสอบ(หรือตามรูปแบบIRT)นั่นเอง โดยในการใช้การเปรียบเทียบคะแนนตามรูปแบบIRT นี้ต้องเริ่มต้นด้วยการใช้โปรแกรมประมวลค่าพารามิเตอร์ของข้อสอบ ที่มีหลากหลายโปรแกรม ซึ่งถ้าผู้ทำการเปรียบเทียบคะแนนความสนใจศึกษาค้นคว้าให้เข้าใจในวิธีการวิเคราะห์ข้อมูลให้ถ่องแท้แล้ว ผู้วิจัยเห็นว่าวิธีการเปรียบเทียบคะแนนตามรูปแบบIRT นี้เหมาะสมอย่างยิ่งต่อการเปรียบเทียบคะแนน O-NET ที่ต่างปีกันและกลุ่มผู้สอบก็ต่างกลุ่มกัน เพราะลักษณะของรูปแบบการสอบสอดคล้องกับการใช้รูปแบบIRT เปรียบเทียบมากที่สุด

### **ข้อเสนอแนะในการศึกษาวิจัยต่อเนื่อง**

1. ควรมีการศึกษาค้นคว้าความแตกต่างของผลการเทียบคะแนนในรูปแบบอื่น ๆ ด้วย เช่น ทำการเปรียบเทียบคะแนนตามแนวตั้งในข้อสอบ O-NET วิชาเดียวกัน แต่ต่างระดับชั้นกัน เป็นต้น
2. ควรทำการศึกษารูปแบบการเปรียบเทียบคะแนน O-NET ที่ต่างปีกัน ในรายวิชาเดียวกัน โดยพิจารณาจำนวนข้อสอบรวมภายในที่อาจแทรกอยู่ในข้อสอบแต่ละฉบับ เพื่อค้นหาคำตอบว่าถ้าข้อสอบรวมมีความยาวแตกต่างกัน ควรใช้วิธีการเปรียบเทียบใดจึงจะเหมาะสมที่สุด
3. ควรใช้รูปแบบIRT เปรียบเทียบคะแนน O-NET ให้เป็นค่าความสามารถที่แท้จริงของผู้สอบแล้วศึกษาวิธีกำหนดเกณฑ์มาตรฐานในการจัดกลุ่มผู้เรียนเป็น กลุ่มเก่ง ปานกลาง อ่อนตามค่าคะแนน O-NET โดยผ่านการเปรียบเทียบคะแนนเป็นค่าความสามารถมาผู้เรียน

## บรรณานุกรม

- กระทรวงศึกษาธิการ, สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ, สำนักงานทิศและพัฒนา  
มาตรฐานการศึกษาแห่งชาติ. (2544). **พัฒนาการของคุณภาพนักเรียน  
ประถมศึกษา และแนวทางการประเมิน.** กรุงเทพฯ : โรงพิมพ์การศาสนา.  
กนิษฐา แสงแก้ว.(2541).**การปรับแก้เฉลยสะสมระดับมัธยมศึกษาตอนปลายด้วยคะแนน  
ความถนัดทางการเรียน : การเปรียบเทียบระหว่างวิธีการเปรียบเทียบอิกวิ  
เปอร์เซนไทล์เชิงเส้นตรงและไออาที.** วิทยานิพนธ์ คม. กรุงเทพฯ : จุฬาลงกรณ์  
มหาวิทยาลัย.
- เกตรินทร์ หมัดอะดัม. (2545).**การเปรียบเทียบคุณภาพของการเทียบมาตรฐานตามแนวตั้ง  
ระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสาม  
พารามิเตอร์.** วิทยานิพนธ์ กศ.ม. สงขลา : มหาวิทยาลัยทักษิณ,
- จิราภา นาครินทร์. (2549). **การเปรียบเทียบผลการเทียบระหว่างวิธีการตอบข้อคำถามที่ใช้  
คะแนนจริง คะแนนสังเกต และวิธีเชิงเส้นตรง.** วิทยานิพนธ์ กศ.ม., มหาวิทยาลัย  
มหาสารคาม.
- ชูชีพ พงษ์สมบุญ. (2528). **การเปรียบเทียบการเทียบมาตรฐานระหว่างรูปแบบที่ใช้ผู้สอบ  
ร่วมกับรูปแบบที่ใช้แบบสอบร่วม.** วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์  
มหาวิทยาลัย.
- ชูศักดิ์ ขัมภลิขิต. (2527). **การปรับเทียบคะแนน.** ในบทคัดสรรทางวิชาการทดสอบ นนทบุรี :ศูนย์  
พัฒนาศึกษาแห่งชาติของประเทศไทย ทบวงมหาวิทยาลัย.
- ชูศักดิ์ ขัมภลิขิต. (2529). **การเปรียบเทียบคะแนนทดสอบ.** ในโครงการพัฒนาศึกษาอาเซียน บทคัด  
สรรทางวิชาการทดสอบ นนทบุรี : โรงพิมพ์มหาวิทยาลัยสุโขทัยธรรมมาธิราช.
- ดร.ณี พรประเสริฐ. (2541).**การเทียบคะแนนระหว่างแบบทดสอบระดับประเทศและ  
แบบทดสอบระดับจังหวัด กลุ่มทักษะคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6  
การศึกษา 2540.** วิทยานิพนธ์ คม. กรุงเทพ : มหาวิทยาลัยรามคำแหง.
- พรพิมล นาคเวช.(2537). **การศึกษาคุณภาพของการเทียบมาตรฐานในแนวตั้งโดยใช้ทฤษฎี  
การตอบสนองข้อสอบ.** วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.
- พิชัย ละแมนชัย.(2538). **ขนาดกลุ่มตัวอย่างขั้นต่ำสำหรับการปรับเทียบคะแนนในแนว  
ระดับตามแนวทฤษฎีการตอบสนองรายข้อ ระหว่างแบบสอบที่มีแบบแผนการ  
ปรับเทียบและความยาวของแบบทดสอบแตกต่างกัน .** วิทยานิพนธ์ ค.ด. (การวัด  
และประเมินผลการศึกษา). กรุงเทพฯ :จุฬาลงกรณ์มหาวิทยาลัย.

- บุญทริกา ดิวงษ์. (2543). การศึกษาคุณภาพของการเทียบคะแนนแบบทดสอบคู่ขนาน ด้วยรูปแบบการเทียบมาตราเชิงเส้นตรง. ปรินซ์นิพนธ์ กศ.ม., มหาวิทยาลัยมหาสารคาม.
- ปราโมทย์ ศรีเพชร. (2549). การศึกษาพัฒนาการความสามารถทางคณิตศาสตร์เรื่องจำนวนนับของนักเรียนระดับช่วงชั้นที่ 2 โดยการเทียบคะแนนในแนวตั้ง. วิทยานิพนธ์ กศ.ม. มหาสารคาม : มหาวิทยาลัยมหาสารคาม.
- ภัทรพร เกษสังข์. (2546). การศึกษาพัฒนาการความสามารถทางคณิตศาสตร์ของนักเรียนชั้นมัธยมศึกษาตอนต้น โดยการเทียบคะแนนในแนวตั้งที่ใช้วิธีการที่เหมาะสม. ปรินซ์นิพนธ์ กศ.ด., มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ภาวิณี ศรีสุขวัฒนานันท์. (2529) การเปรียบเทียบผลจากการใช้รูปแบบการเทียบมาตราที่แตกต่างกันเมื่อแบบทดสอบร่วมมีความยาวต่างกัน. วิทยานิพนธ์ ค.ด., จุฬาลงกรณ์มหาวิทยาลัย.
- เวรดี อินทสระ. (2530). การเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตราระหว่างรูปแบบอิงทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ. ปรินซ์นิพนธ์ กศ.ม., มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- รัตนา ชมานนท์. ( 2545). การพัฒนาแบบทดสอบวัดความถนัดวิชาชีพบัญชีและการเทียบคะแนนตามแนวนอนที่วิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบ. วิทยานิพนธ์ กศ.ม. สงขลา มหาวิทยาลัยทักษิณ.
- รุ่งนภา จันทรา. (2540). การเปรียบเทียบความคลาดเคลื่อนและความเพียงพอของการเทียบมาตราระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์. วิทยานิพนธ์ ศษ.ม. ปัตตานี : มหาวิทยาลัยสงขลานครินทร์.
- วรรณตรี พิชิตเกริกพล. (2537). การเปรียบเทียบคุณภาพของวิธีการเทียบมาตราระหว่างแบบอควิเปอร์เซนไดล์ กับแบบเชิงเส้นตรง. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- วรรณดี แสงประทีปทอง. (2538). การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้นมัธยมศึกษาตอนต้น โดยการเทียบคะแนนตามแนวตั้งตามทฤษฎีการตอบสนองข้อสอบ แบบจำลองโลจิสติกสามพารามิเตอร์. ปรินซ์นิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- วิรัช วรรณรัตน์. (2530). “การเทียบมาตรา (Equating),” วารสารการวัดผลการศึกษา. 9 (26) : 69 – 73; กันยายน-ธันวาคม.

- ศิริชัย กาญจนวาสี. (2541). การปรับเทียบคะแนนระหว่างแบบสอบ (Test Equating).  
กรุงเทพฯ: ศูนย์ตำราและเอกสารทางวิชาการ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- \_\_\_\_\_. (2545). **ทฤษฎีการทดสอบแนวใหม่ (MODERN TEST THIORIES)**. กรุงเทพฯ :  
จุฬาลงกรณ์มหาวิทยาลัย,
- สงบ ลักษณะ. (2525 กันยายน – ธันวาคม). การเทียบระดับคะแนนระหว่างแบบทดสอบ  
(EQUATING) **วารสารวัดผลการศึกษา** 4(2) : 21 – 32
- สถาบันทดสอบทางการศึกษาแห่งชาติ ,<http://www.niets.or.th/pdf/history.pdf> สืบค้นวันที่ 7  
พฤศจิกายน 2551.
- สถาบันทดสอบทางการศึกษาแห่งชาติ ,<http://www.niets.or.th/pdf/obligation.pdf> สืบค้นวันที่ 7  
พฤศจิกายน 2551.
- สถาบันทดสอบทางการศึกษาแห่งชาติ ,<http://www.niets.or.th/pdf/authority.pdf> สืบค้นวันที่ 7  
พฤศจิกายน 2551.
- สมพงษ์ แก้วแฉล้ม. (2539). **การพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการเทียบคะแนนเชิง  
เส้นตรงตามแนวนอนโดยใช้แบบสอบร่วม**. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผล  
การศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- สำนักงานคณะกรรมการการศึกษาแห่งชาติ.สำนักงานกฤษฎมนตรี. (2545). **พระราชบัญญัติ  
การศึกษาแห่งชาติ พ.ศ. 2542 และที่แก้ไขเพิ่มเติม (ฉบับที่ 2) พ.ศ.2545**.  
กรุงเทพฯ: บริษัทพริกหวานจำกัด.
- สำราญ มีแจ้ง. (2542). "การเปรียบเทียบคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบ  
ใหม่ที่มีวิธีแปลงคะแนนผลการเรียนสะสมต่างกัน" **ศึกษาศาสตร์ มหาวิทยาลัยนเรศวร**.  
4(1) พฤษภาคม – สิงหาคม .
- สุจินดา ฝ่องอักษร. (2533). **การศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่ม  
ทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา  
พุทธศักราช 2521ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531)  
โดยใช้การเทียบคะแนนรูปแบบราล์ซ**. ปริญญาานิพนธ์ กศ.ด.กรุงเทพฯ :  
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร,
- สุนิสา จั๋ยม่วงศรี. (2537).**ผลของความยาวของแบบสอบร่วมที่มีต่อคุณภาพของวิธีการ  
เทียบ มาตราเชิงเส้นตรง**. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์  
มหาวิทยาลัย,
- \_\_\_\_\_. (2546).**การศึกษาผลของการเทียบคะแนนข้อสอบที่ตรวจให้คะแนนแบบหลาย  
ค่า** ปริญญาานิพนธ์ กศ.ด. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ,

- สุรัตน์ ขวัญบุญจันทร์. (2531). **การสร้างแบบทดสอบและตารางเทียบคะแนนในแนวนอน วิชาวิทยาศาสตร์ที่วิเคราะห์ด้วยราสซ์โมเดล**. ปรินซ์ตันนิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- สุมนา ไสถติผลอนันต์. (2539). **การศึกษาวิธีการเทียบคะแนนแบบเส้นตรงตามแบบจำลอง คะแนนจริงสัมพันธ์**. ปรินซ์ตันนิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิต วิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- สุภามาส อังศุโชติ. (2543). **การปรับคะแนนเฉลี่ยสะสมระดับมัธยมศึกษาตอนปลาย : การ เปรียบเทียบความตรงเชิงทำนาย**. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผล การศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- สุภาภรณ์ คงทวี. (2541). **การปรับแก้เฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายด้วยคะแนน สอบเข้ามหาวิทยาลัย : การเปรียบเทียบระหว่างวิธีการปรับเทียบแบบอิกวี เเปอร์เซนไทล์เชิงเส้นตรงและไออาร์ที**. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผล การศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
- สุภาภรณ์ หนูเกิด. **การเปรียบเทียบคุณภาพการเทียบมาตราระหว่างวิธีเชิงเส้นตรงกับวิธี อิกวีเปอร์เซนไทล์เมื่อสัดส่วนความยาวของแบบสอบเทียบมาตรต่างกัน**. ปรินซ์ตันนิพนธ์ กศ.ม. ปีที่ 2547. ปัตตานี : มหาวิทยาลัยสงขลานครินทร์, 2547.
- อาภรณ์ กาญจนกิจโสภณ. (2531). **การสร้างแบบสอบและการเทียบระดับคะแนนตาม แนวตั้งในวิชาคณิตศาสตร์**. ปรินซ์ตันนิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรี นครินทรวิโรฒประสานมิตร,
- เอกชัย สุขเคหา.(2549). **ผลการเทียบคะแนนตามแนวนอนที่วิเคราะห์ด้วยทฤษฎีการ ตอบสนองข้อสอบ เมื่อใช้รูปแบบการเก็บรวบรวมข้อมูลที่แตกต่างกัน**. วิทยานิพนธ์ กศ.ม., มหาวิทยาลัยนเรศวร.
- อุทัยวรรณ พงศ์อร่าม. (2545). **การศึกษาขนาดกลุ่มตัวอย่างที่เหมาะสมสำหรับการปรับเทียบ คะแนนด้วยวิธีอิกวีเปอร์เซนไทล์และวิธีเชิงเส้นตรงตามแบบจำลองคะแนนจริง สัมพันธ์ที่มีแบบแผนการปรับเทียบและความยาวของแบบทดสอบแตกต่างกัน**. ปรินซ์ตันนิพนธ์ กศ.ด.,มหาวิทยาลัยศรีนครินทรวิโรฒ.
- Angoff W.H. 1971. **Scales, Norms, and Equivalent Scores**. Educational Measurement. 2<sup>nd</sup> ed. Washington, D.C. : American Council on Education : 508-600 ; In R.L. Thorndike (Ed.),

- \_\_\_\_\_. 1984. **Scales, Norms, and Equivalent Scores**. Princeton, NJ : Educational Measurement,
- Ayerve, R.I. 1992. **The Effectiveness of the Equipercntile Method and IRT Three – Parameter Model on Vertical Equating Under Varying Conditions of Sample Size, Test Length, and Anchor Test Length : A Simulation Study**. Dissertation Abstracts International. 53: 1841 – A; December,
- Caldwell, L.J. (1984) **A Comparison of Equating error in Linear and Rasch Model Test Equating Methods**. Doctoral Dissertation, The Florida State University, Dissertation Abstracts International, 49: 2847.
- Cope, Ronald T.(1987) .**How Well Do the Angoff Design V Linear Equating Methods Compare With the Tucker and Levine Methods**. Applied Psychological Measurement. 11(2): 143-149
- Flanagan, T.C (1951). **Unit, Scores and Norms**. in E.F. Lindquist (Ed.) Educational Measurement. Washington D.C. : American Council on Education, P. 695-763
- Glowacki, M.L. 1991. **An analysis of test equating models for the Alabama high school graduation examination**. Doctoral dissertation, the University of Alabama, Dissertation Abstracts International.. 52(5-A): 1722; November.
- Hambleton, R.K. (1989). **Principles and Selected Applications of Item Response Theory**. in R.L. Linn (Ed) Educational Measurement . 3 rd. New York : Macmillan, p. 9-49
- Hambleton, R.K. and H. Swaminathan (1985). **Item Response Theory**. Kluwer. Nijhoff, Boston.
- Harris, D.J. **Acomparison of Angoff 's Design 1991.I and Design II for vertical equating Using tradition and IRT methodology**. Journal of Education Measurement. 28; Fall,; 221 – 235.
- Hendrickson A.B. & Kolen M.J.2003. **IRT Equating of the MCAT**.University of IOWA, USA.
- Jaeger, R.M. (1981). **Some Exploratory Indicies for Selection of a Test Equating**

- Method.** Journal of Educational Measurement. 18: 23-38.
- Johanson, George A. 1987. **A Study of item Response Theory Equating with an Anchor Test Design**, Doctoral dissertation, University of Massachusetts, Dissertation Abstracts International. 49(3-A): 448; September.
- Kolen. M.J. and Whitney, D.R. 1982. **Comparison of Four Procedure for Equating the Tests of General Education Development.** Journal of Education Measurement. 19 (4) : 279-293 ;
- Kolen. M.J. 1981. **Comparison of traditional and Item Response Theory Methods for Equating test.** Journal of Education Measurement. 18 (1): 1-11; Spring,
- Hulin. C.L, Lissak, R..L. , and Drasgow, F (1982). **Recovery of two – and three Parameter Logistic Item Characteristic Curve : A Monte Carlo Study.** Applied Psychological Measurement . 16 : 153 – 158
- Linden W. J.(2006). **Evaluating Equating Error in Observed-Score Equating.** University of Twente. The Natherlands.
- Livingston S. A. (2004). **Equating Test Scores (Without IRT).** Educational Testing Service, USA.
- Lord, F.M. ( 1977). **Practical Applications of Item Characteristic Curve Theory.** Journal of Educational Measurement . 14 : 117 – 138.
- Lord, F.M. 1980.**Applications of Item Response Theory to Practical Testing Problems.**Hillssdale, N.J. : Erlbaun,
- Loyd, B.H., and Hoover, H.D. 1980. **Vertical equating using the Rasch model.** Journal of Educational Measurement 17: 179 – 192; Fall,
- MacCann, R.A. (1989). **A Comparison of Two Observed-scores Equating Methods that Assume Equally Reliable Congeneric Test..** Applied Psychological Measurement 13(3) : 236-276 .
- Patience, W.M. 1990. **A Comparison of classical and Item Response Theory Methods of Vertical Equating Test..** Dissertation abstracts International. 51 : 486 – A ; August.
- Petersen, N.S., Kolen, M.J., and Hoover, H.D. 1989. **Scaling, Norming, and Equating.** Republication of Chapter 6. In R.L. Linn (Ed.),\_Educational

- Measurement, (3<sup>rd</sup> ed), pp.221-262. Washington D.C. : American Council on Education.
- Petersen, N.S., Marco, C.L., and Stewart, E.E. 1982. **A test of the adequacy of linear Score equating methods.** In P.W. Holland, and D.B. Rubin (ed), Test Equating, pp. 71-135. New York: Academic press.
- Rentz and Bashaw. (1977). **The National Reference Scale for Reading : An Application of the Rasch Model.** Journal of Educational Measurement. 14 : 161 – 178.
- Skaggs, G.E. And Lissits, R.W. (1986). **IRT Test Equating : Relevant Issues and Review of Rrecent Research.** Review of Educational Research . 56 : 495 – 529.
- Sontag, Lauren M. 1984. **Vertical Equating Method: A Comparative Study of Their Efficacy.** Doctoral dissertation, Columbia University Teachers College, Dissertation Abstracts International. 45 (3-B): 1000; September.
- Triscari, R.S. (1990). **A Comparison of Classical and Item Response Theory Equating Methods Using a Composite Score of Direct and Indirect Writing Assesment.** Doctoral Dissertation, The University of Texas at Austin, 1990. Dissertation Abstracts International, 51 : 1996
- Wiley, A. 1999. **An Investigration Into Two Models For Equating Examinations With Multiple Item Formats.** Doctoral dissertation, Fordham University. Dissertation Abstracts International. 60(4-B) : 1910 ; Oct..
- Woodruff, D. (1986). **Deriviations of Observed Score Linear Equating Methods Base on Test Scores Models for the Common Item Nonequivalent Population Design.** Journal of Education Statistic. 11(4) : 245-257
- Yang, Wen-Ling. 1991. **The Effects of Content Homogenety and Equating Method on the Accuracy of Common-Item Test Equating.** Doctoral dissertation, Michigan State University. **Dissertation Abstracts International.** 52(5-A): 1722; January,
- Young, J. W. (1995) . **A Comparison of Two Adjustment Methods for limproving the Prediction of Law School.** Educational and Psychological Measurement. 55: 558-571.

ภาคผนวก

ประวัติผู้วิจัย

## ประวัติผู้วิจัย

### 1. รองศาสตราจารย์ ดร.สำราญ มีแจ้ง

#### 1.1 ประวัติส่วนตัว

1. ชื่อ นายสำราญ มีแจ้ง ตำแหน่ง รองศาสตราจารย์ ระดับ 9 สังกัด คณะศึกษาศาสตร์  
อัตราเงินเดือน 38,940 เลขประจำตำแหน่ง 381 อายุราชการ 31 ปี
2. เกิดเมื่อวันที่ 15 มิถุนายน พ.ศ. 2497 อายุ 54 ปี ณ บ้านเลขที่ 96 ม.3 ตำบลบางคูเวียง  
อำเภอบางกรวย จังหวัดนนทบุรี
3. สถานภาพครอบครัวในปัจจุบัน คู่สมรสชื่อ นางอารีรักษ์ มีแจ้ง รับราชการเป็นข้าราชการ  
มหาวิทยาลัยนเรศวร ตำแหน่งผู้ช่วยศาสตราจารย์ ระดับ 8 ภาควิชาภาษา คณะมนุษยศาสตร์  
มหาวิทยาลัยนเรศวร มีบุตร 1 คน ชื่อ เด็กหญิงอิสริยาภรณ์ มีแจ้ง

#### 1.2 ประวัติการศึกษา

1. จบปริญญาเอก (ค.ด.) การวัดและประเมินผลการศึกษา จุฬาลงกรณ์มหาวิทยาลัย ปี พ.ศ. 2534
2. จบปริญญาโท (ค.ม.) การวัดและประเมินผลการศึกษา จุฬาลงกรณ์มหาวิทยาลัย ปี พ.ศ. 2525
3. จบปริญญาตรี (กศ.บ.) เอกคณิตศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตรปี พ.ศ. 2520
4. จบประกาศนียบัตรการศึกษาชั้นสูง (ป.กศ.สูง) เอกคณิตศาสตร์ วิทยาลัยครูสวนสุนันทา ปี พ.ศ.2518
5. จบประกาศนียบัตรการศึกษาชั้นต้น (ป.กศ.ต้น) วิทยาลัยครูสวนสุนันทา ปี พ.ศ. 2516
6. จบมัธยมศึกษาปีที่ 3 โรงเรียนวัดน้อยใน กทม. ปี พ.ศ. 2514
7. จบชั้นประถมศึกษาปีที่ 7 โรงเรียนวัดโบสถ์ จังหวัดนนทบุรี ปี พ.ศ. 2511
8. จบชั้นประถมศึกษาปีที่ 4 โรงเรียนวัดสิงห์ จังหวัดนนทบุรี ปี พ.ศ. 2508

#### 1.3 ประวัติการทำงาน

1. ครูโรงเรียนประถมศึกษา โรงเรียนบ้านบางอ่า จังหวัดนครสวรรค์ 1 พ.ค.2520-19 ต.ค.2520
2. ครูโรงเรียนมัธยมศึกษาโรงเรียนหล่มเก่าพิทยาคม จังหวัดเพชรบูรณ์ 20 ต.ค. 2520 – 19 ก.ย. 2525
3. อาจารย์วิทยาลัยครู วิทยาลัยครูเลย จังหวัดเลย 20 ก.ย. 2525 – 18 ต.ค. 2538
4. อาจารย์มหาวิทยาลัย มหาวิทยาลัยนเรศวร จังหวัดพิษณุโลก 19 ต.ค. 2538 – ปัจจุบัน

##### 1.3.1 ด้านวิชาการ/สนับสนุนการเรียนการสอน

1. เจ้าหน้าที่ทะเบียนและวัดผล โรงเรียนหล่มเก่าพิทยาคม จังหวัดเพชรบูรณ์ 1 ม.ค. 2521 - 31 พ.ค. 2523

2. ผู้ช่วยเลขานุการกลุ่มโรงเรียนมัธยมศึกษา จังหวัดเพชรบูรณ์ 1 พ.ค. 2525–19 ก.ย. 2525
3. อาจารย์ประจำฝ่ายวางแผนและพัฒนา วิทยาลัยครูเลย จังหวัดเลย 20 ม.ค. 2526 – 31 พ.ค. 2526
4. ผู้ช่วยศาสตราจารย์ วิทยาลัยครูเลย จังหวัดเลย 25 ก.ย. 2535 – 20 กันยายน 2542
5. รองศาสตราจารย์ ภาควิชาการศึกษาศาสตร์ คณะศึกษาศาสตร์ มหาวิทยาลัยนครสวรรค์ จังหวัดพิษณุโลก 21 กันยายน 2542 – ปัจจุบัน

### 1.3.2 ด้านบริหาร

1. หัวหน้าแผนกโสตทัศนศึกษา โรงเรียนห่มเกล้าพิทยาคม จังหวัดเพชรบูรณ์ 1 พ.ค. 2525 - 19 ก.ย. 2525
2. หัวหน้าศูนย์ประสานงาน กศ.บป. วิทยาลัยครูเลย จังหวัดเลย 1 มิ.ย. 2526 – 31 มี.ค. 2527
3. หัวหน้าคณะวิชาครุศาสตร์(คณบดี) วิทยาลัยครูเลย จังหวัดเลย 1 เม.ย. 2527 – 31 พ.ค. 2529
4. หัวหน้าภาควิชาทดสอบและวิจัยการศึกษา วิทยาลัยครูเลย จังหวัดเลย 1 พ.ค. 2534 – 31 พ.ค. 2534
5. หัวหน้าคณะวิชาครุศาสตร์(คณบดี) วิทยาลัยครูเลย จังหวัดเลย 1 มิ.ย. 2534 – 30 มิ.ย. 2538
6. รองคณบดีฝ่ายนโยบายและแผน คณะศึกษาศาสตร์ มหาวิทยาลัยนครสวรรค์ 25 ม.ค. 2539-2 ก.พ. 2534
7. รักษาการในตำแหน่งหัวหน้าภาควิชาพื้นฐานการศึกษา มหาวิทยาลัยนครสวรรค์ 4 เม.ย. 2539 – 30 ต.ค. 2539
8. ผู้อำนวยการศูนย์วิทยบริการจังหวัดพะเยา 1 ก.พ. 2540 – 31 พ.ค. 2540
9. รักษาการในตำแหน่งรองคณบดีฝ่ายวิชาการ คณะศึกษาศาสตร์ มหาวิทยาลัยนครสวรรค์ 15 มิ.ย. 2543 – 2 ก.พ. 2545
10. ผู้อำนวยการศูนย์วิทยบริการจังหวัดเพชรบูรณ์ 1 กุมภาพันธ์ 2544 – 19 มกราคม 2547
11. รองคณบดี คณะศึกษาศาสตร์ มหาวิทยาลัยนครสวรรค์ 20 มกราคม 2547 - ปัจจุบัน

### 1.3.3 กรรมการผู้ทรงคุณวุฒิ

1. กรรมการผู้ทรงคุณวุฒิสภาสถาบันราชภัฏเลย สถาบันราชภัฏเลย จังหวัดเลย 1 มิ.ย. 2540 – 31 พ.ค. 2542
2. ประธานคณะกรรมการสถานศึกษาขั้นพื้นฐาน โรงเรียนวิทยานุกูลนารี จ.เพชรบูรณ์ 1 พ.ค. 2543 – 1 พ.ค. 2547
3. กรรมการผู้ทรงคุณวุฒิสภาสถาบันราชภัฏกำแพงเพชร สถาบันราชภัฏกำแพงเพชร จังหวัดกำแพงเพชร 23 ก.ค. 2544 – 23 ก.ค. 2546
4. กรรมการผู้ทรงคุณวุฒิสภาวิชาการมหาวิทยาลัยราชภัฏกำแพงเพชร จังหวัดกำแพงเพชร 1 มีนาคม 2548 – ปัจจุบัน
5. กรรมการผู้ทรงคุณวุฒิของคณะกรรมการติดตามประเมินผลสำนักงานเขตพื้นที่การศึกษาจังหวัดพิษณุโลก เขต 1

6. กรรมการผู้ทรงคุณวุฒิคณะกรรมการสถานศึกษาขั้นพื้นฐาน โรงเรียนเฉลิมขวัญสตรี จ.พิษณุโลก  
1 ต.ค. 2550 - ปัจจุบัน

#### 1.3.4 กรรมการอื่น ๆ ภายในมหาวิทยาลัยนเรศวร

1. กรรมการออกข้อสอบคัดเลือกเข้าศึกษาในระดับบัณฑิตศึกษาทั้งระดับปริญญาโท และระดับปริญญาเอก
2. กรรมการออกข้อสอบประมวลความรู้ (Comprehensive Examination) หมวดวิชาพื้นฐานทางการศึกษา
3. กรรมการออกข้อสอบวัดคุณสมบัติ (Qualifying Examination) ของนิสิตระดับปริญญาเอกสาขาการวิจัยและประเมินผลการศึกษา
4. กรรมการวิจัยของคณะศึกษาศาสตร์
5. กรรมการจัดทำวารสารศึกษาศาสตร์ คณะศึกษาศาสตร์
6. กรรมการการประกันคุณภาพของมหาวิทยาลัยนเรศวร

#### 1.3.5 กรรมการอื่น ๆ ภายนอกมหาวิทยาลัยนเรศวร

1. กรรมการผู้ทรงคุณวุฒิสอบวิทยานิพนธ์ของสถาบันราชภัฏกำแพงเพชร สถาบันราชภัฏ  
พิบูลสงคราม สถาบันราชภัฏนครสวรรค์ และสถาบันราชภัฏเชียงใหม่
2. ประธานที่ปรึกษาวิทยานิพนธ์ของนิสิตสถาบันราชภัฏกำแพงเพชร และสถาบันราชภัฏนครสวรรค์
3. กรรมการที่ปรึกษาวิทยานิพนธ์ระดับปริญญาเอกของนิสิตมหาวิทยาลัยขอนแก่น  
จุฬาลงกรณ์มหาวิทยาลัย และมหาวิทยาลัยรามคำแหง
4. กรรมการประเมินภายนอกของการประเมินคุณภาพภายนอกระดับการศึกษาขั้นพื้นฐาน  
ของ สมศ.

#### 1.4 การอบรมและศึกษาดูงานต่างประเทศ

1. ศึกษาดูงานที่ Michigan State University และเข้าร่วมประชุมสัมมนา AERA ที่บอสตัน  
ประเทศสหรัฐอเมริกา 31 มีนาคม 2533 – 30 เมษายน 2533
2. ศึกษาดูงานที่ Deakin University ประเทศออสเตรเลียตามโครงการร่วมมือของมหาวิทยาลัยดีกิน  
กับสถาบันราชภัฏเลย 7 สิงหาคม 2537 – 14 สิงหาคม 2537
3. ศึกษาดูงานที่ Deakin University และการจัดการเรียนการสอนระดับมัธยมศึกษาและประถมศึกษา  
ของโรงเรียนในประเทศออสเตรเลีย โดยเป็นหัวหน้าทีม 21 – 29 สิงหาคม 2538
4. อบรมภาษาอังกฤษที่ University of Canberra ประเทศออสเตรเลียตามโครงการ Re-Engineering  
ของมหาวิทยาลัยนเรศวร 16 มีนาคม 2540 – 24 พฤษภาคม 2540
5. ศึกษาและทำวิจัยที่ศูนย์วิจัยและประเมินผลหลักสูตรและการเรียนการสอน (CIRCE) ที่ University of  
Illinois at Urbana – Champaign ประเทศสหรัฐอเมริกา 19 สิงหาคม – 9 ธันวาคม 2544

6. ศึกษาดูงาน ณ ประเทศเกาหลีใต้ 12 – 18 มกราคม 2545
7. อบรมการบริหารแบบมีส่วนร่วม และศึกษาดูงานที่ University of Canberra ประเทศออสเตรเลีย 22,28 กันยายน - 3 ตุลาคม 2546
8. ศึกษาดูงาน ณ ยุโรปตะวันตก ในโครงการพัฒนาคณาจารย์คณะศึกษาศาสตร์ 22 - 30 พฤษภาคม 2550

## 1.5 ผลงานทางวิชาการ

### 1.5.1 ตำรา

1. การประเมินโครงการทางการศึกษา เลย : รุ่งแสงการพิมพ์ 2535.
2. ทฤษฎีการวัดขั้นนำ อัดสำเนา, 2533.
3. สถิติขั้นสูงสำหรับการวิจัย กรุงเทพมหานคร : นิชินแอดเวอร์ ไทซิงกรุ๊ป, 2544.

### 1.5.2 เอกสารประกอบการสอน

1. การประเมินผลและการสร้างแบบทดสอบ อัดสำเนา, 2534.
2. ความรู้พื้นฐานเกี่ยวกับการวิจัย อัดสำเนา, 2537.

### 1.5.3 เอกสารคำสอน

1. การวิเคราะห์ตัวแปรพหุคูณ อัดสำเนา, 2542.

## 1.6 งานวิจัย

1. ผลของคำสั่งและการให้คะแนนที่ต่างกันต่อค่าความเที่ยง ความตรงและอำนาจจำแนกของแบบสอบชนิดเลือกตอบ, 2525.
2. การศึกษาความต้องการจำเป็นในการพัฒนาผู้บริหารโรงเรียนประถมศึกษา สถาบันพัฒนาผู้บริหารการศึกษา สำนักงานปลัดกระทรวง กระทรวงศึกษาธิการ, 2532. (คณะกรรมการร่วม)
2. การพัฒนาดัชนีชี้ความบกพร่องของผู้ตอบแบบทดสอบโดยการวิเคราะห์แบบแผนการตอบข้อกระทง, 2534.
3. รายงานการวิจัยเรื่อง การสังเคราะห์งานวิจัยเกี่ยวกับการฝึกอบรม กองส่งเสริมวิทยฐานะครู กรมการฝึกหัดครู พ.ศ. 2536. (คณะกรรมการร่วม)
4. การประเมินผลโครงการฝึกอบรมครูผู้สอนชั้นประถมศึกษาปีที่ 3 สังกัดสำนักงานคณะกรรมการการศึกษาแห่งชาติ และสำนักงานคณะกรรมการการศึกษาเอกชน ปีงบประมาณ 2536 กองส่งเสริมวิทยฐานะครู กรมการฝึกหัดครู อัดสำเนา, 2536 .(คณะกรรมการร่วม)
5. การวิจัยเพื่อพัฒนาหลักสูตรอบรมครูกลุ่มทักษะคณิตศาสตร์ชั้นประถมศึกษาปีที่ 3-4 ตามหลักสูตรประถมศึกษา พ.ศ. 2521 (ฉบับปรับปรุง พ.ศ. 2533) ระยะที่ 1 การศึกษาปัญหาและ

- ความต้องการการฝึกอบรม กองส่งเสริมวิทยฐานะครู กรมการฝึกหัดครู รุ่งแสงการพิมพ์, 2536.  
(คณะกรรมการร่วม)
6. การวิจัยเพื่อพัฒนาหลักสูตรอบรมครูกลุ่มทักษะคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 3-4 ตามหลักสูตรประถมศึกษา พ.ศ. 2521 (ฉบับปรับปรุง พ.ศ. 2533) ระยะที่ 2 การพัฒนาชุดฝึกอบรม กองส่งเสริมวิทยฐานะครู กรมการฝึกหัดครู รุ่งแสงการพิมพ์, 2537. (คณะกรรมการร่วม)
  7. การประเมินผลโครงการฝึกอบรมครูผู้สอนชั้นประถมศึกษาปีที่ 4 สังกัดสำนักงานคณะกรรมการการศึกษาแห่งชาติ และสำนักงานคณะกรรมการการศึกษาเอกชน ปีงบประมาณ 2537 กองส่งเสริมวิทยฐานะครู กรมการฝึกหัดครู รุ่งแสงการพิมพ์, 2537. (คณะกรรมการร่วม)
  8. การวิจัยเชิงประเมิน โครงการครุทายาทของกระทรวงศึกษาธิการ ปีงบประมาณ 2537 กองส่งเสริมวิทยฐานะครู กรมการฝึกหัดครู, 2537.
  9. ระดับการให้เหตุผลเชิงจริยธรรมของนิสิตสถาบันราชภัฏเลย สถาบันราชภัฏเลย อัดสำเนา, 2538.
  10. การศึกษาค่าใช้จ่ายในการผลิตบัณฑิต สาขาการศึกษา และค่าใช้จ่ายของนิสิตครุทายาทของสถาบันราชภัฏ กองส่งเสริมวิทยฐานะครู สถาบันราชภัฏ อัดสำเนา, 2538. (ประธานโครงการ)
  11. การประเมินผลโครงการฝึกอบรมครูผู้สอนชั้นประถมศึกษาปีที่ 5 สังกัดสำนักงานคณะกรรมการการศึกษาแห่งชาติ และสำนักงานคณะกรรมการการศึกษาเอกชน ปีงบประมาณ 2538 กองส่งเสริมวิทยฐานะครู กรมการฝึกหัดครู อัดสำเนา, 2538. (คณะกรรมการร่วม)
  12. การสังเคราะห์งานวิจัยที่ศึกษาเกี่ยวกับวิธีการเรียนวิธีการสอนและรูปแบบการเรียนการสอน ที่ส่งผลต่อกระบวนการเรียนรู้ที่ยั่งยืนของเด็กไทย วิชาคณิตศาสตร์ ระดับชั้นประถมศึกษา (พ.ศ. 2533-2541) กองวิจัยการศึกษา กรมวิชาการ 2542.
  13. การเปรียบเทียบคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ที่มีวิธีแปลงคะแนน ผลการเรียนสะสมต่างกัน 2542.
  14. การวิเคราะห์ตัวแปรจำแนกโรงเรียนมัธยมศึกษา ที่นักเรียนสามารถสอบคัดเลือกเข้ามหาวิทยาลัยได้ สูงและต่ำ. 2543.
  15. การสังเคราะห์งานวิจัยเกี่ยวกับการจัดการเรียนการสอนแบบบูรณาการ เพื่อพัฒนาการเรียนรู้ของนักเรียน ตามโครงการวิจัยและพัฒนาเพื่อปฏิรูปการศึกษาการเรียนรู้ทั้งโรงเรียน ระยะที่ 1, พฤศจิกายน 2545. (หัวหน้าโครงการ)
  16. รายงานผลการติดตามการดำเนินงานของผู้ประเมินภายนอกของการประเมินคุณภาพภายนอก ระดับการศึกษาขั้นพื้นฐาน ในภาคเหนือ, 2546.
  17. การสำรวจทัศนคติของประชาชนเกี่ยวกับการจัดการน้ำเสียในพื้นที่เทศบาลตำบลชุมแสง จังหวัดนครสวรรค์, 2546.
  18. การสังเคราะห์งานวิจัยเกี่ยวกับการจัดการเรียนการสอนแบบบูรณาการ เพื่อพัฒนาการเรียนรู้

ของนักเรียน ตามโครงการวิจัยและพัฒนาเพื่อปฏิรูปการเรียนรู้ทั้งโรงเรียน ระยะเวลาที่ 2, 2546. (หัวหน้าโครงการ)

19. การสังเคราะห์งานวิจัยเกี่ยวกับการจัดการเรียนการสอนแบบบูรณาการ เพื่อพัฒนาการเรียนรู้ของนักเรียน ตามโครงการวิจัยและพัฒนาเพื่อปฏิรูปการเรียนรู้ทั้งโรงเรียน ระยะเวลาที่ 3, 2547. (หัวหน้าโครงการ)
20. การประเมินหลักสูตรการศึกษามหาบัณฑิต สาขาวิชาการบริหารการศึกษา มหาวิทยาลัยนเรศวร, 2547.
21. การวิเคราะห์ตัวประกอบคุณลักษณะของครูรุ่นใหม่ตามการรับรู้ของนักการศึกษาชั้นนำและครูดีเด่น, 2548.
22. การประเมินการดำเนินงานตามยุทธศาสตร์การแก้ไขปัญหาผู้เสพยาเสพติด (Demand) ครั้งที่ 1, 2548. (กรรมการร่วม นักวิจัยส่วนภูมิภาค)
23. การติดตามผู้ประเมินคุณภาพภายนอกสถานศึกษาขั้นพื้นฐาน อาชีวศึกษา และอุดมศึกษา ในเขตภาคเหนือตอนล่าง ประจำปีการศึกษา 2547-2548, 2548.
24. รายงานผลการติดตามการดำเนินงานของผู้ประเมินภายนอกของการประเมินคุณภาพภายนอก ระดับการศึกษาขั้นพื้นฐาน ในภาคเหนือ ประจำปี 2549, 2549.
25. การพัฒนาตัวชี้วัดศักยภาพการจัดการกองทุนหมู่บ้านและชุมชนเมือง, 2549.
26. การประเมินโครงการศูนย์ซ่อมสร้างเพื่อชุมชนระยะที่ 1 ในเขตจังหวัดเพชรบูรณ์ พิจิตรและอุตรดิตถ์, 2549.
27. การประเมินโครงการพัฒนาผู้นำการเปลี่ยนแปลงเพื่อรองรับการกระจายอำนาจ, 2550.
28. การติดตามผู้ประเมินคุณภาพภายนอกระดับอุดมศึกษา ประจำปี 2549, 2550.
29. กลยุทธ์การยกระดับคุณภาพการศึกษาโดยเครือข่ายการวิจัย:ภาคเหนือตอนล่าง, 2550.

## 1.7 บทความ

### 1.7.1 บทความจากผลงานวิจัย

1. การเปรียบเทียบคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ที่มีวิธีแปลงคะแนนผลการเรียนสะสมต่างกัน. วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 4 ฉบับที่ 1 พฤษภาคม - สิงหาคม 2542.
2. การสังเคราะห์งานวิจัยที่ศึกษาเกี่ยวกับวิธีการเรียน การสอนและรูปแบบการเรียนการสอนที่ส่งผลต่อกระบวนการเรียนรู้ที่ยั่งยืนของเด็กไทย วิชาคณิตศาสตร์ระดับชั้นประถมศึกษา (พ.ศ.2533-2541). วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 4 ฉบับที่ 2 กันยายน - ธันวาคม 2542.

3. การวิเคราะห์ตัวแปรจำแนกโรงเรียนมัธยมศึกษาที่นักเรียนสามารถคัดเลือกเข้ามหาวิทยาลัยได้สูงและต่ำ. วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 5 ฉบับที่ 1 พฤษภาคม – สิงหาคม 2543.
4. การวิเคราะห์การจำแนกพหุ(MCA): วิเคราะห์ถดถอยพหุที่มีตัวแปรพยากรณ์เป็นตัวแปรหุ่น. รวมบทความทางวิชาการด้านการวิจัย สถิติ วัดและประเมินผลการศึกษา, กันยายน 2546.
5. การติดตามผู้ประเมินคุณภาพภายนอกสถานศึกษาขั้นพื้นฐาน อาชีวศึกษา และอุดมศึกษา ในเขตภาคเหนือตอนล่าง ประจำปี 2547 – 2548. วารสารยุทธศาสตร์การพัฒนา มหาวิทยาลัยราชภัฏกำแพงเพชร ปีที่ 1 ฉบับที่ 1 ประจำเดือนมกราคม - มิถุนายน 2549.
6. การประเมินหลักสูตรการศึกษามหาบัณฑิต สาขาวิชาการบริหารการศึกษา มหาวิทยาลัยนเรศวร. วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 8 ฉบับที่ 1 มกราคม - เมษายน 2549.
7. การพัฒนาตัวชี้วัดศักยภาพในการจัดการกองทุนหมู่บ้าน

### 1.7.2 บทความทั่วไป

1. แบบจำลองของอัลคิน รวมบทความทางการประเมินโครงการ สมหวัง พิธิยานุวัฒน์ บรรณาธิการ กทม : สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย, 2528.
2. การวิเคราะห์ข้อมูลด้วยเทคนิค Multiple Regression Analysis ในท่านถาม-เราตอบ. วารสารวิจัยสังคมศาสตร์ ปีที่ 3 ฉบับที่ 2, 2530.
3. CIPP Model เอกสารประกอบการอบรมการวิจัยเชิงประเมิน ศูนย์วิจัยและบริการศึกษาวิทยาลัยครู สกลนคร อัดสำเนา, ธันวาคม 2531.
4. แนะนำหนังสือ การวัดและประเมินผลการเรียนการสอนระดับอุดมศึกษา. วารสารวิธีวิทยาการวิจัย ปีที่ 3 ฉบับที่ 2 พฤษภาคม – สิงหาคม 2531.
5. แนะนำหนังสือ แนวคิดเชิงทฤษฎีการวิจัย การวัดและประเมินผล. วารสารวิธีวิทยาการวิจัย ปีที่ 3 ฉบับที่ 3 กันยายน – ธันวาคม 2531.
6. คำสั่งและการให้คะแนนตามความสามารถที่แท้จริง: การแก้การเดาข้อสอบชนิดเลือกตอบ. วารสารการวัดผลการศึกษา ปีที่ 10 ฉบับที่ 30 มกราคม – เมษายน 2532.
7. ดัชนีชี้วัดความบกพร่องของผู้ตอบแบบทดสอบโดยการวิเคราะห์แบบแผนการตอบข้อกระทง เอกสารการประชุมสัมมนาวิชาการการวัดผลและการวิจัยทางการศึกษาในอนาคต ครั้งที่ 5. ภาควิชา ประเมินผลและวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่ กรกฎาคม 2539.
8. การประเมินผลการปฏิบัติงาน: เทคนิคและวิธีการ. วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 1 ฉบับที่ 1 พฤษภาคม - สิงหาคม 2539.
9. การประเมินผลโครงการฝึกอบรม : ตามรูปแบบของเคริกแพททริก. วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 1 ฉบับที่ 2 กันยายน – ธันวาคม 2539.

10. การประเมินตนเอง. วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 2 ฉบับที่ 1 พฤษภาคม – สิงหาคม 2540.
11. CIPP Model: ผลกระทบของการบิดเบือนเงื่อนไขในการใช้แบบจำลอง. วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 2 ฉบับที่ 2 กันยายน – ธันวาคม 2540.
12. เทคนิคการตัดเกรด. วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ปีที่ 3 ฉบับที่ 2 กันยายน – ธันวาคม 2541.

## 1.8 การนำเสนอผลงานวิจัย

### 1.8.1 ภายในประเทศ

1. การติดตามประเมินคุณภาพภายนอกสถานศึกษาขั้นพื้นฐาน อาชีวศึกษาและอุดมศึกษา ในเขตภาคเหนือตอนล่างประจำปี 2547 – 2548. การประชุมทางวิชาการ “นเรศวรวิจัย” ครั้งที่ 2: ความสำเร็จของการพัฒนาชุดโครงการ ระหว่างวันที่ 28 – 29 กรกฎาคม 2549. มหาวิทยาลัยนเรศวร: พิษณุโลก. 2. การพัฒนาตัวชี้วัดศักยภาพการจัดการกองทุนหมู่บ้านและชุมชนเมือง.
2. การประชุมทางวิชาการ “นเรศวรวิจัย” ครั้งที่ 3: ความสำเร็จของการนำผลงานวิจัยไปใช้ประโยชน์ ระหว่างวันที่ 28 – 29 กรกฎาคม 2550. มหาวิทยาลัยนเรศวร: พิษณุโลก.

### 1.8.2 ต่างประเทศ

1. Mejang, S . (8 -11 January 2007). “Indicators for Measurement of potential in Managing the Village and Urban Community Funds in Thailand”. The Biennial Conference of the Comparative Education Society of Asia (CESA) and The Annual Conference of the Comparative Education Society of Hong Kong (CESHK) . The University of Hong Kong , Hong Kong
2. Mejang, S . (20 -23 March 2008). “The Desirable Characteristics of New Teachers as viewed by Leading Educators and Distinguished Teacher in Thailand”. The 1<sup>st</sup> East-Asian Symposium on Teacher Education Research. Hosei University , Japan.

## 2. นางประภัสสร วงษ์ดี

### 2.1 ประวัติส่วนตัว

1. ชื่อ นางประภัสสร วงษ์ดี นักศึกษาระดับปริญญาเอก สาขาวิจัยและประเมินผลการศึกษา ชั้นปีที่ 1 คณะศึกษาศาสตร์ มหาวิทยาลัยนครสวรรค์
2. เกิดเมื่อวันที่ 31 ตุลาคม พ.ศ. 2512 อายุ 39 ปี ณ บ้านเลขที่ 321 ม.10 ถ.ราชยางกูร ตำบลม่วงสามสิบ อําเภอม่วงสามสิบ จังหวัดอุบลราชธานี
3. สถานภาพครอบครัวในปัจจุบัน คู่สมรสชื่อ นายสนั่น วงษ์ดี รับราชการในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ตำแหน่งผู้อำนวยการ โรงเรียนบ้านหนองแวม อ.เมือง จ.ตาก มีบุตร 2 คน ชื่อ เด็กชายณัฐวุฒิ วงษ์ดี และเด็กหญิงณัฐวดี วงษ์ดี

### 2.2 ประวัติการศึกษา

1. กำลังศึกษาต่อระดับปริญญาเอก (กศ.ด.) การวิจัยและประเมินผลการศึกษา มหาวิทยาลัยนครสวรรค์ เริ่มศึกษาในปีการศึกษา 2551
2. จบปริญญาโท (ค.ม.) การวิจัยการศึกษา จุฬาลงกรณ์มหาวิทยาลัย ปี พ.ศ. 2540
3. จบปริญญาตรี (กศ.บ.) เอกฟิสิกส์ - คณิตศาสตร์ (เกียรตินิยม) มหาวิทยาลัยขอนแก่น ปี พ.ศ. 2534
4. จบมัธยมศึกษาปีที่ 6 โรงเรียนเบ็ญจะมะมหาราช อ.เมือง จ.อุบลราชธานี ปี พ.ศ. 2531
5. จบมัธยมศึกษาปีที่ 3 โรงเรียนม่วงสามสิบอัมพวันวิทยา อ.ม่วงสามสิบ จ.อุบลราชธานี ปี พ.ศ. 2528
6. จบประถมศึกษาปีที่ 6 โรงเรียนม่วงสามสิบอําเภอเบ็ญจะมะมหาราช อ.ม่วงสามสิบ จ.อุบลราชธานี ปี พ.ศ. 2525

### 2.3 รางวัลและเกียรติคุณที่เคยได้รับ

- |           |   |
|-----------|---|
| 2531      | ได้รับทุนการศึกษาจากมหาวิทยาลัยขอนแก่น  |
| 2532-2534 | ได้รับทุนการศึกษาของชาวสวิสเซอร์แลนด์   |
| 2532      | โล่รางวัลนักศึกษาดีเด่นจากมหาวิทยาลัยขอนแก่น  |
| 2532      | เกียรติบัตรผู้มีผลการเรียนดีเยี่ยม ประจำปีการศึกษา 2531 ของ สาขาวิชาการมัธยมศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น |
| 2533      | โล่รางวัลนักศึกษาดีเด่นจากมหาวิทยาลัยขอนแก่น  |
| 2533      | เกียรติบัตรผู้มีผลการเรียนดีเยี่ยม ประจำปีการศึกษา 2532 ของ สาขาวิชาการมัธยมศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น |
| 2533      | เกียรติบัตรผู้มีความประพฤติดีเด่น ประจำปีการศึกษา 2532 ของ  |

สาขาวิชาการมัธยมศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

## 2.4 ประวัติการทำงาน

2.4.1 ปฏิบัติงานที่ สสวท. (2 มกราคม 2535 - 30 กันยายน 2549) ดังนี้

1. ปฏิบัติงานตำแหน่งวิทยากร 1 สาขาวิจัย 2 มกราคม 2535 - 21 ตุลาคม 2541
2. ปฏิบัติงานตำแหน่งวิทยากร 1 / นักวิชาการ ที่โครงการ สรรหาและส่งเสริมนักเรียนที่มีแววเป็นอัจฉริยะทางคณิตศาสตร์ประถมศึกษา 22 ตุลาคม 2541 - 1 ธันวาคม 2543

3. ปฏิบัติงานตำแหน่งนักวิชาการ สาขาวิจัย 2 ธันวาคม 2543 - 19 กรกฎาคม 2544

4. ปฏิบัติงานตำแหน่งนักวิชาการโครงการประเมินผลนักเรียนร่วมกับนานาชาติ (โครงการ OECD/PISA หรือ Programme of International Student Assessment) 20 กรกฎาคม 2544 - 10 ตุลาคม 2548

5. ปฏิบัติงานตำแหน่งนักวิชาการ สาขาวิจัย และเป็นคณะทำงานโครงการศึกษาแนวโน้มการประเมินผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์และคณิตศาสตร์ร่วมกับนานาชาติ (TIMSS 2007) 11 ตุลาคม 2548 - 30 กันยายน 2549

2.4.2 ปฏิบัติงาน ที่วิทยาลัยสารพัดช่างตาก (4 พฤษภาคม 2550 - 30 กันยายน 2550)

สอนวิชาคณิตศาสตร์ ประยุกต์ 1 วิชาอาชีพอนามัยและความปลอดภัย(ปวช.1) และสอนการทำโครงการวิจัย ของนักศึกษาสาขางานการชาย (ปวช.2)

2.4.3 ปฏิบัติงานที่วิทยาลัยลุ่มน้ำปิง อ.เมือง จ.ตาก (18 กุมภาพันธ์ 2551 - ปัจจุบัน) ในตำแหน่ง

อาจารย์ และขณะเดียวกันก็เป็นอาจารย์พิเศษสอนที่มหาวิทยาลัยราชภัฏราชวิทยาลัย วิทยาเขตตาก โดยทำหน้าที่สอนในรายวิชาต่อไปนี้

1. สถิติธุรกิจ(ระดับปริญญาตรีคณะบริหารธุรกิจ)
2. สถิติเบื้องต้นและวิจัย(นักศึกษาระดับปริญญาตรีการจัดการเชิงพุทธ มหาวิทยาลัยราชภัฏราชวิทยาลัย)
3. คณิตศาสตร์พื้นฐาน(นักศึกษาระดับปริญญาตรีการจัดการเชิงพุทธ มหาวิทยาลัยราชภัฏราชวิทยาลัย)
4. สถิติ(ระดับประกาศนียบัตรวิชาชีพชั้นสูงของโรงเรียนธุรกิจการบัญชี วิทยาลัยลุ่มน้ำปิง)

## 2.5 ด้านวิชาการ/สนับสนุนการเรียนการสอน

1. คณะทำงานฝ่ายมาตรฐานการศึกษาและฝ่ายวิจัยของวิทยาลัยสารพัดช่างตาก 4 พ.ค. - 30 ก.ย. 2550
2. คณะทำงานวิจัยและงานมาตรฐานการศึกษา ของวิทยาลัยลุ่มน้ำปิง อ.เมือง จ.ตาก (18 ก.พ. 2551 - พฤศจิกายน 2551)

## 2.6 การเป็นกรรมการ

ปี พ.ศ. 2535

คณะทำงานสร้างแบบทดสอบความถนัดทางการเรียนเพื่อใช้คัดเลือกนักเรียนเข้าโครงการเจ้าภาพ  
ซีววิทยา โอลิมปิก ในปีการศึกษา 2538

คณะทำงานโครงการประชุมปฏิบัติการทดลองใช้รูปแบบการจัดการเรียนการสอนคณิตศาสตร์ เพื่อ  
พัฒนาผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ของนักเรียนระดับประถมศึกษา

ปี พ.ศ. 2537

คณะกรรมการฝ่ายวิจัยและประเมินผลค่ายฝึกอบรมเข้ม ครั้งที่ 1 นักเรียนโครงการจัดส่งเยาวชนไทยไป  
แข่งขันคณิตศาสตร์ วิทยาศาสตร์ โอลิมปิก ระหว่างประเทศ ประจำปี พ.ศ. 2537

คณะกรรมการดำเนินงานวิจัยการจัดการเรียนการสอนวิชาวิทยาศาสตร์และคณิตศาสตร์ ที่เหมาะสม  
กับนักเรียนในโรงเรียนขยายโอกาสทางการศึกษาขั้นพื้นฐาน

ปี พ.ศ. 2541

คณะกรรมการโครงการสรรหาและส่งเสริมนักเรียนที่มีแววเป็นอัจฉริยะทางคณิตศาสตร์ประถมศึกษา

ปี พ.ศ. 2542

คณะกรรมการดำเนินการประเมินผลการดำเนินงานของ สสวท. ปี 2542

ปี พ.ศ. 2543

เป็นคณะทำงานจัดประชุมระดมความคิดการจัดทำแผนแม่บทของ สสวท.

ปี พ.ศ. 2544

เป็นผู้ช่วยเลขานุการและคณะกรรมการประเมินผลการปฏิบัติงานของ สสวท.

ปี พ.ศ. 2548

ประธานกรรมการจัดจ้างผู้ดำเนินการประเมินผลการดำเนินโครงการวิทยาศาสตร์สิ่งแวดล้อมเพื่อ  
สิ่งแวดล้อมโลก (GLOBE Program) โดยวิธีพิเศษ

คณะกรรมการฝ่ายประเมินผลดำเนินการจัดประชุมวิชาการวิทยาศาสตร์และเทคโนโลยี “วิทยาศาสตร์  
และเทคโนโลยี นำวิถีไทย ก้าวไกลสู่สากล” ปี 2548

คณะกรรมการฝ่ายประเมินผลโครงการเทศกาลภาพยนตร์วิทยาศาสตร์เพื่อการเรียนรู้ (Science Film  
Festival) ครั้งที่ 1

คณะทำงานฝ่ายประเมินผลการจัดประชุมวิชาการวิทยาศาสตร์และเทคโนโลยี “วิทยาศาสตร์และ  
เทคโนโลยี นำวิถีไทย ก้าวไกลสู่สากล”

ปี พ.ศ. 2549

คณะทำงานโครงการศึกษาแนวโน้มการประเมินผลสัมฤทธิ์ทางการเรียนวิทยาศาสตร์และคณิตศาสตร์  
ร่วมกับนานาชาติ (TIMSS 2007)

คณะกรรมการประเมินผลการจัดประชุมวิชาการ (โดยวิธี Video Conferences) เนื่องในงานวันสถาปนา สสวท. 1 กันยายน 2549

## 2.7 การอบรมและศึกษาดูงานต่างประเทศ

1. เข้าร่วมการอบรมในหลักสูตร “Development of Evaluation Instruments for Science and Mathematics Teachers (Primary) with Emphasis on Critical and Creative Thinking” ที่ศูนย์ SEAMEO RECSAM (Regional Center for Education in Science And Mathematics) ณ ประเทศมาเลเซีย 22 พฤศจิกายน 2536 - 29 มกราคม 2537

## 2.8 ประสบการณ์ที่เกี่ยวข้องกับงานวิจัย

### หัวหน้าโครงการวิจัย /ผู้วิจัย

1. การศึกษาสภาพปัญหาและผลสำเร็จในการจัดการเรียนการสอนวิชาคณิตศาสตร์ระดับประถมศึกษา.(2541).
2. การศึกษาคุณลักษณะของครูทั่วไปและครูที่ได้รับรางวัลในด้านการจัดการเรียนการสอนวิชาวิทยาศาสตร์และคณิตศาสตร์ และศึกษาความพร้อมของครูในการจัดทำหลักสูตรตาม พ.ร.บ. การศึกษาแห่งชาติ พ.ศ. 2542.(2543).
3. การศึกษาข้อมูลพื้นฐานของนักเรียนในโครงการสรรหาและส่งเสริมนักเรียนที่มีความสามารถพิเศษทางคณิตศาสตร์ ระดับประถมศึกษา ปี พ.ศ. 2542.(2544).
4. การศึกษาสภาพการเรียนการสอนที่ส่งผลต่อทักษะการสำรวจตรวจสอบทางวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาตอนต้น.(2549).

### ผู้วิจัยร่วม(คณะวิจัย)ที่แล้วเสร็จ

1. สุณีย์ คล้ายนิล.(2547). รายงานการวิจัย โครงการประเมินผลนักเรียนนานาชาติ PISA 2000 และ PISA Plus “ความรู้และทักษะของเยาวชนไทยสำหรับโลกวันพรุ่งนี้”. (คณะวิจัย)
2. สุณีย์ คล้ายนิล.(2547). รายงานการวิจัยโครงการ PISA Thailand “ความรู้และทักษะของเยาวชนไทยสำหรับโลกวันพรุ่งนี้” ผลจากการวิจัย โครงการประเมินผลนักเรียนนานาชาติ. (คณะวิจัย)
3. สุณีย์ คล้ายนิล.(2549). รายงานโครงการประเมินผลนักเรียนนานาชาติ “การเรียนรู้เพื่อโลกวันพรุ่งนี้ รายงานการประเมินผลการเรียนรู้จาก PISA 2003”(คณะวิจัย)
4. อรุณี อ่อนสวัสดิ์ และคณะ. (2552). โครงการติดตามและประเมินผลการดำเนินงานโครงการพัฒนาศักยภาพของหมู่บ้าน / ชุมชน(SML) ตามแนวปรัชญาเศรษฐกิจพอเพียง.
5. สำราญ แดงประเสริฐ และคณะ.(2551). การศึกษาความพึงพอใจและความคิดเห็นของผู้รับบริการในจังหวัดตาก ต่อโครงการหลักประกันสุขภาพถ้วนหน้า ปี 2551.

6. สำราญ แดงประเสริฐ และคณะ.(2551). การศึกษาความพึงพอใจและความคิดเห็นของผู้ให้บริการในจังหวัดตากต่อโครงการหลักประกันสุขภาพถ้วนหน้า ปี 2551 .

### **การวิจัยในชั้นเรียน**

1. การพัฒนากระบวนการเรียนรู้ ในวิชาคณิตศาสตร์ประยุกต์ 1 เรื่อง การแปรผันของนักศึกษาระดับชั้นปวช.1 โดยใช้หลักการแก้ปัญหาการแปรผัน. (2550).
2. การพัฒนากระบวนการเรียนรู้ ในวิชาคณิตศาสตร์ประยุกต์ 1 เรื่องสมการเชิงเส้นตัวแปรเดียวของนักศึกษาระดับชั้นปวช.1 โดยใช้แบบฝึกเสริมการเรียนรู้พื้นฐานที่จำเป็นในการแก้สมการ.(2550).

### **2.9 บทความจากผลงานวิจัย**

1. (สุพร เข้มเฮง / ประภัสสร วงษ์ดี) การศึกษาสภาพปัญหาและผลสำเร็จในการจัดการเรียนการสอนวิชาคณิตศาสตร์ระดับประถมศึกษา .วารสาร สสวท. ปีที่ 27 ฉบับที่ 107 ตุลาคม – ธันวาคม 2542.

### 3. นางสาวยุพิน โกณฑทา

#### ประวัติส่วนตัว

1. นางสาวยุพิน โกณฑทา นิสิตปริญญาเอกสาขาวิจัยและประเมินผลการศึกษา มหาวิทยาลัยนเรศวร หน้าที่การงาน ตำแหน่ง อาจารย์ ระดับ 7 คณะครุศาสตร์ มหาวิทยาลัยราชภัฏเพชรบุรี
2. เกิดเมื่อวันที่ 17 มกราคม 2513 อายุ 38 ปี
3. สถานภาพครอบครัวในปัจจุบัน โสด
4. โทรศัพท์มือถือ 085-8412-860
5. e-mail : yuphinkon@yahoo.com

#### ประวัติการศึกษา

1. กำลังศึกษาระดับปริญญาเอก สาขาวิจัยและประเมินผลการศึกษา มหาวิทยาลัยนเรศวร ชั้นปีที่ 1
2. จบปริญญาโท (กศ.ม.) การวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ปี พ.ศ. 2542
3. จบปริญญาตรี (กศ.บ.) การวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ปี พ.ศ. 2535
4. จบชั้นมัธยมศึกษาปีที่ 6 สายวิทย์-คณิต จากโรงเรียนบางสะพานวิทยา อำเภอบางสะพาน จังหวัดประจวบคีรีขันธ์ พ.ศ. 2531
5. จบชั้นมัธยมศึกษาปีที่ 3 จากโรงเรียนบางสะพานวิทยา อำเภอบางสะพาน จังหวัดประจวบคีรีขันธ์ พ.ศ. 2528
6. จบชั้นประถมศึกษาปีที่ 6 โรงเรียนบ้านดอนสง่า อำเภอบางสะพาน จังหวัดประจวบคีรีขันธ์ พ.ศ. 2525

#### ประวัติการทำงาน

1. อาจารย์ 1 ระดับ 3 โรงเรียนบ้านวังน้ำเขียว อำเภอบางสะพาน จังหวัดประจวบคีรีขันธ์ พ.ศ. 2535-2543
2. ศึกษานิเทศก์ สำนักงานการประถมศึกษาอำเภอมืองประจวบคีรีขันธ์ พ.ศ. 2544-2545
3. ศึกษานิเทศก์ สำนักงานเขตพื้นที่การศึกษาประจวบคีรีขันธ์ เขต 1 พ.ศ.2546-2547
4. อาจารย์ คณะครุศาสตร์ มหาวิทยาลัยราชภัฏเพชรบุรี พ.ศ.2548 - ปัจจุบัน

#### ประวัติการวิจัย

1. หัวหน้าโครงการวิจัย : ชื่อโครงการวิจัย
  - 1.1 ยุพิน โกณฑทา และคณะ. (2544). การพัฒนาคุณลักษณะนิสัยใฝ่รู้ใฝ่เรียนของนักเรียนชั้นมัธยมศึกษาตอนต้น โรงเรียนบ้านวังน้ำเขียว

1.2 ยุพิน โภณฑา. (2544). การสร้างเครื่องมือวัดความซื่อสัตย์ด้านการลักขโมยและเจตคติต่อความซื่อสัตย์ด้านการลักขโมย

2. ผู้ร่วมวิจัย : ชื่อโครงการวิจัย

2.1 ปัญญา ทองนิต และคณะ.(2549). การพัฒนาระบบการบริหารจัดการหลักสูตรครู 5 ปี มหาวิทยาลัยราชภัฏเพชรบุรี

2.2 อรุณี อ่อนสวัสดิ์ และคณะ. (2552).โครงการติดตามและประเมินผลการดำเนินงานโครงการพัฒนาศักยภาพของหมู่บ้าน / ชุมชน(SML) ตามแนวปรัชญาเศรษฐกิจพอเพียง.