

## เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้นำเสนอเนื้อหาต่างๆ ตามลำดับดังนี้

### ตอนที่ 1 แนวคิดเกี่ยวกับการเทียบมาตรฐาน

1. ความหมายของการเทียบมาตรฐาน
2. เงื่อนไขของการเทียบมาตรฐาน
3. ประเภทของการเทียบมาตรฐาน
4. แบบแผนการเก็บรวบรวมข้อมูล
5. รูปแบบของการเทียบมาตรฐาน

### ตอนที่ 2 การเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบ

1. ประวัติความเป็นมาของทฤษฎีการตอบสนองข้อสอบ
2. ลักษณะทั่วไปของทฤษฎีการตอบสนองข้อสอบ
3. ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ
4. โมเดลตามทฤษฎีการตอบสนองข้อสอบ
5. วิธีการเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบ
6. การแปลงค่าพารามิเตอร์ให้อยู่บนสเกลเดียวกัน
7. การหาค่าคงที่สำหรับสมการเทียบมาตรฐาน

### ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

1. งานวิจัยต่างประเทศ
2. งานวิจัยภายในประเทศ

### ตอนที่ 1 แนวคิดเกี่ยวกับการเทียบมาตรฐาน

#### 1. ความหมายของการเทียบมาตรฐาน

กัลลิกเสน (Gulliksen, 1950 : 298-304, อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 17-18) ให้ความหมายของการเทียบมาตรฐานว่า เป็นวิธีการทำคะแนนที่ได้จากแบบสอบสองชุดของวิชาเดียวกันให้เป็นคะแนนสมมูลที่เปรียบเทียบกันได้โดยตรง และเสนอว่าถ้าใช้กับกลุ่มผู้สอบ

กลุ่มเดียวทำแบบสอบสองชุดซึ่งมีค่าสัมประสิทธิ์ความโค้ง และความเบ้ของการแจกแจงคล้ายคลึงกันแล้ว สามารถทำได้โดยวิธีการแปลงคะแนนแต่ละชุดให้เป็นคะแนนมาตรฐาน แล้วนำคะแนนแปลงมาเทียบกันโดยตรง

ฟลานานแกน (Flanagan, 1951 : 747-748, อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 18) ให้ความหมายของการเทียบมาตรฐานในเชิงทฤษฎีว่า เป็นวิธีการที่ทำคะแนนจากแบบสอบต่างชุดให้สามารถนำมาเปรียบเทียบกันได้ “ ความสามารถในการเปรียบเทียบกันได้ ” มีความหมายเฉพาะที่ว่า เมื่อกำหนดประชากรให้ ถ้าการแจกแจงของคะแนนจริงจากแบบสอบทั้งสองจากกลุ่มตัวอย่างที่เลือกมาขนาดใหญ่มาก มีลักษณะเหมือนกันแล้ว คะแนนดิบของแบบสอบทั้งสองชุดจึงจะสามารถเปรียบเทียบกันได้ หรือถ้าความเชื่อมั่นของแบบสอบสองชุดนั้นเท่ากันในประชากรนั้นแล้ว ก็สามารถเปรียบเทียบการแจกแจงของค่าที่ได้เช่นกัน และให้ความหมายของการเทียบมาตรฐานในทางปฏิบัติว่า ในประชากรที่กำหนด ถ้าคะแนนจากแบบสอบสองชุดมีค่าเฉลี่ยเท่ากันหรือเกือบเท่ากันในทุกๆ กลุ่มตัวอย่างขนาดใหญ่ใดๆ การเปรียบเทียบจะทำได้ อย่างไรก็ตามการจะทำให้เกิดลักษณะตามนิยามเป็นเรื่องยากมาก สิ่งสำคัญคือแบบสอบต้องเป็นคู่ขนานกัน

แองกอฟ (Angoff, 1971 : 562) ให้ความหมายของการเทียบมาตรฐานว่า เป็นการแปลงระบบของหน่วยการวัดของแบบสอบฉบับหนึ่ง ไปสู่ระบบของหน่วยการวัดของแบบสอบอีกฉบับหนึ่ง ซึ่งหลังจากการแปลงแล้วคะแนนจากแบบสอบทั้งสองฉบับสามารถเทียบคะแนนที่สมมูลกันได้

ปีเตอร์เซ็น และคณะ (Petersen and others, 1989 : 242) ให้ความหมายของการเทียบมาตรฐานโดยกล่าวว่า วิธีการเทียบมาตรฐานเป็นวิธีการเชิงประจักษ์ (Empirical Procedure) ในการสร้างความสัมพันธ์ระหว่างคะแนนของแบบสอบสองฉบับ ทำให้สามารถแปลงคะแนนของแบบสอบฉบับหนึ่ง ไปสู่คะแนนของแบบสอบอีกฉบับหนึ่งได้

มาร์โค (Marco, 1981 : 105) ให้ความหมายของการเทียบมาตรฐานว่า เป็นกระบวนการแปลงคะแนนที่ได้จากแบบสอบฉบับหนึ่ง ให้มีค่าตัวเลขสมมูล (Equivalent) กับตัวเลขที่ได้จากแบบสอบอีกฉบับหนึ่ง

ลอร์ด (Lord, 1980 : 195) ให้ความหมายของการเทียบมาตรฐานว่า เป็นการแปลงคะแนนจากแบบสอบต่างฉบับให้มีความหมายให้สับเปลี่ยนกันได้ และเพื่อความเสมอภาคของผู้เข้าสอบ

สงบ ตัณฑะ (2522 : 22) ให้ความหมายของการเทียบมาตรฐานโดยกล่าวว่า คะแนนจากแบบสอบสองฉบับวัดสิ่งเดียวกัน แต่ไม่จำเป็นต้องเป็นข้อสอบคู่ขนาน จะถือว่าเทียบเท่ากันได้ ถ้าคะแนนจากแบบสอบทั้งสองฉบับนั้นมาจากคะแนนจริง (True Score) หรือความสามารถแท้ (True Ability) ที่เท่ากัน

ซุคักคี ขัมภลิจิต (2527 : 2) ให้ความหมายของการเทียบมาตราโดยได้สรุปเกี่ยวกับการเทียบมาตราไว้ว่า เป็นกระบวนการที่เกี่ยวข้องกับกิจกรรม 2 ประการ คือ

1. กระบวนการที่ทำให้แบบทดสอบสองฉบับใดๆ มีความทัดเทียมกัน หรือเท่ากัน ในเชิงโครงสร้าง
2. การใช้วิธีการทางสถิติเพื่อปรับ (Adjust) คะแนนที่ได้จากแบบสอบแต่ละฉบับให้อยู่ในมาตราเดียวกันและเทียบกันได้

จากความหมายของการเทียบมาตราที่กล่าวมา ผู้วิจัยสรุปได้ว่า การเทียบมาตราเป็นวิธีการเชิงประจักษ์ที่ใช้วิธีการทางสถิติ ที่ทำให้คะแนนจากแบบสอบต่างฉบับที่วัดสิ่งเดียวกันสามารถนำมาเปรียบเทียบกันได้

## 2. เงื่อนไขของการเทียบมาตรา

การเทียบมาตราของแบบสอบสองฉบับจะต้องมีข้อกำหนด (Requirement) ดังนี้ (Hambleton and Swaminathan, 1985 : 199-200)

1) ความเสมอภาค (Equity) ลอร์ด (Lord, 1980 : 195) กล่าวว่า “ ถ้าการเทียบมาตราของแบบสอบฉบับ X และฉบับ Y มีความยุติธรรมต่อผู้สอบแล้ว จะต้องไม่มีความแตกต่างที่ระดับความสามารถ  $\theta$  ไม่ว่าจะสอบด้วยแบบสอบฉบับ X หรือ Y ก็ตาม ” และมีข้อกำหนดดังนี้

1.1) แบบสอบที่วัดคุณลักษณะ (Trait) หรือความสามารถ (Ability) ต่างกันจะไม่สามารถเทียบคะแนนกันได้

1.2) คะแนนแบบสอบที่มีค่าความเชื่อมั่นไม่เท่ากันแล้ว ไม่สามารถเทียบกันได้

1.3) คะแนนแบบสอบที่มีระดับความยากต่างกัน ในการเทียบมาตราในแนวตั้ง ไม่สามารถจะเทียบกันได้ (เนื่องจากคะแนนจริงของแบบสอบจะมีความสัมพันธ์กันแบบไม่เป็นเส้นตรง และทำให้แบบสอบจะมีค่าความเชื่อมั่นไม่เท่ากัน ณ ระดับความสามารถ  $\theta$  ที่แตกต่างกัน)

1.4) การแจกแจงความถี่แบบมีเงื่อนไข (The Conditional Frequency Distribution) ณ ระดับความสามารถ  $\theta$  ของคะแนนจากแบบสอบฉบับ X ( $f_{x|\theta}$ ) จะต้องเป็นเช่นเดียวกับ การแจกแจงความถี่แบบมีเงื่อนไข ณ ระดับความสามารถ  $\theta$  ของคะแนนแปลง  $X(Y) [f_{x(y)|\theta}]$  โดยที่  $X(Y)$  เป็นฟังก์ชันแบบหนึ่งต่อหนึ่งของ Y

1.5) คะแนนที่อาจผิดพลาดจากแบบสอบฉบับ X และฉบับ Y ไม่สามารถจะเทียบกันได้ นอกจากว่าแบบสอบฉบับ X และฉบับ Y คู่ขนานกัน

1.6) แบบสอบที่มีความเชื่อมั่นที่คู่สามารถเทียบกันได้

2) ความไม่แปรเปลี่ยนไปตามกลุ่ม (Invariance Across Group) คือ คะแนนของกลุ่มต่างๆ ที่ผ่านการเทียบแล้ว จะมีค่าไม่แตกต่างกันไม่ว่าจะเทียบจากกลุ่มตัวอย่างใด

3) ความสมมาตร (Symmetry) คือ ผลของการเทียบมาจากรูปแบบสอบฉบับ X ไปสู่รูปแบบสอบฉบับ Y หรือจากฉบับ Y ไปสู่ฉบับ X จะต้องเหมือนกัน

4) ความเป็นมิติเดียวกันของแบบสอบ (Unidimensionality of The Test) คือ แบบสอบทั้งสองฉบับจะต้องวัดคุณลักษณะเดียวกัน

### 3. ประเภทของการเทียบมาตรา

การเทียบมาตราสามารถจำแนกออกได้ 2 ประเภทใหญ่ (วิรัช วรรณรัตน์, 2525 : 69-70, อ้างถึงใน สงบ ลักษณะ, 2525 : 23) ดังนี้

1) การเทียบมาตราในแนวนอน (Horizontal Equating) เป็นกรณีที่มีแบบสอบ 2 ฉบับ คือ แบบสอบฉบับ ก และฉบับ ข มีความยากง่ายปานๆ กัน นำมาสอบกับบุคคลประชากรเดียวกันแล้วต้องการหาว่าที่คะแนนของแบบสอบฉบับ ก จึงเทียบเท่ากับที่คะแนนของแบบสอบฉบับ ข

2) การเทียบมาตราในแนวตั้ง (Vertical Equating) เป็นกรณีที่แบบสอบ 2 ฉบับ หรือมากกว่าสองฉบับมีความยากต่างกัน นำไปสอบประชากรบุคคลต่างระดับความสามารถกัน กล่าวคือ ต้องการเทียบคะแนนจากแบบสอบสองฉบับเมื่อความยากต่างกัน และเมื่อสอบกับบุคคลที่มีความสามารถต่างกัน การเทียบมาตราในลักษณะนี้จะกระทำโดยมีแบบสอบร่วม

### 4. แบบแผนการเก็บรวบรวมข้อมูล

ข้อมูลจากการสอบโดยใช้แบบสอบหลายฉบับ และเก็บรวบรวมข้อมูลโดยใช้วิธีดำเนินการต่างๆ กันนั้น ข้อมูลที่ได้จากบางวิธีไม่สามารถนำมาเทียบกันได้ เช่น ข้อมูลที่ได้จากการสอบโดยใช้แบบสอบ 2 ฉบับ กับผู้สอบที่มีความสามารถแตกต่างกัน 2 กลุ่ม จะนำคะแนนที่ได้จากแบบสอบต่างฉบับมาเปรียบเทียบกันไม่ได้ โดยทั่วไปคะแนนจากแบบสอบต่างฉบับที่จะนำมาเปรียบเทียบกันได้ ต้องได้มาจากผู้สอบกลุ่มเดียวกัน หรือผู้สอบกลุ่มเท่าเทียมกัน หรือได้จากผู้สอบที่สอบแบบสอบร่วมกัน วิธีใดวิธีหนึ่ง แบบแผนการเก็บรวบรวมข้อมูลที่สามารถนำคะแนนจากแบบสอบต่างฉบับมาเปรียบเทียบกันได้ จำแนกได้ดังนี้ (Petersen and others, 1989 : 244-246)

แบบแผนที่ 1 กลุ่มผู้สอบกลุ่มเดียว (Single Group Design) เป็นแบบแผนที่ง่ายที่สุดใน การเก็บรวบรวมข้อมูล โดยใช้กลุ่มตัวอย่างผู้สอบกลุ่มเดียว ทำได้โดยให้ผู้สอบแต่ละคนทำแบบสอบสองฉบับ ผู้สอบจะทำแบบสอบฉบับหนึ่งแล้วตามด้วยแบบสอบอีกฉบับหนึ่ง การให้ทำแบบสอบทั้งสองฉบับ ทำให้ไม่มีปัญหาในเรื่องความยากของแบบสอบต่างฉบับ แต่การสอบครั้งหลัง

ผู้สอบ อาจได้รับผลกระทบจากการสอบครั้งแรก เนื่องจากการเรียนรู้ ความเมื่อยล้า การฝึกปฏิบัติ และประสบการณ์

แบบแผนที่ 2 ผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล (Counterbalanced Random Group Design) ในทางปฏิบัติลำดับของการสอบก่อนหลังย่อมส่งผลกระทบต่อผลการสอบ ดังนั้นแบบแผนนี้จึงปรับมาจากแบบแผนแรกโดยจัดผู้สอบเป็นกลุ่มสุ่ม 2 กลุ่ม แต่ละกลุ่มย่อยได้รับการสอบด้วยแบบสอบ 2 ฉบับ ในลักษณะการจัดลำดับก่อนหลัง ผลิตให้เกิดความสมดุล การจัดแบบนี้เพื่อให้การสอบโดยใช้แบบสอบสองฉบับ ได้รับอิทธิพลจากองค์ประกอบในเรื่องการเรียนรู้ ความเมื่อยล้า และการฝึกปฏิบัติในลักษณะที่สมดุลกัน

แบบแผนที่ 3 ผู้สอบกลุ่มเท่าเทียมกัน (Equivalent Group Design) ในทางปฏิบัติการให้ผู้สอบทำแบบสอบทั้ง 2 ฉบับ อาจทำได้ยาก การใช้ผู้สอบกลุ่มเท่าเทียมกันเป็นวิธีการที่สามารถนำมาใช้ได้ แบบแผนนี้จำเป็นต้องจัดกลุ่มผู้สอบให้มีความคล้ายคลึงกันมากที่สุดเท่าที่จะทำได้ ผู้สอบในแต่ละกลุ่มจะทำแบบสอบฉบับเดียว ดังนั้นจึงหลีกเลี่ยงปัญหาเรื่องการเรียนรู้ ความเมื่อยล้า และการฝึกฝนจากการสอบครั้งแรกได้ แต่จะไม่มีข้อมูลที่จะนำมาใช้ปรับความแตกต่างของกลุ่มตัวอย่างทั้งสอง วิธีการที่จะลดความแตกต่างระหว่างกลุ่มทำได้โดยใช้กลุ่มตัวอย่างขนาดใหญ่

แบบแผนที่ 4 ผู้สอบกลุ่มสุ่มสองกลุ่ม โดยมีแบบสอบร่วม (Anchor Test Random Group Design) การใช้แบบแผนนี้อาจมีข้อจำกัดในเรื่องความแตกต่างระหว่างกลุ่มตัวอย่าง การใช้แบบสอบร่วมด้วย ทำให้สามารถปรับความแตกต่างระหว่างกลุ่มได้ ตามแบบแผนนี้กลุ่มสุ่มแต่ละกลุ่มจะได้รับการสอบกลุ่มละ 1 ฉบับ และทั้งสองกลุ่มจะได้รับการสอบด้วยแบบสอบร่วมด้วย ซึ่งอาจจัดไว้ในแบบสอบเทียบมาตราเรียกว่า แบบสอบร่วมภายใน (Internal Anchor Test) หรือจัดแยกไว้จากแบบสอบเทียบมาตราก็ได้ เรียกว่า แบบสอบร่วมภายนอก (External Anchor Test)

แบบแผนที่ 5 กลุ่มผู้สอบที่ไม่เท่าเทียมกัน โดยมีแบบสอบร่วม (Anchor Test Non-Equivalent Group Design) ในการสอบบางครั้งจำเป็นต้องใช้แบบสอบฉบับใหม่กับผู้สอบกลุ่มใหม่ ในโปรแกรมการสอบเดียวกัน การสอบโดยใช้แบบสอบต่างฉบับ และเวลาสอบต่างกันนี้ ผู้สอบไม่ได้มาจากกลุ่มประชากรเดียวกัน จึงจำเป็นต้องใช้แบบสอบร่วมในการปรับความแตกต่างระหว่างกลุ่ม ตามแบบแผนนี้ผู้สอบในแต่ละกลุ่มจะได้รับแบบสอบ 1 ฉบับ และแบบสอบร่วม ซึ่งต้องมีเนื้อหาและระดับความยากคล้ายกับแบบสอบเทียบมาตราให้มากที่สุดเท่าที่จะเป็นไปได้ คะแนนจากแบบสอบร่วมจะใช้เพื่อลดความลำเอียงในการเทียบมาตราจากกลุ่มที่ต่างกัน

จากแบบแผนการเก็บรวบรวมข้อมูลที่กล่าวมา แสดงให้เห็นดังตารางต่อไปนี้

ตาราง 1 แบบแผนที่ 1 Single Group Design

กลุ่มตัวอย่าง	แบบสอบฉบับที่ 1	แบบสอบฉบับที่ 2
$P_1$	/	/

ตาราง 2 แบบแผนที่ 2 Counterbalanced Random Group Design

กลุ่มตัวอย่าง	แบบสอบฉบับที่ 1		แบบสอบฉบับที่ 2	
	สอบครั้งที่ 1	สอบครั้งที่ 2	สอบครั้งที่ 1	สอบครั้งที่ 2
$P_1$	/			/
$P_2$		/	/	

ตาราง 3 แบบแผนที่ 3 Equivalent Group Design

กลุ่มตัวอย่าง	แบบสอบฉบับที่ 1	แบบสอบฉบับที่ 2
$P_1$	/	
$P_2$		/

ตาราง 4 แบบแผนที่ 4 Anchor Test Random Group Design

กลุ่มตัวอย่าง	แบบสอบฉบับที่ 1	แบบสอบฉบับที่ 2	แบบสอบร่วม
$P_1$	/		/
$P_2$		/	/

ตาราง 5 แบบแผนที่ 5 Anchor Test Non-Equivalent Group Design

กลุ่มตัวอย่าง	แบบสอบฉบับที่ 1	แบบสอบฉบับที่ 2	แบบสอบรวม
$P_1$	/		/
$Q_1$		/	/

5. รูปแบบของการเทียบมาตรา

รูปแบบของการเทียบมาตราเมื่อพิจารณาโดยใช้ทฤษฎีการวัดเป็นเกณฑ์ สามารถแบ่งได้เป็น 2 รูปแบบ คือ การเทียบมาตราตามทฤษฎีคั้งเดิม (Traditional Equating Model) และการเทียบมาตราตามทฤษฎีการตอบสนองข้อสอบ (Item Response Equating Model) แต่ละรูปแบบกล่าวโดยสรุปได้ดังนี้ (ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 20-25)

5.1 การเทียบมาตราตามทฤษฎีคั้งเดิม (Traditional Equating Model)

สามารถแบ่งออกได้เป็นการเทียบมาตราวิธีอควิเปอร์เซ็นไทล์ และการเทียบมาตราเชิงเส้นตรง หลักการแต่ละรูปแบบกล่าวโดยย่อดังนี้

5.1.1 การเทียบมาตราวิธีอควิเปอร์เซ็นไทล์ (Equipercentile Equating) การเทียบมาตราตามวิธีการนี้มีนิยามว่า สำหรับกลุ่มผู้สอบที่กำหนด คะแนนที่ได้จากแบบสอบต่างฉบับกันจะถือว่าเทียบเท่ากัน ถ้าต่างก็ตรงกับคะแนน ณ ตำแหน่งเปอร์เซ็นไทล์เดียวกัน ผลการเทียบมาตราแสดงด้วยกราฟ การเทียบมาตราวิธีอควิเปอร์เซ็นไทล์ประกอบด้วย 2 ขั้นตอน คือ

ขั้นที่ 1 การกระจายความถี่สะสมสัมพัทธ์เป็นตารางหรือเป็นกราฟ สำหรับแบบสอบที่นำมาเทียบมาตรา คือ

- 1) นำคะแนนของกลุ่มผู้สอบที่มีความสามารถทั้งเก่ง ปานกลาง และอ่อน ซึ่งสุ่มกลุ่มตัวอย่างออกเป็น 2 กลุ่ม ให้กลุ่มหนึ่งทำแบบสอบฉบับ X อีกกลุ่มหนึ่งทำแบบสอบฉบับ Y มาแจกแจงคะแนน X และคะแนน Y
- 2) คำนวณหาจุดกลางเปอร์เซ็นไทล์ของแต่ละการแจกแจง
- 3) อ่านและทำเครื่องหมายสำหรับคะแนนของแบบสอบฉบับ X และฉบับ Y ของการแจกแจงความถี่ที่สมมูลกันลงบนกระดาษ Arithmetic Graph Paper โดยแกน X เป็นคะแนนดิบ แกน Y เป็นตำแหน่งเปอร์เซ็นไทล์ ประมาณ 30 จุด ลากเส้นเชื่อมต่อกันเป็นเส้นกราฟ

ขั้นที่ 2 เทียบมาตราคะแนนจากรายละเอียดในข้อ 3 ของขั้นที่ 1 นำมาพล็อตกราฟใหม่ลงบนกระดาษ โดยแกน X เป็นคะแนนของแบบสอบฉบับ X และแกน Y เป็นคะแนนของแบบ

สอบฉบับ Y ปรับเส้นกราฟให้เรียบ เส้นกราฟที่ได้จะใช้อ่านค่าคะแนน X ที่สมมูลกันกับคะแนน Y ซึ่งสามารถนำไปสร้างตารางคะแนนสมมูลกัน

การเทียบมาตรฐานวิธีอิกวิเปอร์เซ็นไทล์ในขั้นตอนที่ 2 อาจมีปัญหา เนื่องจากอาจจะไม่มีคะแนนของแบบสอบฉบับหนึ่งตรงกับคะแนนแบบสอบอีกฉบับหนึ่ง ฉะนั้นตำแหน่งเปอร์เซ็นไทล์เดียวกัน ดังนั้นเพื่อให้สามารถเทียบกันได้ด้วยวิธีอิกวิเปอร์เซ็นไทล์ จึงจำเป็นต้องปรับการแจกแจงคะแนนของแบบสอบให้มีลักษณะต่อเนื่อง

5.1.2 การเทียบมาตรฐานเชิงเส้นตรง (Linear Equating) การเทียบมาตรฐานวิธีการนี้มีนิยามว่า สำหรับกลุ่มผู้สอบที่กำหนด คะแนนที่ได้จากแบบสอบต่างฉบับที่วัดในสิ่งเดียวกันจะเทียบเท่ากันได้ ถ้าต่างก็ตรงกับคะแนนมาตรฐานค่าเดียวกัน วิธีเทียบมาตรฐานเชิงเส้นตรงประกอบด้วย 2 ขั้นตอน ได้แก่

ขั้นที่ 1 คำนวณค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของคะแนนที่ได้จากแบบสอบทั้งสองฉบับของกลุ่มตัวอย่าง

ขั้นที่ 2 เทียบมาตรฐานคะแนนของแบบสอบฉบับหนึ่งไปยังแบบสอบอีกฉบับหนึ่งตามสมการแปลงเชิงเส้นตรงดังนี้

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

เมื่อ	X	คือ	คะแนนของแบบสอบฉบับ X
	Y	คือ	คะแนนของแบบสอบฉบับ Y
	$M_y$	คือ	ค่าเฉลี่ยของแบบสอบฉบับ Y
	$M_x$	คือ	ค่าเฉลี่ยของแบบสอบฉบับ X
	$S_y$	คือ	ส่วนเบี่ยงเบนมาตรฐานของแบบสอบ Y
	$S_x$	คือ	ส่วนเบี่ยงเบนมาตรฐานของแบบสอบ X

การเทียบมาตรฐานเชิงเส้นตรงสามารถนำไปใช้กับข้อมูลการสอบที่ได้จากแบบแผนการเก็บรวบรวมข้อมูลดังกล่าวแล้ว ซึ่งทำให้เกิดรูปแบบเฉพาะของวิธีการ เช่น แบบจำลองของทักเกอร์ และแบบจำลองของเลวิน เป็นต้น

## 5.2 การเทียบมาตรฐานทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Equating)

การเทียบมาตรฐานทฤษฎีการตอบสนองข้อสอบอาศัยรูปแบบความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้ตอบข้อสอบใดๆ ในแบบสอบว่าขึ้นอยู่กับความสามารถของผู้สอบ ซึ่งการประมาณค่าได้จากโค้งลักษณะเฉพาะของข้อสอบ (Item Characteristic Curve : ICC)



ของแต่ละรูปแบบที่ใช้ ซึ่งอาจจะเป็นรูปแบบหนึ่ง สอง หรือสามพารามิเตอร์ ทั้งนี้ไม่ขึ้นกับการแจกแจงของคะแนนของกลุ่มตัวอย่าง (สุจินดา ผ่องอักษร, 2533 : 17-18 , อ้างจาก Lord and Novick, 1968) การเทียบมาตราโดยใช้ทฤษฎีการตอบสนองข้อสอบมีหลักการว่า คะแนนจากแบบสอบต่างฉบับที่วัดในสิ่งเดียวกัน จะถือว่าเทียบกันได้ถ้าผู้สอบทำแบบสอบต่างฉบับแล้วได้คะแนนจริง (True Scores) หรือความสามารถแท้ (Ability) เท่ากัน (สงวน ลักษณะ, 2525 : 22)

การเทียบมาตราตามทฤษฎีการตอบสนองข้อสอบ แบ่งออกเป็น 2 วิธี ได้แก่

- 1) การเทียบโดยใช้คะแนนจริง (True Score Equating)
- 2) การเทียบโดยใช้คะแนนดิบ (Raw Score Equating)

การเทียบมาตราตามทฤษฎีการตอบสนองข้อสอบทั้งสองวิธี มีข้อได้เปรียบกัน คือ วิธีการเทียบโดยใช้คะแนนจริงไม่สามารถอธิบายคะแนนที่อยู่ต่ำกว่าระดับการคาดได้ โดยจะให้ความหมายของคะแนนสมมูลเฉพาะคะแนนที่อยู่เหนือค่าเฉลี่ยของการคาด ซึ่งแม้จะเป็นการเทียบโดยใช้คะแนนจริงก็ยังเป็นเพียงค่าประมาณที่ได้จากสูตรการคำนวณ จึงยังคงมีค่าความคลาดเคลื่อนอยู่ ส่วนการเทียบโดยใช้คะแนนดิบเป็นการเทียบโดยประมาณ แต่สามารถอธิบายคะแนนสมมูลจากแบบสอบทั้งสองฉบับได้ครอบคลุมพิสัยของคะแนนดิบ

สำหรับการวิจัยครั้งนี้ ผู้วิจัยเลือกการเทียบมาตราตามทฤษฎีการตอบสนองข้อสอบโดยใช้วิธีการเทียบความสามารถ แทนการใช้คะแนนจริง โดยใช้แบบสอบร่วมแล้วแปลงค่าความสามารถให้อยู่ในสเกลเดียวกัน

## **ตอนที่ 2 การเทียบมาตราตามทฤษฎีการตอบสนองข้อสอบ**

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีการวัดที่ถูกคิดขึ้นมา โดยมีวัตถุประสงค์เพื่อเสนอหลักการพื้นฐานที่สำคัญสำหรับการทำนาย ประเมินค่า และสรุปอ้างอิงเกี่ยวกับความสามารถหรือคุณลักษณะที่สามารถวัดได้ด้วยแบบสอบ (Hambleton and Swaminathan, 1985 : 53) เพื่อให้เกิดความชัดเจนเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ ผู้วิจัยจึงเสนอรายละเอียดเป็นค่านๆ ก่อนที่จะกล่าวถึงการเทียบมาตราตามทฤษฎีการตอบสนองข้อสอบ ดังนี้

## 1. ประวัติความเป็นมาของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบมีประวัติความเป็นมาเรียงตามลำดับดังนี้ (Hambleton and Swaminathan, 1985 : 4-5)

- ค.ศ. 1916 บิเน็ตและ ซามอน (Binet and Simon) ได้กำหนดระดับผลการปฏิบัติ (Performance levels) ของผู้สอบตามตัวแปรอิสระต่างๆ และใช้ผลการปฏิบัติดังกล่าวไปพัฒนาแบบสอบ
- ค.ศ. 1936 ริชาร์ดสัน (Richardson) ได้หาความสัมพันธ์ระหว่างค่าพารามิเตอร์ของข้อสอบตาม โมเดลทฤษฎีการตอบสนองข้อสอบและทฤษฎีการวัดแบบดั้งเดิม
- ค.ศ. 1943-1944 ลอว์ลีย์ (Lawley) ได้คิดวิธีการใหม่ๆ เพื่อใช้สำหรับการประมาณค่าพารามิเตอร์
- ค.ศ. 1952 ลอร์ด (Lord) ได้นำเสนอโมเดลนอร์มอลโอจีฟ (Normal Ogive Model) แบบสองพารามิเตอร์โดยนำเสนอวิธีการประมาณค่าพารามิเตอร์และการนำโมเดลไปประยุกต์ใช้
- ค.ศ. 1957-1958 เบอรัมบอม (Birnbaum) ได้นำเสนอโมเดลโลจิสติก (Logistic Model) เพื่อใช้แทนโมเดลนอร์มอลโอจีฟ
- ค.ศ. 1960 ราสช์ (Rasch) ได้พัฒนาโมเดลการตอบสนองข้อสอบ 3 โมเดล โดยพรรณนาไว้ในหนังสือ Probabilistic Model for some Intelligence and Attainment Tests ซึ่งผลงานชุดนี้มีอิทธิพลต่อไรท์ (Wright) ซึ่งอยู่ในอเมริกาและนักจิตวิทยาบางคน เช่น แอนเดอร์เซนและฟิชเชอร์ (Andersen and Fisher) ในยุโรป
- ค.ศ. 1967 ไรท์ (Wright) ได้เป็นผู้นำและกระตุ้นให้มีการวิจัยในอเมริกาเกี่ยวกับโมเดลของราสช์ (Rasch) ตลอดช่วงทศวรรษ 1970
- ค.ศ. 1968 ลอร์ดและ โนวิก (Lord and Novick) ได้นำเสนอทฤษฎีคุณลักษณะแฝง (Theory of Latent Trait)
- ค.ศ. 1969 ไรท์และปันชาปากีสาน (Wright and Panchapakesan) ได้พรรณนาวิธีการประมาณค่าพารามิเตอร์ใน โมเดลของราสช์โดยใช้โปรแกรม BICAL และในปีเดียวกันนี้ เซมีจิม่า (Samejima) ได้พรรณนาโมเดลการตอบสนองข้อสอบโมเดลใหม่ พร้อมทั้งประยุกต์ใช้ โมเดลนี้กับข้อมูลที่ได้จากการตอบหลายแบบและแบบต่อเนื่อง (Polychotomous

- and Contineous) และปรับปรุง โมเดลมิติเดียว (Unidimension) ไปสู่ โมเดลแบบหลายมิติ (Multidimension)
- ค.ศ. 1972 บอกซ์ (Bock) ได้สนับสนุนแนวคิดใหม่ๆ หลายแนวคิดเกี่ยวกับการ ประมาณค่าพารามิเตอร์
- ค.ศ. 1974 ลอร์ด (Lord) ได้พรรณนาวิธีการประมาณค่าพารามิเตอร์วิธีใหม่ โดยใช้โปรแกรมโลจิส (LOGIST) และฟิชเชอร์ (Fisher) ได้นำเสนอโปรแกรมการวิจัยเกี่ยวกับ โมเดลโลจิสติกเชิงเส้นตรง (Linear Logistic Model)
- ค.ศ. 1976 ลอร์ด (Lord) ได้พัฒนาโปรแกรม โลจิสเพื่อให้สามารถประมาณ พารามิเตอร์ใน โมเดลโลจิสติก ได้ดียิ่งขึ้น
- ค.ศ. 1977 เบเกอร์ (Baker) นำเสนอผลการทบทวนวิธีการประมาณค่า พารามิเตอร์และ ได้ตีพิมพ์ประเด็นต่างๆ เป็นกรณีพิเศษเกี่ยวกับการใช้ ทฤษฎีการตอบสนองข้อสอบในการวัดผลในวารสาร Journal of Educational Measurement ซึ่งเขียน โดยนักวิจัยหลายคน เช่น บาชอร์ (Bashaw), ลอร์ด (Lord), มาร์โค (Marco), เรนซ์ (Rentz) ยูรี (Urry) และไรท์ (Wright) เป็นต้น
- ค.ศ. 1979 ไรท์และสโตน (Wright and Stone) ได้เขียนหนังสือเกี่ยวกับแบบแผน ของแบบสอบที่ดีที่สุด (Best Test Design) โดยใช้โมเดลของราทซ์ (Rach)
- ค.ศ. 1980 ลอร์ด (Lord) ได้ทบทวนพัฒนาการของทฤษฎีการตอบสนองข้อสอบ และการประยุกต์ใช้โมเดลตามพารามิเตอร์ โดยเขียนในหนังสือ Application of Item Response Theory to Practical Testing Problems
- ค.ศ. 1982 ลอร์ดและคณะ (Lord and Staff) ได้ปรับปรุงโปรแกรมโลจิส ครั้งที่ 2 เพื่อให้ใช้ได้สะดวกกว่าเดิม

## 2. ลักษณะทั่วไปของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่แสดงความสัมพันธ์ระหว่างการตอบข้อสอบกับคุณลักษณะที่ข้อสอบดังกล่าวต้องการจะวัด ทฤษฎีนี้ได้เกิดขึ้นท่ามกลางข้อจำกัดของทฤษฎีการวัดแบบดั้งเดิมหลายประการ (Hambleton and Swaminathan, 1985 : 2) คือ

1) ค่าสถิติประจำตัวของข้อสอบ เช่น ค่าความยากง่าย ค่าอำนาจจำแนก นั้นจะขึ้นอยู่กับลักษณะของกลุ่มผู้สอบ กล่าวคือ ถ้าผู้สอบมีความสามารถสูง ข้อสอบนั้นก็ง่าย แต่ถ้าผู้สอบมีความสามารถต่ำ ข้อสอบนั้นก็ยาก ทั้งๆที่เป็นแบบสอบฉบับเดิม ส่วนค่าอำนาจจำแนกก็จะขึ้นอยู่กับความเป็นเอกพันธ์ของกลุ่มผู้สอบ ถ้าผู้สอบมีความสามารถแตกต่างกันมากๆ ข้อสอบก็จะมีค่าอำนาจจำแนกสูง ซึ่งมีผลทำให้ค่าความเชื่อมั่นของแบบสอบมีค่าสูงตามไปด้วย ทั้งนี้เนื่องจากค่าความเชื่อมั่นของแบบสอบมีความสัมพันธ์โดยตรงกับค่าความแปรปรวนของคะแนน

2) การเปรียบเทียบความสามารถของผู้สอบนั้น จำเป็นต้องใช้แบบสอบฉบับเดียวกันหรือแบบสอบคู่ขนานกันและค่าความเชื่อมั่นของแบบสอบถูกนิยามในรูปของผลที่ได้จากการใช้แบบสอบคู่ขนาน ซึ่งในการปฏิบัติจริงๆ แล้วนับว่าเป็นเรื่องที่ยากมากที่จะให้การสอบทั้งสองครั้งมีสภาพเหมือนกันทุกประการ ถึงแม้ว่าแบบสอบทั้งสองจะคู่ขนานกันจริง เพราะผู้สอบอาจมีลักษณะที่ต่างไปจากการสอบครั้งแรก เช่น แรงจูงใจ ความกังวล หรือการพัฒนาตนเองในบางทักษะ

3) ทฤษฎีการวัดแบบดั้งเดิมไม่สามารถบอกได้ว่าผู้สอบจะตอบข้อสอบได้ผลอย่างไร ยกเว้นว่าจะได้ใช้ข้อสอบนั้นกับผู้สอบที่มีลักษณะคล้ายคลึงกันมาก่อนแล้ว

4) แบบสอบวัดผลสัมฤทธิ์และแบบสอบวัดความถนัดนั้น ส่วนใหญ่แล้วจะเหมาะสมกับผู้สอบที่มีความสามารถปานกลาง ดังนั้นความถูกต้องแม่นยำของการวัดผู้สอบที่มีความสามารถสูงและผู้สอบที่มีความสามารถต่ำจึงลดลง

5) ทฤษฎีการวัดแบบดั้งเดิมใช้ค่าความแปรปรวนของความคลาดเคลื่อนในการวัดเหมือนกันกับผู้สอบทุกคน ในความเป็นจริงแล้วผู้สอบที่มีความสามารถสูงและต่ำจะมีค่าความแปรปรวนดังกล่าวต่างจากผู้สอบที่มีความสามารถปานกลาง

นอกจากที่กล่าวแล้ว ทฤษฎีการวัดแบบดั้งเดิมยังหาคำตอบของปัญหาเกี่ยวกับการสอบไม่ชัดเจนอีกหลายประการ เช่น แบบแผนการสอบ (Design of Test) ความลำเอียงของข้อสอบ (Item Bias) เป็นต้น ทฤษฎีการตอบสนองข้อสอบมีชื่อเรียกหลายชื่อ เช่น ทฤษฎีคุณลักษณะแฝง (Latent Trait Theory) ทฤษฎีโค้งคุณลักษณะข้อสอบ (Item Characteristic Curve Theory) แต่ชื่อที่ค่อนข้างได้รับความนิยม คือ ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT)

### 3. ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ มีข้อตกลงเบื้องต้น ดังนี้

1) ข้อตกลงเกี่ยวกับมิติ (Dimensionality) ทฤษฎีการตอบสนองข้อสอบมีความเชื่อว่า ผู้สอบนั้นมีความสามารถ  $k$  อย่าง ซึ่งความสามารถแต่ละอย่างก็จะตอบสนองต่อสิ่งเร้าต่างๆ กันไป แบบสอบที่จะนำมาวิเคราะห์ด้วยทฤษฎีการตอบสนองข้อสอบนั้น มีข้อตกลงว่าต้องมีความเป็น มิติเดียว (Unidimension) นั่นคือ เป็นแบบสอบที่ประกอบด้วยข้อสอบหลายๆ ข้อสำหรับวัดความสามารถด้านใดด้านหนึ่งของผู้สอบ (สำเร็จ บุญเรืองรัตน์, 2529 : 43) สิ่งที่ยอมรับเป็นดัชนีระบุความเป็นมิติเดียว ก็คือ องค์ประกอบที่เด่นชัด หรือมีความสำคัญต่อการอธิบายผลการตอบแบบสอบ สำหรับการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับความเป็นมิติเดียวนั้น ก็เพื่อประโยชน์ในการตีความหมายของคะแนนที่ได้จากการสอบคุณสมบัติ (Lumsden, 1961, อ้างถึงใน รังสรรค์ มณีเล็ก, 2540 : 42) ได้แนะนำว่า การที่จะได้ข้อสอบที่วัดมิติเดียวนั้นต้องอาศัยวิธีการคัดเลือกข้อสอบและวิธีการตรวจสอบที่ดี ซึ่งวิธีการตรวจสอบความเป็นมิติเดียวนั้นมีหลายวิธีด้วยกัน แต่วิธีที่ได้รับความนิยมมาก ได้แก่ วิเคราะห์องค์ประกอบ (Factor Analysis) โดยใช้อัตราส่วนของความแปรปรวนขององค์ประกอบแรกกับความแปรปรวนขององค์ประกอบสอง เป็นดัชนีความเป็นมิติเดียว (Unidimension) แต่ปัญหาที่พบในการวิเคราะห์องค์ประกอบคือ การเลือกใช้ค่าสหสัมพันธ์ฟี (Phi Correlation) และค่าสหสัมพันธ์เตตราคลอริก (Tetrachoric Correlation) ซึ่งการวิเคราะห์โดยใช้สหสัมพันธ์นั้นทำให้ได้องค์ประกอบหลายองค์ประกอบ และอาจทำให้ได้ องค์ประกอบที่ไม่ใช่องค์ประกอบที่แท้จริง ซึ่งแมคโดนัลและอฮลาวัต (McDonald and Ahlawat, 1974) ชี้ให้เห็นว่า ที่เกิดผลเช่นนี้ก็เพราะข้อมูลมีความสัมพันธ์กันไม่เป็นเส้นตรง (Nonlinearity) ส่วนการใช้ค่าสหสัมพันธ์เตตราคลอริกในการวิเคราะห์นั้นค่อนข้างยุ่งยากในการคำนวณ ไม่เหมาะสมที่จะนำมาใช้ เมื่อความสามารถของผู้สอบไม่แจ่มแจ้งเป็นปกติ ฟังก์ชันของการตอบสนองข้อสอบไม่เป็นแบบปกติสะสม และเมื่อมีการเดาของผู้สอบ แต่สิ่งที่ดีของการใช้ค่าสหสัมพันธ์เตตราคลอริก ก็คือ ผลของการวิเคราะห์จะได้องค์ประกอบที่เด่นชัดเพียงองค์ประกอบเดียว ซึ่งเป็นเงื่อนไขที่เพียงพอ ในการตรวจสอบความเป็นมิติเดียวของกลุ่มข้อสอบที่ถูกวิเคราะห์ อย่างไรก็ตามบางครั้ง การประเมินความเป็นมิติเดียวด้วยวิธีวิเคราะห์องค์ประกอบก็ได้ผล ไม่เป็นที่น่าพอใจ เพราะการวิเคราะห์องค์ประกอบเป็นวิธีการที่ดำเนินการกับข้อมูลที่สัมพันธ์กันเชิงเส้นตรง ดังนั้นหากข้อมูลไม่สัมพันธ์กันเชิงเส้นตรง ผลของการวิเคราะห์ก็อาจไม่ได้มิติเดียว ถึงแม้กลุ่มข้อสอบจะมีความเป็นมิติเดียวก็ตาม และส่วนใหญ่แล้วข้อมูลที่เหมาะ (Fit) กับ โมเดลของทฤษฎีการตอบสนองข้อสอบก็มีลักษณะไม่สัมพันธ์กันเป็นเส้นตรง แบบสอบที่น่าจะมีความเป็นมิติเดียว เช่น แบบสอบการสะกดคำ คำศัพท์ การอ่านเอาเรื่อง เหตุผลเชิงคณิตศาสตร์ อุปมาอุปไมย อนุกรม

ตัวเลข และมิติสัมพันธ์ ส่วนแบบสอบที่ไม่เป็นมิติเดียว เช่น แบบสอบวัดผลสัมฤทธิ์วิชาเคมี ซึ่งต้องอาศัยความสามารถหลายอย่าง ทั้งที่เกี่ยวและไม่เกี่ยวกับคณิตศาสตร์ในการตอบข้อสอบ

2) ข้อตกลงเกี่ยวกับความเป็นอิสระของข้อสอบ (Local Independence) ข้อตกลงข้อนี้กล่าวว่า โอกาสในการตอบข้อสอบข้อใดข้อหนึ่ง ได้ถูกต้องนั้น ไม่เกี่ยวข้องกับการตอบข้อสอบข้ออื่นๆ เขียนเป็นสมการทางคณิตศาสตร์ได้ดังนี้

$$\text{Prob. } [U_1=u_1, U_2=u_2, \dots, U_n=u_n] = \prod_{g=1}^n P_g^u Q_g^{1-u}$$

เมื่อ  $U_g$  แทน ผลการตอบข้อสอบข้อ  $g$

$P_g$  แทน ความน่าจะเป็นที่จะตอบข้อสอบข้อ  $g$  ถูกต้อง

$Q_g$  แทน ความน่าจะเป็นที่จะตอบข้อสอบข้อ  $g$  ผิด

จากสมการนี้กล่าวได้ว่า ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องทุกข้อเท่ากับผลคูณของความน่าจะเป็นในการตอบข้อสอบในแต่ละข้อได้ถูกต้อง การที่จะตรวจข้อสอบแต่ละข้อว่าเป็นไปตามข้อตกลงเบื้องต้นข้อนี้หรือไม่ ก็สามารถตรวจสอบได้จากแบบแผนของการตอบข้อสอบของผู้สอบที่มีความสามารถเท่าๆกันว่า ข้อสอบแต่ละคู่ เช่น ข้อ 1 กับข้อ 2 ข้อ 2 กับข้อ 3 มีลักษณะของการตอบแตกต่างกันหรือไม่ ซึ่งสามารถทดสอบความเป็นอิสระของการตอบข้อสอบเหล่านี้ได้โดยใช้สถิติไคสแควร์ (Chi-Square) (สำเร็จ บุญเรืองรัตน์, 2529 : 46) และสิ่งที่น่าสนใจเป็นอย่างยิ่งคือ ความเป็นอิสระของข้อสอบนี้จะได้มาโดยอัตโนมัติ ถ้าข้อสอบมีความเป็นมิติเดียว

#### 4. โมเดลตามทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างปริมาณของการกระทำหรือคะแนนของผู้สอบกับปริมาณความสามารถของผู้สอบ ซึ่งแสดงได้ดังสมการ

$$P = f(\theta)$$

เมื่อ  $P$  แทน ผลการสอบ (Performance)

$\theta$  แทน ความสามารถหรือคุณลักษณะ (Ability หรือ Trait)

$f$  แทน ฟังก์ชัน (Function)

เนื่องจากข้อมูลที่ได้จากผู้สอบนั้นมีหลายลักษณะ ดังนั้นจึงได้มีการพัฒนาโมเดลที่แสดงความสัมพันธ์นี้ในลักษณะต่างๆ กันเพื่อให้แต่ละโมเดลมีความเหมาะสมกับข้อมูลที่ได้ โมเดลของทฤษฎีการตอบสนองข้อสอบสามารถจำแนกออกได้เป็น 3 กลุ่มใหญ่ๆ ตามลักษณะของข้อมูลที่ได้จากการสอบ 3 ลักษณะ คือ คำตอบที่เป็นข้อมูลแบบแบ่งสอง (Dichotomous)

ข้อมูลแบบพหุ (Multichotomous) และข้อมูลแบบต่อเนื่อง (Contineous) โมเดลที่มีความเหมาะสมกับลักษณะข้อมูลดังกล่าว มีดังนี้ (Hambleton and Swaminathan, 1985 : 33-52 )

ตาราง 6 ลักษณะของข้อมูล โมเดลที่ใช้วิเคราะห์ และผู้นำเสนอโมเดล

ลักษณะของข้อมูล	โมเดลที่ใช้วิเคราะห์	ผู้นำเสนอโมเดล
Dichotomous	Latent Linear	Lazarsfeld & Henry (1968)
	Latent Distance	
	Perfect scale	Guttman (1944)
	One, Two, Three Parameter	Lord (1952)
	Normal Ogive	
	One, Two, Three Parameter	Birnbaum (1957,1958a, 1958b,1968)
	Logistic	Lord & Novick (1968)
	Four Parameter Logistic	Wright & Stone (1979)
Multicategory	Nominal Response	McDonald (1967)
	Graded Response	Barton & Lord (1981)
	Partial Credit Model	Bock (1972)
Contineous	Contineous Response	Samejima (1969)
		Master (1982)
		Samijima (1972)

ในแต่ละโมเดลก็มีฟังก์ชันทางคณิตศาสตร์แตกต่างกันออกไป ดังต่อไปนี้

1) โมเดลนอร์มอลโอจีฟ แบบ 2 พารามิเตอร์ (Two-parameter Normal Ogive Model)  
เสนอโดย ลอร์ด (Lord, 1952, 1953a) ได้ังลักษณะข้อสอบจะหาได้จากสมการ

$$P(\theta) = \int_{-\alpha}^{a(\theta-b)} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

- เมื่อ  $P(\theta)$  แทน ความน่าจะเป็นที่ผู้สอบจะตอบข้อที่  $i$  ได้ถูกต้อง  
 ด้วยความสามารถ  $\theta$
- $b$  แทน ค่าความยากของข้อสอบข้อที่  $i$
- $a$  แทน ค่าอำนาจจำแนกของข้อสอบข้อที่  $i$
- $z$  แทน ความเบี่ยงเบนปกติจากการแจกแจงที่ค่าเฉลี่ย  $b$  และ  
 ความเบี่ยงเบนมาตรฐาน  $1/a$

2) โมเดลโลจิสติกแบบ 2 พารามิเตอร์ (Two-parameter Logistic Model) เบอรัมบอม (Birnbaum) เป็นผู้เสนอโมเดลนี้ ซึ่งโค้งลักษณะข้อสอบจะได้จากฟังก์ชันการแจกแจง ดังนี้

$$P(\theta) = \frac{e^{Da(\theta-b)}}{1 + e^{Da(\theta-b)}}$$

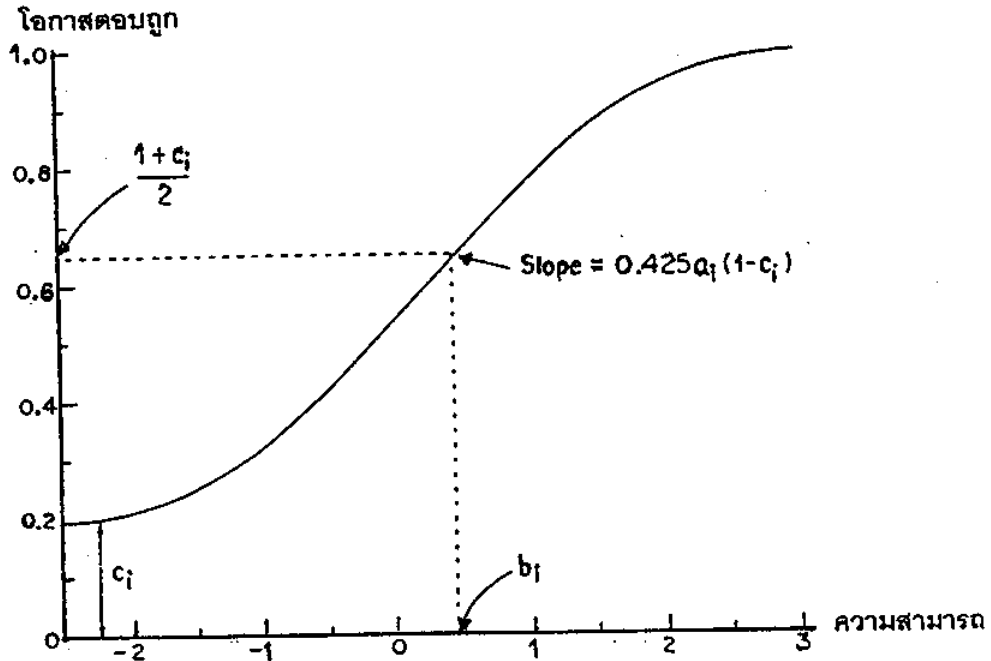
ค่าของ  $P(\theta)$ ,  $a$ ,  $b$  และ  $\theta$  มีความหมายเหมือนเดิม ส่วน  $D$  เป็นค่าที่ใช้ในการปรับ  
 แก่ (Scaling Factor) ซึ่งถ้ากำหนดให้  $D = 1.7$  แล้ว ค่าของ  $P(\theta)$  จากโมเดลนอร์มอลไฮเจส  
 และโมเดลโลจิสติกแบบ 2 พารามิเตอร์จะต่างกันเล็กน้อยกว่า .01 ในทุกๆ ค่าของ  $\theta$  ที่ประมาณได้  
 และโมเดลโลจิสติกนั้นนำไปใช้ได้สะดวกกว่าโมเดลนอร์มอลไฮเจสในการประมาณค่า  $P(\theta)$

3) โมเดลโลจิสติกแบบ 3 พารามิเตอร์ (Three-parameter Logistic Model) โมเดลนี้  
 พัฒนามาจากโมเดลโลจิสติกแบบ 2 พารามิเตอร์ โดยเพิ่มค่าพารามิเตอร์การเดา ( $c$ ) เข้าไปใน  
 ฟังก์ชันอีก 1 ค่า ทำให้ได้ฟังก์ชันของการแจกแจงดังนี้

$$P(\theta) = c + (1-c) \frac{e^{Da(\theta-b)}}{1 + e^{Da(\theta-b)}}$$

พารามิเตอร์  $c$  เป็นค่าความน่าจะเป็นของผู้สอบที่มีความสามารถต่ำมากๆ แต่สามารถ  
 ตอบข้อสอบได้ถูกต้อง ค่าความยากง่ายของข้อสอบ ( $b$ ) จะอยู่บนสเกลความสามารถ ( $\theta$ ) ตรงจุด  
 ที่ค่าความชันของโค้งมีค่าสูงสุด ซึ่งค่าความชันของโค้งที่สูงสุดมีค่าเท่ากับ  $0.425a(1-c)$  โค้ง  
 ลักษณะข้อสอบที่มีค่าอำนาจจำแนกสูง จะมีค่าความชันมากกว่าโค้งลักษณะข้อสอบที่มีค่าอำนาจ  
 จำแนกต่ำ ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้สอบที่มีความสามารถเท่ากับค่า  
 ความยากง่ายของข้อสอบจะเท่ากับ  $(1+c)/2$  ซึ่งโค้งลักษณะของข้อสอบตาม โมเดลนี้ มีลักษณะดัง  
 ภาพประกอบ 1





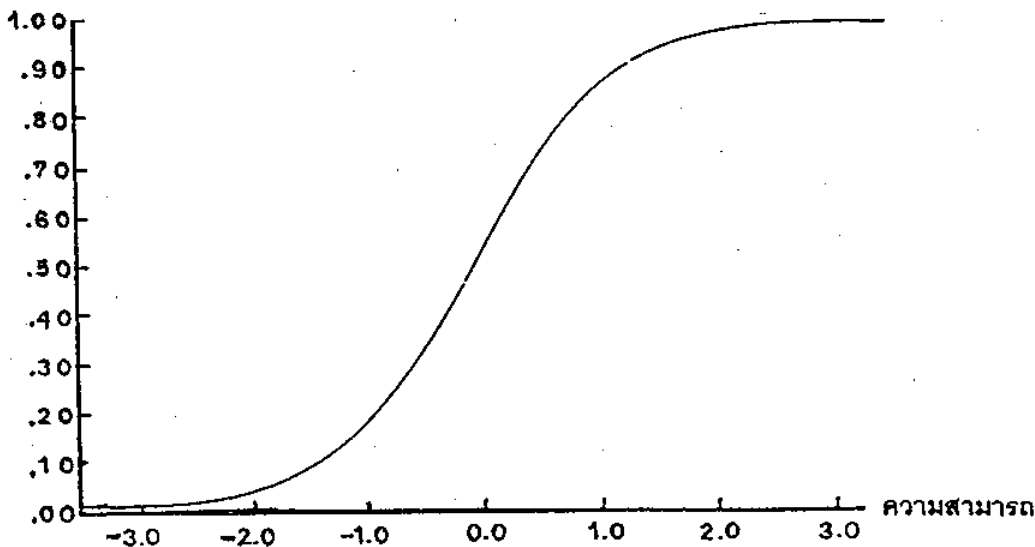
ภาพประกอบ 1 โค้งลักษณะของข้อสอบตามโมเดลโลจิสติกแบบ 3 พารามิเตอร์

4) โมเดลโลจิสติกแบบ 1 พารามิเตอร์ หรือราสช์โมเดล (One-parameter Logistic Model, Rasch Model) ราสช์ (Rasch) นักคณิตศาสตร์ชาวเดนมาร์กเป็นผู้เสนอโมเดลนี้ ซึ่งเป็นโมเดลที่เป็นกรณีเฉพาะของโมเดลโลจิสติก 3 พารามิเตอร์ของเบอร์นบอม (Birnbaum) ที่มีข้อตกลงว่า ข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันและมีค่าการเดาน้อยมากๆ ทำให้ได้ฟังก์ชันของการแจกแจง ดังนี้

$$P(\theta) = \frac{e^{Da(\theta-b)}}{1+e^{Da(\theta-b)}}$$

เมื่อ  $a$  แทน ค่าอำนาจจำแนกของข้อสอบ ซึ่งกำหนดให้มีค่าเท่ากันทุกข้อ มีโค้งลักษณะของข้อสอบ ดังภาพประกอบ 2

โอกาสตอบถูก



ภาพประกอบ 2 โค้งลักษณะของข้อสอบตามโมเดลโลจิสติกแบบ 1 พารามิเตอร์

5) โมเดลโลจิสติกแบบ 4 พารามิเตอร์ (Four-parameter Logistic Model) แมค โคนัล (McDonald) บาร์ตัน (Barton) และลอร์ด (Lord) เป็นผู้เสนอโมเดลนี้ โดยมีความเชื่อว่าผู้สอบที่มีความสามารถสูงๆ นั้น บางครั้งอาจตอบข้อสอบผิดได้ด้วยความสะดวก หรือการเฉลยข้อสอบที่คิดของผู้สร้างข้อสอบ ทำให้ฟังก์ชันของการแจกแจงเป็นดังนี้

$$P(\theta) = c + (\gamma - c) \frac{e^{Da(\theta-b)}}{1 + e^{Da(\theta-b)}}$$

เมื่อ  $\gamma$  เป็นค่าพารามิเตอร์ความสะดวก ซึ่งมีค่าต่ำกว่า 1 เล็กน้อย โมเดลนี้เป็นเพียงโมเดลที่ศึกษาตามทฤษฎีเท่านั้น ส่วนการปฏิบัติจริงนั้นยังไม่ค่อยนำมาใช้

6) โมเดลนอร์มอลโอจิวแบบ 1, 3 และ 4 พารามิเตอร์ (One Three and Four-parameter Normal Ogive Model) ลอร์ด (Lord) เป็นผู้เสนอโมเดลนี้ ซึ่งมีฟังก์ชันของการแจกแจง ดังนี้

$$P(\theta) = \int_{-\alpha}^{\theta-b} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

$$P(\theta) = c + (1-c) \int_{-\alpha}^{a(\theta-b)} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

$$P(\theta) = c + (\gamma - c) \int_{-\alpha}^{a(\theta-b)} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

เมื่อเปรียบเทียบกับ โมเดลโลจิสติกแล้ว พบว่าสมการจะมีลักษณะคล้ายๆ กัน แต่สมการของโมเดลโลจิสติกจะง่ายกว่า ซึ่งในทางปฏิบัติแล้วไม่ค่อยได้ใช้โมเดลนอร์มอลโอใจผีเท่าใดนัก เนื่องจากมีข้อจำกัดในเรื่องความยุ่งยากในการประมาณค่าและจะใช้โมเดลโลจิสติกแทน สำหรับโมเดลนอร์มอลโอใจผีนั้นเป็นการศึกษาตามทฤษฎีเท่านั้น

7) โมเดลการตอบสนองข้อสอบแบบนามบัญญัติ (Nominal Response Model) บ็อกและเซมิจิม่า (Bock and Samejima) เป็นผู้เสนอ โมเดลนี้ เนื่องจากเห็นว่า โมเดลนอร์มอลโอใจผีและโมเดลโลจิสติกนั้นมุ่งตอบสนองข้อมูลที่มีลักษณะแบบแบ่งสอง (Dichotomous) สำหรับ โมเดลนี้เหมาะสำหรับข้อมูลที่ได้จากการทดสอบที่ให้คะแนนแบบพหุ (Multidichotomouse) มีฟังก์ชันทางคณิตศาสตร์หลายฟังก์ชันที่แสดงโค้งลักษณะตัวเลือก ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ  $\theta$  จะเลือกตัวเลือก  $k$  (จาก  $m$  ตัวเลือกในแต่ละข้อ) สำหรับข้อที่  $i$  มีฟังก์ชันของการแจกแจง ดังนี้

$$P_k(\theta) = \frac{e^{b_k + a_k \theta}}{\sum_{k=1}^m e^{b_k + a_k \theta}}$$

เมื่อ  $b$  และ  $a$  เป็นค่าพารามิเตอร์ของข้อสอบที่สัมพันธ์กับตัวเลือก  $k$  (ถ้า  $m=2$  ข้อสอบก็จะถูกให้คะแนนแบบแบ่งสอง การประมาณค่าด้วยโมเดลนี้กับโมเดลโลจิสติกมีค่าเท่ากัน)

8) โมเดลการตอบสนองข้อสอบแบบให้ระดับคะแนน (Grade Response Model) เซมิจิม่า (Samejima) เป็นผู้เสนอ โมเดลนี้ ซึ่งคะแนนของการสอบนั้นจะให้ตามระดับของความถูกต้องในการตอบข้อสอบแต่ละข้อ เช่น ตอบผิดทั้งหมด ตอบผิดบางส่วนหรือตอบได้ถูกต้องทั้งหมด

ซึ่งในการสอบลักษณะนี้ข้อสอบหนึ่งข้อสามารถจำแนกคำตอบได้เท่ากับ  $m+1$  ตัวเลือก ถ้าคะแนนเต็มของข้อสอบนี้ คือ  $m$

9) โมเดลการตอบสนองของสเกลการวัดแบบต่อเนื่อง (Contineouse Scale) โมเดลนี้เป็นกรณีเฉพาะของโมเดลการตอบสนองข้อสอบแบบให้ระดับคะแนน ต่างกันตรงที่ผู้สอบจะตอบลงบนสเกลที่ต่อเนื่อง โมเดลนี้มีประโยชน์อย่างมากต่อนักจิตวิทยาสังคมและผู้ที่จะวัดเจตคติ

## 5. วิธีการเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบ

### 5.1 การเทียบโดยใช้คะแนนจริง

ในการเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบโดยใช้คะแนนจริงทำได้โดยแปลงค่า  $\theta$  ไปสู่คะแนนจริงที่สอดคล้องกันได้ โดยอาศัยสมการสัมพันธ์เชิงคณิตศาสตร์ระหว่างค่าความสามารถ  $\theta$  และคะแนนจริงที่ได้จากแบบสอบแต่ละฉบับ ดังนี้

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

เมื่อ  $P_i(\theta)$  คือ ความน่าจะเป็นในการตอบข้อสอบข้อที่  $i$  ได้ถูกต้อง  
เมื่อผู้สอบมีความสามารถ  $\theta$

$\theta_x$  คือ ความสามารถของผู้สอบในแบบสอบฉบับ  $X$

$\xi_x$  คือ คะแนนจริงที่ได้จากแบบสอบฉบับ  $X$

ในการทำงานเดียวกัน ถ้า  $\theta_y$  คือ ความสามารถของผู้สอบในแบบสอบฉบับ  $Y$  แล้วคะแนนจริงที่สมมูลกัน ในแบบสอบฉบับ  $Y$  คือ

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

เมื่อ  $\theta_y = \alpha\theta_x + \beta$  คือ เส้นความสัมพันธ์ระหว่าง  $\theta_y$  และ  $\theta_x$  สำหรับให้ค่าคะแนนจริงที่สมมูลกันแต่ละคู่ ( $\xi_x, \xi_y$ ) ดังนั้นจึงเป็นไปได้ที่จะเทียบคะแนนจริงจากแบบสอบต่างฉบับโดยคำนวณค่าคงที่ของการเทียบมาตรฐาน  $\alpha$  และ  $\beta$  หลังจากนั้นจะสามารถพล็อตกราฟเพื่อเทียบคะแนนจริงระหว่างแบบสอบได้

ในทางปฏิบัติสามารถหาค่า  $P_i(\theta_x)$  และ  $P_j(\theta_y)$  ได้จากการประมาณค่าพารามิเตอร์ของข้อสอบโดยเลือกว่าจะใช้ชนิดหนึ่ง สอง หรือสามพารามิเตอร์

## 5.2 การเทียบโดยใช้คะแนนจริงด้วยแบบสอบร่วม

การเทียบคะแนนจริงโดยใช้แบบสอบร่วม มีวิธีการคือ ถ้าแบบสอบฉบับ X และฉบับ Y สอบกับผู้สอบต่างกลุ่มกัน จะใช้แบบสอบร่วมเป็นตัวเชื่อมข้อมูลเข้าด้วยกัน ทำให้ค่าพารามิเตอร์ที่วิเคราะห์ห่ออกมาอยู่บนมาตราเดียวกัน ถ้าขาดแบบสอบร่วมแล้วจะไม่สามารถเทียบคะแนนกันได้ นอกจากนี้ว่ากลุ่มผู้สอบแบบสอบฉบับ X และสอบแบบสอบฉบับ Y มีการกระจายความสามารถเหมือนกัน วิธีการเทียบเช่นเดียวกับการเทียบโดยใช้คะแนนจริง

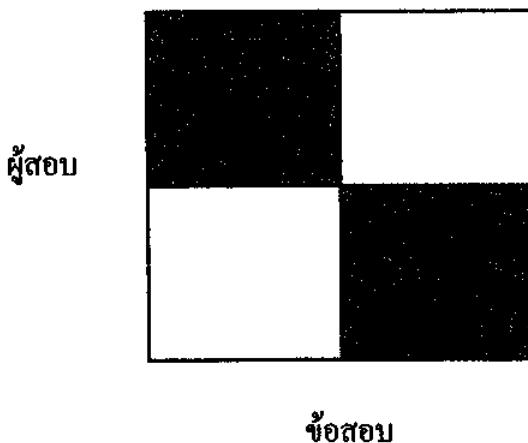
## 6. การแปลงค่าพารามิเตอร์ให้อยู่บนสเกลเดียวกัน (Rescale)

ค่าพารามิเตอร์ของข้อสอบที่ได้จากแบบสอบต่างชุด และผู้สอบต่างกลุ่มกันนั้น ไม่สามารถจะนำมาเปรียบเทียบกันได้ทันที ทั้งนี้เพราะสเกลของการวัดจะต่างกัน จึงจำเป็นต้องแปลงค่าพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเดียวกัน วิธีการแปลงประกอบด้วย 2 ขั้นตอน คือ การเลือกใช้แบบแผนของการใช้ข้อสอบหรือผู้สอบร่วม (Anchoring Design) และวิธีการแปลงคะแนน (Transformation Method) ซึ่งแต่ละขั้นตอนมีรายละเอียดดังนี้

### 6.1 แบบแผนการใช้ข้อสอบร่วมหรือผู้สอบร่วม (Anchoring Design)

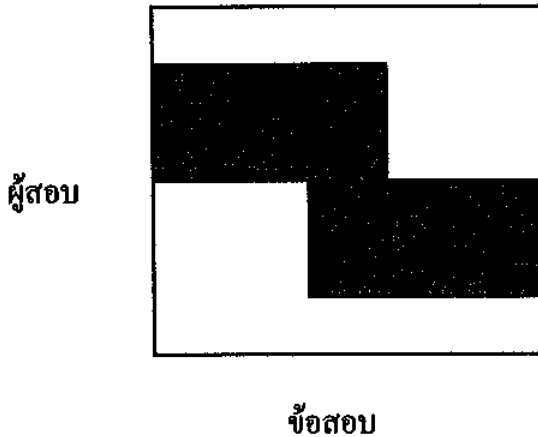
ในขั้นนี้เป็นขั้นตอนการวางแผนเพื่อเก็บข้อมูล ซึ่งจะต้องดำเนินการก่อนที่จะมีการสอบ แบบแผนดังกล่าวมีดังนี้

1) ใช้ข้อสอบต่างชุดและผู้สอบต่างกลุ่ม (Unanchored) วิธีนี้ใช้แบบสอบคู่ขนานกันซึ่งได้จากการสุ่มข้อสอบลงฉบับ แบบสอบเหล่านี้สามารถใช้แทนกันได้ แต่ละฉบับจะถูกนำไปสอบกับกลุ่มผู้สอบต่างกลุ่มกัน ไม่มีข้อสอบข้อใดที่ถูกนำไปสอบกับผู้สอบทั้งสองกลุ่ม และไม่มีผู้สอบคนใดที่ต้องทำแบบสอบหลายชุด ซึ่งมีลักษณะดังภาพประกอบ 3



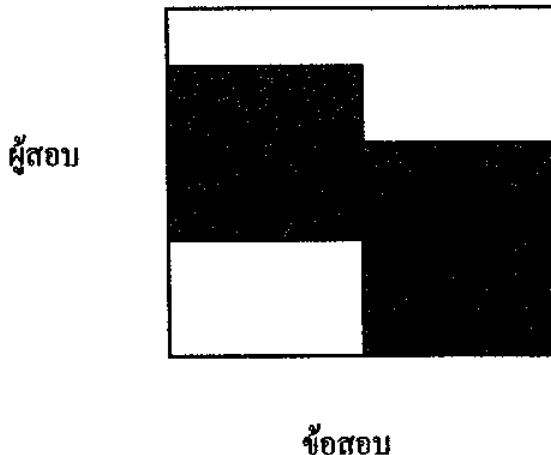
ภาพประกอบ 3 แบบแผนของการใช้ข้อสอบร่วมต่างชุดและผู้สอบต่างกลุ่ม

2) ใช้ข้อสอบร่วมกัน (Anchor-Items) แบบแผนนี้ถือว่าเป็นแบบแผนที่ได้รับความนิยมมาก ข้อสอบส่วนหนึ่งจะใช้เป็นข้อสอบร่วมซึ่งปรากฏอยู่ในแบบสอบทั้งสองฉบับ ค่าพารามิเตอร์ของข้อสอบร่วม จะเป็นตัวปรับค่าพารามิเตอร์ของข้อสอบจากแบบสอบฉบับต่างๆ คู่สเกลที่ใช้ร่วมกัน (Common Scale) ซึ่งมีลักษณะดังภาพประกอบ 4



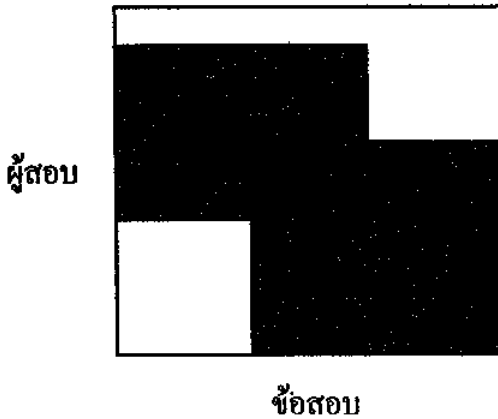
ภาพประกอบ 4 แบบแผนของการใช้ข้อสอบร่วมกัน

3) ใช้ผู้สอบร่วมกัน (Anchor-Group) แบบแผนนี้มีลักษณะคล้ายๆกับแบบแผนการใช้ข้อสอบร่วมกัน แต่ต่างกันตรงที่ใช้ผู้สอบร่วมกัน กล่าวคือ มีผู้สอบบางส่วนต้องทำแบบสอบมากกว่า 1 ฉบับ ซึ่งมีลักษณะดังภาพประกอบ 5



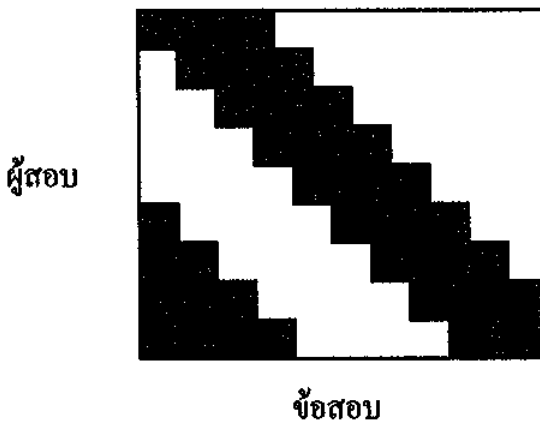
ภาพประกอบ 5 แบบแผนของการใช้ผู้สอบร่วมกัน

4) ใช้ผู้สอบและข้อสอบร่วมกัน (Double anchor Designs) สำหรับแบบแผนนี้ได้นำแบบแผนของการใช้ข้อสอบร่วมกันและผู้สอบร่วมกันมาผสมกัน เพื่อประหยัดค่าใช้จ่ายในการเก็บรวบรวมข้อมูล แต่ในทางปฏิบัติจริงนั้นไม่ค่อยได้มีการเลือกใช้แบบแผนนี้ เพราะยุ่งยากในเรื่องของการแปลงคะแนน แบบแผนมีลักษณะดังภาพประกอบ 6



ภาพประกอบ 6 แบบแผนของการใช้ข้อสอบและผู้สอบร่วมกัน

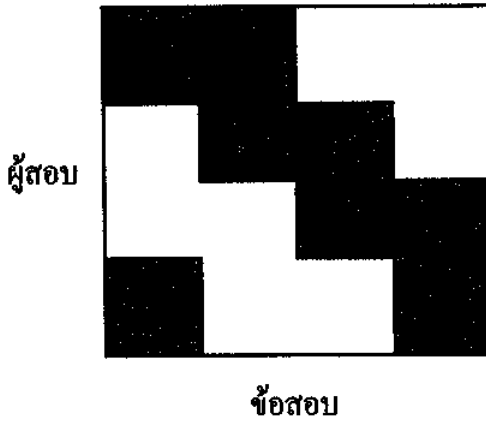
5) ใช้ข้อสอบร่วมกันแบบโยงสลับ (Interlaced) แบบแผนนี้ใช้ข้อสอบร่วมกัน จำนวนแบบสอบจะเท่ากับจำนวนของข้อสอบทั้งหมด และจำนวนข้อสอบในแต่ละฉบับจะเท่ากัน การจัดฉบับของแบบสอบจะมีลักษณะดังภาพประกอบ 7



ภาพประกอบ 7 แบบแผนของการใช้ข้อสอบร่วมกันแบบ โยงสลับ

6) ใช้ข้อสอบร่วมกันแบบ โยงสลับเป็นกลุ่ม (Block interlaced anchor) แบบแผนนี้มีลักษณะคล้ายกับแบบแผนของการใช้ข้อสอบร่วมกันแบบ โยงสลับ เป็นการแก้ปัญหาที่ต้องใช้

แบบสอบหลายฉบับ ดังนั้นจึงจัดผู้สอบเป็นกลุ่มๆ ให้ทำแบบสอบฉบับเดียวกัน ดังภาพประกอบ 8 จึงทำให้จำนวนแบบสอบลดลงกว่าเดิม



ภาพประกอบ 8 แบบแผนของการใช้ข้อสอบร่วมกันแบบโยงสลับเป็นกลุ่ม

## 6.2 วิธีการแปลงคะแนน (Transformation Method)

แบบสอบที่สามารถนำค่าพารามิเตอร์ของข้อสอบมาแปลงคะแนนได้นั้นต้องมีลักษณะดังนี้ (Wainer, 1990 : 146, อ้างถึงใน รังสรรค์ นนดีเล็ก, 2540 : 70)

- 1) แบบสอบทุกฉบับต้องวัด โครงสร้าง (Construct) เดียวกัน
- 2) การแปลงคะแนนต้องมีความเท่าเทียมกัน (Equity) กล่าวคือ การแจกแจงอย่างมีเงื่อนไขของคะแนนของแบบสอบแต่ละฉบับต้องเท่าเทียมกัน
- 3) ผลของการแปลงคะแนนต้องไม่แปรเปลี่ยน (Invariant) เมื่อใช้กลุ่มประชากรต่างกัน
- 4) คะแนนที่แปลงต้องมีลักษณะสมมาตร (Symmetric) เช่น คะแนนที่แปลงจากแบบสอบ X ไปยังแบบสอบ Y จะเท่ากับคะแนนที่แปลงจากแบบสอบ Y ไปยังแบบสอบ X ดังนั้นจึงไม่สามารถแปลงคะแนนโดยใช้ลักษณะของการถดถอย (Regression) ได้

ในการแปลงคะแนนไปสู่สเกลเดียวกันนั้น จะทำหลังจากที่ข้อสอบแต่ละข้อแต่ละฉบับได้รับการประมาณค่าพารามิเตอร์แล้ว การแปลงคะแนนเป็นเพียงการจัดวางตำแหน่งของค่าพารามิเตอร์ของข้อสอบจากแบบสอบฉบับต่างๆ ลงบนสเกลที่ใช้ร่วมกัน (Common Scale) วิธีแปลงคะแนนที่ง่ายที่สุดคือ การแปลงคะแนนเชิงเส้นตรง (Linear Transformation) ซึ่งได้ดังสมการ

$$\theta^* = k\theta + m$$

$$b^* = kb + m$$



เมื่อค่าคงที่ได้จาก

$$k = \frac{S^*}{S}$$

$$m = x^* - x \frac{S^*}{S}$$

- เมื่อ  $\theta^*$  แทน ค่าความสามารถที่ได้รับการแปลงค่า  
 $\theta$  แทน ค่าความสามารถที่ยังไม่ได้รับการแปลงค่า  
 $S^*$  แทน ค่าความคลาดเคลื่อนมาตรฐานที่แปลง  
 $S$  แทน ค่าความคลาดเคลื่อนมาตรฐานที่ยังไม่ได้แปลง  
 $x^*$  แทน คะแนนดิบที่ได้รับการแปลงค่า  
 $x$  แทน คะแนนดิบที่ยังไม่ได้รับการแปลงค่า  
 $k, m$  แทน ค่าคงที่

จากแบบแผนของการใช้ข้อสอบร่วมกันและผู้สอบร่วมกันตามที่กล่าวมาแล้ว สรุปเป็นกลุ่มใหญ่ๆ ได้ 3 กลุ่ม คือ กลุ่มที่ไม่ได้ใช้ข้อสอบและผู้สอบร่วมกันเลย กลุ่มที่ใช้ข้อสอบร่วมกัน และกลุ่มที่ใช้ผู้สอบร่วมกัน ซึ่งมีลักษณะและวิธีการแปลงคะแนน ดังนี้

6.2.1 กลุ่มที่ไม่ได้ใช้ข้อสอบและผู้สอบร่วมกันเลย ผลของการวิเคราะห์ข้อมูลจากผลการสอบของกลุ่มนี้ ค่าพารามิเตอร์ที่ได้ซึ่งได้แก่ คะแนนความสามารถ ค่าอำนาจจำแนก ค่าความยากง่าย และค่าการเดา ไม่สามารถนำมาเปรียบเทียบกันได้ เพราะสเกลของ  $\theta$  จะอยู่บนเมตริกที่ต่างกัน วิธีการแก้ไขคือ ต้องรวมข้อมูลผู้สอบทั้งหมดเข้าด้วยกัน เหมือนเป็นผู้สอบกลุ่มเดียวกันแล้วจึงวิเคราะห์ ส่วนของข้อสอบต่างชุดที่ผู้สอบไม่ได้ทำ ก็ถือเสมือนว่าเป็นข้อสอบที่ผู้สอบทำไม่ถึง (Lord, 1980 : 201)

6.2.2 กลุ่มที่ใช้ข้อสอบร่วมกัน ลักษณะของการสอบคือ จะมีข้อสอบอยู่จำนวนหนึ่งที่เหมือนกันอยู่ในแบบสอบทุกฉบับที่ใช้สอบ ซึ่งมีวิธีการสอบดังนี้

- 1) นำแบบสอบไปให้ผู้สอบ 2 กลุ่ม กลุ่มละ 1 ฉบับ
- 2) นำผลการสอบมาวิเคราะห์ค่าพารามิเตอร์แยกกัน โดยกำหนดค่าเฉลี่ยและค่าส่วนเบี่ยงเบนมาตรฐานของความสามารถของผู้สอบ ซึ่งมักกำหนดค่าให้เท่ากับ 0 และ 1 ตามลำดับ
- 3) หาค่าเฉลี่ยเลขคณิตและค่าเบี่ยงเบนมาตรฐานของค่าความยากง่าย ของข้อสอบรวมในแบบสอบทั้งสองฉบับ
- 4) แปลงค่าความยากง่าย ค่าอำนาจจำแนกของข้อสอบ และค่าความสามารถของผู้สอบ จากแบบสอบฉบับหนึ่ง ไปสู่แบบสอบอีกฉบับหนึ่ง ตามสูตร

$$b_1 = \left( \frac{SD_{b_1}}{SD_{b_2}} \right) b_2 + \bar{b}_1 - \left( \frac{SD_{b_1}}{SD_{b_2}} \right) \bar{b}_2$$

$$a_1 = a_2 \left( \frac{SD_{b_2}}{SD_{b_1}} \right)$$

$$\theta_1 = \left( \frac{SD_{b_1}}{SD_{b_2}} \right) \theta_2 + \bar{b}_1 - \left( \frac{SD_{b_1}}{SD_{b_2}} \right) \bar{b}_2$$

ค่าการเดาของข้อสอบไม่ต้องแปลงเนื่องจากอยู่บนสเกลเดียวกับความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง  $[P(\theta)]$  อยู่แล้ว หากต้องการวิเคราะห์ข้อมูลรวมกันก็สามารถทำได้โดยรวมข้อมูลผลการตอบของผู้สอบทั้งสองกลุ่ม ข้อสอบต่างฉบับที่ผู้สอบไม่ได้ตอบ ก็ถือเสมือนว่าเป็นข้อสอบที่ทำไม่ถึงหรือทำไม่ทัน จากนั้นก็วิเคราะห์ข้อสอบเหมือนกับว่าข้อสอบชุดใหญ่ถูกตอบโดยผู้สอบกลุ่มเดียวกัน

6.2.3 กลุ่มที่ใช้ผู้สอบร่วมกัน กรณีที่ใช้แบบสอบต่างฉบับกัน ผู้สอบต่างกลุ่มกัน แต่จะมีผู้สอบอยู่จำนวนหนึ่งที่ต้องทำแบบสอบทั้งสองฉบับ วิธีการแปลงคะแนนสามารถทำได้ดังนี้

1) รวมกระดาษคำตอบของผู้สอบรวมทั้ง 2 ฉบับ ดังนั้นจะเป็นเสมือนได้กระดาษคำตอบของแบบสอบฉบับยาว 1 ฉบับ

2) วิเคราะห์ค่าพารามิเตอร์แยกเป็นรายฉบับก็จะได้ค่าความสามารถ 2 ค่า หากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถแต่ละฉบับ แล้วแปลงค่าที่ได้ ดังนี้

$$b_1 = \left( \frac{SD_{\theta_1}}{SD_{\theta_2}} \right) b_2 + \bar{\theta}_1 - \left( \frac{SD_{\theta_1}}{SD_{\theta_2}} \right) \bar{\theta}_2$$

$$a_1 = a_2 \left( \frac{SD_{\theta_2}}{SD_{\theta_1}} \right)$$

$$\theta_1 = \left( \frac{SD_{\theta_1}}{SD_{\theta_2}} \right) \theta_2 + \bar{\theta}_1 - \left( \frac{SD_{\theta_1}}{SD_{\theta_2}} \right) \bar{\theta}_2$$

สำหรับค่าการเดานั้น ไม่ต้องมีการแปลงค่า เพราะอยู่บนสเกลเดียวกันอยู่แล้ว ส่วนกรณีที่ไม่มีผู้สอบร่วมซึ่งต้องทำแบบสอบทั้งสองฉบับ แต่ผู้ดำเนินการสอบต้องการที่จะใช้แบบแผนนี้ ก็อาจทำได้โดยนำข้อสอบจากแบบสอบทั้งสองฉบับส่วนหนึ่งมาสร้างเป็นแบบสอบฉบับที่ 3 แล้วนำไปสอบกับผู้สอบกลุ่มที่ 3 แล้วจึงใช้แบบสอบฉบับที่ 3 นี้เป็นข้อสอบร่วมเพื่อปรับค่าพารามิเตอร์ของแบบสอบฉบับที่ 1 และแบบสอบฉบับที่ 2

## 7. การหาค่าคงที่สำหรับสมการเทียบมาตรา

ค่าความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบจากการสอบตั้งแต่ 2 ครั้งขึ้นไป สามารถนำมาจัดให้อยู่บนสเกลเดียวกันได้โดยการใช้สมการเชิงเส้นเพื่อแปลงค่า ซึ่งสมการดังกล่าวมีค่าคงที่ 2 ค่า คือ  $\alpha$  และ  $\beta$  ซึ่งมีวิธีการหาค่า 4 วิธี ดังนี้

(Hambleton and Swaminathan, 1985 : 206-210)

1) วิธี Regression Method ความสัมพันธ์เชิงเส้นระหว่าง 2 ตัวแปร ส่วนใหญ่แล้วมีลักษณะดลดอย ดังสมการ

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

การแปลงค่าความสามารถทำได้โดยให้  $y = \theta_y$  และ  $x = \theta_x$  และในการแปลงค่าความยากง่ายของข้อสอบนั้น ทำได้โดยให้  $y = b_y$  และ  $x = b_x$  ส่วนค่า  $e$  คือ ค่าความคลาดเคลื่อนและค่าสัมประสิทธิ์การดลดอย หาได้โดย

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

$$\hat{\beta} = \bar{y} - \alpha \bar{x}$$

เมื่อ  $r$  แทน สหสัมพันธ์ระหว่าง  $x$  กับ  $y$

$\bar{y}$  แทน ค่าเฉลี่ยของ  $y$

$\bar{x}$  แทน ค่าเฉลี่ยของ  $x$

$S_y$  แทน ค่าเบี่ยงเบนมาตรฐานของ  $y$

$S_x$  แทน ค่าเบี่ยงเบนมาตรฐานของ  $x$

วิธีการนี้ไม่ค่อยเหมาะสมในการเทียบมาตรา เพราะไม่สมมาตรเนื่องจากแบบสอบที่ใช้เป็นฐานจะมีผลต่อค่าสัมประสิทธิ์ การดลดอย

2) วิธี Mean and Sigma Method วิธีการนี้มีลักษณะของตัวแปร ดังสมการ

$$y = \alpha x + \beta$$

$$\bar{y} = \alpha \bar{x} + \beta$$

$$S_y = \alpha S_x$$

$$\alpha = \frac{S_y}{S_x}$$

$$\beta = \bar{y} - \alpha \bar{x}$$

ความสัมพันธ์นี้มีลักษณะสมมาตร หากใช้รูปแบบการวิเคราะห์ข้อมูลโดยราตซ์โมเดล แล้วค่าคงที่ที่ได้จากวิธีนี้จะเท่ากับวิธีการดลดอย

ความสัมพันธ์นี้มีลักษณะสมมาตร หากใช้รูปแบบการวิเคราะห์ข้อมูลโดยวิธีขโมเดล แล้วค่าคงที่ที่ได้จากวิธีนี้จะเท่ากับวิธีการถดถอย

3) วิธี Robust Mean and Sigma Method ถึงแม้ว่าวิธีการหาค่าคงที่วิธี Mean and Sigma Method จะมีลักษณะของการให้ค่าคงที่ที่เปลี่ยนแปลงคะแนนแล้วมีลักษณะสมมาตรก็ตาม แต่ก็ไม่สามารถตอบได้ว่า ค่าพารามิเตอร์ของข้อสอบและค่าความสามารถของผู้สอบจะถูกแปลงค่าได้อย่างถูกต้องแม่นยำกว่า ในการวิเคราะห์ข้อมูลแต่ละครั้งก็จะเกิดความคลาดเคลื่อนมาตรฐานในการประมาณค่าทุกครั้ง การถ่วงน้ำหนักให้กับข้อมูลแต่ละคู่ (x,y) ก็จะเป็นการปรับค่าความแปรปรวนที่หลายๆจากการประมาณ ถ้าเป็นการแปลงค่าความสามารถ ค่าความสามารถที่มีความแปรปรวนหลายๆจะมีค่าถ่วงน้ำหนักน้อยๆ ถ้าใช้ค่าความยากง่ายเพื่อหาค่าคงที่ของสมการแล้ว Information Matrix ของข้อสอบแต่ละข้อจะถูกปรับค่า ส่วนประกอบของ Diagonal ที่เหมาะสมจะถูกใช้เป็นค่าความแปรปรวน วิธีหาค่าคงที่มีขั้นตอนดังนี้

ขั้นที่ 1 กำหนดค่าถ่วงน้ำหนัก (w) ของ x และ y แต่ละคู่

$$w_j = \max \{v(x_j), v(y_j)\} \quad \text{เมื่อ } j = 1, 2, 3, \dots, k$$

เมื่อ  $v(\theta_j)$  แทน ความแปรปรวนของการประมาณค่าความสามารถครั้งที่ j

ขั้นที่ 2 สเกลของค่าถ่วงน้ำหนัก คือ

$$w_j = \frac{w_j}{\sum_{j=1}^n w_j}$$

ขั้นที่ 3 หาค่าที่ได้รับถ่วงน้ำหนักแต่ละคู่

$$x'_j = w'_j x_j$$

$$y'_j = w'_j y_j$$

ขั้นที่ 4 หาค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของค่าที่ถ่วงน้ำหนัก

ขั้นที่ 5 หาค่าคงที่  $\alpha, \beta$  จากสมการ

$$\alpha = \frac{S'_y}{S'_x}$$

$$\beta = y' - \alpha x'$$

ต่อมาได้มีการปรับปรุงค่าถ่วงน้ำหนักให้มีประสิทธิภาพสูงขึ้น โดยนำระยะห่างระหว่างข้อมูลกับเส้นเทียบคะแนน ซึ่งเรียกว่า คิวถ่วงน้ำหนักที่แกร่ง (Robust Weight) มาใช้ในการถ่วงน้ำหนัก

ขั้นที่ 6 เมื่อได้ค่าคงที่ของสมการแปลงคะแนนแล้ว เส้นเทียบคะแนนจะถูกกำหนดขึ้น  
ระยะห่างระหว่างข้อมูลกับเส้นเทียบคะแนน หาได้จากสมการ

$$d_j = \frac{y_j - \alpha x_j - \beta_j}{\sqrt{\alpha^2 + \beta^2}}$$

ขั้นที่ 7 หาค่าถ่วงน้ำหนัก(Turkey Weight) จากสมการ

$$T_j = \left[ 1 - \left( \frac{d_j}{6M} \right)^2 \right]^2 \quad ; \quad d_j < 6M$$

ขั้นที่ 8 ถ่วงน้ำหนักของข้อมูลแต่ละคู่ด้วยค่าถ่วงน้ำหนัก

$$u_j = w_j' T_j \left[ \sum_{j=1}^n w_j' T_j \right]$$

ขั้นที่ 9 ทำซ้ำขั้นตอนที่ 3 แล้วหาค่า  $w_j'$  แทน  $w_j$  แล้วหาค่าคงที่  $\alpha$  และ  $\beta$  ใน  
ขั้นตอนที่ 5

ขั้นที่ 10 ทำซ้ำขั้นตอนที่ 6-9 จนกระทั่งค่าคงที่  $\alpha$  และ  $\beta$  เปลี่ยนแปลงน้อยกว่าที่  
กำหนดไว้

4) วิธี Characteristic Curve Method วิธีหาค่าคงที่ในสมการเทียบมาตรฐานวิธี Roust  
Mean and Sigma Method มีข้อบกพร่องตรงที่จะใช้ความสัมพันธ์  $b_y = \alpha b_x + \beta$  เท่านั้นในการหา  
ค่าคงที่ แต่ไม่ได้ใช้ค่าอำนาจจำแนกในการหาค่า  $\alpha$  แต่อย่างใด ตามวิธีการนี้คะแนนจริงของผู้สอบ  
ที่มีความสามารถ  $\theta$  บนแบบสอบฉบับ X และความสามารถ  $\theta$  บนแบบสอบ Y คือ

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

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

ค่าคงที่  $\alpha$  และ  $\beta$  ที่ดีก็คือค่าที่ทำให้คะแนนจริง  $\xi_{xa}$  และ  $\xi_{ya}$  มีความแตกต่างกันน้อย  
ที่สุดโดยใช้เกณฑ์ของสต็อกกิงและลอร์ด (Stocking and Lord, 1983) ดังนี้

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

เมื่อ N แทน จำนวนผู้สอบ

F แทน ฟังก์ชันของ  $\alpha$  และ  $\beta$  ซึ่งมีค่าน้อยที่สุด เมื่อ

$$\frac{\delta F}{\delta \alpha} = \frac{\delta F}{\delta \beta} = 0$$

### ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

การวิจัยในครั้งนี้ผู้วิจัยได้ศึกษาค้นคว้างานวิจัยเกี่ยวกับการเทียบมาตรฐานในแนวนอน และการเทียบมาตรฐานในแนวตั้ง เพื่อนำข้อมูลประกอบการวิจัย ซึ่งสามารถแยกเป็นงานวิจัยต่างประเทศและงานวิจัยภายในประเทศ ที่กล่าวถึงต่อไปนี้

#### 1. งานวิจัยต่างประเทศ

สไลด์และลินด์ (Slinde and Linn, 1977-1979, อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 68) ศึกษาเพื่อตรวจสอบปัญหาของการเทียบมาตรฐานในแนวตั้ง ของแบบสอบสองชุดที่สร้างขึ้นใช้กับกลุ่มตัวอย่างย่อยที่ความสามารถต่างกัน ผลการศึกษาให้ข้อเสนอแนะว่า การใช้รูปแบบการเทียบมาตรฐาน 3 รูปแบบ คือ การเทียบมาตรฐานเชิงเส้นตรง การเทียบมาตรฐานวิธีอิกวิเปอร์เซ็นไทล์ และการเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบชนิดหนึ่งพารามิเตอร์รูปแบบโลจิสติก การเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบรูปแบบโลจิสติกอาจมีข้อจำกัด โดยเฉพาะเมื่อเทียบมาตรฐานโดยใช้แบบสอบสองชุดที่มีความแตกต่างกันมาก และกลุ่มตัวอย่างสองกลุ่มที่มีระดับความสามารถต่างกันมาก และให้ข้อสังเกตว่า การเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบ ถ้าใช้รูปแบบสามพารามิเตอร์ของโลจิสติก อาจให้ผลการเทียบมาตรฐานที่ดีกว่าในสถานการณ์เดียวกัน

มาร์โค ปีเตอร์สัน และสตีเวอร์ท (Marco, Petersen and Stewart, 1983, อ้างถึงใน วรรณดี แสงประทีปทอง, 2538 : 41-42) นำวิธีการเทียบมาตรฐานต่างๆ ได้แก่ วิธีเชิงเส้นตรง 4 วิธี วิธีอิกวิเปอร์เซ็นไทล์ 2 วิธี และวิธีการเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบแบบจำลองโลจิสติกหนึ่งและสามพารามิเตอร์ มาศึกษาภายใต้เงื่อนไขต่างๆ ได้แก่ กลุ่มตัวอย่างคล้ายคลึงกัน และต่างกัน แบบสอบร่วมภายในและภายนอก คะแนนเกณฑ์และค่าสถิติรวมที่ต่างกัน การศึกษาแบ่งเป็น 2 ส่วน คือ การเทียบมาตรฐานในแนวนอนและการเทียบมาตรฐานในแนวตั้ง โดยใช้แบบสอบร่วม การประเมินความเที่ยงพอของการเทียบมาตรฐานใช้สถิติ 2 อย่าง คือ ความคลาดเคลื่อนรวม และค่าความลำเอียงยกกำลังสอง ผลการศึกษาพบว่าในการเทียบมาตรฐานในแนวนอน เมื่อใช้แบบสอบร่วมภายนอก วิธีเทียบมาตรฐานเชิงเส้นตรงและวิธีเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบให้ผลดี และเมื่อใช้แบบสอบร่วมภายใน วิธีอิกวิเปอร์เซ็นไทล์ให้ผลดีเช่นกัน และวิธีใช้แบบจำลองของราล์สให้ผลดีกว่าวิธีอื่นๆ ส่วนการเทียบมาตรฐานในแนวตั้งนั้นวิธีการเทียบเชิงเส้นตรงมีความคลาดเคลื่อนมากที่สุด รองลงมาคือวิธีใช้แบบจำลองของราล์ส สำหรับการเทียบมาตรฐานในแนวตั้ง โดยใช้ทฤษฎีการตอบสนองข้อสอบสามพารามิเตอร์มีความคลาดเคลื่อนน้อยที่สุด

โกลเลน (Kolen, 1981 : 1-11) ได้ศึกษาเพิ่มเติมจากผลการศึกษาของมาร์โคและคณะ โดยใช้วิธีการเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบโดยใช้การเทียบคะแนนจริงและใช้คะแนนดิบ วิธีเทียบมาตรฐานเชิงเส้นตรง และวิธีเทียบมาตรฐานวิธีอิกวิเปอร์เซ็นไทล์ และใช้กลุ่มตัวอย่างตามผล ผลการศึกษาพบว่า การเทียบมาตรฐานในแนวตั้งวิธีอิกวิเปอร์เซ็นไทล์และวิธีใช้สามพารามิเตอร์ให้ผลดีกว่าวิธีอื่นๆ และวิธีการเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบเมื่อใช้การประมาณค่าจากคะแนนดิบให้ผลดีกว่าใช้คะแนนจริงเล็กน้อย

เพเทียสน (Patience, 1981) ศึกษาการเทียบมาตรฐานในแนวตั้งโดยใช้แบบสอบที่แบ่งย่อยเป็นฉบับยาก ยากปานกลาง และง่าย ใช้กับกลุ่มตัวอย่างนักเรียนระดับ 12, 11 และระดับ 10 ตามลำดับ ใช้วิธีการเทียบมาตรฐานตามทฤษฎีการตอบสนองข้อสอบที่มีหนึ่ง สอง สามพารามิเตอร์ และวิธีอิกวิเปอร์เซ็นไทล์ ดำเนินการเทียบจากแบบสอบฉบับยากและฉบับง่ายไปยังแบบสอบฉบับยากปานกลาง การประเมินใช้ค่าสหสัมพันธ์ระหว่างคะแนนสอบกับคะแนนที่ผ่านการเทียบแล้วพบว่าค่าสหสัมพันธ์ของการเทียบมาตรฐานเมื่อใช้แบบจำลองสามพารามิเตอร์ต่ำกว่าวิธีอื่นๆ

## 2. งานวิจัยภายในประเทศ

ชูชีพ พงษ์สมบุรณ์ (2528 : 112-118) ศึกษาเปรียบเทียบประสิทธิภาพและความคงที่ของการเทียบมาตรฐานระหว่างรูปแบบที่ใช้แบบสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วม โดยใช้วิธีการเทียบ 3 วิธี คือ การเทียบเชิงเส้นตรง การเทียบโดยใช้เปอร์เซ็นไทล์ และการเทียบโดยใช้โค้งลักษณะข้อสอบ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 จำนวน 3,721 คน จากการสอบโครงการตรวจสอบคุณภาพการศึกษาทั่วประเทศ ในวิชาคณิตศาสตร์ ของกรมวิชาการ กระทรวงศึกษาธิการ แบบสอบที่ใช้แบ่งเป็น 2 ฉบับๆ ละ 38 ข้อ มีข้อสอบรวมภายใน 20 ข้อ แบบสอบทั้งสองฉบับมีความยากใกล้เคียงกัน ผลการศึกษาพบว่าประสิทธิภาพของการเทียบมาตรฐานระหว่างรูปแบบที่ใช้แบบสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วมในแต่ละวิธีไม่แตกต่างกัน สำหรับการเปรียบเทียบความคงที่ของการเทียบมาตรฐานทั้งสองรูปแบบได้ผลเช่นเดียวกัน คือ การเทียบเชิงเส้นตรงมีความคงที่ต่ำที่สุด และการเทียบโดยใช้โค้งลักษณะข้อสอบคงที่พอๆ กับการเทียบโดยใช้เปอร์เซ็นไทล์

ภาวิณี ศรีสุขวัฒนานันท์ (2529 : 155-166) ศึกษาเปรียบเทียบผลการใช้รูปแบบการเทียบมาตรฐานที่ต่างกัน 3 รูปแบบ คือ รูปแบบอิกวิเปอร์เซ็นไทล์ รูปแบบเชิงเส้นตรง และรูปแบบอิงทฤษฎีการตอบสนองแบบโลจิสติกตามพารามิเตอร์ โดยใช้ข้อสอบรวมที่ต่างกัน 3 ขนาด คือ ขนาดร้อยละ 20 (7 ข้อ) ขนาดร้อยละ 40 (14 ข้อ) และขนาดร้อยละ 60 (21 ข้อ) ข้อสอบทั้งหมดมี 100 ข้อ จัดแบ่งเป็นแบบสอบ 2 ฉบับๆ ละ 35 ข้อ ซึ่งทั้งสองฉบับมีความยากใกล้เคียงกัน

ที่เหลือนักเป็นข้อสอบพร้อมขนาดต่างๆ สำหรับกลุ่มตัวอย่างแยกเป็น 2 กรณี คือ กรณีการสอบคัดเลือก และกรณีการสอบวัดผลสัมฤทธิ์ กลุ่มตัวอย่างแต่ละกรณีมี 2 กลุ่มๆ ละ 1,500 คน ใช้เป็นกลุ่มเทียบมาตรฐานและกลุ่มสอบทานผล แบบสอบที่ใช้มีกรณีละ 2 ชุด คือ ชุด X และชุด Y ซึ่งมีจำนวนชุดละ 35 ข้อ และมีแบบสอบพร้อมขนาด 21, 14 และ 7 ข้อ โดยยึดหลักแบบสอบพร้อมมีค่าความเชื่อมั่นเท่าเทียมกัน ในแต่ละคู่ใช้รูปแบบการเทียบมาตรฐาน 3 รูปแบบ การประเมินผลของการเทียบมาตรฐาน 2 ลักษณะ คือ เปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานด้วยรูปแบบต่างกันของแบบสอบแต่ละกรณี และตรวจสอบความเพียงพอของการเทียบมาตรฐานแต่ละรูปแบบ โดยใช้ค่าดัชนีการเปรียบเทียบความแตกต่าง (Index : C) ที่ได้จากกลุ่มสอบทานผล ผลการศึกษาพบว่า การเทียบมาตรฐานทั้ง 9 แบบ ได้ผลยอมรับว่า ความยาวแบบสอบพร้อมมีผลต่อความแม่นยำและความเพียงพอของการเทียบมาตรฐาน โดยที่แบบสอบพร้อมมีความยาวมากและแบบสอบทั้งสองชุดมีความเชื่อมั่นสูงและใกล้เคียงกันแล้ว จะทำให้การเทียบมาตรฐานแม่นยำมากขึ้นในแต่ละสถานการณ์การสอบให้ผลการเทียบมาตรฐานแต่ละรูปแบบต่างกัน คือ กรณีการสอบคัดเลือก รูปแบบที่ผลการเทียบมาตรฐานมีความเพียงพอมากที่สุด คือ รูปแบบอิกวิเปอร์เซ็น ไทล์ รูปแบบอิงทฤษฎีการตอบข้อสอบ และรูปแบบเชิงเส้นตรง ตามลำดับ ส่วนกรณีการสอบวัดผลสัมฤทธิ์ รูปแบบให้ผลการเทียบมาตรฐานเพียงพอมากที่สุด คือ รูปแบบเชิงเส้นตรง รูปแบบอิกวิเปอร์เซ็น ไทล์ และรูปแบบอิงทฤษฎีการตอบข้อสอบ ตามลำดับ

เรวดี อินทสระ (2530 : 64-68) ศึกษาเปรียบเทียบความคลาดเคลื่อนมาตรฐานและความเพียงพอของการเทียบมาตรฐานระหว่างรูปแบบอิงทฤษฎีการตอบข้อสอบแบบสามพารามิเตอร์ กับรูปแบบเทคนิคการวิเคราะห์องค์ประกอบ โดยใช้กลุ่มทานผล กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2529 ในจังหวัดพัทลุง จำนวน 2,823 คน แบบสอบที่ใช้เป็นแบบสอบวิชาคณิตศาสตร์ มี 2 ฉบับๆ ละ 45 ข้อ โดยมีแบบสอบพร้อมภายในฉบับละ 15 ข้อ ผลการศึกษาพบว่า ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานระหว่างรูปแบบอิงทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานรูปแบบอิงทฤษฎีการตอบข้อสอบมีค่าน้อยกว่ารูปแบบเทคนิควิเคราะห์องค์ประกอบ ส่วนความเพียงพอของการเทียบมาตรฐานรูปแบบอิงทฤษฎีการตอบข้อสอบแบบสามพารามิเตอร์อยู่ในระดับที่น่าพอใจ รูปแบบเทคนิควิเคราะห์องค์ประกอบอยู่ในระดับปานกลาง

อาภรณ์ กาญจนกิจโสภณ (2531 : 66-69) ศึกษาการสร้างแบบสอบและตารางเทียบคะแนนในแนวตั้ง วิชาคณิตศาสตร์ เรื่อง สมการและอสมการ สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น วิเคราะห์ตามแนวราล์ซ โมเดล กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาตอนต้น



เขตท้องที่การศึกษา 2 กรุงเทพมหานคร จำนวน 2,510 คน แบบสอบที่ใช้สร้างขึ้น 3 ฉบับ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ฉบับละ 40 ข้อ มีข้อสอบร่วมฉบับละ 10 ข้อ โดยคัดเลือกข้อสอบที่มีความยากสูงในแบบสอบชั้นมัธยมศึกษาปีที่ 1 และ 2 เป็นข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 1 กับ 2 และข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 2 กับ 3 ตามลำดับ การสร้างตารางเทียบคะแนน สร้างจากการปรับค่าความสามารถให้อยู่ในสเกลเดียวกัน เพื่อนำไปศึกษาพัฒนาการของนักเรียนกลุ่มเดิมที่ผ่านขึ้นไปเรียนชั้นสูงขึ้นไปว่า นักเรียนมีความก้าวหน้าเป็นไปตามปกติ หรือว่าเบี่ยงเบนไปตามปกติ

สุวัฒน์ ขวัญบุญจันทร์ (2531 : 77-80) ศึกษาการสร้างแบบสอบและตารางเทียบคะแนนในแนวนอนวิชาวิทยาศาสตร์ ที่วิเคราะห์ด้วยราล์ฟ โมเดล กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2531 ในจังหวัดชัยภูมิ จำนวน 1,391 คน แบบสอบที่สร้างขึ้นเป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาวิทยาศาสตร์ เรื่อง พลังงานและการเปลี่ยนแปลง ใช้กลุ่มสอบทานผลตรวจสอบความเพียงพอของการเทียบคะแนน สร้างตารางเทียบคะแนนโดยเทียบหาค่าคะแนนดิบจากความสามารถที่เทียบเท่ากัน ผลการศึกษาพบว่า ภายใต้อำนาจที่เทียบเท่ากัน ได้คะแนนดิบแต่ละค่าเท่ากันทุกคะแนน และค่าดัชนีความเพียงพอของความคลาดเคลื่อนในการเทียบคะแนนอยู่ในระดับที่น่าพอใจมาก

สุจินดา ผ่องอักษร (2533 : 143-157) ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบคะแนนที่วิเคราะห์ตามราล์ฟ โมเดล หลังจากเทียบคะแนนแล้ว มีการเปรียบเทียบความแตกต่างของผลสัมฤทธิ์ทางเรียนระหว่างปีการศึกษาต่างๆ และเปรียบเทียบสัดส่วนจำนวนนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนผ่านเกณฑ์ที่น่าพอใจ ระหว่างการเทียบคะแนนกับไม่เทียบคะแนน การออกแบบเก็บรวบรวมข้อมูลใช้รูปแบบผู้สอบร่วม โดยกลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบหลักสูตรประถมศึกษา พุทธศักราช 2521 ในปีการศึกษา 2532 ในจังหวัดปทุมธานี จำนวน 581 คน แบบสอบที่ใช้เป็นแบบสอบวัดผลสัมฤทธิ์ในกลุ่มทักษะ (วิชาคณิตศาสตร์และภาษาไทย) ซึ่งสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติสร้างขึ้น และใช้สอบนักเรียนเมื่อปีการศึกษา 2529, 2530 และ 2531 ผลการศึกษาพบว่า จากการเทียบคะแนนระหว่างแบบสอบทั้ง 3 ฉบับ คะแนนที่เทียบแล้วของแบบสอบทั้งสามฉบับ ซึ่งตรงกับคะแนนดิบค่าเดียวกัน มีค่าแตกต่างกันทั้งแบบสอบวิชาคณิตศาสตร์และแบบสอบวิชาภาษาไทย เมื่อศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนพบว่า ผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ในปีการศึกษา 2530 สูงกว่าปีการศึกษา 2529 แต่ลดลงในปีการศึกษา 2531 สำหรับผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทย

พบว่า อัตราความก้าวหน้าของคะแนนเฉลี่ยเพิ่มขึ้นทุกปีการศึกษา และอัตราความก้าวหน้าของค่าร้อยละนักเรียนที่มีผลน่าพอใจ เพิ่มขึ้นทุกปีการศึกษา

สุนิสา ชัยม่วงศรี (2537 : 88-92) ศึกษาคุณภาพของการเทียบมาตรฐานเชิงเส้นตรง เมื่อใช้แบบสอบร่วมภายในที่มีความยาวต่างกัน 4 ขนาด โดยเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน และเปรียบเทียบดัชนีความแตกต่างจากการวิเคราะห์กลุ่มทานผล กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2534 ของกลุ่มโรงเรียนกลุ่มที่ 2 ในจังหวัดขอนแก่น จำนวน 927 คน แบบสอบที่ใช้เป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียน ปลายภาคเรียน วิชาคณิตศาสตร์ ค 203 ซึ่งมี 2 ฉบับ สร้างขึ้นใช้ในสถานการณ์การสอบจริงเมื่อปีการศึกษา 2532 และปีการศึกษา 2533 ฉบับละ 60 ข้อ นำไปสอบกับกลุ่มตัวอย่าง โดยเพิ่มข้อสอบร่วมภายใน ฉบับละ 25 ข้อ นำผลการสอบมาวิเคราะห์โดยพิจารณาตามความยาวแบบสอบร่วม เป็นแบบสอบ ฉบับละ 70, 75, 80 และ 85 ข้อ ผลการศึกษาพบว่า ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานเชิงเส้นตรงมีค่าลดลง เมื่อแบบสอบร่วมมีความยาวเพิ่มขึ้น ส่วนค่าดัชนีความแตกต่างของการเทียบมาตรฐานที่ได้จากการใช้แบบสอบร่วมที่ยาวกว่าจะมีค่าน้อยกว่า และจากการใช้แบบสอบ ทั้ง 4 ขนาด ให้ผลอยู่ในระดับที่ยอมรับได้

วรเนตร พิชิตกรีกพล (2537 : 94-98) ศึกษาเปรียบเทียบคุณภาพของวิธีการเทียบมาตรฐานระหว่างวิธีอิกวิเปอร์เซ็นไทล์กับวิธีเชิงเส้นตรง โดยเปรียบเทียบความคลาดเคลื่อนของคะแนนที่ได้จากการเทียบมาตรฐาน ที่วิเคราะห์จากกลุ่มทานผล เมื่อกลุ่มตัวอย่างมีความสามารถใกล้เคียงและแตกต่างกัน สอบแบบสอบที่มีความยากทั้งฉบับใกล้เคียงกันและแตกต่างกัน กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2534 ของจังหวัดพระนครศรีอยุธยา จำนวน 1,140 คน แบบสอบที่ใช้เป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียนในวิชาคณิตศาสตร์ ค 203 จำนวน 3 ฉบับๆ ละ 40 ข้อ โดยมีข้อสอบร่วมฉบับละ 20 ข้อ ผลการศึกษาพบว่า เมื่อใช้แบบสอบที่มีความยากใกล้เคียงกันและแตกต่างกัน กับกลุ่มตัวอย่างที่มีความสามารถแตกต่างกัน การเทียบมาตรฐานวิธีอิกวิเปอร์เซ็นไทล์ให้คุณภาพดีกว่าวิธีเชิงเส้นตรง ส่วนกรณีแบบสอบที่มีความยากใกล้เคียงกันกับกลุ่มตัวอย่างที่มีความสามารถใกล้เคียงกัน วิธีเชิงเส้นตรงให้คุณภาพดีกว่าวิธีอิกวิเปอร์เซ็นไทล์ แต่การเทียบมาตรฐานทั้งสองวิธีอยู่ในระดับไม่น่าพอใจ เมื่อใช้กรณีแบบสอบมีความยากต่างกันกับกลุ่มตัวอย่างที่มีความสามารถแตกต่างกัน

พรพิมล นาคเวช (2537 : 99-101) ศึกษาคุณภาพของการเทียบมาตรฐานในแนวตั้งโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ พิจารณาค่าความคลาดเคลื่อนมาตรฐานและความเพียงพอของการเทียบมาตรฐานโดยใช้กลุ่มสอบทานผล กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้น

มัธยมศึกษาตอนต้น ปีการศึกษา 2535 จังหวัดสุราษฎร์ธานี จำนวน 6,429 คน แบบสอบที่ใช้เป็นแบบสอบวิชาคณิตศาสตร์ เรื่อง สมการและอสมการ ของระดับชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ระดับละ 1 ฉบับ ฉบับละ 25 ข้อ โดยมีข้อสอบร่วมระหว่างระดับชั้นที่ติดกันระดับชั้นละ 5 ข้อ ผลการศึกษาพบว่า ความคลาดเคลื่อนมาตรฐานของการเทียบมาตราในแนวตั้งโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์อยู่ในระดับต่ำ ส่วนความเพิงพองของการเทียบมาตราพิจารณาจากค่าดัชนีความแตกต่าง พบว่าอยู่ในระดับที่น่าพอใจ

วรรณดี แสงประทีปทอง (2538 : 123-130) ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทย ของนักเรียนชั้นมัธยมศึกษาตอนต้น โดยใช้วิธีการเทียบมาตราในแนวตั้งตามทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาตอนต้น ในจังหวัดสิงห์บุรี ปีการศึกษา 2537 แบบสอบที่ใช้มี 4 ฉบับ คือ แบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทย ท 101, ท 203 และ ท 305 และแบบสอบร่วมภายนอกเป็นแบบสอบความสามารถทั่วไปทางภาษาไทย การเก็บรวบรวมข้อมูลใช้ในสถานการณ์จริงซึ่งเป็นการสอบปลายภาคเรียนที่ 1 ของนักเรียน ผลการศึกษาพบว่า ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 3 สูงกว่าชั้น ม. 2 และ ม. 3 ตามลำดับ แต่ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 1 และชั้น ม. 2 ไม่แตกต่างกัน ในกลุ่มนักเรียนที่มีความสามารถสูง ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 3 สูงกว่าชั้น ม. 1 แต่ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 3 และ ม. 2 ไม่แตกต่างกัน และผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 2 และชั้น ม. 1 ไม่แตกต่างกัน ในกลุ่มนักเรียนที่มีความสามารถปานกลาง ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 1, ม. 2 และชั้น ม. 3 ก้าวหน้าขึ้นตามลำดับ โดยผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 3 สูงกว่าชั้น ม. 2 และผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 2 สูงกว่าชั้น ม. 1 ส่วนในกลุ่มนักเรียนที่มีความสามารถต่ำ ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 3 สูงกว่าชั้น ม. 2 และชั้น ม. 1 ตามลำดับ แต่ผลสัมฤทธิ์ทางการเรียนของนักเรียนชั้น ม. 1, ม. 2 และชั้น ม. 3 ไม่แตกต่างกัน

พร โสภา พิณิจผล (2539 : 65-67) ศึกษาเปรียบเทียบผลการเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบโดยใช้วิธีวิเคราะห์ 2 วิธี คือ วิธีการเทียบคะแนนจริงโดยใช้แบบสอบร่วมและวิธีการเทียบคะแนนดิบโดยใช้แบบสอบร่วม จากการเปรียบเทียบค่าความคลาดเคลื่อนและความเพิงพองของการเทียบคะแนน กลุ่มตัวอย่างที่ใช้เป็นนักศึกษาพยาบาลศาสตร์ (ระดับต้น) ชั้นปีที่ 2 ปีการศึกษา 2538 ของวิทยาลัยพยาบาล ในสังกัดสถาบันพัฒนากำลังคนด้านสาธารณสุข จำนวน 980 คน แบบสอบที่ใช้เป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาการพยาบาลสูติศาสตร์ เรื่อง การพยาบาลในระยะตั้งครรภ์ จำนวน 2 ฉบับ ในแต่ละฉบับมีแบบสอบร่วมภายใน ผลการศึกษาพบว่า คะแนนความคลาดเคลื่อนที่ได้จากการเทียบคะแนนทั้งสองวิธีไม่แตกต่างกัน ส่วน

ความเพียงพอของการเทียบคะแนนทั้งสองวิธี อยู่ในเกณฑ์ที่น่าพอใจอย่างยิ่ง โดยวิธีการเทียบคะแนนจริงมีค่าความเพียงพอของการเทียบคะแนนสูงกว่าวิธีการเทียบคะแนนคิด

รุ่งนภา จันทรา (2540 : 71-75) ศึกษาเปรียบเทียบความคลาดเคลื่อนและตรวจสอบความเพียงพอของการเทียบมาตรฐานระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2539 ในจังหวัดเพชรบุรี จำนวน 925 คน แบบสอบที่ใช้เป็นแบบสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ จำนวน 2 ฉบับ การเทียบมาตรฐานทั้งสองรูปแบบใช้คะแนนจริงที่สามารถเดิยวกัน ผลการศึกษาพบว่า ความคลาดเคลื่อนของการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์ มีค่าต่ำกว่าการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์ การตรวจสอบความเพียงพอของการเทียบมาตรฐานโดยกลุ่มทานผลพบว่า ความเพียงพอของการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์อยู่ในระดับ ไม่น่าพอใจ ส่วนความเพียงพอของการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์อยู่ในระดับน่าพอใจ

จากงานวิจัยที่กล่าวมา สรุปได้ว่าการเทียบมาตรฐานในแนวคิดโดยใช้ทฤษฎีการตอบสนองข้อสอบจะให้ผลดีกว่าการเทียบโดยใช้ทฤษฎีดั้งเดิม แต่ยังไม่สามารถสรุปได้แน่ชัดว่าเมื่อใช้แบบจำลองหนึ่งและสามพารามิเตอร์ วิธีใดจะให้ผลที่ดีกว่ากัน ผู้วิจัยจึงต้องการศึกษาว่า เมื่อนำทฤษฎีการตอบสนองข้อสอบโลจิสติก โมเดลที่มีหนึ่งและสามพารามิเตอร์มาใช้ในการเทียบมาตรฐานในแนวคิด รูปแบบใดจะให้ความเพียงพอในการเทียบมาตรฐานมากกว่ากัน โดยมุ่งศึกษาในกรณีที่ใช้แบบสอบร่วมภายนอก และมีจำนวนข้อสอบไม่มากนัก เพื่อศึกษาถึงความเป็นไปได้ในการนำไปใช้จริง ในสถานศึกษาต่างๆ ต่อไป