บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การปรับเทียบคะแนนระหว่างแบบสอบ (Test Equating) เป็นศัพท์เฉพาะที่นักจิตมิติ (Psychometricians) นำมาใช้ในกระบวนการวัดและประเมินผล ในช่วงแรกนักวัดผลโดยส่วนมากไม่ได้ให้ความสนใจเท่าที่ควร จะมีการศึกษากันในเฉพาะกลุ่มนักจิตมิติเท่านั้น (Kolen and Brennan, 1995) ต่อมาในปี ค.ศ. 1980 ได้มีเอกสารเกี่ยวกับเรื่องนี้เพิ่มมากขึ้น ทำให้นักวัดผลเริ่มเห็นความสำคัญและนำมาใช้ให้เป็นประโยชน์กับการทดสอบ 15 ปี ต่อมาได้มีการเพิ่มจำนวนและความหลากหลายของโปรแกรมการสอบที่มีการใช้แบบสอบหลายฉบับ ผู้รับผิดชอบในการทำแบบสอบมีความต้องการที่จะอ้างอิงบทบาทของการปรับเทียบคะแนนเพื่อวิจารณ์การทดสอบ ความเคลื่อนไหวในการตรวจสอบและติดตามผลการจัดการศึกษา และประเด็นความยุติธรรมเกี่ยวกับการทดสอบที่ปรากฏชัดเจนขึ้น ทำให้ศาสตร์ด้านการปรับเทียบคะแนนได้รับการศึกษาและพัฒนาขึ้นมาตามลำดับ การเสนอเนื้อหาสำคัญในส่วนนี้ แยกนำเสนอเป็น 8 ตอน ดังนี้

ตอนที่ 1 แนวคิดเชิงทฤษฎีของการปรับเทียบคะแนนระหว่างแบบสอบ
ตอนที่ 2 ประเภทของการปรับเทียบคะแนน
ตอนที่ 3 ขั้นตอนการดำเนินการปรับเทียบคะแนน
ตอนที่ 4 การออกแบบวิธีการเก็บรวบรวมข้อมูล
ตอนที่ 5 รูปแบบการปรับเทียบคะแนนตามทฤษฎีการวัดแบบดั้งเดิม
ตอนที่ 6 การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ
ตอนที่ 7 เกณฑ์ที่ใช้ในการปรับเทียบเทียบคะแนน
ตอนที่ 8 งานวิจัยที่เกี่ยวข้องกับการปรับเทียบคะแนน

ตอนที่ 1 แนวคิดเชิงทฤษฎีของการปรับเทียบคะแนนระหว่างแบบสอบ

กัลลิคเซน (Gulliksen,1950) ได้กล่าวถึงการปรับเทียบคะแนนระหว่างแบบสอบว่าเป็นกระบวนการทำให้คะแนนจากแบบสอบต่างฉบับที่วัดวิชาเดียวกัน ให้เป็นคะแนนสมมูลกัน สามารถเปรียบเทียบกันได้โดยตรง โดยเสนอวิธีการให้ผู้สอบกลุ่มเดียวทำแบบสอบสองชุดปรับคะแนนแต่ละชุดให้เป็นคะแนนมาตรฐาน คะแนนที่ปรับแล้วนำมาเทียบกันได้โดยตรง แบบสอบที่นำมาปรับเทียบต้องตรวจสอบความเป็นคู่ขนานโดยใช้สถิติของวิคส์ (Wilks) หรือพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์แบบสอบแต่ละชุดกับเกณฑ์ ถ้าสัมประสิทธิ์มีค่าใกล้เคียงกัน ก็จะบ่งบอกความสามารถของความเพียงพอในการปรับเทียบคะแนนระหว่างแบบสอบ

ฟลานาแกน (Flanagan, 1951 cited in Angoff, 1984) ได้ให้ความหมายของการปรับเทียบคะแนนระหว่างแบบสอบว่าเป็นวิธีการที่ทำคะแนนจากแบบสอบต่างชุดให้มีคุณลักษณะที่เปรียบเทียบกันได้โดยแบบสอบทั้งสองฉบับ จะต้องมีการแจกแจงของคะแนนจริงในประชากรเท่ากัน หรือมีความเที่ยงเท่ากันเสียก่อน จึงสามารถนำคะแนนดิบมาเปรียบเทียบกันได้ จากความหมายดังกล่าวเป็นนิยามเชิงทฤษฎี ในทางปฏิบัติฟลานาแกนได้เสนอให้ตรวจสอบจากคะแนนเฉลี่ยของแบบสอบที่ได้จากกลุ่มตัวอย่างขนาดใหญ่ ถ้าคะแนนเฉลี่ยจากแบบสอบทั้งสองไม่แตกต่างกันอย่างมีนัยสำคัญ จึงสามารถนำคะแนนดิบมาเปรียบเทียบกันได้ แต่ถ้าคะแนนเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญ ควรสร้างแบบสอบทั้งสองให้มีคุณสมบัติของความเป็นคู่ขนาน แล้วเลือกวิธีการปรับเทียบคะแนนที่เหมาะสม ซึ่งเสนอไว้ 4 วิธี ดังนี้

1. ใช้ค่าเฉลี่ยโดยคำนวณค่าเฉลี่ยของการแจกแจงคะแนนทั้งสองชุด ถ้าความแตกต่างของค่าเฉลี่ยอยู่ภายในขอบเขตของการแปรผันเชิงสุ่มแล้ว ถือว่าคะแนนทั้งสองชุดเปรียบเทียบกันได้ แต่ถ้ามีความแตกต่างกันอย่างมีนัยสำคัญ ให้ใช้วิธีบวกเข้าหรือลบออกเท่ากับจำนวนที่แตกต่างจากคะแนนชุดที่หนึ่ง เพื่อให้เกิดคะแนนสมมูลกับอีกชุดหนึ่ง

2. ใช้เทคนิคของสมการถดถอย โดยหาค่าประมาณที่ดีที่สุดของคะแนนจากแบบสอบชุดที่หนึ่งซึ่งรู้ค่าของอีกชุดหนึ่ง

3. ใช้คะแนนมาตรฐานที่เท่ากันเป็นตัวปรับเทียบคะแนน

4. ใช้วิธีอิควิเปอร์เซ็นไทล์ โดยคะแนนสมมูลกันเมื่อคะแนนจากแบบสอบต่างฉบับมีตำแหน่งเปอร์เซ็นต์ไทล์เท่ากัน

ลอร์ด (Lord, 1980) ได้กล่าวถึงวัตถุประสงค์ของการปรับเทียบคะแนนระหว่างแบบสอบ เพื่อกำหนดความเท่าเทียมกันระหว่างคะแนนดิบจากแบบสอบต่างฉบับ และเนื่องจากวิธีการปรับเทียบคะแนนระหว่างแบบสอบเป็นวิธีการหาหลักฐานเชิงประจักษ์ จึงจำเป็นต้อง

กำหนดแบบแผนในการเก็บรวบรวมข้อมูลและกฎเกณฑ์ของการปรับคะแนนจากแบบสอบหนึ่งไปสู่แบบสอบอีกฉบับหนึ่ง ลอร์ดได้กำหนดเงื่อนไขของการปรับเทียบคะแนนซึ่งนักวัดผลหลายคนได้เห็นด้วยในกรณีที่ว่า แบบสอบฉบับ X และฉบับ Y จะสามารถนำมาปรับเทียบกันได้ก็ต่อเมื่อแบบสอบทั้งสองมีคุณสมบัติ 4 ประการ คือ

1. แบบสอบทั้งสองฉบับจะต้องวัดความสามารถเดียวกัน (Same ability) คือ แบบสอบทั้งสองวัดในคุณลักษณะเดียวกัน คุณลักษณะนี้อาจเป็นคุณลักษณะแฝง หรือความสามารถหรือทักษะ อย่างใดอย่างหนึ่งก็ได้

2. มีความเสมอภาค (Equity) คือ เมื่อทุกกลุ่มมีความสามารถเดียวกัน การแจกแจงคะแนนของแบบสอบ Y หลังจากที่มีการปรับคะแนนแล้วจะมีการแจกแจงเหมือนกับการแจกแจงของคะแนนจากแบบสอบ X

3. ประชากรไม่แปรเปลี่ยน (Population invariance) คือ การปรับคะแนนต้องเป็นไปในลักษณะเดียวกัน ไม่ว่าคะแนนจะมาจากกลุ่มตัวอย่างใดก็ตาม

4. มีความสมมาตร (Symmetry) คือ การปรับคะแนนสามารถแปรเปลี่ยนกลับได้เช่น การแปลงคะแนนจากฉบับ X ไปยังฉบับ Y มีผลเช่นเดียวกันกับฉบับ Y ไปสู่ฉบับ X

โดยสรุป การปรับเทียบคะแนนระหว่างแบบสอบเป็นกระบวนการแปลงระบบคะแนนของแบบสอบต่างฉบับ ที่มุ่งวัดเนื้อหาเดียวกัน โดยแปลงคะแนนของแบบสอบฉบับหนึ่งให้อยู่บนมาตรหรือสเกลของแบบสอบอีกฉบับหนึ่ง ที่ถือเป็นคะแนนที่สมมูลกัน เพื่อทำให้คะแนนจากแบบสอบต่างฉบับสามารถเปรียบเทียบกันได้โดยตรงและใช้แทนกันได้

ตอนที่ 2 ประเภทของการปรับเทียบคะแนน

ในการใช้เทคนิคการปรับเทียบคะแนนระหว่างแบบสอบตามเงื่อนไขและจุดมุ่งหมายของการทดสอบ จำแนกได้เป็น 2 สถานการณ์ (Hambleton and Swaminathan, 1985) ดังรายละเอียดต่อไปนี้

1. การปรับเทียบคะแนนตามแนวนอน (Horizontal Equating) เป็นการปรับเทียบคะแนนระหว่างแบบสอบสองฉบับที่มุ่งวัดคุณลักษณะเดียวกัน มีระดับความยากใกล้เคียงกัน และกลุ่มผู้สอบมีการแจกแจงความสามารถ อยู่ในกลุ่มประชากรเดียวกัน หรือมีความสามารถใกล้เคียงกัน เป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบสอบเนื้อหาเดียวกันขึ้นมาหลาย ๆ ฉบับเพื่อให้การทดสอบเกิดความยุติธรรมและป้องกันความลับของข้อสอบเมื่อใช้ต่างเวลากันสำหรับผู้สอบกลุ่มขนาดใหญ่ ปรับเทียบว่าคะแนนที่ได้จากฉบับหนึ่งเทียบเป็น

เท่าไรของอีกฉบับหนึ่ง ซึ่งวัดในระดับความสามารถเดียวกัน นำไปใช้ในการปรับเทียบคะแนนระหว่างแบบสอบวิชาเดียวกันต่างฉบับกัน เช่น การปรับเทียบคะแนนระหว่างแบบสอบวิชาวิทยาศาสตร์ 2 ฉบับ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

แบบสอบที่นำมาปรับเทียบคะแนนตามแนวนอนนั้น ควรเป็นแบบสอบที่พยายามสร้างหรือมุ่งให้คู่ขนาน(alternate forms) นอกจากนี้ต้องตรวจสอบการแจกแจงความสามารถของผู้สอบ ทั้งสองกลุ่มควรมีการแจกแจงที่ประมาณได้ว่ามีระดับความสามารถที่ใกล้เคียงกัน

2. การปรับเทียบคะแนนตามแนวตั้ง เป็นการปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน แต่มีระดับความยากแตกต่างกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่คนละกลุ่มประชากรหรือมีความสามารถแตกต่างกัน เหมาะกับสถานการณ์ที่มีความจำเป็นต้องสร้างแบบสอบต่างฉบับวัดเนื้อหาเดียวกัน มุ่งวัดความสามารถของผู้สอบที่ต่างระดับกัน เพื่อปรับเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่งเทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน นำไปใช้ในการปรับเทียบคะแนนระหว่างแบบสอบวิชาเดียวกันแต่ต่างระดับชั้น เช่น การปรับเทียบคะแนนระหว่างแบบสอบวิชาวิทยาศาสตร์ ชั้น ม.1, ม.2 และ ม.3

ตอนที่ 3 ขั้นตอนดำเนินการปรับเทียบคะแนน

การปรับเทียบคะแนนจะเป็นประโยชน์กับการจัดการศึกษา ต้องมีขั้นตอนดำเนินการ ดังต่อไปนี้ (Kolen and Brennan, 1995)

1. กำหนดจุดมุ่งหมายในการปรับเทียบคะแนนว่าจะนำผลจากการปรับเทียบเพื่อดูพัฒนาการเปลี่ยนแปลงทางการศึกษา หรือเพื่อเทียบความสามารถผู้สอบเป็นข้อมูลใช้ตัดสินผลร่วมกัน ใช้แทนกันได้

2. สร้างแบบสอบหลายฉบับ แต่ละฉบับวัดเนื้อหาเดียวกัน และสร้างตามแบบแผนการออกข้อสอบเดียวกัน (Item Specification) เป็นแบบสอบที่มีลักษณะของความเป็นคู่ขนานในด้าน เนื้อเรื่อง โครงสร้าง รูปแบบ ชนิดของข้อสอบ และเวลาที่ใช้สอบ

3. เลือกวิธีการเก็บรวบรวมข้อมูล มีรูปแบบกลุ่มสุ่ม (Random Groups Design) โดยสุ่มกลุ่มตัวอย่างจากประชากรเดียวกัน แต่ละกลุ่มทำแบบสอบคนละชุด รูปแบบผู้สอบกลุ่มเดียว (Single- Group Design) ผู้สอบกลุ่มเดียวทำแบบสอบทั้งสองชุด รูปแบบกลุ่มเดียวที่ได้รับการจัดให้สมดุล (Single–Group Design with Counterbalancing) แบ่งผู้สอบแต่ละกลุ่มแบบสอบแต่ละชุดเป็น 3 ส่วน ให้ผู้สอบกลุ่มย่อยแรกสอบแบบสอบชุดที่ 1 ตอนแรก ตามด้วยแบบสอบชุดที่ 2 ตอนหลัง และผู้สอบกลุ่มย่อยที่ 2 ทำแบบสอบชุดที่ 2 ตอนแรก ตามด้วยแบบ

สอบชุดที่ 1 ตอนหลัง และรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Common-Item Nonequivalent Groups Design) ผู้สอบต่างกลุ่มประชากรทำแบบสอบคนละชุด และผู้สอบทุกคนทำแบบสอบร่วมที่อาจจะแทรกภายในแบบสอบทั้งสองฉบับ (Internal Common Item) หรือแยกออกจากแบบสอบ (External Common Item)

4. เก็บรวบรวมข้อมูลตามรูปแบบที่กำหนดไว้

5. เลือกนิยามเชิงปฏิบัติการของการปรับเทียบคะแนน เพื่อตัดสินใจว่าจะใช้วิธีการปรับเทียบคะแนนเชิงเส้นตรง (Linear Equating Methods) หรือวิธีการปรับเทียบที่ไม่เป็นเชิงเส้นตรง (Nonlinear Equating Methods)

6. เลือกวิธีประมาณค่าสถิติที่ใช้วิเคราะห์ เลือกให้สอดคล้องกับนิยามเชิงปฏิบัติการที่กำหนด มีวิธีการปรับจากค่าเฉลี่ย (Mean Equating) โดยพิจารณาคะแนนสมมูลกันเมื่อคะแนนจากแบบสอบต่างฉบับเบี่ยงเบนไปจากคะแนนเฉลี่ยเท่ากัน วิธีปรับเทียบเชิงเส้นตรง (Linear Equating) พิจารณาคะแนนสมมูลกันเมื่อคะแนนจากแบบสอบต่างฉบับมีคะแนนมาตรฐานเท่ากัน วิธีการปรับเทียบอิควิเปอร์เซ็นต์ไทล์ (Equipercentile Equating) ที่คะแนนสมมูลกันเมื่อคะแนนจากแบบสอบต่างฉบับมีตำแหน่งเปอร์เซ็นต์ไทล์เท่ากัน และวิธีการปรับเทียบคะแนนโดยใช้สมการถดถอย (Regression Equating) เป็นการสร้างสมการทำนายคะแนนคะแนนจากแบบสอบชุดหนึ่งไปยังอีกชุดหนึ่ง หรือได้จากคะแนนสมมูลกันเมื่อคะแนนของแบบสอบแต่ละฉบับทำนายคะแนนเกณฑ์ได้เท่ากัน ทั้ง 4 วิธีนี้ เป็นวิธีการปรับเทียบคะแนนตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ส่วนวิธีการปรับเทียบตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เป็นการหาสัมประสิทธิ์การปรับเทียบ หรือค่าความชัน (Slope) และ ค่าคงที่ ของฟังก์ชันเชิงเส้นตรงที่เป็นความสัมพันธ์ของการปรับเทียบคะแนน วิธีการหาสัมประสิทธิ์การปรับเทียบ มีวิธีใช้ค่าเฉลี่ยของค่าอำนาจจำแนกและค่าเฉลี่ยของค่าความยากของข้อสอบ (Mean and Mean Method) วิธีใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบ (Mean and σ Method) วิธีการทำให้ความแตกต่างระหว่างคะแนนจริงเดิมกับคะแนนจริงที่ปรับแล้วมีค่าน้อยที่สุด โดยใช้สถิติ F-test (Characteristic Curve Method) และวิธีทำให้ความแตกต่างระหว่างพารามิเตอร์ข้อสอบเดิมกับที่ปรับแล้วมีค่าน้อยที่สุด โดยใช้สถิติ χ^2- test (Minimum χ^2)

7. ประเมินผลการปรับเทียบคะแนน มีเกณฑ์ที่ใช้ดังนี้ (Harris and Crouse, 1993) ความเสมอภาค (Weak Equity) ของ Divgi และ Yen ที่พิจารณาจากความเท่าเทียมกันของการแจกแจงตามเงื่อนไขของคะแนนที่ได้จากแบบสอบต่างฉบับหลังจากการปรับเทียบแล้ว ดัชนีสำหรับการเปลี่ยนแปลงคะแนน (Indices) ของ Angoff ความคลาดเคลื่อนมาตรฐาน

(Standard Error) ของ Angoff เป็นการวิเคราะห์เพื่อประมาณความคลาดเคลื่อนของการปรับเทียบจากการสุ่มตัวอย่าง ข้อมูลที่จำลองขึ้น (Generated Data) ของ Lord เพื่อใช้สำหรับการปรับเทียบคะแนน การปรับเทียบคะแนนจากแบบสอบกลับสู่แบบสอบเดิม (Equating a Test to Itself) ของ Lord เป็นการปรับเทียบคะแนนกลับสู่แบบสอบเดิมโดยตรง หรือปรับผ่านแบบสอบอื่นก่อนปรับกลับสู่แบบสอบเดิม กลุ่มตัวอย่างขนาดใหญ่ (Large Sample) ของ Angoff เป็นการใช้กลุ่มตัวอย่างขนาดใหญ่ในการปรับเทียบคะแนนซึ่งคล้ายกับการปรับเทียบคะแนนจากประชากร และใช้เปรียบเทียบกับผลการปรับเทียบคะแนนที่มีขนาดกลุ่มตัวอย่างน้อยกว่า ความคงเส้นคงวา (Consistency) เป็นการประเมินผลการปรับเทียบคะแนนข้ามวิธี เพื่อหาความคงเส้นคงวา ความคงที่ (Stability) ของ Angoff เป็นการปรับเทียบคะแนนซ้ำเพื่อตรวจความคงที่ของผลการปรับเทียบคะแนน

เงื่อนไขที่ส่งเสริมให้การปรับเทียบคะแนนเป็นที่น่าพอใจมีรายละเอียดดังนี้ (Kolen and Brennan, 1995)

1. ลักษณะทั่วไป

1.1 เป้าหมายของการที่นำคะแนนมาปรับเทียบมีการระบุอย่างชัดเจน

1.2 การออกแบบเก็บรวบรวมข้อมูล การวางแผนเชื่อมโยงการปรับเทียบคะแนน วิธีการทางสถิติที่ใช้ และการเลือกผลที่ได้จากการปรับเทียบหลายวิธี มีความเหมาะสมกับความสำเร็จของการปรับเทียบคะแนน

1.3 มีการติดตามและควบคุมคุณภาพของกระบวนการปรับเทียบคะแนนอย่างเพียงพอ

2. การพัฒนาแบบสอบสำหรับการเก็บรวบรวมข้อมูลทุกรูปแบบ

2.1 กำหนดเนื้อหาแบบสอบและกำหนดแบบแผนทางสถิติข้อสอบเป็นอย่างดีมีความคงที่

2.2 การสร้างแบบสอบเพื่อนำมาปรับเทียบพิจารณาจากสถิติข้อสอบ เช่น ค่าความยาก ค่าอำนาจจำแนก จากการสอบครั้งก่อน

2.3 ข้อสอบที่ใช้มีความยาวพอเหมาะ อย่างน้อยมีจำนวน 30 ข้อ หรือมากกว่าตัวเฉลยสำหรับตรวจให้คะแนนมีความคงที่เมื่อใช้กับแบบสอบแบบเลือกตอบ

3. การพัฒนาแบบสอบสำหรับรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม

3.1 ข้อสอบร่วมต้องเป็นตัวแทนของแบบสอบที่ใช้ในการปรับเทียบ ทั้งคุณลักษณะของเนื้อหาและค่าสถิติของข้อสอบ

3.2 ข้อสอบร่วมมีจำนวนพอเหมาะ อย่างน้อย 20 % ของแบบสอบที่นำมาปรับเทียบที่มีจำนวน 40 ข้อ หรือมากกว่า และอย่างน้อย 30 ข้อ เมื่อแบบสอบที่นำมาปรับ

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

เทียบมีขนาดใหญ่

3.3 ข้อสอบร่วมแต่ละข้อสร้างให้มีลักษณะที่ใกล้เคียงกับแบบสอบที่นำมาปรับเทียบ ทั้งตัวคำถามและตัวเลือก

4. กลุ่มตัวอย่าง

4.1 กลุ่มตัวอย่างมีลักษณะเป็นตัวแทน

4.2 กลุ่มตัวอย่างมีความคงที่

4.3 กลุ่มตัวอย่างมีขนาดใหญ่เพียงพอ

4.4 กลุ่มตัวอย่าง 2 กลุ่ม สำหรับกลุ่มผู้สอบไม่เท่าเทียมกันใช้แบบสอบร่วม มีความสามารถไม่แตกต่างกันมากนัก

5. การบริหารการสอบ

5.1 ดำเนินการทดสอบอย่างปลอดภัย

5.2 การสอบในแต่ละครั้งดำเนินการอย่างระมัดระวัง และจัดสภาพการสอบให้เหมือนกัน

6. หลักสูตร เนื้อหาการปฏิบัติ หรือสาระที่นำมาใช้ในการปรับเทียบคะแนนมีความคงที่

ผลของการปรับเทียบคะแนนมีความเชื่อถือได้ หรือมีความถูกต้องแม่นยำขึ้นอยู่กับการเลือกสถิติที่ใช้ในการปรับคะแนนให้มีความเหมาะสมกับแบบแผนการออกข้อสอบ การเก็บรวบรวมข้อมูล และการดูแลการสอบที่เป็นไปอย่างมีมาตรฐาน

ตอนที่ 4 การออกแบบวิธีการเก็บรวบรวมข้อมูล (Designs for Data Collection)

ในการปรับเทียบคะแนนระหว่างแบบสอบแต่ละครั้งจำเป็นต้องมีการออกแบบวิธีการเก็บรวบรวมข้อมูล ให้มีความสอดคล้องและเหมาะสมกับสถิติที่ใช้ในการปรับเทียบคะแนน สามารถจำแนกเป็นแบบแผนการเก็บรวบรวมข้อมูลได้ 4 รูปแบบ (Kolen and Brennan, 1995) ดังนี้

1. รูปแบบผู้สอบกลุ่มเดียว (Single Group Design) รูปแบบนี้ต้องนำแบบสอบสองฉบับไปสอบกับกลุ่มผู้สอบเพียงกลุ่มเดียว ดังนั้นกลุ่มผู้สอบแต่ละกลุ่มจะได้รับการสอบทั้งสองฉบับ ระดับความยากของแบบสอบไม่มีส่วนเกี่ยวข้องกับระดับความสามารถของผู้สอบ อย่างไรก็ตามในรูปแบบนี้ความเมื่อยล้าของผู้สอบอาจมีผลต่อการปรับเทียบคะแนน

2. รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล (Single Group Design with Counterbalancing) รูปแบบนี้ปรับมาจากรูปแบบประเภทกลุ่มเดียว เพื่อขจัดผลของลำดับก่อนหลัง โดยจัดแบ่งผู้สอบออกเป็นสองกลุ่มย่อย และแบ่งแบบสอบทั้งชุด X และชุด Y ออกเป็น

สองส่วน จัดให้กลุ่มผู้สอบย่อยกลุ่มที่ 1 ทำแบบสอบย่อยชุด X ตอนแรก ตามด้วยแบบสอบย่อยชุด Y ตอนหลัง และกลุ่มผู้สอบย่อยกลุ่มที่ 2 ทำแบบสอบย่อยชุด Y ตอนหลัง ตามด้วยแบบสอบย่อยชุด X ตอนแรก

การใช้รูปแบบนี้ กลุ่มผู้สอบจะได้รับอิทธิพลจากลำดับการทดสอบ และความเมื่อยล้าที่สมดุลกันในทั้งสองกลุ่มย่อย

3. รูปแบบกลุ่มสุ่ม (Random Group Design) รูปแบบนี้เป็นการใช้กลุ่มเท่าเทียมกัน กลุ่มที่เลือกมาทำแบบสอบจะเลือกมาโดยการสุ่ม สุ่มแบบสอบเข้ากลุ่มผู้สอบแต่ละกลุ่ม เป็นการแก้ปัญหาความเมื่อยล้าจากการทำแบบสอบได้ อย่างไรก็ตามเนื่องจากกลุ่มผู้สอบไม่เป็นกลุ่มเดียวกัน อาจจะมีความสามารถแตกต่างกันเล็กน้อย จึงอาจเกิดความลำเอียงเกิดขึ้นกับการปรับเทียบคะแนนได้

4. รูปแบบผู้สอบต่างกลุ่มโดยใช้แบบสอบร่วม (Common-Item Nonequivalent Design) รูปแบบนี้เป็นการใช้กลุ่มผู้สอบซึ่งมีลักษณะไม่เท่าเทียมกัน แล้วให้แต่ละกลุ่มทำแบบสอบเพียงฉบับเดียว โดยแบบสอบแต่ละฉบับจะมีข้อสอบร่วมที่เป็นชุดข้อสอบที่จัดอยู่ภายในแบบสอบทั้งสองฉบับเรียกแบบสอบร่วมภายใน (Internal Common Test) หรือ แยกออกเป็นฉบับต่างหากเรียกแบบสอบร่วมภายนอก (External Common Test) แบบสอบร่วมนี้จะนำไปทดสอบทั้งสองกลุ่มพร้อมกับแบบสอบที่ต้องการนำมาปรับเทียบ ตามรูปแบบนี้จะช่วยแก้ปัญหาต่าง ๆ ที่เกิดขึ้นกับสามรูปแบบข้างต้น

ตอนที่ 5 รูปแบบการปรับเทียบคะแนนตามทฤษฎีการวัดแบบดั้งเดิม

การปรับเทียบคะแนนตามทฤษฎีการวัดแบบดั้งเดิมจาก จำแนกเป็น 5 วิธีดังนี้

1. วิธีการปรับเทียบค่าเฉลี่ย (Mean Equating)

วิธีนี้ใช้สำหรับรูปแบบการเก็บรวบรวมข้อมูลแบบกลุ่มสุ่ม เมื่อมีการปรับเทียบคะแนนสังเกตได้ (Observed Score Equating) เป็นการพิจารณาค่าความยากของแบบสอบชุดที่ 1 ต่างจากชุดที่ 2 คงที่ตลอดตั้งแต่คะแนนน้อยที่สุดจนถึงคะแนนมากที่สุด เช่นคะแนนจากแบบสอบชุดที่ 1 ง่ายกว่าชุดที่ 2 อยู่ 3 คะแนนในกลุ่มผู้ที่ได้คะแนนสูง และง่ายกว่าในกลุ่มต่ำ 3 คะแนนเช่นเดียวกัน มีกระบวนการปรับเทียบคะแนนดังนี้

กำหนดแบบสอบ X เป็นแบบสอบใหม่ที่ต้องการปรับเทียบ X เป็นคะแนนที่ได้จากแบบสอบ X ทั้งหมด และ x แทน คะแนนเฉพาะที่ต้องการปรับเทียบ และ Y เป็นแบบสอบชุดเดิม Y แทนคะแนนที่ได้จากแบบสอบ Y ทั้งหมด และ y แทนคะแนนเฉพาะจากแบบ

สอบ Y ให้ $\mu(X)$ เป็นค่าเฉลี่ยของคะแนนที่ได้จากแบบสอบ X และ $\mu(Y)$ เป็นค่าเฉลี่ยของคะแนนที่ได้จากแบบสอบ Y

ความหมายของการปรับเทียบคะแนนคือ คะแนนจาก 2 แบบสอบ มีความแตกต่างจากค่าเฉลี่ยของแต่ละกลุ่มเท่ากัน กำหนดโดย

$$x - \mu(X) = y - \mu(Y)$$

จะได้ $m_y(x) = y = x - \mu(X) + \mu(Y)$

เมื่อ $m_y(x)$ แทน คะแนน x จากแบบสอบ X แปลงไปสู่สเกลของแบบสอบ Y โดยใช้วิธีการปรับเทียบค่าเฉลี่ย

ตัวอย่างเช่น เมื่อคะแนนเฉลี่ยของแบบสอบ X เท่ากับ 72 ค่าเฉลี่ยของแบบสอบ Y เท่ากับ 77 เมื่อแทนค่าในสมการข้างบน ได้ดังนี้

$$m_y(x) = x - 72 + 77 = x + 5$$

คะแนนเฉลี่ยของแบบสอบ X เป็น 72 คะแนน เมื่อพิจารณาแล้วจะอยู่ในระดับความสามารถเท่ากับ 77 คะแนนในแบบสอบ Y และคะแนนจากแบบสอบ X 75 คะแนน จะอยู่ในระดับความสามารถเดียวกันกับแบบสอบชุด Y เท่ากับ 80 คะแนน

2. วิธีปรับเทียบคะแนนเชิงเส้นตรง

วิธีปรับเทียบคะแนนเชิงเส้นตรงเป็นวิธีที่ยึดถือนิยามของการปรับเทียบคะแนนว่า คะแนนจากแบบสอบ 2 ฉบับ X และ Y เมื่อแบบสอบ X และ Y วัดสิ่งเดียวกันและมีความเที่ยงใกล้เคียงกัน จะถือว่าคะแนนสมมูลกัน (Equivalent Score) เมื่อคะแนนของแต่ละฉบับมีคะแนนมาตรฐาน (Standard Score) เท่ากัน

วิธีปรับเทียบคะแนนเชิงเส้นตรงนี้ อาจกล่าวได้ว่าเป็นกรณีเฉพาะของวิธีปรับเทียบเฉพาะของวิธีปรับเทียบคะแนนอีควิเปอร์เซ็นไทล์ เนื่องจากว่าถ้าการแจกแจงคะแนนจากแบบสอบ X และ Y เหมือนกัน วิธีอีควิเปอร์เซ็นไทล์ให้ผลการปรับเทียบคะแนนระหว่างแบบสอบเป็นเส้นตรง ซึ่งเหมือนกับวิธีปรับเทียบเชิงเส้นตรง แต่ถ้าการแจกแจงคะแนนจากแบบสอบ X และ Y แตกต่างกัน และแบบสอบทั้งสองมีความยากแตกต่างกัน ความสัมพันธ์ระหว่างคะแนนสอบจะมีลักษณะเป็นเส้นโค้ง วิธีอีควิเปอร์เซ็นไทล์น่าจะให้ผลการปรับเทียบคะแนนที่ดีกว่า วิธีปรับเทียบคะแนนเชิงเส้นตรงไม่ค่อยเหมาะสำหรับการปรับเทียบคะแนนตามแนวตั้ง (Vertical Equating) เนื่องจากการปรับเทียบดังกล่าวมีกลุ่มตัวอย่างที่ระดับความสามารถไม่

เท่ากัน อีกทั้งการแจกแจงคะแนนอาจมีรูปทรงที่แตกต่างกันได้ทั้งในแง่ความแบ้ (Skewness) และความโด่ง (Kurtosis)

การใช้วิธีปรับเทียบคะแนนเชิงเส้นตรง มีข้อแนะนำดังนี้

1) เนื่องจากมีข้อตกลงเบื้องต้นของการปรับเทียบคะแนนเป็นแบบเส้นตรง จึงควรตรวจสอบความสอดคล้องของการแจกแจงก่อนปรับเทียบและหลังการปรับเทียบ ว่าเป็นความสัมพันธ์เชิงเส้นตรง

2) ถ้าความยากของแบบสอบที่ต้องการนำมาปรับเทียบคะแนนกันมีความแตกต่างกันการแจกแจงคะแนนอาจมีความสัมพันธ์เชิงเส้นโค้ง การใช้วิธีปรับเทียบคะแนนเชิงเส้นตรงคงไม่เหมาะสม การเลือกใช้วิธีปรับเทียบคะแนนอีควิเปอร์เซ็นไทล์น่าจะให้ผลดีกว่า

3. วิธีการปรับเทียบคะแนนโดยใช้สมการถดถอย (Regression Method)

เป็นวิธีที่ใช้เทคนิคการวิเคราะห์สมการถดถอย (Regression Equation) สำหรับปรับเทียบคะแนนระหว่างแบบสอบ การปรับเทียบคะแนนโดยวิธีนี้เป็นการใช้สมการเส้นตรงในการทำนายตัวแปรตาม (Dependent Variable) จากตัวแปรอิสระ (Independent Variable) ซึ่งมีลักษณะที่ไม่สมมาตร (Symmetry) กล่าวคือ สมการทำนาย Y จาก X หรือการทำนาย X จาก Y เป็นสมการที่ไม่สมมาตร หรือให้ผลทำนายไปในทิศทางเดียว นอกจากนี้คะแนนจากแบบสอบที่ใช้เป็นตัวทำนายยังมีข้อตกลงเบื้องต้นว่าจะต้องมีค่าความเที่ยงเป็น 1

การนำวิธีปรับเทียบคะแนนวิธีนี้ไปใช้ มีข้อแนะนำดังนี้

1) แบบสอบที่ต้องการนำมาปรับเทียบคะแนนกัน ต้องเป็นแบบสอบที่วัดคุณลักษณะเดียวกัน มีความเป็นคู่ขนาน และมีความเที่ยงสูง

2) แบบสอบที่ต้องการนำมาปรับเทียบคะแนนกัน ต้องมีความสัมพันธ์กับคะแนนเกณฑ์เท่าเทียมกัน มิเช่นนั้นแล้วจะทำให้คะแนนจากแบบสอบฉบับหนึ่งจะสามารถทำนายคะแนนเกณฑ์ได้แม่นยำกว่าอีกฉบับหนึ่ง ทำให้การปรับเทียบคะแนนมีความผันแปรตามกลุ่มที่ศึกษา

4. การปรับเทียบอิควิเปอร์เซ็นไทล์ (Equipercentile Equating)

การปรับเทียบคะแนนแบบอีควิเปอร์เซ็นไทล์ใช้นิยามของการปรับเทียบว่าคะแนนจากแบบสอบ X และแบบสอบ Y จะมีความเท่าเทียมกันก็ต่อเมื่อคะแนนทั้งสองอยู่ที่ตำแหน่งเปอร์เซ็นไทล์เดียวกัน และลักษณะการแจกแจงคะแนนของประชากรผู้สอบจะต้องมีลักษณะเหมือนกัน

วิธีอีควิเปอร์เซ็นไทล์เป็นการสร้างคะแนนสมมูลระหว่างแบบสอบบนพื้นฐานความเท่าเทียมกันของค่าโมเมนต์ทั้ง 3 ทางสถิติ ซึ่งได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการแจกแจงของคะแนน จึงเป็นกระบวนการแปลงคะแนนแบบไม่ใช่เส้นตรง (Nonlinear Transformation) ยังผลให้ความสัมพันธ์ระหว่างคะแนนดิบและคะแนนแปลงไม่เป็นเส้นตรงใน

การนำวิธีปรับเทียบคะแนนนี้ไปใช้มีข้อแนะนำพอสรุปได้ดังนี้ (Angoff, 1984)

1) ควรใช้กลุ่มผู้สอบขนาดใหญ่ ควรเป็นกลุ่มที่มีความสามารถค่อนข้างกระจายและกระจายพอ ๆ กัน ถ้าใช้กลุ่มตัวอย่างขนาดเล็กทำให้คะแนนมีความไวต่อความแปรปรวนเชิงสุ่ม ผลการปรับเทียบคะแนนสามารถแปรผันไปตามกลุ่มผู้สอบได้

2) การสร้างกราฟเปอร์เซ็นไทล์ เพื่อปรับเทียบคะแนนระหว่างแบบสอบควรสร้างด้วยความระมัดระวังและไม่ลำเอียง

3) แบบสอบที่ต้องการนำมาปรับเทียบคะแนนกัน ควรมีความเที่ยงใกล้เคียงกัน ถ้าต่างกันมาก จะทำให้การปรับคะแนนขาดความคงที่

4) การปรับเทียบคะแนนควรอยู่ในช่วงพิสัยของคะแนนสังเกต การปรับเทียบคะแนนที่อยู่นอกพิสัยของคะแนนสังเกตจะมีความคลาดเคลื่อนสูง

สำหรับการวิเคราะห์ในการปรับเทียบคะแนนอิควิเปอร์เซ็นต์ไทล์ มีขั้นตอนดังนี้

1) ทำกราฟตำแหน่งเปอร์เซ็นต์ไทล์ของคะแนนที่ได้จากแบบสอบแต่ละฉบับลงบนแกนเดียวกัน

2) หาจุดที่เท่าเทียมกันของกราฟทั้งสอง โดยพิจารณาจากตำแหน่งเปอร์เซ็นต์ไทล์เดียวกันจะตกกับคะแนน x เท่าใด คะแนน y เท่าใด

3) ทำกราฟแสดงความสัมพันธ์ของความเท่าเทียมกันของคะแนนโดยแกนหนึ่งเป็นคะแนนจากแบบสอบชุด X ส่วนอีกแกนหนึ่งเป็นคะแนนจากแบบสอบชุด Y ที่มีตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน

ตอนที่ 6 รูปแบบการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ

6.1 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีการวัดที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายใน หรือความสามารถที่มีอยู่ภายในตัวบุคคลกับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้นว่ามีโอกาสตอบข้อสอบถูกมากน้อยเพียงใด เมื่อนำมาประยุกต์ใช้กับการปรับเทียบคะแนน แบบสอบที่นำมาใช้ต้องมีคุณลักษณะตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

แฮมเบิลตันและสวามินาธาน (Hambleton and Swaminathan, 1985) ได้เสนอข้อตกลงเบื้องต้น ของทฤษฎีการตอบสนองข้อสอบ (IRT) ไว้ 4 ประการ คือ

6.1.1 ความเป็นเอกมิติ (Unidimentional)

โดยทั่วไปมีข้อสมมุติว่ามีเพียงลักษณะเดียวหรือความสามารถเดียวเท่านั้นที่จำเป็นต่อการอธิบายคะแนนที่ได้จากการตอบของผู้สอบ ตามโมเดลของการตอบสนองรายข้อ

กำหนดว่า การที่มีความสามารถเดียวหรือคุณลักษณะเดียวคือความเป็นเอกมิติ โดยทั่วไปแล้ว ข้อตกลงข้อนี้เป็นไปได้ค่อนข้างยาก เนื่องจากมีปัจจัยที่มีผลต่อคะแนนสอบ เช่น ปัจจัยด้านความรู้ความเข้าใจ (Cognitive) บุคลิกภาพ และปัจจัยเกี่ยวกับการจัดการสอบ ปัจจัยเหล่านี้อาจรวมถึงแรงจูงใจ ความวิตกกังวลในการสอบ ความสามารถในการทำงานได้รวดเร็ว ความรู้เกี่ยวกับการใช้กระดาษคำตอบ เมื่อเป็นเช่นนี้สิ่งที่ทำให้ข้อตกลงนี้เป็นไปได้ คือการพิจารณาว่าแบบสอบฉบับนั้นมีองค์ประกอบใดหรือปัจจัยใดที่เด่นที่สุด ก็ถือว่าแบบสอบได้วัดในสิ่งนั้น

จากฟังก์ชันความน่าจะเป็นในการตอบข้อสอบถูก ณ ระดับความสามารถต่าง ๆ (θ) ที่แสดงด้วยฟังก์ชันโลจิสติกหรือโอไจฟ์ปกติ ฟังก์ชันทั้งสองยืนยันว่าความน่าจะเป็นในการทำข้อสอบแต่ละข้อถูกขึ้นอยู่กับพารามิเตอร์ของข้อสอบ 3 ตัว ถ้าเราต้องการทราบความสามารถของคนเพื่อกำหนดความน่าจะเป็นที่จะทำข้อสอบถูกในข้อใดข้อหนึ่ง ก็จะไม่ช่วยให้ทราบความน่าจะเป็นที่ทำข้อสอบข้ออื่น ๆ ถูก ถ้าข้อสอบที่เพิ่มขึ้นเข้าไปไม่ใช่การวัดความสามารถเดียวกับข้อสอบเดิม ผลการสอบจากข้อสอบเหล่านั้นจะขึ้นอยู่กับคุณลักษณะบางอย่างมากกว่าความสามารถของผู้สอบที่มุ่งวัด ดังนั้นแบบสอบตามแนวทฤษฎีนี้จึงควรมีเพียงมิติเดียว ซึ่งโดยทั่วไป ความเป็นมิติเดียวนี้ หมายถึง ข้อสอบทั้งหลายที่วัดเพียงเนื้อเดียวของความรู้หรือความสามารถ ข้อตกลงเกี่ยวกับความเป็นมิติเดียวนี้เป็นเรื่องที่ซับซ้อนและยุ่งยากมากที่สุด ของทฤษฎีการตอบสนองข้อสอบ ในการทดสอบถึงความเป็นมิติเดียวของแบบสอบส่วนใหญ่แล้วจะใช้วิธีการวิเคราะห์ตัวประกอบ แล้วสังเกตค่าไอเกน (Eigen Value) ของตัวประกอบที่นำมาลงจุดในกราฟกับอันดับของตัวประกอบ ถ้าค่าไอเกนในตัวประกอบแรกมากกว่าตัวประกอบตัวที่ 2 และค่าไอเกนของตัวประกอบที่เหลือมีค่าใกล้เคียงกันจะแสดงถึงความเป็นมิติเดียว

6.1.2 ความเป็นอิสระในการตอบข้อสอบ (Local Independence)

ข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระในการตอบข้อสอบ กล่าวว่าคำตอบของผู้สอบที่ตอบข้อสอบแต่ละข้อมีความเป็นอิสระจากกันอย่างมีนัยสำคัญ ข้อตกลงเบื้องต้นข้อนี้จะเป็นจริงก็ต่อเมื่อ คะแนนของการตอบข้อหนึ่งข้อใดจะไม่มีผลต่อคะแนนในการตอบข้ออื่น ๆ ตัวอย่างเช่น เนื้อหาในข้อสอบต้องไม่เป็นการแนะคำตอบสำหรับข้ออื่น เมื่อเกิดความเป็นอิสระเช่นนี้จริง ความน่าจะเป็นของแบบแผนการตอบข้อสอบแต่ละข้อจึงเป็นไปได้หลาย ๆ แบบ เช่น ความน่าจะเป็นที่เกิดขึ้นในแบบแผนการตอบของข้อสอบ 5 ข้อ $u = (1\ 0\ 1\ 1\ 0)$ เมื่อ 1 แทนคำตอบที่ถูกต้อง และ 0 แทนคำตอบที่ผิด มีค่าเท่ากับ $P_1 (1 - P_2) P_3 P_4 (1 - P_5)$ เมื่อ P_i เป็นความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบข้อที่ i ได้ถูกต้อง และ $(1 - P_i)$ เป็นความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบผิด เมื่อเป็นเช่นนี้การจัดเรียงข้อสอบในลักษณะต่าง ๆ จะไม่กระทบกระเทือนต่อผลการสอบ

ถ้าให้ U_i เมื่อ $i = 1, 2,\ldots,n$ แทนคำตอบแบบทวิวิภาค (ตอบถูกได้ 1 ตอบผิดได้ 0) ของผู้สอบที่ทำข้อสอบ n ข้อ P_i เป็นความน่าจะเป็นในการตอบถูกของข้อสอบ i และ $Q = 1 - P$ จากข้อตกลงเกี่ยวกับความเป็นอิสระจะได้

$$\text{Prob}[U_1 = u_1, U_2 = u_2, \ldots, U_n = u_n \mid \theta] = \text{Prob}[U_1 = u_1 \mid \theta]\ \text{Prob}[U_2 = u_2 \mid \theta] \ldots \text{Prob}[U_n = u_n \mid \theta]$$

ถ้ากำหนดว่า $P_i(\theta) = \text{Prob}[U_i = 1 \mid \theta]$ และ $Q_i(\theta) = \text{Prob}[U_i = 0 \mid \theta]$

จากนั้นจะได้ $\text{Prob}[U_1 = u_1, U_2 = u_2, \ldots, U_n = u_n \mid \theta]$

$$= P_1(\theta)^{u1} Q_1(\theta)^{1-u1} P_2(\theta)^{u2} Q_2(\theta)^{1-u2} \ldots P_n(\theta)^{un} Q_n(\theta)^{1-un}$$

$$= \prod_{i=1}^{n} P_i(\theta)^{ui} Q_i(\theta)^{1-ui}$$

จากสมการนี้หมายความว่าความน่าจะเป็นของแบบแผนการตอบของผู้สอบแต่ละคนมีค่าเท่ากับผลคูณของความน่าจะเป็นที่เกี่ยวข้องกับคำตอบของผู้สอบในแต่ละข้อ

ผลของข้อตกลงเบื้องต้นในเรื่องความเป็นอิสระนี้คือ ค่าความถี่ของคะแนนผู้สอบที่ระดับความสามารถใด ๆ เป็นดังนี้

$$f(x \mid \theta) = \sum \prod_{i=1}^{n} P_i(\theta)^{ui} Q_i(\theta)^{1-ui} \qquad (1)$$

เมื่อ x เป็นคะแนนจากการสอบซึ่งมีค่าระหว่าง 0 ถึง n

สิ่งหนึ่งที่อธิบายได้เกี่ยวกับข้อตกลงเบื้องต้นด้านความเป็นอิสระ ในกรณีที่เมื่อ θ มีความเป็นเอกมิติ หรือข้อตกลงเบื้องต้นเกี่ยวกับคุณลักษณะแฝงเป็นจริงคือ ประการแรกสมมุติว่าแบบวัดความสามารถร่วม สำหรับผู้สอบที่มีความสามารถ θ คำตอบที่ได้มามีความเป็นอิสระอย่างมีนัยสำคัญ แต่ถ้าไม่มีความเป็นอิสระ จะเกิดผลว่าผู้สอบบางคนจะมีคะแนนที่คาดหวังสูงกว่าผู้สอบคนอื่นที่มีความสามารถระดับเดียวกัน สิ่งที่ตามมาคือ จำเป็นต้องใช้ความสามารถหลาย ๆ อย่างมาใช้อธิบายคะแนนจากแบบสอบ ลักษณะนี้เป็นการละเมิดข้อตกลงด้านความเป็นเอกมิติ ประการที่สอง ข้อตกลงด้านความเป็นอิสระแสดงให้เห็นเป็นนัยว่า เมื่อเกิดความเป็นอิสระที่ระดับความสามารถใด ๆ จะมีเพียงหนึ่งความสามารถเท่านั้นที่อธิบายความสัมพันธ์ระหว่างชุดของข้อสอบในแบบสอบได้

เนื่องจากความเท่าเทียมกันระหว่างข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระและความเป็นเอกมิติของคุณลักษณะแฝง การตรวจสอบข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระสามารถทำได้โดยใช้ การวิเคราะห์องค์ประกอบ (Factor Analysis) และการตรวจสอบความมีนัยสำคัญทางสถิติสามารถทำได้โดยใช้ χ^2

6.1.3 โค้งคุณลักษณะข้อสอบ (Item Characteristic Curves)

เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ของความน่าจะเป็นในการทำข้อสอบได้ถูกกับระดับความสามารถของผู้ตอบ และเป็นฟังก์ชันการถดถอยในลักษณะที่ไม่เป็นเส้นตรง โค้งคุณลักษณะของข้อสอบไม่มีการเปลี่ยนแปลงในกลุ่มประชากร กล่าวคือสำหรับคุณลักษณะแฝงของกลุ่มที่ผู้สอบที่ระดับความสามารถใด ๆ จะไม่แปรเปลี่ยนถ้าการแจกแจงมีลักษณะเหมือนกัน จากนั้นโค้งที่เชื่อมโยงค่าเฉลี่ยของการแจกแจงเหล่านี้ต้องมีลักษณะเหมือนกันด้วย นั่นคือโค้งคุณลักษณะข้อสอบจึงไม่แปรเปลี่ยนทั้งกลุ่มประชากร

6.1.4 ความเร็วในการทำแบบสอบ

ข้อตกลงเบื้องต้นในข้อนี้แสดงเป็นนัยว่า โมเดลการตอบสนองรายข้อจะมีความเหมาะสมก็ต่อเมื่อการสอบไม่อยู่ภายใต้เงื่อนไขของความเร็วในการทำแบบสอบ กล่าวคือผู้สอบที่ทำข้อสอบผิดจะต้องมีสาเหตุมาจากความสามารถของเขาเท่านั้น ไม่ได้มีสาเหตุมาจากการทำข้อสอบไม่ทัน ข้อตกลงเบื้องต้นนี้จำเป็นต้องมีเพราะเป็นส่วนที่แฝงอยู่ในข้อตกลงเบื้องต้นเกี่ยวกับความเป็นเอกมิติ เนื่องจากเมื่อความเร็วในการสอบเข้ามามีส่วนเกี่ยวข้องในการสอบ หมายถึงมีคุณลักษณะหรือความสามารถอย่างน้อย 2 ประการที่เข้ามาเกี่ยวข้องกับการสอบ ได้แก่ ความเร็ว กับคุณลักษณะที่วัดโดยเนื้อหาของข้อสอบ การที่จะพิจารณาว่าแบบสอบเป็นแบบสอบที่เกี่ยวข้องกับความเร็วในการสอบหรือไม่ อาจทำได้โดยการนับจำนวนผู้ที่ทำข้อสอบไม่ครบทุกข้อก็จะสามารถทราบได้

ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เป็นทฤษฎีทางการวัดผลที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะภายในของบุคคลกับพฤติกรรมการตอบสนองข้อสอบแต่ละข้อว่ามีความน่าจะเป็นในการตอบถูกเพียงใด ทฤษฎีนี้ตั้งอยู่บนหลักพื้นฐานที่สำคัญ 2 ประการ (Hambleton, Swarminathan and Roger, 1991) คือ

1) ความสามารถของบุคคลในการตอบข้อสอบได้ถูกหรือผิด สามารถอธิบายได้ด้วยคุณลักษณะภายในหรือความสามารถ (Latent Trait or Ability) ของบุคคลนั้น ๆ

2) ความสัมพันธ์ระหว่างความสามารถในการตอบข้อสอบได้ถูกต้องกับความ

สามารถของผู้สอบที่วัดจากแบบสอบสามารถอธิบายได้ด้วยโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC)

พัฒนาการของทฤษฎีนี้เริ่มมาจากการที่ ลอว์ลี (Lawley) ได้เสนอบทความเกี่ยวกับปัญหาการสร้างและการเลือกข้อสอบ ในปี ค.ศ.1943 (ศิริชัย กาญจนวาสี, 2535) โดยเขาได้เสนอแนวคิดในรูปของโมเดลนอร์มัลโอไจฟ์ (Normal Ogive Model) ผู้ที่พัฒนาต่อมาในปี ค.ศ. 1950 คือ ราส์ช (Rasch) นักคณิตศาสตร์ชาวเดนมาร์กซึ่งได้เสนอโมเดลราส์ช (Rasch Model) แบบ 1 พารามิเตอร์ โดยมีแนวคิดว่าความยากของข้อสอบ (b_j) เป็นสิ่งเดียวที่มีอิทธิพลต่อการตอบสนองข้อสอบ

ในปี ค.ศ. 1952 ลอร์ด (Lord) ได้เสนอฟังก์ชั่นนอร์มัลโอไจฟ์ แบบ 2 พารามิเตอร์ โดยเพิ่มพารามิเตอร์อำนาจจำแนก (a_j) และในปี ค.ศ. 1968 เบิร์นบอม (Birnbaum) ได้พัฒนาโมเดลโลจิสติก (Logistic Model) เป็นฟังก์ชั่นที่สามารถคำนวณได้ง่ายขึ้นใน ปี ค.ศ. 1974 ลอร์ด (Lord) ได้เสนอโมเดลโลจิสติก แบบ 3 พารามิเตอร์ โดยเพิ่มพารามิเตอร์โอกาสการเดาข้อสอบ (c_j) ในการประมาณค่าพารามิเตอร์ของผู้สอบด้วย

ปัจจุบันทฤษฎีการตอบสนองข้อสอบได้พัฒนาไปมากโดย บาร์ตันและลอร์ด (Barton and Lord) ได้เสนอโมเดลโลจิสติก แบบ 4 พารามิเตอร์ขึ้น โดยพวกเขาเชื่อว่า นักเรียนที่มีความสามารถสูงมากไม่จำเป็นที่จะต้องตอบข้อสอบถูกเสมอไป แม้ข้อสอบจะง่ายหรือจะระมัดระวังอย่างมากก็ตามก็ยังไม่สามารถใช้สารสนเทศ ที่มีตอบข้อสอบได้ถูกต้องตามเฉลยของผู้ออกข้อสอบได้เสมอไป (Hambleton, 1982) โมเดลนี้ให้ความสนใจศึกษาเชิงทฤษฎีเท่านั้น โดยผู้เสนอเองยังไม่ได้ศึกษาในเชิงปฏิบัติหรือนำไปใช้ สำหรับการนำแนวคิดของ IRT ไปใช้กับการวิเคราะห์มาตรประมาณค่าหรือข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค (Polytomous) ในปัจจุบันใช้แบบ 1 หรือ 2 พารามิเตอร์เท่านั้น

6.2 การเปลี่ยนแปลงคะแนนตามทฤษฎีการตอบสนองข้อสอบ

เมื่อมีการปรับเทียบคะแนนในรูปแบบการเก็บรวบรวมข้อมูลจากกลุ่มที่ไม่เท่าเทียมกัน พารามิเตอร์ข้อสอบจากแบบสอบต่างกันต้องทำให้อยู่บนสเกลเดียวกัน อย่างไรก็ตามพารามิเตอร์ที่ประมาณจากกระบวนการประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบเป็นสเกลที่แตกต่างกัน ตัวอย่างเช่นกำหนดค่าพารามิเตอร์จากโมเดลการตอบสนองข้อสอบที่ประมาณจากแบบสอบชุด X ที่ใช้ทดสอบกับประชากรกลุ่มที่ 1 ส่วนแบบสอบ Y ใช้ทดสอบกับประชากรกลุ่มที่ 2 เมื่อประชากรทั้งสองกลุ่มไม่เท่าเทียมกัน ในการวิเคราะห์แยกกลุ่ม โปรแกรมคอมพิวเตอร์จะกำหนดให้ค่าเฉลี่ยของค่าความสามารถ (θ) เป็น 0 และส่วนเบี่ยงเบน

มาตรฐานเป็น 1 เพื่อให้ให้ข้อมูลได้ถูกวิเคราะห์ ในกรณีนี้ความสามารถแต่ละกลุ่มจะเป็นสเกลที่มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 แม้ว่าแต่ละกลุ่มจะมีความสามารถที่แตกต่างกัน การเปลี่ยนแปลงของสเกลตามทฤษฎีการตอบสนองข้อสอบจึงเกิดขึ้น

ถ้าโมเดลการตอบสนองข้อสอบมีความเหมาะสมกับข้อมูลแล้ว การเปลี่ยนแปลงเชิงเส้นตรงของสเกลความสามารถก็จะมีความเหมาะสมกับข้อมูล และพารามิเตอร์ข้อสอบต้องมีการเปลี่ยนแปลงด้วยเหมือนกัน ตามโมเดลทฤษฎีการตอบสนองข้อสอบ การประมาณค่าพารามิเตอร์จากการวิเคราะห์ด้วยคอมพิวเตอร์แยกกัน ผลที่ได้ให้สเกลความสามารถ (θ) มีความสัมพันธ์เชิงเส้นตรง สมการเส้นตรงที่ได้สามารถใช้ในการเปลี่ยนแปลงพารามิเตอร์ที่ประมาณได้ให้อยู่ในสเกลเดียวกัน

6.3 สมการการเปลี่ยนแปลง

กำหนดสเกล I และ สเกล J เป็นสเกลที่แตกต่างกันตามทฤษฎีการตอบสนองข้อสอบ 3 พารามิเตอร์ โดยมีการเปลี่ยนแปลงเชิงเส้นตรง เมื่อ θ คือ ค่าความสามารถของ 3 สเกล มีความสัมพันธ์กันดังนี้

$$\theta_{Ji} = A\theta_{Ii} + B, \qquad (2)$$

เมื่อ A และ B เป็นค่าคงที่ในสมการเส้นตรง θ_{Ji} และ θ_{Ii} เป็นค่าความสามารถคนที่ i ในสเกล J และ สเกล I พารามิเตอร์ข้อสอบของ 2 สเกล มีความสัมพันธ์ดังนี้

$$a_{Jj} = \frac{a_{Ij}}{A}, \qquad (3)$$

$$b_{Jj} = Ab_{Ij} + B, \qquad (4)$$

และ $$c_{Jj} = c_{Ij}, \qquad (5)$$

เมื่อ a_{Jj} , b_{Jj} และ c_{Jj} คือพารามิเตอร์ข้อสอบ j ในสเกล J และ a_{Ij} , b_{Ij} และ c_{Ij} คือพารามิเตอร์ข้อสอบ j ในสเกล I

การเปลี่ยนแปลงสเกลตามทฤษฎีการตอบสนองข้อสอบ จากฟังก์ชันการตอบสนองข้อสอบ

$$P_{ij} = P_{ij}(\theta_i; a_j, b_j, c_j) = c_j + (1 - c_j)\frac{\exp[Da_j(\theta_i - b_j)]}{1 + \exp[Da_j(\theta_i - b_j)]} \qquad (6)$$

มีค่า A และ B ที่เป็นผลจากการเปลี่ยนแปลงสเกลจากการเปลี่ยนแปลงพารามิเตอร์จากสเกล I ไปสู่สเกล J พิจารณาทางขวาของฟังก์ชันสำหรับสเกล J เท่ากับ

$$c_{Jj} + (1 - c_{Jj}) \frac{\exp[Da_{Jj}(\theta_{Ji} - b_{Jj})]}{1 + \exp[Da_{Jj}(\theta_{Ji} - b_{Jj})]} \qquad (7)$$

แทน $(\theta_i; a_j, b_j, c_j)$ ในสมการที่ 3 — 5 ได้

$$= c_{Ij} + (1 - c_{Ij}) \frac{\exp\{D\frac{a_{Ij}}{A}[A\theta_{Ii} + B - (Ab_{Ij} + B)]\}}{1 + \exp\{D\frac{a_{Ij}}{A}[A\theta_{Ii} + B - (Ab_{Ij} + B)]\}}$$

$$= c_{Ij} + (1 - c_{Ij}) \frac{\exp[Da_{Ij}(\theta_{Ii} - b_{Ij})]}{1 + \exp[Da_{Ij}(\theta_{Ii} - b_{Ij})]} \qquad (8)$$

ผลที่ได้เท่ากับข้างขวาของฟังก์ชันสำหรับสเกล I ที่แสดงว่า A และ B ในสมการที่ 3 - 5 ช่วยในการเปลี่ยนแปลงสเกล

6.4 การหาค่าคงที่ A และ B

การหาค่า A และ B สำหรับบุคคล i และ i' หรือข้อสอบ j กับ j' A และ B ในสมการที่ 3 – 5 สามารถหาค่าได้ดังนี้ (Kolen and Brennan, 1995)

$$A = \frac{\theta_{Ji} - \theta_{Ji*}}{\theta_{Ii} - \theta_{Ii*}} = \frac{b_{Jj} - b_{Jj*}}{b_{Ij} - b_{Ij*}} = \frac{a_{Ij}}{a_{Jj}} \qquad (9)$$

และ $B = b_{Jj} - Ab_{Ij} = \theta_{Ji} - A\theta_{Ii}$, (10)

ตารางที่ 1 พารามิเตอร์ข้อสอบและค่าความสามารถของผู้สอบสำหรับ 2 สเกลของแบบสอบ

	สเกลที่ 1			สเกลที่ 2		
พารามิเตอร์ ข้อสอบ						
ข้อที่	a_{Ij}	b_{Ij}	c_{Ij}	a_{Jj}	b_{Jj}	c_{Jj}
j = 1	1.30	-1.30	.10	2.60	-1.15	.10
j = 2	.60	-.10	.17	1.20	-.55	.17
j = 3	1.70	.90	.18	3.40	-.05	.18
ความสามารถผู้สอบ						
คนที่	θ_{Ii}			θ_{Ji}		
i = 1	-2.00			-1.50		
i = 2	1.00			.00		
ค่าคงที่ที่ใช้ในการเปลี่ยนแปลงสเกล						
A = .5 B = -.05						
ความน่าจะเป็นที่จะตอบข้อสอบได้ถูก						
	$P_{Ij}(\theta_{Ii}; a_{Ij}, b_{Ij}, c_{Ij})$			$P_{Ij}(\theta_{Ji}; a_{Jj}, b_{Jj}, c_{Jj})$		
	คนที่			คนที่		
ข้อที่	i = 1	i = 2		i = 1	i = 2	
j = 1	.26	.99		.26	.99	
j = 2	.27	.80		.27	.80	
j = 3	.18	.65		.18	.65	

จากตารางแทนค่าได้ดังนี้

$$A = \frac{(-1.15)-(-.55)}{(-1.30)-(-.10)} = \frac{-.6}{-1.2} = .5$$

หรือใช้ความชันจากพารามิเตอร์ของข้อสอบข้อที่ 1 ได้

$$A = \frac{1.3}{2.6} = .5$$

ใช้สมการที่ 10 กับพารามิเตอร์ความยากข้อสอบข้อที่ 1 ได้

$$B = (-1.15) - (.5)(-1.30) = -.5$$

ค่า A และ B ใช้ในการเปลี่ยนสเกล I ไปสู่สเกล J ได้จากสมการ 2 - 5 ตัวอย่างเช่นจะเปลี่ยนค่าความสามารถของผู้สอบคนที่ 1 จากสเกล I ไปสู่สเกล J ได้ดังนี้

$$\theta_{J1} = A\theta_{I1} + B = .5(-2.00) + (-.5) = -1.5$$

ค่าความสามารถที่ได้ดังแสดงไว้ในตารางที่ 1

เมื่อต้องการเปลี่ยนพารามิเตอร์ของข้อสอบข้อที่ 3 จากสเกล I ไปสู่สเกล J ใช้สมการ 3 - 5 จะได้

$$a_{J3} = \frac{a_{I3}}{A} = \frac{1.7}{.5} = 3.4,$$

$$b_{J3} = Ab_{I3} + b = .5(.90) - .5 = -.05,$$

และ $c_{J3} = c_{I3} = .18$

ค่า P_{ij} เป็นความน่าจะเป็นที่ตอบข้อสอบได้ถูก ค่าที่ได้สามารถคำนวณจากค่าความสามารถและพารามิเตอร์ข้อสอบในตารางที่ 1 ค่า P_{ij} เหมือนกันทั้งสเกล I และสเกล J และจะเป็นเอกลักษณ์ซึ่งกันและกัน เมื่อสเกลมีความสัมพันธ์เชิงเส้นตรง

6.5 การหาค่า A และ B ในกลุ่มของข้อสอบหรือกลุ่มบุคคล

เมื่อต้องการหาความสัมพันธ์ระหว่างสเกลในกลุ่มของข้อสอบหรือกลุ่มผู้สอบ โดยการปรับสมการที่ 9 และ 10 ได้ดังนี้

$$A = \frac{\sigma(b_J)}{\sigma(b_I)} \qquad (11a)$$

$$= \frac{\mu(a_I)}{\mu(a_J)} \qquad (11b)$$

$$= \frac{\sigma(\theta_J)}{\sigma(\theta_I)} \qquad (11c)$$

$$B = \mu(b_J) - A\mu(b_I), \qquad (12a)$$

$$= \mu(\theta_J) - A\mu(\theta_I). \qquad (12b)$$

ตัวอย่างการคำนวณ ใช้ข้อมูลจากตารางที่ 1

$$\mu(b_I) = -.1667, \sigma(b_I) = .8994, \mu(a_I) = 1.2, \mu(b_J) = -.5833, \sigma(b_J) = .4497$$

และ $\mu(a_J) = 2.4$ แทนค่าในสมการที่ 11 และ 12 ได้

$$A = \frac{\sigma(b_J)}{\sigma(b_I)} = \frac{\mu(a_I)}{\mu(a_J)} = \frac{.4497}{.8994} = \frac{1.2000}{2.4000} = .5000,$$

และ $B = \mu(b_J) - A\mu(b_I) = -.5833 - .5000(-.1667) = -.5000$

ในการปรับเทียบคะแนนในกลุ่มไม่เท่าเทียมกัน การประมาณพารามิเตอร์จากข้อสอบร่วมจะนำมาใช้เป็นประโยชน์สำหรับสำหรับผู้สอบทั้งสองกลุ่ม การประมาณค่าพารามิเตอร์ของข้อสอบร่วม ใช้หาค่าคงตัวของสเกล โดยการแทนลงในสมการเพื่อประมาณหาค่าคงที่

เมื่อนำแบบสอบชุด 1 ไปสอบกับประชากรกลุ่มที่ 1 และนำแบบสอบชุด J ไปทดสอบกับกลุ่มที่ 2 ที่มีความแตกต่างจากกลุ่มที่ 1 ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของความสามารถไม่สามารถนำไปใช้ในการปรับสเกลได้ในสมการที่ 11c และ 12b เพราะสมการเหล่านี้ใช้กับกลุ่มผู้สอบกลุ่มเดียวที่ทำแบบสอบทั้งสองฉบับ

6.6 การประมาณค่าพารามิเตอร์เพื่อใช้ในการเปลี่ยนแปลงสเกลตามทฤษฎีการตอบสนองข้อสอบ

การประมาณค่าพารามิเตอร์ข้อสอบเป็นขั้นตอนที่สำคัญมากในการเปลี่ยนแปลงสเกลตามทฤษฎีการตอบสนองข้อสอบ ในการประมาณพารามิเตอร์ความสามารถแต่ละครั้งมีการกำหนดให้มีค่าเฉลี่ยเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 และกระบวนการประมาณค่าพารามิเตอร์ขึ้นอยู่กับรูปแบบการเก็บรวบรวมข้อมูล ดังต่อไปนี้

6.6.1 รูปแบบการเก็บรวบรวมข้อมูลแบบกลุ่มสุ่ม

เมื่อประมาณค่าพารามิเตอร์ทั้งสองกลุ่มที่ทำแบบสอบ X และแบบสอบ Y แยกกัน กำหนดให้แต่ละกลุ่มมีค่าเฉลี่ยความสามารถเป็น 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 พารามิเตอร์ที่ประมาณได้จากทั้งสองกลุ่มอยู่บนสเกลเดียวกัน ไม่มีการเปลี่ยนแปลงสเกล เพราะผู้สอบแต่ละกลุ่มได้จากการสุ่มให้มีความเท่าเทียมกัน และได้กำหนดให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถทั้งสองกลุ่มเท่ากัน แต่ถ้าไม่ได้กำหนดให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถให้เท่ากันแล้ว จำเป็นต้องมีการปรับสเกล โดยใช้สมการที่ 11c และ 12b

6.6.2 รูปแบบการเก็บรวบรวมข้อมูลกลุ่มเดียวจัดให้สมดุล

พารามิเตอร์ความสามารถผู้สอบและพารามิเตอร์ข้อสอบทั้งสองกลุ่มประมาณร่วมกัน เพราะพารามิเตอร์สำหรับแบบสอบทั้งสองฉบับถูกประมาณร่วมกันในกลุ่มตัวอย่างกลุ่มเดียว พารามิเตอร์ที่ประมาณได้กำหนดให้อยู่บนสเกลเดียวกัน ถ้าหากแยกประมาณคนละกลุ่มโดยกำหนดให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความสามารถเท่ากันทั้งสองกลุ่ม ผลที่ได้จะมีลักษณะเช่นเดียวกันกับการเก็บรวบรวมข้อมูลแบบกลุ่มสุ่ม

6.6.3 รูปแบบการเก็บรวบรวมข้อมูลกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม

เมื่อประมาณพารามิเตอร์ความสามารถและพารามิเตอร์ข้อสอบแยกกันทั้งกลุ่มที่ทำแบบสอบชุด X และกลุ่มที่ทำแบบสอบชุด Y ต้องมีการปรับเทียบคะแนนจากแบบสอบชุด X ให้อยู่บนสเกลเดียวกันกับชุด Y เพราะผู้สอบที่ทำแบบสอบชุด X ไม่เท่าเทียมกับผู้ทำแบบสอบชุด Y พารามิเตอร์ที่ประมาณได้จึงไม่อยู่บนสเกลเดียวกัน อย่างไรก็ตามมีการกำหนดข้อสอบร่วมสำหรับแบบสอบทั้งสองชุด และใช้ค่าพารามิเตอร์ที่ประมาณได้จากแบบสอบร่วม ใช้ในการประมาณเพื่อเปลี่ยนแปลงสเกล

6.7 วิธีการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ

การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ มีการปรับเปลี่ยนค่าความสามารถและพารามิเตอร์ของข้อสอบต่างกลุ่มให้อยู่บนสเกลเดียวกัน ความสัมพันธ์ของการปรับเปลี่ยนเป็นความสัมพันธ์เชิงเส้นตรง ฉะนั้นต้องมีการหาค่าความชัน (A) และ ค่าคงที่ (B) ในสมการความสัมพันธ์เชิงเส้นตรง วิธีการคำนวณหาค่าความชัน และค่าคงที่ หรือค่าสัมประสิทธิ์ของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ มีดังต่อไปนี้

6.7.1 วิธีใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความยาก (Mean and Sigma Method) เมื่อต้องการปรับคะแนนจากสเกล I ให้อยู่ในสเกล J หาค่าความชัน (A) และค่าคงที่ (B) จากสมการต่อไปนี้

$$A = \sigma(b_J) / \sigma(b_I)$$

$$B = \mu(b_J) - A\mu(b_I)$$

เมื่อ $\sigma(b_J)$ เป็น ส่วนเบี่ยงเบนมาตรฐานของค่าความยากจากแบบสอบชุด J

$\sigma(b_I)$ เป็น ส่วนเบี่ยงเบนมาตรฐานของค่าความยากจากแบบสอบชุด I

$\mu(b_J)$ เป็น ค่าเฉลี่ยของค่าความยากจากแบบสอบชุด J

I20877808

$\mu(b_I)$ เป็น ค่าเฉลี่ยของค่าความยากจากแบบสอบชุด I

6.7.2 วิธีใช้ค่าเฉลี่ยของค่าความยากและค่าอำนาจจำแนก (Mean and Mean Method) เมื่อต้องการปรับคะแนนจากสเกล I ให้อยู่ในสเกล J หาความชัน (A) และค่าคงที่ (B) จากสมการต่อไปนี้

$$A = \mu(a_I) / \mu(a_J)$$

$$B = \mu(b_J) - A\mu(b_I)$$

เมื่อ $\sigma(b_J)$ เป็น ส่วนเบี่ยงเบนมาตรฐานของค่าความยากจากแบบสอบชุด J

$\sigma(b_I)$ เป็น ส่วนเบี่ยงเบนมาตรฐานของค่าความยากจากแบบสอบชุด I

$\mu(b_J)$ เป็น ค่าเฉลี่ยของค่าความยากจากแบบสอบชุด J

$\mu(b_I)$ เป็น ค่าเฉลี่ยของค่าความยากจากแบบสอบชุด I

6.7.3 วิธีโค้งคุณลักษณะข้อสอบ (The Characteristic Curve Method)

ตามทฤษฎีการตอบสนองข้อสอบ (IRT) เมตริกซ์การเปลี่ยนแปลงเบื้องต้นกำหนดโดย

$$\theta_j^* = A\theta_j + B$$

เมื่อ A เป็นความชัน

B เป็นค่าคงที่

θ_j เป็นร ะดับความสามารถของผู้สอบในแบบสอบปัจจุบัน

θ_j^* เป็นร ะดับความสามารถของผู้สอบในแบบสอบที่เป็นเป้าหมาย

การใช้สัมประสิทธิ์ A และ B ค่าพารามิเตอร์ของข้อสอบในชุดปัจจุบัน เมื่อ a_i เป็นค่าการกระจายของข้อสอบ และ b_i เป็นค่าความยากของข้อสอบ สามารถเปลี่ยนเป็นค่าพารามิเตอร์ในแบบสอบเป้าหมายได้ดังนี้

$$a_i^* = \frac{a_i}{A} \qquad (13)$$

และ $$b_j^* = Ab_j + B \qquad (14)$$

เมื่อ i = 1, 2, 3,... แทนจำนวนข้อสอบ

ฮาบารา (Haebara, 1980 cite in Baker, 1997) ได้พัฒนาใช้กับแบบสอบที่มีการให้คะแนนแบบทวิวิภาคหรือแบบ 0-1 เป็นคนแรก ต่อมาสต๊อคกิงและลอร์ด (Stocking and Lord, 1983) ปรับสูตรใหม่ในรูปแบบโค้งคุณลักษณะแบบสอบ ผลที่ได้คือค่าคงที่หรือสัมประสิทธิ์การปรับเทียบคะแนนเป็นตัวปรับคะแนนระหว่างแบบสอบ

วิธีโค้งคุณลักษณะข้อสอบนี้ เป็นวิธีที่ใช้ทั้งการปรับเทียบคะแนนแนวตั้งและแนวนอน สำหรับการตรวจให้คะแนนแบบทวิวิภาค หรือ แบบ 0-1 ในกรณีที่จำเป็นต้องประมาณค่าพารามิเตอร์ของข้อสอบแยกกัน เป็นวิธีการที่ใช้โค้งคุณลักษณะข้อสอบเพื่อหาสัมประสิทธิ์การปรับเทียบ (Equating Coefficients) มีรายละเอียดดังนี้

ในกรณีการตอบแบบสอบที่มีการตรวจให้คะแนน 0-1 กำหนดให้ $P_i(\theta_j)$ แทนความน่าจะเป็นของการตอบถูกของข้อสอบร่วม (Anchor Test) ในแบบสอบที่เป็นหลัก และ $P_i^*(\theta_j)$ แทนความน่าจะเป็นของการตอบถูกของข้อสอบร่วมในแบบสอบที่ปรับคะแนนแล้ว ถ้าหากว่า $[P_i(\theta_j) - P_i^*(\theta_j)] = 0$ แสดงว่าการปรับเทียบคะแนนมีความสมบูรณ์ ดังนั้นการปรับเทียบคะแนนคือการหาสัมประสิทธิ์การปรับเทียบคะแนน A และ B ได้จากเกณฑ์ที่กำหนดค่าผลต่างจากสมการ

$$F = \frac{1}{N}\sum_{j=1}^{N}\{\sum_{i=1}^{n}[P_i(\theta_j) - P_i^*(\theta_i)]\}^2, \qquad (15)$$

เมื่อ N คือ จำนวนจุดบน θ ที่กำหนดขึ้นเอง และ n คือ จำนวนข้อสอบร่วมของทั้งสองแบบสอบ การกำหนด $P_i(\theta_j)$ และ $P_i^*(\theta_j)$ ขึ้นอยู่กับโมเดลที่ใช้ในฟังก์ชั่นการตอบสนองข้อสอบ ถ้าใช้โมเดล 2 พารามิเตอร์ จะได้

$$P_i(\theta_j) = \frac{1}{1 + \exp[-a'_i(\theta_j - b'_i)]}$$

และ $$P_i(\theta_j) = \frac{1}{1 + \exp[-a_i^*(\theta_j - b_i^*)]}$$

เมื่อ a_i' และ b_i' เป็นค่าพารามิเตอร์ของข้อสอบร่วมจากแบบสอบหลัก

a_i^* และ b_i^* เป็นผลจากการใช้สมการที่ 13 และ 14 เพื่อประมาณค่าพารามิเตอร์ข้อสอบที่ปรับเทียบ

θ_j แทน เมตริกซ์ร่วมของคะแนนความสามารถ

ส่วนความสัมพันธ์ของผลรวมอันดับ 2 ในสมการที่ 15 กำหนดเป็นคะแนนจริงทั้งสองแบบสอบ สัมประสิทธิ์การปรับเทียบคะแนนสามารถหาได้จาก A Quadratic Loss Function กำหนดในรูปของโค้งคุณลักษณะข้อสอบดังนี้ (Stocking and Lord, 1983)

$$F = \frac{1}{N}\sum_{j=1}^{N}(T_j - T_j^*)^2,$$

เมื่อ $$T_j = \sum_{i=1}^{n} P_i(\theta_j),$$

และ $$T_j^* = \sum_{i=1}^{n} P_i^*(\theta_j)$$

ตามการหาค่าสัมประสิทธิ์การปรับเทียบคะแนน A และ B ที่ได้จากการใช้เทคนิคการค้นหาหลายตัวแปรหลายครั้ง (The Iterative Multivariate Search Technique) ซึ่งเดวิดดอน (Davidon, 1959 cited in Baker, 1997) และ เฟรซเชอร์กับเพาเวลล์ (Fletcher and Powell, 1963 cited in Baker, 1997) พบว่าเมื่อค่า F ลดลงมากที่สุดจะได้ค่าสัมประสิทธิ์การปรับเทียบ ถ้าการปรับเทียบคะแนนสมบูรณ์ค่า F = 0

6.7.4 วิธีประมาณค่าไคว์สแคว์น้อยที่สุด (The Minimum χ^2 Method)

วิธีประมาณค่าไคว์สแคว์น้อยที่สุด เสนอโดย ดิฟกิ (Divgi, 1985 cited in Kim and Cohen, 1995) เป็นวิธีที่ใช้ทั้งการปรับเทียบคะแนนแนวตั้งและแนวนอน สำหรับการตรวจให้คะแนนแบบทวิวิภาค หรือ แบบ 0-1 ในกรณีที่จำเป็นต้องประมาณค่าพารามิเตอร์ของข้อสอบแยกกัน เป็นวิธีที่วิเคราะห์ใช้เกณฑ์ที่ทำให้ค่า χ^2 น้อยที่สุดเพื่อจะให้ได้ค่าสัมประสิทธิ์การปรับเทียบคะแนน มีรายละเอียดดังนี้

ลอร์ด (Lord, 1980 cite in Kim and Cohen, 1995) ได้แสดงให้เห็นว่าใน IRT ความสัมพันธ์ระหว่างเมตริกซ์พารามิเตอร์ข้อสอบของสองข้อสอบที่เปรียบเทียบเป็นเชิงเส้นตรง เมื่อปรับค่าพารามิเตอร์ข้อสอบจากฉบับที่ 2 ให้เทียบกับเมตริกซ์ฉบับที่ 1 การประมาณค่าอำนาจจำแนก (a) และความยากของข้อสอบ (b) ของข้อสอบ i หาได้จาก

$$a_{i2}^* = a_{i2} / A$$

และ $$b_{i2}^* = Ab_{i2} + B,$$

เมื่อ * แทนค่าที่ปรับของแบบสอบชุดที่ 1

A คือ ความชัน

B คือ ค่าคงที่

คะแนนความสามารถ (θ) ของคนที่ j แสดงได้ดังนี้

$$\theta_{j2}^{*} = A\theta_{j2} + B$$

การปรับเทียบสองเมตริกซ์ต้องหาสัมประสิทธิ์การปรับเทียบ A และ B ที่เหมาะสม โดยใช้วิธีประมาณค่าไคว์สแคว์ต่ำสุด เพื่อให้ได้สัมประสิทธิ์การปรับเทียบคะแนนจากสมการ

$$\chi^2 = \sum_{i=1}^{n} \chi_i^2 = \sum_{i=1}^{n} \xi'_i \sum_{i}^{-1} \xi_i,$$

เมื่อ n แทน จำนวนข้อสอบทั้งหมด

ξ_i แทน เวกเตอร์ความแตกต่างระหว่างพารามิเตอร์ข้อสอบครั้งแรกและครั้งที่ 2 ที่ปรับแล้ว

และ $\sum_i$ แทน เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วม ของ ξ_i

สำหรับโมเดล IRT 2 พารามิเตอร์ มีตัวอย่างดังนี้

$$\xi_i = \xi_{i1} - \xi_{i2}^{*}$$

และ $\Sigma_i = \Sigma_{i1} - \Sigma_{i2}^{*}$

เมื่อ $\xi_{i1} = (a_{i1}, b_{i1})'$,

และ $\xi_{i2}^{*} = (a_{i2}^{*}, b_{i2}^{*})'$

$\sum_{i1}$ แทน เมตริกซ์ความแปรปรวนร่วม มิติ 2 x 2 ของความคลาดเคลื่อนของการสุ่มตัวอย่าง สำหรับข้อสอบ i จากแบบสอบชุดแรก และ

$\sum_{i2}^{*}$ แทน เมตริกซ์ความแปรปรวนร่วมที่ปรับจากแบบสอบชุดที่ 2 จะได้

$$\chi^2 = \sum_{i=1}^{n} (a_{i1} - a_{i2}^{*}, b_{i1} - b_{i2}^{*})(\Sigma_{i1} + \Sigma_{i2}^{*})^{-1}(a_{i1} - a_{i2}^{*}, b_{i1} - b_{i2}^{*})' \quad (15)$$

χ^2 นี้เป็นฟังก์ชันของสองสัมประสิทธิ์การปรับเทียบที่ยังไม่ทราบค่าคือ A และ B ได้ B จากการได้มาบางส่วน(Partial Derivative) ของสมการที่ 16 กับการพิจารณาที่ B : $\partial\chi^2 / \partial B = 0$ การได้มาบางส่วนมีความสัมพันธ์เชิงเส้นตรงกับการพิจารณา B และ

ทำให้หาค่า A ได้ง่ายขึ้น แสดงให้เห็นว่า S_{iab} และ S_{ibb} เป็นส่วนประกอบเฉพาะจากเมตริกซ์

$$S_i = \Sigma_i^{-1} = (\Sigma_{i1} + \Sigma_{i2}^*)^{-1}$$

เมื่อ $$B = \frac{\sum_{i=1}^{n}[(a_{i1} - a_{i2}/A)S_{iab} + (b_{i1} - Ab_{i2})S_{ibb}]}{\sum_{i=1}^{n} S_{ibb}}$$

เมื่อค่า B ได้จากสมการนี้แทนจากที่ได้ χ^2 ทำให้หาค่า A ได้ง่ายขึ้น โดยใช้วิธีของนิวตัน (Newton's Method) จากสมการดังต่อไปนี้

$$A^{(s)} = A^{(s-1)} - [H^{(s-1)}]^{-1} f^{(s-1)},$$

เมื่อ s เป็นจำนวนรอบที่ทำซ้ำ

โดยที่ $$f^{(s-1)} = \frac{\partial \chi^2}{\partial A}_{A(s-1)},$$

และ $$H^{(s-1)} = \frac{\partial^2 \chi^2}{\partial A^2}_{A(s-1)}$$

ตอนที่ 7 เกณฑ์ที่ใช้ในการปรับเทียบคะแนน

การปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับที่มีคุณภาพ ให้ผลการปรับเทียบมีความเชื่อถือได้ ต้องมีการติดตามและประเมินผลการปรับเทียบคะแนน มีเกณฑ์ที่ใช้ดังต่อไปนี้

1. ความเสมอภาค (Weak Equity)

ลอร์ด (Lord, 1980) ได้เสนอคุณลักษณะที่ดีของการปรับเทียบคะแนนด้านความเสมอภาค (Equity) คือ เมื่อทุกกลุ่มผู้สอบมีความสามารถเดียวกัน การแจกแจงคะแนนจากแบบสอบชุดที่ 1 ที่ปรับสเกลเข้าสู่แบบสอบชุดที่ 2 จะมีการการแจกแจงคะแนนเหมือนกับแบบสอบชุดที่ 2 ดิฟกิ (Divgi, 1981) และเยน (Yen, 1983) ได้นำคุณลักษณะความเสมอภาคของลอร์ดไปประยุกต์เป็นเกณฑ์ โดยกำหนดคุณภาพที่ดีของการปรับเทียบคะแนนนั้นพิจารณาจากความเท่ากันของการแจกแจงตามเงื่อนไขของคะแนนของแบบสอบแต่ละฉบับหลังจากการปรับเทียบแล้ว แฮนสัน (Hanson, 1991b) ได้เสนอเกณฑ์ในลักษณะเดียวกันนี้ใช้ชื่อว่าคะแนน

ความคาดหวังที่เท่าเทียมกัน (Equivalent Expected Score) โดยมีข้อตกลงเบื้องต้นว่าคะแนนจริงของแบบสอบชุดเดิมที่ไม่ได้ปรับ (X) และคะแนนจริงของแบบสอบชุดใหม่ที่ต้องการปรับไปสู่แบบสอบชุด X มีความสัมพันธ์กันเชิงฟังก์ชัน เกณฑ์จะพิจารณาคะแนนจากแบบสอบชุด X ที่ปรับคะแนนมาจากแบบสอบชุด Y ว่าจะมีความเท่าเทียมกันกับคะแนนจากแบบสอบ X แสดงได้ดังสมการต่อไปนี้

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

เมื่อ τ แทน คะแนนจริงจากแบบสอบเดิมชุด X

$E[e(Y)|\gamma(\tau)]$ แทน คะแนนจากแบบสอบใหม่(Y) ที่ปรับเทียบแล้ว

$E(X|\tau)$ แทน คะแนนจากแบบสอบชุดเดิมชุด X

2.1 ค่าดัชนี (Indices)

เกณฑ์นี้เสนอโดยแองกอฟ จำแนกได้ดังนี้

2.1 ดัชนีที่ถ่วงน้ำหนักด้วยความถี่ของคะแนน

เมื่อกำหนดให้

A_i แทน คะแนน i ที่ปรับจากแบบสอบใหม่ให้อยู่ในสเกลกับแบบสอบเดิม

B_i แทน คะแนนจริงหรือคะแนนที่ใช้เป็นเกณฑ์

f_i แทน ความถี่ของคะแนนดิบจากแบบสอบชุดใหม่

2.1.1 ดัชนี RMS (Root Mean Square) หาจากสูตร

$$RMS = \left(\frac{\sum_i f_i(A_i - B_i)^2}{\sum_i f_i}\right)^{\frac{1}{2}}$$

2.1.2 ดัชนี MAD (Mean Absolute Difference) หาจากสูตร

$$MAD = \frac{\sum_i f_i|A_i - B_i|}{\sum_i f_i}$$

2.1.3 ดัชนี MSD (Mean Signed Difference) หาได้จากสูตร

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

2.2 ดัชนีที่ไม่ได้ถ่วงน้ำหนักด้วยความถี่ของคะแนน

เนื่องจากดัชนีที่ถ่วงน้ำหนักด้วยความถี่ของคะแนน ให้ความสำคัญในส่วนของความแตกต่างของคะแนนที่มีความถี่ แต่ไม่ให้ความสำคัญในส่วนของความแตกต่างของ

คะแนนที่ไม่มีผู้ได้คะแนน อย่างไรก็ตามควรใช้ความแตกต่างของคะแนนที่ปรากฏทั้งหมดบนสเกลของคะแนน เมื่อกำหนด k คือ จำนวนจุดของคะแนน ดัชนีที่ไม่ได้ถ่วงน้ำหนักด้วยความถี่ของคะแนน มีดังนี้

2.2.1 ดัชนี RMS_U หาได้จากสูตร

$$RMS_u = (\frac{\sum_i (A_i - B_i)^2}{k})^{\frac{1}{2}}$$

2.2.2 ดัชนี MAD_U หาได้จากสูตร

$$MAD_u = \frac{\sum_i |A_i - B_i|}{k}$$

2.2.3 ดัชนี MSD_U หาได้จากสูตร

$$MSD_u = \frac{\sum_i (A_i - B_i)}{k}$$

3. ความคลาดเคลื่อนมาตรฐาน (Standard Errors of Equating)

ความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน เป็นการประมาณขนาดของความคลาดเคลื่อนจากการสุ่มตัวอย่าง เมื่อขนาดของกลุ่มตัวอย่างเข้าใกล้ค่าอนันต์แล้วความคลาดเคลื่อนมาตรฐานจะเข้าใกล้ศูนย์

4. ข้อมูลที่จำลองขึ้น (Generated Data)

การจำลองข้อมูลเป็นเทคนิคที่ใช้กันมากในการศึกษาเกี่ยวกับการปรับเทียบคะแนน เป็นเพราะสามารถจำลองข้อมูลขึ้น รู้ความสัมพันธ์กับการปรับเทียบคะแนนอย่างแท้จริง และทำให้นักวิจัยเสียค่าใช้จ่ายน้อยในการทำวิจัย

5. การปรับเทียบคะแนนกลับสู่แบบสอบเดิม (Test Equated to Itself or Circular Equating)

ลอร์ด (Lord, 1980) ได้เสนอการปรับเทียบกลับสู่แบบสอบเดิม การปรับเทียบคะแนนกลับสู่แบบสอบเดิมถูกนำมาใช้บ่อยมากในงานวิจัย ไม่ว่าการปรับเทียบโดยตรงหรือการปรับเทียบผ่านแบบสอบอื่นแล้วปรับกลับสู่แบบสอบเดิม มีการเปลี่ยนแปลงการปรับเทียบคะแนนมากกว่า 1 ครั้ง เช่น การปรับเทียบคะแนนจากแบบสอบชุด A ไปยังชุด B ปรับชุด B ต่อไปยังชุด C และปรับจากชุด C กลับสู่แบบสอบชุด A เดิม

ผลการเปลี่ยนแปลงคะแนนจากการปรับเทียบคะแนนกลับสู่แบบสอบเดิมควรเป็นเอกลักษณ์ คือคะแนนดิบจากแบบสอบชุด A เดิม กับคะแนนจากแบบสอบชุด A ที่ปรับผ่านจากชุด B และชุด C มีค่าเท่ากัน

6. เกณฑ์กลุ่มตัวอย่างขนาดใหญ่ (Large Sample Criterion)

แองกอฟ (Angoff, 1985) ได้เสนอเกณฑ์กลุ่มตัวอย่างขนาดใหญ่ โดยเสนอว่าถ้าหากรู้การปรับเทียบคะแนนจากประชากร ทำให้มีเกณฑ์ที่แท้จริงเพื่อประเมินผลการปรับเทียบคะแนน ในบางงานวิจัยใช้ผลการปรับเทียบคะแนนจากกลุ่มตัวอย่างขนาดใหญ่เหมือนกับการปรับเทียบคะแนนจากประชากร เพื่อเปรียบเทียบกับผลการปรับเทียบคะแนนจากกลุ่มตัวอย่างที่เล็กกว่า

7. ความคงเส้นคงวา (Consistency)

การพิจารณาความคงเส้นคงวาของการปรับเทียบคะแนน ได้จากการนำวิธีการปรับเทียบคะแนนหลายวิธีไปใช้ในกระบวนการปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับ เพื่อดูผลจากการปรับเทียบคะแนนว่าแตกต่างหรือเหมือนกัน ถ้าให้ผลเหมือนกันแสดงว่ามีความคงที่ของการปรับเทียบคะแนน แต่ไม่ได้ชี้ชัดลงไปว่ามีความถูกต้องแม่นยำ

8. การใช้กลุ่มสอบทานผล (Replication Samples or Cross-Validation Studies)

กลุ่มสอบทานผลเป็นกลุ่มผู้สอบที่มีคุณลักษณะเช่นเดียวกับกลุ่มผู้สอบที่ทำแบบสอบเพื่อการปรับเทียบคะแนน กลุ่มนี้จะทำแบบสอบทั้งสองชุดและจะเป็นกลุ่มที่ตรวจสอบความคลาดเคลื่อนที่เกิดจากการปรับเทียบคะแนน เพราะเมื่อมีตารางปรับเทียบคะแนนจากแบบสอบสองชุดแล้วนำคะแนนจากกลุ่มสอบทานผลไปปรับสเกลโดยใช้ตารางปรับเทียบเดิม ถ้าการปรับเทียบมีคุณภาพดีแล้ว ผลต่างระหว่างคะแนนที่ไม่ได้ปรับสเกลกับคะแนนที่ปรับสเกลจะมีค่าน้อย

9. เกณฑ์ของปิเตอร์เซนและคณะ

ปิเตอร์เซน มาร์โค และ สตีเวอร์ท (Pertersen, Marco, and Stewart, 1982) ได้เสนอเกณฑ์ในการตัดสินคุณภาพของการปรับเทียบคะแนน โดยคำนวณค่าดัชนี MSE ซึ่งเป็นค่าเฉลี่ยกำลังสองของผลต่างของคะแนนฐานกับคะแนนที่ปรับแล้ว ถ่วงน้ำหนักด้วยความแปรปรวนคะแนนฐาน และได้กำหนดเกณฑ์ตามความเห็นของคณะผู้วิจัย เป็นดังนี้

ระดับน่าพอใจอย่างมาก เมื่อ $MSE < (.05\ S_X)^2$

ระดับน่าพอใจ เมื่อ $(.05\ S_X)^2 \leq MSE < (.10\ S_X)^2$

ระดับปานกลาง เมื่อ $(.10\ S_X)^2 \leq MSE < (.15\ S_X)^2$

ระดับไม่น่าพอใจ เมื่อ $(.15\ S_X)^2 \leq MSE < (.20\ S_X)^2$

ระดับไม่น่าพอใจอย่างมาก เมื่อ $(.20\ S_X)^2 \leq MSE$

เมื่อ S_X คือ ส่วนเบี่ยงเบนมาตรฐานของคะแนนฐาน

การนำเกณฑ์ที่กล่าวมาแล้วไปใช้ในการปรับเทียบคะแนนสามารถนำไปใช้ได้หลายรูปแบบดังนี้ นำเกณฑ์เดียวไปใช้เลย เช่น ค่าดัชนีบอกความแตกต่างระหว่างวิธีการปรับเทียบสองวิธี หรือค่าดัชนีบอกความแตกต่างระหว่างเงื่อนไขการปรับเทียบจำนวนข้อสอบร่วมต่างกันจากการใช้วิธีปรับเทียบวิธีเดียวกัน นำสองเกณฑ์ใช้ร่วมกัน เช่น ใช้กลุ่มสอบทานผลกับค่าดัชนีบอกความแตกต่างระหว่างวิธีการปรับเทียบคะแนนสองวิธี ค่าดัชนีน้อยกว่าจะบ่งบอกถึงคุณภาพการปรับเทียบคะแนนที่ดีกว่า และนำสามเกณฑ์ใช้ร่วมกัน เช่น การจำลองข้อมูล การปรับเทียบกลับสู่แบบสอบเดิมกับค่าดัชนีบอกความแตกต่างระหว่างเงื่อนไขจำนวนข้อสอบร่วมต่างกันจากการปรับเทียบคะแนนด้วยวิธีเดียวกัน มีผู้วิจัยเกี่ยวกับเกณฑ์ ดังนี้ (Harris and Crouse, 1993)

แพเทียน (Patience, 1981) ได้เปรียบเทียบวิธีการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ และวิธีอิควิเปอร์เซ็นต์ไทล์ ในการศึกษาการปรับเทียบคะแนนตามแนวตั้ง เปรียบเทียบวิธีการโดยการสร้างกราฟและคำนวณความสัมพันธ์ระหว่างคะแนนสเกลที่ได้มากับคะแนนสเกลที่เป็นเกณฑ์

ลอร์ด (Lord, 1982a) ได้ดำเนินการหาความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนอิควิเปอร์เซ็นต์ไทล์ ในสถานการณ์ปรับเทียบคะแนน 4 สถานการณ์ โดยการจำลองข้อมูลมาศึกษา มีการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนเชิงเส้นตรงกับวิธีอิควิเปอร์เซ็นต์ไทล์ เมื่อใช้ข้อมูลจริง

ปิเตอร์เซน และคณะ (Petersen and others, 1983) ใช้ Scale Drift ในการปรับเทียบคะแนนตามทฤษฎีการวัดแบบดั้งเดิมและทฤษฎีการตอบสนองข้อสอบหลายวิธี ปรับเทียบคะแนนจากแบบสอบมาตรฐานด้านภาษา ถ้าหากว่าการปรับเทียบคะแนนไม่มีความคลาดเคลื่อนเชิงระบบและเชิงสุ่ม แล้วความสัมพันธ์ที่ได้จากการปรับเทียบคะแนนกลับสู่แบบสอบเดิมควรเป็นเอกลักษณ์ (Identity) ใช้กราฟของคะแนนที่เปลี่ยนแปลงสุดท้ายลบด้วยเกณฑ์ ดัชนี MSD ถ่วงน้ำหนักด้วยความถี่ และค่าความลำเอียง ช่วยในการแปลผล ใช้ส่วนเบี่ยงเบนมาตรฐานของคะแนนจากแบบสอบ เป็นเครื่องวัดที่บอกขนาด

อัลโดซารี (Al-Dosary, 1983) ได้ศึกษาวิธีการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ วิธี Mean and Sigma และวิธี Characteristic curve มีการสร้างกราฟของการประมาณพารามิเตอร์ข้อสอบ กราฟของฟังก์ชันสารสนเทศข้อสอบ ใช้ดัชนี RMSE ความแตกต่างระหว่างแบบสอบพื้นฐานและแบบสอบที่ปรับแล้ว และดัชนี MAE

ดิฟกิ (Divgi, 1983) ได้แบ่งครึ่งกลุ่มตัวอย่างแต่ละกลุ่ม แล้วปรับเทียบคะแนนจากกลุ่มที่แบ่งเป็น 2 ครั้ง ความแตกต่างกำลังสองระหว่างกลุ่มถูกใช้เป็นความคลาดเคลื่อนเชิงสุ่ม

ฟิลลิปส์ (Phillips, 1983) แนะนำว่าการปรับเทียบคะแนนควรทำสะท้อนกลับ (Replication) ในการปรับเทียบคะแนนกลับสู่แบบสอบเดิม ถูกใช้เป็นเช่นเดียวกับเกณฑ์

เยน (Yen, 1984) กล่าวว่าควรมีความรอบคอบในการใช้กลุ่มสอบทานผล (Cross-Validation Samples) ในการปรับเทียบคะแนนสำหรับการสอบที่มีการวัดหลายมิติ

คุกและคณะ (Cook and others, 1985) ได้ศึกษาความสัมพันธ์ระหว่างคุณสมบัติของการวัดมิติเดียว และคุณภาพของการปรับเทียบคะแนนจริง ตามทฤษฎีการตอบสนองข้อสอบด้วยการตรวจสอบ Scale Drift กำหนดให้ Scale Drift จะเกิดขึ้นถ้าหากว่าผลการปรับเทียบคะแนนจากแบบสอบ D ไปสู่แบบสอบ A โดยตรง มีความแตกต่างจากการปรับเทียบคะแนนจากแบบสอบชุด Dไปสู่ชุด A โดยผ่ายชุด B และชุด C และใช้ดัชนี MSD ที่ถ่วงน้ำหนักด้วยความถี่สรุปผลความแตกต่างระหว่าง 2 การเปลี่ยนแปลง

เคลอินและจาร์เจอรา (Klein and Jarjoura, 1985) ใช้การปรับเทียบกลับสู่แบบสอบเดิมผ่านวงจรลูกโซ่เพื่อตรวจสอบความสำคัญของการเป็นตัวแทนเนื้อหาของแบบสอบร่วม เมื่อกำหนดให้จำนวนข้อสอบร่วมแตกต่างกัน เกณฑ์ที่ใช้คือดัชนี RMSE ที่ถ่วงน้ำหนักด้วยความถี่และค่าความลำเอียง

จาร์เจอราและโคเลน (Jarjoura and Kolen, 1985) ได้ศึกษาหาความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนวิธีอิควิเปอร์เซ็นต์ไทล์ ในรูปแบบการเก็บรวบรวมข้อมูลกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม มีการเปรียบเทียบความคลาดเคลื่อนที่ได้มากับความคลาดเคลื่อน Bootstrap จากการจำลองข้อมูลทางคอมพิวเตอร์

เคลอินและโคเลน (Klein and Kolen, 1985) ได้ปรับเทียบคะแนนกลับสู่แบบสอบเดิมเพื่อตรวจสอบผลของการใช้จำนวนข้อสอบร่วมต่างกัน ในการปรับเทียบคะแนนโดยใช้ข้อสอบร่วมที่กลุ่มผู้สอบไม่ได้จากการสุ่ม เกณฑ์ที่ใช้คือดัชนี RMSE และ ค่าความลำเอียง

คุกและไอเนอร์ (Cook and Eignor, 1985) ได้ศึกษาการปรับเทียบคะแนนจากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนตามทฤษฎีการตอบสนองข้อสอบ มี Scale Drift ค่าความแปรปรวนของความแตกต่างและกำลังสองของค่าความลำเอียงเป็นเกณฑ์ประเมินสำหรับแต่ละวิธี

การปรับเทียบคะแนนแบบสอบ

เยน (Yen, 1985) ใช้กลุ่มตัวอย่างครึ่งหนึ่งเช่นเดียวกันกับกลุ่มตรวจสอบข้ามกลุ่ม กับการศึกษาการปรับเทียบคะแนนในแนวตั้ง จากเกณฑ์ความแตกต่างของค่าเฉลี่ยที่เป็นมาตรฐานอัตราส่วนของส่วนเบี่ยงเบนมาตรฐาน ดัชนี RMSD ค่าสหสัมพันธ์และความลำเอียง

โคเลน (Kolen, 1985a) ใช้ดัชนี RMSD ความแตกต่างค่าเฉลี่ย ค่าความลำเอียง ส่วนเบี่ยงเบนมาตรฐานของความแตกต่าง และความชันกับค่าคงที่ของสมการการปรับเทียบคะแนน ในการศึกษาเปรียบเทียบวิธีการปรับเทียบคะแนนเชิงเส้นตรง ใช้แบบสอบร่วม ประเมินความคงที่ของการปรับเทียบคะแนนโดยใช้การปรับเทียบกับสู่แบบสอบเดิม

คิงสตันและฮอลแลนด์ (Kingston and Holland , 1986) ใช้การปรับเทียบกลับสู่แบบสอบเดิม ตรวจสอบทางเลือกของการปรับเทียบคะแนนจากการทดสอบในระดับบัณฑิตวิทยาลัย จากวิธีการปรับเทียบหลายวิธี ตรวจสอบจากดัชนี RMSE ค่าความลำเอียง และ กราฟ

สต๊อกกิงและไอเนอร์ (Stocking and Eignor, 1986) ได้ใช้การจำลองข้อมูลตรวจสอบผลกระทบของการแจกแจงความสามารถในการดำเนินการปรับเทียบคะแนนก่อน (Preequating) ตามทฤษฎีการตอบสนองข้อสอบ

ราจูและคณะ (Raju and others, 1986) ได้ศึกษาเกี่ยวกับความยาวแบบสอบร่วม ปรับเทียบคะแนนด้วยวิธี Rasch และวิธี IRT ให้ผู้สอบแต่ละคนที่เป็นกลุ่มสอบทานผลทำแบบสอบทั้งสองชุด และประมาณความสามารถผู้สอบแต่ละคนที่สอบแบบสอบทั้งสองฉบับใช้ความแตกต่างระหว่างการประมาณค่าความสามารถ เป็นเกณฑ์พิจารณาความคลาดเคลื่อนของการปรับเทียบ

สเก็กส์และลิสสิท (Skaggs and Lissitz, 1986b) ได้เปรียบเทียบความแกร่งของวิธีการปรับเทียบคะแนน 4 วิธี เกณฑ์ที่ใช้คือ ดัชนี MSE ทั้งที่ถ่วงและไม่ถ่วงน้ำหนักด้วยความถี่ เมื่อพิจารณาดูผลจากการใช้เกณฑ์ด้วยดัชนี MSE ที่ถ่วงน้ำหนักด้วยความถี่โดยใช้กราฟ พบว่าในบางกรณีอธิบายผลได้ไม่เพียงพอ เพราะมีความถี่ของคะแนนน้อยที่ช่วงคะแนนที่ต่ำสุด เมื่อมีความคลาดเคลื่อนของการปรับเทียบคะแนนมากที่สุดเกิดขึ้น

โคเลนและจาร์เจอรา (Kolen and Jarjoura, 1987) ได้ประมาณความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนวิธีอิควิเปอร์เซ็นต์ไทล์ปรับให้เรียบ Cubic Spline ใช้กระบวนการ Bootstrapแบ่งความคลาดเคลื่อนของการปรับเทียบคะแนนออกเป็น ความคลาดเคลื่อนเชิงสุ่ม และ ความคลาดเคลื่อนเชิงระบบ ตามระดับความแตกต่างของการปรับให้เรียบ จุดมุ่งหมายของการศึกษาคือ เปรียบเทียบระดับที่ต่างกันของการปรับให้เรียบกับฟังก์ชัน Cubic Spline และคำนวณดัชนีของความคลาดเคลื่อนทั้งหมด ทั้งความคลาดเคลื่อนเชิงสุ่ม และความคลาดเคลื่อนเชิงระบบเพื่อช่วยในการเปรียบเทียบ

ค็อบ (Cope, 1987) ใช้การปรับเทียบกลับสู่แบบสอบเดิม ได้ศึกษาเมื่อปี 1986 และ 1987 เพื่อตรวจสอบการใช้กลุ่มตัวอย่างที่ซ้ำในการปรับเทียบคะแนนใช้แบบสอบร่วม และการปรับเทียบคะแนนเชิงเส้นตรงใช้แบบสอบร่วม มีการใช้รูปแบบและวงจรการปรับเทียบที่หลากหลาย เกณฑ์ที่ใช้คือดัชนี RMSE ที่ถ่วงน้ำหนักด้วยความถี่ และค่าความลำเอียง

แฮร์ริส (Harris, 1987) ได้เปรียบเทียบวิธีการปรับเทียบคะแนน 4 วิธี ศึกษากับกลุ่มที่เปรียบเทียบได้กับกลุ่มที่เปรียบเทียบได้น้อย ในการปรับเทียบคะแนนที่ใช้แบบสอบร่วม ใช้การเปรียบเทียบกลุ่มตัวอย่างรวมเป็นเกณฑ์

แฮร์ริสและฮูเวอร์ (Harris and Hoover, 1987) ได้ประยุกต์ใช้การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ 3 พารามิเตอร์ สำหรับการปรับเทียบคะแนนในแนวตั้ง เปรียบเทียบโดยตรงและการปรับเทียบผ่านแบบสอบอื่นก่อน ผลที่ได้ถูกนำไปเปรียบเทียบกับผลจากการใช้วิธีการปรับเทียบคะแนนด้วยวิธีอิควิเปอร์เซ็นต์ไทล์และวิธีของ Rasch โดยพิจารณาจากค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน

ลิวิงสตันและเฟอร์ย็อค (Livigston and Feryok, 1987) ใช้ความถูกต้องแม่นยำ โดยการคำนวณผลการปรับเทียบคะแนน ในประชากร 50,000 คน เป็นเกณฑ์

เฮิรส (Hirsch, 1989) ใช้ทั้งข้อมูลจำลองและข้อมูลจริงในการศึกษาการปรับเทียบคะแนนที่มีหลายมิติ (Multidimensional Equating) ผลที่ได้ชี้ให้เห็นประสิทธิผลของการปรับเทียบคะแนน เมื่อวัดโดยการเปรียบเทียบความสามารถของคะแนนจริงที่เป็นไปได้กับเทคนิคที่ศึกษาแต่เมื่อความคงที่ของการประมาณความสามารถถูกตรวจสอบ ผลที่ได้ไม่เป็นที่พึงพอใจ นอกจากนี้ยังได้เปรียบเทียบผลของความคลาดเคลื่อนของการวัดที่ช่วยในการแปลความหมายขนาดและหาความสัมพันธ์ระหว่างการประมาณความสามารถ

กาฟนิและเมแลม (Gafni and Melamed,1990) ใช้วิธีเชิงเส้นตรงสำหรับการปรับเทียบกลับสู่แบบสอบเดิม ที่มีการจัดการสอบในสถานการณ์ที่ต่างกัน 2 สถานการณ์ เกณฑ์ที่ใช้ในการประเมินความคลาดเคลื่อนของการปรับเทียบคะแนน คือ ความแตกต่างของค่าเฉลี่ยระหว่างคะแนนที่แท้จริงกับคะแนนที่ปรับเทียบแล้ว และดัชนี RMSE ข้อค้นพบคือ สิ่งที่ชี้ให้เห็นปฏิสัมพันธ์ระหว่างวิธีการปรับเทียบคะแนนและรูปแบบที่ใช้ที่ครอบคลุมการแปลผล และสรุปว่าการใช้รูปแบบวงจรการปรับเทียบไม่มีความเหมาะสมเสมอไป สำหรับการประมาณความคลาดเคลื่อนของการปรับเทียบคะแนน

แฮร์ริสและโคเลน (Harris and Kolen,1990) ใช้ดัชนี RMSE ที่ถ่วงน้ำหนักด้วยความถี่ ดัชนี MAD และดัชนี MSD ประเมินความแตกต่างระหว่างวิธีการปรับเทียบคะแนนวิธีอิควิเปอร์เซ็นต์ไทล์ 2 วิธี รูปแบบใช้แบบสอบร่วม ได้ตรวจสอบการเลือกความเท่าเทียมกันของคะแนนดิบ ความแตกต่างระหว่างความเท่าเทียมกัน ความถี่ของความคลาดเคลื่อนมาตรฐาน

ที่ประมาณได้ และอัตราของความแตกต่างที่สังเกตได้ของความคลาดเคลื่อนมาตรฐาน

แมคแคน (MacCann, 1990) ได้ปรับเทียบคะแนนกลับสู่แบบสอบเดิมโดยตรงโดยไม่ผ่านแบบสอบอื่น แต่ผ่านแบบสอบร่วมภายนอก เพราะความสัมพันธ์ของคะแนนที่ปรับควรเป็นเอกลักษณ์ (Identity) เกณฑ์การปรับเทียบคะแนนคือเส้นตรงที่มีความชันเป็น 1 และค่าคงที่เป็น 0 มีการตรวจสอบวิธีการปรับเทียบคะแนน แบบสอบร่วมและความแตกต่างความสามารถของประชากร โดยใช้ดัชนี RMSE มาตรฐานในการหาความแตกต่างรวม

ไรท์และดูแรนด์ (Wright and Dorans,1990) ได้ประเมินความถูกต้องแม่นยำของการปรับเทียบคะแนนกลุ่มเท่าเทียมกัน วิธีการปรับเทียบอิควิเปอร์เซ็นต์ไทล์ โดยคำนวณจากกลุ่มประชากรย่อยที่มากกว่า 115,000 คน ดัชนีสำหรับการประเมินผลคือ ดัชนี RMSD ที่ถ่วงน้ำหนักด้วยความถี่

ลอเรนซ์และดูแรนซ์ (Lawrence and Dorans, 1990) ได้เปรียบเทียบผลจากการปรับเทียบคะแนนจาก 5 วิธี ใช้แบบสอบร่วม โดยเปรียบเทียบผลภายในและข้ามกลุ่มตัวอย่าง

ไอเนอร์และคณะ (Eignor and others, 1990a) ใช้การจำลองข้อมูลและปรับเทียบคะแนนกลับสู่แบบสอบเดิม เพื่อตรวจสอบผลกระทบของการใช้เกณฑ์ที่ผิดพลาดในการปรับเทียบคะแนนสังเกตได้และคะแนนจริง

เคร้าส์ (Crouse, 1991) ได้เปรียบเทียบวิธีการเก็บรวบรวมข้อมูลในการปรับเทียบคะแนน 3 วิธี ใช้การประมาณ Bootstrap เป็นการประมาณความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนเป็นเกณฑ์

แฮนสันและคณะ (Hanson and others, 1991) ใช้การจำลองข้อมูลเปรียบเทียบวิธี Presmoothing และวิธี Postsmoothing สำหรับการปรับเทียบคะแนนวิธีอิควิเปอร์เซ็นต์ไทล์

เบคเกอร์ (Baker, 1991) ได้เปรียบเทียบวิธีการปรับเทียบอิควิเปอร์เซ็นต์ไทล์ 3 วิธี ใช้ดัชนี RMSD ที่ถ่วงน้ำหนักด้วยความถี่ และความเท่าเทียมกันแบบอิควิเปอร์เซ็นต์ไทล์เป็นเกณฑ์เปรียบเทียบ

เบคเกอร์และอัลคาร์นิ (Baker and Al-Karni, 1991) ใช้ทั้งข้อมูลจริงและข้อมูลที่จำลองขึ้น เพื่อเปรียบเทียบ 2 กระบวนการคำนวณสัมประสิทธิ์การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบใช้การปรับเทียบคะแนนกลับสู่แบบสอบเดิม ใช้ดัชนีสรุปผลคล้ายกับดัชนี RMS

เวย์และทาง (Way and Tang, 1991) ได้เปรียบเทียบวิธีการปรับเทียบคะแนนเชิงตรรกวิทยา (Logistic Equating Methods) ใช้ทั้งข้อมูลจริงและข้อมูลที่จำลองขึ้น

เซง (Zeng, 1991) ได้ศึกษาความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนเชิงเส้นตรง สำหรับการเก็บรวบรวมข้อมูลใช้ผู้สอบกลุ่มเดียว มีการจำลองข้อมูลเพื่อใช้ในการหา

ความคลาดเคลื่อนมาตรฐาน Bootstrap และใช้ข้อมูลจริงประเมินคุณลักษณะของการประมาณความคลาดเคลื่อนมาตรฐาน

เวย์และรีส (Way and Reese, 1991) ได้ใช้โมเดล IRT ศึกษาการทำสเกลคะแนนและการปรับเทียบคะแนนจากแบบสอบ TOEFL โดยการเปรียบเทียบผลการทดลองกับผลจากที่ใช้โมเดล IRT 3 พารามิเตอร์ การเปรียบเทียบดูจากความสัมพันธ์ระหว่างพารามิเตอร์ที่ประมาณได้และพารามิเตอร์ที่จำลองขึ้น ความสอดคล้องของการเปลี่ยนแปลงคะแนนที่จำลองขึ้น กับการเปลี่ยนแปลงจากพารามิเตอร์ที่สร้างขึ้น โดยใช้ดัชนี RMSD

แฮร์ริส (Harris, 1991c) ได้เปรียบเทียบวิธีการจัดแบบสอบร่วมที่เป็นตัวแทน ในการปรับเทียบคะแนนที่ใช้แบบสอบร่วม ใช้ดัชนี RMSD ดัชนี MAD และดัชนี MSD ที่ถ่วงน้ำหนักด้วยความถี่เป็นเกณฑ์

ตอนที่ 8 งานวิจัยที่เกี่ยวข้องกับการปรับเทียบคะแนน

โจแฮนสัน (Johanson, 1988) ศึกษาผลของความยากของแบบทดสอบร่วม ความสามารถของผู้สอบ และวิธีการประมาณค่าพารามิเตอร์ที่มีต่อการเทียบคะแนนตามแนวตั้ง ตามทฤษฎีการตอบสนองข้อสอบ ข้อมูลการศึกษาได้จากการจำลองขึ้นตามแบบจำลองโลจิสติก ประชากรมี 3 กลุ่มแต่ละกลุ่มประกอบด้วยกลุ่มตัวอย่าง 3 กลุ่มย่อย การประมาณค่าพารามิเตอร์ของประชากรใช้โปรแกรม LOGIST แบบทดสอบร่วมที่ใช้มีความยาว 3 ขนาด ในการศึกษาจะศึกษาผลของแบบทดสอบร่วมไปพร้อมกัน ลักษณะประชากร และวิธีการประมาณค่าพารามิเตอร์ ความยากของแบบทดสอบร่วม มีการแจกแจงแบบยูนิฟอร์ม วิธีการประมาณค่าพารามิเตอร์ใช้วิธีประมาณค่าพร้อมกัน วิธีใช้โค้งลักษณะข้อสอบ วิธีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน วิธีกำลังสองน้อยที่สุดแบบออธอกอนอล และวิธีกำลังสองน้อยที่สุดแบบปกติ

ผลการวิจัยพบว่า วิธีปรับเทียบคะแนนโดยใช้โค้งลักษณะข้อสอบเป็นวิธีที่มีความถูกต้องมากที่สุด นอกจากนี้วิธีปรับเทียบคะแนนโดยใช้โค้งลักษณะข้อสอบยังเป็นวิธีเดียวที่ให้ผลการปรับเทียบในระดับที่ยอมรับได้โดยใช้แบบทดสอบร่วมจำนวนน้อยที่สุด สำหรับวิธีการปรับเทียบคะแนนทุกวิธีเมื่อใช้แบบสอบร่วมที่มีจำนวนมากที่สุด และใช้กลุ่มผู้สอบที่มีความแตกต่างกันน้อยที่สุด ให้ผลการปรับเทียบคะแนนที่ยอมรับได้ เมื่อใช้กลุ่มตัวอย่างที่มีความแตกต่างกันมากการประมาณค่าพารามิเตอร์ของข้อสอบทำได้ไม่ดีนัก โดยสรุปแล้วความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ของข้อสอบ อาจทำให้ยากในการปรับเทียบคะแนนระหว่างแบบสอบที่มีระดับความยากแตกต่างกันมาก ถ้าแบบสอบร่วมที่ใช้สั้นกว่าและแบบทดสอบฉบับรวมมีขนาดสั้น

ทริสคาริ (Triscari,1990) ได้ทำการศึกษาเปรียบเทียบวิธีการปรับเทียบคะแนน 5 วิธี โดยใช้แบบแผนการเก็บรวบรวมข้อมูลแบบไม่เท่าเทียมกันของกลุ่มประชากร (Nonequivalent Population) และใช้ข้อสอบร่วมภายในโดยมีการเชื่อมโยงข้อสอบแบบเลือกตอบและแบบเรียงความ วิธีการปรับเทียบที่ใช้คือตามทฤษฎีการวัดแบบดั้งเดิมคือ แบบเส้นตรงและแบบอีควิเปอร์เซ็นไทล์ และแบบ IRT ชนิด 1, 2 และ 3 พารามิเตอร์ ซึ่งใช้การให้คะแนนตามแนว Partial Credit Model การเปรียบเทียบใช้วิธีการจำลองสถานการณ์เมื่อทราบค่าคะแนนจริงที่เท่าเทียมกันของแบบสอบที่เป็นฐาน การประเมินประสิทธิผลใช้ค่าความแตกต่างระหว่างคะแนนจริงของแบบสอบฐาน นอกจากนี้ยังได้เก็บข้อมูลจริงเพื่อหาความสอดคล้องกับข้อมูลที่จำลองขึ้นมาด้วย ผลการวิจัยพบว่า การปรับเทียบตามแนว IRT ที่ใช้ Partial Credit Model และมีการวิเคราะห์รวมให้ความคลาดเคลื่อนน้อยที่สุด และการปรับเทียบคะแนนแบบเส้นตรงให้ความคลาดเคลื่อนมากที่สุด จากการเก็บข้อมูลจริงพบว่ามีความสอดคล้องกับการจำลองข้อมูลร้อยละ 70

เพเทียน (Patiene, 1990) ศึกษาเปรียบเทียบผลของการปรับเทียบคะแนนระหว่างระดับชั้น โดยใช้วิธีปรับเทียบคะแนน 5 วิธี คือวิธีการตามทฤษฎีการวัดแบบดั้งเดิม 2 วิธี ได้แก่ อีควิเปอร์เซ็นไทล์ กับวิธีของเทอร์สโตน และวิธีการตามทฤษฎีการตอบสนองข้อสอบ 2 วิธี คือ วิธีใช้แบบจำลองโลจิสติกแบบหนึ่ง สอง และสามพารามิเตอร์ โดยในการวิจัยครั้งนี้ต้องการศึกษาว่าวิธีการที่ใช้วิธีใดให้ผลการเทียบคะแนนที่เหมาะสมที่สุด แบบทดสอบที่ใช้ในการวิจัยแบ่งเป็น 3 ระดับ คือ ง่าย ยากง่ายปานกลาง และยาก การเทียบคะแนนแบบทดสอบแต่ละฉบับใช้วิธีการปรับเทียบคะแนนทั้ง 5 วิธี แบบทดสอบฉบับง่ายใช้กับนักเรียนระดับ 9 แบบยากง่ายปานกลางใช้กับนักเรียนระดับ 10 และแบบสอบฉบับยากใช้กับนักเรียนระดับ 11 คะแนนที่ได้จากการทดสอบแต่ละระดับจำนำไปเปรียบเทียบกับคะแนนโดยรวมของนักเรียน วิธีการปรับเทียบคะแนนที่เหมาะสมที่สุดกำหนดว่า เป็นวิธีการที่ให้ผลการปรับเทียบคะแนนใกล้เคียงกับผลที่ได้จากคะแนนของนักเรียนจากแบบสอบรวมทุกระดับ

ผลการวิจัยพบว่า วิธีการปรับเทียบทั้ง 5 วิธี ให้ผลการปรับเทียบคะแนนที่คล้ายคลึงกันเมื่อใช้กับแบบสอบฉบับง่าย และฉบับยากง่ายปานกลาง เมื่อใช้กับแบบสอบฉบับยากพบว่าการปรับเทียบคะแนนโดยวิธีอีควิเปอร์เซ็นไทล์ วิธีการของเทอร์สโตน และวิธีสามพารามิเตอร์ให้ผลคล้ายคลึงกับคะแนนที่ได้จากแบบสอบรวม มากกว่าวิธีหนึ่งพารามิเตอร์ และวิธีสองพารามิเตอร์

กลอวากิ (Glowacki,1991) ได้ตรวจสอบโมเดลของการปรับเทียบคะแนนที่มีความเหมาะสมกับการสอบของบัณฑิตวิทยาลัยแห่งมหาวิทยาลัยอลาบามา ปัญหาในการวิจัยคือ

โมเดลของการปรับเทียบที่ตรวจสอบ มีการแจกแจงคะแนนดิบ หรือคะแนนที่ผ่านจากการสอบแบบสอบการอ่าน และคณิตศาสตร์ แตกต่างกันหรือไม่ โมเดลที่ใช้ในการตรวจสอบ คือ โมเดลเชิงเส้นตรง อีควิเปอร์เซนต์ไทล์ และ IRT ชนิด 1, 2 และ 3 พารามิเตอร์ ผลการวิจัยพบว่า วิธีการปรับเทียบคะแนนทั้ง 5 โมเดล ในการสอบการอ่านและคณิตศาสตร์ให้ผลที่คล้ายคลึงกัน แสดงว่าโมเดลทั้งหมดสามารถนำมาใช้กับการปรับเทียบคะแนนได้ โดยไม่มีโมเดลใดดีที่สุด

แฮริส (Harris, 1991) ศึกษาผลของแบบแผนการรวบรวมข้อมูล 2 วิธี คือแบบแผนที่ 2 ของแองกอฟ (ผู้สอบกลุ่มเดียวจัดให้สมดุล) และแบบแผนที่ 2 ของแองกอฟ (ผู้สอบกลุ่มเท่าเทียมกัน) ที่มีต่อการเทียบคะแนนตามแนวตั้งโดยใช้วิธีปรับเทียบคะแนน 2 วิธี คือ วิธีอีควิเปอรเซ็นไทล์ และวิธีตามทฤษฎีการตอบสนองข้อสอบสามพารามิเตอร์ แบบทดสอบที่ใช้ คือ แบบทดสอบวิชาคณิตศาสตร์ 2 ฉบับ คือ EACT มีจำนวน 60 ข้อ และฉบับ PACT มีจำนวน 40 ข้อ แบบสอบทั้งสองฉบับมีเนื้อหาคาบเกี่ยวกัน แต่มีเนื้อหาเฉพาะและระดับความยากต่างกัน กลุ่มตัวอย่างเป็นนักเรียนระดับ 11 และ 12 ซึ่งแบ่งเป็นกลุ่มย่อยหลายกลุ่ม เกณฑ์ที่ใช้ในการประเมิน คือ ดัชนีความแตกต่าง RMSD ผลการวิจัยสรุปได้ว่าผลการปรับเทียบคะแนนตามแนวตั้งโดยใช้แบบแผนการเก็บข้อมูลทั้ง 2 แบบ มีความเพียงพอและให้ผลคล้ายคลึงกัน โดยรูปแบบที่ใช้ผู้สอบกลุ่มเดียวให้ผลคงที่กว่ารูปแบบที่ใช้ผู้สอบกลุ่มเท่าเทียมกันเล็กน้อย

อเยอเว (Ayerve, 1992) ได้ศึกษาเปรียบเทียบประสิทธิภาพของการปรับเทียบคะแนนตามแนวตั้งด้วยวิธีอีควิเปอร์เซ็นไทล์ และวิธีสามพารามิเตอร์ โดยใช้การจำลองข้อมูลภายใต้เงื่อนไขของขนาดตัวอย่าง ความยาวของแบบสอบ และความยากของแบบสอบร่วมการเปรียบเทียบทำ 3 กรณี คือ 1) เปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบคะแนนทั้งสองภายใต้เงื่อนไขทุกเงื่อนไข 2) เปรียบเทียบประสิทธิภาพของวิธีการปรับเทียบคะแนนแต่ละวิธีภายใต้เงื่อนไขของตัวแปรอิสระแต่ละตัว 3) ตรวจสอบผลของตัวแปรอิสระแต่ละตัวในการปรับเทียบคะแนนแต่ละวิธี ขนาดกลุ่มตัวอย่างที่ใช้มี 3 ขนาด คือ 200, 500 และ 1,000 ความยาวของข้อสอบมี 2 ขนาด คือ 30 ข้อ และ 60 ข้อ ส่วนความยาวของแบบทดสอบร่วมมี 2 ขนาด คือ 5 ข้อ และ 10 ข้อ สถิติใช้เป็นเกณฑ์การประเมินมี 2 อย่าง คือ ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนซึ่งถ่วงน้ำหนักแล้ว (Weighted Mean Square Error : WMSE) และค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนซึ่งไม่ถ่วงน้ำหนัก (UMSE)

ผลการวิจัยพบว่า การปรับเทียบคะแนนตามเงื่อนไขต่าง ๆ โดยใช้วิธีการปรับเทียบคะแนนทั้งสองวิธีไม่แตกต่างกัน สำหรับการเทียบคะแนนโดยวิธีอีควิเปอร์เซ็นไทล์ พบว่า ความยาวของแบบสอบกับความยาวของแบบสอบร่วมเป็นองค์ประกอบสำคัญ ส่วนการเทียบคะแนน

วิธีสามพารามิเตอร์ พบว่า ขนาดกลุ่มตัวอย่างเป็นองค์ประกอบสำคัญ ขนาดตัวอย่างเล็ก (200) มีแนวโน้มที่จะให้ผลการเทียบคะแนนที่ไม่ถูกต้อง ในขณะที่ขนาดกลุ่มตัวอย่างที่ใหญ่ขึ้น (500 และ 1,000) จะให้ผลการปรับเทียบคะแนนที่มีความถูกต้องมากกว่า

คิมและโคเฮน (Kim and Cohen, 1995) ได้ศึกษาเปรียบเทียบวิธีการปรับเทียบคะแนนระหว่างแบบสอบที่มีการตรวจให้คะแนนแบบพหุวิภาค ตามโมเดล Graded Response Model วิธีประมาณค่าไคว์สแคว์ต่ำสุด (A Minimum χ^2 Method) 2 วิธี คือ วิธี Diagonal χ^2 Method กับวิธี Full χ^2 Method และ วิธี. โค้งคุณลักษณะข้อสอบ ใช้สถานการณ์จำลองข้อมูล ศึกษาปรับเทียบแนวนอน ออกแบบการเก็บข้อมูลใช้รูปแบบผู้สอบกลุ่มเท่าเทียมกัน ใช้แบบสอบกลุ่มละ 30 ข้อ กลุ่มตัวอย่าง 300 คน พบว่า สัมประสิทธิ์การปรับเทียบคะแนนหรือค่าคงที่จากทั้งสามวิธีมีค่าใกล้เคียงกัน นอกจากนี้ใช้ในการปรับเทียบในแนวตั้ง ออกแบบการเก็บข้อมูลรูปแบบใช้ข้อสอบร่วมภายใน พบว่า สัมประสิทธิ์การปรับเทียบหรือค่าคงที่ที่ได้จากทั้งสามวิธีมีค่าใกล้เคียงกันด้วย

ภาวิณี ศรีสุขวัฒนานันท์ (2528) ได้ทำการศึกษาเปรียบเทียบผลการปรับเทียบคะแนนระหว่างแบบสอบ วิธีอีควิเปอร์เซนต์ไทล์ วิธีเชิงเส้นตรง และวิธี IRT 3 พารามิเตอร์ โดยใช้ข้อสอบร่วมที่มีขนาดต่างกัน 3 ขนาด คือ ร้อยละ 60 ร้อยละ 40 และร้อยละ 20 ในการสอบผลสัมฤทธิ์ทางการเรียน และการสอบคัดเลือก โดยใช้กลุ่มตัวอย่างกลุ่มละ 1,500 คน การประเมินผลการปรับเทียบคะแนน ใช้การวิเคราะห์หาความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนน และการวิเคราะห์กลุ่มสอบทานผล ผลการศึกษาพบว่า 1) วิธีการปรับเทียบทุกวิธีให้ผลในระดับที่ยอมรับได้ 2) ความยาวของแบบสอบร่วมเป็นปัจจัยที่มีผลกระทบต่อความแม่นยำและความเพียงพอของวิธีการปรับเทียบ 3) สถานการณ์การสอบผลสัมฤทธิ์ทางการเรียนและการสอบคัดเลือกภายใต้วิธีการปรับเทียบเดียวกันให้ผลที่แตกต่างกัน

วรรณ์ดี แสงประทีปทอง (2538) ได้ทำการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียนชั้น ม.1, ม.2 และ ม.3 โดยใช้วิธีการปรับเทียบคะแนนตามแนวตั้ง ตามทฤษฎีการตอบสนองข้องสอบ แบบจำลองโลจิสติกสามพารามิเตอร์ ใช้ข้อสอบร่วมภายนอก ปรับคะแนนให้อยู่บนสเกลเดียวกัน แล้วเปรียบเทียบผลสัมฤทธิ์ทางการเรียนของนักเรียนแต่ละชั้น พบว่า 1) ผลสัมฤทธิ์ทางการเรียนภาษาไทยของนักเรียน ม.2 ไป ม.3 ก้าวหน้ามากกว่าปกติ แต่ผลสัมฤทธิ์ทางการเรียนภาษาไทย ม.1 ไป ม.2 ก้าวหน้าตามปกติ 2) ในกลุ่มนักเรียนที่มีความสามารถสูง ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยจาก ม.2 ไป ม.3 และ จาก

ม.1 ไป ม.2 ก้าวหน้าตามปกติ 3) ในกลุ่มนักเรียนที่มีความสามารถปานกลาง ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของ ม.2 ไป ม.3 และ ม.1 ไป ม.2 ก้าวหน้ามากกว่าปกติ 4) ในกลุ่มนักเรียนที่มีผลสัมฤทธิ์ทางการเรียนต่ำ ผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทยของนักเรียน ม.2 ไป ม.3 ก้าวหน้ามากกว่าปกติ แต่ผลสัมฤทธิ์ทางการเรียนของ ม.1 ไป ม.2 ก้าวหน้าตามปกติ

พรพิมล นาคเวช (2538) ได้ทำการศึกษา เพื่อตรวจสอบคุณภาพของการปรับเทียบคะแนนระหว่างแบบสอบในแนวตั้ง โดยใช้ทฤษฎีตอบสนองข้อสอบ สำหรับผู้ที่มีความสามารถต่างกัน 3 ระดับ คือ ม.1 ม.2 และ ม.3 กลุ่มตัวอย่างเป็นนักเรียนสังกัดกรมสามัญศึกษา จำนวน 6,429 คน เครื่องมือที่ใช้เป็นแบบสอบวิชาคณิตศาสตร์ 3 ฉบับ ฉบับละ 25 ข้อ โดยมีข้อสอบร่วมภายใน 5 ข้อ พบว่า ความคลาดเคลื่อนมาตรฐานของการปรับเทียบคะแนนสำหรับผู้มีความสามารถต่างกัน 3 ระดับ ค่อนข้างต่ำ มีค่าอยู่ระหว่าง 0.031 – 0.051 และความเพียงพอของการปรับเทียบคะแนนระหว่างแบบสอบตามแนวตั้ง สำหรับผู้มีความสามารถต่างกัน 3 ระดับ พบว่าค่าดัชนีความแตกต่างของการปรับเทียบคะแนน (C) อยู่ในระดับน่าพอใจ

พิชัย ละแมนชัย (2538) ได้ศึกษาเพื่อหาขนาดกลุ่มตัวอย่างขั้นต่ำที่ทำให้การปรับเทียบคะแนนในแนวระดับ ตามแนวทฤษฎีการตอบสนองรายข้อ โมเดลหนึ่ง และสามพารามิเตอร์ในแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วม ที่มีความยาวแบบสอบต่าง ๆ กัน คือ 30, 50, 70, 90, 110, 130 และ 150 ข้อ เพื่อให้ผลการปรับเทียบคะแนนในแนวระดับเกิดประสิทธิภาพ ข้อมูลที่ใช้ในการวิจัยจำลองจากโปรแกรม IRTDATA เกณฑ์ที่ใช้ในการตัดสินคือการหาค่าดัชนีความแตกต่างที่อยู่ในระดับน่าพอใจอย่างมาก พบว่า โมเดลการตอบสนองรายข้อ 1 และ 3 พารามิเตอร์ ทั้งแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วม ขนาดกลุ่มตัวอย่างขั้นต่ำในแต่ละความยาวแบบสอบแตกต่างกัน โดยโมเดล 1 พารามิเตอร์ใช้ขนาดกลุ่มตัวอย่างขั้นต่ำน้อยกว่า และโมเดลการตอบสนองรายข้อที่ใช้ 1 และ 3 พารามิเตอร์ทั้งแบบแผนกลุ่มสมมูลและแบบแผนข้อสอบร่วมขนาดกลุ่มตัวอย่างขั้นต่ำในแต่ละความยาวแบบสอบแตกต่างกัน โดยแบบแผนข้อสอบร่วมใช้ขนาดกลุ่มตัวอย่างขั้นต่ำน้อยกว่า

จากผลงานวิจัยที่รวบรวมมานี้ แสดงให้เห็นว่าในการศึกษาเกี่ยวกับการปรับเทียบคะแนนระหว่างแบบสอบ ในด้านการเปรียบเทียบรูปแบบการปรับเทียบคะแนน ศึกษาความยาวของแบบสอบ ความยาวแบบสอบร่วม ขนาดกลุ่มตัวอย่าง รูปแบบการเก็บรวบรวมข้อมูล โดยมีการศึกษาทั้งใช้ข้อมูลจริงและข้อมูลที่จำลองขึ้นเพื่อหารูปแบบที่เหมาะสม งานวิจัยที่ใช้กลุ่มตัวอย่างเป็นตัวแปรได้ผลไปในทำนองเดียวกันว่า กลุ่มตัวอย่างขนาดใหญ่จะให้ผลในการ

ปรับเทียบที่ดีกว่า ความยาวของแบบสอบร่วมได้ผลไปในลักษณะเดียวกันว่า จำนวนข้อสอบร่วมที่มากกว่าจะให้ผลการปรับเทียบคะแนนที่แม่นยำกว่า ในเรื่องวิธีการปรับเทียบคะแนนนั้นยังได้ข้อสรุปที่ขัดแย้งกันอยู่ ทั้งนี้ขึ้นอยู่กับสถานการณ์ที่ทำการศึกษา และมีการใช้เกณฑ์ที่หลากหลายเพื่อพิจารณาความแตกต่างหรือตัดสินผลการปรับเทียบคะแนน ที่ใช้กับการปรับเทียบวิธีเดียวกันแต่ต่างบริบท หรือการปรับเทียบที่ใช้วิธีปรับเทียบแตกต่างกัน ยังไม่มีข้อสรุปว่าเกณฑ์ใดมีคุณภาพดีที่สุด