บทที่ 2

วรรณคดีที่เกี่ยวข้อง

จากการศึกษา เอกสาร บทความ และงานวิจัยที่เกี่ยวกับการเทียบมาตรา ผู้วิจัยขอเสนอวรรณคดีที่เกี่ยวข้อง ตามหัวข้อต่อไปนี้

1. แนวคิดเชิงทฤษฎีและนิยามของการเทียบมาตรา
2. วิธีเทียบมาตราวิธีอิควิเปอร์เซนไทล์และวิธีเชิงเส้นตรง
3. การออกแบบการรวบรวมข้อมูล
4. การหาคุณภาพของการเทียบมาตรา
5. งานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบวิธีการเทียบมาตรา

แนวคิดเชิงทฤษฎีและนิยามของการเทียบมาตรา

กิลลิคเสน (Gulliksen 1950: 298-304) ได้ให้ความหมายของการเทียบมาตราว่า คือ วิธีการทำคะแนนจากแบบสอบสองชุดของวิชาเดียวกันให้เป็นคะแนนสมมูลที่เปรียบเทียบกันได้โดยตรง โดยเสนอวิธีการให้ผู้สอบกลุ่มเดียวทำแบบสอบสองชุดและใช้วิธีง่าย ๆ คือ แปลงคะแนนแต่ละชุดให้เป็นคะแนนมาตรฐานแล้วนำคะแนนที่แปลงแล้วมาเทียบกันโดยตรง แต่ก่อนอื่นให้ตรวจสอบความเป็นคู่ขนานของแบบสอบโดยใช้สถิติของวิคส์ (Wilks) หรือ พิจารณาจากค่าสหสัมพันธ์ระหว่างแบบสอบแต่ละชุดกับเกณฑ์ ถ้ามีค่าเท่ากันโดยประมาณ ($r_{yc} = r_{xc}$) ก็จะสามารถบอกถึงความเพียงพอในการเทียบมาตราแบบสอบทั้งสองชุด

ฟลานาแกน (Flanagan 1951: 747-748) ให้นิยามการเทียบมาตราว่า เป็นวิธีการในการทำคะแนนจากแบบสอบต่างชุดให้สามารถนำมาเปรียบเทียบกันได้ คำว่า "ความสามารถในการเปรียบเทียบกันได้" มีความหมายเฉพาะที่ว่า เมื่อกำหนดประชากรให้ ถ้าการแจกแจงของคะแนนจริงจากแบบสอบทั้งสองชุดซึ่งสอบกับกลุ่มตัวอย่างที่เลือกมาขนาดใหญ่ ๆ มี

ลักษณะเหมือนกันแล้วคะแนนดิบจากแบบสอบทั้งสองชุดจึงจะสามารถเปรียบเทียบกันได้ การแปลงระบบหน่วยการวัดของแบบสอบชุดหนึ่งไปสู่ระบบหน่วยการวัดของแบบสอบอีกชุดหนึ่ง ซึ่งหลังจากการแปลงแล้วก็จะสามารถเทียบหาคะแนนสมมูลระหว่างแบบสอบต่างชุดได้

มาร์โค (Marco 1981) ได้สรุปนิยามของการเทียบมาตรา ไว้ 6 รูปแบบ คือ

นิยามที่ 1 คะแนนจากแบบสอบ X และแบบสอบ Y นำมาเทียบมาตรากันได้

ถ้า $M_{y'} = M_x$ และ

$SD_{y'} = SD_x$ สำหรับประชากร P

เมื่อ M และ SD คือ ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน

y' คือ คะแนนแปลงที่ได้จากฟังชั่นการเทียบมาตรา $e_x(Y)$

ในกรณีเช่นนี้วิธีการเทียบมาตราเชิงเส้นตรงสามารถแปลงคะแนน Y ไปสู่คะแนนของ X อย่างเพียงพอ ด้วยสมการ

$$y' = (SD_x/SD_y)\, y + M_x - (SD_x/SD_y)\, M_y$$

โดยทั่วไป M_y ไม่เท่ากับ M_x และ SD_y ไม่เท่ากับ SD_x เพราะว่ามีความหมายแตกต่างกันในความยากของคำถามในแบบสอบต่างชุดกัน การแปลงเชิงเส้นตรงจะปรับ M_y ให้เท่ากับ M_x และ SD_y เท่ากับ SD_x และเป็นการอ้างถึงการจัดการให้คะแนนเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานให้เท่ากัน นิยามนี้ไม่มีข้อกำหนดว่า แต่ละรายบุคคลจะต้องมี y เท่ากับ x เพียงแต่กำหนดให้คะแนนเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานจะต้องเท่ากันในประชากรที่ศึกษาและโดยนิยามนี้ไม่มีข้อกำหนดว่า แบบสอบ X และ Y จะต้องวัดสิ่งเดียวกัน เช่น ความรู้ หรือความสามารถ หรือทักษะต่าง ๆ

นิยามที่ 2 คะแนนจากแบบสอบ X และแบบสอบ Y นำมาเทียบมาตรากันได้ ถ้า y และ x มีตำแหน่งเปอร์เซนไทล์เดียวกันในการแจกแจงของคะแนน X และ Y ในประชากร P

นิยามนี้ตรงกับนิยามวิธีอิควิเปอร์เซนไทล์ ฉะนั้นใช้วิธีคำนวณตำแหน่งเปอร์เซนไทล์ของคะแนนการแจกแจงใน X และ Y แล้วหาตำแหน่งคะแนนที่สมนัยกันในทางปฏิบัติการแจกแจงของ X และ Y จะถือเสมือนหนึ่งเป็นการแจกแจงต่อเนื่องมากกว่าเป็นคะแนนขาดตอนและตำแหน่งเปอร์เซนไทล์ใช้การคำนวณด้วย linear interpolation การแจกแจงน่าจะปรับเสียก่อนการคำนวณ นิยามนี้ไม่มีข้อกำหนดว่า X และ Y ต้องวัดในสิ่งเดียวกัน

นิยามที่ 3 คะแนนจากแบบสอบ X และแบบสอบ Y จะนำมาเทียบมาตรากันได้ เมื่อ $M_{y'} = M_x$ สำหรับประชากรกลุ่มย่อยที่มีความสามารถ ความรู้ และทักษะที่ทำการวัดอยู่ในระดับเดียวกัน นั่นคือ

$$M_{y'/x} = M_{x/a}$$

a คือ ระดับความสามารถที่กล่าวถึง

ฟังชั่นการเทียบมาตราที่เป็นไปตามนิยามต้องเป็นฟังชั่นที่ไม่ใช่เส้นตรง(curvilinear) การใช้วิธีเชิงเส้นตรงอาจให้การประมาณที่เป็นประโยชน์ได้ ส่วนวิธีอิควิเปอร์เซนไทล์อาจให้ผลที่คงเส้นคงวาตามนิยามหรืออาจไม่เป็นเช่นนั้นก็ได้แต่ที่ชัดเจน คือต้องใช้วิธีที่สลับซับซ้อนยิ่งขึ้น จึงเพียงพอที่จะให้เกิดการแปลงที่ให้คะแนนสมมูลได้ เช่น วิธีที่อาศัยวิธีอิงทฤษฎีการตอบข้อสอบ (Lord 1980) ซึ่งจะให้ค่าประมาณของระดับความสามารถที่อยู่ภายในกลุ่ม นิยามที่ 3 นี้ ได้กำหนดให้ X และ Y วัดในสิ่งเดียวกันวิธีทางสถิติที่จะให้ผลลัพธ์อย่างน่าพอใจตามนิยามก็ต่อเมื่อได้เป็นไปตามข้อตกลงเบื้องต้นของนิยาม ไม่มีข้อกำหนดใด ๆ ที่เกี่ยวกับความคมชัดของแบบสอบ ฉะนั้นจึงสามารถใช้เทียบมาตรากับแบบสอบที่มีความเที่ยงต่างกันได้

นิยามที่ 4 คะแนนจากแบบสอบ X และแบบสอบ Y นำมาเทียบมาตรากันได้ เมื่อ $M_{y'} = M_x$ ในทุก ๆ กลุ่มย่อยที่มีความสามารถระดับเดียวกันในประชากร P นั่นคือ $M_{y'/a} = M_{x/a}$ และคะแนนความคลาดเคลื่อนมาตรฐานของการวัด y และ x เท่ากันในประชากร P

นิยามนี้เพิ่มข้อกำหนดจากนิยามที่ 3 อีกหนึ่งข้อ คือ แบบสอบ X และ Y ต้องมีความแม่นยำในการวัดเท่ากัน ซึ่งหมายถึง มีค่าความเที่ยงเท่ากัน ในกรณีนี้แบบสอบ X และ Y ควรประกอบด้วยข้อสอบจำนวนที่เท่ากัน

นิยามที่ 5 คะแนนจากแบบสอบ X และแบบสอบ Y นำมาเทียบมาตรากันได้ เมื่อ $M_{y'} = M_x$ และ $SD_{y'} = SD_x$ สำหรับทุกๆกลุ่มย่อยที่มีระดับความสามารถเดียวกันในประชากร P $M_{y'/a} = M_{x/a}$ และ $SD_{y'/a} = SD_{x/a}$ สำหรับทุก ๆ a

นิยามนี้ยากแก่การจัดการให้มีข้อกำหนดตามที่ระบุไว้ เพราะไม่ใช่เฉพาะแต่ความคลาดเคลื่อนมาตรฐานโดยส่วนรวมจะต้องเท่ากัน แต่หมายถึงความคลาดเคลื่อนมาตรฐานของทุก ๆ กลุ่มย่อยต้องเท่ากันหมดด้วย โดยทางปฏิบัตินิยามนี้ไม่มีทางทำได้เลย นอกจากต้องอาศัยการดำเนินการสอบด้วยคอมพิวเตอร์ และแต่ละคนจะต้องได้รับจำนวนคำถามที่มากพอที่จะให้การประมาณที่แม่นยำเพื่อให้เป็นไปตามนิยาม

นิยามที่ 6 คะแนนจากแบบสอบ X และแบบสอบ Y จะเทียบมาตรากันได้ ถ้าการแจกแจงความถี่ของ y และ x เหมือนกันในแต่ละกลุ่มตัวอย่างย่อยที่มีความสามารถอยู่ในระดับเดียวกัน

นิยามที่ 6 นี้ มีข้อกำหนดมากขึ้นกว่านิยามที่ 5 ในเรื่องของรูปร่างของการแจกแจงของคะแนน y อย่างมีเงื่อนไขที่ระดับความสามารถใด ๆ ทั้งนี้ย่อมหมายถึงว่า $M_{y'/a}$ และ $SD_{y'/a}$ จะต้องเข้าชุดกับรูปร่างของการแจกแจงอย่างมีเงื่อนไขของ X ลอร์ด (Lord 1980)ได้แสดงให้เห็นว่า นิยามนี้เป็นไปได้ต่อเมื่อคำถามในแบบสอบ Y มีความเท่าเทียมหรือสมมูลในเชิงหน้าที่กับคำถามในแบบสอบ X เท่านั้นหรือว่าแบบสอบทั้งสองฉบับสามารถให้ผลการวัดเป็นคะแนนสมบูรณ์ ในกรณีเช่นนี้ $y = x$ นิยามนี้ดูเหมือนจะไม่มีประโยชน์ แต่เป็นนิยามในเชิงทฤษฎีที่ต้องการให้มีความหมายของความไม่แตกต่างของการสอบ X หรือ Y ของบุคคลใด ๆอย่างแท้จริง

สงบ ลักษณะ (2522: 22) ได้กล่าวว่า คะแนนจากแบบทดสอบ 2 ฉบับ วัดสิ่งเดียวกันแต่ไม่จำเป็นต้องเป็นข้อสอบคู่ขนาน จะถือว่าเทียบเท่ากันได้ถ้าคะแนนจากแบบทดสอบทั้ง 2 ฉบับนั้นมาจากคะแนนจริง (True score) หรือความสามารถแท้ (True ability) ที่เท่ากัน

ชูศักดิ์ ขัมภลิขิต (2527: 2) ได้สรุปเกี่ยวกับการเทียบมาตราว่าเป็นกระบวนการที่เกี่ยวโยงกับกิจกรรม 2 ประการ คือ

1. กระบวนการที่ทำให้แบบทดสอบ 2 ฉบับใด ๆ มีความเท่าเทียมกันหรือเท่ากันในเชิงโครงสร้าง

2. การใช้วิธีการทางสถิติเพื่อปรับ (Adjust) คะแนนที่ได้จากแบบทดสอบแต่ละฉบับให้อยู่ในมาตราเดียวกันและเทียบกันได้

วิธีเทียบมาตราแบบอิควิเปอร์เซนไทล์

มรินทัศน์ของการเทียบมาตราวิธีอิควิเปอร์เซนไทล์ เริ่มจากการแจกแจงของคะแนนจากแบบสอบชุด X และชุด Y ที่มีลักษณะคล้ายกันหรือแตกต่างเกิดขึ้นบ้าง ก็มีเพียงเล็กน้อย เช่น ต่างกันที่ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน การเทียบมาตราสมมูลทำได้โดยใช้คะแนน ณ ตำแหน่งเปอร์เซนไทล์เดียวกันของคะแนน 2 ชุดนั้น ผลการเทียบคะแนนแสดงด้วยกราฟ ขั้นตอนการแปลงคะแนนมีดังนี้ คือเลือกกลุ่มผู้สอบที่มีความสามารถกระจาย มีทั้งเก่ง ปานกลาง และอ่อน แล้วแบ่งกลุ่มย่อยโดยสุ่ม 2 กลุ่มให้กลุ่มหนึ่งทำแบบสอบ X และอีกกลุ่มหนึ่งทำแบบสอบ Y ผลการสอบนำมาทำการแจกแจงคะแนน X และ Y คำนวณหาจุดกลางเปอร์เซน-ไทล์ของแต่ละการแจกแจง อ่านและทำเครื่องหมายสำหรับค่า X และ Y จากการแจกแจงคะแนนที่สมนัยกันบนกระดาษกราฟ ทำจุดบนกระดาษประมาณ 30 จุด แล้วลากเส้นเชื่อมจุดจะเกิดเป็นเส้นกราฟ ทำการปรับเส้นให้เรียบเส้นกราฟนี้จะใช้อ่านค่า x ที่สมนัยกับ y หรืออ่านค่า y ที่สมนัยกับ x ก็ได้ จากนั้นทำตารางสำเร็จเพื่ออ่านค่าคะแนนแปลง โดยปกติการเทียบมาตราวิธีอิควิเปอร์เซนไทล์ให้ภาพของการแปลงที่สะท้อนถึงระดับความยากง่ายของแบบสอบสองชุด ถ้าแบบสอบสองชุดมีความยากใกล้เคียงกันเส้นกราฟจะมีลักษณะใกล้เคียงเส้นตรง แต่ถ้าแบบสอบมีความยากต่างกันเส้นกราฟจะเป็นเส้นโค้ง (curvilinear) คะแนนสมมูลที่เกิดขึ้นจะถูกยืดหรือหย่นคะแนนดิบเพื่อให้คงรักษาคะแนนให้เหมือนชุดก่อนตามต้องการ (Angoff 1984: 97-101)

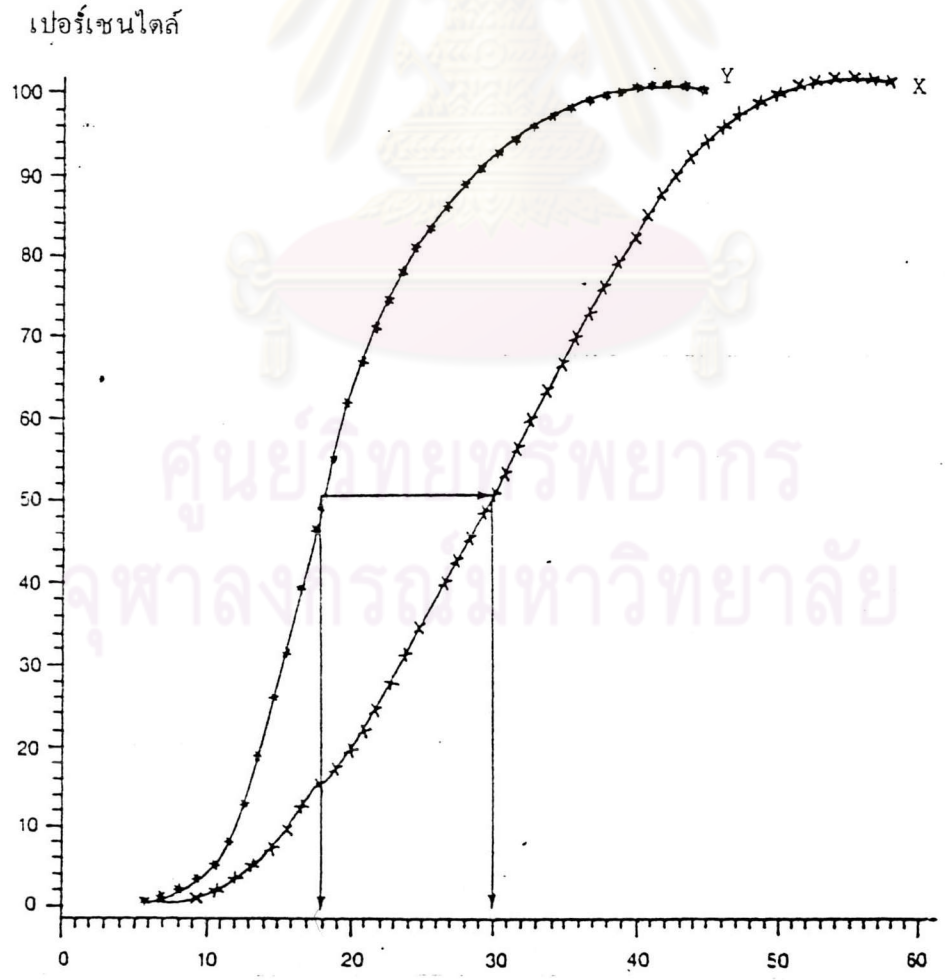
ขั้นตอนในการเทียบมาตราวิธีอิควิเปอร์เซนไทล์ สรุปเป็นขั้นตอนได้ 2 ขั้นตอน (two-stage) คือ

ขั้นตอนที่หนึ่ง การกระจายความถี่สะสมที่มีความสัมพันธ์เป็นตารางหรือเป็นกราฟสำหรับสองแบบสอบที่นำมาเทียบมาตรา คือ

1) นำคะแนนของกลุ่มผู้สอบที่มีความสามารถกระจายทั้ง เก่ง ปานกลางและอ่อน ซึ่งถูกแบ่งเป็นกลุ่มย่อยสองกลุ่มโดยการสุ่ม ให้กลุ่มหนึ่งทำแบบสอบ X และอีกกลุ่มทำแบบสอบ Y มาทำการแจกแจงคะแนน X และ Y

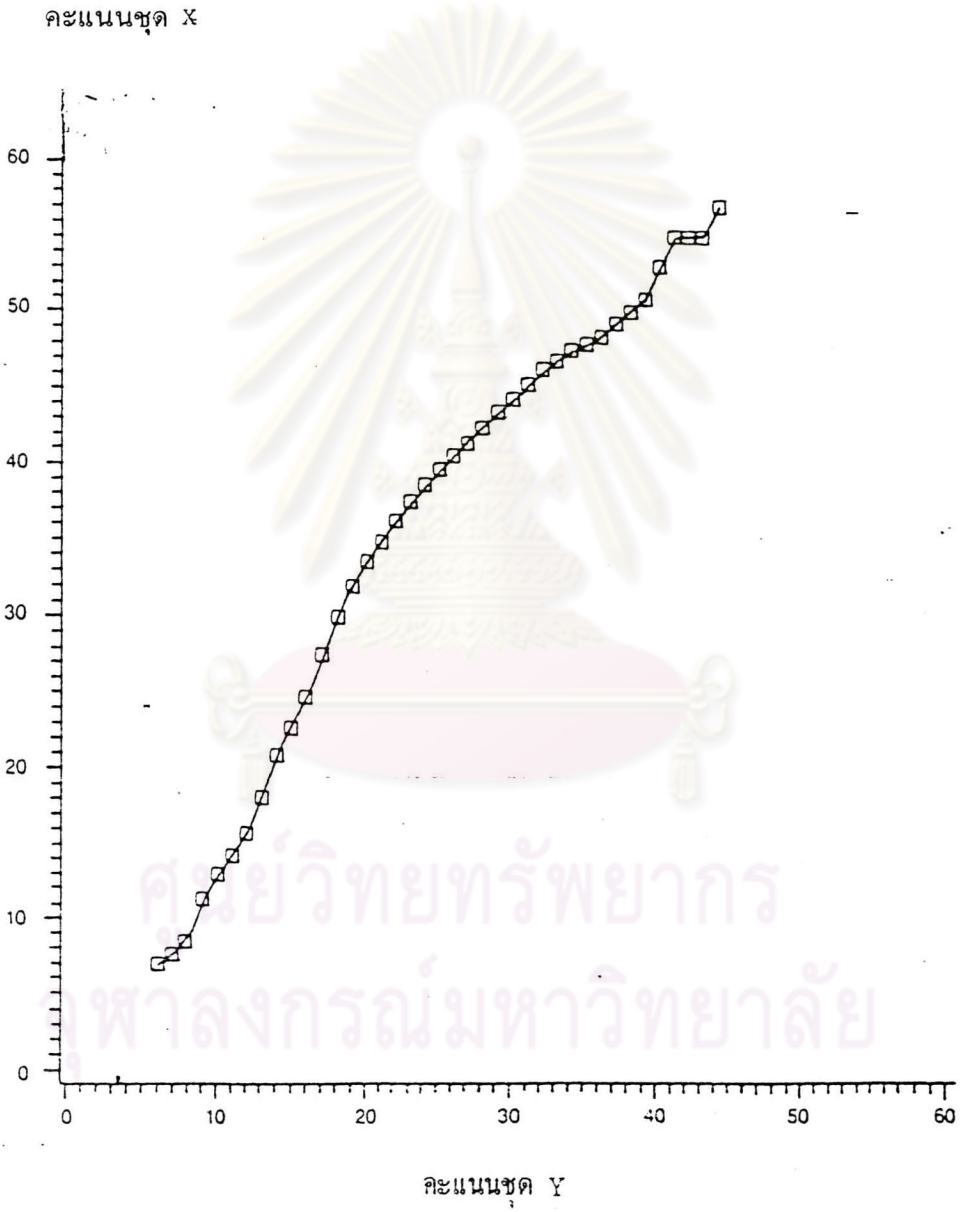
2) คำนวณหาจุดกลางเปอร์เซนไทล์ของแต่ละการแจกแจง

3) อ่านและทำเครื่องหมายสำหรับค่าคะแนนของแบบสอบฉบับ X และฉบับ Y ของการแจกแจงที่สมมูลกันบนกระดาษกราฟ ซึ่งแองกอฟฟ์แนะนำให้ใช้กระดาษ Arithmatic graph paper (โดยแกนนอนเป็นคะแนนดิบ แกนตั้งเป็นตำแหน่งเปอร์เซนไทล์ดังภาพที่ 1) ประมาณ 30 จุด และลากเส้นเชื่อมเกิดเป็นเส้นกราฟ



ภาพที่ 1 : กระบวนการเทียบมาตราวิธีอิควิเปอร์เซนไทล์

ขั้นตอนที่สอง เทียบมาตราคะแนนจากรายละเอียดในข้อ (3) ขั้นตอนที่หนึ่ง นำมาพล็อตกราฟใหม่ลงบนกระดาษกราฟ โดยแกนนอนเป็นคะแนนฉบับ Y แกนตั้งเป็นคะแนนฉบับ X ทำการปรับเส้นกราฟให้เรียบ เส้นกราฟนี้จะใช้อ่านค่า X ที่สมมูลกันกับ Y จากนั้นสร้างตารางสำเร็จ เพื่ออ่านค่าคะแนนแปลงจากกราฟ ดังภาพที่ 2



ภาพที่ 2 : การแปลงคะแนนของการเทียบมาตราวิธีอิควิเปอร์เซนไทล์

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

การเทียบมาตราวิธีอิควิเปอร์เซนไทล์โดยใช้แบบสอบร่วม

ลอร์ดและเลวิน ได้เสนอขั้นตอนการเทียบมาตราวิธีอิควิเปอร์เซนไทล์โดยใช้แบบสอบร่วม สำหรับกลุ่มที่ไม่ได้สุ่มซึ่งมีความสามารถไม่แตกต่างกันมาก โดยที่กลุ่ม α สอบฉบับ X และฉบับ U กลุ่ม β สอบฉบับ Y และฉบับ U ขั้นตอนในการเทียบมาตรา คือ ประมาณค่าความถี่ในแบบสอบฉบับ X และ Y สำหรับกลุ่ม t $(\alpha + \beta)$ มีขั้นตอนดังนี้

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

<u>วิธีเทียบมาตราแบบเชิงเส้นตรง</u>

วิธีเชิงเส้นตรงนั้นอยู่ในนิยามที่ว่า คะแนนจากแบบสอบ 2 ชุดโดยที่แบบสอบทั้ง 2 ชุดต่างเป็นแบบสอบที่วัดในสิ่งเดียวกันจะพิจารณาตัดสินให้เทียบกันได้ ก็ต่อเมื่อคะแนนมาตรฐานทั้งสองชุดมีค่าเท่ากัน

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

เมื่อ X คือ คะแนนจากแบบสอบฉบับที่ 1

Y คือ คะแนนจากแบบสอบฉบับที่ 2

M_x คือ ค่าเฉลี่ยของแบบสอบฉบับที่ 1

M_y คือ ค่าเฉลี่ยของแบบสอบฉบับที่ 2

S_x คือ ส่วนเบี่ยงเบนมาตรฐานของแบบสอบฉบับที่ 1

S_y คือ ส่วนเบี่ยงเบนมาตรฐานของแบบสอบฉบับที่ 2

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

ในกรณีวิธีการแจกแจงของคะแนนแบบสอบฉบับที่ 1 เหมือนกับแบบสอบฉบับที่ 2 ใช้วิธีการเทียบแบบเชิงเส้นตรงได้สะดวกและมีความเป็นปรนัย เพราะจะเป็นไปตามข้อตกลงเบื้องต้น

การเทียบมาตราวิธีเชิงเส้นตรงโดยใช้แบบสอบร่วม

ทักเกอร์ (Tucker) เป็นผู้คิดค้นวิธีเชิงเส้นตรงโดยใช้แบบสอบร่วมขึ้น วิธีนี้เริ่มจากข้อตกลงเบื้องต้นของการเทียบมาตรามีฟังชั่นเป็นเส้นตรง ดังนี้

$$e_x(Y) = \mu_x + \frac{\sigma_x}{\sigma_y}(Y - \mu_y)$$

เมื่อ μ_x, σ_x, μ_y, σ_y เป็นค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของคะแนนจากแบบสอบฉบับที่ 1 และฉบับที่ 2 ตามลำดับ และในการหาค่า μ_x, σ_x, μ_y, σ_y จะหาได้ก็ต่อเมื่อได้กำหนดเงื่อนไขเป็นข้อตกลงเบื้องต้น 2 ประการคือ

1) ค่าเฉลี่ยและความแปรปรวนอย่างมีเงื่อนไขของ x เมื่อกำหนดแบบสอบร่วม(v) เดียวกับ y

2) เป็นข้อตกลงตามรูปแบบ (Model Assumtion) เพราะเป็นการยอมรับการทำค่าคาดหมาย (E) และความแปรปรวน (Var) ให้มีรูปที่ง่ายขึ้นข้อตกลงเบื้องต้นเหล่านี้เป็นแบบฉบับที่กำหนดในการวิเคราะห์การถดถอยว่า ค่าคาดหมายอย่างมีเงื่อนไขเป็นเส้นตรงและความแปรปรวนอย่างมีเงื่อนไขเป็นค่าคงที่

ผลจากข้อตกลงเบื้องต้น สรุปหลักการเทียบมาตราได้ดังนี้ ถ้าค่าคาดหมายเฉลี่ยและค่าความแปรปรวนในแบบสอบร่วมเท่ากันแล้ว จะได้ค่า μ_x, σ_x, μ_y, σ_y เพื่อจะนำไปแทนค่าในสมการข้างต้น จะได้ฟังชั่นของการเทียบมาตราเชิงเส้นตรงของ x ไปสู่ y หรือ ทำการเทียบมาตราจาก y ไปสู่ x

<u>การออกแบบรวบรวมข้อมูล</u>

แองกอฟฟ์ (Angoff 1971 :569-586) ได้รวบรวมและเสนอโมเดลของการเทียบมาตรา ซึ่งได้บรรยายการออกแบบเพื่อรวบรวมข้อมูลและจัดกระทำในทางสถิติเพื่อระบุการแปลงคะแนนไว้ 6 รูปแบบ คือ

รูปแบบที่ 1 (Random group-one test administered to each group) รูปแบบนี้ผู้สอบมาจากการสุ่มจากประชากรกลุ่มเดียวกัน แบ่งเป็น 2 กลุ่ม คือ กลุ่มสอบฟอร์ม X กลุ่ม สอบฟอร์ม Y

1. *การเทียบมาตราเชิงเส้นตรงสำหรับแบบสอบที่มีความเที่ยงเท่ากัน*

ขั้นแรกคำนวณค่าเฉลี่ยและค่าเบี่ยงเบนมาตราฐานของคะแนนแบบสอบฟอร์ม X โดยกลุ่ม α และแบบสอบฟอร์ม Y โดยกลุ่ม β จัดให้คะแนนมาตรฐานของทั้งสองกลุ่มเท่ากัน

$$\frac{Y - M_{y\beta}}{S_{y\beta}} = \frac{X - M_{x\alpha}}{S_{x\alpha}} \quad \ldots\ldots\ldots [1]$$

โดยที่ Y คือ คะแนนสอบที่ได้จากฟอร์ม Y

$M_{y\beta}$ คือ ค่าเฉลี่ยของคะแนนฟอร์ม Y จากการสอบโดยกลุ่ม β

$S_{y\beta}$ คือ ค่าเบี่ยงเบนมาตรฐานของคะแนนฟอร์ม Y จากการสอบโดยกลุ่ม β

X คือ คะแนนสอบที่ได้จากฟอร์ม X

$M_{x\alpha}$ คือ ค่าเฉลี่ยของคะแนนฟอร์ม X จากการสอบโดยกลุ่ม α

$S_{x\alpha}$ คือ ค่าเบี่ยงเบนมาตรฐานของคะแนนฟอร์ม X จากการสอบโดยกลุ่ม α

เมื่อจัดเทอมในสมการ [1] เสียใหม่ จะได้สมการเชิงเส้นตรง คือ

$$y = \frac{S_{y\beta}}{S_{x\alpha}} X + M_{y\beta} - \frac{S_{y\beta}}{S_{x\alpha}} M_{x\alpha} \quad \ldots\ldots [2]$$

ซึ่งจะอยู่ในรูป $Y = AX + B$

เมื่อ $A = S_{y\beta} / S_{x\alpha}$

$B = M_{y\beta} - AM_{x\alpha}$

ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา

เช่นเดียวกับวิธีการทางสถิติอื่น ๆ การเทียบมาตราจะมีความคลาดเคลื่อนแบบสุ่มซึ่งเกิดจากการแกว่ง (fluctuation)ของค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของคะแนนทั้งฟอร์ม X และฟอร์ม Y ซึ่งเป็นผลมาจากการสร้างแบบสอบลอร์ด (Lord 1950) ได้อธิบายความคลาดเคลื่อนมาตรฐานในลักษณะของค่าเบี่ยงเบนมาตรฐานของคะแนนแปลงในสเกล y ที่ตรงกับคะแนน x ที่กำหนด ซึ่งคะแนนที่แปลงแล้ว y ได้มาจากกลุ่มตัวอย่าง α และ β ที่เป็นอิสระต่อกัน ความคลาดเคลื่อนมาตรฐานของการเทียบคะแนนของรูปแบบที่ 1 มีสมการ คือ

$$S.E^2.y^* = \frac{2\,S_y^2}{N_t}\,(Z_x^2 + 2) \quad \text{.....[3]}$$

โดยที่ $S.E^2\ y^*$ คือความแปรปรวนของความคลาดเคลื่อนของคะแนนที่แปลงแล้ว

N_t คือจำนวนผู้สอบทั้งสองกลุ่ม $(N_t = N_\alpha + N_\beta)$

Z_x คือคะแนนมาตรฐาน ได้จาก $Z_x = (X - M)/S_x$

การเทียบมาตราโดยใช้เปอร์เซนไทล์สำหรับแบบสอบที่มีความเที่ยงเท่ากัน

จากการกระจายของคะแนนทั้งสองฟอร์ม ฟอร์ม x สอบโดยกลุ่ม α ฟอร์ม y สอบโดยกลุ่ม β คำนวณจุดกลางตำแหน่งเปอร์เซนไทล์ สำหรับคะแนนแต่ละชุด พล็อตลงบนกระดาษกราฟ Arithmetic probability paper โดยแกนนอนเป็นคะแนนดิบแกนตั้งเป็นตำแหน่งเปอร์เซนไทล์ พล็อตตำแหน่งเปอร์เซนไทล์ทั้งสองชุดลงบนกระดาษกราฟแผ่นเดียวกันอ่านคะแนนทั้งสองฟอร์มที่มีตำแหน่งเปอร์เซนไทล์เดียวกันประมาณ 30 จุดแล้วนำมาพล็อตใหม่ลงบนกระดาษกราฟธรรมดา โดยแกนนอนเป็นคะแนนฟอร์ม x แกนตั้งเป็นคะแนนฟอร์ม Y สร้างตารางเทียบมาตราโดยอ่านจากกราฟที่พล็อตใหม่

2. การเทียบมาตราเชิงเส้นตรงสำหรับแบบสอบที่มีความเที่ยงไม่เท่ากัน

ถ้าแบบสอบฟอร์ม X และฟอร์ม Y มีความเที่ยงแตกต่างกันอย่างมีนัยสำคัญแล้วจะไม่มีหนทางที่จะพัฒนาการแปลงคะแนนให้มีลักษณะเทียบเคียงกันได้เลย อย่างไรก็ตามสถานการณ์ในทางปฏิบัติที่เรียกว่า การเทียบมาตรานั้น เราสามารถที่จะนำเอาความเที่ยงของแบบสอบทั้งสองฟอร์มรวมเข้าไปด้วย ดังสมการ

$$\frac{Y - M_{y\beta}}{S_{\tilde{y}\beta}} = \frac{X - M_{x\alpha}}{S_{\tilde{x}\alpha}} \quad \text{.....[4]}$$

เมื่อ $S_{\tilde{y}\beta} = S_y\sqrt{r_{yy}}$

$S_{\tilde{x}\alpha} = S_x\sqrt{r_{xx}}$

โดยที่ Y คือ คะแนนสอบที่ได้จากฟอร์ม Y

$M_{y\beta}$ คือ ค่าเฉลี่ยของคะแนนฟอร์ม Y จากการสอบโดยกลุ่ม β

S_y คือ ค่าเบี่ยงเบนมาตรฐานของคะแนนสอบ (observe score)ของฟอร์ม Y

$S_{\tilde{y}\beta}$ คือ ค่าเบี่ยงเบนมาตรฐานของคะแนนจริง (True score)ของฟอร์ม Y

r_{xy} คือ ความเที่ยงของแบบสอบฟอร์ม Y

X คือ คะแนนสอบที่ได้จากฟอร์ม X

$M_{x\alpha}$ คือ ค่าเฉลี่ยของคะแนนฟอร์ม X จากการสอบโดยกลุ่ม α

S_x คือ ค่าเบี่ยงเบนมาตรฐานของคะแนนสอบของฟอร์ม X

$S_{\tilde{x}_x}$ คือ ค่าเบี่ยงเบนมาตรฐานของคะแนนจริงของฟอร์ม X

r_{xx} คือ ความเที่ยงของแบบสอบฟอร์ม X

สมการสำหรับแปลงคะแนนจากสเกลฟอร์ม X ไปสู่สเกลฟอร์ม Y จะเป็น

$$y = \frac{S_{\tilde{y}_\beta}}{S_{\tilde{x}_\alpha}} X + M_y - \frac{S_{\tilde{y}_\beta}}{S_{\tilde{x}_\alpha}} M_x \quad \ldots\ldots\ldots [5]$$

ซึ่งอยู่ในรูป $Y = AX + B$

เมื่อ $A = S_{\tilde{y}_\beta} / S_{\tilde{x}_\alpha}$

$B = M_{y_\beta} - AM_{x\alpha}$

<u>รูปแบบที่ 2</u> Random groups, both forms administered to each group, counterbalanced

ลักษณะเช่นเดียวกับรูปแบบที่ 1 เลือกกลุ่มตัวอย่างขนาดใหญ่แล้วแบ่งแบบสุ่มออกเป็น 2 กลุ่ม โดยกลุ่มหนึ่งเป็น α สอบฟอร์ม X แล้วตามด้วยฟอร์ม Y อีกกลุ่มหนึ่งเป็นกลุ่ม β สอบฟอร์ม Y แล้วตามด้วยฟอร์ม X ในการที่จะป้องกันความคลาดเคลื่อนที่จะเกิดขึ้นในการสอบโดยเฉพาะอย่างยิ่ง เมื่อสอบแบบฟอร์ม X และฟอร์ม Y พร้อมกัน ควรจะเย็บแบบสอบทั้งสองฟอร์มติดกันโดยครึ่งหนึ่งให้ฟอร์ม X อยู่ด้านบนและอีกครึ่งหนึ่งให้ฟอร์ม X อยู่ด้านบนและอีกครึ่งหนึ่งให้ฟอร์ม Y อยู่ด้านบน และในการสอบควรจะแจกสลับกันเรียงตามผู้สอบ คือ XY, YX , XY , YX , ...

วิธีการเทียบมาตราเชิงเส้นตรงกรณีแบบสอบมีความเที่ยงเท่ากัน

ลอร์ด (Lord 1950)ได้เสนอวิธีการนี้โดยมีข้อตกลงว่า ผลกระทบที่เกิดจากการฝึกที่มีต่อแบบสอบฟอร์ม Y อันเป็นผลมาจากการสอบฟอร์ม X ก่อน และผลกระทบที่มีต่อฟอร์ม X อันเป็นผลมาจากการสอบฟอร์ม Y ก่อน ต่างก็เป็นสัดส่วนของความเบี่ยงเบนมาตรฐานของแบบสอบทั้งสองฉบับ

$$K_x / S_x = K_y / S_y = H$$

การประมาณค่าที่ดีที่สุดของ H คือ การหาค่าเฉลี่ยของความแตกต่างระหว่างค่าเฉลี่ยของฟอร์ม X และค่าเฉลี่ยของฟอร์ม Y เมื่อแต่ละฟอร์มอยู่ในหน่วยของค่าเบี่ยงเบนมาตรฐาน

$$H = 1/2 \left(\frac{M_{x\beta} - M_{x\alpha}}{S_x} + \frac{M_{y\alpha} - M_{y\beta}}{S_y} \right) \quad \ldots\ldots\ldots [6]$$

ซึ่งสมการนี้ก็คือ ต้นแบบของสมการสำหรับเทียบเชิงเส้นตรง สมการ [1]

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

รูปฟอร์มที่ใช้สำหรับแทนค่าในสมการ [1] คือ

$$M_x = 1/2 \ (M_{x\alpha} + M_{x\beta} - K_x) \quad \ldots\ldots\ldots [7]$$

$$M_y = 1/2 \ (M_{y\alpha} + M_{y\beta} - K_y) \quad \ldots\ldots\ldots [8]$$

$$S_x^2 = 1/2 \ (S_{x\alpha}^2 + S_{x\beta}^2) \quad \ldots\ldots\ldots [9]$$

$$S_y^2 = 1/2 \ (S_{y\alpha}^2 + S_{y\beta}^2) \quad \ldots\ldots\ldots [10]$$

เมื่อสมการ [7] ถึง [10] แทนลงในสมการ [1] จะได้สมการเชิงเส้นในรูปของ $Y = AX + B$ โดยที่

$$A = \sqrt{\frac{S_y^2 + S_y^2}{S_x^2 + S_x^2}}$$

$$B = 1/2 \ (M_{y\alpha} + M_{y\beta}) - \frac{A}{2} (M_{x\alpha} + M_{x\beta})$$

ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา

ลอร์ด (Lord 1950) ได้ตั้งข้อตกลงเบื้องต้นไว้ว่า ค่าเบี่ยงเบนมาตรฐานของแต่ละฟอร์มและความสัมพันธ์ระหว่างฟอร์มเหมือนกันในประชากรแล้ว ความแปรปรวนของความคลาดเคลื่อนของคะแนนแปลงในสเกล y คือ

$$S.E_y^2. = S_y^2 \ (1 - r_{xy}) \ Z_x^2 \ (1 + r_{xy}) + 2 \quad \ldots\ldots\ldots [11]$$

$$\overline{\qquad\qquad\qquad\qquad}$$

$$Nt$$

เมื่อ $Z_x = (X - M_x) / S_x$

การเทียบมาตราด้วยรูปแบบนี้มีความถูกต้องสูงโดยอาจจะสังเกตจากการเปรียบเทียบความแปรรวนของความคลาดเคลื่อนในสมการ [11] กับสมการ [3] ตัวอย่างเช่น แบบสอบ 2 ฟอร์ม มีค่าสหสัมพันธ์ .80 ความแปรปรวนของความคลาดเคลื่อนของคะแนนแปลง y ที่จุด $Z_x = 0$ ของรูปแบบที่ 3 มีค่า 1/10 ของขนาดความคลาดเคลื่อนของรูปแบบที่ 1 นั่นคือ อาจจะกล่าวได้ว่ารูปแบบที่ 1 จะได้รับความถูกต้องเหมือนรูปแบบที่ 2 จะต้องใช้ผู้สอบจำนวน 10 เท่าของจำนวนเดิม แสดงว่า การใช้รูปแบบที่ 2 มีข้อดีในด้านของความประหยัดใช้ผู้สอบน้อยแต่มีความถูกต้องมากกว่า

การเทียบมาตราเชิงเส้นโค้งกรณีแบบสอบมีความเที่ยงเท่ากัน

การเทียบเชิงเส้นโค้ง หรือการเทียบโดยใช้เปอร์เซนไทล์ของรูปแบบนี้ใช้วิธีการเช่นเดียวกับรูปแบบที่ 1 ซึ่งข้อมูลของฟอร์ม X และฟอร์ม Y จะรวมกันเสียก่อนสำหรับแบบสอบแต่ละฟอร์ม นั่นคือรวมกลุ่มคะแนนของผู้สอบกลุ่ม α และ β สำหรับฟอร์ม X และรวมคะแนนของผู้สอบกลุ่ม α และ β สำหรับฟอร์ม Y แล้วจึงใช้วิธีเทียบโดยใช้เปอร์เซนไทล์ดังได้อธิบายไว้ในรูปแบบที่ 1 แล้ว

2. การเทียบมาตราเชิงเส้นตรงสำหรับแบบสอบที่มีความเที่ยงไม่เท่ากัน

เมื่อฟอร์ม X และฟอร์ม Y มีความเที่ยงไม่เท่ากัน ค่าเฉลี่ยของความแปรปรวนในกลุ่มไม่ควรจะคำนวณจากคะแนนที่ได้จากการสอบ ในสมการ [9] และ [10] แต่ควรจะคำนวณจากคะแนนจริง โดยค่าที่สมนัยกับสมการ [9] และ [10] คือ สมการ[12] และสมการ [13]

$$S_x^2 = 1/2 \, (S_{x_\alpha}^2 \, r_{xx\alpha} + S_x^2 \, r_{xx\beta}) \quad \ldots\ldots [12]$$

$$S_y^2 = 1/2 \, (S_{y_\alpha}^2 \, r_{yy\alpha} + S_y^2 \, r_{yy\beta}) \quad \ldots\ldots [13]$$

<u>รูปแบบที่ 3</u> Random groups - one test administered to each group, common equating test administered to both groups

การเทียบมาตราเชิงเส้นตรงเมื่อแบบสอบมีความเที่ยงเท่ากัน

คะแนนสอบที่ได้ส่วนใหญ่ขึ้นกับลักษณะ 2 ประการ คือ ความสามารถรายบุคคลและลักษณะของแบบสอบ ในการที่จะเปรียบเทียบผลการสอบของแต่ละบุคคลซึ่งสอบแบบสอบที่แตกต่างกัน จึงจำเป็นที่จะต้องปรับคะแนนเสียก่อน กล่าวคือ การเทียบมาตรานั้นจะเป็นเพียงการเทียบความแตกต่างของคะแนนอันเป็นผลมาจากความแตกต่างของแต่ละบุคคล หรือของกลุ่ม ถ้ากลุ่ม α

และ β ไม่ได้สุ่มมาจากประชากรเดียวกัน ความแตกต่างระหว่างกลุ่มอาจจะเป็นองค์ประกอบที่มีนัยสำคัญมากในการแปรเปลี่ยนของค่า A และ B สมการ [2] , [4] และจะเป็นผลทำให้เกิดอคติ (bias) ในวิธีการเทียบมาตรา แม้ว่ากลุ่มจะถูกเลือกมาโดยวิธีการสุ่ม ความแตกต่างเล็กน้อยระหว่างกลุ่มอาจจะเกิดขึ้นได้ ถ้าไม่ป้องกันก็จะเกิดอคติในสมการแปลงคะแนน จะเป็นผลต่อการเปรียบเทียบภายหลังในการที่จะควบคุมให้บังเกิดผล การเทียบมาตราจำเป็นต้องใช้คะแนนของแบบฟอร์ม U ซึ่งเป็นข้อกระทงที่เพิ่มขึ้น หรือเป็นข้อร่วมกันระหว่างฟอร์ม X และฟอร์ม Y แบบสอบร่วมนี้ใช้สำหรับปรับความแตกต่างที่อาจจะพบระหว่างกลุ่ม α และ β ในการบริหารแบบสอบ กลุ่ม α จะได้รับแบบสอบฟอร์ม X และฟอร์ม U และกลุ่ม β จะได้รับแบบฟอร์ม Y และฟอร์ม U

ลอร์ด (Lord 1955a) ได้พัฒนาสมการสำหรับรูปแบบนี้โดยใช้การประมาณค่าความเป็นไปได้สูงสุด (maximum likelihood estimates) ประมาณค่าเฉลี่ยและความแปรปรวนของฟอร์ม X และ Y ดังนี้

$$\hat{\mu}_x = M_{x\alpha} + b_{xu\alpha}\,(\hat{\mu}_u - M_{u\alpha}) \quad \ldots\ldots\ldots [14]$$

$$\hat{\mu}_y = M_{y\beta} + b_{yu\beta}\,(\hat{\mu}_u - M_{u\beta}) \quad \ldots\ldots\ldots [15]$$

$$\hat{\sigma}^2_x = S^2_{x\alpha} + b^2_{xu\alpha}\,(\hat{\sigma}^2_u - S^2_{u\alpha}) \quad \ldots\ldots\ldots [16]$$

$$\hat{\sigma}^2_y = S^2_{y\beta} + b^2_{yu\beta}\,(\hat{\sigma}^2_u - S^2_{u\beta}) \quad \ldots\ldots\ldots [17]$$

เมื่อ $\hat{\mu}_u = M_{ut}$ และ $\hat{\sigma}_u^2 = S^2_{ut}$ และ $t = \alpha + \beta$

ซึ่งการประมาณค่าเหล่านี้ใช้กับสมการ [1] $Y = AX + B$

โดยที่ $A = \hat{\sigma}_y / \hat{\sigma}_x$ และ $B = \hat{\mu}_y - A\hat{\mu}_x$

สำหรับรูปแบบนี้ ลอร์ด (Lord 1950)ได้เสนอสูตรในการหาความแปรปรวนของความคลาดเคลื่อน คือ

$$SE^2_{y^*} = 2\,\hat{\sigma}^2_y(1 - \hat{r}^2)\,\frac{(1 + \hat{r}^2)\,Z^2_x + 2}{N_t} \quad \ldots\ldots\ldots [18]$$

โดยถือว่า $r = \dfrac{b_{xu\alpha}\,\hat{\sigma}_u}{\hat{\sigma}_x} = \dfrac{b_{yu\beta}\,\hat{\sigma}_u}{\hat{\sigma}_y}$

จากสูตรนี้จะเห็นได้ว่า เมื่อ $r = 0$ ความแปรปรวนของความคลาดเคลื่อนจะเหมือนกับการเทียบเชิงเส้นตรงในรูปแบบที่ 1 จากสมการ[14] ถึง [17] ถ้ากลุ่ม α และกลุ่ม β มีค่าเฉลี่ยในฟอร์ม U เท่ากันแล้ว ค่าในวงเล็บของสมการ [14],[15] จะพบว่าเป็นศูนย์ นั่นคือ การปรับค่าความแตกต่างระหว่างกลุ่มก็ไม่จำเป็นและการประมาณค่าเฉลี่ยประชากรของคะแนน ฟอร์ม X และฟอร์ม Y ที่ดีที่สุด คือ ค่าเฉลี่ยที่สังเกตได้จริง ๆ ของกลุ่ม α และกลุ่ม β

เป็นที่น่าสังเกตว่า สมการเหล่านี้ไม่มีข้อจำกัดเกี่ยวกับธรรมชาติของฟอร์ม U เลย การจะใช้ประโยชน์ของฟอร์ม U ในการเทียบมาตรานั้น จะขึ้นกับความสัมพันธ์ของฟอร์ม U กับฟอร์ม X และฟอร์ม Y ตัวอย่างเช่น ถ้า $r_{xu_\alpha} = 0$ และ $r_{yu_\beta} = 0$ (เนื่องจาก X และ Y คู่ขนานกัน)จะเห็นได้ว่า คะแนนหรือค่าที่สังเกตได้ของฟอร์ม U ไม่เกี่ยวข้องกับการวัดโดยฟอร์ม X และฟอร์ม Y เลย ดังนั้นจึงไม่มีประโยชน์ในการปรับคะแนนโดยใช้คะแนนจากฟอร์ม U เลย

การเทียบมาตราด้วยวิธีนี้มีประโยชน์มากสามารถยืดหยุ่นและปรับให้เข้ากับสถานการณ์อื่น ๆ ได้ เช่น ฟอร์ม U อาจจะสอบรวม หรือแยกจากฟอร์ม X และฟอร์ม Y ฟอร์ม U อาจจะเป็นส่วนหนึ่งร่วมกับฟอร์ม X แต่แยกจากฟอร์ม Y และฟอร์ม U อาจจะรวมเข้าไปภายในฟอร์ม X และฟอร์ม Y โดยกระจายทั่วไปตลอดทั้งฉบับ อย่างไรก็ตามแบบสอบร่วมควรจะมีความยาวและความเที่ยงเพียงพอที่จะให้ข้อมูลไปปรับความแตกต่างระหว่างกลุ่มตามที่คาดหวังซึ่งแองกอฟฟ์ (Angoff 1971: 578)ได้เสนอกฎเกณฑ์เกี่ยวกับจำนวนข้อของแบบสอบร่วม คือไม่ควรน้อยกว่า 20 ข้อ หรือ 20% ของจำนวนข้อในแต่ละฟอร์มแล้วแต่จำนวนใดจะมากกว่าให้ใช้จำนวนนั้น และเพื่อที่จะหลีกเลี่ยงเกี่ยวกับแบบสอบที่ใช้วัดความเร็ว ควรกระจายฟอร์ม U ปะปนไปในทั้งฟอร์ม X และ Y หลักเกณฑ์ทั่ว ๆไปเกี่ยวกับการใช้ฟอร์ม U กับกลุ่ม α และกลุ่ม β คือจะต้องเป็นการสอบเหมือนกันทั้งสองกลุ่ม ในกรณีของการแยกสอบคนละเวลา ตัวอย่างเช่น จัดสอบภายหลังฟอร์ม X และฟอร์ม Y หรือ จัดสอบก่อนฟอร์ม X และฟอร์ม Y เพราะจะได้มีผลเท่ากันในทางปฏิบัติและข้อกระทงในฟอร์ม U ไม่ควรซ้ำกันกับข้อกระทงของฟอร์ม X และฟอร์ม Y

การเทียบมาตราด้วยรูปแบบที่ 3 นี้ยังมีวิธีการที่ยืดหยุ่นได้ คือ สามารถเทียบแบบสอบที่มากกว่า 2 ฟอร์มขึ้นไป อาจจะเป็น 3 ฟอร์ม 4 ฟอร์ม หรือมากกว่านั้น เช่น กรณี 3 ฟอร์ม คือ ฟอร์ม X, Y และ Z แบบสอบทั้งสามฟอร์มนี้สอบโดยกลุ่ม α , β และ γ ตามลำดับ ค่าเฉลี่ยและความแปรปรวนสำหรับกลุ่มรวม ($\alpha + \beta + \gamma$)จะประมาณได้โดยใช้ข้อตกลงเกี่ยวกับการเทียบมาตราของแบบสอบสองฟอร์ม

แบบฟอร์มร่วม ฟอร์ม U ไม่จำเป็นที่จะต้องเป็นตัวแปรเดียวเสมอไป อาจจะเป็น 2,3,4,... หรือเท่ากับคะแนนที่ใช้ปรับระหว่างกลุ่มตัวอย่าง เช่น ฟอร์ม X และฟอร์ม Y เป็นแบบสอบที่ประกอบด้วย ข้อกระทงที่เป็นภาษาคณิตศาสตร์ด้วยคะแนนของกระทงทั้งสองชนิดอาจจะเป็นคะแนนรวมกัน คือ ฟอร์ม U จะประยุกต์ใช้วิธีการเดียวกับที่กล่าวมาแล้ว และอีกกรณีหนึ่งคือแยกเป็นคะแนนฟอร์ม V และ ฟอร์ม M แล้วใช้การรวมแบบ (multiple combination) สมการที่ใช้สำหรับประมาณค่าเฉลี่ย และความแปรปรวนของกลุ่ม t คือ

$$\mu_x = M_x + b_{xv.m}(\mu_v - M_v) + b_{xm.v}(\mu_m - M_m) \quad \ldots[19]$$

$$\mu_y = M_y + b_{yv.m}(\mu_v - M_v) + b_{ym.v}(\mu_m - M_m) \quad \ldots[20]$$

$$\hat{\sigma}^2_x = S^2_x + b^2_{xv.m}(\sigma^2_v - S^2_v) + b^2_{xm.v}(\sigma^2_m - S^2_m) + 2b_{xv.m}\, b_{xm.v}(\sigma^2_{vm} - S^2_{vm}) \quad \ldots[21]$$

$$\hat{\sigma}^2_y = S^2_y + b^2_{yv.m}(\sigma^2_v - S^2_v) + b^2_{ym.v}(\sigma^2_m - S^2_m) + 2b_{yv.m}\, b_{ym.v}(\sigma^2_{vm} - S^2_{vm}) \quad \ldots[22]$$

เมื่อ $b_{xv.m}$ คือ สัมประสิทธิ์กระถดถอย สำหรับการทำนาย X จาก V โดยที่ M เป็นค่าคงที่ และ $S_{vmx} = r_{vmx} \cdot S_{vx} \cdot S_{mx}$,

$\mu_v = M_{vt}$, $\mu_m = M_{mt}$, $\sigma^2_v = S^2_{vt}$, $\sigma^2_m = S_{mt}$, $\sigma^2_{vm} = S^2_{vmt}$

ซึ่งอาจกล่าวได้ว่า การประมาณค่าพารามิเตอร์ของประชากรสำหรับตัวแปรร่วม V และ M ได้มาโดยตรงจากค่าสถิติที่สังเกตได้สำหรับกลุ่มรวม t ($t = \alpha + \beta$)

อีกกรณีหนึ่งซึ่งยืดหยุ่นได้ คือ แบบสอบร่วมฟอร์ม U ที่สอบโดยกลุ่ม α ไม่จำเป็นต้องเหมือนกันกับฟอร์ม U ที่สอบโดยกลุ่ม β อาจจะเป็นแบบสอบร่วมครึ่งหนึ่ง(quasi-common test) หรืออาจจะเป็นแบบสอบสองฟอร์มที่แตกต่างกัน(ฟอร์ม U และ W)แต่ทั้งสองฟอร์มต้องวัดในเรื่องเดียวกัน กระบวนการในการเทียบมาตรา คือ แปลงแบบสอบร่วมฟอร์ม W ให้อยู่ในสเกลเดียวกับฟอร์ม U ก่อน หรือทั้งสองฟอร์มอาจจะแปลงไปเป็นสเกลใหม่เลยก็ได้แล้วจึงใช้วิธีการที่กล่าวมาแล้วตอนต้นปรับความแตกต่างระหว่างกลุ่ม

การเทียบมาตราเชิงเส้นตรงเมื่อแบบสอบมีความเที่ยงไม่เท่ากัน

เลวิน (Levine 1955) ได้เสนอไว้ในกรณีของกลุ่มสุ่ม แบบสอบฟอร์ม X และฟอร์ม Y มีความเที่ยงไม่เท่ากัน ซึ่งเหมาะสมกับการแปลง โดยคะแนนจริงมากกว่าคะแนนที่สอบได้ ข้อตกลงที่เพิ่มเติมขึ้นอีก คือ ฟอร์ม U จะต้องมีลักษณะคู่ขนานกันทางโครงสร้างกับทั้งฟอร์ม X

และฟอร์ม Y ภายใต้เงื่อนไขนี้ เมื่อฟอร์ม U แยกออกจากฟอร์ม X และ Y ความชัน (slope) และจุดตัดแกน (intercept) ของสมการ $y = AX + B$ คือ

$$A = b_{yu\beta} / b_{xu\alpha}$$

$$B = \hat{\mu}_y - A\hat{\mu}_x$$

เมื่อ $b_{xu\alpha}$ และ $b_{yu\beta}$ คือ สัมประสิทธิ์กระถดถอย (regression coefficients) ที่พบในกลุ่ม α และ β สำหรับการทำนาย X จาก U และ Y จาก U และ $\hat{\mu}_x$, $\hat{\mu}_y$ คำนวณจากสมการที่ [14],[15]

เลวินเพิ่มข้อตกลงเบื้องต้นอีกว่า ฟอร์ม U ต้องเป็นฉบับคู่ขนานกับแบบสอบ X และ Y เมื่อฟอร์ม U รวมอยู่ในส่วนของ X และ Y แล้ว

$$A = (b_{xu\alpha}\ \hat{\sigma}^2_y) / (b_{yu\beta}\ \hat{\sigma}^2_x)$$

และ $$B = \hat{\mu}_y - A\hat{\mu}_x$$

ส่วนค่าของ $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\sigma}^2_x$ และ $\hat{\sigma}^2_y$ คำนวณได้จากสมการที่ [14]-[17]

<u>รูปแบบที่ 4</u> Nonrandom group - one test to each group, common equating test administered to both groups

สถานการณ์ในการสอบบางครั้งถือเป็นความลับที่ไม่เปิดเผย ไม่สามารถที่จะแนะนำแบบสอบฟอร์มใหม่ก่อนที่จะสอบได้ และคำสั่งของโปรแกรมการสอบไม่อนุญาตให้อธิบายนอกเหนือไปกว่าที่สอบ ภายใต้สถานการณ์เหล่านี้ การเทียบมาตราโดยการสุ่มจาก 3 วิธีการไม่สามารถนำมาใช้ได้และข้อมูลที่ใช้สำหรับเทียบมาตราต้องมาจากผลการสอบของกลุ่มที่สอบจริงๆแต่ละฟอร์ม ซึ่งการสอบต่างฟอร์มกันและเวลาสอบต่างกันผู้สอบ อาจจะไม่ได้มาจากกลุ่มประชากรเดียวกัน ตัวอย่างเช่น มีการจัดสอบฟอร์มใหม่ในเดือนกันยายน 2527 และต้องการเทียบกับคะแนนของแบบสอบฟอร์มเก่าที่จัดสอบในเดือนกันยายน 2526 ซึ่งผู้สอบอาจจะมีลักษณะเหมือนกันหลาย ๆ ประการ แต่ยังไม่แน่ใจว่า กลุ่มผู้สอบทั้งสองฟอร์มมาจากประชากรเดียวกัน ดังนั้นเมื่อมีความระมัดระวังเกิดขึ้นดังตัวอย่าง การเลือกกลุ่ม α และ β เพื่อที่จะลดความแตกต่างให้น้อยลงใช้วิธีการ คือ

วิธีการเทียบมาตราเชิงเส้นตรง (สำหรับกลุ่มที่มีความสามารถแตกต่างกันน้อย)

วิธีการที่จะเสนอต่อไปนี้เป็นการรวมวิธีการที่เสนอโดยลอร์ด (Lord 1955a)ที่ใช้วิธีหาความเป็นไปได้สูงสุด (Maximum likelihood method) กับวิธีการที่เสนอโดยเลวิน

(Levine 1955) ที่ใช้การปรับด้วยคะแนนจริง (true-score adaptation)โดยมีรูปแบบดังต่อไปนี้

กลุ่ม α สอบฟอร์ม X กลุ่ม β สอบฟอร์ม Y และทั้งสองกลุ่มสอบฟอร์ม U ซึ่งเป็นแบบสอบร่วม และใช้ปรับความแตกต่างที่อาจจะเกิดขึ้นระหว่างกลุ่ม การประมาณค่าเฉลี่ยและความแปรปรวนของทั้งฟอร์ม X และฟอร์ม Y กระทำโดยรวมกลุ่ม α และ β เป็นกลุ่ม t และประยุกต์ใช้กับสมการที่ [1] ซึ่งเป็นสมการเชิงเส้นที่เป็นความสัมพันธ์ของคะแนนดิบฟอร์ม X และฟอร์ม Y (Gulliksen 1950 chap 19 also Angoff 1961a)สมการที่จะเสนอต่อไปจะขึ้นกับข้อตกลงเบื้องต้น 3 ประการของทฤษฎี univariate selection theory นั่นคือ

1. จุดตัดแกนของคะแนนฟอร์ม X จากฟอร์ม U โดยกลุ่ม t เหมือนกับกลุ่ม

$$M_{xt} - b_{xut} M_{ut} = M_x - b_{xu\alpha} M_u \quad \text{.....[23]}$$

2. สัมประสิทธิ์กระถดถอย (regression coefficient)ของฟอร์ม X จากฟอร์ม U สำหรับกลุ่ม t เหมือนกับกลุ่ม α

$$b_{xut} = b_{xu\alpha} \quad \text{.....[24]}$$

3. ความแปรปรวนของความคลาดเคลื่อนในการประมาณค่าของ X จาก U ของกลุ่ม t เหมือนกับกลุ่ม α

$$S^2_{xt} (1 - r^2_{xut}) = S^2_{x\alpha} (1 - r^2_{xu\alpha}) \quad \text{.....[25]}$$

แทนค่าสมการ [24] ลงใน [23] เพื่อหาค่า M_{xt}

$$\hat{M}_{xt} = M_{x\alpha} + b_{xu\alpha} (M_{ut} - M_{u\alpha}) \quad \text{.....[26]}$$

(สัญญลักษณ์ (^) ใช้แสดงค่าประมาณ)

แทนค่าลงในสมการ [25] โดยที่ $b_{xut}S_{ut}$ แทน $r_{xut}S_{xt}$ และ $b_{xu\alpha}S_{u\alpha}$ แทนค่า $r_{xu\alpha}S_{x\alpha}$ เพื่อหาค่า S_{xt}

$$\hat{S}^2_{xt} = S^2_{x\alpha} + b^2_{xu\alpha} (S^2_{ut} - S^2_{u\alpha}) \quad \text{.....[27]}$$

ด้วยข้อตกลงเบื้องต้นลักษณะเดียวกันของฟอร์ม X และฟอร์ม Y นำมาหาความสัมพันธ์ระหว่าง ฟอร์ม Y และฟอร์ม U กลุ่ม β และกลุ่ม t ได้ผลลักษณะเช่นเดียวกับสมการ [26], [27] คือ

$$\hat{M}_{yt} = M_{y\beta} + b_{yu\beta} (M_{ut} - M_{u\beta}) \quad \text{.....[28]}$$

และ

$$\hat{S}^2_{yt} = S^2_{y\beta} + b^2_{yu\beta} (S^2_{ut} - S^2_{u\beta}) \quad \text{.....[29]}$$

(สัญญลักษณ์ $\hat{M}_{xt}$, $\hat{S}^2_{xt}$, $\hat{M}_{yt}$ และ $\hat{S}^2_{yt}$ ที่ใช้ในสมการ [26] ถึง [29] นำมาใช้แทนสัญญลักษณ์ μ_x, σ^2_x, μ_y และ σ^2_y ตามลำดับ เพราะว่าในที่นี้เป็นการประมาณค่ากลุ่มรวม ไม่ใช่ประมาณค่าสำหรับประชากร)

สมการ [26] ถึง [29] แทนค่าในสมการ [1] คือ

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

ซึ่งจะได้สมการสำหรับแปลงคะแนน $Y = AX + B$ โดยที่

$$A = \hat{S}_{yt} / \hat{S}_{xt}$$

$$B = \hat{M}_{yt} - A\hat{M}_{xt}$$

เป็นที่สังเกตได้ว่า วิธีการคำนวณสำหรับการประมาณค่าในสมการ [26]-[29] เหมือนกับสมการ [14]-[17] ตามลำดับ แม้ว่าสมการทั้ง 2 ชุดจะแตกต่างกัน

ลักษณะเช่นเดียวกันนี้ สามารถที่จะยืดหยุ่นใช้กับสถานการณ์อื่นได้อีก เช่น ฟอร์ม U อาจจะสอบรวมหรือแยกจากฟอร์ม X และฟอร์ม Y หรือเป็นส่วนหนึ่งของฟอร์ม X และเป็นส่วนหนึ่งของฟอร์ม Y หรืออาจแยกจากฟอร์ม X แต่เป็นส่วนหนึ่งของฟอร์ม Y หรืออาจจะกระจายข้อกระทงของฟอร์ม U ปะปนกับฟอร์ม X และฟอร์ม Y ก็ได้ แต่ต้องสามารถหาคะแนนรวมของแต่ละฟอร์มได้ถูกต้อง

วิธีการประมาณค่าความเป็นไปได้สูงสุด (Maximum likelihood) นี้ไม่จำกัดเฉพาะการเทียบแบบสอบ 2 ฟอร์มเท่านั้น แต่สามารถเทียบแบบสอบตั้งแต่ 3 ฟอร์มขึ้นไปได้ ซึ่งในการบริหารแบบสอบแต่ละฟอร์มอาจจะสอบแยกกลุ่มกันแต่ฟอร์ม U ซึ่งเป็นแบบสอบร่วมต้องสอบทุกกลุ่มแล้วการประมาณค่าเฉลี่ย และความแปรปรวนของแต่ละฟอร์มที่จะเทียบกันก็ใช้กลุ่มรวมสอบฟอร์ม U ซึ่งสมการ [26]-[29] สามารถใช้หลาย ๆ ครั้งเท่ากับจำนวนแบบสอบที่จะมาเทียบกัน และอีกประการหนึ่ง ฟอร์ม U ก็ไม่จำกัดเฉพาะการวัดตัวแปรเดียว แต่อาจจะใช้ได้หลาย ๆ ตัวแปร ซึ่งมีลักษณะเดียวกับสมการ [19] ถึง [22] แบบสอบร่วมอาจจะเป็นลักษณะครึ่งหนึ่ง (quasi-common) ซึ่งอาจจะกล่าวได้ว่าเป็นแบบสอบที่แตกต่างกัน 2 ฟอร์ม ฟอร์มหนึ่งสอบกับกลุ่ม α และอีกฟอร์มหนึ่งสอบกับกลุ่ม β มีข้อจำกัดอยู่ว่าทั้งสองฟอร์มจะต้องวัดเรื่องเดียวกัน คะแนนอยู่บนสเกลเดียวกัน

วิธีการเทียบมาตราเชิงเส้นตรง (สำหรับกลุ่มตัวอย่างที่มีความสามารถแตกต่างกัน)

1) แบบสอบมีความเที่ยงเท่ากัน

เลวิน (Levine) ได้เสนอว่า เมื่อกลุ่ม α และ β มีความสามารถแตกต่างกันมาก ข้อตกลงเบื้องต้นที่เป็นพื้นฐานของทฤษฎีการเลือกนั้น ไม่มีความเหมาะสมกับสถานการณ์นี้ควรใช้กับข้อตกลงเบื้องต้นแบบอื่นต่อภายใต้ข้อจำกัดที่ว่า ฟอร์ม U มีลักษณะคู่ขนานกับทั้งฟอร์ม X และฟอร์ม Y นั่นคือ

1. จุดตัดแกนของเส้นกระถดถอย (regression line) ที่เกี่ยวกับคะแนนจริงของฟอร์ม X และฟอร์ม U โดยกลุ่ม t เหมือนกับกลุ่ม

$$M_{xt} - \frac{S_{\tilde{x}t}}{S_{ut}} M_{ut} = M_x - \frac{S_{\tilde{x}\alpha}}{S_u} M_{u\alpha} \quad \text{.....[30]}$$

โดยที่ $S_x = S_x \sqrt{r_{xx}}$ และ $S_u = S_u \sqrt{r_{uu}}$

2. ความชันของเส้นสัมพันธ์สำหรับกลุ่ม t เหมือนกับกลุ่ม

$$\frac{S_{\tilde{x}t}}{S_{\tilde{u}t}} = \frac{S_{\tilde{x}\alpha}}{S_{\tilde{u}\alpha}} \quad \text{.....[31]}$$

3. ความแปรปรวนของความคลาดเคลื่อนของการวัดจากฟอร์ม X กลุ่ม t เหมือนกับกลุ่ม ๆ

$$S^2_{xt}(1-r_{xxt}) = S^2_{x\alpha}(1-r_{xx\alpha}) \quad \text{.....[32]}$$

จากสมการ [30],[31],[32] จะได้

$$\hat{M}_{xt} = M_{x\alpha} + \frac{S_{\tilde{x}\alpha}}{S_u}(M_{ut} - M_{u\alpha}) \quad \text{.....[33]}$$

$$\hat{S}^2_{xt} = S^2_{x\alpha} + \frac{S^2_{\tilde{x}\alpha}}{S^2_{\tilde{u}\alpha}}(S^2_{ut} - S^2_{u\alpha}) \quad \text{.....[34]}$$

ด้วยข้อตกลงเบื้องต้นแบบเดียวกันสำหรับฟอร์ม X และฟอร์ม Y จะได้

$$\hat{M}_{yt} = M_{y\beta} + \frac{S_{\tilde{y}\beta}}{S_u}(M_{ut} - M_{u\beta}) \quad \text{.....[35]}$$

$$\hat{S}^2_{yt} = S^2_{y\beta} + \frac{S^2_{\tilde{y}\beta}}{S^2_{\tilde{u}\beta}}(S^2_{ut} - S^2_{u\beta}) \quad \text{.....[36]}$$

สมการที่แสดงความสัมพันธ์ระหว่างคะแนนของฟอร์ม X และฟอร์ม Y คือ

$$Y = AX + B$$

โดยที่ $A = \hat{S}_{yt}/\hat{S}_{xt}$, $B = \hat{M}_{yt} - A\hat{M}_{xt}$

เพื่อเป็นการง่ายในการคำนวณ แองกอฟฟ์ (Angoff)ได้เสนอข้อมูลที่ได้ทำจากการทดลองเทียบมาตราโดยใช้การประมาณค่าอัตราส่วนของความเบี่ยงเบนมาตรฐานของคะแนนจริง นั่นคือ $S_{\tilde{x}}/S_{\tilde{u}} = n$ ซึ่งเป็นอัตราส่วนของความยาวของแบบสอบฟอร์ม X กับฟอร์ม U เมื่อฟอร์ม U เป็นการสอบรวมในฟอร์ม X และฟอร์ม U คู่ขนานกับฟอร์ม X

$$n = S_x / r_{xu}S_u = 1 / b_{ux}$$

เมื่อฟอร์ม U เป็นการสอบแยกจากฟอร์ม X แล้ว

$$n = (S_x^2 + S_{ux}) / (S_u^2 + S_{ux})$$

ซึ่งเมื่อใช้กับฟอร์ม Y และฟอร์ม U ก็ปรับสมการในลักษณะเช่นเดียวกัน

2) แบบสอบมีความเที่ยงไม่เท่ากัน

เมื่อฟอร์ม U สอบแยกจากฟอร์ม X และฟอร์ม Y จะได้

$$Y = AX + B$$

เมื่อ $A = (b_{yu\beta}\, r_{uu\alpha}) / (b_{xu\alpha}\, r_{uu\beta})$

$$B = M_{y\beta} - AM_{x\alpha} + \frac{b_{yu\beta}}{r_{uu\beta}}(M_{u\alpha} - M_{u\beta})$$

เมื่อฟอร์ม U สอบรวมกับฟอร์ม X และฟอร์ม Y จะได้

$$A = b_{ux\alpha} / b_{uy\beta}$$

$$B = M_{y\beta} - AM_{x\alpha} + [(M_{u\alpha} - M_{u\beta}) / b_{uy\beta}]$$

วิธีการเทียบมาตราแบบอิควิเปอร์เซนไทล์ (สำหรับกลุ่มที่มีความสามารถไม่แตกต่างกันมาก)

ลอร์ด (Lord 1957)และเลวิน (Levine 1958) ได้เสนอวิธีการเทียบมาตราโดยใช้ตำแหน่งเปอร์เซนไทล์ ซึ่งมาจากพื้นฐานข้อตกลงเบื้องต้นในสมการที่ [23],[24]และ[25] โดยที่กลุ่ม α สอบ ฟอร์ม X และฟอร์ม U กลุ่ม β สอบฟอร์ม Y และฟอร์ม U ขั้นตอนของกระบวนการเทียบมาตราคือ ประมาณค่าความถี่ในแบบสอบฟอร์ม X และ Y สำหรับกลุ่ม t $(\alpha + \beta)$ โดยมีวิธีการดังนี้

1. รวมคะแนนของฟอร์ม U ที่สอบโดยกลุ่ม t ($t = \alpha + \beta$)
2. หาสัดส่วนความถี่ $f_{it}/f_{i\alpha}$ และ $f_{it}/f_{i\beta}$ ที่ทุกช่วงคะแนน i
3. กระจายความถี่ของฟอร์ม U ที่ทุกช่วงคะแนนของฟอร์ม X และ Y
4. คูณความถี่ของฟอร์ม U ในแต่ละช่วงคะแนนของฟอร์ม X ด้วยสัดส่วน $f_{it}/f_{i\alpha}$
5. คูณความถี่ของฟอร์ม U ในแต่ละช่วงของฟอร์ม Y ด้วยสัดส่วน $f_{it}/f_{i\beta}$
6. หาตำแหน่งเปอร์เซนไทล์ของความถี่ที่ประมาณค่าใหม่ของทั้งสองชุด
7. ดำเนินการเทียบโดยเทียบฟอร์ม X ฟอร์ม Y ที่ตำแหน่งเปอร์เซนไทล์เดียวกัน

รูปแบบที่ 5 Other methods involving score data

1. แบบสอบฟอร์ม X และฟอร์ม Y เทียบไปสู่แบบสอบร่วม

วิธีการเทียบมาตราแบบเชิงเส้นตรง

วิธีนี้เกี่ยวข้องกับการบริหารแบบสอบร่วมฟอร์ม U ซึ่งอาจจะสอบฟอร์ม U ก่อนแล้วตามด้วยฟอร์ม X และฟอร์ม Y หรืออาจจะสอบฟอร์ม X และฟอร์ม Y ก่อน แล้วตามด้วยฟอร์ม U ถ้าหากฟอร์ม X และฟอร์ม Y เป็นแบบสอบที่มีลักษณะคู่ขนานกันก็มีเหตุผลที่จะถือได้ว่า แต่ละฟอร์มมีผลต่อฟอร์ม U เหมือนกัน เมื่อฟอร์ม U สอบภายหลังหรือฟอร์ม U มีผลต่อฟอร์ม X และฟอร์ม Y เหมือนกันเมื่อ ฟอร์ม U สอบก่อน ฟอร์ม X ฟอร์ม Y การเทียบมาตราโดยวิธีการนี้คือ เทียบจากฟอร์ม X ไปสู่ฟอร์ม U และเทียบจากฟอร์ม Y ไปสู่ฟอร์ม U ซึ่งคะแนนที่เทียบแล้วของทั้งสองฟอร์มที่ระดับคะแนนของฟอร์ม U เดียวกัน ถือว่าเท่าเทียมกัน ดังนั้น

ถ้า $X = A_{xu}U + B_{xu}$ เมื่อ $A_{xu} = S_{x\alpha} / S_{u\alpha}$

$B_{xu} = M_{x\alpha} - A_{xu}M_{u\alpha}$

และ ถ้า $Y = A_{yu}U + B_{yu}$ เมื่อ $A_{yu} = S_{y\beta} / S_{u\beta}$

$B_{yu} = M_{y\beta} - A_{yu} M_{u\beta}$

แล้วจะได้ $Y = A_{yx}X + B_{yx}$ เมื่อ $A_{yx} = A_{yu}/A_{xu}$

$B_{yx} = B_{yu} - A_{yu}B_{xu}$

ในการที่ดูถึงความเหมาะสมของสมการแปลง $Y = A_{yx}X + B_{yx}$ ฟอร์ม U ต้องมีลักษณะคู่ขนานกับฟอร์ม X และ Y โดยที่มีอิสระในการเลือกกลุ่ม α และ β แต่ถ้าฟอร์ม U ไม่มีลักษณะคู่ขนานกับฟอร์ม X และ Y แล้ว กลุ่ม α และ β ต้องมาจากการสุ่มจากประชากรเดียวกัน

ภายใต้เงื่อนไขของการสุ่มตัวอย่าง ลอร์ด (Lord 1950) ได้เสนอความแปรปรวนของความคลาดเคลื่อนของการเทียบมาตราของวิธีนี้ คือ

$$S.E.^2_{y^*} = 4S^2_{y\beta}(1-r)\ \frac{Z^2_x(1+r)+2}{Nt} \quad \ldots\ldots\ldots[37]$$

เมื่อมีข้อตกลงว่า $r = r_{xu\alpha} = r_{yu\beta}$

การเทียบมาตราเชิงเส้นโค้ง

การเทียบเชิงเส้นโค้งก็มีลักษณะเดียวกันกับรูปแบบอื่น ๆ ที่อธิบายไว้แล้ว แต่มีข้อแตกต่างกัน คือ เทียบระหว่างแบบสอบร่วมฟอร์ม U กับฟอร์ม X โดยใช้ตำแหน่งเปอร์เซนไทล์เดียวกันและใช้วิธีเดียวกันเทียบมาตราระหว่างฟอร์ม U กับฟอร์ม Y แล้วจึงนำคะแนนฟอร์ม X และฟอร์ม Y ที่มีตำแหน่งเปอร์เซนไทล์บนสเกลของฟอร์ม U มาพล็อตกราฟและสร้างตารางแปลงคะแนน

2. แบบสอบร่วมฟอร์ม U เป็นตัวทำนาย (predict)ฟอร์ม X และฟอร์ม Y

การเทียบเชิงเส้นตรง

วิธีการบริหารการสอบ คือ กลุ่ม α สอบฟอร์ม X กลุ่ม β สอบฟอร์ม Y และทั้งสองกลุ่มสอบฟอร์ม U ซึ่งวิธีการในการกำหนดคะแนนที่เทียบเคียงกันทั้ง 2 ฟอร์ม โดยการทำนายของคะแนนฟอร์ม U เดียวกัน ดังนั้น

ถ้า $\hat{X} = b_{xu}U + D_{xu}$ เมื่อ $b_{xu} = r_{xu\alpha}(S_{x\alpha}/S_{u\alpha})$

$$D_{xu} = M_{x\alpha} - b_{xu}M_{u\alpha}$$

และถ้า $\hat{Y} = b_{yu}U + D_{yu}$ เมื่อ $b_{yu} = r_{yu\beta}(S_{y\beta}/S_{u\beta})$

$$D_{yu} = M_{y\beta} - b_{yu}M_{u\beta}$$

แล้วจะได้ $Y = AX + B$ เมื่อ $A = b_{yu}/b_{xu}$ .....[38]

$$B = D_{yu}/A\,D_{xu} \quad [39]$$

การเทียบมาตราเชิงเส้นโค้ง

ในการเทียบเชิงเส้นโค้งโดยใช้ตำแหน่งเปอร์เซนไทล์ ใช้วิธีการดังนี้ คือ

ขั้นที่หนึ่ง ที่ทุก ๆช่วงคะแนนของฟอร์ม U หาความถี่ของฟอร์ม X ที่กระจายทุกช่วงคะแนน แล้วหาค่าเฉลี่ยของคะแนน X

ขั้นที่สอง ที่ทุก ๆช่วงคะแนนของฟอร์ม U หาความถี่ของฟอร์ม Y ที่กระจายทุกช่วงคะแนน แล้วหาค่าเฉลี่ยของคะแนน Y

ขั้นที่สาม นำค่าเฉลี่ยของทั้งสองสเกลมาพล็อตกราฟ ซึ่งค่าเฉลี่ย M_{xu} กับ M_{yu} จะเป็นตัวกำหนดความสัมพันธ์ระหว่างคะแนนฟอร์ม X กับคะแนน ฟอร์ม Y

3. แบบสอบฟอร์ม X และฟอร์ม Y เป็นตัวทำนายแบบสอบร่วม

การเทียบมาตราเชิงเส้นตรง

กลุ่ม α สอบฟอร์ม X กลุ่ม β สอบฟอร์ม Y แล้วทั้งสองกลุ่มสอบฟอร์ม U

ถ้า $U = b_{ux}X + D_{ux}$ เมื่อ $b_{ux} = r_{ux\alpha}(S_{u\alpha}/S_{x\alpha})$

$D_{ux} = M_{u\alpha} - b_{ux} M_{x\alpha}$

และถ้า $U = b_{uy}Y + D_{uy}$ เมื่อ $b_{uy} = r_{uy\beta}(S_{u\beta}/S_{y\beta})$

$D_{uy} = M_{u\beta} - b_{uy\beta} M_{y\beta}$

แล้วจะได้ $Y = AX + B$ เมื่อ $A = b_{ux}/b_{uy}$ .....[40]

$B = (D_{ux} - D_{uy})/b_{uy}$ .....[41]

การเทียบมาตราเชิงเส้นโค้ง

ใช้ตำแหน่งเปอร์เซนไทล์ที่เท่ากันเทียบโดยมีขั้นตอนดังนี้

1) ที่ทุก ๆ ช่วงคะแนนของฟอร์ม X หาความถี่และค่าเฉลี่ยของฟอร์ม U เมื่อกำหนดให้ X คงที่
2) ที่ทุก ๆ ช่วงคะแนนของฟอร์ม Y หาความถี่และค่าเฉลี่ยของฟอร์ม U เมื่อกำหนดให้ Y คงที่
3) หาความถี่สะสมและตำแหน่งเปอร์เซนไทล์ของทั้งสองสเกลแล้วนำมาพล็อตกราฟ
4) ค่าของ X และ Y อ่านจากกราฟที่มีค่า U เท่ากัน เป็นตัวกำหนดความสัมพันธ์

<u>รูปแบบที่ 6</u> การเทียบมาตราที่ขึ้นกับลักษณะของข้อกระทง

วิธีการของเธอร์สโตน (Thurstone's absolute scaling method)

เธอร์สโตน (Thurstone 1925) ได้เสนอวิธีการโดยมีสถานการณ์คือ กลุ่ม α สอบฟอร์ม X กลุ่ม β สอบฟอร์ม Y แบบสอบฟอร์ม X และแบบสอบฟอร์ม Y มีข้อกระทงชุดหนึ่งซึ่งมีลักษณะร่วมกันคือ ค่าความยาก (P) ซึ่งต่างจากค่า P ทั่ว ๆไป คือ แปลงให้มีลักษณะเช่นเดียวกับการกระจายแบบปกติ ข้อตกลงเบื้องต้นเช่นเดียวกับวิธีอื่น ๆ คือ ฟอร์ม X และฟอร์ม Y มีลักษณะคู่ขนานกัน ดังนั้นจึงสามารถแปลงเป็นสเกลร่วมได้ และข้อตกลงอีกอย่างหนึ่งคือ กลุ่ม α และ β มีลักษณะการกระจายเป็นแบบปกติ

จุดประสงค์ของวิธีนี้ คือ ความสัมพันธ์ระหว่างความยากของข้อกระทงที่ร่วมกันของทั้งสองกลุ่มจะนำไปสู่สมการแปลงคะแนนดิบของฟอร์ม X ไปสู่ฟอร์ม Y นั่นคือ

$$Y = AX + B$$

เมื่อ $A = S_{y\alpha} / S_{x\alpha}$ $(= S_{y\beta} / S_{x\beta})$

$B = M_{y\alpha} - AM_{x\alpha}$ $(= M_{y\beta} - AM_{x\beta})$

ถ้าการกระจายของคะแนนความสามารถเป็นโค้งปกติ ภายในกลุ่ม α และ β แล้วความสัมพันธ์ระหว่างข้อกระทงและความสามารถ คือ

$$Z_{i\alpha} = \frac{y_i - M_{y\alpha}}{S_{y\alpha}} \quad \text{.....[42]}$$

และ

$$Z_{i\beta} = \frac{y_i - M_{y\beta}}{S_{y\beta}} \quad \text{.....[43]}$$

เมื่อ $M_{y\alpha}$, $M_{y\beta}$ คือ ค่าเฉลี่ยของคะแนนความสามารถของกลุ่ม α และกลุ่ม β ตามลำดับ

$S_{y\alpha}$, $S_{y\beta}$ คือ ค่าเบี่ยงเบนมาตรฐานของคะแนนกลุ่ม α และกลุ่ม β ตามลำดับ

$Z_{i\alpha}$, $Z_{i\beta}$ คือ ค่าคะแนนมาตรฐานของตำแหน่งของข้อกระทงที่ i ของสเกลสำหรับกลุ่ม α และกลุ่ม β ตามลำดับ

ให้สมการที่ (42) และ (43) เท่ากัน

$$Z_{i\beta} = (S_{y\alpha} / S_{y\beta}) Z_{i\alpha} + \frac{M_{y\alpha} - M_{y\beta}}{S_{y\beta}} \quad \text{.....[44]}$$

ถ้าพล็อตความเบี่ยงเบนปกติระหว่าง Z_i กับ Z_i ความสัมพันธ์ระหว่างทั้งสองสามารถอธิบายได้ด้วย

$$Z_i' = (S_\beta' / S_\alpha') Z_{i\alpha} + M_\beta' - (S_\beta' / S_\alpha') M_\alpha' \quad \text{.....[45]}$$

เมื่อ M_α', M_β' คือ ค่าเฉลี่ยของความเบี่ยงเบนปกติของทั้งสองกลุ่ม

S_α', S_β' คือ ค่าเบี่ยงเบนมาตรฐานของความเบี่ยงเบนปกติของทั้งสองกลุ่ม

สมมุติให้สมการ [44] และ [45] เป็นการอธิบายความสัมพันธ์เหมือนกัน จึงสรุปได้ว่าความชันเท่ากัน และจุดตัดแกนเท่ากัน คือ

$$S_{y\alpha} / S_{y\beta} = S_\beta' / S_\alpha' \quad \text{.....[46]}$$

$$(M_{y\alpha} - M_{y\beta}) / S_{y\beta} = M_\beta' - (S_\beta' / S_\alpha') M_\alpha' \quad \text{.....[47]}$$

I17031965

จากสมการ [46],[47] ค่าประมาณ $\hat{S}_{y\alpha}$ และ $\hat{M}_{y\alpha}$ ซึ่งจะใช้สำหรับคำนวณค่าความชัน คือ

$$A = \hat{S}_{y\alpha} / \hat{S}_{x\alpha}$$

และจุดตัดแกน คือ $B = M_{y\alpha} - AM_{x\alpha}$

สำหรับสมการแปลงคะแนนจากฟอร์ม X ไปสู่ ฟอร์ม Y หาได้จาก

$$\hat{S}_{y\alpha} = S_{y\beta}(S_\beta / S_\alpha)$$

และ $$\hat{M}_{y\alpha} = S_{y\beta}[(M'_\beta - (S'_\beta / S'_\alpha) M'_\alpha] + M_{y\beta}$$

วิธีการเทียบมาตราของแฟน (Swine-Fan method of equating)

วิธีการนี้คล้ายกับของเธอร์สโตน ซึ่งแฟน (Fan) จัดดำเนินการโดยแบบสอบมีกลุ่มข้อกระทงร่วมกันเหมือนกันทั้งสองฉบับ (ฟอร์ม X และ Y) กลุ่ม สอบฟอร์ม X กลุ่ม สอบฟอร์ม Y มีการคำนวณความชันและจุดตัดแกน คือ

$$A = S_{y\alpha} / S_{x\alpha}$$

$$B = M_{y\alpha} - AM_{x\alpha}$$

เพื่อจะจัดเข้าสมการ $Y = AX + B$

การหาค่าของ M_y และ S_y โดยหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานจากคะแนนดิบของข้อกระทงที่ทำถูก (W) โดย

$$M_w = \sum_{i=1}^{n} p_i \quad \text{.....[48]}$$

$$S_w = \sum_{i=1}^{n} p_i d_i \quad \text{.....[49]}$$

เมื่อ p_i คือ สัดส่วนของการตอบถูกข้อกระทงที่ i

d_i คือ $(M_i - M_w)/ S_w$

M_i คือ ค่าเฉลี่ยของคะแนนดิบที่ตอบข้อกระทงที่ i ถูก

สมการที่ [49] นำมาจาก กิลลิคเสน (Gulliksen) ซึ่งแสดงไว้ว่า

$$S_w = \sum_{i=1}^{n} r_{iw} \sqrt{p_i q_i}$$

เมื่อ r_{iw} คือ พอยท์ไบซีเรียลของข้อกระทง (point-biserial item-test correlation)

และ $r_{iw}\sqrt{p_i q_i} = p_i (M_i - M_w)/ S_w$

$$Z'_\alpha = (S'_\alpha / S'_\beta) Z'_\beta + M'_\alpha - (S'_\alpha / S'_\beta) M'_\beta \quad \text{.....[50]}$$

เมื่อ M', S' เป็นค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่า Z'_α

การใช้เส้นของค่า Z'_α อาจประมาณได้จาก Z'_β จากข้อกระทง ค่า Z'_α ไปยังค่า p สำหรับ

ข้อกระทงทั้งหมดในฟอร์ม Y ประมาณค่าในกลุ่ม α แล้วรวมจำนวนข้อกระทงทั้งหมดจากฟอร์ม Y มาคำนวณค่า $M_{y\alpha}$ ซึ่งแสดงในสมการที่ [48] พล็อคกราฟค่า d_i จากข้อกระทงที่ร่วมกัน

$$d = (S'_{\alpha}/S'_{\beta}) d_{\beta} + M'_{\alpha} - (S'_{\alpha}/S'_{\beta}) M'_{\beta} \ldots [51]$$

เมื่อ M", S" เป็นค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานจากค่า d จากสมการที่ [50],[51] ถูกใช้คำนวณค่า d จากข้อกระทงที่ไม่ร่วมกันในฟอร์ม Y จากกลุ่ม α คูณค่า d_i กับ p_i จากข้อกระทงที่ร่วมกันและคำนวณค่า d_i กับ p_i จากข้อกระทงที่ร่วมกัน แล้วรวมผลที่ได้ก็จะได้ตามสมการที่ [49] ซึ่งเป็นสมการที่ประมาณค่า $S_{y\alpha}$ และการคำนวณค่า $\hat{M}_{y\alpha}$, $\hat{S}_{y\alpha}$ ใช้ค่า $M_{x\alpha}$ และ $S_{x\alpha}$ มาเป็นพารามิเตอร์ในสมการ $Y = AX + B$

จากรูปแบบการรวบรวมข้อมูลของแองกอฟฟ์ (Angoff) ทั้ง 6 รูปแบบ มาร์โคได้บรรยายซ้ำ แต่ได้แยกประเด็นของการออกแบบเพื่อรวบรวมข้อมูลกับการจัดกระทำทางสถิติออกจากกันทั้งนี้เพราะวิธีการทางสถิติบางวิธีใช้ได้กับแบบแผนการรวบรวมข้อมูลมากกว่าหนึ่งแบบดังนี้คือ

แบบแผนที่ 1 กลุ่มเดียวทำแบบสอบทั้งสองชุด คือ ชุดใหม่และชุดเก่า การทดสอบอาจจัดทำในวันเดียว หรือ 2 วันติดต่อกัน แต่หลักการที่สำคัญ คือ ช่วงระยะห่างของการทำแบบสอบ 2 ชุด ต้องเป็นเวลาสั้น ๆ เพื่อมิให้เกิดประสบการณ์แทรกซ้อนที่มีผลกระทบต่อคะแนน ลำดับการทดสอบชุดใดก่อนไม่ใช่ปัญหา เพราะถือว่าองค์ประกอบต่าง ๆ อันได้แก่ การเรียนรู้ การฝึกฝน ความล้า มีผลต่อคะแนนน้อยมาก

แบบแผนที่ 2 กลุ่มสุ่มสองกลุ่ม แต่ละกลุ่มได้รับการทดสอบทั้งสองชุดในลักษณะการจัดลำดับการสอบก่อนหลังสลับให้เกิดความสมดุลกัน แบบแผนที่ 2 เป็นแบบแผนที่จะรับประกันอิทธิพลขององค์ประกอบการเรียนรู้ การฝึกฝน และความล้า เป็นการขจัดความลำเอียงที่จะเกิดขึ้นกับผลการสอบชุดใดชุดหนึ่ง

แบบแผนที่ 3 กลุ่มสุ่มสองกลุ่ม ให้แต่ละกลุ่มทำการสอบเพียงชุดเดียว สิ่งสำคัญของแบบแผนที่ 3 คือ ต้องมีกลุ่มสุ่ม 2 กลุ่มที่มีความคล้ายคลึงกันทั้งทางด้านความรู้ ความสามารถหรือทักษะที่ต้องการวัดเพื่อให้คะแนนสอบที่ได้มาไม่ใช่เป็นผลของความแตกต่างของความสามารถของกลุ่ม

แบบแผนที่ 4 กลุ่มสุ่มสองกลุ่มแต่ละกลุ่มทำแบบสอบเพียงชุดเดียวและทำแบบสอบร่วมเหมือนกันอีกส่วนหนึ่ง แบบแผนนี้มีข้อได้เปรียบกว่าแบบแผนที่ 3 ในประเด็นที่ว่าได้ข้อสนเทศจากแบบสอบร่วมจากกลุ่มสุ่มทั้งสอง แบบสอบร่วมอาจเป็นฉบับแยกจากแบบสอบทั้งสองชุด(external)

หรือเป็นส่วนที่ผนวกเข้าในแบบสอบทั้งสองชุด (internal) ก็ได้ส่วนที่เป็นแบบสอบร่วมควรประกอบด้วยคำถามที่คล้ายคลึงกับคำถามในแบบสอบที่ต้องการเทียบคะแนนทั้งสองชุด ประโยชน์จากแบบสอบร่วมที่มีต่อการเทียบมาตราของแบบสอบสองชุดจะมีมากหรือน้อยขึ้นอยู่กับความสัมพันธ์ของแบบสอบร่วมกับแบบสอบสองชุดนั้น เหตุผลที่แบบแผนที่ 4 มีข้อได้เปรียบเพราะคะแนนชุดเก่าและใหม่สามารถรับการปรับเพื่อสะท้อนให้เห็นความแตกต่างที่ปรากฏอยู่ในกลุ่มสุ่มทั้งสอง ตามผลของแบบสอบร่วมได้

แบบแผนที่ 5 กลุ่มไม่ได้จากการสุ่มสองกลุ่มแต่ละกลุ่มทำแบบสอบเพียงชุดเดียวและทำแบบสอบร่วมเหมือนกันอีกส่วนหนึ่งเช่นเดียวกับแบบแผนที่ 4 กลุ่มผู้สอบที่ไม่ได้มาจากการสุ่มมักจะเกิดขึ้นกับการสอบตามธรรมชาติ ปัจจัยสำคัญของแบบแผนนี้อยู่ที่แบบสอบร่วมซึ่งจะต้องให้มีความคล้ายคลึงมากที่สุดกับแบบสอบสองชุดที่ต้องการทำการเทียบมาตรา โดยปกติแล้วไม่มีวิธีการทางสถิติใดที่จะทำการปรับคะแนนในกรณีกลุ่มที่ไม่ได้มาจากการสุ่ม แต่จะทำได้ต่อเมื่อมีแบบสอบร่วมที่มีความเป็นคู่ขนานกับแบบสอบชุดใหม่และเก่า หน้าที่ของแบบสอบร่วม คือ กำหนดจุดอ้างอิงให้เกิดขึ้นกับทั้งสองชุดในทำนองกับการกำหนดจุดเดือด หรือ จุดเยือกแข็งของน้ำ เพื่อใช้กำหนดมาตราบนเทอร์โมมิเตอร์ฟาเรนไฮท์ให้เท่ากับองศาบนเทอร์โมมิเตอร์แบบเซลเซียสได้แบบแผนนี้ได้นำไปใช้กับการเทียบมาตราในแบบสอบ SAT มาแล้ว

จากแบบแผนการเก็บรวบรวมข้อมูลและการใช้สถิติในการเทียบมาตรานี้พบว่าแบบแผนที่สำคัญที่ก่อให้เกิดการประยุกต์ใช้ได้อย่างมาก คือ แบบแผนที่มีการใช้แบบสอบร่วม ซึ่งยังคงใช้อยู่ในโครงการทดสอบของ SAT และสอดคล้องกับเงื่อนไขในสภาพแวดล้อมปัจจุบันด้วย จากสถานการณ์ที่ผู้วิจัยจัดกระทำตรงกับรูปแบบที่ 3 ของแองกอฟฟ์ (Angoff)ในกรณีแบบสอบมีความเที่ยงไม่เท่ากัน

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย

<u>การหาคุณภาพของวิธีเทียบมาตรา</u>

การเทียบมาตรารูปแบบใดก็ตามจะมีคุณภาพดีที่สุดเมื่อคะแนนที่ได้จากแบบสอบที่ต้องการเป็นไปตามเงื่อนไขต่าง ๆ ที่กำหนดไว้ของแต่ละรูปแบบ แต่ในสถานการณ์จริงมีข้อจำกัดทำให้ไม่สามารถได้ข้อมูลตามเงื่อนไข ดังนั้นจึงจำเป็นจะต้องตรวจสอบความเพียงพอ (Adequacy) ของการเทียบมาตราหรือประเมินคุณภาพของรูปแบบที่แตกต่างกันซึ่งวิธีการประเมินความเพียงพอมีผู้เสนอแนวคิดและวิถีปฏิบัติไว้ดังนี้

การประเมินก่อนดำเนินการ

ดัชนีตรวจสอบความเพียงพอของเจเกอร์ (Jaeger)

เจเกอร์ (Jaeger 1981: 26) ได้เสนอดัชนี 5 ตัว เพื่อตรวจสอบความเพียงพอของการใช้วิธีเชิงเส้นตรงว่า เทคนิควิธีที่ได้นำมาใช้ในการเทียบเพียงพอกับการปรับความแตกต่างระหว่างการแจกแจงของคะแนนจากแบบสอบหรือไม่ หรือจำเป็นต้องมีวิธีอื่นที่มีความสลับซับซ้อนมากขึ้น ดัชนี 5 ตัว มีดังนี้ คือ

1) ดัชนีความคล้ายคลึงของการแจกแจงคะแนนสะสมของแบบสอบเก่า และ ชุดใหม่ โดยการปรับความแตกต่างระหว่างค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน การทดสอบความเหมือนของการแจกแจงใช้การทดสอบด้วย the Kolmogorov-Smirnov two-sample test

2) รูปแบบของการแจกแจงคะแนนดิบกับคะแนนแปลง(Shape of the raw score tranformation) เหตุผลของการใช้ดัชนีตัวนี้ คือ ถ้าการเทียบมาตราด้วยวิธีเชิงเส้นตรงสามารถอธิบายความแตกต่างในการแจกแจงของคะแนนดิบทั้งสองชุดอย่างเพียงพอ ก็เป็นเหตุผลอย่างเพียงพอเช่นกันที่จะยอมรับว่าการแปลงคะแนนดิบจากแบบสอบชุดใหม่ไปยังแบบสอบชุดเก่าเป็นเส้นตรงอย่างแน่นอน

3) ความคงเส้นคงวาของผลลัพธ์ของการเทียบมาตราตามวิธีเชิงเส้นตรงกับการเทียบที่ตำแหน่งเปอร์เซนไทล์ (Consistency of linear and equipercentile equating results) การวิเคราะห์นี้อาศัยข้อตกลงที่เป็นสมมุติฐานเบื้องต้นที่ว่า ถ้าวิธีเชิงเส้นตรงมีความเพียงพอแล้ว ฟังชั่นของวิธีการเทียบที่ตำแหน่งเปอร์เซนไทล์จะแปรผันไปรอบ ๆ ฟังชั่นของวิธีเชิงเส้นตรงที่สมนัยกัน

4) ความคล้ายคลึงของการแจกแจงความยากของข้อกระทง (Similarity of item difficulty distributions) โดยอาศัยหลักการที่ว่า การใช้วิธีเชิงเส้นตรงมีความเพียงพออย่างแท้จริงกับแบบสอบที่มีคุณสมบัติเป็นคู่ขนานกัน ถ้ามีการเบี่ยงเบนจากความเป็นคู่ขนานมากเท่าใด แสดงว่า ต้องการวิธีการเทียบมาตราที่ซับซ้อนขึ้นเพราะการแจกแจงของแบบสอบที่ไม่ใช่คู่ขนานจะมีความแตกต่างเกิดขึ้นในระดับโมเมนต์ที่สูงขึ้น

5) ความคล้ายคลึงของค่าอำนาจจำแนกของข้อกระทง (Similarity of item discrimination distributions) เหตุผลทำนองเดียวกับดัชนีตัวที่ 4

การประเมินผลการเทียบมาตรา

1. ดัชนีเปรียบเทียบเปอร์เซนไทล์ (the percentile comparison index)

โคเลนและวิทนีย์ (Kolen and Whitney 1982: 284)ได้แนะนำให้ใช้ดัชนีเปรียบเทียบเปอร์เซนไทล์ เพื่อประเมินความเพียงพอของการเทียบมาตราทั้งนี้โดยอาศัยกลุ่มตัวอย่างตรวจสอบผล (cross-validation sample examinees) ซึ่งเป็นกลุ่มอิสระ อีกกลุ่มหนึ่งที่สุ่มมาจากประชากรเดียวกันกับกลุ่มผู้สอบที่ใช้สร้างตารางการเทียบมาตรา วิธีการหาค่าดัชนีเปรียบเทียบเปอร์เซนไทล์นี้ คือ การวัดความไม่สอดคล้อง (dissimilarity) ระหว่างการแจกแจงคะแนนที่ได้จากกลุ่มตรวจสอบผล คือ คะแนนจากแบบสอบชุด X กับคะแนนซึ่งได้มาจากการแปลงคะแนนจากแบบสอบชุด Y ไปยังสเกลของแบบสอบชุด X ตามตารางการเทียบมาตราที่ได้จากวิธีการที่ใช้เทียบ โดยหาค่าสมมูล ณ ที่ตำแหน่งเปอร์เซนไทล์เดียวกันดัชนีดังกล่าวคำนวณได้จาก

$$C = \sum_{i} (x_i - x_i^{*})^2 / nk$$

เมื่อ n คือ จำนวนคะแนนดิบในกลุ่มตรวจสอบผล

k คือ จำนวนข้อสอบในแบบสอบร่วมที่ใช้

ถ้าค่า C น้อยกว่า แสดงว่าวิธีการเทียบมาตราที่นำมาใช้สร้างตารางการแปลงคะแนนนั้นมีความเพียงพอที่จะนำมาใช้ให้ได้ผลการเทียบที่คงที่มากกว่า

2. ดัชนีความแตกต่าง (Discrepancy Indicies)

ปีเตอร์เสน มาร์โค และสตีเวอร์ท (Petersen, Marco and Stewart 1982: 91) ได้เสนอ วิธีการประเมินความเพียงพอของการเทียบมาตราด้วยการใช้ดัชนีความแตกต่าง ซึ่งมีแนวคิดว่าคะแนนดิบแต่ละตัวจากแบบสอบชุด X จะมีคะแนนเกณฑ์ t และคะแนนแปลง t' ที่ได้จากการประมาณค่า โดยใช้วิธีการเทียบมาตราเฉพาะอันหนึ่ง ค่าความแตกต่าง d ระหว่าง t และ t' เป็นความคลาดเคลื่อนในการเทียบมาตรา ซึ่งค่าที่น้อยกว่าแสดงว่า วิธีการเทียบมาตรานั้นมีความเหมาะสมมากกว่า

ค่าดัชนีความแตกต่าง คือค่าความคลาดเคลื่อนรวมในการเทียบมาตราของวิธีการเทียบมาตราที่ใช้นั้นซึ่งคำนวณจากค่าเฉลี่ยยกกำลังสองของความแตกต่างระหว่างคะแนนแปลงกับคะแนนเกณฑ์และถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ (the standardized weighted mean-square difference) ซึ่งเท่ากับค่าความแปรปรวนของความแตกต่างรวมกับค่ากำลังสองของความลำเอียง

$$\text{total error} = {}_j\sum f_j d_j \,/\, nS_t^2$$

เมื่อ $d_j = (t - t')$

n = จำนวนคะแนนที่ใช้

S_t^2 = ความแปรปรวนของคะแนน t

ดัชนีเปรียบเทียบความแตกต่างนี้เป็นค่ามาตรฐาน (แสดงในรูปสัดส่วนของความเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ ค่าความคลาดเคลื่อนรวมที่ได้นี้จึงสามารถเปรียบเทียบกันได้โดยตรงระหว่างค่าที่ได้จากการใช้วิธีและสถานการณ์การเทียบที่แตกต่างกัน(Petersen,Marco and Stewart 1982: 91)

3. **ดัชนีความแตกต่างมาตรฐาน** (Index of standard discripancy)

ภาวิณี ศรีสุขวัฒนานันท์ (2528) ได้เสนอวิธีการประเมินความเพียงพอโดยใช้ดัชนีความแตกต่าง ซึ่งดัดแปลงจากสูตรของโคเลนได้ใช้ข้อมูลคะแนนจากผู้สอบเองเป็นเกณฑ์ในการหาความแตกต่าง ข้อมูลเหล่านี้ได้มาจากการออกแบบด้วยการใช้กลุ่มตัวอย่างตรวจสอบผล ซึ่งผู้สอบในกลุ่มตัวอย่างนี้ได้รับการทดสอบด้วยแบบสอบทั้งสองชุด ดังนั้น การใช้คะแนนของตนเองเป็นเกณฑ์จึงมีความอิสระไม่ขึ้นกับกระบวนการแปลงคะแนนอื่น ๆ และใช้แนวคิดของปีเตอร์เซนและคณะ คือ ใช้ค่าความแปรปรวนเป็นตัวถ่วงน้ำหนักเพื่อให้ค่าที่ได้มีหน่วยเป็นมาตรฐาน

$$C = {}_i\sum (X_i - X_i^*)^2 \,/\, nS_x^2$$

เมื่อ X_i คือ คะแนนเกณฑ์หรือคะแนนจากการสอบชุด X ของคนที่ i

X_i คือ คะแนนที่ได้เทียบด้วยตารางแปลงคะแนนที่สมนัยกันของคนที่ i

n คือ จำนวนในกลุ่มตัวอย่างตรวจสอบผลที่นำมาวิเคราะห์

S_x^2 คือ ค่าความแปรปรวนของคะแนน X

จากงานวิจัยนี้ ผู้วิจัยได้หาคุณภาพของการเทียบมาตราจากดัชนีความแตกต่างมาตรฐาน ซึ่งดัชนีความแตกต่างมาตรฐานนี้ โดยใช้แนวคิดของภาวิณี ศรีสุขวัฒนานันท์

งานวิจัยที่เกี่ยวข้อง

โคเลน (Kolen 1981: 1-11) ได้เปรียบเทียบผลการเทียบมาตราระหว่างวิธีดั้งเดิมสองวิธีกับวิธี IRT 7 วิธี โดยใช้ข้อมูลจากผลการสอบนักเรียนในรัฐไอโอวา ปี ค.ศ. 1978 ตามโครงการชื่อ The Iowa Tests of Education Development แบบสอบที่ใช้คือ แบบสอบที่พิมพ์ใหม่ครั้งที่ 7 เทียบไปสู่ แบบสอบชุดเก่าที่พิมพ์ครั้งที่ 6 แบบสอบที่พิมพ์ครั้งที่ 6 มีหนึ่งชุด คือ X6 ใช้สอบนักเรียนในทั้งสองระดับ คือระดับที่ 1 และ ระดับ 2 ซึ่งระดับที่ 1 เป็นนักเรียนเกรด 9 และ 10 , ระดับ 2 เป็นนักเรียนเกรด 11 และ 12 ส่วนแบบสอบที่พิมพ์ครั้งที่ 7 มีสองชุดที่คู่ขนานกันคือ X7 และ Y7 แต่ละชุดแบ่งแบบสอบเป็น 2 ระดับ ซึ่งมีความยากแตกต่างกันเพื่อแยกสอบนักเรียนแต่ละคน โดยมีแผนการรวบรวมข้อมูลดังนี้

นักเรียนระดับ 1 สอบแบบสอบ X6, X7 ระดับ 1 และ Y7 ระดับ 1

นักเรียนระดับ 2 สอบแบบสอบ X6, X7 ระดับ 2 และ Y7 ระดับ 2

ทั้งนี้นักเรียนแต่ละคนจะได้รับแบบสอบเพียง 1 ชุด คือ X6 , X7 หรือ Y7 โดยการสุ่ ดังนั้นจำนวนนักเรียนที่สอบแบบสอบแต่ละชุดจะมี 1 ใน 3 ของจำนวนนักเรียนรวมสองระดับสำหรับเกณฑ์ในการเปรียบเทียบผลใช้ค่าดัชนีจากกลุ่มตรวจสอบผล ซึ่งเป็นนักเรียนทุก ๆ คนที่ 3 จากการสอบแบบสอบแต่ละชุดและแต่ละระดับ

การเทียบมาตราแบบดั้งเดิมใช้วิธีเชิงเส้นตรงและอิควิเปอร์เซนไทล์ ซึ่งจัดกระทำตามวิธีของแองกอฟฟ์ (Angoff) ส่วนวิธี IRT ใช้ชนิดหนึ่ง, สอง และสามพารามิเตอร์ โดยแต่ละโมเดลเทียบ 2 แบบ คือ เทียบด้วยค่าประมาณคะแนนจริง และเทียบด้วยค่าประมาณคะแนนสังเกต ส่วนอีกวิธีหนึ่งคือราสช์โมเดล

ดัชนีสำหรับตรวจสอบผล โคเลนใช้วิธีคำนวณจากค่าเฉลี่ยกำลังสองของความแตกต่าง (สำหรับผู้สอบแบบสอบ X6 ในกลุ่มตรวจสอบผล) ระหว่างคะแนนจากแบบสอบ X6 กับคะแนนจากแบบสอบ X7 และ Y7 ที่แปลงแล้ว ณ ตำแหน่งเปอร์เซนไทล์เดียวกัน ทั้งนี้โดยใช้วิธีการเทียบมาตราแบบต่าง ๆ ทั้งเก้าวิธี หากค่าดัชนีจากกลุ่มตรวจสอบผลที่ได้จากวิธีการเทียบใดมีค่าน้อยกว่า แสดงว่าวิธีนั้นมีความสอดคล้องระหว่างการแจกแจงคะแนนของแบบสอบชุดเก่ากับคะแนนของแบบสอบชุดใหม่ที่แปลงไปยังสเกลของชุดเก่ามากกว่านั่นคือ วิธีนั้นมีความคงเส้นคงวามากกว่า ทั้งนี้โดยโคเลนทดสอบความแตกต่างของดัชนีที่ได้จากวิธีการเทียบมาตราแบบต่าง ๆ เพื่อเปรียบเทียบผลของแต่ละวิธีโดยใช้ค่าสถิติทดสอบฟรายด์แมน (Friedman test)

ผลการวิจัยพบว่า วิธีการเทียบมาตราแบบต่าง ๆให้ผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 เมื่อใช้แบบสอบระดับ 1 และ 2 ตามลำดับ ถ้าพิจารณาแบบสอบระดับที่ 1 พบว่า วิธีที่ให้ผลคงเส้นคงวามากที่สุดคือ วิธี IRT ชนิดสามพารามิเตอร์ที่ประมาณค่าสังเกต รองลงมาคือวิธี IRT ชนิดหนึ่งพารามิเตอร์ และให้ผลน้อยที่สุดคือ วิธีเชิงเส้นตรง สำหรับแบบสอบระดับที่ 2 พบว่า วิธีการที่ให้ผลคงเส้นคงวามากที่สุดคือ วิธี IRT ชนิดสาม-พารามิเตอร์ที่ประมาณค่าคะแนนจริง วิธีของราสช์ให้ผลคงเส้นคงวามากกว่าวิธี IRT ชนิดหนึ่งพารามิเตอร์ที่ประมาณค่าคะแนนจริง

สำหรับวิธีเทียบมาตราแบบ IRT ชนิดสามพารามิเตอร์ให้ผลที่เพียงพอมากที่สุดยกเว้นการเทียบมาตราที่อยู่ต่ำกว่าระดับการเดา ดังนั้นการเทียบมาตราในส่วนนี้จึงต้องอาศัยวิธีการอื่น เช่น วิธีเชิงเส้นตรง

ส่วนวิธีเทียบแบบอิควิเปอร์เซนไทล์ พบว่าให้ผลที่เพียงพอที่จะนำมาใช้เทียบมาตราแบบสอบที่มีความยากต่างกัน แต่วิธีเทียบเชิงเส้นตรงพบว่าให้ผลที่ไม่น่าพอใจในกรณีนี้ ซึ่งจากผลการวิจัยของสไลด์และลินน์ (Slinde and Linn 1977: 23-31) ก็ได้พบว่า การเทียบมาตราแบบอิควิเปอร์เซนไทล์ให้ผลดีกว่าการเทียบเชิงเส้นตรงในกรณีที่แบบสอบมีค่าความยากต่างกัน

คุก ดันบาร์ และไอเนอร์ (Cook,Dunbar and Eignor 1981: 1-52)ได้ศึกษาผลของการเทียบมาตราโดยใช้ทฤษฎีการตอบข้อสอบ (IRT)เปรียบเทียบกับการเทียบมาตราแบบดั้งเดิม 2 วิธี คือ การเทียบมาตราเชิงเส้นตรงและการเทียบมาตราแบบอิควิเปอร์เซนไทล์ การเทียบมาตราแบบ IRT ใช้โมเดลแบบสามพารามิเตอร์ซึ่งประมาณค่าพารามิเตอร์โดยใช้โปรแกรม LOGIST ด้วยวิธีความเป็นไปได้สูงสุด (maximum likelihood procedure) สำหรับการประมาณค่าที่อยู่ต่ำกว่าระดับการเดาใช้วิธีการเชิงเส้นตรง(linear interpolation) ในการเทียบมาตราศึกษาทั้งกรณีแบบสอบร่วมและไม่มีแบบสอบร่วม โดยใช้กลุ่มตัวอย่างที่ไม่ได้สุ่มมาจากประชากรเดียวกัน ข้อมูลที่ใช้ศึกษาเป็นผลการสอบ 2 ครั้ง ที่ดำเนินการแล้วโดย The Collage Board Admissions Testing Program ซึ่งใช้แบบสอบ 2 ชุด ที่มีความยากและความยาวแตกต่างกัน การเปรียบเทียบพิจารณา 2 ประการ คือ ความสอดคล้องสัมพัทธ์ (relative agreement) ระหว่างวิธี IRT กับ วิธีดั้งเดิมแต่ละวิธี โดยพิจารณาจากการพล็อตกราฟ อีกประการหนึ่งคือ ดัชนีความแตกต่าง (discrepancy indices) สำหรับการแจกแจงคะแนนรวมและแต่ละส่วนของการแจกแจง (3 ส่วน) ซึ่งคำนวณจากค่าที่ถ่วงน้ำหนักแล้ว

ของค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างค่าประมาณจากการเทียบมาตราแต่ละวิธีที่ศึกษากับคะแนนเกณฑ์ เกณฑ์ในที่นี้คือคะแนนแปลงด้วยการเทียบมาตราแบบ IRT ชนิดสามพารามิเตอร์ ซึ่งเคยวิจัยมาก่อนและได้แนะนำว่ามีความเหมาะสมในการเทียบมาตรา ในกรณีที่แบบสอบมีความยากต่างกัน สอบกับกลุ่มตัวอย่างที่มีความสามารถต่างกัน

ผลการวิจัยพบว่า วิธีการเทียบมาตราแบบดั้งเดิมกับวิธีการเทียบแบบ IRT สอดคล้องกันมาก เนื่องจากการแจกแจงคะแนนดิบของแบบสอบ 2 ชุดมีรูปร่างคล้ายคลึงกันมาก อย่างไรก็ตามจากความแตกต่างที่เกิดขึ้นที่ส่วนปลายของวิธีการเทียบแบบดั้งเดิมทุกวิธีที่ต่างจากการเทียบแบบ IRT เป็นผลจากข้อมูลในส่วนนี้มีน้อยและหายากมาก ในทางทฤษฎีแนะนำให้ใช้การเทียบแบบ IRT เพราะไม่มีผลเนื่องจากขาดแคลนข้อมูลในส่วนปลายการแจกแจง แต่การเทียบแบบ IRT ก็ไม่สามารถเตรียมการเทียบมาตราในส่วนที่อยู่ปลายด้านล่างที่ต่ำกว่าระดับการเดาได้ ซึ่งจำเป็นต้องใช้วิธีอื่นประมาณค่าเพิ่มเติม นอกจากนี้ผลการวิจัยพบว่า การเทียบมาตราแบบ IRT โดยตรงเมื่อไม่มีแบบสอบร่วมและเมื่อกลุ่มตัวอย่างไม่ได้เป็นกลุ่มที่เท่าเทียมกันโดยสุ่มนั้นให้ผลแตกต่างจากการเทียบเมื่อมีแบบสอบร่วมเพียงเล็กน้อย

ปีเตอร์เสน มาร์โค และสตีเวอร์ท (Petersen, Marco and Stewart 1982: 71-135) ได้ศึกษาความเพียงพอของการเทียบมาตราเชิงเส้นตรงหลาย ๆ วิธีและวิธีอิควิเปอร์เซนไทล์ ในสถานการณ์ต่าง ๆ โดยใช้ข้อมูลจากผลการสอบระดับชาติสองครั้ง คือ ครั้งที่สอบเมื่อเดือนเมษายน ปี ค.ศ.1975 เป็นแบบสอบชุด SAT (Scholastic Aptitude Test) ซึ่งมี 3 ฉบับ คือ ee, fe และ fm อีกครั้งหนึ่งสอบเมื่อเดือนพฤศจิกายน ปี ค.ศ.1975 เป็นแบบสอบชุด TSWE (Test of Standard Written English) ซึ่งมี 3 ฉบับ คือ eg, fg และ fo ซึ่งบันทึกผลการตอบข้อสอบรายข้อของผู้สอบทั้งหมดไว้ในคอมพิวเตอร์ กลุ่มตัวอย่างสร้างขึ้นตามลักษณะเฉพาะที่กำหนดไว้ โดยทำการสุ่มมาจำนวน 6 กลุ่ม แต่ละกลุ่มตัวอย่างมีจำนวนผู้สอบ 4,731 คน แบ่งแต่ละกลุ่มออกเป็น 3 กลุ่มย่อยโดยไม่ให้ซ้ำกัน ทั้งนี้ให้มีลักษณะต่างกันดังนี้ กลุ่มสุ่มอย่างง่าย กลุ่มที่มีความสามารถคล้ายคลึงกัน (ใช้ค่าเฉลี่ยความสามารถทางภาษาเป็นเกณฑ์) และกลุ่มที่มีความสามารถไม่คล้ายคลึงกัน ดังนั้นจึงมีกลุ่มย่อยทั้งหมด 54 กลุ่ม การออกแบบการเทียบมาตราใช้แบบสอบร่วมลักษณะต่าง ๆ กันกับแบบสอบที่ต้องการเทียบ คือ ต่างกันในตำแหน่งของแบบสอบร่วม (ชนิดภายนอกและภายใน) เนื้อหา (คล้ายคลึงและไม่คล้ายคลึง) และความยาก (คล้ายคลึงและไม่คล้ายคลึง) โดยทำการเทียบแบบสอบชุด SAT

ส่วนที่เป็นภาษา (verbal) กับตัวเองเหมือนกับว่าเป็นการเทียบแบบสอบสองชุดโดยใช้แบบสอบร่วมลักษณะต่าง ๆเพื่อศึกษาระดับความคล้ายคลึงระหว่างแบบสอบที่ต้องการเทียบกับแบบสอบร่วม และระดับความคล้ายคลึงระหว่างกลุ่มตัวอย่างสองกลุ่มที่ใช้ในการเทียบที่มีต่อผลการเทียบมาตรา ทั้งนี้การจัดกระทำตัวแปรต่าง ๆดำเนินการอย่างเป็นระบบ(systematicaly manipulated) นอกจากนี้ได้มีการเทียบมาตราในระหว่างแบบสอบ (SAT กับ TSWE)โดยใช้แบบสอบร่วมภายในเพื่อศึกษาระดับความคล้ายคลึงระหว่างแบบสอบที่ต้องการเทียบทั้งสองชุด และระดับของความคล้ายคลึงระหว่างกลุ่มตัวอย่างที่ใช้เทียบสองกลุ่มที่มีผลต่อการเทียบมาตรา โดยแบบสอบที่เทียบมีลักษณะที่แตกต่างกันด้านต่าง ๆคือ เนื้อหา ความยากและความยาวของแบบสอบ การวิเคราะห์ข้อสอบทำการเลือกคะแนนตามจุดมุ่งหมาย แล้วคำนวณค่าสถิติรายข้อและข้อมูลอื่น ๆที่ต้องการใช้เพื่อเทียบแบบสอบในสถานการณ์ต่าง ๆ

เกณฑ์ที่ใช้ในการประเมินความเพียงพอของวิธีการเทียบมาตรา ใช้ค่าดัชนีความแตกต่าง (discrepancy indices) ซึ่งเป็นค่าที่ถ่วงน้ำหนักแล้วของค่าเฉลี่ยกำลังสองของความแตกต่าง ระหว่างค่าประมาณหรือคะแนนแปลงกับคะแนนเกณฑ์ คะแนนเกณฑ์ที่ใช้คือคะแนนแปลงจากคะแนนดิบชุดเดียวกัน ดัชนีนี้เป็นค่ามาตรฐานโดยอยู่ในรูปสัดส่วนของความเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ ดังนั้นจึงสามารถเปรียบเทียบผลการเทียบมาตราระหว่างสถานการณ์ต่าง ๆ และระหว่างวิธีการเทียบมาตราแต่ต่าง ๆ ได้ ทั้งนี้ค่าดัชนีความแตกต่างหรือความคลาดเคลื่อนรวม (total error;TE)ซึ่งคำนวณได้จากสมการแสดงอยู่ในสเกลที่มีค่าเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์เท่ากับ 100 และจำแนกตามระดับคุณภาพสำหรับการประเมินความเพียงพอของวิธีการเทียบมาตรา ดังนี้

TE	≤ 25	คือ	น่าพอใจมาก
25 < TE	≤ 100	คือ	น่าพอใจ
100 < TE	≤ 225	คือ	ปานกลาง
225 < TE	≤ 400	คือ	ไม่น่าพอใจ
400 < TE		คือ	ไม่น่าพอใจมาก

ผลการวิจัยสรุปได้ดังนี้

1. วิธีการเทียบมาตราเชิงเส้นตรงโดยใช้เทคนิคของพอททอฟ (Potthoff) ให้ผลไม่น่าพอใจในเกือบทุกสถานการณ์ที่ศึกษา

2. กรณีที่กลุ่มตัวอย่างเท่ากัน วิธีอิควิเปอร์เซนไทล์ให้ผลที่มีความคลาดเคลื่อนรวมมากกว่า วิธีเชิงเส้นตรง

3. วิธีการเทียบมาตราเชิงเส้นตรงโดยเทคนิคของทัคเกอร์ 1 (Tucker 1)ให้ผลการเทียบที่มีความคลาดเคลื่อนมากกว่าวิธีเชิงเส้นตรงอื่น ๆ ในกรณีที่กลุ่มตัวอย่างแตกต่างกัน ยกเว้น สำหรับวิธีเชิงเส้นตรงโดยใช้เทคนิคของพอททอฟ

4. วิธีอิควิเปอร์เซนไทล์ให้ผลการเทียบดีกว่าวิธีเชิงเส้นตรง เมื่อแบบสอบสองชุดมีความสัมพันธ์เชิงเส้นโค้ง อันเนื่องมาจากแบบสอบมีความยากแตกต่างกันและจากกลุ่มตัวอย่างที่ไม่คล้ายคลึงกัน

5. ค่าความยากระหว่างแบบสอบร่วมและแบบสอบที่เทียบมาตราแตกต่างกันจะทำให้ความคลาดเคลื่อนในการเทียบมีมากกว่าความแตกต่างในเนื้อหา

6. ค่าความยากที่ต่างกันระหว่างแบบสอบทั้งสองชุดที่จะเทียบมาตรา จะทำให้ความคลาดเคลื่อนในการเทียบมีมากกว่าความแตกต่างในเนื้อหา

7. แบบสอบร่วมที่สร้างให้มีลักษณะคล้ายคลึงกับแบบสอบที่จะเทียบมาตราอย่างมาก หรือ อาจเรียกว่าเป็นฉบับย่อแล้ว การเทียบมาตราจะให้ผลดีที่สุด

โคป (Cope 1987: 143-149)ได้เปรียบเทียบมาตราเชิงเส้นตรง 5 วิธีระหว่างวิธีของแองกอฟฟ์ (Angoff 1971) ตามวิธีที่ 5 ซึ่งมี 3 วิธี กับวิธีของทักเกอร์ (Tucker) และเลวิน (Levine) ในกรณีที่แบบสอบมีความเชื่อมั่นเท่ากันโดยใช้แบบสอบร่วม แบบสอบที่ใช้เทียบมาตราเป็นแบบสอบเพื่อให้ประกาศนียบัตรในวิชาชีพ จำนวน 5 ชุด สอบกับกลุ่มประชากรที่ไม่เท่าเทียมกัน 5 กลุ่ม ๆ ละ 1 ชุด ปีละ 2 ครั้ง การเปรียบเทียบผลการเทียบระหว่างแบบสอบแต่ละคู่ (single link) ภายใต้วิธีการเชิงเส้นตรงที่แตกต่างกัน 5 วิธี ทั้งนี้ไม่มีการเปรียบเทียบความถูกต้องของวิธีการต่าง ๆ เพราะว่าสเกลคะแนนดิบของแบบสอบสองชุดถูกคาดว่าจะต่างกัน อีกแบบหนึ่งคือ การเปรียบเทียบผลการเทียบคะแนนแบบริช่ ซึ่งเป็นการเปรียบเทียบมาตราแบบสอบแต่ละชุดกับตัวเอง โดยผ่านแบบสอบชุดอื่นที่เลือก ทั้งนี้เพื่อทำการเปรียบเทียบความถูกต้องของวิธีการเหล่านั้นด้วย

การวัดความคลาดเคลื่อนของการเทียบมาตรา (equating error) ใช้สถิติ 2 ตัว คือ ค่าประมาณความลำเอียง (estimate bias) และค่าประมาณที่ถ่วงน้ำหนักแล้วของรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนในการเทียบมาตรา(estimated weighted root squared error; RMSE) ซึ่งใช้เปรียบเทียบระหว่างวิธีที่แตกต่างกัน คำนวณจากสมการ

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

เมื่อ $\bar{X}$ คือ ค่าเฉลี่ยของคะแนนดิบของแบบสอบชุดหนึ่ง

$\bar{X}'$ คือ ค่าเฉลี่ยของคะแนนดิบที่เท่าเทียมกันโดยการเทียบแบบสอบชุดนั้นไปยังตัวเอง

และ
$$RMSE = \frac{\sum (x_i' - x_i)^{1/2}}{\sum n_i}$$

เมื่อ x_i คือ คะแนนดิบของคนที่ i

x_i' คือ คะแนนดิบที่เท่าเทียมกัน

n คือ จำนวนผู้สอบ

ในกรณีที่มีผู้สอบจำนวนมาก การคำนวณ RMSE จะง่ายขึ้นโดยใช้สมการ

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

เมื่อ m คือ ความชันของฟังก์ชั่นการเทียบ $x' = mx + b$

Var(X) คือ ความแปรปรวนของ X

ผลการวิจัยสรุปว่า ในสถานการณ์ที่ศึกษาภายใต้รูปแบบที่ 5 ตามวิธีของแองกอฟฟ์ให้คะแนนดิบ (raw score equivalents) ใกล้เคียงกับวิธีการของทัคเกอร์และเลวิน โดยการเทียบมาตราแบบสายรัชทั้ง 5 วิธี ให้ค่าประมาณความลำเอียงและรากที่สองของค่าเฉลี่ยยกกำลังสองของความคลาดเคลื่อนที่ใกล้เคียงกัน ส่วนการเทียบมาตราจากวิธีของทักเกอร์และเลวิน ผู้วิจัยกล่าวว่าผลการวิจัยนี้สนับสนุนวิธีการเทียบมาตราเชิงเส้นตรงตามรูปแบบที่ 5 ซึ่งมีข้อตกลงที่มีข้อจำกัดน้อยกว่าวิธีของทักเกอร์และเลวิน ดังนั้นในทางปฏิบัติควรใช้วิธีการตามรูปแบบที่ 5 วิธีการใดวิธีหนึ่งใน 3 วิธี

โคเลน และวิทนีย์ (Kolen and Whitney 1982 : 279-293) ได้เปรียบเทียบความเพียงพอของวิธีการเทียบมาตรา 4 วิธี คือ วิธีอิควิเปอร์เซนไทล์ วิธีเชิงเส้นตรง วิธี IRT ชนิดหนึ่งพารามิเตอร์โลจิสติคโมเดล หรือราสช์โมเดล และวิธี IRT ชนิดสามพารามิเตอร์ โดยใช้แบบสอบ GED (Tests of General Educational Development)ซึ่งเป็นแบบสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียนจำนวน 5 วิชา แต่ละวิชามีเนื้อหาแตกต่างกัน ดังนี้ ทักษะการเขียน(80 ข้อ) การศึกษาสังคม(60 ข้อ) วิทยาศาสตร์(60 ข้อ) ทักษะการอ่าน(40 ข้อ) และคณิตศาสตร์(50 ข้อ) ซึ่งใช้สำหรับตัดสินให้ประกาศนียบัตรแก่นักเรียนระดับอุดมศึกษาทั่วประเทศที่มีความรู้เทียบเท่าระดับที่แสดงว่า ได้เรียนสำเร็จแล้ว แบบสอบแต่ละวิชามี 12 ชุด ดำเนินการสอบผู้สอบกลุ่มละ 2 ชุด และทุกกลุ่มต้องสอบแบบสอบชุดที่ 12 ซึ่งแยกออกมาเป็นแบบสอบร่วม การจัดลำดับแบบสอบให้กับแต่ละกลุ่มใช้วิธีสมดุล (couterbalance) ทำการเทียบมาตราแต่ละคู่ของแบบสอบในแนวนอน ด้วยวิธีการเทียบ 4 วิธีดังกล่าว

กลุ่มตัวอย่างสุ่มจากการสอบในปี ค.ศ.1980 ซึ่งมีจำนวนมากกว่า 800,000 คน โดยสุ่มมากลุ่มละประมาณ 200 คน คิดเป็นสัดส่วนโรงเรียนละ 22 คน ด้วยวิธีสุ่มแบบแบ่งชั้น (statified random sampling) กลุ่มตรวจสอบผลก็ใช้วิธีสุ่มเช่นเดียวกันโดยสุ่มมาจากผู้สอบแต่ละวิชาและแต่ละชุดให้ได้ผู้สอบ 20 คน (ประมาณร้อยละ 10 ของผู้เข้าสอบทั้งหมด) การเทียบมาตราเชิงเส้นตรงและอิควิเปอร์เซนไทล์ กระทำตามวิธีของแองกอฟฟ์ ส่วนวิธีการ IRT ประมาณค่าพารามิเตอร์โดยใช้โปรแกรม LOGIST เฉพาะแบบสอบร่วมก่อน จากนั้นใช้ค่าพารามิเตอร์ความสามารถของผู้สอบ ที่กำหนดให้คงที่ในการประมาณค่าพารามิเตอร์จากแบบสอบอีก 11 ชุด แล้วสร้างตารางคะแนนสมมูลระหว่างแบบสอบทั้งหมดโดยใช้การประมาณค่าคะแนนจริง ส่วนคะแนนซึ่งอยู่ต่ำกว่าระดับการเดาตามวิธีสามพารามิเตอร์นั้นอาศัยวิธีการเชิงเส้นตรงมาประมาณค่าเพิ่มเติม

เกณฑ์ในการตรวจสอบความเพียงพอของวิธีการเทียบมาตรา ใช้ค่าสถิติจากกลุ่มตรวจสอบผล ซึ่งผู้สอบแต่ละคนจะต้องสอบแบบสอบทั้งสองชุดที่นำมาเทียบ ดัชนีนี้คือดัชนีเปรียบเทียบเปอร์เซนไทล์ ซึ่งเป็นค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างคะแนน X ที่ได้จากแบบสอบร่วมกับคะแนน Y ที่แปลงไปสู่สเกลของคะแนน X โดยใช้วิธีเทียบแต่ละวิธี และทดสอบความมีนัยสำคัญของค่าอันดับจากการแปลงคะแนนตามปริมาณของค่าดัชนีด้วยการทดสอบฟรายด์แมน

ผลการวิจัยพบว่า วิธีการเทียบมาตราแบบอิควิเปอร์เซนไทล์ให้ผลการเทียบที่คงเส้นคงวาน้อยกว่าการเทียบมาตราเชิงเส้นตรง ส่วนการเทียบแบบ IRT ชนิดหนึ่งพารามิเตอร์หรือราสช์โมเดลให้ผลที่คงเส้นคงวาเหมือนกับวิธีเทียบเชิงเส้นตรง ขณะที่วิธีการเทียบแบบ IRT ชนิดสามพารามิเตอร์และอิควิเปอร์เซนไทล์ให้ผลน้อยกว่า ดังนั้นการวิจัยนี้จึงให้ผลการเปรียบเทียบแตกต่างจากงานวิจัยอื่น ได้แก่ การศึกษาของโคเลน (Kolen 1981: 1-11) ที่พบว่าวิธีการเทียบมาตราแบบอิควิเปอร์เซนไทล์ให้ผลมากที่สุด เมื่อเปรียบเทียบกับวิธีเชิงเส้นตรงและวิธี IRT ชนิดหนึ่งพารามิเตอร์ ทั้งนี้การศึกษาของโคเลนใช้กลุ่มตัวอย่างขนาดใหญ่ ซึ่งแตกต่างจากงานวิจัยครั้งนี้ที่ใช้กลุ่มตัวอย่างขนาดเล็ก นอกจากนี้แบบสอบที่ใช้เทียบในงานวิจัยนี้มีความยากใกล้เคียงกันน้อยกว่า ส่วนผลการเปรียบเทียบของคนอื่น ๆ (Kolen and Whitney 1982 : 291 citing Marco, Petersen and Stewart 1979 ; Petersen , Cook and Stocking 1981) ซึ่งต่างพบว่า เมื่อแบบสอบที่เทียบโดยใช้แบบสอบร่วมกระทำกับกลุ่มที่ไม่ได้มาจากการสุ่มซึ่งมีขนาดใหญ่นั้น วิธีการIRT ชนิดสามพารามิเตอร์ จะให้ผลการเทียบที่มีความเพียงพอมากที่สุด อย่างไรก็ตาม วิธีเชิงเส้นตรงเพียงพอมากที่สุดหากแบบสอบที่เทียบมีความคล้ายคลึงกันมาก ผลการศึกษาที่แตกต่างเหล่านี้โคเลนและวิทนีย์ได้สรุปว่า วิธีการเทียบที่มีความเพียงพอขึ้นกับองค์ประกอบหลายประการ ได้แก่ คุณลักษณะต่าง ๆ ของแบบสอบวิธีที่ใช้ในการเทียบและลักษณะของกลุ่มตัวอย่างที่ใช้เทียบ เป็นต้น

เยาวดี รางชัยกุล (RANGCHAIKUL,1975) ได้ทำการเทียบมาตราผลการสอบในประเทศไทยที่กระทรวงศึกษาธิการได้จัดดำเนินการสอบให้กับนักเรียนในระบบโรงเรียน เมื่อเดือนมีนาคม พ.ศ.2516 และ พ.ศ.2517 และการสอบที่จัดให้กับบุคคลทั่วไปที่อยู่นอกระบบโรงเรียนเมื่อเดือนสิงหาคม พ.ศ.2516 และ พ.ศ.2517 โดยมีจุดมุ่งหมายเพื่อสำรวจขอบเขตแบบสอบทั้ง 4 ครั้ง ซึ่งใช้แบบสอบต่างชุดกันนี้ว่ามีความเท่าเทียมกันเพียงใดที่จะนำมาใช้ตัดสินว่านักเรียนสมควรได้ผ่านชั้นมัธยมศึกษาปีที่ 5 หรือไม่และเพื่อเสนอแนะวิธีที่ดีที่สุดสำหรับการเทียบมาตราแบบสอบต่างชุด กลุ่มตัวอย่างที่ใช้คือ คะแนนผลการสอบของนักเรียนแต่ละครั้ง ครั้งละ 500 คน เทียบมาตราในวิชาภาษาไทย ภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์ โดยใช้วิธีเทียบมาตราเชิงเส้นตรง วิธีที่ 1 ที่เสนอโดยแองกอฟฟ์ (Angoff) ทำการเทียบมาตราแบบสอบแต่ละชุดไปสู่สเกลร่วม (common scale) ที่มีช่วงคะแนน 1-175 ค่าเฉลี่ย 100 ค่าเบี่ยงเบนมาตรฐาน 25

ผลการวิจัยพบว่า คะแนนจากแบบสอบทั้ง 4 ชุด ไม่เท่าเทียมกันทุกวิชา กล่าวคือ จุดตัด (cut-off score) ของแบบสอบแต่ละชุดไม่สามารถเปรียบเทียบกันได้ เช่น การสอบเมื่อเดือนมีนาคม พ.ศ.2517 พบว่าคะแนนจุดตัดสำหรับการผ่านในสเกลร่วมเท่ากับ 88 แต่การสอบเมื่อเดือนมีนาคม พ.ศ.2516 มีคะแนนจุดตัดสำหรับการผ่านในสเกลร่วมเท่ากับ 78 และภายหลังการตรวจสอบสัดส่วนการผ่านของผู้สอบแล้ว พบว่า นักเรียนที่ควรสอบได้แต่ได้รับการตัดสินให้ตก ซึ่งผู้วิจัยได้เสนอแนะวิธีการซึ่งควรใช้ในการเทียบมาตราในสถานการณ์การสอบนี้ว่าควรใช้รูปแบบที่ 4 ตามที่แองกอฟฟ์ได้เสนอไว้ นั่นคือ รูปแบบที่มีข้อสอบร่วม

ชูชีพ พงษ์สมบูรณ์ (2528) ได้ทำการเปรียบเทียบประสิทธิภาพและความคงที่ของการเทียบมาตรา 3 วิธี คือ การเทียบเชิงเส้นตรง การเทียบแบบอิควิเปอร์เซนไทล์ การเทียบโดยใช้โค้งลักษณะของข้อสอบ จากรูปแบบที่ใช้แบบทดสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วม กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศ สุ่มมาจำนวน 3,721 คน ที่เข้าสอบโครงการตรวจสอบคุณภาพของกรมวิชาการ กระทรวงศึกษาธิการ ซึ่งแบบทดสอบในโครงการมีจำนวน 60 ข้อ นำมาแบ่งเป็น 2 ฉบับ ฉบับละ 38 ข้อ มีแบบทดสอบร่วมภายใน จำนวน 20 ข้อ ผลการเปรียบเทียบได้ว่าประสิทธิภาพของการเทียบมาตราระหว่างรูปแบบทดสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วมในแต่ละวิธีไม่แตกต่างกัน ความคงที่ของวิธีการเทียบมาตราโดยใช้รูปแบบที่ใช้ผู้สอบร่วมได้ผลว่า วิธีการเทียบโดยใช้วิธีอิควิเปอร์เซนไทล์มีความคงที่มากกว่าวิธีเชิงเส้นตรง แต่มีความคงที่พอ ๆ กับวิธีใช้โค้งลักษณะข้อสอบ ความคงที่ของวิธีการเทียบมาตราในรูปแบบที่ใช้แบบทดสอบร่วมได้ผลเหมือนกับรูปแบบใช้ผู้สอบร่วม

ภาวิณี ศรีสุขวัฒนานันท์ (2529) ได้เปรียบเทียบผลของการใช้วิธีการเทียบมาตรา 3 วิธี คือวิธีอิควิเปอร์เซนไทล์ วิธีเชิงเส้นตรง วิธีอิงทฤษฎีการตอบข้อสอบแบบสามพารามิเตอร์ โดยใช้แบบสอบร่วมภายในที่ต่างกัน 3 ขนาด คือ ร้อยละ 20 (7 ข้อ) ร้อยละ 40 (14 ข้อ) ร้อยละ 60 (21 ข้อ)โดยใช้ข้อกระทงแต่ละฉบับ 35 ข้อ(ไม่ร่วมกัน) การประเมินผลการเทียบมาตรามี 2 ลักษณะ

1) ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (SEE) ด้วยวิธีที่ต่างกันของแบบสอบแต่ละกรณี และ

2) ตรวจสอบความเพียงพอของวิธีการเทียบมาตราแต่ละวิธี โดยใช้ดัชนีเปรียบเทียบความแตกต่าง (Index C) ที่ได้วิเคราะห์จากกลุ่มตรวจทานผล

กลุ่มตัวอย่างแยกเป็น 2 กรณี คือ กรณีแบบสอบคัดเลือก 2 กลุ่ม ๆ ละ 1,500 คน และกลุ่มตรวจสอบผลอีก 1 กลุ่ม จำนวน 1,500 คน อีกกรณีหนึ่งคือ กรณีแบบสอบวัดผลสัมฤทธิ์ ซึ่งมีกลุ่มตัวอย่างเช่นเดียวกับกรณีแรก

ผลการวิจัยพบว่า การใช้แบบสอบร่วมที่ยาวกว่าให้ประสิทธิภาพของการเทียบมาตราสูงกว่า (คลาดเคลื่อนน้อยกว่า)และให้ผลในระดับที่น่าพอใจมากกว่า ทั้งสองกรณีแบบสอบ การศึกษาวิธีของการเทียบมาตรา พบว่าในกรณีแบบสอบคัดเลือกนั้นวิธีที่ให้ความเพียงพอมากที่สุด คือ วิธีอิควิเปอร์เซนไทล์ รองลงมาคือ วิธีทฤษฎีการตอบข้อสอบ ซึ่งให้ผลที่แตกต่างกับในกรณีแบบสอบวัดผลสัมฤทธิ์ ที่พบว่า วิธีที่ให้ความเพียงพอมากที่สุด คือ วิธีเชิงเส้นตรง รองลงมาคือ วิธีอิควิเปอร์เซนไทล์และเมื่อทดสอบความแตกต่างของค่าดัชนี C เป็นรายคู่ พบว่า กรณีแบบสอบวัดผลสัมฤทธิ์ วิธีเชิงเส้นตรงมีความเพียงพอมากกว่าวิธีที่ใช้วิธีทฤษฎีการตอบข้อสอบอย่างมีนัยสำคัญทางสถิติที่ .05 ส่วนการเปรียบเทียบวิธีคู่อื่นไม่สามารถสรุปความแตกต่างอย่างมีนัยสำคัญ ซึ่งผู้วิจัยอภิปรายผลว่า วิธีการเทียบมาตราที่ให้ความเพียงพอต่างกันในกรณีแบบสอบคัดเลือกและแบบสอบวัดผลสัมฤทธิ์ อาจเนื่องมาจากปัจจัยด้านแบบทดสอบและปัจจัยด้านประชากรผู้สอบ

เรวดี อินทะสะระ(2530)ได้เปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบมาตราและความเพียงพอของวิธีการเทียบมาตรา 2 วิธีคือ วิธีอิงทฤษฎีการตอบข้อสอบสามพารามิเตอร์กับวิธีการวิเคราะห์องค์ประกอบ กลุ่มตัวอย่างคือนักเรียนมัธยมศึกษาปีที่ 1 ปีการศึกษา 2529 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดพัทลุง โดยนำแบบทดสอบ 2 ชุด ๆละ 30 ข้อ มีแบบสอบร่วมอีก 15 ข้อ นำไปดำเนินการสอบกับกลุ่มตัวอย่างจำนวน 1,557 คน โดยทุกคนต้องทำแบบสอบทั้ง 2 ชุด หลังจากนั้นจัดสุ่มใช้คะแนนผลการสอบจากแบบสอบชุดที่ 1 จำนวน 779 คนและชุดที่ 2 จำนวน 778 คน ส่วนจำนวนที่เหลือใช้เป็นคะแนนตรวจสอบผลของแต่ละวิธี

ผลการวิจัยพบว่าการเทียบมาตราวิธีอิงทฤษฎีการตอบข้อสอบให้ความเพียงพอในระดับที่น่าพอใจ ส่วนวิธีการใช้เทคนิคการวิเคราะห์องค์ประกอบให้ความเพียงพอในระดับปานกลาง สำหรับการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา พบว่า เทคนิคการวิเคราะห์องค์ประกอบมีความคลาดเคลื่อนสูงกว่าอย่างมีนัยสำคัญทางสถิติที่ .01 แต่เมื่อพิจารณาในแต่ละช่วงคะแนน พบว่า วิธีทั้งสองให้คะแนนสมมูลใกล้เคียงกันในช่วงตรงกลางของการแจกแจงคะแนน (19-23) ส่วนช่วงคะแนนต่ำและสูง พบว่าวิธีทั้งสองแตกต่างกันในทางตรงข้าม กล่าวคือ ในช่วงคะแนนต่ำ (1-18) วิธีอิงทฤษฎีการตอบข้อสอบมีคะแนนสมมูลสูงกว่า ส่วนช่วงคะแนน 24-25 วิธีวิเคราะห์องค์ประกอบมีคะแนนสมมูลสูงกว่า

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังกล่าวมา พอสรุปได้ดังนี้

1. ผลการศึกษาที่ผ่านมา พบว่า ไม่มีวิธีการเทียบมาตราวิธีใดเพียงวิธีเดียวที่ให้ผลดีที่สุดในทุกสถานการณ์ ดังนั้นในทางปฏิบัติจำเป็นต้องศึกษาข้อกำหนดต่าง ๆ ของวิธีการเทียบมาตราแต่ละวิธี ได้แก่ ลักษณะการแจกแจงของคะแนนที่ได้จากแบบสอบต่างชุด ลักษณะของแบบสอบ และลักษณะของกลุ่มตัวอย่าง เป็นต้น ทั้งนี้เพื่อเป็นแนวทางในการตัดสินใจเลือกวิธีการในการเทียบมาตราที่เหมาะสมในแต่ละสถานการณ์การเทียบ สรุปได้ดังนี้

1.1 เมื่อแบบสอบต่างชุดที่นำมาเทียบมาตรามีระดับความยากคล้ายคลึงกัน และกลุ่มตัวอย่างเป็นกลุ่มที่เท่าเทียมกัน ควรใช้วิธีการเชิงเส้นตรง ซึ่งสะดวกในการแปลงคะแนนและมีความเป็นปรนัยมากกว่า แต่ผลการเทียบมาตราขึ้นอยู่กับกลุ่มตัวอย่างที่ใช้

1.2 เมื่อแบบสอบมีความยากต่างกัน กลุ่มตัวอย่างเป็นกลุ่มที่เท่าเทียมกันแล้ว วิธีการเทียบมาตราแบบอิควิเปอร์เซนไทล์จะเหมาะสมกว่า แต่ถ้ากลุ่มตัวอย่างเป็นกลุ่มที่ไม่เท่าเทียมกัน รูปร่างการแจกแจงคะแนนที่ได้จากแบบสอบต่างชุดจะแตกต่างกันมากและมีความคลาดเคลื่อนมาก

1.3 เมื่อใช้แบบสอบที่มีความยากต่างกันและกลุ่มตัวอย่างไม่ได้เป็นกลุ่มที่เท่าเทียมกัน ควรใช้วิธีการแบบ IRT ชนิดสามพารามิเตอร์ ซึ่งให้ผลที่น่าพอใจกว่าวิธีอื่น

1.4 เมื่อกลุ่มตัวอย่างค่อนข้างน้อย และแบบสอบวัดในสิ่งเดียวกันแล้ว วิธีเชิงเส้นตรงและวิธีการ IRT ชนิดหนึ่งพารามิเตอร์ ให้ผลที่น่าพอใจกว่าวิธีอิควิเปอร์เซนไทล์และ IRT ชนิดสามพารามิเตอร์