บทที่ 2

วรรณคดีที่เกี่ยวข้อง

ในการนำเสนอวรรณคดีที่เกี่ยวข้องกับการวิจัยนี้ ผู้วิจัยได้นำเสนอโดยจำแนกออกเป็นตอน ๆ ดังนี้

1. ความหมายของการเทียบมาตรา
2. รูปแบบของการเทียบมาตรา
3. รูปแบบของการเก็บรวบรวมข้อมูล
4. การเทียบค่าความสามารถของผู้สอบ
5. ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา
6. ความเพียงพอของการเทียบมาตรา
7. งานวิจัยที่เกี่ยวข้องกับการเทียบมาตรา

ความหมายของการเทียบมาตรา

การเทียบมาตรา (Equating) เป็นการศึกษาเชิงประจักษ์ โดยใช้วิธีการทางการวัดผลที่มีผู้คิดค้นเพื่อทำให้คะแนนจากแบบสอบต่างฉบับเทียบกันได้ มีผู้ให้นิยามของการเทียบมาตราไว้หลายท่าน ดังนี้

กัลลิคเซน (Gulliksen,1950) กล่าวถึงการเทียบมาตราว่า เป็นวิธีการทำคะแนนที่ได้จากแบบสอบสองฉบับที่วัดวิชาเดียวกัน ให้เป็นคะแนนสมมูล (Equivalent Score) ที่เทียบกันได้โดยตรง โดยเสนอวิธีการให้ผู้สอบกลุ่มเดียวทำแบบสอบสองชุด และใช้วิธีง่าย ๆ คือ แปลงคะแนนแต่ละชุดให้เป็นคะแนนมาตรฐาน แล้วนำคะแนนที่แปลงแล้ว

มาเทียบกันโดยตรง แต่ก่อนอื่นให้ตรวจสอบความเป็นคู่ขนานของแบบสอบโดยใช้สถิติของวิลค์ (Wilks) หรือพิจารณาจากค่าสหสัมพันธ์ระหว่างแบบสอบแต่ละชุดกับเกณฑ์ ถ้ามีค่าเท่ากันโดยประมาณ ($R_{xc} = R_{yc}$) ก็จะบอกความสามารถของความเพียงพอในการเทียบคะแนนแบบสอบทั้งสองชุด

ฟลานาแกน (Flanagan, 1951:747-748) ให้ความหมายของการเทียบมาตราว่า เป็นวิธีการทำคะแนนจากแบบสอบต่างฉบับกัน ให้สามารถนำมาเปรียบเทียบกันได้ คำว่า "ความสามารถในการเปรียบเทียบกันได้" มีความหมายเฉพาะที่ว่า เมื่อกำหนดประชากรให้ถ้าการแจกแจงของคะแนนจริงจากแบบสอบทั้งสองชุด ซึ่งสอบกับกลุ่มตัวอย่างที่เลือกมาขนาดใหญ่ใด ๆ มีลักษณะเหมือนกันแล้ว คะแนนดิบจากแบบสอบทั้งสองชุดจึงจะสามารถเปรียบเทียบกันได้ หรือถ้าความเชื่อมั่น (Reliability) ของแบบสอบทั้งสองชุดเท่ากันในประชากรนั้นแล้ว ก็สามารถเปรียบเทียบการแจกแจงของค่าที่ได้เช่นกัน จากความหมายดังกล่าวเป็นนิยามเชิงทฤษฎี ในทางปฏิบัติได้นิยามไว้ว่าในประชากรที่กำหนด ถ้าคะแนนจากแบบสอบสองชุดมีค่าเฉลี่ยเท่ากัน หรือเกือบเท่ากันในทุก ๆ กลุ่มตัวอย่างขนาดใหญ่ใด ๆ แล้ว การเปรียบเทียบจะทำได้ อย่างไรก็ตาม จะทำให้เกิดลักษณะตามที่นิยามเป็นเรื่องที่ยากมาก สิ่งที่สำคัญคือแบบสอบต้องเป็นคู่ขนานกัน ฟลานาแกนจึงแนะนำว่า ควรเริ่มต้นตั้งแต่การสร้างแบบสอบให้มีคุณสมบัติให้เป็นคู่ขนานกัน แล้วเลือกวิธีเทียบคะแนนที่เหมาะสม ซึ่งเสนอไว้ 4 วิธี คือ

1. ใช้ค่าเฉลี่ย โดยคำนวณค่าเฉลี่ยของการแจกแจงคะแนนทั้งสองชุด ถ้าความแตกต่างของค่าเฉลี่ยอยู่ภายในขอบเขตของการแปรผันเชิงสุ่มแล้ว ถือว่าคะแนนทั้งสองชุดนั้นเปรียบเทียบกันได้ แต่ถ้าความแตกต่างมีนัยสำคัญ ให้ใช้วิธีบวกเข้าหรือลบออกเท่าจำนวนที่แตกต่างจากคะแนนชุดที่หนึ่ง เพื่อให้เกิดคะแนนสมมูลกับอีกชุดหนึ่ง

2. ใช้เทคนิคของสมการถดถอย โดยหาค่าประมาณที่ดีที่สุดของคะแนนจากแบบสอบชุดที่หนึ่งซึ่งรู้ค่าของอีกชุดหนึ่ง

3. ใช้คะแนนมาตรฐาน ซึ่งเป็นวิธีปรับคะแนนอย่างคงที่ตลอดการแจกแจง

4. ใช้วิธีอิควิเปอร์เซนไทล์ ซึ่งเป็นวิธีหาค่าคะแนนโดยการเปรียบเทียบตามสัดส่วนของการแจกแจงคะแนน

มาร์โค (Marco, 1981) ได้อธิบายความหมายของการเทียบมาตราว่า เป็นกระบวนการแปลงคะแนน ที่ได้จากแบบสอบฟอร์มหนึ่ง ให้มีค่าตัวเลขสมมูล (Equivalent) กับตัวเลขที่ได้จากแบบสอบอีกฟอร์มหนึ่ง

แองกอฟ (Angoff, 1971) ให้ความหมายของการเทียบมาตราว่า เป็นการแปลงระบบของหน่วยการวัดของแบบสอบฉบับหนึ่ง ไปสู่ระบบหน่วยการวัดของแบบสอบอีกฉบับหนึ่ง คะแนนที่ผ่านการแปลงแล้วจะให้ความหมายของการเทียบกันโดยตรง แองกอฟได้เสนอวิธีการในการเทียบมาตราไว้ 2 รูปแบบ คือ การเทียบมาตราแบบอิควิเปอร์เซนไทล์ และการเทียบมาตราเชิงเส้นตรง

ลอร์ด (Lord, 1980) ให้ความหมายของการเทียบมาตราว่าเป็นการแปลงคะแนนจากแบบสอบต่างฟอร์ม ให้มีความหมายให้สับเปลี่ยนกันได้ และเพื่อความเสมอภาคของผู้ที่รับการสอบ

สงบ ลักษณะ (2522) กล่าวว่า คะแนนจากแบบสอบสองฉบับ วัดสิ่งเดียวกัน แต่ไม่จำเป็นต้องเป็นแบบสอบคู่ขนาน จะถือว่าเทียบกันได้ถ้าคะแนนจากแบบสอบทั้งสองมาจากคะแนนจริง(True Score) หรือความสามารถแท้ (True Ability) ที่เท่ากัน

ชูศักดิ์ ขัมภลิขิต (2527:2) ได้สรุปความหมายของการเทียบมาตราว่า การเทียบมาตราเป็นกิจกรรมที่เกี่ยวโยงกับกิจกรรมสองประการ คือ

1. กระบวนการที่ทำให้แบบสอบ 2 ฉบับใด ๆ มีความทัดเทียมกันหรือเท่ากันในเชิงโครงสร้าง

2. การใช้วิธีการทางสถิติ เพื่อปรับคะแนนที่ได้จากแบบสอบแต่ละฉบับให้อยู่ในมาตราเดียวกัน และเทียบกันได้

จากความหมายข้างต้นพอสรุปได้ว่า การเทียบมาตราเป็น วิธีการทางการวัดผล เพื่อปรับคะแนนที่ได้จากแบบสอบหนึ่งไปยังอีกแบบสอบหนึ่ง ให้สามารถเปรียบเทียบกันได้ โดยถือเอาความสามารถในการตอบข้อสอบของผู้สอบเป็นเกณฑ์

รูปแบบการเทียบมาตรา

รูปแบบการเทียบมาตรามีผู้เสนอไว้หลายรูปแบบดังนี้

1. การเทียบมาตรารูปแบบอิควิเปอร์เซนไตล์ (Equipercentile Equating) รูปแบบอิควิเปอร์เซนไตล์เริ่มจากการแจกแจงของคะแนนจากแบบสอบ 2 ฉบับ มีลักษณะคล้ายกัน หรือแตกต่างกันบ้างก็เพียงเล็กน้อย การเทียบมาตรา ณ ตำแหน่งเดียวกันของคะแนน 2 ชุดนั้น ผลของการเทียบมาตราแสดงด้วยกราฟ โดยปกติการเทียบมาตรารูปแบบนี้ให้ภาพการแปลงคะแนนที่สะท้อนถึงระดับความยากง่ายของแบบสอบ 2 ฉบับ ถ้าแบบสอบมีความยากใกล้เคียงกัน เส้นกราฟจะมีลักษณะใกล้เคียงเส้นตรง แต่ถ้ามีค่าความยากแตกต่างกันเส้นกราฟจะเป็นเส้นโค้ง คะแนนสมมูลที่เกิดขึ้นจะถูกยืดหยุ่นตามคะแนนดิบ เพื่อให้คงรักษาให้เหมือนชุดก่อนตามต้องการ (Angoff, 1984)

บรูนและฮอลแลนด์ (Braun and Holland, 1982:19-22 อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์) ได้แสดงถึงความสัมพันธ์ของแบบสอบร่วมในการเทียบมาตราตามรูปแบบอิควิเปอร์เซนไตล์ ไว้ดังนี้ ให้กลุ่มตัวอย่าง 2 กลุ่ม คือ P และ Q ที่สุ่มมาอย่างอิสระจากประชากรทั้งหมด (T) เขียนเป็นสมการได้ว่า

$$T = fP + (1-f)Q$$

โดย P และ Q ขึ้นกับน้ำหนักของ f และ (1-f) ตามลำดับ การคำนวณการแจกแจงของตัวอย่างรวม ทำได้โดยการคำนวณการแจกแจงตามเงื่อนไขของกลุ่มตัวอย่างกลุ่มที่ 1 (P) และกลุ่มที่ 2 (Q) ก่อน แล้วหาค่าเฉลี่ยที่ถ่วงน้ำหนักด้วย f และ (1-f)ซึ่งได้จากสมการ คือ

$$f = N_1/(N_1+N_2)$$

เมื่อ N_1 และ N_2 คือ ขนาดของกลุ่มตัวอย่างที่ 1 และที่ 2 ตามลำดับ ค่าของ f คือค่าเฉลี่ยของกลุ่มตัวอย่างที่ 1 และกลุ่มตัวอย่างที่ 2

ในการเทียบมาตราด้วยรูปแบบอิควิเปอร์เซนไทล์ มีข้อตกลงเพื่อหาค่าประมาณการแจกแจงของแบบสอบฉบับที่ 1 ($F_T(X)$) และการแจกแจงของแบบสอบฉบับที่ 2 (G_TX)) ในประชากรทั้งหมด (T) ซึ่ง $F_T(X)$ และ $G_T(X)$ หาได้ในเทอมของปริมาณจากข้อมูลที่รวบรวมได้ ฟังก์ชั่นของการเทียบคะแนนเป็นดังนี้

$$e_x(Y) = F_T^{-1}(G_T(Y))$$

วิธีการเทียบมาตราแบบอิควิเปอร์เซนไทล์นี้ มีข้อจำกัดที่ต้องคำนึงถึงอยู่หลายประการ คือ

1) วิธีนี้ใช้การเขียนเส้นกราฟเพื่อหาคะแนนสมมูลซึ่งจะต้องปรับหรือเกลาเส้นโค้งให้เรียบ ดังนั้นประสบการณ์และการตัดสินใจเกลาเส้นให้เรียบด้วยข้อมูลเหล่านั้น อาจนำไปสู่ความคลาดเคลื่อนอย่างมาก การเทียบมาตราด้วยวิธีนี้จึงเป็นการเทียบไปยังจำนวนเต็มที่ใกล้ที่สุด (Angoff,1971)

2) รูปแบบอิควิเปอร์เซนไทล์ มีความไวต่อความแปรปรวนเชิงสุ่มมาก โดยเฉพาะกลุ่มตัวอย่างที่มีขนาดเล็ก (Angoff,1971;Potthoff,1982)

3) ในกรณีที่แบบสองสองชุดมีความเชื่อมั่นต่างกันมาก ผลของการเทียบมาตราจะขาดความคงที่ (Potthoff,1982)

4) การเทียบคะแนนสมมูลของแบบสอบสองชุด ทำได้เฉพาะในช่วงพิสัยของคะแนนที่มีความถี่ของคะแนนสังเกตเพียงพอ ส่วนช่วงที่มีความถี่ของคะแนนน้อยจะมีความคลาดเคลื่อนสูงมาก

2. <u>การเทียบมาตรารูปแบบเชิงเส้นตรง</u> (Linear Equating)

วิธีนี้เป็นวิธีที่ง่ายในขั้นตอนของการแปลงคะแนน โดยมีข้อตกลงว่า แบบสอบสองชุดที่คู่ขนานกันจะมีการแจกแจงคะแนนดิบเป็นอย่างเดียวกัน ยกเว้นสำหรับความแตกต่างในค่าเฉลี่ยและความแปรปรวน นั่นคือ นอกจากความแตกต่างในสองโมเมนต์แรกแล้วโมเมนต์มาตรฐานของการแจกแจงคะแนนดิบของแบบสอบสองชุดของกลุ่มผู้สอบที่กำหนดให้ จะต้องเหมือนกันหมด (Angoff, 1971)

จากจุดล่นของการเทียบมาตราโดยใช้ลิควิเปอร์เซนไทล์ คือ การประมาณค่า $F(x)$ และ $G(y)$ ฉะนั้นโมเดลที่ง่ายและลาศัยการประมาณค่าพารามิเตอร์ที่เป็นไปได้ในแง่การปฏิบัติ คือ โมเดลเชิงเส้นตรง (Linear Model)

$$e_x(y) = ay + b \qquad (1)$$

จากโมเดลของฟังก์ชั่นการเทียบเชิงเส้นนี้ มีพารามิเตอร์ต้องประมาณค่า 2 ตัว คือ ค่าความชัน (a) และ ค่าคงที่ (b)

ในกรณีของสมการที่ (1) ซึ่งตรงกับฟังก์ชั่นการเทียบโดยใช้เปอร์เซนไทล์ ถ้า

$$F^{-1}(G(y)) = ay+b \qquad (2)$$

หรือเช่นเดียวกันถ้า

$$G(y) = F(ay+b) \qquad (3)$$

สมการที่ (3) ตรงกับข้อตกลงที่ให้ F และ G อยู่ในรูปของ

$$F(X) = H((x- \mu_x)/ \sigma_x) \quad (4)$$

และ

$$G(Y) = H((Y- \mu_y)/ \sigma_y) \quad (5)$$

สำหรับฟังก์ชั่นการกระจายบางอย่าง H ในสมการ (4) และ (5) μ_x, μ_y, σ_x, σ_y คือค่าเฉลี่ยและความแปรปรวนของแบบสอบฟอร์ม X และฟอร์ม Y ถ้าประยุกต์สมการ (4) และ (5) กับสมการ (2) เราจะได้รูปฟอร์มของ a และ b ในสมการ (2) คือ

$$a = \sigma_x/\sigma_y \quad (6)$$

$$b = \mu_x - (\sigma_x/\sigma_y)\mu_y \quad (7)$$

นำไปใช้กับสมการ (1) เราจะได้ฟังก์ชั่นการเทียบเชิงเส้นตรง คือ

$$e(y) = \mu_x + (\sigma_x/\sigma_y)(Y - \mu_y)$$

3. <u>การเทียบมาตรารูปแบบลิงทฤษฎีตอบสนองข้อสอบ</u>

การเทียบมาตรารูปแบบลิงทฤษฎีตอบสนองข้อสอบ ลอร์ด (Lord, 1980) ได้กล่าวว่า แบบสอบสองฉบับใด ๆ ที่เทียบคะแนนกันต้องเป็นแบบสอบที่มีมิติเดียวกัน และเทียบคะแนนจากความสามารถ (θ) ที่เท่ากัน ซึ่งจะมีข้อกำหนดที่สำคัญ 3 ประการ ดังนี้

1) ความเสมอภาค (Equity) คือ ถ้าพิจารณาที่ระดับความสามารถ (θ) ใด ๆ การแจกแจงความถี่อย่างมีเงื่อนไขของคะแนนแปลง $x(y)$ หรือ y^{*} (คะแนนจากแบบสอบ X ที่แปลงมาอยู่สเกลเดียวกับแบบสอบ Y) ที่ θ ที่กำหนด ให้ด้วยเทคนิคการเทียบคะแนน ต้องเหมือนกันกับการแจกแจงความถี่อย่างมีเงื่อนไขของคะแนนจากแบบสอบที่ต้องการเทียบ (x)

2) ความไม่ผันแปรตามกลุ่ม (Invariance Across Groups) คือ คะแนนแปลง $x(y)$จะเหมือนกันโดยไม่ขึ้นกับตัวแปรลึก ๆ ของประชากรที่นำมาสร้างสมการเทียบมาตรา

3) ความสมมาตร (Symmetry) คือ คะแนนเทียบมาตราจะต้องเหมือนกัน ไม่ว่าการเทียบนั้นจะเทียบจากแบบสอบชุดที่ 1ไปชุดที่ 2 หรือจากแบบสอบชุดที่ 2 ไปชุดที่ 1

การเทียบมาตราโดยใช้ทฤษฎีตอบสนองข้อสอบ มีการเทียบมาตราโดยใช้คะแนนจริงและการเทียบมาตราโดยใช้คะแนนสังเกต ทั้งสองวิธีนี้มีทั้งข้อได้เปรียบเสียเปรียบกล่าวคือ การเทียบด้วยคะแนนจริงไม่สามารถอธิบายคะแนนที่อยู่ต่ำกว่าระดับการเดาได้ โดยจะให้ความหมายของคะแนนสมมูลเฉพาะคะแนนที่อยู่เหนือค่าเฉลี่ยของการเดา ซึ่งถึงแม้จะเป็นการเทียบโดยใช้คะแนนจริง แต่ยังเป็นค่าที่ประมาณได้จากสูตรในการคำนวณ ดังนั้นจึงยังคงมีความคลาดเคลื่อนอยู่ ส่วนการเทียบโดยใช้คะแนนสังเกตนั้นเป็นการเทียบคะแนนโดยประมาณ ซึ่งอธิบายคะแนนสมมูลจาก X และ Y ได้ครอบคลุมพิสัยของคะแนนที่สังเกตได้ ลอร์ดได้กล่าวว่า ทั้งสองวิธีนี้มีความสอดคล้องกันมาก แต่การสรุปผลในการอ้างอิงต้องทำอย่างรอบคอบ (Lord, 1980) อย่างไรก็ตามการเทียบมาตราโดยใช้คะแนนสังเกตนี้ ยังนำไปใช้ในงานวิจัยน้อยมาก อาจเป็นเพราะว่า วิธีนี้มีความซับซ้อนและการลงทุนแพงกว่าวิธีการเทียบมาตราโดยใช้คะแนนจริง (Lord and Wingersky, 1984)

ลอร์ด (Lord, 1980) ได้แบ่งรูปแบบการเทียบมาตราโดยอิงทฤษฎีตอบสนองข้อสอบออกเป็นหลายวิธีดังนี้

1. <u>วิธีการเทียบมาตราโดยใช้คะแนนจริง</u> (True-Score Equating)

พิจารณาจากความสัมพันธ์เชิงคณิตศาสตร์ ระหว่างความสามารถและจำนวนข้อที่ทำถูก ซึ่งจะได้คะแนนจริงจากแบบสอบ 2 ฉบับ ดังนี้

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

$$\eta = \eta(\theta) = \sum_{j=1}^{n} P_j(\theta_y)$$

เมื่อ ξ แทน คะแนนจริงจากแบบสอบฉบับ X

η แทน คะแนนจริงจากแบบสอบฉบับ Y

m แทน จำนวนข้อของแบบสอบฉบับ X

n แทน จำนวนข้อของแบบสอบฉบับ Y

θ_x แทน ค่าความสามารถจากแบบสอบฉบับ X

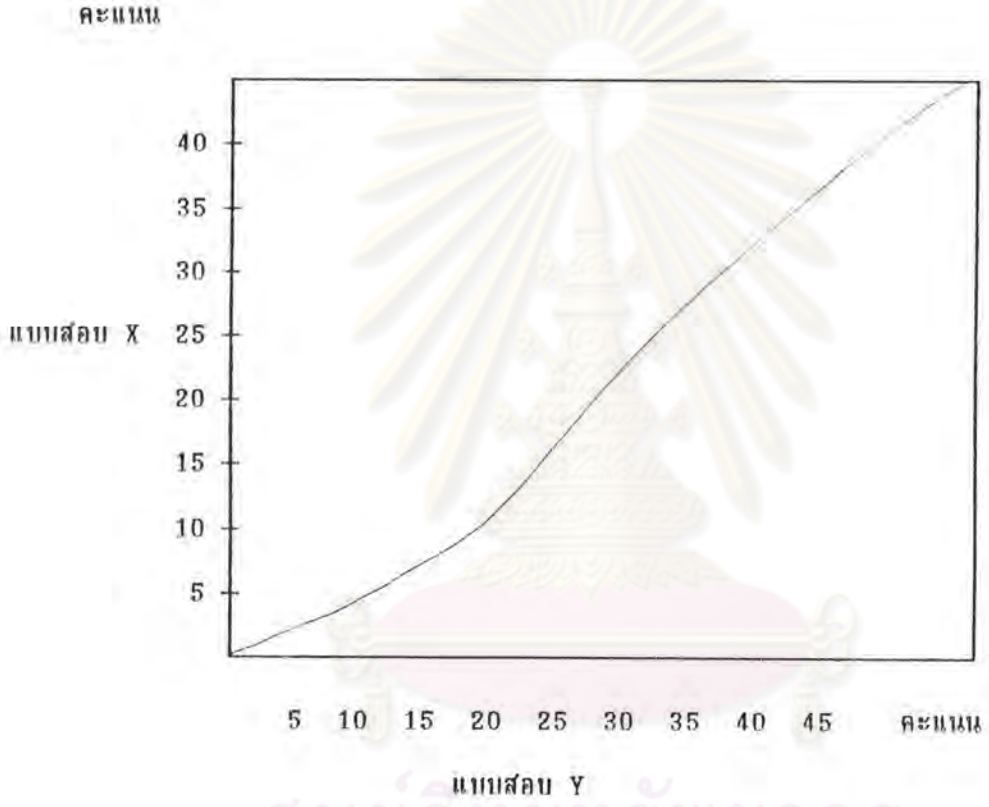
θ_y แทน ค่าความสามารถจากแบบสอบฉบับ Y

ในทางปฏิบัติ เราจะหาค่า $P_i(\theta)$ และ $P_j(\theta)$ ได้จากการประมาณค่าพารามิเตอร์ของข้อกระทงโดยเลือกว่าจะใช้ชนิดหนึ่งพารามิเตอร์, สองพารามิเตอร์ หรือ สามพารามิเตอร์ จากนั้น นำคะแนนจริงของผลการสอบจากแบบสอบฉบับ X และแบบสอบ Y มาหาความสัมพันธ์กัน โดยใช้ค่าความสามารถ (θ) ที่ระดับเดียวกัน

2. <u>วิธีการเทียบมาตราโดยใช้คะแนนจริงด้วยแบบสอบร่วม</u> (True-score Equating with an Anchor Test) การใช้แบบสอบร่วมทำได้ 2 กรณี คือ ใช้แบบสอบร่วมรวมเข้าเป็นชุดเดียวกับแบบสอบที่ต้องการเทียบ เรียกว่า แบบสอบร่วมภายใน (Internal Anchor Test) ส่วนในกรณีที่จัดแยกเป็นชุดแบบสอบต่างหาก จากแบบสอบที่ต้องการเทียบเรียกว่า แบบสอบร่วมภายนอก (External Anchor Test) จุดมุ่งหมาย

ของการใช้แบบสอบร่วม คือ เพื่อใช้แบบสอบร่วมเป็นตัวเชื่อมข้อมูลเข้าด้วยกัน ทำให้ค่าพารามิเตอร์ที่วิเคราะห์ออกมาอยู่บนมาตราเดียวกัน ซึ่งถ้าขาดแบบสอบร่วมแล้ว จะไม่สามารถเทียบคะแนนกันได้นอกเสียจากว่า กลุ่มตัวอย่างที่สอบแบบสอบ X และ Y จะมีการกระจายความสามารถเหมือนกัน

ในทางปฏิบัติจะหาค่าความสามารถ (θ) ของกลุ่มตัวอย่างที่ทำแบบสอบร่วมไปเทียบมาตรากับแบบสอบฉบับ X และแบบสอบฉบับ Y ดังภาพที่ 1



ภาพที่ 1 การเทียบระดับคะแนนโดยใช้คะแนนจริงจากแบบสอบร่วม

3. <u>วิธีการเทียบมาตราโดยใช้คะแนนดิบด้วยแบบสอบร่วม</u> (Raw-Score Equating with an Anchor) ปัญหาของการเทียบมาตราโดยใช้คะแนนจริง คือ ไม่สามารถทราบคะแนนจริงของแต่ละคนได้ นอกจากใช้วิธีประมาณจากผลการสอบโดยสมการ

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

เมื่อ $P_i(\theta)$ เป็นความน่าจะเป็นที่ทำข้อสอบข้อที่ i ถูกที่ระดับความสามารถ θ และ n เป็นจำนวนข้อสอบทั้งหมด ซึ่งค่าที่ได้เป็นค่าประมาณเท่านั้น ยังไม่มีคุณสมบัติเป็นคะแนนจริง ดังนั้นการเทียบมาตราโดยใช้คะแนนจริงจึงคล้ายกับการเทียบมาตราโดยใช้คะแนนดิบชนิดใหม่นั่นเอง สำหรับการเทียบมาตราโดยใช้คะแนนดิบในข้อนี้ได้ศึกษาข้อมูลจากการตอบแบบสอบร่วม (Anchor Test)

การดำเนินการเทียบมาตราโดยวิธีนี้ จะเริ่มด้วยการประมาณลักษณะการแจกแจงของความสามารถของกลุ่มผู้สอบรวม $r(\theta)$ ซึ่งหมายถึง ผู้สอบทั้งหมดที่ทำแบบสอบร่วม การแจกแจง (θ)ในกลุ่มเป็นการประมาณการกระจายของ $r(\theta)$

การประมาณการแจกแจงของคะแนนดิบ X สำหรับกลุ่ม $\phi_x(x)$ ได้จากสมการ

$$\hat{\phi}_x(x) = \frac{1}{N} \Sigma \phi_x(x/\hat{\theta}_a)$$

เมื่อ a = 1,2,3,...,N คือ ผู้สอบแต่ละระดับ และ $\phi_x(x/\hat{\theta}_a)$ เป็นการแจกแจงของคะแนนดิบ x สำหรับผู้ที่มีความสามารถ θ

ถ้า $r(\theta)$ มีลักษณะต่อเนื่อง (Continous) สามารถประมาณด้วยสมการ

$$\hat{\phi}_x(x) = \int_{-\alpha}^{\alpha} \phi_x(x/\theta)\, r(\theta)\, d\theta$$

เมื่อ $\phi_x(x/\hat{\theta}_a)$ ได้มาจากการประมาณค่าพารามิเตอร์โดยใช้ทฤษฎีคลาสสิคของข้อสอบ (Item Parameter) ของข้อสอบในแบบสอบฉบับ X และสมการต่าง ๆ สามารถนำมาประยุกต์ใช้กับแบบสอบฉบับ Y ได้เช่นกัน

เนื่องจากแบบสอบฉบับ X และฉบับ Y เป็นอิสระจากกัน เมื่อความสามารถ (θ) คงที่การเทียบมาตราต้องคำนวณการกระจายร่วม (Joint Distribution) จากสมการ

$$\hat{\phi}(x,y) = \frac{1}{N} \Sigma \phi_x(x/\theta_a)\phi_y(y/\theta_a)$$

$$\hat{\phi}(x,y) = \int_{-\alpha}^{\alpha} \phi_x(x/\theta)\phi_y(y/\theta) Y\, d\theta \qquad (8)$$

จากสมการข้างต้นจะเห็นบทบาทของแบบสอบร่วมได้ชัดเจน ทำให้สามารถประมาณการแจกแจงร่วม (Joint Distribution) ของ x และ y ถึงแม้จะไม่มีผู้ใดทำแบบสอบทั้งสองฉบับ

สมการ (8) เป็นการแจกแจงร่วมกันของตัวแปรสามตัว คือ θ, x และ y เนื่องจาก θ เป็นตัวระบุคะแนนจริง ξ และ η การแจกแจงนี้นับว่าเป็นการแจกแจงร่วมกันของตัวแปร 4 ตัว คือ ξ , η , x และ y สมการนี้ไม่มีทางคลี่ประมาณได้ นอกจากระบุความสัมพันธ์เชิงดิควิเปลร์เซนต์ไทล์ ระหว่าง x และ y จากสองสมการสุดท้าย

อย่างไรก็ตาม รูปแบบการเทียบมาตราด้วยทฤษฎีตอบสนองข้อสอบยังแบ่งออกเป็น 3 แบบ ซึ่งแตกต่างไปตามจำนวนพารามิเตอร์ที่ใช้ในแต่ละรูปแบบ ดังนี้ (Hambleton and Swaminathan, 1985)

1. แบบที่ใช้พารามิเตอร์ 3 ตัว (Three-Parameters Logistic Model) ฟังก์ชั่นที่ใช้กับเส้นลักษณะของข้อสอบ เป็นดังนี้

$$P_i(\theta) = c_i + (1-c_i)\frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad ; \; i = 1,2,3,\ldots,n$$

เมื่อ $P_i(\theta)$ คือ ความน่าจะเป็นในการตอบข้อสอบข้อที่ i ถูกเมื่อผู้สอบมีความสามารถ θ

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

b_i คือ ค่าความยากของข้อสอบข้อที่ i

c_i คือ ค่าการเดาของข้อสอบข้อที่ i

D คือ scaling factor มีค่า 1.702

e คือ ค่าคงที่มีค่าประมาณ 2.7182818...

2. แบบที่ใช้พารามิเตอร์ 2 ตัว (Two-Parameters Logistic Model) ฟังก์ชั่นที่ใช้เป็นดังนี้

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1+e^{Da_i(\theta-b_i)}} \quad ; i=1,2,3,\ldots,n$$

3. แบบที่ใช้พารามิเตอร์ 1 ตัว (One-Parameter Logistic Model) ซึ่งราส์ช (Rasch) ได้พัฒนาขึ้นมาใช้ โดยรูปแบบนี้เป็นกรณีพิเศษของแบบ 2 พารามิเตอร์ ที่เบิร์นบลม (Birnbuam) ได้พัฒนาขึ้น ข้อตกลงเบื้องต้นที่เพิ่มขึ้น คือ ค่าอำนาจจำแนกของข้อสอบ (a) เป็นค่าคงที่สำหรับข้อสอบแต่ละข้อ และค่าความยากของข้อสอบแปรเปลี่ยนไป ฟังก์ชั่นที่ใช้ เป็นดังนี้

$$P_i(\theta) = \frac{e^{D\bar{a}(\theta-b_i)}}{1+e^{D\bar{a}(\theta-b_i)}} \quad ; i=1,2,,3,\ldots,n$$

<u>รูปแบบในการเก็บรวบรวมข้อมูล</u>

ในการเทียบคะแนนแบบสอบต่างชุดซึ่งสอบกับผู้สอบต่างกลุ่มนั้น การประมาณค่าพารามิเตอร์แยกกันระหว่างแบบสอบแต่ละชุดจะไม่อยู่ในสเกลเดียวกัน ดังนั้นจึงจำเป็นต้องลอกแบบในการเก็บรวบรวมข้อมูลภายใต้รูปแบบที่สามารถจะจัดกระทำทางสถิติ เพื่อหาค่าคงที่สำหรับการแปลงค่าพารามิเตอร์ที่ประมาณค่าได้จากแบบสอบต่างชุดให้อยู่ในสเกลร่วมกัน และสามารถเทียบหาคะแนนสมมูลระหว่างแบบสอบได้ สำหรับรูปแบบในการเก็บรวบรวมข้อมูลเพื่อเทียบคะแนนนั้นพอแบ่งได้เป็น 3 รูปแบบดังนี้ (Lord, 1982 :Hambleton, 1985)

1. รูปแบบที่ใช้กลุ่มผู้สอบกลุ่มเดียวหรือใช้ผู้สอบร่วม โดยกลุ่มเดียวกันทำแบบสอบทั้งสองชุด การประมาณค่าความยาก จึงไม่เกี่ยวข้องกับความแตกต่างในระดับความสามารถ

ของกลุ่มผู้สอบแต่ลำดับการสอบชุดใดก่อนหลังอาจทำให้เกิดการเรียนรู้ การฝึกฝน และความเหนื่อยล้า ซึ่งจะมีผลกระทบต่อคะแนนผลการสอบได้ ดังนั้นเพื่อขจัดความลำเอียงเหล่านี้จึงดัดแปลงรูปแบบนี้ โดยสุ่มผู้สอบเป็นสองกลุ่ม แต่ละกลุ่มทำแบบสอบทั้งสองชุดในลักษณะของการจัดลำดับก่อนหลังต่างกัน โดยวิธีสุ่ม

2. รูปแบบที่ใช้กลุ่มที่เทียบเท่ากัน รูปแบบนี้กลุ่มผู้สอบที่เทียบเท่ากันสองกลุ่ม ซึ่งอาจได้มาโดยการสุ่ม หรือเป็นกลุ่มที่ได้จากการจับคู่ (Matching) ทางด้านความรู้และความสามารถ ทั้งนี้เพื่อให้คะแนนผลการสอบจากแบบสอบแต่ละชุดที่นำมาเทียบคะแนนนั้นไม่มีผลของความแตกต่างในความสามารถระหว่างกลุ่มผู้สอบ อย่างไรก็ตาม กลุ่มผู้สอบทั้งสองอาจมีความแตกต่างในการแจกแจงความสามารถ ถึงแม้จะมีเพียงเล็กน้อยก็อาจทำให้เกิดความลำเอียงในการเทียบคะแนนได้

3. รูปแบบที่ใช้แบบสอบร่วม รูปแบบนี้กลุ่มผู้สอบสองกลุ่มที่แตกต่างกัน (อาจเป็นกลุ่มที่ได้มาจากการสุ่มหรือไม่ได้จากการสุ่ม) แต่ละกลุ่มทำแบบสอบเพียง 1 ชุด ซึ่งแต่ละชุดจะประกอบด้วยข้อสอบร่วมจำนวนหนึ่ง โดยอาจสอบรวมหรือสอบแยกต่างหากจากแบบสอบทั้งสองชุดก็ได้ แบบสอบร่วมจะต้องคู่ขนาน หรือเป็นตัวแทนแบบสอบทั้งสองชุดนั้นให้มากที่สุด แบบสอบร่วมจะกำหนดสเกลร่วมของคะแนนจากแบบสอบแต่ละชุดที่สอบโดยผู้สอบทั้งสองกลุ่มเพื่อนำไปปรับค่าความยากของข้อสอบ หรือค่าความสามารถของผู้สอบจากแบบสอบต่างชุดให้อยู่ในสเกลร่วมกัน

<u>การเทียบค่าประมาณความสามารถของผู้สอบ</u>

การเทียบค่าประมาณความสามารถ จะแตกต่างไปตามรูปแบบในการเก็บรวบรวมข้อมูลซึ่งมีรายละเอียดในการเทียบ ดังนี้

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

วิธีที่ 1 รวมข้อมูลทั้งสองชุดและประมาณค่าพารามิเตอร์ของข้อสอบและค่าความสามารถ ซึ่งจะได้ค่าพารามิเตอร์อยู่ในสเกลของแบบสอบชุด X และแบบสอบชุด Y

วิธีที่ 2 แยกข้อมูลแต่ละชุด (ถ้าจำเป็นต้องประมาณค่าพารามิเตอร์แยกกันสำหรับแบบสอบแต่ละชุด) แล้วกำหนดค่า θ ให้เหมือนกันในการประมาณค่าพารามิเตอร์แต่ละชุด ด้วยวิธีนี้ค่าประมาณความสามารถจะอยู่บนสเกลร่วมกัน

2. รูปแบบที่ใช้กลุ่มเทียบเท่ากัน กรณีนี้ไม่มีข้อสอบร่วมหรือผู้สอบร่วม จึงจำเป็นต้องประมาณค่าพารามิเตอร์แยกกันสำหรับแบบสอบแต่ชุด โดยมีลำดับขั้นดังนี้

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

2.2 ประมาณค่าพารามิเตอร์แยกกันสำหรับกลุ่มผู้สอบแต่ละกลุ่ม โดยกำหนดค่า θ ให้เหมือนกันในการประมาณค่าทั้งสองกลุ่ม

2.3 เรียงลำดับค่าประมาณ θ และจับคู่ระหว่างค่าประมาณที่ต่ำสุดของ θ_X กับ θ_Y ที่ต่ำสุดทีละคู่เรื่อยไป

2.4 พล็อตกราฟระหว่างค่า θ_X กับค่า θ_Y ซึ่งถ้าเป็นไปตามข้อตกลงของทฤษฎีตอบสนองข้อสอบแล้ว ผลที่ได้จะเป็นเส้นตรง อย่างไรก็ตามในทางปฏิบัติใช้ค่าประมาณพารามิเตอร์ ดังนั้นผลที่ได้อาจไม่เป็นเส้นตรง โดยเฉพาะส่วนปลายของค่า θ_X และ θ_Y จะมีความคลาดเคลื่อนในการประมาณค่ามากกว่าส่วนอื่น

3. รูปแบบที่ใช้แบบสอบร่วม ในรูปแบบนี้ผู้สอบ N_X ทำแบบสอบชุด X ซึ่งมี n_X ข้อ และแบบสอบร่วม n_A ข้อ เช่นเดียวกับผู้สอบ N_Y ทำแบบสอบชุด Y ซึ่งมี n_Y ข้อ และข้อสอบร่วมลีก n_A ข้อ มีวิธีการที่เหมาะสม 2 วิธี สำหรับการเทียบมาตรา คือ

วิธีที่ 1 ประมาณค่าแยกกันสำหรับแบบสอบแต่ละชุด คือ ชุด X มีผู้สอบ N_X คน และมีข้อสอบ $(n_X + n_A)$ ข้อ ส่วนชุด Y มีผู้สอบ N_Y คน และมีข้อสอบ $(n_Y + n_A)$ ข้อ กำหนดค่า θ ให้เหมือนกัน ในการประมาณค่าทั้งสองชุด แล้วหาเส้นความสัมพันธ์ระหว่างแบบสอบโดยผ่านข้อสอบร่วม n_A ข้อ ที่สอบกับกลุ่มผู้สอบทั้งสองกลุ่ม $(N_X + N_Y)$ คน ตามวิธีที่อธิบายในข้อ 1 ดังนั้น ความสามารถของผู้สอบจะเทียบกันได้โดยใช้เส้นความสัมพันธ์นี้

วิธีที่ 2 ใช้โปรแกรม LOGIST ประมาณค่าพารามิเตอร์ของข้อสอบและค่าความสามารถของผู้สอบพร้อมกันทั้งสองชุด ดังนี้

1) จัดข้อมูลรวมกับว่าผู้สอบ $N_x + N_y$ คน ทำแบบสอบ $(n_x + n_y + n_a)$ ข้อ

2) จัดข้อสอบ n_y ข้อ ซึ่งผู้สอบ N_x คน ไม่ได้ทำแบบสอบให้เป็นข้อสอบที่ผู้สอบทำไม่ถึง (not reached) ส่วนข้อสอบ n_x ข้อ ซึ่งผู้สอบ N_y คน ไม่ได้ทำข้อสอบก็จะจัด ให้เป็นข้อสอบที่ผู้สอบทำไม่ถึงเช่นกัน

3) ประมาณค่าพารามิเตอร์ความสามารถ สำหรับผู้สอบ $(N_x + N_y)$ คน และพารามิเตอร์ของข้อสอบ สำหรับข้อสอบ $(n_x + n_y + n_a)$ ข้อ พร้อมกันครั้งเดียวด้วยวิธีนี้ ค่าประมาณพารามิเตอร์ที่ได้จากแบบสอบต่างชุดจะอยู่ในสเกลเดียวกันทั้งหมด และค่าประมาณความสามารถระหว่างผู้สอบแบบสอบต่างชุดสามารถนำมาเปรียบเทียบได้

ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา

การประเมินความคลาดเคลื่อนของการเทียบมาตราที่สมบูรณ์ คือ การวิเคราะห์แหล่งความลำเอียง (bias) และความแปรผัน (variability) แล้วนำเสนอให้เห็นเป็นภาพเดียวกัน เพื่อช่วยในการตัดสินใจว่า ในการนำการเทียบมาตราด้วยรูปแบบนั้น ๆ ไปใช้มีความเหมาะสมหรือไม่อย่างไร (Braun and Hooland ,1982)

1. ความลำเอียงของการเทียบมาตรา (Bias error)

ความลำเอียงเป็นองค์ประกอบหนึ่งของความคลาดเคลื่อน ซึ่งจะต้องพิจารณากันอย่างรอบคอบ เมื่อต้องการนำผลของการเทียบมาตราไปใช้ บรูนและฮอลแลนด์ (1982) ได้เสนอบทวิเคราะห์ ไว้ว่า ความลำเอียงทางสถิติ มีความหมาย 2 ประการ คือ

1) ความแตกต่างของค่าเฉลี่ยของค่าตัวประมาณตลอดจนการทำการสุ่มซ้ำ ๆ จากประชากรเดียวกัน กับค่าของประชากรที่ถูกประมาณ

2) ความแตกต่างของฟังก์ชั่นของการเทียบมาตราโดยประมาณ (Estimated equating function) และค่าที่แท้จริงของฟังก์ชั่นการเทียบมาตรา (equating function)

แหล่งของความลำเอียงของการเทียบมาตราที่สำคัญมี 2 แหล่ง คือ

2.1 ความผันแปรของประชากร (Population variability) เนื่องจากการเทียบมาตราไม่ได้จัดกระทำกับประชากรกลุ่มเดียวโดยตลอด ฟังก์ชั่นของการเทียบมาตราที่สร้างขึ้นกับเพื่อใช้กับประชากรหนึ่ง อาจนำไปใช้กับประชากรหนึ่งไม่ได้ เช่น ถ้า $X_P^*(Y)$ เทียบมาตราจาก Y ไปยังมาตรา X ในประชากร P และ $Y_Q^*(Z)$ เทียบมาตราจาก Z ไปยังมาตรา Y ในประชากร Q แล้ว จะสรุปว่า ฟังก์ชั่นเชิงซ้อน $X_P^*(Y_Q^*(Z))$ เทียบมาตราจาก Z ไปยัง X ในประชากร P หรือ Q ก่อนไม่ได้ ในแบบแผนการเทียบมาตราที่ใช้แบบสอบร่วมได้แก้ไขปัญหานี้ โดยการสร้างประชากรสังเคราะห์ (Systhesis population) ซึ่งเป็นผลรวมของประชากร P และ Q ตามสัดส่วน แต่ยังไม่รับประกันผลว่า จะไม่มีความลำเอียงในทุก ๆ ประชากรที่แปรเปลี่ยนไป

2.2 ความคลาดเคลื่อนของรูปแบบ (Model error) ความคลาดเคลื่อนนี้ เกิดขึ้นจากการจัดการกระทำข้อสมมุติฐานที่ว่าด้วยรูปการแจกแจงให้ง่ายขึ้นอย่างผิดพลาด ดังแบ่งเป็น 2 พวก คือ

1) ความคลาดเคลื่อนของรูปแบบที่ทดสอบได้ (Testable model error) เช่น ความเป็นเส้นตรงของฟังก์ชั่นการกดถอย สามารถทดสอบได้จากข้อมูลจำนวนมาก ซึ่งจะเห็นผลได้ว่า มีความเพียงพอหรือไม่

2) ความคลาดเคลื่อนรูปแบบที่ทดสอบไม่ได้ (Nontestable model errors) เป็นกรณีที่ไม่สามารถหาข้อมูลมาทดสอบเพื่อพิสูจน์ว่า สมมติฐานที่ว่าไว้เพียงพอหรือไม่

2. ความแปรผันเชิงสุ่ม และความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา

วิธีการเทียบมาตราทุกวิธี ไม่ว่าอยู่ในรูปแบบการเทียบมาตราใด เมื่อกลุ่มตัวอย่างผู้สอบเป็นกลุ่มที่สุ่มมาจากประชากรเดียว หรือหลายประชากร ก่อนมีความผันแปรเชิงสุ่มเกิดขึ้น จึงนิยมใช้เทคนิคการประมาณค่าความคลาดเคลื่อนเชิงสุ่ม (Estimate sampling error) กับการเทียบมาตราต่าง ๆ ความคลาดเคลื่อนเชิงสุ่มมีสมมุติฐานเบื้องต้นว่า กลุ่มตัวอย่างมาจากการสุ่ม และใช้ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (Standard error of equating : SEE) เป็นการวัดความแปรผันประเภทนี้

ความคลาดเคลื่อนมาตรฐานโดยอิงรูปแบบทฤษฎีตอบสนองข้อสอบ ค่าประมาณคะแนนจำนวนข้อที่ตอบถูก (Number-right score) ξ ของผู้สอบที่ทำแบบสอบ X และ η ของผู้ที่ทำแบบสอบ Y มีค่าเท่ากับฟังก์ชั่นคุณลักษณะที่ประเมินที่ระดับความสามารถ (θ) เดียวกัน ให้ความหมายของการเท่ากันในเชิงคะแนนสมมูล ในทางปฏิบัติจึงนำความสัมพันธ์เชิงฟังก์ชั่น ξ และ η มาใช้เทียบคะแนน X และ Y จากคะแนนที่สอบได้จริงของคนทั้งสองกลุ่ม การประมาณเพื่อให้เทียบ η ไปยัง ξ ตามฟังก์ชั่น จะต้องใช้ค่าประมาณของประชากรข้อสอบ ค่าประมาณเหล่านี้ คือ ที่มาของความคลาดเคลื่อนเชิงสุ่ม (Sampling error)ในการเทียบมาตรา (Lord ,1981)

<u>ความเพียงพอของการเทียบมาตรา</u>

วิธีการเทียบมาตราแต่ละวิธี ประกอบขึ้นด้วยรูปแบบของการเทียบมาตรา (Model) ซึ่งมีข้อตกลงที่ว่าด้วยสมมติฐานเบื้องต้น (Assumptions) ของแต่ละรูปแบบ และประกอบด้วยการลอกแบบ (Designs) เพื่อจัดเก็บข้อมูลให้เป็นไปตามข้อตกลงต่าง ๆ ถ้าหากทุกอย่างเป็นไปตามเงื่อนไข เชื่อได้ว่า ผลของการเทียบมาตราจะมีความถูกต้อง (Accurate) และความแม่นยำ (Precise) ตามทฤษฎี แต่ความเป็นจริงของการทดสอบมักไม่ได้เป็นไปตามอุดมการณ์เพราะมีหลายสิ่งหลายอย่างอยู่นอกเหนือการควบคุม เช่น โปรแกรมการทดสอบระดับชาติ ซึ่งมีทั้งนโยบาย และกฎเกณฑ์เป็นตัวกำหนด จึงไม่สามารถควบคุมการทดสอบให้เป็นไปตามข้อตกลงเชิงทฤษฎี ตัวแบบสอบเองก็จำเป็นต้องมีการเปลี่ยนชุดไป ดังนั้นข้อมูลที่นำมาใช้ศึกษา จึงไม่สามารถระบุว่า เป็นตัวอย่างประชากรได้อย่างชัดเจน และโดยความเป็นจริงเป็นการจัดกระทำกับประชากรมากกว่า (Population quantites) ไม่ใช่ค่าประมาณ (Sample estimates) (Braun and Holland,1982:10)

ด้วยสภาพความเป็นจริงดังกล่าว ทำให้การเทียบมาตราที่จัดกระทำอยู่นั้นอยู่ในภาวะที่มีเงื่อนไขน้อยกว่าความพอดีตามข้อตกลงในแต่ละรูปแบบ ดังนั้น การเทียบมาตราที่ได้พัฒนาขึ้น จึงจำเป็นต้องมีการตรวจสอบความเพียงพอของรูปแบบ (The adequacy of equating models)(Petersen, Marco and Stewart,1982:71) วิธีการประเมินความเพียงพอมีผู้เสนอแนวคิดและวิธีการปฏิบัติไว้ดังนี้

1. ดัชนีตรวจสอบความเพียงพอของเจเกอร์ (Jaeger)

เจเกอร์ (Jaeger,1981:26) ได้เสนอดัชนี 5 ตัว เพื่อตรวจสอบความเพียงพอของการใช้รูปแบบเชิงเส้นตรงว่า เทคนิควิธีที่ได้นำมาใช้ในการเทียบมาตราเพียงพอกับการปรับความแตกต่างระหว่างการแจกแจงของคะแนนจากแบบสอบหรือไม่ หรือจำเป็นต้องมีรูปแบบสิ่นที่มีความสลับซับซ้อนมากขึ้น ดัชนี 5 ตัว มีดังนี้ คือ

1.1 ดัชนีความคล้ายคลึงของการแจกแจงคะแนนสะสมของแบบสอบเก่าและชุดใหม่ โดยการปรับความแตกต่างระหว่างค่าเฉลี่ยและส่วนเบี่ยงเบนมาตราฐาน การทดสอบความเหมือนของการแจกแจงใช้ทดสอบด้วย The Kolmogorov-smirnov two-sample test (Smirnov,1948 cited by Jaeger,1981:27)

1.2 รูปแบบของการแจกแจงคะแนนดิบกับคะแนนแปลง (Shape of the raw score tranformation) ดัชนีตัวนี้บอกความเพียงพอของการเทียบมาตราด้วยรูปแบบเชิงเส้น ซึ่งสามารถอธิบายความแตกต่างในการแจกแจงคะแนนดิบทั้งสองชุดในการเทียบมาตราด้วยรูปแบบเชิงเส้น ถ้าสามารถอธิบายความแตกต่างในการแจกแจงคะแนนดิบทั้งสองชุดอย่างเพียงพอ ก็สามารถยอมรับได้ว่า การแปลงคะแนนดิบจากแบบสอบชุดใหม่ไปยังแบบสอบชุดเก่าเป็นเส้นตรงอย่างแน่นอน

1.3 ความคงเส้นคงวาของผลลัพธ์ของการเทียบมาตราตามรูปแบบเชิงเส้นกับการเทียบที่ตำแหน่งเปอร์เซนไตล์ (Consistency of linear and equipercentile equating results) การวิเคราะห์นี้ อาศัยข้อตกลงที่เป็นสมมติฐานเบื้องต้นที่ว่า ถ้ารูปแบบเชิงเส้นตรงมีความเพียงพอแล้ว ฟังก์ชั่นของรูปแบบการเทียบมาตราที่ตำแหน่งเปอร์เซนไตล์จะแปรผันไปโดยสุ่มรอบ ๆ ฟังก์ชั่นของรูปแบบเชิงเส้นตรงที่สมนัยกัน

1.4 ความคล้ายคลึงของการแจกแจงความยากของข้อสอบ (Similarity of item difficulty distributions) โดยอาศัยหลักการที่ว่า การใช้รูปแบบเชิงเส้นตรงมีความเพียงพออย่างแท้จริงกับแบบสอบที่มีคุณสมบัติเป็นคู่ขนานกัน ถ้ามีความเบี่ยงเบนจากความเป็นคู่ขนานมากเท่าใด แสดงว่า ต้องการรูปแบบการเทียบมาตราที่ซับซ้อนขึ้น เพราะการแจกแจงของแบบสอบที่ไม่ใช่คู่ขนานจะมีความแตกต่างเกิดขึ้นในระดับโมเมนต์ที่สูงขึ้น

1.5 ความคล้ายคลึงของค่าอำนาจจำแนกของข้อสอบ (Similarity of item discrimination distributions) มีเหตุผลทำนองเดียวกับดัชนีตัวที่ 4

2. <u>ดัชนีความแตกต่าง</u> (Discrepancy indecies)

ปีเตอร์สัน มาร์โค และ สตีเวอร์ท (Petersen, marco and Stewart, 1982) ได้เสนอวิธีประเมินความเพียงพอของรูปแบบการเทียบมาตราโดยการหาความแตกต่างระหว่างคะแนนแปลง (An estimated criterion score : t') ซึ่งเป็นผลจากการเทียบมาตรา กับคะแนนเกณฑ์ (Criterion score : t) ที่สมนัยกัน ถ้าความแตกต่างมีค่าน้อย มีความหมายว่า ความคลาดเคลื่อนที่เกิดขึ้นจากการใช้รูปแบบการเทียบมาตรานั้นน้อยด้วย รูปแบบการเทียบมาตราดังกล่าวย่อมมีความเหมาะสมกับสถานการณ์ที่ใช้

ดัชนีความแตกต่างที่นำมาเปรียบเทียบระหว่างรูปแบบและสถานการณ์ที่ต่าง ๆ กัน คือ ค่ารากที่สองของค่าเฉลี่ยของความแตกต่างที่ถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ (The weighted mean-square difference) เป็นค่ามาตรฐาน ค่าที่คำนวณออกมานี้ เรียกว่า ความแปรปรวนรวม (The total error) ซึ่งมีสูตรดังนี้

$$\text{total error} = \Sigma f_i d^2_i / nS^2_t \quad \text{.....(9)}$$

เมื่อ $d_i = (t_i - t')$

n = จำนวนคะแนนที่ใช้

S^2_t = ความแปรปรวนของคะแนน t

3. <u>ดัชนีเปรียบเทียบเปอร์เซนไตล์</u> (The percentile comparision index) เป็นมาตราการวัดความไม่สอดคล้องระหว่างการแจกแจงของคะแนนในแบบสอบชุด X กับแบบสอบชุด Y ที่ได้แปลงไปอยู่ในมาตราคะแนนของ X แล้ว ตามวิธีที่ได้พัฒนาขึ้น ดัชนีเปรียบเทียบเปอร์เซนไตล์ คือ ค่าเฉลี่ยกำลังสองของความแตกต่าง (The mean-squared difference) ที่ได้จากการแจกแจงของคะแนนต่าง ๆ ของเกณฑ์ X กับคะแนนแปลง X^* ที่แปลงมาจาก Y ด้วยวิธีเทียบมาตราที่ระบุไว้ ณ ที่ตำแหน่งเปอร์-เซนไตล์เดียวกัน ดัชนีได้เสนอโดย โคเลน (Kolen,1982) ได้แนะนำให้ใช้ข้อมูลจากกลุ่มตัวอย่างสอบทานผล ซึ่งได้สุ่มมาจากประชากรเดียวกันกับกลุ่มตัวอย่างที่ได้พัฒนาตารางคะแนนแปลง สูตรการคำนวณมีดังนี้ คือ

$$C = \Sigma (X_i - X_i^*)^2 / nk \quad \text{.....(10)}$$

เมื่อ n คือ จำนวนของคะแนนดิบของกลุ่มสอบทานผล

k คือ จำนวนข้อสอบในแบบสอบร่วมที่ใช้

ค่า C ที่ได้ ถ้ามีค่าน้อยจะให้ความหมายว่า รูปแบบการเทียบมาตราที่นำมาสร้างตารางคะแนนแปลงนั้นมีความเหมาะสมและเพียงพอที่จะให้ผลการแปลงคะแนนอย่างคงเส้นคงวา

วิธีการประเมินความเพียงพอที่ได้กล่าวมานี้ อาจจำแนกเป็นสองพวก พวกแรกเป็นการประเมินก่อนดำเนินการ เช่น ดัชนีความคลาดคลึงของการแจกแจงคะแนนสะสมของแบบสอบสองชุด ดัชนีความคลาดคลึงของการแจกแจงของค่าความยากของข้อสอบ เป็นต้น พวกหลังเป็นการประเมินผลของการเทียบมาตรา ซึ่งอาศัยคะแนนเกณฑ์ที่เลือกสรรแล้วเป็นหลักในการเทียบหาความแตกต่าง สำหรับดัชนีที่เสนอโดย ปีเตอร์สันและคณะนั้น (1982) คะแนนเกณฑ์ที่ใช้ คือ ผลการแปลงคะแนนด้วยรูปแบบอิงทฤษฎีตอบสนองข้อสอบ

ค่าดัชนีที่คำนวณมีลักษณะเป็นหน่วยมาตรฐานแล้ว สามารถนำค่าเหล่านี้ที่ได้จากการใช้รูปแบบที่ต่างกัน ตลอดจนสถานการณ์ที่ได้ข้อมูลที่ต่างกันมาเปรียบเทียบกันโดยตรงได้ ส่วนดัชนีที่เสนอโดย โคเลน (Kolen,1982) ได้ใช้ข้อมูลจากผู้สอบเองเป็นเกณฑ์ในการหาความแตกต่าง ข้อมูลเหล่านี้ได้จากการออกแบบด้วยการใช้กลุ่มตัวอย่างสอบทานผล ซึ่งผู้สอบในกลุ่มตัวอย่างนี้ได้รับการทดสอบด้วยแบบสอบทั้งสองชุด ดังนั้น การใช้คะแนนของตนเองเป็นเกณฑ์จึงมีความเป็นอิสระ ไม่ขึ้นกับกระบวนการแปลงคะแนนอื่น ๆ เช่นวิธีที่เสนอโดย ปีเตอร์สันและคณะ ด้วยเหตุผลดังกล่าว ผู้วิจัยจึงได้เลือกวิธีการประเมินความเพียงพอจากการวิเคราะห์ผลในกลุ่มตัวอย่างสอบทานผล แต่การหาค่าดัชนีได้ดัดแปลงจากสูตรของโคเลน ไปใช้ตามแนวความคิดของปีเตอร์สันและคณะ คือ ใช้ค่าความแปรปรวนเป็นตัวถ่วงน้ำหนักเพื่อให้ค่าที่ได้มีหน่วยเป็นมาตรฐาน

สูตรที่ดัดแปลงและนำมาใช้ในการศึกษาครั้งนี้ คือ

$$C = \Sigma (X - X_i^{*})^2 / nS_x^2 \qquad \text{.....(11)}$$

เมื่อ X_i คือ คะแนนเกณฑ์หรือคะแนนจากการสอบชุด X ของคนที่ 1

X_i^{*} คือ คะแนนที่ได้เทียบด้วยตารางคะแนนที่สมนัยกันของคนที่ 1

n คือ จำนวนคนในกลุ่มตัวอย่างสอบทานผลที่นำมาวิเคราะห์

S_x^2 คือ ค่าความแปรปรวนของคะแนน X

งานวิจัยที่เกี่ยวข้องกับการเทียบมาตรา

งานวิจัยที่เกี่ยวข้องกับการเทียบมาตราโดยอิงทฤษฎีตอบสนองข้อสอบในต่างประเทศ สเกจส์และลิชชิล (Skaggs Gary and Lissitz Robertes .1986:498-501) ได้สรุปไว้ดังนี้

ตารางที่ 1 การเทียบมาตราตามแนวตั้งโดยอิงทฤษฎีตอบสนองข้อสอบ

ปีที่ศึกษา	ผู้ศึกษา	แบบสอบ	กลุ่มตัวอย่าง	วิธีการ	ตัวแปรที่เทียบมาตรา	วิธีประเมิน
1968	Wright	LSAT	นักเรียนกฎหมาย (980)	ราล์ช	แบบสอบฉบับง่ายกับฉบับยาก และความสามารถของกลุ่มต่ำกับกลุ่มสูง	-เปรียบเทียบความแตกต่างมาตราฐานของแบบสอบ -ลักษณะโค้งความสัมพันธ์ b
1968	Anderson, Kearney และ Everett	แบบสอบเชาว์ปัญญา	ทหาร (610 875	ราล์ช	ความยากของแบบสอบจากกลุ่มตัวอย่าง 2 กลุ่ม	ความสัมพันธ์ b

ตารางที่ 1 (ต่อ) การเทียบมาตราตามแนวตั้งโดยอิงทฤษฎีตอบสนองข้อสอบ

ปีที่ศึกษา	ผู้ศึกษา	แบบสอบ	กลุ่มตัวอย่าง	วิธีการ	ตัวแปรที่เทียบมาตรา	วิธีประเมิน
1975	Tinsley และ Dawis	คำอุปมาอุปมัย	นักเรียนมัธยมศึกษา, ลูกจ้างชั้นดี	ราส์ช	ความยากของข้อสอบจากกลุ่มตัวอย่างที่ต่างกัน	ความสัมพันธ์ b
1978	Slinde และ Linn	แบบสอบผลสัมฤทธิ์ทางคณิตศาสตร์	นักศึกษาปีที่ 1 (390-810)	ราส์ช	เทียบมาตราในแนวตั้งจากกลุ่มตัวอย่างย่อยที่ความสามารถต่างกัน	ความแตกต่างมาตรฐานของความสามารถ
1979	Slinde และ Linn	ความเข้าใจทางการอ่านและคำศัพท์	นักเรียนระดับ 5 (510-570)	ราส์ช	เทียบมาตราในแนวตั้งจากกลุ่มตัวอย่างย่อยที่มีความสามารถต่างกัน	ความแตกต่างมาตรฐานของความสามารถ
1979	Marco, Peterson และ Stewart	SAT-ภาษา	นักเรียนมัธยม (ส่วนมาก) (1580)	ราส์ช, 3 พารามิเตอร์, อิควิเปอร์เซนไทล์ และเชิงเส้นตรง	เทียบมาตราในแนวตั้ง, กลุ่มตัวอย่างที่สุ่มกับที่คล้ายกัน, แบบสอบร่วมภายนอกกับภายใน, ความยากของแบบสอบร่วม, IRT กับอิควิเปอร์เซนต์ไทล์, เปรียบเทียบรูปแบบการเทียบมาตรา	ค่าความคลาดเคลื่อนกำลังสองและดัชนีความลำเอียง

ตารางที่ 1 (ต่อ) การเทียบมาตราตามแนวตั้งโดยอิงทฤษฎีตอบสนองข้อสอบ

ปีที่ศึกษา	ผู้ศึกษา	แบบสอบ	กลุ่มตัวอย่าง	วิธีการ	ตัวแปรที่เทียบมาตรา	วิธีประเมิน
1980	Lord และ Hoover	ITBS, ความเข้าใจทางคณิตศาสตร์	นักเรียนระดับ 6 - 8 (245-300)	ราส์ช	เทียบมาตราในแนวตั้งจากตัวอย่างที่มีระดับความสามารถต่างกัน	เปรียบเทียบจากกราฟ
1981	Guskey	ITBS: ความเข้าใจในการอ่าน	นักเรียนที่มีความสามารถสูงในระดับ 6-8	ราส์ช	เทียบมาตราในแนวตั้ง: เปรียบเทียบสเกลความสามารถโดยราส์ชกับคะแนนเกรดที่เทียบเท่ากัน	เปรียบเทียบจากกราฟของคะแนนดิบกับคะแนนแปลง
1981	Divgi	MAT: การอ่าน	นักเรียนระดับ 6-8 (5500)	ราส์ช	เทียบมาตราในแนวตั้งของแบบสอบย่อย	ความแตกต่างมาตรฐาน, ความลำเอียงจากฟังก์ชั่นของคะแนนรวม

ตารางที่ 1 (ต่อ) การเทียบมาตราตามแนวตั้งโดยอิงทฤษฎีตอบสนองข้อสอบ

ปีที่ศึกษา	ผู้ศึกษา	แบบสอบ	กลุ่มตัวอย่าง	วิธีการ	ตัวแปรที่เทียบมาตรา	วิธีประเมิน
1981	Kolen	ITED:คำศัพท์เกี่ยวกับความยาวของสระ	นักเรียนระดับ 9-12 (1580-1925)	ราส์ช,2,3 พารามิเตอร์, อิควิเปอร์เซนไทล์, เส้นตรง	เทียบมาตราในแนวนอนและแนวตั้ง,เปรียบเทียบโมเดลการเทียบมาตรา,คำศัพท์กับความยาวของสระ	Cross-Validation,ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน
1981	Peterson, Cook, Stocking	SAT:ภาษาและคณิตศาสตร์	นักเรียนมัธยมศึกษา (2670)	3 พารามิเตอร์, เส้นตรง อิควิเปอร์เซนไทล์	เปรียบเทียบรูปแบบการเทียบมาตรา,เปรียบเทียบเมตริกซ์การแปลงของ 3 พารามิเตอร์, ข้อสอบภาษากับคณิต	ความคลาดเคลื่อนกำลังสองเฉลี่ยและดัชนีความลำเอียง,แบบสอบเทียบมาตรา
1981	Cook, Dubar, Eignor	SAT/PSAT ภาษาและคณิตศาสตร์	นักเรียนมัธยมศึกษา (2000)	3 พารามิเตอร์, เส้นตรง, อิควิเปอร์เซนไทล์	เปรียบเทียบรูปแบบการเทียบมาตรา,ข้อสอบภาษาและคณิตศาสตร์	ความคลาดเคลื่อนกำลังสองเฉลี่ย-การเทียบมาตราโดยใช้ 3 พารามิเตอร์เป็นเกณฑ์

ตารางที่ 1 (ต่อ) การเทียบมาตราตามแนวตั้งโดยอิงทฤษฎีตอบสนองข้อสอบ

ปีที่ศึกษา	ผู้ศึกษา	แบบสอบ	กลุ่มตัวอย่าง	วิธีการ	ตัวแปรที่เทียบมาตรา	วิธีประเมิน
1981	Kolen, Whitney	GED Test	ผู้ใหญ่ (200)	3 พารา-มิเตอร์, ราส์ช, เส้นตรง, อิควิเปอร์เซนไตล์	เปรียบเทียบรูปแบบการเทียบมาตราด้วยการสลับฉบับกัน	ความคลาดเคลื่อนกำลังสองเฉลี่ย Cross-Validation
1981	Cowell	TOELT-SLEP	นักเรียนมัธยม-ศึกษา (n_1=290-320 n_2=2070-3170)	3 พารา-มิเตอร์, ราส์ช, เส้นตรง	กลุ่มตัวอย่างใหญ่และเล็ก, เปรียบเทียบรูปแบบการเทียบมาตราด้วยการสลับฉบับกัน	สถิติโดยสรุปต่าง ๆ-วิธีการเปรียบเทียบรายคู่
1981	Patience	TTED-วิธีเทียน	นักเรียนระดับ 9-10 (1000)	ราส์ช, 2, 3 พารา-มิเตอร์, อิควิเปอร์เซนไตล์	เปรียบเทียบการเทียบมาตราในแนวตั้ง	ความสัมพันธ์ระหว่างมาตรคะแนนของผู้สอบด้วยจุดเริ่มต้น, เกณฑ์บนพื้นฐานวิธีเทียบมาตราแบบอิควิเปอร์เซนไตล์กับคะแนนรวม

ตารางที่ 1 (ต่อ) การเทียบมาตราตามแนวตั้งโดยอิงทฤษฎีตอบสนองข้อสอบ

ปีที่ศึกษา	ผู้ศึกษา	แบบสอบ	กลุ่มตัวอย่าง	วิธีการ	ตัวแปรที่เทียบมาตรา	วิธีประเมิน
1982	Modu	CB-Adv. ฟิสิกส์กาย-ภาพ	นักเรียน ม.ต้น	3 พารา-มิเตอร์, เส้นตรง ลิควิเปอร์ เซนไตล์	เทียบมาตราของฟอร์มที่เลือกภายใต้เงื่อนไข Mutidimensiona-lity	เปรียบเทียบกราฟลองฟังก์ชั่นการเทียบมาตรา
1983	Holmes, Doody, Bogan	CTBS-ศัพท์ ความเข้า-ใจในการอ่าน	นักเรียน ระดับ 4 5,8,9	3 พารา-มิเตอร์	วิธีการ 3 วิธีของการสร้างเมตริกซ์ธรรมดา	ค่าความคลาด-เคลื่อนกำลังสองประยุกต์ไปสู่กลุ่มทานผล
1983	Doody, Bogan, Yen	การแปล	(2000)	3 พารา-มิเตอร์	-เทียบมาตราในแนวตั้ง -เมตริกข้อมูลมิติเดียว และหลายมิติ	เปรียบเทียบโดยประมาณความสามารถจากแบบสอบย่อย
1983	Loyd	ITBS ความคิดรวบยอด ทางคณิต-ศาสตร์	นักเรียน ระดับ 3-8 (250)	ราส์ช, 3 พารา-มิเตอร์	เปรียบเทียบรูปแบบการเทียบมาตรา, แบบสอบร่วมภายในและภายนอก, การเทียบมาตราในแนวตั้ง	เทียบโดยใช้กราฟ

ตารางที่ 1 (ต่อ) การเทียบมาตราตามแนวตั้งโดยลิงทฤษฎีตอบสนองข้อสอบ

ปีที่ศึกษา	ผู้ศึกษา	แบบสอบ	กลุ่มตัวอย่าง	วิธีการ	ตัวแปรที่เทียบมาตรา	วิธีประเมิน
1984	Cook, Eignor, Taft	ชีววิทยา	นักเรียนมัธยมศึกษา (2400-3900)	3 พารามิเตอร์, เชิงเส้นตรง, อิควิเปอร์เซนไตล์	IRT และแบบดั้งเดิม เทียบบนฐานของกลุ่มตัวอย่างที่ต่างกัน, เปรียบเทียบโมเดลการเทียบมาตรา	ฟังก์ชั่นของการเปรียบเทียบการเทียบมาตรา
1985	Harris, Kolen	ACT-การใช้คณิตศาสตร์	นักเรียนมัธยมศึกษา (3870-3970)	3 พารามิเตอร์, อิควิเปอร์เซนไตล์, เชิงเส้นตรง	เปรียบเทียบโมเดลการเทียบมาตรา, กลุ่มตัวอย่างที่มีความสามารถสูง-ต่ำ	เปรียบเทียบจากกลุ่มที่แตกต่าง - ความคลาดเคลื่อนกำลังสองและดัชนีความลำเอียง

สไลด์และลินด์ (Slinde and Linn,1977-1979) ได้ทำการตรวจสอบปัญหาของการเทียบมาตราในแนวตั้ง (Vertical Equating) ของแบบสอบสองชุดที่สร้างขึ้นเพื่อใช้กับประชากรที่มีความสามารถต่างกันด้วยวิธีอ้อม จากการศึกษาได้ให้ข้อเสนอแนะว่า การใช้รูปแบบการเทียบมาตรา 3 รูปแบบ คือ เชิงเส้นตรง อิควิเปอร์เซนไตล์ และลิงทฤษฎีตอบสนองข้อสอบชนิดหนึ่งพารามิเตอร์ รูปแบบโลจีสอาจมีข้อจำกัดในกระบวนการเทียบมาตราในแนวตั้ง โดยเฉพาะอย่างยิ่งเมื่อทำการเทียบมาตราโดยใช้แบบสอบสองชุดที่มีความแตกต่างกันมาก และกลุ่มตัวอย่าง

สองกลุ่มที่มีระดับความสามารถต่างกันมาก จากการศึกษานี้ได้ให้ข้อสังเกตว่า ถ้าใช้รูปแบบลิงทฤษฎีตอบสนองข้อสอบชนิดสามพารามิเตอร์ของโลจีส อาจให้ผลการเทียบมาตราที่ดีกว่าในสถานการณ์เช่นนั้น

มาร์โค ปีเตอร์สัน และสตีเวอร์ท (Marco,Petersen and Stewart, 1979) ได้ประเมินและเปรียบเทียบวิธีการเทียบมาตราเชิงเส้นตรง วิธีการเทียบมาตราโดยใช้เปอร์เซนไตล์ และวิธีการใช้โค้งลักษณะข้อสอบโดยใช้แบบสอบ SAT ผลปรากฏว่าแบบสอบที่มีความยากแตกต่างกัน วิธีการเทียบมาตราโดยใช้โค้งลักษณะข้อสอบให้ผลดีที่สุด และวิธีเทียบมาตราเชิงเส้นตรงให้ผลน้อยที่สุด และเมื่อปีเตอร์สันและคณะทำซ้ำอีก ในปี ค.ศ.1983 โดยวิเคราะห์ซ้ำกับปี ค.ศ.1979 ผลคงปรากฏเหมือนเดิม แต่เมื่อโคเลน (Kolen,1981)ได้ศึกษารูปแบบการเทียบมาตราระหว่างรูปแบบดั้งเดิม 2 วิธี คือ รูปแบบเชิงเส้นตรงและรูปแบบอิควิเปอร์เซนไตล์ กับรูปแบบการเทียบมาตราโดยใช้โค้งลักษณะข้อสอบที่ศึกษาทั้ง 1,2,3 พารามิเตอร์ แต่ละวิธีเทียบด้วยค่าประมาณจากค่าจริง (Estimated true score equating) และเทียบด้วยการประมาณจากค่าสังเกต (Estimated observed score equating) ใช้ข้อมูลจากโครงการ IOWA Test of Education Development (ITED) ค.ศ.1978 ศึกษา 2 รูปแบบ คือ เทียบจากฟอร์มที่มีความยากเท่าเทียมกัน และเทียบจากฟอร์มที่มีความยากแตกต่างกัน กลุ่มตัวอย่างเป็นนักเรียนระดับที่ 1 คือ ระดับเกรด 9 และ 10 ระดับ 2 คือ ระดับเกรด 11 และ 12 จำนวนทั้งหมด 10,728 คน จากโรงเรียนในรัฐไอโอวา 34 แห่ง การศึกษานี้ใช้เกณฑ์ Cross-Validation criterion ผลการวิจัยพบว่า การเทียบมาตราทั้งหมดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 การเทียบมาตราโดยใช้วิธีอิควิเปอร์เซนไตล์เหมาะสมที่สุดกับแบบสอบที่มีความยากแตกต่างกัน ส่วนการเทียบมาตราโดยใช้โค้งลักษณะข้อสอบพบว่า กรณีใช้พารามิเตอร์เดียว ถึงเป็นวิธีที่ไม่ถูกต้องสำหรับแบบสอบที่ต่างกัน ซึ่งเนื่องมาจากการเดาเกิดขึ้นมาก ส่วนกรณีใช้สามพารามิเตอร์ มีปัญหาที่เกิดขึ้นอย่างหลีกเลี่ยงไม่ได้ คือ การประมาณค่าไม่ถูกต้องของ Lower asymtote parameter

คุ๊ก ดันบาร์ และไอเนอร์ (Cook, Dunbar and Eignor, 1981) ได้ศึกษาผลการเทียบมาตราแบบใช้ทฤษฎีตอบสนองข้อสอบ เปรียบเทียบกับการเทียบมาตราแบบดั้งเดิม 2 วิธี คือ การเทียบมาตราเชิงเส้นตรง และการเทียบมาตราแบบอิควิเปอร์เซนไทล์ การเทียบมาตราโดยใช้ทฤษฎีตอบสนองข้อสอบใช้โมเดลสามพารามิเตอร์ ซึ่งประมาณค่าพารามิเตอร์โดยใช้โปรแกรม LOGIST ด้วยวิธีความเป็นไปได้สูงสุด (Maximum lielihood procedure) สำหรับการประมาณค่าที่อยู่ต่ำกว่าระดับการเดา ใช้วิธีการเชิงเส้นตรง (Linear interpolation) ในการเทียบมาตราศึกษาทั้งสองกรณี คือ มีแบบสอบร่วม และไม่มีแบบสอบร่วม โดยใช้ประชากรที่ไม่ได้สุ่มมาจากประชากรเดียวกัน ทั้งนี้เพื่อศึกษาความเป็นไปได้ (Feasibility) ของวิธีการที่ใช้ทฤษฎีตอบสนองข้อสอบ ในการนำไปแก้ปัญหาทางปฏิบัติ การเปรียบเทียบพิจารณา 2 ประการ คือ ความสอดคล้องสัมพันธ์ (Relative agreement) ระหว่างวิธีที่ใช้ทฤษฎีตอบสนองข้อสอบกับวิธีดั้งเดิมแต่ละวิธี โดยพิจารณาจากการพล็อตกราฟ อีกประการหนึ่งคือ ดัชนีความแตกต่างสำหรับการแจกแจงคะแนนรวม และแต่ละส่วนของการแจกแจง (สูงกว่า 20 เปอร์เซนต์, ตรงกลาง 60 เปอร์เซนต์ และต่ำกว่า 20 เปอร์เซนต์ของการแจกแจง) ซึ่งคำนวณจากค่าที่ถ่วงน้ำหนักของค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างค่าประมาณจากการเทียบมาตราแต่ละวิธีที่ศึกษากับคะแนนเกณฑ์ เกณฑ์ที่ใช้ศึกษาในที่นี้ คือ คะแนนแปลงด้วยการเทียบมาตราที่ใช้ทฤษฎีตอบสนองข้อสอบชนิดสามพารามิเตอร์ ซึ่งงานวิจัยที่มีมาก่อน (Kolen,1981 : Slind and Linn,1977) แนะนำว่ามีความเหมาะสมในการเทียบมาตราในกรณีที่แบบสอบมีความยากต่างกัน สอบกับกลุ่มตัวอย่างที่มีความสามารถต่างกัน

ข้อมูลที่ใช้ในการศึกษาครั้งนี้ เป็นผลการสอบสองครั้งที่ดำเนินการแล้ว โดย The Collage Board Admissions Testing Program ซึ่งใช้แบบสอบ 2 ชุด ที่มีความแตกต่างในความยาวและความยาก แต่ละชุดมีข้อสอบ 2 ตอน คือ ข้อสอบภาษาและคณิตศาสตร์ ใช้กับกลุ่มตัวอย่างที่ไม่ได้เท่าเทียมกันโดยการสุ่ม

ผลการศึกษาพบว่า วิธีการเทียบมาตราแบบดั้งเดิม กับวิธีเทียบมาตราโดยใช้ทฤษฎีตอบสนองข้อสอบมีความสอดคล้องกันมาก เนื่องจากโดยความจริงแล้ว การแจกแจงคะแนนดิบจากแบบสอบทั้งสองชุดมีรูปร่างคล้ายคลึงกันมาก จึงทำให้รูปแบบเชิงเส้นตรง

และเชิงเส้นโค้งหรือลิควิเปอร์เซนไตล์มีผลใกล้เคียงกัน โดยการเทียบเชิงเส้นโค้งมีผลใกล้เคียงกับการเทียบแบบใช้ทฤษฎีตอบสนองข้อสอบมากกว่าการเทียบเชิงเส้นตรง อย่างไรก็ตาม ความแตกต่างที่เกิดขึ้นที่ส่วนปลายของการเทียบแบบดั้งเดิมทุกวิธีที่ต่างไปจากการเทียบที่ใช้ทฤษฎีตอบสนองข้อสอบ เป็นผลจากข้อมูลส่วนนี้มีน้อยและหายากมาก ในทางทฤษฎีจึงแนะนำให้ใช้การเทียบที่ใช้ทฤษฎีตอบสนองข้อสอบเพราะไม่มีผลเนื่องจากขาดแคลนข้อมูลในส่วนปลายของการแจกแจง อย่างไรก็ตาม การเทียบที่ใช้ทฤษฎีตอบสนองข้อสอบก็ไม่สามารถเตรียมการเทียบคะแนนในส่วนที่อยู่ปลายด้านล่างที่ต่ำกว่าระดับการเดาได้ ซึ่งจำเป็นต้องใช้วิธีอื่นประมาณค่าเพิ่มเติม นอกจากนี้ผลการวิจัยพบว่า การเทียบมาตราโดยใช้ทฤษฎีตอบสนองข้อสอบโดยตรงเมื่อไม่มีแบบสอบร่วม และเมื่อกลุ่มตัวอย่างไม่ได้เป็นกลุ่มที่เท่าเทียมกันโดยการสุ่มนั้น ให้ผลแตกต่างจากการเทียบเมื่อมีแบบสอบร่วมเพียงเล็กน้อย

เมื่อโคเลนและวิทนีย์ (Kolen and Whitney, 1982) ได้เปรียบเทียบความถูกต้องของการเทียบมาตราโดยใช้แบบสอบร่วม 4 วิธี คือ วิธีอิควิเปอร์เซนไตล์ วิธีเชิงเส้นตรง พารามิเตอร์ตัวเดียว และพารามิเตอร์สามตัว ใช้แบบสอบ General Education Development (GED) ซึ่งเป็นแบบสอบผลสัมฤทธิ์ เพื่อใช้ในการตัดสินให้ประกาศนียบัตรที่ต้องการเทียบความรู้ระดับเตรียมอุดมศึกษาทั่วประเทศ แบบสอบมี 12 ชุด โดยใช้ฟอร์มที่ 12 เป็นแบบสอบร่วม ที่เหลืออีก 11 ฟอร์ม เป็นแบบสอบที่ใช้เทียบมาตรา ผู้เข้าสอบแต่ละคนต้องทำแบบสอบ 2 ฟอร์ม ในจำนวน 11 ฟอร์ม ส่วนฟอร์มที่ 12 ต้องทำทุกคน กลุ่มตัวอย่างมาจากการสอบปี 1980 จำนวนมากกว่า 800,000 คน ทำการสุ่มแบบแบ่งชั้นหลายชั้นจากประเภทของโรงเรียน เขตทางภูมิศาสตร์ สถานภาพทางสังคม สุดท้ายได้นักเรียนที่เป็นตัวอย่างโรงเรียนละ 22 คน จำนวนคนในแต่ละชุดของแบบสอบประมาณ 200 คน ประมาณค่าพารามิเตอร์โดยใช้โปรแกรม LOGIST แล้วสร้างตารางสมมูลระหว่างแบบสอบอีก 11 ชุด การเทียบผลการวิเคราะห์โดยใช้ Cross-Validation ผลการศึกษาพบว่า วิธีการเทียบมาตราโดยใช้อิควิเปอร์เซนไตล์และใช้โค้งลักษณะข้อสอบได้ผลไม่เป็นที่ยอมรับ ขณะเดียวกันวิธีการเทียบมาตราโดยใช้พารามิเตอร์ตัวเดียวให้ผลเพียงพอเทียบเท่ากับรูปแบบสามพารามิเตอร์ ในการประเมินผลการศึกษา

ครั้งนี้มีการเปรียบเทียบอื่น ๆ ซึ่งสรุปได้ว่า ความถูกต้องของการเทียบมาตราขึ้นอยู่กับองค์ประกอบหลายอย่าง เช่น ลักษณะของแบบสอบ รูปแบบของการเทียบมาตรา ขนาดของกลุ่มตัวอย่าง เป็นต้น

สัทเทน (Hutten,1982) ได้ศึกษาความเหมาะสมของข้อมูลจริงกับรูปแบบทฤษฎีตอบสนองข้อสอบ 2 รูปแบบ คือ รูปแบบของราส์ช กับรูปแบบ 3 พารามิเตอร์ ในการประมาณค่าพารามิเตอร์ ความสามารถ และความยากของข้อสอบ สำหรับกลุ่มตัวอย่างขนาดเล็ก (250 คน) และแบบสอบสั้น (20 ข้อ) ผลการศึกษาพบว่า รูปแบบของราส์ช และรูปแบบ 3 พารามิเตอร์ มีความเหมาะสมกับข้อมูลประมาณร้อยละ 80 ทั้งสองรูปแบบเมื่อเปรียบเทียบคะแนนที่ทำนายกับคะแนนที่วัดได้ โดยใช้สถิติ Kolmogorov-Smirnov แล้ว รูปแบบของราส์ชเหมาะสมกับข้อมูลทั้งหมดดีกว่ารูปแบบ 3 พารามิเตอร์ ผู้วิจัยสนับสนุนให้ใช้รูปแบบ 3 พารามิเตอร์ กับกลุ่มตัวอย่างที่มีขนาดตั้งแต่ 1,000 คนขึ้นไป เพื่อจะได้ความแม่นตำในการประมาณค่าพารามิเตอร์ต่าง ๆ ผลการวิจัยนี้สอดคล้องกับที่ดักลาสส์ (Douglass,1981) ได้ศึกษาเปรียบเทียบการใช้รูปแบบทฤษฎีตอบสนองข้อสอบกับผลการสอบในระดับห้องเรียน โดยใช้รูปแบบทฤษฎีตอบสนองข้อสอบประมาณค่าพารามิเตอร์ของข้อสอบ 100 ข้อ ที่เป็นข้อสอบปลายปีของ 4 ปีการศึกษา จากกลุ่มตัวอย่างระดับมหาวิทยาลัยจำนวน 594-1,082 คน จัดข้อสอบแบ่งเป็น 4 ชุด จากนั้นทำการสุ่มข้อสอบชุดละ 43-53 ข้อ สุ่มผู้สอบจำนวน 200, 600 และ 800 คน มาวิเคราะห์ ปรากฏว่าการวิเคราะห์ข้อสอบด้วยรูปแบบของราส์ชมีความคงที่ของค่าพารามิเตอร์ดีกว่ารูปแบบ 3 พารามิเตอร์ และเมื่อมีการเทียบคะแนนของแบบสอบข้ามกลุ่มตัวอย่าง รูปแบบของราส์ชก็มีความคงที่ดีกว่าด้วย

ลอร์ด และวิงเกอร์สกี้ (Lord and Wingersky,1984) ได้เปรียบเทียบการเทียบมาตราระหว่างรูปแบบทฤษฎีตอบสนองข้อสอบที่เทียบโดยใช้คะแนนจริง และรูปแบบทฤษฎีตอบสนองข้อสอบที่เทียบคะแนนสังเกตโดยใช้อิควิเปอร์เซนไทล์ จากแบบสอบ SAT ด้านภาษา จำนวน 6 ชุด เทียบเข้าสู่กันในลักษณะโซ่ โดยใช้แบบสอบร่วม ทั้งนี้แบบสอบชุดแรกและชุดสุดท้ายเป็นชุดเดียวกัน การออกแบบรวบรวมข้อมูลดำเนินการดังภาพที่ 1

ภาพที่ 2 การออกแบบรวมรวมข้อมูลของลอร์ดและวิงเกอร์สกี้

จากภาพ แบบสอบ V4 เทียบไปสู่แบบสอบ X2 โดยใช้แบบสอบร่วม fe ส่วนแบบสอบ X2 เทียบไปสู่แบบสอบ Y3 โดยใช้แบบสอบร่วม fm ซึ่งจะได้ผลการเทียบจาก V4 ไป Y3 กระทำการเทียบคะแนนในลักษณะเดียวกันไปยังแบบสอบชุดอื่น สุดท้ายจะได้ผลการเทียบ V4 กับ V4 ซึ่งเป็นการเทียบแบบสอบกับตัวเอง และใช้เป็นเกณฑ์ในการเปรียบเทียบความคงเส้นคงวาของวิธีการเทียบมาตราทั้ง 2 วิธี ทั้งนี้เกณฑ์นี้แตกต่างจากที่โคเลน (Kolen,1981) ได้เคยใช้ศึกษาความคงเส้นคงวาของวิธีเทียบมาตราแบบต่าง ๆ โคเลนใช้เกณฑ์จากกลุ่มทานผลซึ่งเป็นกลุ่มอิสระอีกกลุ่มหนึ่ง โดยผู้วิจัยกล่าวไว้ว่า เกณฑ์ที่โคเลนใช้ ไม่ใช่เกณฑ์ที่มีคุณสมบัติสำหรับการเลือกวิธีการเทียบมาตราที่ดีที่สุด เพราะวิธีเทียบมาตราที่ไม่ถูกต้อง อาจให้ผลการเทียบคงเส้นคงวามากกว่าวิธีการที่ถูกต้อง เกณฑ์การเทียบแบบสอบกับตัวเองที่ใช้ในการศึกษานี้อาจไม่เที่ยงตรงในเชิงทฤษฎี และไม่น่าพอใจในการปฏิบัติสำหรับการเทียบแบบสอบสองฉบับที่แตกต่างกัน แต่เป็นเกณฑ์ที่มีประโยชน์ต่อการเปลี่ยนวิธีการเทียบมาตราใหม่ หากพบว่าคะแนนที่แปลงแล้วในขั้นสุดท้ายไม่เป็นอย่างเดียวกันกับคะแนนเดิมของตนเอง

ผลการวิจัยยังไม่มีการตัดสินใจเชิงทฤษฎีที่ชัดเจนในการเลือกวิธีการเทียบมาตราแบบทฤษฎีตอบสนองข้อสอบที่ใช้คะแนนจริง กับแบบทฤษฎีตอบสนองข้อสอบที่ใช้คะแนนสังเกตโดยอิควิเปอร์เซนไทล์ เนื่องจากทั้งสองวิธีนี้มีความสอดคล้องกัน

โดรานส์ และคิงสตัน (Dorrans and Kingston,1985) ได้ศึกษาผลการฝ่าฝืนเกี่ยวกับความมีมิติเดียว ในการประมาณค่าพารามิเตอร์ของข้อสอบ และพารามิเตอร์ความสามารถ โดยใช้การเทียบมาตราแบบสามพารามิเตอร์กับแบบสอบวัดความถนัดชุด GRE ด้านภาษา จำนวน 4 ชุด แต่ละชุดวัด 2 ด้าน คือ ความเข้าใจในการอ่าน และหลักภาษา การออกแบบการเทียบมาตรามี 2 แบบ คือ เทียบคะแนนรวมทั้งชุดซึ่งมีความเป็นวิวิธพันธ์ (Heterogeneous) และเทียบคะแนนรวมแต่ละด้านซึ่งมีความเป็นเอกพันธ์ (Homogeneous) โดยการออกแบบการรวบรวมข้อมูล เมื่อใช้กลุ่มที่เท่าเทียมกัน และเมื่อใช้แบบสอบร่วม จำนวนกลุ่มตัวอย่างมีตั้งแต่ 2,579 ถึง 4,351 คน

ผลการวิจัยพบว่า การเทียบมาตราแบบทฤษฎีตอบสนองข้อสอบ เมื่อฝ่าฝืนข้อตกลงเกี่ยวกับแบบสอบต้องมิติเดียวนั้น อาจมีผลกระทบ (impact) ต่อการเทียบมาตรา แต่ไม่อาจสืบสันได้ เนื่องจากมีความคล้ายคลึงกัน ระหว่างผลการเทียบเมื่อข้อสอบเป็นเอกพันธ์และวิวิธพันธ์ ดังนั้นผู้วิจัยจึงแนะนำว่า การเทียบมาตราโดยใช้ทฤษฎีตอบสนองข้อสอบมีความแกร่งเพียงพอต่อการฝ่าฝืนการมีมิติเดียว

สำหรับงานวิจัยที่เกี่ยวกับการเทียบมาตราในประเทศไทย มีผู้ศึกษาเกี่ยวกับความถูกต้องในการเทียบมาตรา โดย ชูชีพ พงศ์สมบูรณ์ (2528) ได้ศึกษาเปรียบเทียบรูปแบบการเทียบมาตรา 3 วิธี คือ การเทียบเชิงเส้นตรง การเทียบแบบอิควิเปอร์เซนไตล์ และการเทียบโดยการใช้โค้งลักษณะข้อสอบ จากรูปแบบที่ใช้แบบทดสอบร่วม กับรูปแบบที่ใช้ผู้สอบร่วม เพื่อต้องการเปรียบเทียบประสิทธิภาพและความคงที่ของการเทียบมาตราทั้ง 3 วิธี ใช้กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศสุ่มมาจำนวน 3,721 คน ที่เข้าสอบโครงการตรวจคุณภาพการศึกษาของกรมวิชาการ กระทรวงศึกษาธิการ ในจำนวน 60 ข้อ แบ่งเป็น 2 ฉบับ ฉบับละ 38 ข้อ มีแบบสอบร่วมชนิดแบบสอบร่วมภายใน จำนวน 20 ข้อ ผลการเปรียบเทียบได้ว่าประสิทธิภาพของการเทียบมาตราระหว่างรูปแบบทดสอบร่วม กับรูปแบบที่ใช้ผู้สอบร่วมในแต่ละวิธีประสิทธิภาพไม่แตกต่างกัน ความคงที่ของวิธีการเทียบมาตราโดยรูปแบบที่ใช้ผู้สอบร่วมได้ผลว่า วิธีการเทียบโดยใช้อิควิเปอร์เซนไตล์มีความคงที่มากกว่าการเทียบเชิงเส้นตรง แต่มีความคงที่พอ ๆ กับการเทียบโดยใช้โค้งลักษณะข้อสอบ ความคงที่ของวิธีการเทียบมาตราในรูปแบบที่ใช้แบบสอบร่วมได้ผลเหมือนกับรูปแบบใช้ผู้สอบร่วม

ส่วนภาวิณี ศรีสุขวัฒนานันท์ (2528) ได้เปรียบเทียบผลการใช้รูปแบบการเทียบมาตรา 3 รูปแบบ คือ รูปแบบอิควิเปอร์เซนไตล์ รูปแบบเชิงเส้นตรง รูปแบบ IRT แบบสามพารามิเตอร์ โดยใช้แบบสอบร่วมภายในต่างกัน 3 ขนาด คือ ขนาดร้อยละ 20 (7 ข้อ) ร้อยละ 40 (14 ข้อ) ร้อยละ 60 (21 ข้อ) โดยใช้ข้อสอบทั้งหมด 35 ข้อ

การประเมินผลของการเทียบมาตรา 2 ลักษณะ คือ ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (SEE) ซึ่งเป็นการวิเคราะห์ในเชิงทฤษฎีของรูปแบบ และดัชนีการเปรียบเทียบความแตกต่าง (Index:C) ที่ได้จากวิเคราะห์กลุ่มตัวอย่างสอบทานผล ซึ่งเป็นการวิเคราะห์ความเพียงพอของการเทียบมาตราโดยตรง กับผลการเทียบสุดท้าย ส่วนประชากรและกลุ่มตัวอย่างแยกเป็น 2 กรณี คือ 1) กรณีสอบเลือก ประชากรคือ ผลการสอบรายบุคคลผู้มีคุณสมบัติขั้นต่ำตามกำหนดระเบียบการสอบคัดเลือกที่กำหนดไว้ 15,875 คน สุ่มตัวอย่างสำหรับการเทียบมาตรา 2 กลุ่ม กลุ่มละ 1,500 คน สำหรับสอบทานผลกลุ่มละ 1,500 คน 2) กรณีแบบสอบผลสัมฤทธิ์ ประชากรคือ ผลการสอบรายบุคคลในวิชาที่กำหนดในระดับอุดมศึกษาจำนวน 12,383 คน สุ่มตัวอย่างเหมือนกรณีแรก แบบสอบมีกรณีละ 2 ชุด โดยมีชุดละ 35 ข้อ มีข้อสอบร่วม 21 ข้อ, 14 ข้อ, และ 7 ข้อ โดยมีแบบสอบร่วมมีค่าความเชื่อมั่นเท่าเทียมกันในแต่ละคู่ใช้รูปแบบการเทียบมาตรา 3 รูปแบบ สร้างตารางแปลงคะแนนสมมูลจากแบบสอบชุดที่ 1 ไปยังแบบสอบชุดที่ 2 แล้ววิเคราะห์ความคลาดเคลื่อนมาตราฐานของการเทียบมาตรา ค่าดัชนีประสิทธิภาพสัมพันธ์ การวิเคราะห์กลุ่มสอบทานผล โดยคำนวนค่าดัชนีเปรียบเทียบความแตกต่าง (C) ผลการศึกษาพบว่า การเทียบระดับคะแนน 9 รูปแบบ ได้ผลเป็นที่ยอมรับว่า ความยาวของแบบสอบร่วมมีผลต่อความแม่นยำ และความเพียงพอของการเทียบมาตรา แต่ในกรณีแบบสอบคัดเลือกมีผลไม่ชัดเจน ในด้านมิติเกี่ยวกับรูปแบบการเทียบมาตราที่ได้ผลดีตามลำดับ คือ รูปแบบอิควิเปอร์เซนไตล์ รูปแบบ IRT และเชิงเส้นตรง กรณีการสอบวัดผลสัมฤทธิ์ รูปแบบที่ได้ผลคือ รูปแบบเชิงเส้นตรง รูปแบบอิควิเปอร์เซนไตล์ และรูปแบบ IRT

ต่อมา เรวดี อินทสะระ (2530) ได้เปรียบเทียบความคลาดเคลื่อนของการเทียบมาตราและความเพียงพอจากการใช้รูปแบบอิงทฤษฎีการตอบสนองข้อสอบ กับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ จากกลุ่มตัวอย่างจำนวน 2,823 คน ใช้แบบทดสอบผลสัมฤทธิ์ทางคณิตศาสตร์ จำนวน 150 ข้อ วิเคราะห์โดยใช้โปรแกรม LOGIST 5 นำค่าพารามิเตอร์แยกเป็น 2 ฉบับ ฉบับละ 45 ข้อ ทั้งสองมีข้อสอบร่วม 15 ข้อ ไปทดสอบกับกลุ่มตัวอย่าง แล้วนำมาเทียบระดับคะแนนและทานผลการเทียบมาตราได้ผลว่า

เมื่อเทียบมาตราจากฉบับที่ 2 ไปฉบับที่ 1 เปรียบเทียบความสมมูลของคะแนนทั้งสองได้ว่า ช่วงคะแนน 1 - 18 รูปแบบ IRT มีคะแนนสมมูลสูงกว่ารูปแบบเทคนิคการวิเคราะห์องค์ประกอบ ช่วงคะแนน 19 - 23 ทั้งสองรูปแบบมีคะแนนสมมูลใกล้เคียงกัน ช่วงคะแนน 24 - 45 รูปแบบเทคนิควิเคราะห์องค์ประกอบมีคะแนนสมมูลสูงกว่ารูปแบบ IRT เมื่อตรวจสอบผลได้ว่าการทานผลการเทียบระดับคะแนนแบบอิงทฤษฎีตอบสนองข้อสอบ มีดัชนีความเพียงพออยู่ในระดับความพอใจ ส่วนรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ มีสัดส่วนความคลาดเคลื่อนมาตรฐานทั้งสองฉบับสูงกว่ารูปแบบอิงทฤษฎีตอบสนองข้อสอบ

อาภรณ์ กาญจนกิจโสภณ (2531) ได้ศึกษาการสร้างแบบทดสอบและตารางเทียบมาตราตามแนวตั้ง ในวิชาคณิตศาสตร์ เรื่องสมการและอสมการ โดยวิธีเทียบคะแนนแบบราส์ช สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น กลุ่มตัวอย่างที่ใช้เทียบคะแนนเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1,2 และ 3 จำนวนระดับชั้นละ 490, 482 และ 415 คน ตามลำดับ ผู้วิจัยได้นำแบบสอบระดับชั้นละ 90 ข้อ ไปทดสอบกับกลุ่มตัวอย่างอื่น แล้วคัดข้อสอบที่มีคุณภาพ เรียงตามความยากที่ต้องการระดับชั้นละ 40,30 และ 30 ข้อ โดยแต่ละชั้นจะมีข้อสอบร่วมระหว่างชั้นอยู่ 10 ข้อ หลังจากนั้นนำไปทดสอบกับกลุ่มตัวอย่างที่ใช้เทียบคะแนนและสร้างตารางเทียบคะแนน เพื่อใช้ศึกษาพัฒนาการของนักเรียนกลุ่มเดิมที่ผ่านชั้นไปเรียนในชั้นที่สูงขึ้นว่า นักเรียนมีความก้าวหน้าเป็นไปตามปกติหรือเบี่ยงเบนไปจากปกติ

จากการศึกษาเอกสารและงานวิจัยที่ผ่านมาพบว่า ไม่มีวิธีการเทียบมาตราวิธีใดเพียงวิธีเดียวที่ให้ผลที่ดีที่สุดในทุกสถานการณ์ ดังนั้นในทางปฏิบัติจึงจำเป็นต้องศึกษาข้อตกลงต่าง ๆ ของวิธีการเทียบมาตราแต่ละวิธี เช่น ลักษณะการแจกแจงของคะแนนที่ได้จากแบบสอบต่างชุดกัน ลักษณะของแบบสอบ ลักษณะของกลุ่มตัวอย่าง เป็นต้น เพื่อเป็นแนวทางในการตัดสินใจเลือกวิธีการในการเทียบมาตราที่เหมาะสม ในแต่ละสถานการณ์ที่ต้องการเทียบ จึงพอสรุปได้ดังนี้

1. เมื่อแบบสอบมีความยากต่างกัน แต่กลุ่มตัวอย่างเป็นกลุ่มสมมูลแล้ว วิธีการของรูปแบบลิควิเปอร์เซนไตล์ ซึ่งในความสัมพันธ์ของการแปลงที่ไม่เป็นเส้นตรงให้ความเหมาะสมต่อการแปลงได้ แต่ถ้าไม่เป็นกลุ่มสมมูล บางครั้งอาจเกิดจากการแจกแจงคะแนนที่ต่างมาก จนทำให้เกิดความคลาดเคลื่อนที่ทำให้ผลการเทียบขาดความเที่ยงพอของการสร้างคะแนนสมมูล

2. เมื่อแบบสอบมีความคล้ายคลึงในระดับความยาก และกลุ่มตัวอย่างเป็นกลุ่มที่สมมูล ควรใช้วิธีการของรูปแบบเชิงเส้นตรง ซึ่งสะดวกและให้ความแน่นอนกว่า

3. การเทียบมาตราโดยทฤษฎีดั้งเดิม ในกรณีแบบสอบมีความยากแตกต่างกันและกลุ่มตัวอย่างมีความสามารถต่างกันมากแล้ว การใช้วิธีลิควิเปอร์เซนไตล์ดีกว่าการเทียบเชิงเส้นตรง

4. เมื่อนำแบบสอบที่มีความยากแตกต่างกันมาเทียบ กลุ่มตัวอย่างไม่สมมูลรูปแบบทฤษฎีตอบสนองข้อสอบสามพารามิเตอร์น่าจะให้ผลเป็นที่น่าพอใจกว่าวิธีอื่น

5. เมื่อกลุ่มตัวอย่างค่อนข้างน้อย และแบบสอบวัดในสิ่งเดียวกันแล้ว วิธีการเชิงเส้นตรง และวิธีใช้ทฤษฎีตอบสนองข้อสอบชนิดหนึ่งพารามิเตอร์ให้ผลเป็นที่น่าพอใจกว่าวิธีลิควิเปอร์เซนไตล์

จากข้อสรุปข้างต้น จะเห็นว่า ในกรณีที่นำแบบสอบที่มีความยากแตกต่างกันมาสอบกับกลุ่มตัวอย่างที่มีระดับความสามารถแตกต่างกัน แล้วนำผลที่ได้มาเทียบมาตรากัน การเทียบมาตราโดยใช้รูปแบบลิงทฤษฎีตอบสนองข้อสอบ แบบ 3 พารามิเตอร์ น่าจะให้ผลเป็นที่น่าพอใจกว่าวิธีอื่น ดังนั้นผู้วิจัยจึงทำการศึกษาคุณภาพของการเทียบมาตราโดยใช้รูปแบบลิงทฤษฎีการตอบสนองข้อสอบ แบบ 3 พารามิเตอร์