

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การนำเสนอเอกสารและงานวิจัยที่เกี่ยวข้อง ได้นำเสนอแยกเป็นหัวข้อดังนี้

1. ตอนที่ 1 การปรับเทียบมาตรา
 - 1.1 ความหมายของการปรับเทียบมาตรา
 - 1.2 เงื่อนไขของการปรับเทียบมาตรา
 - 1.3 ประเภทของการปรับเทียบมาตรา
 - 1.3.1 การปรับเทียบมาตราแนวนอน (Horizontal Equating)
 - 1.3.2 การปรับเทียบมาตราแนวตั้ง (Vertical Equating)
 - 1.4 การออกแบบการปรับเทียบมาตรา (Equating Designs)
 - 1.4.1 การออกแบบวิธีการเก็บรวบรวมข้อมูล
 - 1.4.2 การออกแบบวิธีการปรับเทียบมาตรา
 - 1.5 แบบทดสอบร่วมในการปรับเทียบมาตรา
 - 1.5.1 ประเภทของแบบทดสอบร่วม
 - 1.5.2 ความยาวของแบบทดสอบร่วม
2. ตอนที่ 2 การประเมินคุณภาพของการปรับเทียบมาตรา
 - 2.1 ความคลาดเคลื่อนของการปรับเทียบมาตรา
 - 2.2 ความเพียงพอของการปรับเทียบมาตรา
 - 2.3 ปัจจัยที่ส่งผลต่อคุณภาพของการปรับเทียบมาตรา
3. ตอนที่ 3 งานวิจัยที่เกี่ยวข้องกับการปรับเทียบมาตรา

ตอนที่ 1 การปรับเทียบมาตรา

1. ความหมายของการปรับเทียบมาตรา

ความหมายของการปรับเทียบมาตรา ได้มีผู้ให้คำนิยามไว้หลายคน ได้แก่

ศิริชัย กาญจนวาสี (2541, หน้า 2) ได้นิยามการปรับเทียบมาตรานี้ว่าเป็นกระบวนการแปลงคะแนนของแบบทดสอบต่างฉบับ ที่มุ่งวัดลักษณะเดียวกันให้เป็นคะแนนที่สมมูลกัน เพื่อให้คะแนนสามารถเปรียบเทียบกันได้โดยตรง

กัลลิกเซน (Gulliksen, 1950) ได้นิยามการปรับเทียบมาตรฐานว่า หมายถึง วิธีการปรับคะแนนที่ได้จากแบบทดสอบสองฉบับที่วัดเนื้อหาเดียวกัน ให้เป็นคะแนนสมมูล (Equivalent Scores) ที่เปรียบเทียบกันได้โดยตรง และกัลลิกเซนได้เสนอวิธีการง่าย ๆ โดยให้ผู้สอบกลุ่มเดียวกันทำแบบทดสอบทั้งสองฉบับที่พยายามสร้างให้คู่ขนานกัน แล้วคำนวณคะแนนมาตรฐานของแต่ละฉบับ คะแนนแปลงที่ตรงกับคะแนนมาตรฐานเดียวกัน ถือว่าเป็นคะแนนที่สมมูลกัน สามารถนำมาเปรียบเทียบกันได้โดยตรง

ฟลานาแกน (Flanagan, 1951, pp. 747 - 748) ได้ให้นิยามเชิงทฤษฎีของการปรับเทียบมาตรฐานไว้ว่า หมายถึง วิธีการปรับคะแนนที่ได้จากแบบทดสอบต่างฉบับกัน ให้สามารถนำมาเปรียบเทียบกันได้ โดยแบบทดสอบต่างฉบับกันนั้นจะต้องมีการแจกแจงของคะแนนจริงในประชากรเท่ากัน หรือมีความเที่ยงเท่ากันเสียก่อน จึงจะสามารถนำคะแนนดิบมาเปรียบเทียบกันได้ ในทางปฏิบัติ ฟลานาแกน ได้เสนอให้ตรวจสอบจากคะแนนเฉลี่ยของแบบทดสอบที่ได้จากกลุ่มตัวอย่างขนาดใหญ่ ถ้าคะแนนเฉลี่ยจากแบบทดสอบไม่แตกต่างกันอย่างมีนัยสำคัญในทุกกลุ่มตัวอย่างขนาดใหญ่ใด ๆ แล้ว จึงจะสามารถนำคะแนนมาเปรียบเทียบกันได้ แต่ถ้าคะแนนเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญ ฟลานาแกน เสนอให้ใช้วิธีบวกหรือลบคะแนนด้วยจำนวนที่แตกต่าง ก่อนนำมาเทียบกัน หรืออาจใช้เทคนิคการปรับเทียบด้วยวิธีการสร้างสมการถดถอย วิธีปรับด้วยคะแนนมาตรฐาน หรือ วิธีอีควิเปอร์เซ็นไทล์

ลอร์ด (Lord, 1980) ได้นิยามการปรับเทียบมาตรฐานว่า เป็นวิธีการแปลงคะแนนของแบบทดสอบต่างฉบับ (Forms) ให้มีความหมายที่สับเปลี่ยนกันได้ เพื่อให้เกิดความเสมอภาคระหว่างผู้สอบ

แองกอฟฟ์ (Angoff, 1982) ได้นิยามการปรับเทียบมาตรฐานว่า เป็นกระบวนการแปลงระบบคะแนนของแบบทดสอบฉบับหนึ่งไปสู่ระบบคะแนนของแบบทดสอบอีกฉบับหนึ่ง ซึ่งวัดสิ่งเดียวกัน เพื่อให้คะแนนแปลงจากแบบทดสอบทั้งสองฉบับสามารถเทียบเท่ากัน (Equivalent) และใช้แทนกันได้

ปีเตอร์เซน และคณะ (Petersen & Others, 1989) ได้นิยามการปรับเทียบมาตรฐานว่าเป็นกระบวนการเชิงประจักษ์สำหรับสร้างระบบความสัมพันธ์ระหว่างคะแนนของแบบทดสอบต่างฉบับให้สามารถแปลงคะแนนของแบบทดสอบฉบับหนึ่งไปสู่คะแนนของแบบทดสอบอีกฉบับหนึ่งได้

โคเลน และเบรนแนน (Kolen & Brennan, 1995, p. 3) ได้นิยามการปรับเทียบมาตรฐานว่า เป็นกระบวนการทางสถิติที่ใช้ในการปรับคะแนนจากแบบทดสอบต่างฉบับ เพื่อให้คะแนนสามารถใช้แทนกัน เมื่อแบบทดสอบแต่ละฉบับมีความยากแตกต่างกัน แต่ถ้ามีเนื้อหาต่างกันจะไม่

สามารถปรับเทียบกันได้ ดังคำกล่าวที่ว่า “Equating Adjusts to Differences in Difficulty, not for Differences in Content”

จากนิยามที่กล่าวมาข้างต้น สามารถสรุปได้ว่า การปรับเทียบมาตรา เป็นการปรับคะแนนจากแบบทดสอบฉบับหนึ่งไปสู่แบบทดสอบอีกฉบับหนึ่งให้เป็นคะแนนที่สมมูลกัน และสามารถเปรียบเทียบกันได้

2. เงื่อนไขของการปรับเทียบมาตรา

การปรับเทียบมาตราระหว่างแบบทดสอบสามารถกระทำได้ภายใต้เงื่อนไขบางประการ ซึ่งมีผู้เสนอเงื่อนไขต่าง ๆ ที่จำเป็นดังนี้

กัลลิกเซน (Gulliksen, 1950) ได้เสนอว่าแบบทดสอบที่จะนำมาปรับเทียบมาตรากันได้นั้น ต้องเป็นแบบทดสอบคู่ขนานกัน ซึ่งสร้างโดยการจับคู่ข้อสอบที่เหมือนกัน แล้วสุ่มแยกเป็นแบบทดสอบคู่ขนาน 2 ฉบับ

เลวาย (Levine, 1955) ได้เสนอว่า แบบทดสอบ 2 ฉบับที่ต้องการนำมาเปรียบเทียบให้อยู่ในมาตราเดียวกันนั้น ต้องเป็นแบบทดสอบคู่ขนานกันในด้าน

- 1) เนื้อหา (Subject Matter)
- 2) โครงสร้าง (Structure)
- 3) รูปแบบ (Formats)
- 4) ชนิดของข้อสอบ (Item Type)
- 5) เวลาที่ใช้สอบ (Timing)

ลอร์ด (Lord, 1977) ได้เสนอแนวคิดว่าการปรับเทียบมาตราระหว่างแบบทดสอบจะต้องก่อให้เกิดผลที่ไม่แตกต่างกับผู้สอบ ไม่ว่าผู้สอบจะได้รับการสอบจากแบบทดสอบฉบับใดก็ตาม ซึ่งต่อมอลอร์ด (Lord, 1980, pp. 195-198) ได้พัฒนาหลักการของความเสมอภาค (Equity) ในการปรับเทียบมาตราระหว่างแบบทดสอบดังนี้

การปรับเทียบมาตราระหว่างแบบทดสอบ X กับแบบทดสอบ Y จะถือว่ามีความเสมอภาคสำหรับผู้สอบก็ต่อเมื่อ ไม่ว่าผู้สอบจะมีความสามารถระดับใดก็ตาม เมื่อสอบแบบทดสอบฉบับ X หรือ Y แล้วจะต้องไม่ก่อให้เกิดผลที่แตกต่างกันกับผู้สอบ” นั่นคือคะแนนที่ผ่านการปรับเทียบแล้วจะต้องมีการแจกแจงเหมือนกับคะแนนที่กลุ่มนั้นทำได้จริง หลักการดังกล่าวสะท้อนให้เห็นเงื่อนไขของการปรับเทียบมาตราระหว่างแบบทดสอบที่สามารถทำได้อย่างมีประสิทธิภาพ เมื่อแบบทดสอบฉบับต่าง ๆ ที่ต้องการนำมาปรับเทียบมาตรากัน ต้องมีคุณสมบัติดังนี้

- 1) วัดคุณลักษณะเดียวกัน

- 2) มีค่าความเที่ยงเท่ากัน
- 3) มีระดับความยากเท่ากัน
- 4) การแจกแจงคะแนนก่อนแปลงและหลังแปลงเหมือนกัน
- 5) เป็นแบบทดสอบคู่ขนาน

แองกอฟฟ์ (Angoff, 1971) และลอร์ด (Lord, 1980) ได้เสนอคุณสมบัติเพิ่มเติมของการปรับเทียบมาตราระหว่างแบบทดสอบไว้อย่างสอดคล้องกันคือ

1) เป็นการปรับเทียบมาตราที่ไม่ขึ้นกับกลุ่มผู้สอบ (Invariance Across Groups) กล่าวคือ คะแนนที่ปรับเทียบแล้วไม่แปรผัน ไม่ว่าจะปรับเทียบจากกลุ่มใด จึงสามารถนำผลไปใช้กับกลุ่มอื่นๆ ได้

2) เป็นการปรับเทียบมาตราที่สมมาตร (Symmetry) กล่าวคือ ผลการปรับเทียบมาตราจะต้องเหมือนกัน ไม่ว่าเป็นการปรับเทียบมาตราจากแบบทดสอบฉบับที่ 1 ไปยังแบบทดสอบฉบับที่ 2 หรือปรับเทียบมาตราจากแบบทดสอบฉบับที่ 2 ไปยังแบบทดสอบที่ 1

แองกอฟฟ์ (Angoff, 1984) ได้เสนอแนะว่า แบบทดสอบที่จะนำมาปรับเทียบมาตรากันได้นั้น ต้องเป็นแบบทดสอบที่วัดคุณลักษณะเดียวกัน และใช้วัดกลุ่มผู้สอบที่มีระดับความสามารถใกล้เคียงกัน

ดังนั้นจึงพอสรุปได้ว่าทั้งแบบทดสอบและวิธีการปรับเทียบมาตราต้องมีลักษณะพิเศษบางประการ โดยแบบทดสอบฉบับที่ต้องการนำมาปรับเทียบมาตรากันต้องมีคุณลักษณะที่สำคัญดังนี้

- 1) วัดคุณลักษณะเด่นเพียงคุณลักษณะเดียวเหมือนกัน (Unidimensionality)
- 2) มีลักษณะของความเป็นคู่ขนานในด้าน
 - 2.1) เนื้อหา
 - 2.2) โครงสร้าง
 - 2.3) รูปแบบ
 - 2.4) ชนิดของข้อสอบ
 - 2.5) เวลาที่ใช้สอบ

3) มีค่าความเที่ยงสูงทัดเทียมกัน

นอกจากนี้ วิธีการปรับเทียบมาตราระหว่างแบบทดสอบต้องเป็นวิธีการที่มีคุณสมบัติที่สำคัญดังนี้

- 1) ความเสมอภาค
- 2) ความไม่ผันแปรตามกลุ่มผู้สอบ

3) ความสมมาตร

3. ประเภทของการปรับเทียบมาตรา

การทดสอบที่มีความจำเป็นต้องใช้เทคนิคการปรับเทียบมาตราระหว่างแบบทดสอบสามารถจำแนกได้เป็น 2 สถานการณ์ (ศิริชัย กาญจนวาสี, 2541, หน้า 5 อ้างอิงจาก Hambleton & Swaminathan, 1985) ซึ่งต้องใช้รูปแบบการปรับเทียบมาตรา (Equating Forms) 2 รูปแบบ ดังนี้

3.1 การปรับเทียบมาตราแนวนอน

การปรับเทียบมาตราแนวนอนเป็นการปรับเทียบมาตราระหว่างแบบทดสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน มีระดับความยากใกล้เคียงกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ในประชากรเดียวกัน หรือมีความสามารถใกล้เคียงกัน

การปรับเทียบมาตราแนวนอนเป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบทดสอบเนื้อหาเดียวกันขึ้นมาหลาย ๆ ฉบับ ด้วยเหตุผลของการนำไปใช้ในการทดสอบให้เกิดความยุติธรรมและป้องกันความลับของข้อสอบเมื่อใช้ต่างเวลากันสำหรับผู้สอบกลุ่มขนาดใหญ่ เพื่อปรับเทียบว่าคะแนนที่ได้จากฉบับหนึ่งเทียบเป็นเท่าไรของอีกฉบับหนึ่งซึ่งวัดในระดับเดียวกัน จึงเป็นการปรับเทียบมาตราระหว่างแบบทดสอบต่างฉบับของวิชาเดียวกันสำหรับกลุ่มผู้สอบระดับชั้นเดียวกัน เช่น การปรับเทียบมาตราระหว่างแบบทดสอบวิชาคณิตศาสตร์ 2 ฉบับ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 (ดังภาพที่ 1) เป็นต้น

แบบทดสอบต่างฉบับที่ต้องการนำมาปรับเทียบมาตราแนวนอนนั้น ถึงแม้แบบทดสอบเหล่านั้นยากที่จะมีความเท่าเทียมกันในทุกด้าน แต่ก็ควรเป็นแบบทดสอบที่พยายามสร้างหรือมุ่งให้คู่ขนานกัน (Alternate Forms) นอกจากนี้การแจกแจงความสามารถของผู้สอบที่ทำการสอบด้วยแบบทดสอบแต่ละฉบับ ควรมีการแจกแจงที่ประมาณได้ว่ามีระดับความสามารถที่ใกล้เคียงกัน



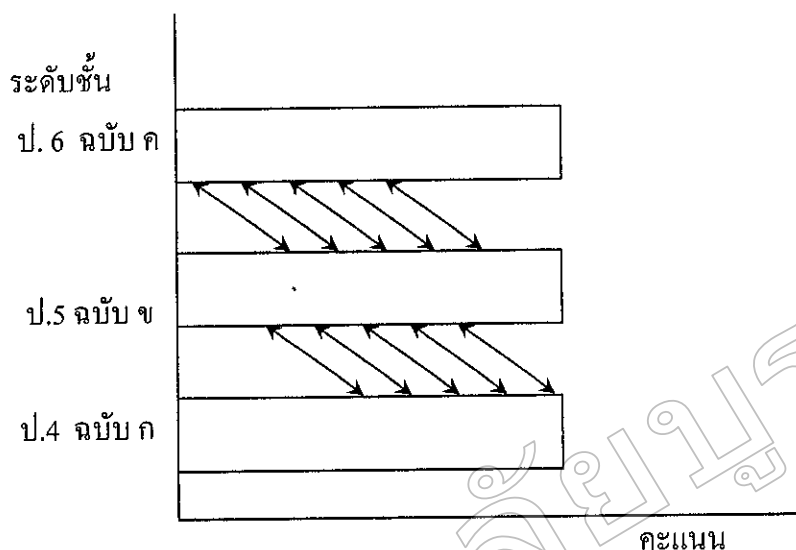
ภาพที่ 1 การเปรียบเทียบมาตรฐานแนวนอน

3.2 การเปรียบเทียบมาตรฐานแนวนอน

การเปรียบเทียบมาตรฐานแนวนอน เป็นการเปรียบเทียบมาตรฐานระหว่างแบบทดสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน แต่มีระดับความยากแตกต่างกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ต่างประชากรกัน หรือมีความสามารถแตกต่างกัน

การเปรียบเทียบมาตรฐานแนวนอนเป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบทดสอบเนื้อหาเดียวกัน แต่ต่างฉบับต่างมุ่งวัดความสามารถของผู้สอบที่ต่างระดับกัน เพื่อเปรียบเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่งเทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน จึงเป็นการเปรียบเทียบมาตรฐานระหว่างแบบทดสอบต่างระดับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบต่างระดับชั้นกัน เช่น การเปรียบเทียบมาตรฐานระหว่างแบบทดสอบคณิตศาสตร์ชั้น ป.4, ป.5 และ ป.6 (ดังภาพที่ 2) เป็นต้น

แบบทดสอบต่างฉบับที่วัดเนื้อหาเดียวกัน แต่ต่างระดับกันที่ต้องการนำมาเปรียบเทียบมาตรฐานนั้น แบบทดสอบแต่ละฉบับจึงมีระดับความยากแตกต่างกันไป รวมทั้งกลุ่มผู้สอบแต่ละฉบับมีการแจกแจงความสามารถอยู่ต่างประชากร หรือมีการแจกแจงความสามารถที่อยู่ในระดับที่แตกต่างกัน ดังนั้นการเปรียบเทียบมาตรฐานแนวนอนจึงมีความสลับซับซ้อนกว่าการเปรียบเทียบมาตรฐานแนวนอนทั้งในแง่ทฤษฎี และการปฏิบัติ



ภาพที่ 2 การปรับเทียบมาตราแนวดัง

4. การออกแบบการปรับเทียบมาตรา

แบบทดสอบต่างฉบับกันสามารถนำไปใช้ทดสอบเพื่อเก็บรวบรวมข้อมูลจากกลุ่มผู้สอบได้หลากหลายรูปแบบ แต่ถ้าการนำแบบทดสอบไปใช้ทดสอบกับกลุ่มผู้สอบโดยไม่ได้รับการออกแบบที่เหมาะสม คะแนนที่ได้จากแบบทดสอบต่างฉบับกันนั้นย่อมไม่สามารถนำมาเปรียบเทียบกันได้อย่างมีความหมาย ตัวอย่างเช่น

เมื่อนำแบบทดสอบฉบับหนึ่ง (X) มาทำการทดสอบกับผู้สอบกลุ่มหนึ่ง และนำแบบทดสอบอีกฉบับหนึ่ง (Y) ไปทำการทดสอบกับผู้สอบอีกกลุ่มหนึ่ง ถ้าเราไม่ทราบลักษณะของแบบทดสอบทั้งสอง ฉบับและผู้สอบทั้งสองกลุ่ม เราย่อมไม่สามารถหาทศนิยมเปรียบเทียบคุณภาพข้อสอบจากฉบับหนึ่งกับอีกฉบับหนึ่งได้ รวมทั้งย่อมไม่สามารถหามาตรการปรับเทียบคะแนนสอบฉบับหนึ่งของผู้สอบกลุ่มหนึ่งกับคะแนนสอบอีกฉบับหนึ่งของผู้สอบอีกกลุ่มหนึ่งได้เลย

ลอร์ด (Lord, 1975) ได้เสนอแนวคิดในการเก็บรวบรวมข้อมูลสำหรับการปรับเทียบมาตราระหว่างแบบทดสอบว่า การปรับเทียบมาตราระหว่างแบบทดสอบจะมีความเป็นไปได้ เมื่อมีการเก็บรวบรวมข้อมูลในกรณีใดกรณีหนึ่ง ดังต่อไปนี้

1. ผู้สอบกลุ่มหนึ่งและผู้สอบกลุ่มสองจะต้องมีสมาชิกร่วมกันอยู่จำนวนหนึ่ง
2. ผู้สอบกลุ่มหนึ่ง และผู้สอบกลุ่มสองต่างเป็นตัวแทนของผู้สอบที่มาจากประชากรเดียวกัน

3. แบบทดสอบฉบับหนึ่ง และแบบทดสอบอีกฉบับหนึ่งมีข้อสอบร่วมกันอยู่จำนวน

หนึ่ง กล่าวคือ การเปรียบเทียบมาระหว่างแบบทดสอบจะกระทำได้อีกต่อเมื่อมีการออกแบบการทดสอบเพื่อเก็บรวบรวมข้อมูลที่เหมาะสม โดยการทดสอบนั้นจะต้องมี “ข้อสอบร่วม” อยู่ในแบบทดสอบแต่ละฉบับ หรือมี “ผู้สอบร่วม” อยู่ในผู้สอบแต่ละกลุ่ม หรือใช้กลุ่มผู้สอบแต่ละกลุ่มที่เป็น “กลุ่มเท่าเทียมกัน” จึงจะสามารถหาวิธีการทางสถิติมาใช้ในการเปรียบเทียบคะแนนระหว่างแบบทดสอบได้อย่างมีความหมาย

ดังนั้น การเปรียบเทียบมาระหว่างแบบทดสอบ จึงต้องอาศัยเทคนิคสำคัญสองประการ ได้แก่

4.1 การออกแบบวิธีการเก็บรวบรวมข้อมูล (Designs for Data Collection)

การเก็บรวบรวมข้อมูลสำหรับนำผลมาใช้ในการเปรียบเทียบสามารถจำแนกได้ 3 รูปแบบ (Kolen & Brennan, 1995; ศิริชัย กาญจนวาลี, 2541) ดังนี้

4.1.1 แบบแผนกลุ่มเดียว (Single Group Design) แบ่งเป็น 2 รูปแบบ ได้แก่

4.1.1.1 แบบแผนกลุ่มเดียวที่ไม่ได้รับการจัดให้สมดุล (Single Group Design with Uncounterbalancing)

รูปแบบนี้เป็นรูปแบบที่ง่ายที่สุด ใช้ผู้สอบกลุ่มเดียวทำหน้าที่เป็นผู้สอบร่วม วิธีการเก็บรวบรวมข้อมูลกระทำโดยใช้กลุ่มตัวอย่างผู้สอบกลุ่มเดียว แต่ละคนทำแบบทดสอบทั้ง 2 ฉบับ โดยให้ผู้สอบทำแบบทดสอบฉบับหนึ่งแล้วตามด้วยแบบทดสอบอีกฉบับหนึ่ง ดังภาพที่ 3

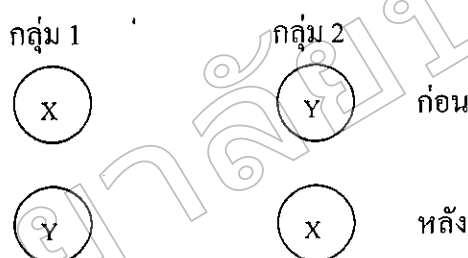


ภาพที่ 3 แบบแผนกลุ่มเดียวที่ไม่ได้รับการจัดให้สมดุล

การให้ผู้สอบกลุ่มเดียวกันทำแบบทดสอบทั้ง 2 ฉบับ ระดับความสามารถของผู้สอบแบบสอบทั้ง 2 ฉบับเท่ากัน จึงไม่เป็นปัญหาแทรกซ้อน รวมทั้งไม่มีปัญหาในเรื่องความยากของแบบทดสอบต่างฉบับ คะแนนจากแบบทดสอบจึงสามารถเปรียบเทียบกันได้บนพื้นฐานความสามารถที่เท่ากันจากคนกลุ่มเดียวกัน แต่ในทางปฏิบัติ การสอบฉบับหลัง ผู้สอบอาจได้รับปัจจัยที่ส่งผลกระทบต่องานจากการสอบฉบับแรก เช่น การเรียนรู้ การฝึกฝน ความเมื่อยล้าจากการทำแบบทดสอบฉบับแรก เป็นต้น ซึ่งอาจมีอิทธิพลต่อการเปรียบเทียบมาตรา

4.1.1.2 แบบแผนกลุ่มเดียวที่ได้รับการจัดให้สมดุล (Single Group Design with Counterbalancing)

รูปแบบนี้ปรับมาจากรูปแบบแรก เพื่อขจัดผลของลำดับการทดสอบก่อน-หลัง โดยการสุ่มผู้สอบแยกเป็น 2 กลุ่มย่อย แต่ละกลุ่มย่อยได้รับการทดสอบด้วยแบบทดสอบทั้ง 2 ฉบับ ในลักษณะที่กลุ่มย่อยหนึ่งให้ทำแบบทดสอบฉบับที่ 1 แล้วตามด้วยแบบทดสอบฉบับที่ 2 ส่วนอีกกลุ่มย่อยหนึ่งให้ทำแบบทดสอบฉบับที่ 2 แล้วตามด้วยแบบทดสอบฉบับที่ 1 สลับกันเพื่อให้เกิดความสมดุล ดังภาพที่ 4

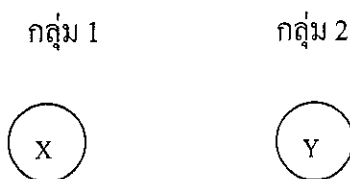


ภาพที่ 4 แบบแผนกลุ่มเดียวที่ได้รับการจัดให้สมดุล

การใช้รูปแบบนี้ กลุ่มผู้สอบจะได้รับอิทธิพลจากลำดับการทดสอบ การเรียนรู้ การฝึกฝน และความเมื่อยล้าที่สมดุลกันในทั้งสองกลุ่มย่อย

4.1.2 แบบแผนกลุ่มสมมูล (Equivalent Group Design)

รูปแบบนี้ใช้กลุ่มผู้สอบที่มีความสามารถเท่าเทียมกัน โดยการจับกลุ่มผู้สอบให้มีความคล้ายคลึงกันมากที่สุดเท่าที่จะทำได้ แล้วให้ผู้สอบในแต่ละกลุ่มทำแบบทดสอบฉบับเดียว ดังภาพที่ 5



ภาพที่ 5 แบบแผนกลุ่มสมมูล

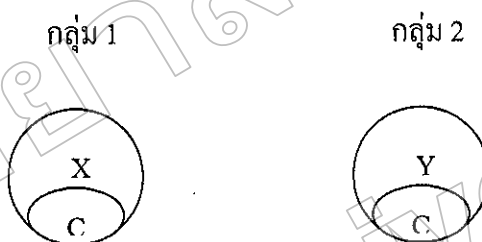
แม้การจัดกลุ่มจะไม่เหมือนกันทุกประการ (Identical Groups) แต่ก็ควรมีลักษณะทั่วไปคล้ายกันมากที่สุด กลุ่มดังกล่าวอาจได้มาโดยการสุ่ม (Random) ข้อดีของแบบแผนนี้คือ แต่ละกลุ่มทำแบบทดสอบฉบับเดียว จึงสามารถหลีกเลี่ยงปัญหาเรื่องการเรียนรู้ การฝึกฝน และความ

เมื่อพิจารณาจากการทำแบบทดสอบฉบับแรก แต่ปัญหาที่ตามมาคือ กลุ่มที่ใช้ไม่เหมือนกัน อาจมีการแจกแจงความสามารถที่แตกต่างกัน และไม่มีข้อมูลที่น่ามาใช้ปรับความแตกต่างของกลุ่ม ความแตกต่างที่เกิดขึ้นแม้เพียงเล็กน้อยย่อมส่งผลกระทบต่อความลำเอียงในการเปรียบเทียบได้ วิธีการที่จะช่วยลดความแตกต่างระหว่างกลุ่มได้โดยการให้กลุ่มตัวอย่างกลุ่มขนาดใหญ่

4.1.3 แบบแผนแบบทดสอบร่วม (Anchor Test Design) แบ่งเป็น 2 รูปแบบ ได้แก่

4.1.3.1 แบบแผนแบบทดสอบร่วมภายใน (Internal Anchor Test Design)

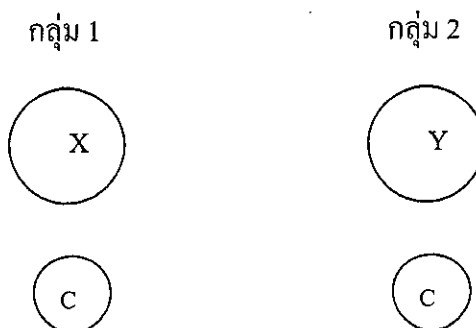
รูปแบบนี้เป็นการใช้กลุ่มผู้สอบสองกลุ่ม โดยแต่ละกลุ่มทำแบบทดสอบคนละชุด โดยที่แบบทดสอบแต่ละชุดจะมีแบบทดสอบร่วมรวมอยู่ด้วย ดังภาพที่ 6



ภาพที่ 6 แบบแผนแบบทดสอบร่วมภายใน

4.1.3.2 แบบแผนแบบทดสอบร่วมภายนอก (External Anchor Test Design)

รูปแบบนี้เป็นการใช้กลุ่มผู้สอบสองกลุ่ม แล้วให้ผู้สอบแต่ละกลุ่มทำแบบทดสอบคนละชุดและต้องทำแบบทดสอบร่วมแยกอีกหนึ่งฉบับ ดังภาพที่ 7



ภาพที่ 7 แบบแผนแบบทดสอบร่วมภายนอก

4.2 การออกแบบวิธีการปรับเทียบมาตรา (Designs for Equating Methods)

การออกแบบวิธีการปรับเทียบมาตราเป็นการกำหนดเงื่อนไข และกฎเกณฑ์การแปลงคะแนนจากแบบทดสอบต่างชุดที่ต้องการนำมาปรับเทียบกันให้สามารถเปรียบเทียบกันได้อย่างมีความหมาย วิธีที่ใช้ในการปรับเทียบมาตรามีหลายวิธี สามารถจำแนกตามแนวคิดพื้นฐานของการปรับเทียบได้เป็นรูปแบบหลักสองรูปแบบ (ศิริชัย กาญจนวาสี, 2541, หน้า 11) ได้แก่

4.2.1 รูปแบบการปรับเทียบโดยใช้ทฤษฎีดั้งเดิม (Classical Models of Equating)

วิธีที่นิยมใช้กันทั่วไปมี 3 วิธี ดังนี้

4.2.1.1 วิธีการปรับเทียบมาตราแบบอีควิเปอร์เซ็นต์ไทล์ (Equipercentile Equating)

นิยาม คะแนนจากแบบทดสอบ 2 ฉบับ ฉบับ X และ Y (เมื่อแบบทดสอบ X และ Y วัดสิ่งเดียวกันและมีความเที่ยงสูงพอ) จะถือว่าเป็นคะแนนสมมูลกัน (Equivalent Scores) เมื่อคะแนนของแต่ละฉบับนั้นมีตำแหน่งเปอร์เซ็นต์ไทล์เท่ากัน (Flanagan, 1951)

นั่นคือ X_i และ Y_j เป็นคะแนนสมมูลกัน เมื่อ $X_{PRK} = Y_{PRK}$

โดย X_{PRK} คือ คะแนนจากแบบทดสอบฉบับ X ตรงตำแหน่งเปอร์เซ็นต์ไทล์ที่ k

Y_{PRK} คือ คะแนนจากแบบทดสอบฉบับ Y ตรงตำแหน่งเปอร์เซ็นต์ไทล์ที่ k

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

เมื่อ L_o = คะแนนขีดจำกัดล่างที่แท้จริงของคะแนนในชั้นเปอร์เซ็นต์ไทล์ที่ k

n = จำนวนสมาชิกในกลุ่มตัวอย่าง

f = ความถี่ของคะแนนในชั้นเปอร์เซ็นต์ไทล์ที่ k

F_b = ความถี่สะสมก่อนถึงคะแนนในชั้นเปอร์เซ็นต์ไทล์ที่ k

c = อัตรากาชัน

วิธีการปรับเทียบมาตราระหว่างแบบทดสอบ X กับแบบทดสอบ Y โดยการใช้วิธีอีควิเปอร์เซ็นต์ไทล์ กระทำได้ดังนี้

1. สุ่มผู้สอบให้เป็นตัวแทนประชากรเป้าหมายที่ต้องการนำแบบทดสอบไปใช้ โดยให้ได้กลุ่มผู้สอบที่มีความสามารถกระจายครอบคลุมลักษณะประชากรเป้าหมาย แล้วทำการสุ่มแยกผู้สอบออกเป็น 2 กลุ่ม โดยให้กลุ่มหนึ่งทำแบบทดสอบฉบับ X และอีกกลุ่มหนึ่งทำแบบทดสอบฉบับ Y

2. นำคะแนนของแบบทดสอบแต่ละฉบับมาคำนวณค่าคะแนน ณ ตำแหน่งเปอร์เซ็นต์ไทล์ต่าง ๆ

3. สร้างตารางเปรียบเทียบคะแนนระหว่างแบบทดสอบ X กับ Y โดยถือว่าคะแนน ณ ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกันของแบบทดสอบทั้ง 2 ฉบับ เป็นคะแนนสมมูลกัน

วิธีอิกวิเปอร์เซ็นต์ไทล์เป็นการสร้างคะแนนสมมูลระหว่างแบบทดสอบบนพื้นฐานความเท่าเทียมกันของค่าโมเมนต์ทั้ง 3 ทางสถิติ ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และการแจกแจงของคะแนน จึงเป็นกระบวนการแปลงคะแนนแบบไม่ใช่เส้นตรง (Nonlinear Transformation) ยังผลให้ความสัมพันธ์ระหว่างคะแนนดิบกับคะแนนสมมูลไม่เป็นเส้นตรง ในการนำวิธีเปรียบเทียบคะแนนนี้ไปใช้มีข้อแนะนำพอสรุปได้ดังนี้ (ศิริชัย กาญจนวาสี, 2541 อ้างอิงจาก Potthoff, 1982; Angoff, 1984)

1. ควรใช้กลุ่มผู้สอบขนาดใหญ่ ควรเป็นกลุ่มที่มีความสามารถค่อนข้างกระจาย และกระจายพอ ๆ กัน ถ้าใช้กลุ่มตัวอย่างขนาดเล็กจะทำให้คะแนนมีความไวต่อความแปรปรวนเชิงสุ่ม ผลการเปรียบเทียบคะแนนสามารถแปรผันไปตามกลุ่มผู้สอบได้ (Group Dependent)

2. การสร้างกราฟเปอร์เซ็นต์ไทล์ เพื่อเปรียบเทียบคะแนนระหว่างแบบทดสอบ ควรสร้างด้วยความระมัดระวัง และไม่ลำเอียง

3. แบบทดสอบที่นำมาเปรียบเทียบคะแนนกัน ควรมีความเที่ยงสูงพอกัน ถ้าต่างกันมาก จะทำให้การเปรียบเทียบคะแนนขาดความคงที่

4. การเปรียบเทียบคะแนนควรอยู่ในช่วงพิสัยของคะแนนสังเกตได้ การเปรียบเทียบคะแนนที่อยู่นอกพิสัยของคะแนนสังเกตได้จะมีความคลาดเคลื่อนสูง

4.2.1.2 วิธีการปรับเทียบมาตราเชิงเส้นตรง (Linear Equating)

นิยาม คะแนนจากแบบทดสอบสองฉบับ ฉบับ X และ Y (เมื่อแบบทดสอบ X และ Y วัดสิ่งเดียวกันและมีความเที่ยงสูงพอกัน) จะถือว่าเป็นคะแนนสมมูลกัน (Equivalent Scores) เมื่อคะแนนของแต่ละฉบับมีคะแนนมาตรฐาน (Standard Scores) เท่ากัน (ศิริชัย กาญจนวาสี, 2541, หน้า 13)

นั่นคือ X_i และ Y_i เป็นคะแนนสมมูลกัน เมื่อ $X_{zk} = Y_{zk}$

โดย X_{zk} คือ คะแนนจากแบบทดสอบฉบับ X ตรงตำแหน่งคะแนนมาตรฐาน k

Y_{zk} คือ คะแนนจากแบบทดสอบฉบับ Y ตรงตำแหน่งคะแนนมาตรฐาน k

$$\text{คะแนน ณ ตำแหน่ง } Z_k = \frac{X_i - \bar{X}}{S_x} = \frac{Y_i - \bar{Y}}{S_y}$$

สามารถเขียนเป็นสมการแปลงคะแนนได้ดังนี้

$$(X_i - \bar{X}) \frac{S_y}{S_x} = Y_i - \bar{Y}$$

$$\therefore T_{(xy)} = Y_i = \bar{Y} + (X_i - \bar{X}) \frac{S_y}{S_x}$$

$$= \bar{Y} + S_y S_x$$

$$\text{หรือ } T_{(xy)} = \left[\bar{Y} - \left(\frac{S_y}{S_x} \right) \bar{X} \right] + \left(\frac{S_y}{S_x} \right) X_i$$

$$\therefore T_{(xi)} = B_{yx} + A_{yx} X_i$$

เมื่อ $T_{(xi)}$ = คะแนนแปลงของ X_i ที่มีความสมมูลกับ Y_i

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

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

จากสูตรจะเห็นได้ว่าวิธีปรับเทียบมาตราเชิงเส้นตรงเป็นการแปลงคะแนนจากแบบทดสอบ X ให้สมมูลกับ Y โดยให้คะแนนแปลงของ X มีค่าเฉลี่ยเป็น $\bar{Y}$ และส่วนเบี่ยงเบนมาตรฐานเป็น S_y นั่นเอง

วิธีการปรับเทียบมาตราระหว่างแบบทดสอบ X กับแบบทดสอบ Y โดยการใช้วิธีเชิงเส้นตรง กระทำได้ดังนี้

1. สุ่มกลุ่มผู้สอบให้เป็นตัวแทนประชากรเป้าหมายที่ต้องการนำแบบทดสอบไปใช้ แล้วสุ่มแยกเป็น 2 กลุ่ม โดยให้กลุ่มหนึ่งทำแบบทดสอบฉบับ X และอีกกลุ่มหนึ่งทำแบบทดสอบฉบับ Y

2. นำคะแนนของแบบทดสอบแต่ละฉบับมาคำนวณค่า $\bar{X}$, S_x , $\bar{Y}$, S_y เพื่อคำนวณหาค่าความชันและค่าคงที่

3. สร้างตารางปรับเทียบคะแนนระหว่างแบบทดสอบ X กับ Y โดยแทนค่า X_i ลงในสมการเพื่อหาคะแนนสมมูล Y_i

วิธีการปรับเทียบมาตราเชิงเส้นตรง เป็นการสร้างคะแนนสมมูลระหว่างแบบทดสอบบนพื้นฐานความเท่าเทียมกันของค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน วิธีนี้จึงคำนวณง่าย สะดวกในการนำไปใช้ อาศัยเพียงการคำนวณคะแนนสมมูลเชิงเส้นตรงจากแบบทดสอบหนึ่งไปยังอีกแบบทดสอบหนึ่ง โดยให้มีค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานเท่ากัน

วิธีปรับเทียบมาตราเชิงเส้นตรงนี้อาจกล่าวได้ว่าเป็นกรณีเฉพาะของวิธีปรับเทียบคะแนนอิกวิเปอร์เซ็นไทล์ เนื่องจากว่าถ้าการแจกแจงคะแนนจากแบบทดสอบ X และ Y เหมือนกัน วิธีอิกวิเปอร์เซ็นไทล์ให้ผลการปรับเทียบคะแนนระหว่างแบบทดสอบเป็นเส้นตรง ซึ่งเหมือนกับวิธีการปรับเทียบมาตราเชิงเส้นตรง

4.2.1.3 วิธีการปรับเทียบมาตราโดยใช้สมการถดถอย (Regression Method)

เราสามารถนำเทคนิคการวิเคราะห์สมการถดถอย (Regression Equation) สำหรับปรับเทียบคะแนนระหว่างแบบทดสอบได้เช่นเดียวกัน วิธีการปรับเทียบคะแนนโดยใช้สมการถดถอย กระทำได้ ดังนี้

สร้างสมการทำนายคะแนนของแบบทดสอบฉบับหนึ่ง จากคะแนนของแบบทดสอบอีกฉบับหนึ่ง วิธีการนี้ถือว่า คะแนนจากแบบทดสอบ 2 ฉบับ คะแนน X_i กับ Y_i เป็นคะแนนสมมูลกัน เมื่อคะแนน Y_i เป็นคะแนนที่ทำนายได้จากคะแนน X_i โดยใช้สมการทำนาย ซึ่งมีลักษณะการคำนวณคล้ายกับวิธีการปรับเทียบมาตราเชิงเส้นตรง แต่มีวิธีการเก็บรวบรวมข้อมูลต่างกัน

วิธีการปรับเทียบมาตราระหว่างแบบทดสอบ X กับแบบทดสอบ Y โดยการสร้างสมการถดถอยทำนายคะแนนของแบบทดสอบ Y จากคะแนนของแบบทดสอบ X กระทำได้ดังนี้

1. สุ่มกลุ่มผู้สอบให้เป็นตัวแทนประชากรเป้าหมายที่ต้องการนำแบบทดสอบไปใช้ แล้วให้กลุ่มตัวอย่างทำแบบทดสอบทั้งฉบับ X และ Y
2. นำคะแนนของแบบทดสอบแต่ละฉบับมาคำนวณค่า X , S_x , Y , S_y เพื่อคำนวณหาค่าความชันและค่าคงที่เพื่อสร้างสมการทำนาย
3. สร้างตารางปรับเทียบคะแนนระหว่างแบบทดสอบ X และ Y โดยแทนค่า X_i ลงในสมการเพื่อหาคะแนนสมมูล Y_i

การปรับเทียบคะแนนโดยใช้สมการถดถอยวิธีนี้เป็นการใช้สมการเส้นตรงในการทำนายตัวแปรตาม จากตัวแปรอิสระ ซึ่งมีลักษณะที่ไม่สมมาตร กล่าวคือ สมการทำนาย Y จาก X หรือการทำนาย X จาก Y เป็นสมการที่ไม่สมมาตร หรือให้ผลการทำนายไปในทิศทางเดียว นอกจากนี้ยังมีข้อดคลงเบื้องต้นว่าคะแนนจากแบบทดสอบที่ใช้เป็นตัวทำนายจะต้องมีค่าความเที่ยงเป็น 1.00

การนำวิธีปรับเทียบคะแนนวิธีนี้ไปใช้ มีข้อแนะนำที่สำคัญดังนี้

1. แบบทดสอบที่ต้องการนำมาปรับเทียบคะแนนกัน ต้องเป็นแบบทดสอบที่วัดคุณลักษณะเดียวกัน มีความเป็นคู่ขนานกัน และมีความเที่ยงสูง
2. แบบทดสอบที่ต้องการนำมาปรับเทียบคะแนนกัน ต้องมีความสัมพันธ์กับคะแนนเกณฑ์เท่าเทียมกัน มิเช่นนั้นแล้วจะทำให้ความสามารถในการทำนายคะแนนเกณฑ์โดยใช้คะแนนจากแบบทดสอบทั้งสองฉบับไม่มีความแม่นยำ อีกฉบับหนึ่งอาจทำนายได้ดีกว่าอีกฉบับหนึ่ง ทำให้การปรับเทียบคะแนนมีความผันแปรตามกลุ่มที่ใช้ศึกษา

4.2.2. รูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Equating Through Item Response Theory)

การปรับเทียบมาตราโดยใช้ทฤษฎีการตอบสนองข้อสอบ มีการปรับเทียบมาตราโดยใช้คะแนนจริงและการปรับเทียบมาตราโดยใช้คะแนนสังเกตได้ ทั้งสองวิธีนี้มีทั้งข้อได้เปรียบและข้อเสียเปรียบ กล่าวคือ การปรับเทียบด้วยคะแนนจริงไม่สามารถอธิบายคะแนนที่อยู่ต่ำกว่าระดับการเดาได้ โดยจะให้ความหมายของคะแนนสมมูลเฉพาะคะแนนที่อยู่เหนือค่าเฉลี่ยของการเดา ถึงแม้จะเป็นการปรับเทียบโดยใช้คะแนนจริง แต่ยังเป็นค่าที่ประมาณได้จากสูตรในการคำนวณ ดังนั้นจึงยังคงมีความคลาดเคลื่อนอยู่ ส่วนการปรับเทียบโดยใช้คะแนนสังเกตได้นั้นเป็นการปรับเทียบคะแนนโดยประมาณ ซึ่งอธิบายคะแนนสมมูลจาก X และ Y ได้ครอบคลุมพิสัยของคะแนนสังเกตได้

ลอร์ด (Lord, 1980) ได้กล่าวว่า การปรับเทียบมาตราโดยใช้คะแนนจริงและการปรับเทียบมาตราโดยใช้คะแนนสังเกตได้มีความสอดคล้องกันมาก แต่การสรุปผลในการอ้างอิงต้องทำอย่างรอบคอบ อย่างไรก็ตามการปรับเทียบมาตราโดยใช้คะแนนสังเกตได้นี้ยังนำไปใช้ในงานวิจัยน้อยมาก อาจเป็นเพราะว่าวิธีนี้มีความซับซ้อนกว่าวิธีการปรับเทียบมาตราโดยใช้คะแนนจริง

วิธีการปรับเทียบมาตราระหว่างแบบทดสอบ X กับแบบทดสอบ Y โดยการใช้วิธีทฤษฎีการตอบสนองข้อสอบ กระทำได้ดังนี้ (Kolen & Brennan, 1995; ศิริชัย กาญจนวาสี, 2541, หน้า 36)

1. เลือกแบบแผนการปรับเทียบ (Equating Design)

แบบแผนการปรับเทียบคะแนนอาจเป็นแบบแผนกลุ่มเดี่ยว แบบแผนกลุ่มสมมูล และแบบแผนแบบทดสอบรวม สำหรับการเลือกแบบแผนการปรับเทียบที่เหมาะสมขึ้นอยู่กับธรรมชาติของแบบทดสอบและกลุ่มผู้สอบ

2. กำหนดโมเดลการตอบสนองข้อสอบ (Item Response Model)

โมเดลการตอบสนองข้อสอบประเภทให้คะแนนรายข้อแบบ 0, 1 ที่นิยมใช้กันมี 3 โมเดล ได้แก่ โมเดล 1, 2 และ 3 พารามิเตอร์ การตัดสินใจเลือกใช้โมเดลใดขึ้นอยู่กับลักษณะของการปรับเทียบและขนาดกลุ่มตัวอย่าง

ในกรณีกลุ่มตัวอย่างมีขนาดไม่ใหญ่มาก การเลือกใช้โมเดล 1 พารามิเตอร์ น่าจะมีความเหมาะสมกับการปรับเทียบในแนวนอน แต่ถ้ากลุ่มตัวอย่างมีขนาดใหญ่ การใช้โมเดล 2 พารามิเตอร์ หรือ 3 พารามิเตอร์น่าจะเหมาะสมกว่า สำหรับการปรับเทียบในแนวตั้ง การใช้โมเดล 1 พารามิเตอร์ไม่ค่อยเหมาะสม เนื่องจากลักษณะข้อสอบอาจไม่สอดคล้องกับโมเดล (Slinde & Linn, 1979) การใช้โมเดล 2 พารามิเตอร์ หรือ 3 พารามิเตอร์ จะมีความเหมาะสมมากกว่า

3. สร้างเมทริกซ์สเกลร่วมของความสามารถและพารามิเตอร์ข้อสอบ

เนื่องจากพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบมีความสัมพันธ์กันเชิงเส้นตรง จึงทำให้สามารถสร้างเมทริกซ์สหสัมพันธ์ได้ โดยการคำนวณค่าคงที่ในสมการ สำหรับการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบให้อยู่บนสเกลเดียวกัน เพื่อเปรียบเทียบคะแนนความสามารถ (θ) ให้อยู่บนสเกลที่สามารถเปรียบเทียบกันได้โดยตรง

4. กำหนดสเกลสำหรับรายงานคะแนนสอบ (Test Score) ที่ปรับเทียบแล้ว

การรายงานผลคะแนนสอบที่ปรับเทียบแล้ว สามารถเลือกสเกลสำหรับการรายงานเป็นคะแนนความสามารถ คะแนนจริง และคะแนนดิบ

5 แบบทดสอบร่วมในการปรับเทียบมาตรา

5.1 ประเภทของแบบทดสอบร่วม

แบบทดสอบร่วมเป็นกลุ่มของข้อสอบที่ผู้สอบต้องทำเหมือนกัน นอกเหนือจากแบบทดสอบที่ผู้สอบต้องทำคนละชุด โดยแบ่งเป็น 2 ประเภท ดังนี้

5.1.1 แบบทดสอบร่วมภายใน เป็นแบบทดสอบที่มีข้อสอบร่วมที่ผู้สอบต้องทำเหมือนกันจัดรวมไว้อยู่ในแบบทดสอบที่ต้องการนำมาปรับเทียบมาตรา

5.1.2 แบบทดสอบร่วมภายนอก เป็นแบบทดสอบที่มีข้อสอบร่วมที่ผู้สอบต้องทำเหมือนกัน ถูกจัดแยกออกมาต่างหากจากตัวแบบทดสอบที่ต้องการนำมาปรับเทียบมาตรา ซึ่งกลุ่มผู้สอบจะต้องทำและจับเวลาแยกออกจากตัวแบบทดสอบที่ต้องการนำมาปรับเทียบมาตรา

5.2 ความยาวของแบบทดสอบร่วม

แบบทดสอบร่วมควรมีความยาว และค่าความเที่ยงเพียงพอที่จะได้ข้อมูลไปปรับความแตกต่างระหว่างกลุ่มตามที่คาดหวัง ซึ่งแองกอฟฟ์ (Angoff, 1984) ได้เสนอเกณฑ์เกี่ยวกับจำนวนข้อของแบบทดสอบร่วม คือ ไม่ควรมีน้อยกว่า 20 ข้อ หรือ 20% ของจำนวนข้อในแต่ละแบบทดสอบ แล้วแต่จำนวนใดจะมากกว่าให้ใช้จำนวนนั้น และถ้าจะให้มีประสิทธิภาพในระดับที่น่าพอใจ จะต้องเป็นแบบทดสอบที่มีค่าความเที่ยงสูงกว่า .866

ตอนที่ 2 การประเมินคุณภาพของการปรับเทียบมาตรา

1. ความคลาดเคลื่อนของการปรับเทียบมาตรา

การประเมินความคลาดเคลื่อนของการปรับเทียบมาตราที่สมบูรณ์ คือ การวิเคราะห์แหล่งความลำเอียง (Bias) และความแปรผัน (Variability) แล้วนำเสนอให้เห็นเป็นภาพเดียวกัน เพื่อช่วยในการตัดสินใจว่าการนำรูปแบบการปรับเทียบมาตรานั้น ๆ ไปใช้ มีความเหมาะสมหรือไม่อย่างไร

1.1 ความลำเอียงของการปรับเทียบมาตรา (Bias)

ความลำเอียงเป็นองค์ประกอบหนึ่งของความคลาดเคลื่อน ซึ่งจะต้องพิจารณากันอย่างรอบคอบ เมื่อต้องการนำผลของการปรับเทียบมาตราไปใช้ เบอรันและฮอลแลนด์ (Braun & Holland, 1982) ได้เสนอบทวิเคราะห์ไว้ว่า ความลำเอียงทางสถิติมีความหมาย 2 ประการ ดังนี้

1.1.1 ความแตกต่างของค่าเฉลี่ยของค่าตัวประมาณค่า ตลอดจนการสุ่มซ้ำ ๆ จากประชากรเดียวกันกับค่าของประชากรที่ถูกต้อง

1.1.2 ความแตกต่างของฟังก์ชันของการปรับเทียบมาตราโดยประมาณ (Estimated Equating Function) และค่าที่แท้จริงของฟังก์ชันการปรับเทียบมาตรา (Equating Function)

แหล่งของความลำเอียงของการปรับเทียบมาตราที่สำคัญมี 2 แหล่ง ได้แก่

1) ความผันแปรของประชากร (Population Variability) เนื่องจากการปรับเทียบมาตราไม่ได้จัดกระทำกับประชากรกลุ่มเดียวโดยตลอด ฟังก์ชันของการปรับเทียบมาตราที่สร้างขึ้นเพื่อใช้กับประชากรหนึ่ง อาจนำไปใช้กับอีกประชากรหนึ่งไม่ได้ เช่น ถ้า $X_p(Y)$ ปรับเทียบมาตราจาก Y ไปยังมาตรา X ในประชากร P และ $Y_q(Z)$ ปรับเทียบมาตราจาก Z ไปยังมาตรา Y ในประชากร Q แล้ว จะสรุปได้ว่า ฟังก์ชันเชิงซ้อน $X_p(Y_q(Z))$ ปรับเทียบมาตราจาก Z ไปยัง X ในประชากร P หรือ Q ย่อมไม่ได้ ในแบบแผนการปรับเทียบมาตราที่ใช้แบบทดสอบร่วมได้แก้ไขปัญหานี้ โดยการสร้างประชากรสังเคราะห์ (Synthesis Population) ซึ่งเป็นการรวมของประชากร P และ Q ตามสัดส่วน แต่ยังไม่รับประกันผลว่าจะไม่มีความลำเอียงในทุก ๆ ประชากรที่แปรเปลี่ยนไป

2) ความคลาดเคลื่อนของโมเดล (Model Error) ความคลาดเคลื่อนนี้เกิดขึ้นจากการต้องการจัดกระทำกับข้อมูลให้ง่ายขึ้น แต่เกิดการผิดพลาด แบ่งเป็น 2 ชนิด ได้แก่

- ก. ความคลาดเคลื่อนของโมเดลที่ทดสอบได้ เช่น ความเป็นเส้นตรงของฟังก์ชันการถดถอย สามารถทดสอบได้จากข้อมูลจำนวนมาก ซึ่งจะเห็นผลได้ว่า มีความเพียงพอหรือไม่
- ข. ความคลาดเคลื่อนของโมเดลที่ทดสอบไม่ได้ เป็นกรณีที่ไม่สามารถหาข้อมูลมาทดสอบเพื่อพิสูจน์ว่า สมมติฐานที่ตั้งไว้เพียงพอหรือไม่

1.2 ความแปรผันเชิงสุ่ม (Variability Sampling)

วิธีการปรับเทียบมาตราทุกวิธี ไม่ว่าอยู่ในรูปแบบการปรับเทียบมาตราใด เมื่อกลุ่มผู้สอบเป็นกลุ่มที่สุ่มมาจากประชากรเดียว หรือหลายประชากร ย่อมมีความผันแปรเชิงสุ่มเกิดขึ้น จึงนิยมใช้เทคนิคการประมาณค่าความคลาดเคลื่อนเชิงสุ่ม (Estimate sampling error) กับการปรับเทียบมาตราต่าง ๆ ความคลาดเคลื่อนเชิงสุ่มมีสมมติฐานเบื้องต้นว่า กลุ่มตัวอย่างมาจากการสุ่มและใช้ความคลาดเคลื่อนมาตรฐานของการปรับเทียบมาตราในการวัดความแปรผันประเภทนี้

2. ความเพียงพอของการเปรียบเทียบมาตรา

การเปรียบเทียบมาตราวิธีใดก็ตามจะมีคุณภาพดีที่สุดในเมื่อคะแนนที่ได้จากแบบทดสอบที่ต้องการเป็นไปตามเงื่อนไขต่าง ๆ ที่กำหนดไว้ของแต่ละวิธี แต่ในสถานการณ์จริงมีข้อจำกัดทำให้ไม่สามารถได้ข้อมูลตามเงื่อนไข ดังนั้นจึงจำเป็นจะต้องตรวจสอบความเพียงพอของการเปรียบเทียบมาตราหรือประเมินคุณภาพของวิธีที่แตกต่างกัน ซึ่งวิธีการประเมินความเพียงพอ มีผู้เสนอแนวคิดและวิธีปฏิบัติดังนี้

การประเมินก่อนดำเนินการเปรียบเทียบมาตรา

เจเกอร์ (Jaeger, 1981) ได้เสนอดัชนี 5 ตัวเพื่อตรวจสอบดัชนีความเพียงพอของการใช้วิธีเชิงเส้นตรงว่าเทคนิควิธีที่ได้นำมาใช้ในการเปรียบเทียบเพียงพอกับการปรับความแตกต่างระหว่างการแจกแจงของคะแนนจากแบบทดสอบหรือไม่ หรือจำเป็นต้องมีวิธีอื่นที่มีความสลับซับซ้อนมากขึ้น ดัชนี 5 ตัว มีดังนี้

ตัวที่ 1 ดัชนีความคล้ายคลึงของการแจกแจงคะแนนสะสมของแบบทดสอบชุดเก่า และชุดใหม่โดยการปรับความแตกต่างระหว่างค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน การทดสอบความเหมือนของการแจกแจง

ตัวที่ 2 รูปแบบของการแจกแจงคะแนนดิบกับคะแนนแปลง เหตุผลของการใช้ดัชนีตัวนี้ คือ ถ้าการเปรียบเทียบมาตราด้วยวิธีเชิงเส้นตรงสามารถอธิบายความแตกต่างในการแจกแจงของคะแนนดิบทั้งสองชุดอย่างเพียงพอ ก็เป็นเหตุผลอย่างเพียงพอเช่นกันที่จะยอมรับว่าการแปลงคะแนนดิบจากแบบทดสอบชุดใหม่ไปยังแบบทดสอบชุดเก่าเป็นเส้นตรงอย่างแน่นอน

ตัวที่ 3 ความคงเส้นคงวาของผลลัพธ์ของการเปรียบเทียบมาตราตามวิธีเชิงเส้นตรงกับ การเปรียบเทียบที่ตำแหน่งเปอร์เซ็นต์ไทล์ การวิเคราะห์นี้อาศัยข้อตกลงเบื้องต้นว่า ถ้าวิธีเชิงเส้นตรงมีความเพียงพอแล้ว ฟังก์ชันของวิธีเชิงเส้นตรงก็มีความสมนัยกัน

ตัวที่ 4 ความคล้ายคลึงของการแจกแจงความยาวของข้อสอบ โดยอาศัยหลักการที่ว่า การใช้วิธีการเชิงเส้นตรงมีความเพียงพออย่างแท้จริงกับแบบทดสอบคู่ขนานกัน ถ้ามีความเบี่ยงเบนจากความเป็นคู่ขนานมากเท่าไร แสดงว่าต้องการวิธีการเปรียบเทียบมาตราที่ซับซ้อนขึ้น เพราะการแจกแจงของแบบทดสอบที่ไม่ใช่คู่ขนานจะมีความแตกต่างเกิดขึ้นในรูปแบบโมเมนต์ที่สูงขึ้น

ตัวที่ 5 ความคล้ายคลึงของค่าอำนาจจำแนกของข้อสอบ มีเหตุผลทำนองเดียวกับดัชนีตัวที่ 4

การประเมินผลการเปรียบเทียบมาตรา

การประเมินผลการเปรียบเทียบมาตราเป็นการประเมิน โดยอาศัยคะแนนเกณฑ์ที่ได้เลือกสรรไว้แล้วเป็นหลักในการเปรียบเทียบเพื่อหาความแตกต่าง มีรายละเอียดดังนี้

1. ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ (The Percentile Comparison Index)

โคเลน และวิทนี (Kolen & Whitney, 1982, p. 284) ได้แนะนำให้ใช้ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ เพื่อประเมินความเพียงพอของการปรับเทียบมาตรา ทั้งนี้โดยอาศัยกลุ่มตัวอย่างสอบทานผล (Cross-Validation Sample Examinees) ซึ่งเป็นกลุ่มอิสระอีกกลุ่มหนึ่งที่สุ่มมาจากประชากรเดียวกันกับกลุ่มผู้สอบที่ใช้สร้างตารางการปรับเทียบมาตรา วิธีการหาค่าดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์นี้ คือ การวัดความไม่สอดคล้อง (Dissimilarity) ระหว่างการแจกแจงคะแนนที่ได้จากกลุ่มตรวจสอบผล คือ คะแนนจากแบบทดสอบชุด X กับคะแนนที่ได้จากการแปลงคะแนนจากแบบทดสอบชุด Y ไปยังสเกลของแบบทดสอบชุด X ตามตารางการปรับเทียบมาตราที่ได้จากวิธีการใช้ปรับเทียบโดยหาค่าสมมูล ณ ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกัน ดัชนีดังกล่าวคำนวณได้ดังนี้

$$C = \sum (X_i - X_i^*)^2 / nk$$

เมื่อ C คือ ดัชนีความเพียงพอ

X_i คือ คะแนนจากการสอบชุด X ของผู้สอบคนที่ i

X_i^* คือ คะแนนที่ได้เทียบด้วยตารางคะแนนที่สมนัยกันของคน i

n คือ จำนวนคะแนนดิบในกลุ่มตัวอย่างสอบทานผล

k คือ จำนวนข้อสอบในแบบทดสอบรวมที่ใช้

การเปรียบเทียบ ถ้าค่า C น้อยกว่า แสดงว่าวิธีการปรับเทียบมาตราที่นำมาใช้สร้างตารางการแปลงคะแนนนั้นมีความเหมาะสมที่จะนำมาใช้ เพราะให้ผลการปรับเทียบที่คงที่มากกว่า

2. ดัชนีความแตกต่าง (Discrepancy Indices)

ปีเตอร์เซน มาร์โก และสจวร์ต (Petersen, Marco, & Stewart, 1982) ได้เสนอวิธีการประเมินความเพียงพอของการปรับเทียบมาตราด้วยการใช้ดัชนีความแตกต่าง ซึ่งมีแนวคิดว่าจะคะแนนดิบแต่ละตัวจากแบบทดสอบชุด X จะมีคะแนนเกณฑ์ t และคะแนนแปลง t' ที่ได้จากการประมาณค่าโดยใช้วิธีการปรับเทียบมาตราเฉพาะรูปแบบหนึ่ง ค่าความแตกต่างระหว่าง t และ t' เป็นความคลาดเคลื่อนในการปรับเทียบมาตรา ซึ่งค่าที่น้อยกว่า แสดงว่า วิธีการปรับเทียบมาตรานั้นมีความเหมาะสมมากกว่า

ค่าดัชนีความเพียงพอ คือ ค่าความคลาดเคลื่อนรวมในการปรับเทียบมาตราของวิธีการปรับเทียบมาตราที่ใช้นั้น ซึ่งคำนวณจากค่าเฉลี่ยยกกำลังสองของความแตกต่างระหว่างคะแนนแปลงกับคะแนนเกณฑ์และถ่วงน้ำหนักด้วยความแปรปรวนของคะแนนเกณฑ์ ซึ่งเท่ากับค่าความแปรปรวนของความแตกต่างรวมกับกำลังสองของความลำเอียง

$$\text{Total error} = \sum f_j d_j^2 / nS_t^2$$

เมื่อ f_j = ความถี่ของคะแนน

$$d_j^2 = (t - t')^2$$

n = จำนวนคะแนนที่ใช้

S_t^2 = ความแปรปรวนของคะแนน t

ดัชนีเปรียบเทียบความแตกต่างนี้เป็นค่ามาตรฐานที่แสดงในรูปสัดส่วนของความเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์ ค่าความคลาดเคลื่อนรวมที่ได้นี้จึงสามารถเปรียบเทียบกันได้โดยตรงระหว่างค่าที่ได้จากการใช้วิธีกับสถานการณ์การเปรียบเทียบมาตรวัดต่างกัน

3. ปัจจัยที่ส่งผลต่อคุณภาพของการเปรียบเทียบมาตรา (Kolen & Brennan, 1995)

3.1 หลักการทั่วไป

3.1.1 มีการกำหนดวัตถุประสงค์ของการเปรียบเทียบมาตรา

3.1.2 มีการออกแบบวิธีการเก็บรวบรวมข้อมูล และวิธีการเปรียบเทียบมาตรา

3.1.3 มีกระบวนการประเมินคุณภาพ

3.2 การพัฒนาแบบทดสอบ (ทุกรูปแบบ)

3.2.1. แบบทดสอบที่นำมาเปรียบเทียบมาตรามีเนื้อหา โครงสร้าง และลักษณะทางสถิติเหมือนกัน

3.2.2 แบบทดสอบมีความยาวเหมาะสม

3.2.3 คำตอบหรือเฉลยมีความถูกต้องแน่นอน

3.3 การพัฒนาแบบทดสอบ (กรณีแบบทดสอบร่วม)

3.3.1 แบบทดสอบร่วมต้องเป็นตัวแทนของแบบทดสอบทั้งหมดในด้านเนื้อหา และลักษณะทางสถิติ

3.3.2 แบบทดสอบร่วมมีจำนวนเพียงพอ คือ อย่างน้อย 20% ของแบบทดสอบทั้งหมด

3.3.3 แบบทดสอบร่วมในแต่ละฉบับอยู่ในตำแหน่งเดียวกัน เช่น ข้อ 1, 2, 3,... เป็นต้น นอกจากนั้นคำถามและตัวเลือกของแบบทดสอบร่วมต้องเหมือนกันทุกฉบับ

3.4 กลุ่มผู้สอบ

3.4.1 ผู้สอบแต่ละกลุ่มเป็นตัวแทนของผู้สอบทั้งหมด

3.4.2 จำนวนผู้สอบมีเพียงพอ

3.4.3 ในรูปแบบแบบทดสอบร่วม ผู้สอบที่ทำแบบทดสอบร่วมแต่ละฉบับต้องมีความสามารถไม่แตกต่างกันมากนัก

3.5 การดำเนินการสอบ

3.5.1 แบบทดสอบที่ใช้ต้องเป็นความลับ ไม่รั่วไหล

3.5.2 การดำเนินการสอบต้องมีการดำเนินการอย่างเป็นมาตรฐาน

ตอนที่ 3 งานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบมาตรฐาน

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการเปรียบเทียบมาตรฐาน ผู้วิจัยสามารถรวบรวมงานวิจัยที่สำคัญได้ดังนี้

ชูชีพ พงษ์สมบุญ (2528) ได้ศึกษาเปรียบเทียบรูปแบบการปรับเทียบระดับมาตรฐาน 3 วิธี คือ การปรับเทียบเชิงเส้นตรง การปรับเทียบแบบอควิเปอร์เซ็นไทล์ และการปรับเทียบแบบ 3 พารามิเตอร์ จากรูปแบบที่ใช้แบบทดสอบร่วมกับรูปแบบที่ใช้ผู้ตอบรวม เพื่อต้องการเปรียบเทียบประสิทธิภาพและความคงที่ของการปรับเทียบระดับมาตรฐานทั้ง 3 วิธี กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศสุ่มมาจำนวน 3,721 คน แบบทดสอบที่ใช้แบ่งเป็น 2 ฉบับ ๆ ละ 38 ข้อ มีแบบทดสอบร่วมภายใน 20 ข้อ แบบทดสอบทั้ง 2 ฉบับ มีความยากใกล้เคียงกัน ผลการศึกษาปรากฏว่าประสิทธิภาพของการปรับเทียบมาตรฐานระหว่างรูปแบบผู้ตอบรวมและรูปแบบแบบทดสอบร่วมไม่แตกต่างกัน

ภาวิณี ศรีสุขวัฒนานนท์ (2529) ได้เปรียบเทียบผลของการใช้รูปแบบการปรับเทียบระดับมาตรฐานต่างกัน 3 รูปแบบ เช่นเดียวกับ ชูชีพ พงษ์สมบุญ แต่ในการวิจัยครั้งนี้จะใช้แบบทดสอบร่วมภายในต่างกัน 3 ขนาด คือ ขนาดร้อยละ 20 (7 ข้อ) ขนาดร้อยละ 40 (14 ข้อ) ขนาดร้อยละ 60 (21 ข้อ) โดยใช้แบบทดสอบ 2 ฉบับ ๆ ละ 35 ข้อ แบบทดสอบทั้ง 2 ฉบับ มีความยากใกล้เคียงกัน การประเมินผลการปรับเทียบมาตรฐานมี 2 ลักษณะ

1. ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบมาตรฐานด้วยวิธีที่ต่างกันของแบบทดสอบแต่ละกรณี
 2. ตรวจสอบความพอเพียงของวิธีการปรับเทียบมาตรฐานแต่ละวิธี โดยใช้ดัชนีเปรียบเทียบความแตกต่าง (Discrepancy Index: C) ที่ใช้ในการวิเคราะห์จากกลุ่มที่ใช้ในการสอบทานผล
- กลุ่มตัวอย่างที่ใช้จำแนกเป็น 2 กรณี คือ กรณีการสอบคัดเลือกเพื่อเข้าทำงานและการสอบวัดผลสัมฤทธิ์ กลุ่มตัวอย่างแต่ละกรณีมี 2 กลุ่ม ๆ ละ 1,500 คน ซึ่งเป็นกลุ่มที่ใช้ในการปรับเทียบมาตรฐาน 1 กลุ่ม และใช้ในการสอบทานผลการปรับเทียบมาตรฐาน 1 กลุ่ม การประเมินผลการปรับเทียบมาตรฐานใช้ในการวิเคราะห์หาความคลาดเคลื่อนมาตรฐานของการปรับเทียบมาตรฐาน และการวิเคราะห์กลุ่มสอบผลทวน ปรากฏว่าการใช้แบบทดสอบร่วมที่ยาวกว่าให้ประสิทธิภาพในการปรับเทียบมาตรฐานสูงกว่า ความคลาดเคลื่อนน้อยกว่า และให้ผลในระดับที่น่าพอใจมาก ทั้งสองกรณีแบบทดสอบ การศึกษาการปรับเทียบมาตรฐาน ปรากฏว่าในกรณีแบบทดสอบคัดเลือกลักษณะวิธีที่ให้ความเพียงพอมากที่สุด คือ วิธีอควิเปอร์เซ็นไทล์ รองลงมาคือ วิธีของทฤษฎีการตอบสนองข้อสอบ ซึ่งให้ผลที่แตกต่างกันในกรณีแบบทดสอบวัดผลสัมฤทธิ์ที่ปรากฏว่า วิธีที่ให้ความเพียงพอ

มากที่สุด คือ วิธีเชิงเส้นตรง รองลงมา คือ วิธีอีควิเปอร์เซ็นไทล์และเมื่อทดสอบความแตกต่างของดัชนี C เป็นรายคู่ ปรากฏว่ากรณีที่แบบทดสอบวัดผลสัมฤทธิ์ วิธีเชิงเส้นตรงมีความเพียงพอมากกว่าวิธีของทฤษฎีการตอบสนองข้อสอบอย่างมีนัยสำคัญทางสถิติ .05 ส่วนการเปรียบเทียบวิธีคู่อื่นไม่สามารถสรุปความแตกต่างอย่างมีนัยสำคัญ

เรวดี อินทสระ (2530) ได้เปรียบเทียบความคลาดเคลื่อนมาตรฐานของการปรับเทียบมาตราและความเพียงพอจากการใช้รูปแบบอิงทฤษฎีการตอบสนองข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ โดยใช้กลุ่มทานผล เครื่องมือใช้แบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 เรื่องอัตราส่วนร้อยละ การวิเคราะห์ค่าพารามิเตอร์ของแบบทดสอบตามทฤษฎีการตอบสนองข้อสอบ โดยใช้โปรแกรมโลจิส 5 แล้วจัดข้อสอบออกเป็น 2 ฉบับ ๆ ละ 30 ข้อ และแบบทดสอบร่วมฉบับละ 15 ข้อ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2529 โรงเรียนสังกัดกรมสามัญศึกษาจังหวัดพัทลุง จำนวน 2,823 คน โดยใช้กลุ่มตัวอย่างทำแบบทดสอบทั้งสองฉบับแล้วสุ่มมากลุ่มละ 1 ฉบับ เพื่อนำคะแนนมาปรับเทียบมาตราทั้งสองรูปแบบและคะแนนอีกฉบับจะเป็นคะแนนทานผล ผลการศึกษาปรากฏว่า ความคลาดเคลื่อนมาตรฐานของการปรับเทียบมาตราของรูปแบบอิงทฤษฎีการตอบสนองข้อสอบมีค่าน้อยกว่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบมาตราของรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบส่วนความเพียงพอของความคลาดเคลื่อนการปรับเทียบมาตรารูปแบบอิงทฤษฎีการตอบสนองข้อสอบอยู่ในระดับที่น่าพอใจมากกว่าการปรับเทียบมาตรารูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ

อาภรณ์ กาญจนกิจโสภณ (2531) ได้ศึกษาการสร้างแบบทดสอบและตารางปรับเทียบมาตราคะแนนตามแนวคิดในวิชาคณิตศาสตร์ เรื่องสมการและอสมการ สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น โดยใช้วิธีการปรับเทียบมาตรารูปแบบราสส์ กลุ่มตัวอย่างที่ใช้ปรับเทียบมาตราเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 จำนวนระดับชั้นละ 490, 482 และ 415 คน ตามลำดับ ผู้วิจัยได้นำแบบทดสอบที่สร้างขึ้นระดับชั้นละ 90 ข้อ ไปทดลองกับกลุ่มตัวอย่างอื่น แล้วคัดเลือกข้อสอบที่มีคุณภาพเรียงตามค่าความยากที่ต้องการระดับชั้นละ 40, 30 และ 30 ข้อ โดยแต่ละชั้นจะมีข้อสอบร่วมระหว่างชั้นอยู่ 10 ข้อ หลังจากนั้นนำไปทดสอบกับกลุ่มตัวอย่างที่ใช้ปรับเทียบมาตรา และสร้างตารางปรับเทียบมาตราเพื่อใช้ศึกษาพัฒนาการของนักเรียนกลุ่มเดิมที่ผ่านขึ้นไปเรียนในชั้นสูงขึ้นไป นักเรียนมีความก้าวหน้าเป็นไปตามปกติหรือเบี่ยงเบนไปจากปกติ

พรพิมล นาคเวช (2537) ได้ตรวจสอบคุณภาพของการปรับเทียบมาตราตามแนวคิดโดยใช้ทฤษฎีการตอบสนองข้อสอบ สำหรับผู้มีความสามารถต่างกัน 3 ระดับ ซึ่งได้แก่นักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 โดยวิเคราะห์ความคลาดเคลื่อนมาตรฐานและความพอเพียงของ

การเปรียบเทียบมาตรา กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 สังกัดกองการมัธยมศึกษา กรมสามัญศึกษา จำนวน 6,429 คน ที่ได้จากกลุ่มตัวอย่างแบบแบ่งชั้นหลายชั้นตอน แบ่งกลุ่มตัวอย่างเปรียบเทียบมาตราเพื่อนำผลสอบไปวิเคราะห์เทียบมาตราระหว่างแบบทดสอบ จำนวน 3,279 คน และกลุ่มตัวอย่างสอบทานผลเพื่อนำผลไปตรวจสอบคุณภาพของการเปรียบเทียบมาตราจำนวน 3,105 คน เครื่องมือที่ใช้เป็นแบบทดสอบคณิตศาสตร์ เรื่องสมการและอสมการ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 รวม 3 ฉบับ ๆ ละ 25 ข้อ โดยใช้ข้อสอบร่วมระหว่างระดับชั้นมัธยมศึกษาปีที่ 1 และ 2 จำนวน 5 ข้อ ข้อสอบร่วมระหว่างระดับชั้นมัธยมศึกษาปีที่ 2 และ 3 จำนวน 5 ข้อ สรุปผลได้ว่าผลการวิเคราะห์ความคลาดเคลื่อนมาตรฐานและความพอเพียงของการเปรียบเทียบมาตราแนวตั้งสำหรับผู้มีความสามารถต่างกัน 3 ระดับ ปรากฏว่าความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบมาตราก่อนข้างต่ำ และค่าความพอเพียงของการเปรียบเทียบมาตราอยู่ในระดับที่น่าพอใจ

รุ่งนภา จันทรา (2540) ได้เปรียบเทียบความคลาดเคลื่อน และความพอเพียงของการเปรียบเทียบมาตราระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์ กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดเพชรบุรี จำนวน 925 คน เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลเป็นแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์แบบเลือกตอบชนิด 4 ตัวเลือก จำนวน 2 ฉบับ ๆ ละ 18 ข้อ ผลการศึกษาปรากฏว่า ค่าดัชนีความพอเพียงของการเปรียบเทียบมาตรารูปแบบที่มีสามพารามิเตอร์อยู่ในระดับที่น่าพอใจมากกว่าหนึ่งพารามิเตอร์ และรูปแบบที่มีสามพารามิเตอร์มีความคลาดเคลื่อนน้อยกว่ารูปแบบหนึ่งพารามิเตอร์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และการเปรียบเทียบมาตราโดยวิธีทฤษฎีการตอบสนองข้อสอบที่ไม่ใช้ค่าการเดา จะไม่สามารถอธิบายผลการเปรียบเทียบคะแนนได้ และการเปรียบเทียบมาตราเช่นนี้ควรใช้วิธีการเปรียบเทียบเชิงเส้นตรงจะได้ผลที่ดีกว่า

เกตรินทร์ หมัดอะดัม (2545) ได้เปรียบเทียบความคลาดเคลื่อนและความพอเพียงของการเปรียบเทียบมาตราแนวตั้งระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์ กับสามพารามิเตอร์ กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชั้นมัธยมศึกษาตอนต้น ปีการศึกษา 2544 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดสงขลา จำนวน 2,100 คน เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูลเป็นแบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์แบบเลือกตอบชนิด 4 ตัวเลือก เรื่อง สมการและอสมการ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 รวม 3 ฉบับ ๆ ละ 25 ข้อ โดยใช้ข้อสอบร่วมระหว่างระดับชั้นมัธยมศึกษาปีที่ 1 และ 2 จำนวน 5 ข้อ ข้อสอบร่วมระหว่างระดับชั้นมัธยมศึกษาปีที่ 2 และ 3 จำนวน 5 ข้อ ผลการศึกษาปรากฏว่า ค่าดัชนีความพอเพียงของ

การเปรียบเทียบมาตรฐานรูปแบบที่มีสามพารามิเตอร์อยู่ในระดับที่น่าพอใจ ส่วนค่าดัชนีความพอเพียงของการเปรียบเทียบมาตรฐานรูปแบบที่มีหนึ่งพารามิเตอร์อยู่ในระดับที่ไม่น่าพอใจ ซึ่งมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และรูปแบบที่มีสามพารามิเตอร์มีความคลาดเคลื่อนน้อยกว่ารูปแบบหนึ่งพารามิเตอร์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

โคเลน (Kolen, 1981, pp. 1-11) ได้เปรียบเทียบวิธีการปรับเทียบมาตรฐานเชิงเส้นตรง อีคิวเปอร์เซ็นต์ไทล์ และทฤษฎีการตอบสนองข้อสอบ หนึ่ง สอง และสามพารามิเตอร์ ผลการศึกษาปรากฏว่า การปรับเทียบมาตรฐานทั้งหมดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 การปรับเทียบมาตรฐานโดยใช้วิธีอีคิวเปอร์เซ็นต์ไทล์เหมาะสมที่สุดกับแบบทดสอบที่มีความยากต่างกัน ส่วนการปรับเทียบมาตรฐานโดยใช้ทฤษฎีการตอบสนองข้อสอบ กรณีสองพารามิเตอร์ ยังเป็นวิธีที่ไม่ถูกต้องสำหรับแบบทดสอบที่ต่างกัน ซึ่งเนื่องมาจากการเดาเกิดขึ้นมาก ส่วนสามพารามิเตอร์มีปัญหาที่เกิดขึ้นอย่างหลีกเลี่ยงไม่ได้ คือ การประมาณค่าไม่ถูกต้องของ Lower Asymptote Parameter

โคเลน และวิทนี (Kolen & Whitney, 1982, pp. 279 - 292) ได้เปรียบเทียบความถูกต้องของการปรับเทียบระดับมาตรฐานโดยใช้แบบทดสอบร่วม 4 วิธี คือ เชิงเส้นตรง อีคิวเปอร์เซ็นต์ไทล์ และทฤษฎีการตอบสนองข้อสอบแบบหนึ่งพารามิเตอร์กับสามพารามิเตอร์ ผลการศึกษาปรากฏว่า วิธีการปรับเทียบระดับมาตรฐานโดยใช้อีคิวเปอร์เซ็นต์ไทล์ และทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ได้ผลไม่เป็นที่ยอมรับ ขณะเดียวกันวิธีการปรับเทียบระดับมาตรฐานโดยใช้แบบหนึ่งพารามิเตอร์ให้ผลความเพียงพอเท่ากับแบบสามพารามิเตอร์ และพบว่าความถูกต้องของวิธีการปรับเทียบระดับมาตรฐานขึ้นอยู่กับองค์ประกอบหลายอย่าง เช่น ลักษณะของแบบทดสอบ รูปแบบของการปรับเทียบขนาดของกลุ่มตัวอย่าง เป็นต้น

ฮัทเทน (Hutten, 1982) ได้ศึกษาความเหมาะสมของข้อมูลจริงกับรูปแบบทฤษฎีการตอบสนองข้อสอบ 2 รูปแบบ คือ หนึ่งพารามิเตอร์รูปแบบของราล์ชกับรูปแบบสามพารามิเตอร์ ในการประมาณค่าพารามิเตอร์ความสามารถ และความยากของข้อสอบ ผลการศึกษาปรากฏว่าการวิเคราะห์ข้อสอบด้วยรูปแบบของราล์ชมีความคงที่ของค่าพารามิเตอร์ดีกว่ารูปแบบสามพารามิเตอร์ และเมื่อมีการปรับเทียบมาตรฐานของแบบทดสอบข้ามกลุ่มตัวอย่างรูปแบบของราล์ชก็มีความคงที่ดีกว่า

โจฮันสัน (Johanson, 1988) ได้ศึกษาผลของความยากของแบบทดสอบร่วมความสามารถของผู้สอบ และวิธีการประเมินค่าพารามิเตอร์ที่มีต่อการปรับเทียบมาตรฐานตามแนวตั้งตามทฤษฎีการตอบสนองข้อสอบ ผลการศึกษาปรากฏว่า วิธีการปรับเทียบมาตรฐานโดยใช้ค่าลักษณะข้อสอบเป็นวิธีที่มีความถูกต้องมากที่สุด นอกจากนี้วิธีปรับเทียบคะแนนโดยใช้ค่าลักษณะข้อสอบยังเป็น

วิธีการเดียวที่ให้ผลการเปรียบเทียบมาตรฐานในระดับที่ยอมรับได้โดยใช้แบบทดสอบร่วมน้อยที่สุด สำหรับการเปรียบเทียบมาตรฐานทุกวิธีเมื่อใช้แบบทดสอบร่วมนที่มีจำนวนมากที่สุดและใช้กลุ่มผู้สอบที่มีความสามารถต่างกันน้อยที่สุดให้ผลการเปรียบเทียบมาตรฐานเป็นที่ยอมรับได้ เมื่อใช้กลุ่มตัวอย่างที่มีความสามารถต่างกันมาก การประมาณค่าพารามิเตอร์ของข้อสอบทำได้ไม่ดีนัก โดยสรุปแล้ว ความคลาดเคลื่อนในการประมาณค่าพารามิเตอร์ของข้อสอบอาจทำได้ยากในการเปรียบเทียบมาตรฐานแบบทดสอบที่มีระดับความยากต่างกันมากถ้าแบบทดสอบร่วมนที่ใช้สั้นกว่าและแบบทดสอบฉบับร่วมนมีขนาดสั้น

แฮริส (Harris, 1991) ได้ศึกษาผลของแบบแผนการรวบรวมข้อมูล 2 วิธี คือ ผู้สอบกลุ่มเดียวจัดให้สมดุลกับผู้สอบกลุ่มสมดุล ที่มีต่อการเปรียบเทียบมาตรฐานแนวดิ่ง โดยใช้วิธีการเปรียบเทียบมาตรฐาน 2 วิธี คือ วิธีอีควิเปอร์เซ็นไทล์ และวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ ผลการศึกษาปรากฏว่า ผลการเปรียบเทียบมาตรฐานแนวดิ่งโดยใช้แบบแผนการเก็บรวบรวมข้อมูลทั้ง 2 แบบมีความพอเพียงและให้ผลคล้ายคลึงกัน โดยรูปแบบที่ใช้ผู้สอบกลุ่มเดียวให้ผลคงที่กว่า รูปแบบที่ใช้ผู้สอบกลุ่มสมดุลเล็กน้อย

อเยเว (Ayerbe, 1992) ได้เปรียบเทียบประสิทธิภาพของการเปรียบเทียบมาตรฐานแนวดิ่งด้วยวิธีอีควิเปอร์เซ็นไทล์และวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ ผลการศึกษาปรากฏว่า การเปรียบเทียบมาตรฐานแนวดิ่งต่าง ๆ โดยใช้วิธีการเปรียบเทียบมาตรฐานทั้งสองวิธีไม่แตกต่างกัน สำหรับการเปรียบเทียบมาตรฐานโดยใช้วิธีอีควิเปอร์เซ็นไทล์ พบว่า ความยาวของแบบทดสอบกับความยาวของแบบทดสอบร่วมนเป็นองค์ประกอบที่สำคัญ ถ้ากลุ่มตัวอย่างมีขนาดเล็ก (200 คน) มีแนวโน้มที่จะให้ผลการเปรียบเทียบมาตรฐานที่ไม่ถูกต้อง ในขณะที่กลุ่มตัวอย่างที่มีขนาดใหญ่ขึ้น (500 คน และ 1,000 คน) จะให้ผลการเปรียบเทียบมาตรฐานที่มีความถูกต้องมากกว่า

ยัง (Young, 1995) ได้เปรียบเทียบวิธีการเปรียบเทียบผลการเรียนระหว่างวิธีทฤษฎีการตอบสนองข้อสอบกับวิธีเชิงเส้นตรง ที่จะทำให้ประสิทธิภาพในการทำนายผลการเรียนในโรงเรียนกฎหมายสูงขึ้น โดยเก็บข้อมูลในโรงเรียนกฎหมาย 4 แห่งของสหรัฐอเมริกา ผลการศึกษาปรากฏว่า วิธีการเปรียบเทียบมาตรฐานทั้งสองวิธีทำให้ประสิทธิภาพในการทำนายของคะแนนสอบคัดเลือกเข้าโรงเรียนกฎหมาย (LSAT) และผลการเรียนระดับปริญญาตรีสูงขึ้น โดยเฉพาะในโรงเรียนที่ 4 มีความสัมพันธ์ระหว่างคะแนนสอบเข้าโรงเรียนกฎหมายกับผลการเรียนในโรงเรียนกฎหมายสูงกว่าแห่งอื่น ๆ ซึ่งเป็นเพราะความแปรปรวนของ LSAT ในโรงเรียนที่ 4 มีค่ามากที่สุดนั่นเอง

เหลียง ชิง ออน (Leung Shing On, 2003) ได้ศึกษาวิธีการเปรียบเทียบในแนวดิ่ง 2 วิธี คือ วิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์และวิธีเชิงเส้นตรง โดยใช้แบบทดสอบร่วมนกลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 1 – 6 ในมณฑลฮ่องกง ซึ่งแบบทดสอบที่ใช้ทดสอบมี

เนื้อหาที่ใช้ทดสอบความเชี่ยวชาญในการใช้ภาษาจีน โดยสร้างตามหลักสูตรในแต่ละระดับชั้น การเปรียบเทียบจะทำเป็นคู่ จำนวน 5 คู่ คือ ระหว่างชั้นประถมศึกษาปีที่ 1 กับชั้นประถมศึกษาปีที่ 2 ระหว่างชั้นประถมศึกษาปีที่ 2 กับชั้นประถมศึกษาปีที่ 3 ระหว่างชั้นประถมศึกษาปีที่ 3 กับชั้นประถมศึกษาปีที่ 4 ระหว่างชั้นประถมศึกษาปีที่ 4 กับชั้นประถมศึกษาปีที่ 5 และระหว่างชั้นประถมศึกษาปีที่ 5 กับชั้นประถมศึกษาปีที่ 6 ผลการศึกษาปรากฏว่าความสามารถของนักเรียนในชั้นประถมศึกษาปีที่ 1 และชั้นประถมศึกษาปีที่ 2 มีความแตกต่างกันมากที่สุด และความสามารถของนักเรียนในชั้นประถมศึกษาปีที่ 5 และชั้นประถมศึกษาปีที่ 6 มีความแตกต่างกันน้อยที่สุด ซึ่งแสดงให้เห็นว่านักเรียนสามารถเรียนรู้ได้อย่างรวดเร็วเมื่ออยู่ในวัยเด็ก และการใช้วิธีการเปรียบเทียบคะแนนทั้งสองวิธีเป็นวิธีที่ใช้ได้ดี เพราะให้ค่าสหสัมพันธ์ที่สูงและมีค่าใกล้เคียง 1.00 มาก

จากผลงานวิจัยในประเทศและต่างประเทศที่กล่าวมา จะเห็นว่าการเปรียบเทียบมาตรฐานิยมศึกษาในเรื่องเทคนิคและวิธีการเปรียบเทียบมาตรฐาน เพื่อให้ได้ผลการเปรียบเทียบมาตรฐานที่มีความถูกต้องแม่นยำและมีคุณภาพ และพบว่าไม่มีวิธีการใดเพียงวิธีเดียวที่ให้ผลการเปรียบเทียบมาตรฐานที่ดีที่สุดในทุกสถานการณ์ ดังนั้นในทางปฏิบัติจึงจำเป็นต้องศึกษาข้อกำหนดต่าง ๆ ของการเปรียบเทียบมาตรฐานแต่ละวิธีเข้ามาพร้อมด้วย เช่น ความยาวของแบบทดสอบ การกระจายค่าความยากของแบบทดสอบ ค่าพารามิเตอร์ความสามารถของกลุ่มตัวอย่าง เป็นต้น