บทที่ 2

วรรณคดีที่เกี่ยวข้อง

จากการศึกษา เอกสาร บทความ งานวิจัยที่เกี่ยวข้องกับการเทียบมาตรา ผู้วิจัยขอเสนอวรรณคดีที่เกี่ยวข้อง ตามหัวข้อต่อไปนี้

ตอนที่ 1 ความหมายของการเทียบมาตรา

ตอนที่ 2 รูปแบบการเทียบมาตรา

2.1 การเทียบมาตราเชิงเส้นตรง

2.2 การเทียบมาตรารูปแบบอีควิเปอร์เซ็นไทล์

2.3 การเทียบมาตราโดยใช้ทฤษฎีการตอบข้อสอบ

ตอนที่ 3 แบบสอบร่วมในการเทียบมาตรา

3.1 ประเภทของแบบสอบร่วม

3.2 ความยาวของแบบสอบร่วม

ตอนที่ 4 การประเมินคุณภาพของการเทียบมาตรา

4.1 ความคลาดเคลื่อนมาตรฐานของการเทียบมาตราเชิงเส้นตรง

4.2 การประเมินความเพียงพอของการเทียบมาตรา

ตอนที่ 5 งานวิจัยที่เกี่ยวข้องในการเทียบมาตรา

5.1 งานวิจัยการเทียบมาตราต่างประเทศ

5.2 งานวิจัยการเทียบมาตราในประเทศ

ตอนที่ 1 ความหมายของการเทียบมาตรา

กัลลิคเสน (Gulliksen 1950: 298-304 ภาวิณี ศรีสุขวัฒนานันท์ 2529: 17) ได้ให้ความหมายของการเทียบมาตราไว้ว่า เป็นวิธีการทำคะแนนที่ได้จากแบบสอบสองชุดในวิชาเดียวกัน ให้เป็นคะแนนสมมูล (Equivalent Score) ที่เปรียบเทียบกันได้โดยตรง โดยเสนอวิธีการให้กลุ่มผู้สอบกลุ่มเดียวทำแบบสอบสองชุด แล้วแปลงคะแนนแต่ละชุดให้เป็นคะแนนมาตรฐาน แล้วนำคะแนนแปลงมาเทียบกันโดยตรง

ฟลานาแกน (Flanagan 1951: 747-748) ได้ให้ความหมายของวิธีการเทียบมาตราไว้ว่า เป็นวิธีการทำคะแนนที่ได้จากแบบสอบต่างชุด ให้มีคุณลักษณะที่เปรียบเทียบกันได้ "คุณลักษณะที่เปรียบเทียบกันได้" หมายความว่า เมื่อกำหนดประชากรให้ ถ้าการแจกแจงของคะแนนจริงจากแบบสอบทั้งสองชุดที่ได้จากกลุ่มตัวอย่างซึ่งเลือกมาขนาดใหญ่ใด ๆ มีลักษณะเหมือนกันแล้ว คะแนนดิบของแบบสอบทั้งสองชุด จึงจะสามารถเปรียบเทียบกันได้ หรือถ้าความเที่ยงของแบบสอบสองชุดนั้นเท่ากันในประชากรแล้ว ก็สามารถเปรียบเทียบการแจกแจงของค่าที่ได้เช่นกันความหมายดังกล่าวเป็นนิยามเชิงทฤษฎี ในทางปฏิบัติได้นิยามไว้ว่า ในประชากรที่กำหนดถ้าคะแนนจากแบบสอบสองชุดใด ๆ มีค่าเฉลี่ยเท่ากัน หรือเกือบเท่ากันในทุก ๆ กลุ่มตัวอย่างขนาดใหญ่ใด ๆ แล้วการเปรียบเทียบจะทำได้

เลวิน (Levine 1955 Cited by Holland and Rubin 1982: 10) ได้ให้ข้อตกลงเบื้องต้นของการเทียบมาตราไว้ว่า แบบสอบสองชุดใด ๆ จะเทียบมาตรากันได้ก็ต่อเมื่อแบบสอบทั้งสองนั้นคู่ขนานกัน (Parallel) ในด้านต่อไปนี้

1) โครงสร้าง (Structure)

2) เวลา (Timing)

3) รูปแบบ (Format)

4) ชนิดของข้อกระทง (Item type)

5) เนื้อหาวิชา (Subject Matter)

แองกอฟฟ์ (Angoff 1971: 562) ได้ให้ความหมายของการเทียบมาตราไว้ว่าเป็นวิธีการแปลงระบบหน่วยคะแนนของแบบสอบชุดหนึ่ง ไปสู่ระบบหน่วยคะแนนของแบบสอบอีกชุดหนึ่ง คะแนนที่ผ่านการแปลงแล้ว จะให้ความหมายของการสมมูลกันโดยตรง

ลอร์ด (Lord 1980: 195) ได้ให้ความหมายของการเทียบมาตราไว้ว่า เป็นวิธีการแปลงคะแนนจากแบบสอบต่างชุด ที่มีความยากต่างกันจากความสามารถเดียวกัน

ชูศักดิ์ ขัมภลิขิต (2527: 2) ได้สรุปความหมายของการเทียบมาตราไว้ว่า เป็นกระบวนการที่เกี่ยวโยงกับกิจกรรม 2 ประการ คือ

1) กระบวนการที่ทำให้แบบสอบ 2 ฉบับใด ๆ มีความทัดเทียมกัน หรือเท่ากันในเชิงโครงสร้าง

2) การใช้วิธีการทางสถิติเพื่อปรับ (adjust) คะแนนที่ได้จากแบบสอบแต่ละฉบับให้อยู่ในมาตราเดียวกัน และเทียบกันได้

จากความหมายดังกล่าว พอสรุปได้ว่าการเทียบมาตราคือกระบวนการทางสถิติ เพื่อปรับคะแนนที่ได้จากแบบสอบต่างชุดที่มีโครงสร้างเดียวกันให้เปรียบเทียบกันได้

ตอนที่ 2 รูปแบบการเทียบมาตรา (Equating Model)

รูปแบบการเทียบมาตราพิจารณาจากลักษณะรูปแบบของทฤษฎีการวัดแล้ว สามารถจำแนกได้ 2 รูปแบบ คือ การเทียบมาตรารูปแบบดั้งเดิม (Traditional Model) และรูปแบบอิงทฤษฎีการตอบข้อสอบ (TRT Model) ซึ่งมีหลักการในการเทียบมาตราที่เหมือนกัน แตกต่างกันในวิธีการเทียบมาตรา ผู้วิจัยขอเสนอรูปแบบในการเทียบมาตรา 3 รูปแบบดังต่อไปนี้

1. **การเทียบมาตราเชิงเส้นตรง** (Linear Equating)

การเทียบมาตราเชิงเส้นตรงมีนิยามว่า คะแนนของแบบสอบ 2 ชุด จะเท่าเทียมกัน ถ้าต่างก็ตรงกับคะแนนมาตรฐานค่าเดียวกัน (Angoff 1971: 564; Petersen and Other 1982: 73) เนื่องจากการเทียบมาตราเป็นการศึกษาเชิงประจักษ์ เพื่อกำหนดคะแนนแปลงที่ได้ จากแบบสอบชุดหนึ่งไปสู่แบบสอบอีกชุดหนึ่ง จึงมีองค์ประกอบหลายประการเข้ามาเกี่ยวข้อง ได้แก่ การออกแบบเพื่อเก็บรวบรวมข้อมูล และการจัดกระทำข้อมูลในทางสถิติเพื่อการแปลงคะแนน แองกอฟฟ์ (Angoff 1984: 93-121) ได้เสนอแบบแผนการรวบรวมข้อมูลและการจัดกระทำข้อมูลในทางสถิติของการเทียบมาตราเชิงเส้นตรงไว้ 6 รูปแบบ รูปแบบที่ 1-2 เป็นการเทียบมาตราโดยไม่ได้ใช้แบบสอบร่วม รูปแบบที่ 3-6 เป็นการเทียบมาตราโดยใช้แบบสอบร่วม ซึ่งแต่ละรูปแบบจะมีวิธีการประมาณค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของแบบสอบแตกต่างกันไป ตามเงื่อนไขของรูปแบบการรวบรวมข้อมูล แต่ทุกรูปแบบจะตัดสินคะแนนสมมูล จากค่าคะแนนมาตรฐานค่าเดียวกัน คือ

$$\frac{Y - M_Y}{S_Y} = \frac{X - M_X}{S_X}$$

เมื่อ X, Y คือ คะแนนจากแบบสอบฉบับ X, Y

M_X, M_Y คือ ค่าเฉลี่ยจากคะแนนแบบสอบฉบับ X, Y

S_X, S_Y คือ ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนแบบสอบฉบับ X, Y

จะได้สมการเทียบมาตราเชิงเส้นตรง

$$Y^* = e_Y(X) = AX + B$$

ในแต่ละรูปแบบการรวบรวมข้อมูลของแองกอฟฟ์ ซึ่งเสนอรายละเอียดแยกระหว่างกรณีที่แบบสอบเทียบมาตรามีความเที่ยงเท่ากัน และเมื่อแบบสอบเทียบมาตรามีความเที่ยงไม่เท่ากันจึงขอเสนอแบบแผนการรวบรวมข้อมูลของแองกอฟฟ์ 6 รูปแบบ โดยแยกประเด็นการจัดกระทำข้อมูลทางสถิติออก ดังมีภาพประกอบที่ 1 และรายละเอียดต่อไปนี้

a. Equivalent-Groups Design

	แบบสอบ	
กลุ่มตัวอย่าง	X	Y
P_1	/	
P_2		/

b. Counterbalanced Random-Groups Design

	แบบสอบ			
	X		Y	
กลุ่มตัวอย่าง	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 1	ครั้งที่ 2
P_1	/			/
P_2		/	/	

c. Anchor-Test-Random-Groups Design

	แบบสอบ		
กลุ่มตัวอย่าง	X	Y	U
P_1	/		/
P_2		/	/

d. Anchor-Test-Nonequivalent-Groups Design

	แบบสอบ		
กลุ่มตัวอย่าง	X	Y	U
P_1	/		/
Q_1		/	/

ภาพที่ 1: แสดงการออกแบบรวบรวมข้อมูลในการเทียบมาตรา (Petersen et al 1989: 244)

แบบแผนการรวบรวมข้อมูล (Data Collection Design)

รูปแบบที่ 1: กลุ่มสุ่มสองกลุ่มที่ถูกสุ่มมาจากประชากรเดียวกัน แต่ละกลุ่มทำแบบสอบเพียงหนึ่งชุด กล่าวคือ กลุ่ม $P_1(\alpha)$ ทำแบบสอบฉบับ X กลุ่ม $P_2(\beta)$ ทำแบบสอบฉบับ Y (ภาพที่ 1: a) กลุ่มทั้งสองควรมีความคล้ายคลึงกันทางความรู้ ความสามารถ เพื่อให้คะแนนสอบที่ได้ไม่ได้เกิดจากผลของความแตกต่างของกลุ่ม กลุ่มตัวอย่างที่ใช้ในรูปแบบนี้ควรมีขนาดใหญ่เพื่อลดความแตกต่างระหว่างกลุ่ม รูปแบบนี้จะมีข้อได้เปรียบคือไม่ได้รับอิทธิพลของการเรียนรู้ ความเหนื่อยล้า การฝึกฝน

รูปแบบที่ 2: กลุ่มสุ่มสองกลุ่มที่ถูกสุ่มมาจากประชากรเดียวกัน ทำแบบสอบทั้งสองชุดในลักษณะของการจัดลำดับการสอบก่อนหลังสลับกันให้เกิดความสมดุล กล่าวคือ กลุ่ม $P_1(\alpha)$ ทำแบบสอบฉบับ X แล้วตามด้วยฉบับ Y กลุ่ม $P_2(\beta)$ ทำแบบสอบฉบับ Y แล้วตามด้วยฉบับ X (ภาพที่ 1 :b) รูปแบบนี้คะแนนจากแบบสอบทั้งสองฉบับ จะได้รับผลกระทบจากองค์ประกอบการเรียนรู้ การฝึกฝน และความเหนื่อยล้าได้อย่างเท่าเทียมกัน

รูปแบบที่ 3: กลุ่มสุ่มสองกลุ่มที่ถูกสุ่มจากประชากรเดียวกัน แต่ละกลุ่มทำแบบสอบเพียงฉบับเดียว และทำแบบสอบร่วมที่เหมือนกันอีกส่วนหนึ่ง กล่าวคือกลุ่ม $P_1(\alpha)$ ทำแบบสอบฉบับ X และแบบสอบร่วม (U) อีกกลุ่ม $P_2(\beta)$ ทำแบบสอบฉบับ Y และแบบสอบร่วม (U) (ภาพที่ 1 :c) โดยแบบสอบร่วมอาจจะเป็นฉบับที่แยกจากแบบสอบชุด X และ Y (External Anchor Test) หรือรวมอยู่ในแบบสอบทั้งสองชุด (Internal Anchor Test) แบบสอบร่วมจะต้องถูกบริหารในลำดับที่เหมือนกันทั้งสองกลุ่ม แบบสอบร่วมควรจะคู่ขนานหรือเป็นตัวแทนของแบบสอบทั้งสองชุดนั้นให้มากที่สุด รูปแบบที่ 3 จะมีข้อได้เปรียบเพราะแบบสอบร่วมจะทำหน้าที่กำหนดสเกลร่วมของคะแนนจากแบบสอบแต่ละชุดที่สอบโดยผู้สอบทั้งสองกลุ่ม เพื่อนำไปปรับค่าความยากของข้อสอบ หรือค่าความสามารถของผู้สอบจากแบบสอบต่างชุดให้อยู่ในสเกลร่วมกัน แบบสอบร่วมจะลดความแปรปรวนของความคลาดเคลื่อนในการเทียบมาตรา เมื่อเปรียบเทียบกับรูปแบบที่ 1

รูปแบบที่ 4: กลุ่มที่ไม่ได้มาจากการสุ่มสองกลุ่ม แต่ละกลุ่มทำแบบสอบเพียงฉบับเดียว และทำแบบสอบร่วมที่เหมือนกันอีกส่วนหนึ่ง กล่าวคือ กลุ่ม $P_1(\alpha)$ ทำแบบสอบฉบับ X และแบบสอบร่วม U อีกกลุ่มคือ $Q_1(\beta)$ ทำแบบสอบฉบับ Y และแบบสอบร่วม U (ภาพที่ 1: d) กลุ่มผู้สอบที่ไม่ได้มาจากการสุ่มเป็นกลุ่มที่เกิดขึ้นในสถานการณ์จริง ซึ่งจะต้องสอบแบบสอบต่างชุดในเวลาต่างกัน ผู้สอบจึงไม่ได้สุ่มมาจากกลุ่มประชากรเดียวกัน เช่นการจัดสอบในแบบสอบฟอร์มใหม่ในปี 2527 และต้องการเทียบคะแนนไปยังแบบสอบฟอร์มเก่าที่สอบในปี 2526 แบบสอบร่วมจึงต้องมีความคล้ายคลึงกับแบบสอบสองชุดที่จะเทียบมาตรามากที่สุด (Klein & Jarjoura 1985) เพื่อลดอคติของการเทียบมาตราจากความแตกต่างในความสามารถระหว่างกลุ่ม

รูปแบบที่ 5: เป็นวิธีการเทียบมาตราโดยใช้แบบสอบร่วมที่เกี่ยวข้องกับคะแนน มีลักษณะการเทียบเป็นลักษณะเฉพาะคือ

5.1 แบบสอบฟอร์ม X และฟอร์ม Y เทียบไปสู่แบบสอบร่วม (Form X and Y equated to a common test) เป็นการบริหารแบบสอบร่วม ซึ่งอาจจะสอบแบบสอบร่วม แล้วตามด้วยแบบสอบชุด X และ Y หรือสอบแบบสอบชุด X และ Y และตามด้วยแบบสอบร่วม การเทียบมาตราโดยวิธีนี้คือการเทียบจากแบบสอบชุด X ไปยังแบบสอบร่วม และเทียบจากแบบสอบชุด Y ไปยังแบบสอบร่วม ซึ่งคะแนนที่เทียบแล้วของทั้งสองแบบสอบที่ระดับคะแนนของแบบสอบร่วมเดียวกันถือว่าเท่าเทียมกัน โดยแบบสอบร่วม U จะต้องมีลักษณะคู่ขนานกันกับแบบสอบฉบับ X และ Y แต่ถ้าแบบสอบร่วม U ไม่มีลักษณะคู่ขนานกับแบบสอบฉบับ X และ Y แล้วกลุ่ม α และ β ต้องมาจากการสุ่มจากประชากรเดียวกัน

5.2 แบบสอบร่วมเป็นตัวทำนายแบบสอบชุด X และ Y (Forms X and Y predicted by a common test) ซึ่งมีสมการแปลงคะแนนแตกต่างจากข้อ 5.1

5.3 แบบสอบชุด X และ Y เป็นตัวทำนายแบบสอบร่วม (Form X and Y predicting a common test) ซึ่งมีสมการแปลงคะแนนแตกต่างจากข้อ 5.1 และ 5.2

รูปแบบที่ 6: เป็นวิธีการเทียบมาตราที่ขึ้นกับลักษณะข้อกระทง ได้แก่ วิธีการของเธอร์สโตนและแฟน (Thorstone 1925 and Fan 1957 cited by Angoff 1984: 118-121)โดยให้แบบสอบทั้งสองชุดที่จะนำมาเทียบมาตรามีข้อสอบชุดหนึ่งร่วมกันคือค่าความยาก

ในการศึกษาครั้งนี้ ผู้วิจัยได้ใช้แบบแผนการรวบรวมข้อมูลรูปแบบที่ 3 ซึ่งเป็นกลุ่มสุ่ม 2 กลุ่ม แต่ละกลุ่มทำแบบสอบฉบับ X (ปีการศึกษา 2532) หรือฉบับ Y (ปีการศึกษา 2533) เพียงฉบับเดียว และทั้งสองกลุ่มทำแบบสอบร่วม (U) เนื่องจากแบบสอบที่นำมาเทียบมาตราเป็นแบบสอบที่สร้างขึ้นในแต่ละปีการศึกษา จึงมีความเที่ยงไม่เท่ากันการจัดกระทำข้อมูลตามวิธีการทางสถิติจึงใช้รูปแบบ 3B ที่แองกอฟฟ์ (Angoff 1984: 104-109) รวบรวมไว้ซึ่งเสนอการแปลงคะแนนตามความเที่ยงของแบบสอบที่นำมาเทียบมาตรา ดังมีสถิติที่ใช้ในการจัดกระทำข้อมูลที่รวบรวมจากรูปแบบที่ 3 ดังต่อไปนี้

สถิติที่ใช้ในการเทียบมาตรารูปแบบที่ 3 (Random groups - one test administered to each group, common equating test administered to both groups)

A. **แบบสอบที่มีความเที่ยงเท่ากัน** (Equally reliable tests)

คะแนนสอบที่ได้ส่วนใหญ่ขึ้นกับลักษณะ 2 ประการ คือ ความสามารถรายบุคคลและลักษณะของแบบสอบ ในการที่จะเปรียบเทียบผลการสอบของแต่ละบุคคลซึ่งสอบแบบสอบที่แตกต่างกัน จึงจำเป็นที่จะต้องปรับคะแนนเสียก่อน กล่าวคือการเทียบมาตรานั้นจะเป็นเพียงการเทียบความแตกต่างของคะแนน (ในค่าสถิติ) อันเป็นผลมาจากความแตกต่างของแต่ละบุคคลหรือของกลุ่ม ถ้ากลุ่ม α และ β ไม่ได้สุ่มมาจากประชากรเดียวกัน ความแตกต่างระหว่างกลุ่มอาจจะเป็นองค์ประกอบที่มีนัยสำคัญมากในการแปรเปลี่ยนของค่า A และ B ของสมการเส้นตรง และจะเป็นผลทำให้เกิดอคติ (bias) ในวิธีการเทียบมาตรา แม้ว่ากลุ่มจะถูกเลือกมาโดยวิธีการสุ่มความแตกต่างเล็กน้อยระหว่างกลุ่มอาจจะเกิดขึ้นได้ ซึ่งถ้าไม่ป้องกันก็จะเกิดอคติในสมการแปลงคะแนนซึ่งจะเป็นผลต่อการเปรียบเทียบภายหลัง

ในการที่จะควบคุมให้บังเกิดผล การเทียบมาตราจำเป็นต้องใช้คะแนนของแบบสอบร่วมฟอร์ม U ซึ่งเป็นข้อกระทงที่เพิ่มขึ้นหรือเป็นข้อร่วมกันระหว่างฟอร์ม X และฟอร์ม Y แบบสอบร่วมนี้ใช้สำหรับปรับความแตกต่างที่อาจจะพบระหว่างกลุ่ม α และ β ในการบริหารแบบสอบกลุ่ม α จะได้รับแบบสอบฟอร์ม X และฟอร์ม U และกลุ่ม β จะได้รับแบบสอบฟอร์ม Y และฟอร์ม U

ลอร์ด (Lord 1955a cited by: Angoff 1984: 105-106) ได้พัฒนาสมการสำหรับรูปแบบนี้ โดยใช้การประมาณค่าความเป็นไปได้สูงสุด (Maximum likelihood estimates) ประมาณค่าเฉลี่ย และค่าความแปรปรวนของแบบสอบฟอร์ม X และฟอร์ม Y โดยสมการเหล่านี้ คือ

$$\hat{\mu}_X = M_{X\alpha} + b_{XU\alpha}(\hat{\mu}_U - M_{U\alpha}) \quad (1)$$

$$\hat{\mu}_Y = M_{Y\beta} + b_{YU\beta}(\hat{\mu}_U - M_{U\beta}) \quad (2)$$

$$\hat{\sigma}_X^2 = S_{X\alpha}^2 + b_{XU\alpha}^2(\hat{\sigma}_U^2 - S_{U\alpha}^2) \quad (3)$$

$$\hat{\sigma}_Y^2 = S_{Y\beta}^2 + b_{YU\beta}^2(\hat{\sigma}_U^2 - S_{U\beta}^2) \quad (4)$$

เมื่อ $\hat{\mu}_U = M_{Ut}$ และ $\hat{\sigma}_U^2 = S_{Ut}^2$ และ $t = \alpha + \beta$

ซึ่งการประมาณค่าเหล่านี้ใช้กับสมการ $Y = Ax + B$

โดย $A = \hat{\sigma}_Y | \hat{\sigma}_X$ และ $B = \hat{\mu}_Y - A\hat{\mu}_X$

การเทียบมาตราด้วยรูปแบบที่ 3 นี้ ยังมีวิธีการที่ยืดหยุ่นได้ คือ สามารถเทียบแบบสอบที่มีมากกว่า 2 ฟอร์มขึ้นไป อาจจะเป็น 3 ฟอร์ม 4 ฟอร์ม หรือมากกว่านั้น เช่น กรณี 3 ฟอร์ม คือ ฟอร์ม X, Y และ Z แบบสอบทั้งสามฟอร์มนี้สอบโดยกลุ่ม α , β และ r ตามลำดับ ค่าเฉลี่ยและความแปรปรวนสำหรับกลุ่มรวม $(\alpha + \beta + r)$ จะประมาณได้โดยใช้ข้อตกลงเกี่ยวกับการเทียบมาตราของแบบสอบสองฟอร์ม

B. **แบบสอบที่มีความเที่ยงไม่เท่ากัน (Unequally reliable tests)**

เลวิน (Levine 1955 cited by Angoff 1984: 109) ได้เสนอไว้เมื่อแบบสอบฟอร์ม X และ Y มีความเที่ยงไม่เท่ากัน ซึ่งเหมาะสมในการแปลงโดยคะแนนจริงมากกว่าคะแนนที่สอบได้ (observed scores) ข้อตกลงที่เพิ่มเติมขึ้นสำหรับแบบสอบร่วม (ฟอร์ม U) นั้น จะต้องมีลักษณะคู่ขนานกันในโครงสร้าง กับแบบสอบฟอร์ม X และ Y

ภายใต้เงื่อนไขนี้ เมื่อฟอร์ม U แยกออกจากฟอร์ม X และ Y (External Anchor test) ความชัน (Slope) และจุดตัดแกน (Intercept) ของสมการ $Y = Ax + B$ คือ

$$A = b_{Y U \beta} | b_{X U \alpha} \quad ; \quad B = \hat{\mu}_Y - A\hat{\mu}_X$$

เมื่อ $b_{XU\alpha}$ และ $b_{YU\beta}$ เป็นสัมประสิทธิ์ถดถอย (Regression Coefficient) ในกลุ่ม α และ β สำหรับการทำนาย X จาก U และ Y จาก U

$\hat{\mu}_Y$ และ $\hat{\mu}_X$ คำนวณเช่นเดียวกับสมการ (1) และ (2)

สำหรับฟอร์ม U เมื่อรวมเข้าเป็นส่วนหนึ่งของฟอร์ม X และฟอร์ม Y มีลักษณะเป็นแบบสอบร่วมภายใน (Internal Anchor Test) จะได้ค่า

$$A = (b_{X U \alpha} \hat{\sigma}_Y^2) | (b_{Y U \beta} \hat{\sigma}_X^2)$$

$$B = \hat{\mu}_Y - A\hat{\mu}_X$$

ค่าของ $\hat{\mu}_X$, $\hat{\mu}_Y$, $\hat{\sigma}_X^2$, $\hat{\sigma}_Y^2$ คำนวณเช่นเดียวกับสมการ (1) ถึง (4)

2. **การเทียบมาตรารูปแบบอีควิเปอร์เซ็นไตล์** (Equipercentile equating)

ลอร์ด และฟลานาแกน (Lord 1950 and Flanagan 1951 cited by Angoff 1984: 86) ได้กล่าวว่า คะแนนสองชุด คะแนนตัวหนึ่งมาจากแบบสอบชุด X และอีกคะแนนมาจากแบบสอบชุด Y เมื่อ X และ Y วัดในสิ่งเดียวกัน ด้วยระดับความเที่ยงที่เท่ากัน จะสมมูลกัน ถ้าต่างก็เป็นค่าที่ตรงกับตำแหน่งเปอร์เซ็นไตล์เดียวกัน โดยปกติการเทียบมาตรารูปแบบอีควิเปอร์เซ็นไตล์นี้ หากแบบสอบสองชุดที่นำมาเทียบมาตรามีความยากใกล้เคียงกัน จะให้การเทียบมาตราที่เป็นเส้นตรง หากแบบสอบสองชุดมีความยากต่างกันจะให้การเทียบมาตราเป็นเส้นโค้ง (Curvilinear) คะแนนสมมูลที่เกิดขึ้นจะทำการย่อหรือขยายคะแนนดิบ เพื่อคงมาตราคะแนนให้เหมือนชุดก่อนตามต้องการ (Angoff 1984: 86-87)

ขั้นตอนในการเทียบมาตรารูปแบบอีควิเปอร์เซ็นไตล์มี 2 ขั้นตอน(two-stage) (Kolen, 1984) คือ

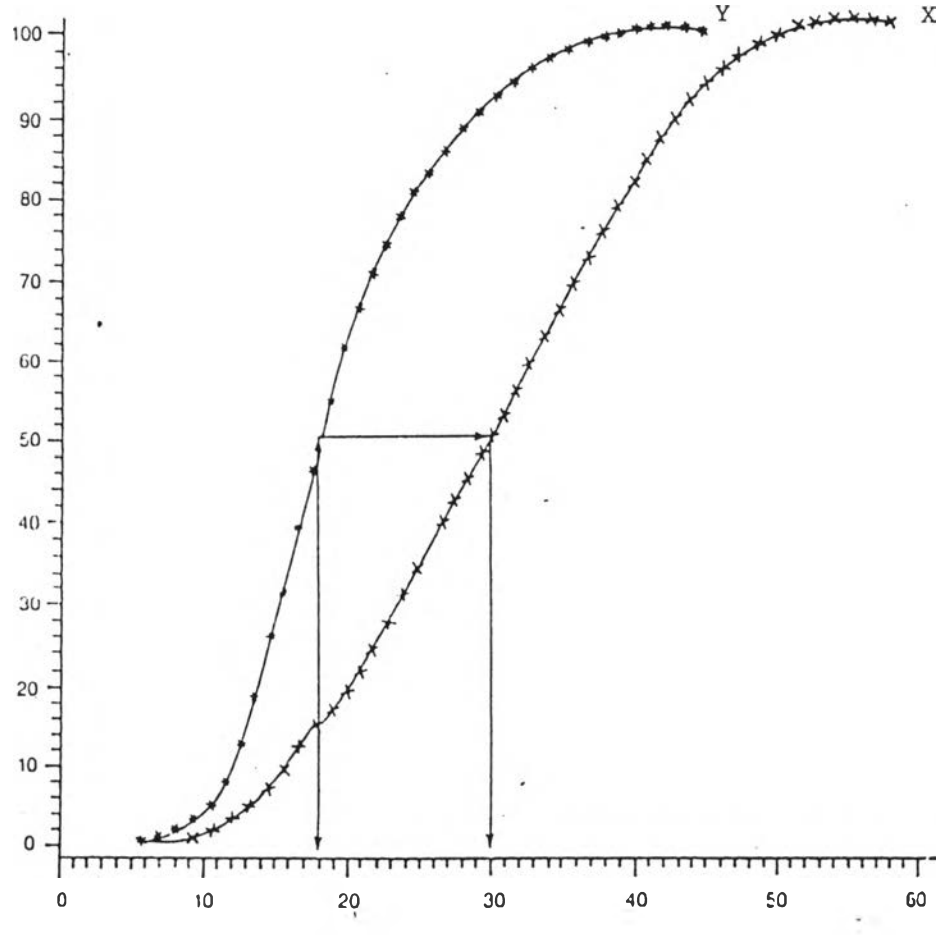
ขั้นตอนที่หนึ่ง การกระจายความถี่สะสมที่มีความสัมพันธ์เป็นตารางหรือเป็นกราฟสำหรับสองแบบสอบที่จะนำมาเทียบมาตรา คือ

1) นำคะแนนของกลุ่มผู้สอบที่มีความสามารถกระจายทั้ง เก่ง ปานกลาง และอ่อน ซึ่งถูกแบ่งเป็นกลุ่มย่อยสองกลุ่มโดยการสุ่ม ให้กลุ่มหนึ่งทำแบบสอบ X และอีกกลุ่มทำแบบสอบ Y มาทำการแจกแจงคะแนน X และ Y

2) คำนวณหาจุดกลางเปอร์เซ็นไตล์ของแต่ละการแจกแจง

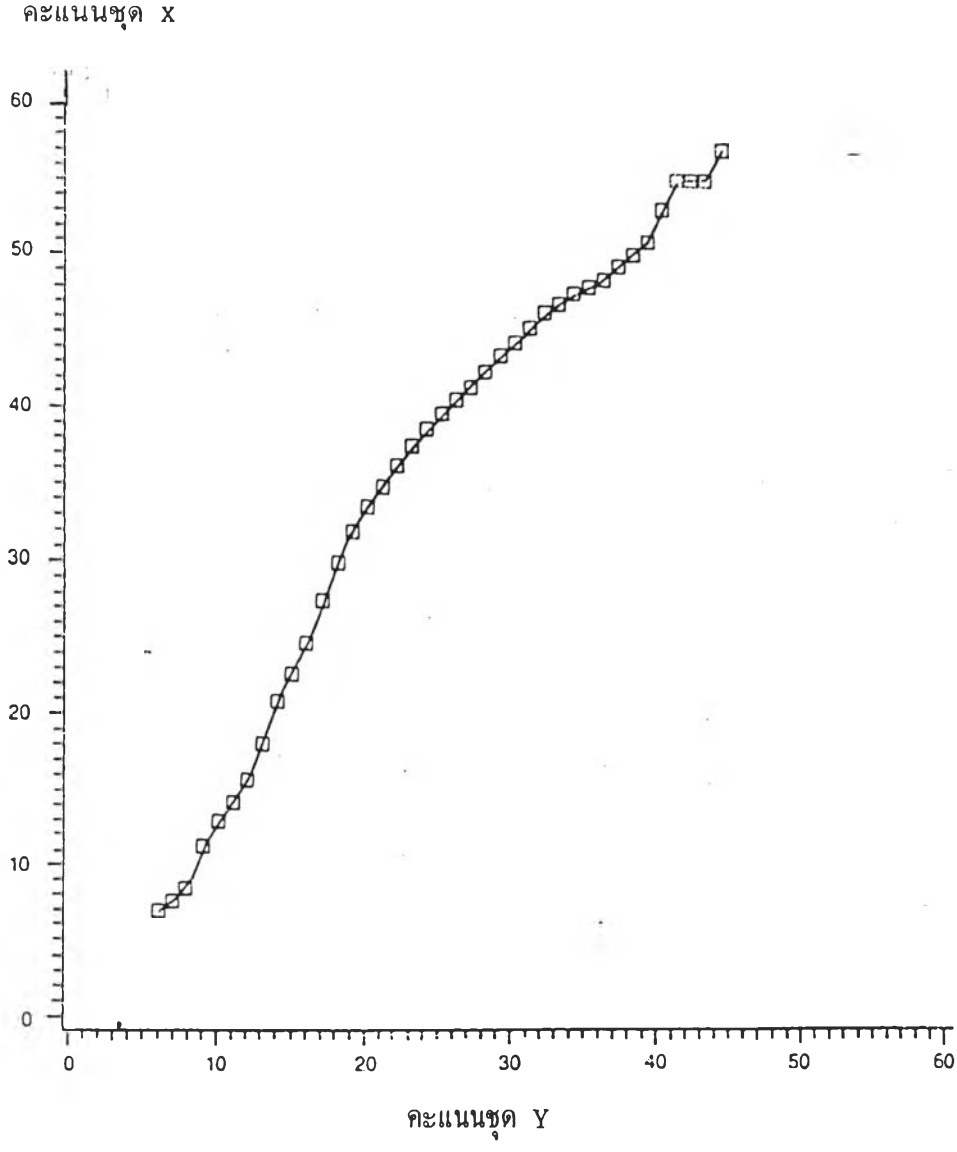
3) อ่านและทำเครื่องหมายสำหรับค่าคะแนนของแบบสอบชุด X และชุด Y ของการแจกแจงที่สมมูลกันบนกระดาษกราฟ ซึ่งแองกอฟฟ์ (Angoff 1984: 98) แนะนำให้ใช้กระดาษ Arithmetic graph paper (โดยแกนนอนเป็นคะแนนดิบ แกนตั้งเป็นตำแหน่งเปอร์เซ็นไตล์ดังภาพที่ 2) ประมาณ 30 จุด และลากเส้นเชื่อมเกิดเป็นเส้นกราฟ

เปอร์เซ็นไตล์



ภาพที่ 2: กระบวนการเทียบมาตรารูปแบบอีควิเปอร์เซ็นไตล์

ขั้นตอนที่สอง เทียบมาตราคะแนน จากรายละเอียดในข้อ (3) ขั้นตอนที่ 1 นำมาพล็อตกราฟใหม่ลงบนกระดาษกราฟ โดยแกนนอนเป็นคะแนนฟอร์ม Y แกนตั้งเป็นคะแนนฟอร์ม X ทำการปรับเส้นกราฟให้เรียบ เส้นกราฟนี้จะใช้อ่านค่า X ที่สมมูลกันกับ Y จากนั้นสร้างตารางสำเร็จ เพื่ออ่านค่าคะแนนแปลงจากกราฟ ดังภาพที่ 3



ภาพที่ 3: การแปลงคะแนนของการเทียบมาตรารูปแบบอีควิเปอร์เซ็นไทล์

ในการเทียบมาตรารูปแบบอีควิเปอร์เซ็นไตล์ หากรวบรวมข้อมูลโดยใช้รูปแบบ Counterbalance นั้น จะต้องรวมข้อมูลผู้เข้าสอบที่สอบในแบบสอบฟอร์ม X และ Y ก่อน คือรวมกลุ่มคะแนนของผู้สอบกลุ่ม α และ β สำหรับฟอร์ม X และรวมคะแนนของผู้สอบกลุ่ม α และ β สำหรับฟอร์ม Y แล้วจึงดำเนินการเทียบมาตราตามวิธีการข้างต้น

การเทียบมาตรารูปแบบอีควิเปอร์เซ็นต์ไตล์โดยใช้แบบสอบร่วม

ลอร์ดและเลวิน (Lord 1957 and Levine 1958 cited by Angoff 1984: 113-115) ได้เสนอขั้นตอนการเทียบมาตรารูปแบบอีควิเปอร์เซ็นไตล์โดยใช้แบบสอบร่วมสำหรับกลุ่มที่ไม่ได้สุ่มซึ่งมีความสามารถไม่แตกต่างกันมาก (Curvilinear method for groups not widely different in ability) โดยที่กลุ่ม α สอบฟอร์ม X และฟอร์ม U กลุ่ม β สอบฟอร์ม Y และฟอร์ม U ขั้นตอนในการเทียบมาตราคือประมาณค่าความถี่ในแบบสอบฟอร์ม X และ Y สำหรับกลุ่มรวม $t(\alpha + \beta)$ มีขั้นตอนดังนี้

1. รวมคะแนนของฟอร์ม U ที่สอบโดยกลุ่ม α, β และกลุ่ม t
2. หาสัดส่วนความถี่ $f_{it}/f_{i\alpha}$ และ $f_{it}/f_{i\beta}$ ที่ทุกช่วงคะแนน i
3. กระจายความถี่ของฟอร์ม U ที่ทุกช่วงคะแนนของฟอร์ม X และ Y
4. คูณความถี่ของฟอร์ม U ในแต่ละช่วงคะแนนของฟอร์ม X ด้วยสัดส่วน $f_{it}/f_{i\alpha}$
5. คูณความถี่ของฟอร์ม U ในแต่ละช่วงคะแนนของฟอร์ม Y ด้วยสัดส่วน $f_{it}/f_{i\beta}$
6. หาตำแหน่งเปอร์เซ็นไตล์ของความถี่ที่ประมาณค่าใหม่ของทั้งสองชุด
7. ดำเนินการเทียบตามขั้นตอนที่สองของการเทียบโดยวิธีอีควิเปอร์เซ็นไตล์ดังกล่าวแล้วข้างต้น

3. **การเทียบมาตราโดยใช้ทฤษฎีการตอบข้อสอบ**

ลอร์ด (Lord 1980: 199-203) ได้กล่าวถึงการเทียบมาตราโดยใช้ทฤษฎีการตอบข้อสอบไว้ว่า แบบสอบสองฉบับใด ๆ ที่ใช้เทียบมาตรากันต้องเป็นแบบสอบที่วัดคุณลักษณะเดียวและเทียบมาตราคะแนนที่สมมูลกันที่ระดับความสามารถเดียวกัน ซึ่งมีข้อกำหนด 3 ประการดังนี้

1. ความเสมอภาค (Equity) หมายถึง สำหรับทุก ๆ ความสามารถ (θ) การแจกแจงความถี่แบบมีเงื่อนไขของคะแนนแปลง ต้องเหมือนกับการแจกแจงความถี่แบบมีเงื่อนไขของคะแนนแบบสอบที่ต้องการเทียบ

2. ความไม่แปรผันตามกลุ่ม (Invariance Across Groups) หมายถึง คะแนนแปลงจะคงที่ โดยไม่แปรเปลี่ยนไปตามประชากรที่นำมาสร้างสมการเทียบมาตรา

3. ความสมมาตร (Symmetry) หมายถึง คะแนนแปลงที่ได้จากการเทียบมาตราจะต้องเหมือนกัน ไม่ว่าจะเป็นการเทียบมาตราจากแบบสอบชุด X ไปยังแบบสอบชุด Y หรือจากแบบสอบชุด Y ไปยังแบบสอบชุด X ก็ตาม

นอกจากนี้ ลอร์ดยังได้แบ่งรูปแบบการเทียบมาตราโดยใช้ทฤษฎีการตอบข้อสอบออกเป็นสองวิธีการใหญ่ ๆ ดังนี้

1. **วิธีการเทียบมาตราโดยใช้คะแนนจริง** (True-score equating)

แบบสอบที่มีความยากต่างกัน ถึงแม้ว่าจะนำไปสอบกับประชากรที่มีความสามารถเท่ากันหรือคุณลักษณะเดียวกัน แต่คะแนนที่ได้จะมีการแจกแจงแตกต่างกัน ดังนั้นถ้าสามารถหาคะแนนจริงได้ ก็จะสามารถเทียบมาตราคะแนนระหว่างความสามารถ (θ) กับคะแนนจริง ได้จากความสัมพันธ์เชิงคณิตศาสตร์ (Mathematical Model) ดังนี้

$$\xi = \xi(\theta) = \sum_{i=1}^{n_X} P_i(\theta) \qquad (5)$$

$$\eta = \eta(\theta) = \sum_{j=1}^{n_Y} P_j(\theta) \qquad (6)$$

เมื่อ n_x, n_y คือจำนวนข้อสอบในแบบสอบชุด X และ Y ตามลำดับ

ξ , η คือคะแนนจริงจากแบบสอบฟอร์ม X และ Y ตามลำดับ

ดังนั้น ถ้าแทนค่าความสามารถ (θ) ใด ๆ ลงในสมการที่ θ เท่ากัน จะหาคะแนนสมมูลของแบบสอบฉบับ X และฉบับ Y ได้ โดยคำนวณจากค่า $P_i(\theta)$ และ $P_j(\theta)$ และในทางตรงกันข้ามก็สามารถประมาณค่าความสามารถ (θ) เมื่อทราบคะแนนจริง

การหาค่า $P_i(\theta)$ และ $P_j(\theta)$ ได้จากการประมาณค่าพารามิเตอร์ของข้อสอบ โดยเลือกว่าจะใช้ชนิด one, two, three-parameter จากนั้นนำคะแนนจริงของแบบสอบฉบับ X และฉบับ Y มาหาความสัมพันธ์กันโดยใช้ค่าความสามารถ (θ) ที่ระดับเดียวกัน

ในการเทียบมาตราโดยใช้คะแนนจริงนี้ ใช้กับรูปแบบที่มีแบบสอบร่วมด้วย จุดมุ่งหมายของการใช้แบบสอบร่วม ก็คือ การเชื่อมข้อมูลเข้าด้วยกัน เพื่อทำให้ค่าพารามิเตอร์ที่วิเคราะห์ออกมาอยู่บนมาตราเดียวกัน ถ้าขาดข้อสอบร่วมแล้ว จะไม่สามารถเทียบกันได้ นอกเสียจากว่ากลุ่มผู้สอบฉบับ X และฉบับ Y จะมีการกระจายความสามารถเหมือนกัน (Lord 1980: 202) ในทางปฏิบัติจะหาค่าความสามารถ (θ) ของกลุ่มตัวอย่างที่ทำแบบสอบร่วม ไปเทียบมาตรากับแบบสอบฟอร์ม X และ Y

2. การเทียบมาตราโดยใช้คะแนนสังเกต (Observed scord equating)

ปัญหาการเทียบมาตราโดยใช้คะแนนจริง คือ ไม่สามารถทราบคะแนนจริงของแต่ละคนได้ นอกจากใช้วิธีประมาณจากผลการสอบ โดยสมการ

$$\xi = \sum_{i=1}^{n} P_i(\theta) \qquad (7)$$

ซึ่งคะแนนที่ได้เป็นเพียงค่าประมาณเท่านั้น ยังไม่มีคุณสมบัติเป็นคะแนนจริง (Lord 1980: 197) ดังนั้นการเทียบมาตราโดยใช้คะแนนจริง จึงเสมือนการเทียบมาตรา โดยใช้คะแนนสังเกตชนิดใหม่

ลอร์ดจึงเสนอการเทียบมาตราโดยใช้คะแนนสังเกต โดยอาศัยข้อมูลจากการตอบแบบสอบร่วม (Anchor test) การเทียบมาตราวิธีนี้ เริ่มด้วยการประมาณการกระจายความสามารถของกลุ่มรวม $r(\theta)$ ซึ่งหมายถึงผู้เข้าสอบทั้งหมดที่ทำแบบสอบร่วม การกระจายความสามารถ (θ) ในกลุ่ม เป็นการประมาณการกระจายของ $r(\theta)$

การประมาณการกระจายของคะแนนสังเกต X สำหรับกลุ่มโดยสมการ

$$\hat{\phi}_X(X) = \frac{1}{N} \sum_{a=1}^{N} \hat{\phi}_X(X|\hat{\phi}_a) \qquad (8)$$

เมื่อ a = 1, 2, ..., N คือผู้สอบในแต่ละระดับ θ ของกลุ่มรวม

ถ้า $\hat{r}(\theta)$ เป็นค่าต่อเนื่องจะประมาณ $\phi(X)$ จากสมการ

$$\hat{\phi}_X(X) = \int_{-\alpha}^{\alpha} \hat{\phi}_X(X|\theta)\hat{r}(\theta)d\theta \qquad (9)$$

เมื่อ $\hat{\phi}_X(X|\hat{\theta}_a)$ ได้จากการประมาณค่าพารามิเตอร์ โดยใช้ทฤษฎีการตอบข้อสอบ item parameter ของข้อสอบในแบบสอบฉบับ X และสมการต่าง ๆ สามารถนำมาประยุกต์ใช้กับแบบสอบฉบับ Y ได้เช่นเดียวกัน

เนื่องจากแบบสอบ X และ Y เป็นอิสระจากกัน เมื่อกำหนดความสามารถ (θ) คงที่ ก็สามารถประมาณส่วนที่เป็นการกระจายร่วม (joint distribution) ของคะแนนจากแบบสอบฉบับ X และ Y การกระจายร่วมของความสามารถประมาณได้จากสมการ

$$\hat{\phi}(X, Y) = \frac{1}{N} \sum_{a=1}^{N} \hat{\phi}_X(X|\hat{\theta}_a)\hat{\phi}_Y(Y|\hat{\theta}_a) \qquad (10)$$

หรือ

$$\hat{\phi}(X, Y) = \int_{-\alpha}^{\alpha} \hat{\phi}_X(X|\theta)\hat{\phi}_Y(Y|\theta)\hat{r}(\theta)d\theta \qquad (11)$$

จากสมการข้างต้น จะเห็นบทบาทของแบบสอบร่วมชัดเจน ทำให้สามารถประมาณการแจกแจงร่วมของ X และ Y ถึงแม้จะไม่มีผู้ใดสอบทั้ง 2 ฉบับ

สมการ (11) เป็นการแจกแจงร่วมกันของตัวแปร 3 ตัว คือ θ , X และ Y จากที่ θ เป็นตัวระบุคะแนนจริง ξ และ η การแจกแจงนี้ จึงเป็นการแจกแจงร่วมกันของตัวแปร 4 ตัว คือ ξ η X และ Y ซึ่งประมาณได้โดยการระบุความสัมพันธ์เชิงอิควิเปอร์เซ็นไตล์ระหว่าง X และ Y

จากการเทียบมาตราทั้ง 2 แบบ คือ การเทียบมาตราด้วยคะแนนจริงและการเทียบมาตราด้วยคะแนนสังเกต มีทั้งข้อดีและข้อเสีย กล่าวคือ วิธีเทียบมาตราด้วยคะแนนจริงไม่สามารถอธิบายคะแนนที่อยู่ต่ำกว่าระดับการเดาได้ โดยจะให้ความหมายของคะแนนสมมูลเฉพาะคะแนนที่อยู่เหนือค่าเฉลี่ยของการเดา ส่วนวิธีการเทียบมาตราด้วยคะแนนสังเกต ก็มีจุดอ่อนที่เป็นเพียงการเทียบมาตราอย่างประมาณ ยังคงมีความคลาดเคลื่อนอยู่ ถึงแม้การเทียบมาตราด้วยคะแนนสังเกต จะสามารถอธิบายคะแนนสมมูลจาก X และ Y ได้ครอบคลุมพิสัยของคะแนนที่สังเกตได้ก็ตาม

ลอร์ดได้กล่าวถึง การเทียบมาตราทั้งสองวิธีนี้ว่ามีความสอดคล้องกันมาก แต่การสรุปผล เพื่อนำไปใช้อ้างอิงต้องทำอย่างพิถีพิถัน (Lord 1980: 202-203) อย่างไรก็ตามการเทียบมาตราโดยใช้คะแนนสังเกตนี้ยังนำไปใช้ในงานวิจัยน้อยมาก เนื่องจากวิธีนี้มีความซับซ้อน และลงทุนแพงกว่าวิธีการเทียบมาตราโดยใช้คะแนนจริง (Lord and Wingersky 1984: 453)

ตอนที่ 3 แบบสอบร่วมในการเทียบมาตรา (anchor test)

แบบสอบร่วมถูกออกแบบมาเพื่อใช้ในการปรับความแตกต่างในการสุ่มตัวอย่างที่อาจจะพบระหว่างกลุ่ม เพื่อให้การเทียบมาตราเป็นการเทียบความแตกต่างของคะแนน อันเป็นผลมาจากการแตกต่างของแต่ละบุคคลหรือของกลุ่ม ในการออกแบบสำหรับเทียบมาตรา ถ้ากลุ่ม α และ β ไม่ได้สุ่มมาจากประชากรเดียวกันความแตกต่างระหว่างกลุ่มอาจจะเป็นองค์ประกอบที่มีนัยสำคัญจะเป็นผลทำให้เกิดอคติ (bias) ในการเทียบมาตราแม้ว่ากลุ่มจะถูกเลือกมาอย่างสุ่ม ความแตกต่างเล็กน้อยระหว่างกลุ่มอาจจะเกิดขึ้นได้ ถ้าไม่ป้องกันก็จะเกิดอคติ ในสมการแปลงคะแนนซึ่งจะเป็นผลต่อการเปรียบเทียบภายหลัง

แบบสอบร่วมเป็นข้อกระทงที่เพิ่มขึ้นหรือเป็นข้อร่วมกันระหว่างแบบสอบฟอร์ม X และฟอร์ม Y ในการออกแบบรวบรวมข้อมูล กลุ่มหนึ่งจะได้รับแบบสอบฟอร์ม X และแบบสอบร่วม กลุ่มสอง จะได้รับแบบสอบฟอร์ม Y และแบบสอบร่วม แบบสอบร่วมจึงถูกสอบจากทั้งสองกลุ่มในลำดับที่เหมือนกันทั้ง 2 กลุ่ม ดังนั้น คะแนนบนแบบสอบร่วมในฟอร์ม X และ Y จะมีผลกระทบในทางเดียวกัน เช่น การเรียนรู้ ความเหนื่อยล้า และผลกระทบจากการฝึกฝน

ประเภทของแบบสอบร่วม

การเทียบมาตราด้วยการใช้แบบสอบร่วมนี้มีประโยชน์มาก สามารถยืดหยุ่นและปรับให้เข้ากับสถานการณ์อื่น ๆ ได้ โดยแบบสอบร่วมอาจจะรวมหรือแยกออกจากฟอร์ม X และฟอร์ม Y ได้ดังนี้

1. แบบสอบร่วมภายใน (Internal Anchor test) เป็นแบบสอบย่อยของข้อกระทง ซึ่งบรรจุในแบบสอบทั้งคู่ที่จะถูกเทียบมาตรา คะแนนในกลุ่มของข้อกระทงร่วม จะถูกใช้ในการคำนวณคะแนนบนแบบสอบทั้งหมด การกระจายแบบสอบร่วมควรกระจายตลอดทั้งฉบับ รูปแบบการใช้แบบสอบร่วมไม่ได้ทำให้ใช้เวลาทดสอบมากขึ้นเมื่อเทียบกับรูปแบบ equivalent-group design (รูปแบบที่ 1)

2. แบบสอบร่วมภายนอก (External Anchor test) คะแนนบนแบบสอบร่วมภายนอกจะไม่ถูกใช้ในการคำนวณบนแบบสอบทั้งหมด การใช้แบบสอบร่วมภายนอกต้องใช้เวลาในการทำการทดสอบมากกว่ารูปแบบ equivalent-group design แต่ถ้าแบบสอบร่วมมีขนาดสั้นก็ไม่จำเป็นต้องใช้เวลามากเท่ากับรูปแบบ Counterbalanced random-group designs (รูปแบบที่ 2) ในกรณีการแยกสอบคนละเวลา ตัวอย่างเช่น จัดสอบภายหลัง การสอบฟอร์ม X และ Y หรือจัดสอบก่อนฟอร์ม X และ Y จะต้องจัดลำดับให้เหมือนกัน เพราะจะมีผลเท่ากันในทางปฏิบัติข้อกระทงในฟอร์ม U ไม่ควรซ้ำกันกับข้อกระทงของฟอร์ม X และฟอร์ม Y

กรณีซึ่งยืดหยุ่นได้ คือ แบบสอบร่วมที่สอบโดยกลุ่ม α ไม่จำเป็นต้องเหมือนกับแบบสอบร่วมที่สอบโดยกลุ่ม β อาจจะเป็นแบบสอบร่วมครึ่งหนึ่ง (quasi-common test) หรืออาจจะเป็นแบบสอบสองฟอร์มที่แตกต่างกัน แต่แบบสอบร่วมทั้งสองฟอร์มจะต้องวัดในเรื่องเดียวกัน กระบวนการในการเทียบมาตราคือแปลงแบบสอบร่วมชุดที่หนึ่งให้อยู่บนสเกลเดียวกับแบบสอบร่วมชุดที่สอง หรือทั้งสองฟอร์มอาจจะแปลงไปเป็นสเกลใหม่เลยก็ได้ แล้วจึงใช้วิธีการในรูปแบบที่ 3 และที่ 4 ของแองกอฟฟ์ ดังที่กล่าวมาแล้วตอนต้น ปรับความแตกต่างระหว่างกลุ่ม

ความยาวของแบบสอบร่วม

แบบสอบร่วมควรจะมีความยาว และความเที่ยงเพียงพอที่จะให้ข้อมูลไปปรับความแตกต่างระหว่างกลุ่มตามที่คาดหวัง ซึ่งแองกอฟฟ์ (Angoff 1984: 107) ได้เสนอกฎเกณฑ์เกี่ยวกับจำนวนข้อของแบบสอบร่วม คือ ไม่ควรน้อยกว่า 20 ข้อ หรือ 20% ของจำนวนข้อในแต่ละฟอร์มแล้วแต่จำนวนใดจะมากกว่าให้ใช้จำนวนนั้น

ขณะเดียวกันไรท์และสโตน (Wright and Stone 1979:98) ได้กล่าวว่าแบบสอบร่วมควรเป็นข้อสอบที่วัดเรื่องเดียวกันกับแบบสอบที่ต้องการศึกษา และมีเพียง 10 ข้อ ก็เพียงพอแล้ว

ต่อมาปี ค.ศ.1985 บูเดส (Budescu) ได้ศึกษาวิเคราะห์เชิงทฤษฎีชี้ให้เห็นความสัมพันธ์ของความยาวของแบบสอบร่วมกับประสิทธิภาพการเทียบมาตรา ซึ่งพบว่าตัวแปรสำคัญที่ส่งผลถึงประสิทธิภาพของการเทียบมาตรา คือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบสอบร่วมกับแบบสอบสองชุดที่ต้องการเทียบมาตรา ซึ่งอธิบายได้ดังต่อไปนี้

การวิเคราะห์เชิงทฤษฎีของบูเดส

การเทียบมาตราโดยใช้แบบสอบร่วมนั้น ออกแบบรวบรวมข้อมูลโดยให้กลุ่ม α ทำแบบสอบชุด X กลุ่ม β ทำแบบสอบชุด Y และทั้งสองกลุ่มทำแบบสอบร่วม U เหมือนกัน การหาค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแบบสอบทั้งสองชุด และของแบบสอบร่วมจากกลุ่มตัวอย่าง α , β และกลุ่มตัวอย่างรวม $T(\alpha + \beta)$ หาได้โดยตรงจากผลการสอบโดยใช้วิธีการประมาณค่าของทัคเคอร์ (Tucker equating Method) กรณีที่กลุ่มผู้สอบไม่ได้มาจากการสุ่ม และใช้วิธีการประมาณค่าแบบวิธีเป็นไปได้สูงสุด (Maximum likelihood) กับกลุ่มผู้สอบอย่างสุ่ม

จะได้ค่าสหสัมพันธ์ระหว่างคะแนนในแบบสอบร่วม U กับคะแนนจากแบบสอบ X และ Y และค่าความแตกต่างระหว่างคะแนนในแบบสอบร่วม U ของกลุ่มตัวอย่างทั้งสอง ถ้าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบสอบร่วมกับชุด X หรือ Y เป็นศูนย์ แบบสอบร่วมที่นำมาใช้ไม่มีความหมาย ตรงกันข้ามถ้าสัมประสิทธิ์สหสัมพันธ์ดังกล่าวมีค่าเพิ่มสูงขึ้น จะมีผลต่อการเพิ่มคุณภาพของตัวประมาณประชากรรวมอย่างสูงขึ้นตามกัน (monotonic increasing) การจะให้ได้มาซึ่งความสัมพันธ์สูง ๆ จะต้องดำเนินการสร้าง และจัดชุดแบบสอบร่วมอย่างพิถีพิถันที่ให้มีลักษณะคู่ขนานกับแบบสอบทั้งสองชุด

บูเดส (1985: 15) ได้วิเคราะห์ความยาวของแบบสอบร่วม และแสดงให้เห็นความสำคัญของความยาวของแบบสอบร่วม ดังนี้

กลุ่ม α ทำชุด X และ U ได้คะแนนรวม Z

ดังนั้น Z ของแต่ละคน คือ $Z = X + U$ (12)

ให้ความเที่ยงของแบบสอบรวม (X + U) เป็น r_{zz} แสดงในเทอมของพารามิเตอร์ของ X และ U ดังนี้ (Horst 1968: 281 cited by Budescu 1985: 16)

$$r_{ZZ} = \frac{(S_X^2 r_{XX} + S_U^2 r_{UU} + 2r_{XU} S_U S_X)}{S_X^2 + S_U^2 + 2r_{XU} S_U S_X} \qquad (13)$$

เมื่อ r_{XX} และ r_{UU} เป็นค่าประมาณความเที่ยงของแบบสอบชุด X และแบบสอบร่วม U ให้ $p(o<p<1)$ เป็นสัดส่วนของข้อกระทงของแบบสอบชุด X ที่อยู่ใน Z และ $q = (1-p)$ เป็นสัดส่วนของแบบสอบร่วม U ในแบบสอบรวม Z ถ้า 2 แบบสอบ X และ U ต่างคู่ขนานกัน ความเที่ยงเขียนในรูปของฟังชั่นของ r_{zz} โดยอาศัยสูตร Spearman-Brown ได้

$$r_{xx} = \frac{pr_{zz}}{(1 - qr_{zz})} \quad \text{และ} \quad r_{uu} = \frac{qr_{zz}}{(1 - pr_{zz})} \tag{14}$$

ในทำนองเดียวกัน ถ้า s_z^2 เป็นความแปรปรวนของแบบสอบฉบับรวม Z ความแปรปรวนของส่วนประกอบสามารถเขียนในรูปของฟังชั่นของความสัมพันธ์ของความยาวได้ดังนี้

$$s_x^2 = s_z^2\, p(1 - qr_{zz}) \quad \text{และ} \quad s_u^2 = s_z^2\, q(1 - pr_{zz}) \tag{15}$$

แทนสมการ (14) และ (15) ใน (13) จะพบว่า r_{xu} เป็นฟังชั่นของความเที่ยงของแบบสอบฉบับรวม Z และความยาวของส่วนประกอบ X และ U ดังนี้

$$r_{xu} = r_{zz} \left[\frac{pq}{(1 - pr_{zz})(1 - qr_{zz})} \right]^{\frac{1}{2}} \tag{16}$$

นิพจน์ในสูตร (16) เป็นอิสระจากค่าเฉลี่ยและความแปรปรวนรวมของแบบสอบที่เกี่ยวข้องทั้ง 2 ดังนั้น สามารถเอาไปใช้กับกลุ่มผู้สอบอีกกลุ่ม (β) ได้เหมือน ๆ กัน สรุปจากการวิเคราะห์ได้ว่า โดยธรรมชาติแล้วสหสัมพันธ์ระหว่างแบบสอบ X กับแบบสอบร่วมเป็นฟังก์ชั่นเพิ่มขึ้นตามความเที่ยงของแบบสอบฉบับรวม (r_{zz}) เมื่อความเที่ยงของแบบสอบรวม (r_{zz}) คงที่ค่าสหสัมพันธ์ระหว่างแบบสอบ X กับแบบสอบร่วม (r_{xu}) มีค่าสูงสุดที่ความยาวแบบสอบทั้งสองเท่ากัน ($p = q = .5$) ค่า r_{xu} มีลักษณะสมมาตรและเพิ่มขึ้นตามผลคูณของ p และ q จากจุดสำคัญนี้ทำให้ได้ค่าอ้างอิง คือ ประสิทธิภาพสูงสุดของการเทียบมาตราเมื่อจัดสัดส่วนของแบบสอบที่ต้องการเทียบกับแบบสอบร่วมให้มีสัดส่วนเท่ากัน ในกรณีเช่นนี้จะได้สหสัมพันธ์ระหว่างแบบสอบที่ต้องการเทียบกับแบบสอบร่วม (r_{xu}) มีค่าเท่ากับ $r_{zz}/(2 - r_{zz})$ (1985: 16-17)

จากการวิเคราะห์ที่กล่าวมาข้างต้น บูเดส (1985: 17) ได้นำมาสร้างเป็นดัชนีของประสิทธิภาพสัมพัทธ์ของการเทียบมาตรา เรียกว่า ความพร่องสัมพัทธ์ (relative deficiency) ตามนิยามของสมการ (16) ดังนี้

$$R.D.\ (q) = 1 - \frac{r_{XU(q)}}{r_{XU(0.5)}} = 1 - \left[\frac{(2 - r_{zz})^2 pq}{(1 - pr_{zz})(1 - qr_{zz})} \right]^{\frac{1}{2}} \tag{17}$$

มาตรการนี้จะแสดงให้เห็นถึงการเพิ่มประสิทธิภาพในเชิงเปรียบเทียบกับค่าอ้างอิงสูงสุด เมื่อ $p = q = .5$ ดังนั้น กระบวนการเทียบมาตราใด ๆ ที่มีค่าความพร่องสัมพัทธ์ต่ำกว่าหรือเท่ากับ 0.10 จัดว่าเป็นการเทียบมาตราที่สามารถให้ผลที่น่าพอใจ (1985: 17)

ในกรณีสูตรสำเร็จของแองกอฟฟ์ที่ใช้ 20% ถ้าจะให้มีประสิทธิภาพสัมพัทธ์ในระดับที่น่าพอใจแล้ว จะต้องเป็นแบบสอบที่มีค่าความเที่ยงสูงกว่า 0.866 (1985: 17)

กล่าวโดยสรุป การวิเคราะห์ของบูเดสเป็นการวิเคราะห์เชิงทฤษฎีที่ชี้ให้เห็นถึงความสัมพันธ์ของความยาวแบบสอบร่วมกับประสิทธิภาพของการเทียบมาตรา ในเชิงปฏิบัติไม่สามารถให้ข้อกำหนดทั่วไปที่จะถือปฏิบัติกัน ทั้งนี้เพราะการปฏิบัติจะต้องคำนึงถึงตัวแปรแวดล้อมอื่น ๆ อีก เช่น เวลา การลงทุน และข้อจำกัดในทางปฏิบัติอื่น ๆ อีก แต่สิ่งที่ควรพิจารณา คือ ให้ประสิทธิภาพของการเทียบมาตรานั้น ๆ อยู่ในระดับของการยอมรับเฉพาะกรณี (1985: 19)

จากการศึกษาของบูเดส (Budescu) ที่เกี่ยวกับความยาวของแบบสอบร่วมนี้ พบว่าตัวแปรที่มีความสัมพันธ์และร่วมส่งผลต่อประสิทธิภาพของการเทียบมาตราที่สำคัญ คือ คุณภาพของแบบสอบร่วม

ตอนที่ 4 การประเมินคุณภาพของการเทียบมาตรา

ในการวิจัยครั้งนี้ ผู้วิจัยประเมินคุณภาพของการเทียบมาตรา โดยใช้ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา (Standard error of Equating: SEE) และค่าดัชนีความแตกต่าง (C) ของความเพียงพอในการเทียบมาตรา ดังนี้

1. **ความคลาดเคลื่อนมาตรฐานของการเทียบมาตราเชิงเส้นตรง** (Standard Error of Linear Equating)

การเทียบมาตราจะมีความคลาดเคลื่อนแบบสุ่มซึ่งเกิดจากการแกว่ง (fluctuation) ของค่าเฉลี่ยและค่าเบี่ยงเบนมาตรฐานของคะแนนทั้งแบบสอบฉบับ X และฉบับ Y

ลอร์ด (Lord 1950 cited by Angoff 1984: 96-97) ได้อธิบายความคลาดเคลื่อนมาตรฐานในลักษณะของค่าเบี่ยงเบนมาตรฐานของคะแนนแปลงในสเกล Y ที่ตรงกับคะแนน X ที่กำหนด ซึ่งคะแนนที่แปลงแล้ว Y^* ได้มาจากกลุ่มตัวอย่างที่เป็นอิสระต่อกัน ความคลาดเคลื่อนมาตรฐานของการเทียบมาตราเชิงเส้นตรง มีสมการตามรูปแบบการรวบรวมข้อมูลของแองกอฟฟ์ (Angoff 1984: 94-121) ตามที่เสนอไว้ก่อนหน้านี้ คือ

รูปแบบที่ 1: $S.E^2y' = \frac{2S_Y^2}{N_t}(Z_X^2 + 2)$ (18)

โดยที่ $S.E^2y'$ คือความแปรปรวนของความคลาดเคลื่อนของคะแนนที่แปลงแล้ว

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

Z_x คือคะแนนมาตรฐานได้จาก $Z_x = \frac{X - M_X}{S_X}$

รูปแบบที่ 2: $S.E^2y' = S_Y^2(1 - r_{XY}) \frac{Z_X^2 (1 + r_{XY}) + 2}{N_t}$ (19)

รูปแบบที่ 2 นี้ ลอร์ดมีข้อตกลงเบื้องต้นไว้ว่า ค่าเบี่ยงเบนมาตรฐานของแต่ละฟอร์ม และความสัมพันธ์ระหว่างฟอร์มเหมือนกันในประชากร

รูปแบบที่ 3: $S.E^2y' = 2\hat{\sigma}_Y^2 (1 - \hat{r}^2) \frac{(1 + \hat{r}^2)Z_X^2 + 2}{N_t}$ (20)

โดยถือว่า $\hat{r} = \frac{b_{XU\alpha}\hat{\sigma}_U}{\hat{\sigma}_X} = \frac{b_{YU\beta}\hat{\sigma}_U}{\hat{\sigma}_Y}$

รูปแบบที่ 5: แบบสอบฟอร์ม X และฟอร์ม Y เทียบไปสู่แบบสอบร่วม

$$S.E^2y' = 4S_{Y\beta}^2(1 - r) \frac{Z_X^2 (1 + r) + 2}{N_t} \quad (21)$$

เมื่อ $r = r_{XU\alpha} = r_{YU\beta}$

จากการพิจารณาเปรียบเทียบความแปรปรวนของความคลาดเคลื่อนในสมการ (18) กับสมการ (19) ดังตัวอย่างเช่น แบบสอบ 2 ฟอร์มมีค่าสหสัมพันธ์ .80 ความแปรปรวนของความคลาดเคลื่อนของคะแนนแปลง Y ที่จุด $Z_x = 0$ ของรูปแบบที่ 2 มีค่า $\frac{1}{10}$ ของขนาดความคลาดเคลื่อนของรูปแบบที่ 1 อาจจะกล่าวได้ว่ารูปแบบที่ 1 จะได้รับความถูกต้องเหมือนรูปแบบที่ 2 จะต้องใช้ผู้สอบจำนวน 10 เท่าของจำนวนเดิม แสดงว่าการใช้รูปแบบที่ 2 มีข้อดีในด้านความประหยัดใช้ผู้สอบน้อย แต่มีความถูกต้องมากกว่า

หากพิจารณาจากสูตรในรูปแบบที่ 3 เมื่อ $\hat{r} = 0$ ความแปรปรวนของความคลาดเคลื่อนจะเหมือนกับการเทียบมาตราในรูปแบบที่ 1

2. การประเมินความเพียงพอของการเทียบมาตรา

การเทียบมาตรารูปแบบใดก็ตาม จะให้ผลดีที่สุดเมื่อคะแนนที่ได้จากแบบสอบที่นำมาเทียบมาตรามีคุณสมบัติเป็นไปตามเงื่อนไขต่าง ๆ ที่กำหนดไว้ในรูปแบบการเทียบมาตราแต่ละรูปแบบ แต่ในสภาพการณ์จริงอาจมีข้อจำกัดบางประการซึ่งไม่สามารถทำให้ได้ข้อมูลตรงตามเงื่อนไขได้ จึงมีความจำเป็นที่จะต้องทำการตรวจสอบความเพียงพอของการเทียบมาตรา ซึ่งเป็นการประเมินประสิทธิภาพของวิธีการเทียบมาตรา ซึ่งมีผู้เสนอแนวคิดไว้หลายวิธี

ภาวิณี ศรีสุขวัฒนานันท์ (2529) ได้ประยุกต์ดัชนีในการประเมินความเพียงพอโดยใช้ดัชนีเปรียบเทียบเปอร์เซ็นไตล์ (the percentile comparision index) ของโคเลนและวิทนีย์ (Kolen and Whitney 1982) ดัชนีความแตกต่าง (Discrepancy Indecies) ของปีเตอร์เซนและคณะ (Petersen and others 1982) ตามคำแนะนำของสเตราด์ (Stroud 1982: 138) มาประยุกต์ใช้

จากแนวคิดของโคเลนและวิทนีย์ที่ใช้ดัชนีเปรียบเทียบเปอร์เซ็นไตล์นั้น เป็นการใช้ข้อมูลคะแนนจากผู้สอบเป็นเกณฑ์ในการหาความแตกต่าง ข้อมูลเหล่านี้ได้มาจากการใช้กลุ่มสอบทานผล ซึ่งเป็นกลุ่มตัวอย่างที่สุ่มมาจากประชากรเดียวกันกับกลุ่มตัวอย่างเทียบมาตรา และไม่มีหน่วยตัวอย่างซ้ำกันเลย ให้ได้รับการทดสอบด้วยแบบสอบเทียบมาตราทั้งสองชุดคือ ฉบับ X และฉบับ Y ให้คะแนนในแบบสอบชุด X เป็นคะแนนเกณฑ์แล้วนำคะแนนของแบบสอบฉบับ Y ไปแปลงคะแนนให้อยู่ในมาตราคะแนนของ X (X^*) ด้วยวิธีการเทียบมาตราที่ระบุไว้ ณ ตำแหน่งเปอร์เซ็นไตล์เดียวกัน มีสูตรการคำนวนค่าดัชนี ดังนี้

$$C = \sum_i (X_i - X_i^*)^2 \,|\, nk \qquad (22)$$

เมื่อ n คือ จำนวนคะแนนดิบของกลุ่มสอบทานผล

k คือ จำนวนข้อสอบในแบบสอบร่วมที่ใช้

ถ้าค่า C ที่ได้มีค่าน้อย หมายความว่า รูปแบบการเทียบมาตราที่นำมาสร้างคะแนนสมมูลนั้นมีความเหมาะสมและเพียงพอที่ให้ผลการแปลงคะแนนคงเส้นคงวา

สำหรับแนวคิดของปีเตอร์เซนและคณะ ที่ใช้ดัชนีความแตกต่างนั้น คะแนนเกณฑ์ที่ใช้คือผลการแปลงคะแนนด้วยรูปแบบอิงทฤษฎีการตอบข้อสอบ มีสูตรการคำนวณค่าดัชนี คือ

$$\text{total error} = \sum_j f_j d_j^2 / n S_t^2 \qquad (23)$$

เมื่อ $d_j = (t - t')$

n = จำนวนคะแนนที่ใช้

S_t^2 = ความแปรปรวนของคะแนน t

ค่าดัชนีที่ได้มีลักษณะเป็นค่ามาตรฐาน ค่าเหล่านี้สามารถนำมาเปรียบเทียบกันได้โดยตรง ถึงแม้ในสถานการณ์ที่ได้ข้อมูลมาต่างกันก็ตาม

จะเห็นว่าจากแนวคิดของโคเลนและวิทนีย์ ที่ใช้คะแนนเกณฑ์ของผู้สอบเองเป็นเกณฑ์ ทำให้มีความเป็นอิสระ ไม่ขึ้นกับกระบวนการแปลงคะแนนในรูปแบบอื่น ๆ เช่นวิธีที่ปีเตอร์เซนและคณะเสนอ ผู้วิจัยจึงเลือกวิธีการประเมินความเพียงพอจากการวิเคราะห์กลุ่มสอบทานผล โดยใช้ค่าดัชนีความแตกต่างตามแนวคิดของภาวิณี ศรีสุขวัฒนานันท์ ซึ่งดัดแปลงจากสูตรของโคเลนไปใช้ตามแนวความคิดของปีเตอร์เซนและคณะ คือใช้ค่าความแปรปรวนเป็นตัวถ่วงน้ำหนัก เพื่อให้ได้ค่าที่เป็นมาตรฐาน คือ

$$C = \sum_i (X_i - X_i^*)^2 / n S_X^2 \qquad (24)$$

เมื่อ X_i คือ คะแนนจากแบบสอบฉบับ X ของคนที่ i

X_i^* คือ คะแนนจากแบบสอบฉบับ X ที่ได้จากการนำคะแนนจากแบบสอบฉบับ Y ไปแปลงจากตารางเทียบมาตราของคนที่ i

n คือ จำนวนคนของกลุ่มสอบทานผล

S_X^2 คือ ค่าความแปรปรวนของคะแนนจากแบบสอบฉบับ X

ตอนที่ 5 งานวิจัยที่เกี่ยวข้องในการเทียบมาตรา

ผู้วิจัยขอแบ่งงานวิจัยออกเป็น 2 ส่วน คือ งานวิจัยที่เกี่ยวข้องกับการเทียบมาตราในต่างประเทศ และงานวิจัยในประเทศ ดังมีรายละเอียดต่อไปนี้

งานวิจัยการเทียบมาตราต่างประเทศ

โคเลน (Kolen 1981: 1-11) ได้เปรียบเทียบวิธีการเทียบมาตราระหว่างรูปแบบดั้งเดิมสองวิธีคือ วิธีการเชิงเส้นตรงและแบบอิควิเปอร์เซ็นไตล์ ซึ่งจัดกระทำตามวิธีของแองกอฟฟ์ (Angoff 1971) กับรูปแบบทฤษฎีการตอบข้อสอบ (IRT) เจ็ดวิธี คือ ใช้ชนิดหนึ่ง, สอง และสามพารามิเตอร์ โลจิสติคโมเดล โดยแต่ละโมเดลเทียบมาตรา 2 แบบ คือ แบบเทียบด้วยค่าประมาณคะแนนจริง (Estimated true score equating) และเทียบด้วยค่าประมาณคะแนนสังเกต (Estimated observed score equating) และอีกวิธีหนึ่ง คือ ราสช์โมเดล

โดยใช้ข้อมูลจากผลการสอบนักเรียนในรัฐไอโอวา ปี 1978 ตามโครงการชื่อ The Iowa Tests of Education Development (ITED) แบบสอบที่ใช้เป็นแบบสอบวัดผลสัมฤทธิ์ เป็นแบบสอบที่พิมพ์ใหม่มี 2 ฉบับที่คู่ขนานกันคือ X7 และ Y7 แต่ละฉบับแบ่งออกเป็น 2 ระดับ ที่มีค่าความยากต่างกัน และแบบสอบชุดเก่า X6 เป็นการศึกษาการเทียบมาตราจากแบบสอบฉบับที่มีความยากเท่าเทียมกันและแตกต่างกัน

กลุ่มตัวอย่างเป็นนักเรียนในสองระดับ คือ ระดับ 1 นักเรียนเกรด 9 และ 10 ระดับ 2 เป็นนักเรียนเกรด 11 และ 12 แบบสอบฉบับ X6 ใช้ สอบกับนักเรียนทั้งสองระดับ จะได้ว่านักเรียนระดับ 1 สอบแบบสอบฉบับ X6 ฉบับ X7ระดับ 1 ฉบับ Y7ระดับ 1 ส่วนนักเรียนระดับ 2 สอบแบบสอบฉบับ X6 ฉบับ X7ระดับ 2 ฉบับ Y7ระดับ 2 ทั้งนี้นักเรียนแต่ละคนจะได้รับแบบสอบเพียง 1 ฉบับโดยการสุ่ม ดังนั้นจำนวนนักเรียนที่สอบแบบสอบแต่ละชุดจะมี 1 ใน 3 ของจำนวนนักเรียนในแต่ละระดับ แต่แบบสอบฉบับ X6 จะมีนักเรียนจำนวน 1 ใน 3 ของจำนวนนักเรียนรวมสองระดับ

การเปรียบเทียบผล การเทียบมาตราใช้ค่าดัชนีจากกลุ่มสอบทานผลซึ่งเป็นนักเรียนทุก ๆ คนที่ 3 จากการสอบแบบสอบแต่ละชุดในแต่ละระดับ ดัชนีที่ใช้ในการตรวจสอบผลคำนวณจากค่าเฉลี่ยกำลังสองของความแตกต่าง (สำหรับผู้สอบแบบสอบ X6 ในกลุ่มสอบทานผล) ระหว่างคะแนนจากแบบสอบ X6 กับคะแนนจากแบบสอบ X7 และ Y7 ที่แปลงแล้ว ณ ที่ตำแหน่งเปอร์เซ็นไตล์เดียวกัน ทั้งนี้โดยใช้วิธีการเทียบมาตราแบบต่าง ๆ 9 วิธี และทดสอบความแตกต่างของดัชนีที่ได้จากการเทียบมาตราแบบต่าง ๆ โดยใช้ค่าสถิติทดสอบฟรายด์แมน (Friedman test)

ผลการวิจัยพบว่า วิธีการเทียบมาตราแบบต่าง ๆ ให้ผลแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 เมื่อใช้แบบสอบระดับ 1 และ 2 ตามลำดับ ถ้าพิจารณาแบบสอบระดับที่ 1 ซึ่งมีความยากต่างกัน พบว่าวิธีที่ให้ผลคงเส้นคงวามากที่สุด คือ วิธี IRT ชนิดสามพารามิเตอร์ที่ประมาณค่าคะแนนสังเกต ส่วนวิธีการที่ให้ผลคงเส้นคงวาน้อยที่สุด คือ วิธีการเชิงเส้นตรง ถัดขึ้นมาคือวิธี IRT ชนิดหนึ่งพารามิเตอร์ สำหรับแบบสอบระดับที่ 2 พบว่าวิธีการที่ให้ผลคงเส้นคงวามากที่สุด คือ วิธี IRT ชนิดสามพารามิเตอร์ที่ประมาณค่าคะแนนจริง วิธีการของราสช์ให้ผลคงเส้นคงวามากกว่าวิธี IRT ชนิดหนึ่งพารามิเตอร์ที่ประมาณค่าคะแนนจริง

โคเลน และวิทนีย์ (Kolen and Whitney 1982: 279-293) ได้เปรียบเทียบความถูกต้องของวิธีการเทียบมาตรา 4 วิธี คือ วิธีอีควิเปอร์เซ็นไตล์ วิธีเชิงเส้นตรง วิธี IRT ชนิดหนึ่งพารามิเตอร์ และสามพารามิเตอร์ โดยใช้แบบสอบ GED (Tests of General Educational Development) ซึ่งเป็นแบบสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียน ซึ่งมี 12 ชุด เป็นแบบสอบร่วมหนึ่งชุด และอีก 11 ชุด เทียบเข้าสู่แบบสอบร่วม

กลุ่มตัวอย่างสุ่มจากการสอบในปี 1980 ซึ่งมีจำนวนมากกว่า 800,000 คน โดยสุ่มมาขนาดเล็ก ได้กลุ่มตัวอย่างฉบับละประมาณ 200 คน กลุ่มสอบทานผลก็ใช้วิธีสุ่มเช่นเดียวกัน โดยสุ่มมาจากผู้สอบแต่ละวิชา และแต่ละฉบับให้ได้ผู้สอบ 20 คน (ประมาณร้อยละ 10 ของผู้เข้าสอบทั้งหมด) การเทียบมาตราเชิงเส้นตรงและอีควิเปอร์เซ็นไตล์ กระทำตามวิธีของแองกอฟฟ์ คือ IA-1 และ IA-2 ส่วนวิธีการ IRT ประมาณค่าพารามิเตอร์โดยใช้โปรแกรมโลจีสเฉพาะแบบสอบร่วมก่อน จากนั้นใช้ค่าพารามิเตอร์ความสามารถของผู้สอบที่กำหนดให้เป็นค่าคงที่ในการประมาณค่าพารามิเตอร์จากแบบสอบอีก 11 ชุด แล้วสร้างตารางคะแนนสมมูลระหว่างแบบสอบทั้งหมดโดยใช้การประมาณค่าคะแนนจริง ส่วนคะแนนซึ่งอยู่ต่ำกว่าระดับการเดาตามรูปแบบสามพารามิเตอร์นั้นอาศัยวิธีการเชิงเส้นตรง (linear interpolation) มาประมาณค่าเพิ่มเติม

เปรียบเทียบผลของวิธีการเทียบมาตรา โดยใช้ดัชนีเปรียบเทียบเปอร์เซ็นไตล์ ซึ่งวิเคราะห์จากกลุ่มสอบทานผล แต่ละคนต้องสอบแบบสอบทั้งสองชุดที่นำมาเทียบมาตรา คำนวณค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างคะแนนที่ได้จากแบบสอบฉบับ X กับคะแนน Y ที่แปลงไปสู่สเกลของคะแนน X โดยใช้วิธีเทียบแต่ละวิธี และทดสอบความมีนัยสำคัญของค่าอันดับจากการแปลงคะแนนตามปริมาณของค่าดัชนีด้วยการทดสอบฟรายด์แมน ผลการวิจัยพบว่า รูปแบบเชิงเส้นตรงให้ผลการเทียบมาตราที่มีความเพียงพอมากกว่ารูปแบบอีควิเปอร์เซ็นไตล์และรูปแบบ IRT สามพารามิเตอร์ ขณะเดียวกันรูปแบบ IRT พารามิเตอร์เดียว ให้ผลการเทียบมาตราที่เพียงพอกว่ารูปแบบ IRT สามพารามิเตอร์

ผู้วิจัยได้สรุปว่าวิธีการเทียบมาตราที่มีความเพียงพอ ขึ้นกับองค์ประกอบหลายประการ ได้แก่คุณลักษณะต่าง ๆ ของแบบสอบ รูปแบบในการเทียบมาตรา ลักษณะของกลุ่มตัวอย่าง และขนาดของกลุ่มตัวอย่างที่ใช้ เป็นต้น ตามสถานการณ์ที่ศึกษานี้ ถ้าขนาดของกลุ่มตัวอย่างมีขนาดเล็ก เป็นการเทียบมาตราตามแนวระนาบ (Horizontal equating) กลุ่มตัวอย่างสุ่มสมมูลแล้ว รูปแบบการเทียบมาตราที่เหมาะสม ควรเป็นวิธีการเทียบมาตราเชิงเส้นตรง ซึ่งเป็นวิธีที่ง่าย และให้ผลการเทียบที่แน่นอนกว่า

ปีเตอร์เซน มาร์โค และสตีเวอร์ท (Petersen, Marco and stewart. 1982: 71-135) ได้ศึกษาความเพียงพอของการเทียบมาตราเชิงเส้นตรงหลายวิธี และวิธีอีควิเปอร์เซ็นไทล์ในสถานการณ์ต่าง ๆ เทียบมาตราจากแบบสอบชุด SAT (Scholastic Aptitude Test) ที่จัดสอบเดือนเมษายน 1975 และแบบสอบชุด TSWE (Test of Standard Written English) ที่จัดสอบเมื่อเดือนพฤศจิกายน 1975 ได้กลุ่มตัวอย่างในการสอบครั้งละ 3 กลุ่ม แบ่งแต่ละกลุ่มโดยไม่ให้ซ้ำกัน คือ กลุ่มสุ่มอย่างง่าย กลุ่มที่มีความสามารถคล้ายคลึงกัน และกลุ่มที่มีความสามารถไม่คล้ายคลึงกัน โดยใช้ค่าเฉลี่ยความสามารถทางภาษาเป็นเกณฑ์ ได้กลุ่มตัวอย่างกลุ่มละ 1,577 คน ในการศึกษาครั้งนี้คณะผู้วิจัยได้ออกแบบการเทียบในลักษณะต่าง ๆ คือ เทียบมาตราโดยใช้แบบสอบ SAT ส่วนที่เป็นภาษา มาแยกออกเป็น 2 ชุด แล้วนำมาเทียบมาตราโดยใช้แบบสอบร่วมลักษณะต่าง ๆ กัน คือ ต่างกันในตำแหน่งของแบบสอบร่วม (ชนิดภายในและภายนอก) เนื้อหา (คล้ายคลึงกัน และไม่คล้ายคลึงกัน) และความยาก (คล้ายคลึงกัน และไม่คล้ายคลึงกัน) โดยการเทียบแบบสอบชุด SAT ส่วนที่เป็นภาษากับตัวเองเหมือนกับว่าเป็นการเทียบแบบสอบสองชุดโดยใช้แบบสอบร่วมลักษณะต่าง ๆ เพื่อศึกษาระดับความคล้ายคลึงระหว่างแบบสอบที่ต้องการเทียบกับแบบสอบร่วม และระดับความคล้ายคลึงระหว่างกลุ่มตัวอย่างสองกลุ่มที่ใช้ในการเทียบที่มีผลต่อการเทียบมาตรา และเทียบมาตราระหว่างแบบสอบ (SAT กับ TSWE) โดยใช้แบบสอบร่วมภายใน เพื่อศึกษาระดับความคล้ายคลึงระหว่างแบบสอบที่ต้องการเทียบทั้งสองชุด และระดับของความคล้ายคลึงระหว่างกลุ่มตัวอย่างที่ใช้เทียบสองกลุ่มที่มีผลต่อการเทียบมาตรา โดยแบบสอบที่เทียบมีลักษณะแตกต่างกันด้านต่าง ๆ คือ เนื้อหา ความยาก และความยาวของแบบสอบ การวิเคราะห์ข้อสอบทำการเลือกคะแนนตามจุดมุ่งหมาย

เกณฑ์ที่ใช้ในการประเมินความเพียงพอของวิธีการเทียบมาตราใช้ ค่าดัชนีความแตกต่าง (Discrepancy Index) ซึ่งเป็นค่าที่ถ่วงน้ำหนักแล้วของค่าเฉลี่ยกำลังสองของความแตกต่างระหว่างค่าประมาณหรือคะแนนแปลงกับคะแนนเกณฑ์ คะแนนเกณฑ์ที่ใช้คือคะแนนแปลงจากคะแนนดิบชุดเดียวกันด้วยรูปแบบการเทียบมาตราแบบ IRT (I.C.C. equating) ดัชนีนี้เป็นค่ามาตรฐานโดยอยู่ในรูปสัดส่วนของความเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์

I17C79469

ผลการวิจัยสรุปได้ดังนี้

1. กรณีที่กลุ่มตัวอย่างขนาดเท่ากัน รูปแบบอีควิเปอร์เซ็นไตล์ให้ผลที่มีความคลาดเคลื่อนรวมมากกว่ารูปแบบเชิงเส้นตรง

2. รูปแบบการเทียบมาตราเชิงเส้นตรงโดยเทคนิคของทัคเกอร์ 1 (Tucker 1) ให้ผลการเทียบที่มีความคลาดเคลื่อนรวมมากกว่าวิธีเชิงเส้นตรงอื่น ๆ ในกรณีที่กลุ่มตัวอย่างแตกต่างกัน ยกเว้นสำหรับรูปแบบเชิงเส้นตรงโดยใช้เทคนิคของพอททอฟ (Potthoff) ให้ผลไม่น่าพอใจในเกือบทุกสถานการณ์ที่ศึกษา

3. รูปแบบอีควิเปอร์เซ็นไตล์ให้ผลการเทียบดีกว่ารูปแบบเชิงเส้นตรง เมื่อแบบสอบสองชุดมีความสัมพันธ์เชิงเส้นโค้งอันเนื่องมาจากแบบสอบมีความยากแตกต่างกัน และจากกลุ่มตัวอย่างที่ไม่คล้ายคลึงกัน

4. ค่าความยากระหว่างแบบสอบร่วม และแบบสอบที่เทียบมาตราแตกต่างกันจะทำให้ความคลาดเคลื่อนในการเทียบมีมากกว่าความแตกต่างในเนื้อหา และค่าความยากที่ต่างกันระหว่างแบบสอบที่จะเทียบมาตราทั้งสองชุด จะทำให้ความคลาดเคลื่อนในการเทียบมีมากกว่าความแตกต่างในเนื้อหา

5. แบบสอบร่วมที่สร้างให้มีลักษณะคล้ายคลึงกับแบบสอบ ที่จะเทียบมาตราอย่างมาก หรืออาจเรียกว่าเป็นฉบับย่อส่วนแล้ว การเทียบมาตราจะให้ผลดีที่สุด

ผู้วิจัยได้ให้ข้อสังเกตว่า การสรุปผลการวิจัยเพื่อนำไปใช้ทั่วไปต้องทำอย่างรอบคอบ โดยพิจารณาจากความเหมาะสมของรูปแบบ สถานการณ์ และลักษณะของกลุ่มตัวอย่าง

ปีเตอร์เซน คุ๊ก และสตอกคิง (Petersen, Cook and Stocking 1983) ได้ทำการศึกษาเปรียบเทียบความคงที่ของการเทียบมาตราด้วยรูปแบบ IRT 3 วิธี คือ วิธี Concurrent วิธี partial Precalibration with fixed bs และวิธี Partial Precalibration with equated bs กับรูปแบบดั้งเดิมสองวิธีคือรูปแบบอิควิเปอร์เซ็นไตล์ และรูปแบบเชิงเส้นตรง ที่นำมาศึกษามี 3 วิธี คือ Tucker Model, Levine Equally Reliable Model และ Levine Unequally Reliable Model

การศึกษาการเทียบมาตราครั้งนี้ ต้องการประเมินขนาดของการเบี่ยงเบนของมาตราในกระบวนการเทียบมาตราที่เป็นลูกโซ่ การเบี่ยงเบนของมาตรานี้ คือ ความแตกต่างของคะแนนแปลงในแบบสอบกับคะแนนเดิมก่อนแปลง กลุ่มตัวอย่างสุ่มจากการทดสอบ 9 ครั้ง ในรอบ 6 ปี แยกเป็นกลุ่มฉบับภาษาและกลุ่มฉบับคณิตศาสตร์แต่ละฉบับมีกลุ่มตัวอย่างย่อย 12 กลุ่ม กลุ่มละ 2,670 คน แบบสอบร่วมมีเนื้อหาแน่นอน การเปรียบเทียบความคลาดเคลื่อนใช้พิจารณาจากกราฟ และดัชนีความแตกต่าง (Discrepancy Index) ซึ่งเป็นค่าที่ถ่วงน้ำหนักแล้วของค่าเฉลี่ยความแตกต่างกำลังสอง (The weight mean square differrence) ความแตกต่างคำนวณจากผลต่างของคะแนนเกณฑ์กับคะแนนที่แปลงด้วยวิธีต่าง ๆ ตามที่ศึกษา ดัชนีนี้ใช้ในการประเมินประสิทธิภาพของรูปแบบการเทียบมาตราต่าง ๆ

ผลการวิจัยพบว่า ในการใช้รูปแบบเชิงเส้นตรงกับการเทียบมาตราสายแบบสอบภาษาและแบบสอบคณิตศาสตร์ของ SAT มีผลคล้ายคลึงกัน Tucker Model ให้ผลความคลาดเคลื่อนมากที่สุด ส่วน Levine equally reliable model ให้ค่าความคลาดเคลื่อนน้อยที่สุด รูปแบบอีควิเปอร์เซ็นไตล์ ยังให้ผลไม่ดีพอเมื่อเปรียบเทียบกับรูปแบบเชิงเส้นตรง เพราะรูปแบบอิควิเปอร์เซ็นไตล์ ไม่สามารถให้อิน ฟอร์เมชัน (Information) ได้ตลอดพิสัยของคะแนนแบบสอบ สำหรับรูปแบบ IRT ซึ่งทำการศึกษาทั้งสามวิธี พบว่าผลการเทียบมาตราที่ทำกับแบบสอบภาษา และแบบสอบคณิตศาสตร์มีความแตกต่างกัน วิธีที่ใช้ค่าความยากที่เทียบแล้ว (Partial Precalibration with equated bs) ให้ผลน่าพอใจที่สุดกับแบบสอบภาษา แต่ให้ผลตรงกันข้ามกับแบบสอบคณิตศาสตร์ สำหรับวิธี IRT ปัจจุบัน (Concurrent) ให้ผลด้อยกว่าวิธีที่ใช้ค่าความยากเทียบแล้วเพียงเล็กน้อยในแบบสอบภาษา แต่ในแบบสอบคณิตศาสตร์วิธี IRT ปัจจุบันให้ผลดีกว่ามาก

ในการเปรียบเทียบผลของการเทียบมาตรารูปแบบ IRT กับรูปแบบดั้งเดิม พบว่าความสัมพันธ์ที่เกิดขึ้นกับแบบสอบภาษาต่างกับแบบสอบคณิตศาสตร์ ในสายของภาษารูปแบบ IRT ให้ผลการเทียบดีกว่า แต่ในสายคณิตศาสตร์รูปแบบ Levine และ IRT ปัจจุบัน ให้ผลคล้ายคลึงกันมาก ผลการวิจัยสรุปได้ว่า ถ้าจำเป็นต้องเลือกรูปแบบการเทียบมาตราเพียงรูปแบบเดียวในการเทียบวิชาทางภาษาและคณิตศาสตร์แล้ว รูปแบบ IRT ปัจจุบัน เป็นวิธีที่เหมาะสมที่สุดใน 7 วิธีที่ได้ศึกษามา แต่ไม่ได้หมายความว่ารูปแบบ IRT ปัจจุบันจะให้ผลดีกว่าเสมอไป เมื่อข้อมูลที่ใช้ต่างไปจากแบบสอบ SAT

ฮิลล์ สับฮิยาท และเฮิร์ช (Hills, Subhiyah and Hirsch 1988: 221-231) ได้ศึกษาเปรียบเทียบวิธีการเทียบมาตรา 5 วิธีการ คือ วิธีเทียบมาตราเชิงเส้นตรง และวิธี IRT 4 วิธีการ ได้แก่วิธี IRT ชนิด 3 พารามิเตอร์ คือ Concurrent Method (TRTCON) Fixed-parameter Method (IRTFIX) และ Formula Method (IRTFOR) และวิธีการราสช์โมเดล และศึกษาขนาดความยาวของแบบสอบร่วม 6 ขนาด คือ ขนาด 30 25 20 15 10 และ 5 ข้อ ซึ่งสุ่มจากแบบสอบร่วม 30 ข้อ ว่าแบบสอบร่วมทั้ง 5 ขนาด มีประสิทธิภาพเทียบเท่ากับแบบสอบร่วม 30 ข้อ โดยการใช้การเทียบมาตราแบบ Concurrent IRT

โดยใช้ข้อมูลผลการสอบของนักเรียนในรัฐฟลอริดาปี 1986 และปี 1984 ตามโครงการสอบของ Florida's statewide Student Assessment Test Part II (SSAT-II) แบบสอบที่ใช้เป็นแบบสอบวัดความสามารถขั้นต่ำ(Minimum-Competency Test) ใช้กับโรงเรียนระดับมัธยมศึกษาตอนปลาย แบบสอบประกอบด้วยแบบสอบฉบับย่อยเป็นแบบสอบสายภาษา (Communications) และสายคณิตศาสตร์ (Mathematics) การศึกษาใช้แบบสอบปี 1984 เป็นเกณฑ์และใช้แบบสอบปี 1986 เทียบไปสู่แบบสอบปี 1984 แบบสอบในสายภาษาและคณิตศาสตร์ ในแต่ละปีแต่ละแบบสอบประกอบด้วยข้อสอบจำนวน 75 ข้อ เป็นข้อสอบร่วมในแต่ละปีจำนวน 30 ข้อ แบบสอบแต่ละฉบับย่อยของทุก ๆ ปี จะเหมือนกันในด้านเนื้อหาและเท่าเทียมกันในคุณสมบัติทางสถิติ กลุ่มตัวอย่างเป็นนักเรียนระดับ 9-11 ของโรงเรียนในรัฐฟลอริดา

จำนวน 6,000 คน สอบแบบสอบในแต่ละปี ๆ ละ 3,000 คน ซึ่งทำทั้งแบบสอบย่อยสายภาษาและคณิตศาสตร์

ผลการประเมินวิธีการเทียบมาตราทั้ง 5 วิธี ไม่มีการสรุปอย่างเป็นแบบแผน เพราะได้นำผลการเทียบมาตราเชิงเส้นตรงมาเป็นพื้นฐานในการเปรียบเทียบนั่นคือ เมื่อเทียบมาตราในแต่ละวิธีแล้ว นำมาหาความแตกต่างของคะแนนที่เทียบตามวิธีการต่าง ๆ กับคะแนนที่ได้จากการเทียบมาตราเชิงเส้นตรง แต่ผู้วิจัยได้อภิปรายผลไว้ว่า ในโปรแกรมการทดสอบทั่ว ๆ ไป วิธีการ IRT ไม่ใช่วิธีการเดียวที่จะใช้ในการพัฒนาข้อสอบและประเมินข้อกระทง แต่วิธีการเชิงเส้นตรงก็เป็นวิธีการที่ดีที่ควรนำมาใช้ เป็นวิธีการที่รู้จักกันอย่างกว้างขวาง สำหรับสถานการณ์ของการทดสอบความสามารถขั้นต่ำนี้ วิธีการแบบ IRT ก็สามารถใช้ได้กับการกระจายที่มีความเบ้มาก และวิธีการ IRTCON กับราสช์ ยังให้ผลที่ใกล้เคียงกัน ซึ่งวิธีการเทียบมาตราแบบราสช์จะง่ายกว่า

สำหรับจำนวนข้อกระทงของแบบสอบร่วมที่ใช้ในการเทียบมาตรา ข้อสอบร่วมตั้งแต่ 10 ข้อขึ้นไป ที่ถูกสุ่มจาก 30 ข้อ โดยการเทียบมาตราแบบ IRTCON มีประสิทธิภาพเพียงพอในการเทียบมาตราเท่ากับแบบสอบร่วมขนาด 30 ข้อ แต่แบบสอบร่วม 5 ข้อ ไม่มีประสิทธิภาพเท่า ผู้วิจัยได้อภิปรายผลว่า ผลที่ได้นี้ต่างจากงานวิจัยของ วิงเกอร์สกี้ และลอร์ด (Wingersky and Lord 1984) ราจู เอ็ดเวิร์ด และออสเบิกร์ก (Raju, Edwards and Osberg 1983) และราจู โบด ลาเซน และสเตนเฮิร์ท (Raju, Bode, Larsen and Steinhaus 1986) ซึ่งชี้ให้เห็นว่าแบบสอบร่วมจำนวน 5-6 ข้อ ก็เพียงพอในการเทียบมาตราด้วยวิธีการ IRT ชนิด 3 พารามิเตอร์ และผู้วิจัยได้ชี้ให้เห็นว่าเมื่อใช้รูปแบบ IRTCON เป็นวิธีการเทียบมาตรา ผู้สร้างแบบสอบสามารถลดจำนวนข้อสอบร่วมได้อย่างมาก การใช้ข้อสอบร่วมจำนวนน้อยถือเป็นการลดข้อกระทงที่ซ้ำจากแบบสอบฟอร์มหนึ่งไปยังแบบสอบอีกฟอร์มหนึ่ง เพื่อการเก็บรักษาข้อสอบอย่างเป็นความลับ การค้นพบนี้ไม่ได้ประยุกต์ใช้ แต่ก็สามารถนำไปใช้กับการเทียบมาตราวิธีการอื่น ๆ ได้

งานวิจัยการเทียบมาตราในประเทศ

เยาวดี รางชัยกุล (Yavadee Rangchaikul 1975: 87-94) เป็นบุคคลแรกของไทยที่เห็นความสำคัญของการเทียบมาตรา ได้ทำการเทียบมาตราคะแนนผลการสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 5 ซึ่งใช้แบบสอบต่างชุดกันว่ามีความเท่าเทียมกันเพียงใดที่จะนำมาใช้ตัดสินผลการสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 5 โดยใช้ข้อมูลผลการสอบซึ่งกระทรวงศึกษาธิการได้จัดดำเนินการสอบให้กับนักเรียนในระบบโรงเรียนเมื่อเดือนมีนาคมปี 2516 และปี 2517 และจัดให้กับบุคคลทั่วไปที่อยู่นอกระบบโรงเรียนเมื่อเดือนสิงหาคม พ.ศ. 2516 และปี 2517 เทียบมาตราคะแนนของแบบสอบทั้ง 4 ชุด กับมาตราคะแนนร่วม (Common Scale score) ที่มีช่วง

คะแนน 1-175 ค่าเฉลี่ย 100 ส่วนเบี่ยงเบนมาตรฐาน 25 กลุ่มตัวอย่างที่ใช้เป็นนักเรียนที่สอบแต่ละครั้งครั้งละ 500 คน เทียบมาตราในวิชาภาษาไทยภาษาอังกฤษ สังคมศึกษา คณิตศาสตร์และวิทยาศาสตร์ โดยใช้เทคนิคการเทียบมาตราเชิงเส้นตรงรูปแบบที่ 1 ซึ่งเสนอโดยแองกอฟฟ์ (Angoff 1971)

ผลการศึกษาพบว่าคะแนนที่เทียบกับมาตราคะแนนร่วมทั้ง 4 ฟอร์ม มีคะแนนไม่เท่าเทียมกันทุกวิชา กล่าวคือ นักเรียนคนหนึ่งสอบได้คะแนนรวมของวิชาภาษาไทยฟอร์มที่หนึ่ง (มีนาคม 2516) สูงกว่าคะแนนรวมของวิชาภาษาไทยฟอร์มที่สอง (มีนาคม 2517) และพบว่าคะแนนจุดตัดของแบบสอบแต่ละชุดไม่สามารถเปรียบเทียบกันได้ เช่น พบว่าคะแนนจุดตัดสำหรับการสอบผ่านเมื่อเดือนมีนาคม พ.ศ.2516 ในมาตราคะแนนร่วมเท่ากับ 88 แต่การสอบเมื่อเดือนมีนาคม 2517 มีคะแนนจุดตัดสำหรับการสอบผ่านในมาตราคะแนนร่วมเท่ากับ 78 ภายหลังการเทียบมาตรา ได้ตรวจสอบสัดส่วนของผู้สอบที่ได้รับการตัดสินได้และตกด้วยเกณฑ์ร้อยละ 50 พบว่าสัดส่วนนักเรียนที่ควรสอบได้ แต่ได้รับการตัดสินให้ตกมีจำนวนน่าสนใจ ในส่วนการศึกษาในรายวิชาทั้ง 5 วิชา พบว่าแต่ละวิชามีความแปรเปลี่ยนของคะแนนแตกต่างกัน วิชาภาษาไทยมีความแปรเปลี่ยนน้อยที่สุด วิชาคณิตศาสตร์มีความแปรเปลี่ยนของคะแนนมากที่สุด ผู้วิจัยได้เสนอแนะวิธีการ ซึ่งควรใช้ในการเทียบมาตราในสถานการณ์การสอบนี้ว่าควรใช้รูปแบบที่ 4 ตามแองกอฟฟ์เสนอไว้ ซึ่งเป็นรูปแบบการใช้แบบสอบร่วมกับกลุ่มตัวอย่างที่ไม่ได้สุ่มจะมีความเหมาะสม

ชูชีพ พงษ์สมบูรณ์ (2528: 112-118) ได้เปรียบเทียบประสิทธิภาพและความคงที่ของการเทียบมาตราระหว่างรูปแบบที่ใช้แบบสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วม ในการเทียบมาตรา 3 วิธี คือ การเทียบมาตราเชิงเส้นตรง การเทียบโดยใช้อิควิเปอร์เซ็นไทล์และการเทียบโดยใช้โค้งลักษณะข้อสอบ กลุ่มตัวอย่างที่ใช้เป็นผลการสอบของ นักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศที่เข้าสอบตามโครงการตรวจสอบคุณภาพการศึกษาของกรมวิชาการ กระทรวงศึกษาธิการ ซึ่งสุ่มมาจำนวน 3,721 คน แบ่งศึกษาเป็นกลุ่มย่อย 6 กลุ่ม ๆ ละ 620 คน แต่ละกลุ่มมีความสามารถใกล้เคียงกัน แบบสอบที่ใช้เป็นแบบสอบคณิตศาสตร์มีข้อสอบทั้งหมด 60 ข้อ แบ่งออกเป็น 2 ฉบับ ๆ ละ 38 ข้อ มีแบบสอบร่วมภายใน 20 ข้อ แบบสอบทั้งสองชุดมีความยากใกล้เคียงกัน

การเปรียบเทียบประสิทธิภาพของการเทียบมาตรา พิจารณาจากค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรา ส่วนการเปรียบเทียบความคงที่ของการเทียบมาตรานั้นพิจารณาจากความไม่แปรผันไปตามกลุ่มตัวอย่าง ความสมมาตร และความเสมอภาค โดยทำการทดสอบนัยสำคัญของความแตกต่างโดยใช้สถิติทดสอบโครโมโกรอฟ-สมิรนอฟ (Kolmogorov-Smirnov two sample test)

ผลการวิจัยพบว่าประสิทธิภาพของการเทียบมาตราระหว่างรูปแบบที่ใช้แบบสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วมในแต่ละวิธีของการเทียบมาตราไม่แตกต่างกัน ส่วนผลการเปรียบเทียบความคงที่ของการเทียบมาตรารูปแบบที่ใช้ผู้สอบร่วม พบว่าการเทียบมาตราโดยใช้อิควิเปอร์เซ็นไทล์มีความคงที่มากกว่าการเทียบเชิงเส้นตรง และมีความคงที่พอ ๆ กับการเทียบโดยใช้โค้งลักษณะข้อสอบ ส่วนการเปรียบเทียบความคงที่ของการเทียบมาตราโดยรูปแบบที่ใช้แบบสอบร่วมนี้ให้ผลเช่นเดียวกับรูปแบบที่ใช้ผู้สอบร่วม

ภาวิณี ศรีสุขวัฒนานันท์ (2529: 155-170) ได้เปรียบเทียบผลของการใช้รูปแบบการเทียบมาตราที่ต่างกัน 3 รูปแบบ คือ รูปแบบอีควิเปอร์เซ็นไทล์ รูปแบบเชิงเส้นตรง และรูปแบบอิงทฤษฎีการตอบข้อสอบแบบสามพารามิเตอร์ โดยใช้แบบสอบร่วมภายในที่ต่างกัน 3 ขนาด คือ ขนาดร้อยละ 20 (7 ข้อ), ขนาดร้อยละ 40 (14 ข้อ) และขนาดร้อยละ 60 (21 ข้อ) โดยมีข้อสอบทั้งหมด 100 ข้อ จัดแบ่งเป็น 2 ชุด ชุดละ 35 ข้อ ซึ่งทั้งสองชุดมีความยากง่ายใกล้เคียงกันที่เหลือจัดเป็นแบบสอบร่วมขนาดต่าง ๆ ดังกล่าว สำหรับกลุ่มตัวอย่างแยกเป็น 2 กรณี คือ กรณีแบบสอบคัดเลือก 2 กลุ่ม กลุ่มละ 1,500 คน และกลุ่มสอบทานผลอีก 1 กลุ่ม จำนวน 1,500 คน อีกกรณีหนึ่งคือกรณีแบบวัดผลสัมฤทธิ์ ซึ่งมีจำนวนกลุ่มตัวอย่างเช่นเดียวกับกรณีแรก

ในการประเมินผลการเทียบมาตราใช้เปรียบเทียบความคลาดเคลื่อนมาตรฐาน ของการเทียบมาตรา (SEE) ตามวิธีการเทียบมาตราที่ต่างกันของแบบสอบแต่ละกรณี ในแต่ละความยาวของแบบสอบร่วม และตรวจสอบความเพียงพอของวิธีการเทียบมาตราแต่ละวิธีในแต่ละความยาวของแบบสอบร่วม โดยใช้ค่าดัชนีเปรียบเทียบความแตกต่าง (index C) ที่ได้จากกลุ่มสอบทานผล

ผลการวิจัยพบว่าการใช้แบบสอบร่วมที่ยาวกว่าให้ค่าความคลาดเคลื่อนมาตรฐานที่น้อยกว่า และให้ผลในระดับที่น่าพอใจมากกว่าทั้งสองกรณีของแบบสอบ ส่วนการเปรียบเทียบรูปแบบของการเทียบมาตรา พบว่าในกรณีของแบบสอบคัดเลือกนั้น รูปแบบที่ให้ผลที่มีความเพียงพอมากที่สุด คือ รูปแบบอีควิเปอร์เซ็นไทล์ รองลงมาคือรูปแบบทฤษฎีการตอบข้อสอบ และรูปแบบเชิงเส้นตรงตามลำดับ ซึ่งให้ผลที่แตกต่างกับในกรณีของแบบสอบวัดผลสัมฤทธิ์ที่พบว่า รูปแบบที่ให้ผลที่มีความเพียงพอมากที่สุด คือ รูปแบบเชิงเส้นตรง รองลงมาคือรูปแบบอีควิเปอร์เซ็นไทล์ และรูปแบบทฤษฎีการตอบข้อสอบตามลำดับ และเมื่อทดสอบความแตกต่างของค่าดัชนี C เป็นรายคู่ พบว่ากรณีแบบสอบวัดผลสัมฤทธิ์วิธีการที่ใช้รูปแบบเชิงเส้นตรงมีความเพียงพอมากกว่า วิธีที่ใช้รูปแบบทฤษฎีการตอบข้อสอบ อย่างมีนัยสำคัญทางสถิติที่ .05 ส่วนการเปรียบเทียบรูปแบบคู่อื่นไม่สามารถสรุปความแตกต่างอย่างมีนัยสำคัญ

เรวดี อินทสะระ (2530: 64-69) ได้เปรียบเทียบความคลาดเคลื่อนและความเพียงพอในการเทียบมาตรา 2 รูปแบบ คือ การเทียบมาตรารูปแบบอิงทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2529 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดพัทลุง โดยแบบสอบที่ใช้เทียบมาตราเป็นแบบสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ เรื่องอัตราส่วน ร้อยละซึ่งผู้วิจัยสร้างขึ้นแล้วนำข้อสอบที่วิเคราะห์หาคุณภาพแล้วมาจัดแบ่งเป็นแบบสอบ 2 ชุด ชุดละ 45 ข้อ และแต่ละชุดมีแบบสอบร่วมภายใน 15 ข้อ นำไปดำเนินการสอบกลุ่มตัวอย่างอื่น จำนวน 1,557 คน โดยทุกคนต้องทำแบบสอบทั้ง 2 ชุด หลังจากนั้นจัดสุ่มใช้คะแนนผลการสอบจากแบบสอบชุดที่ 1 จำนวน 779 คน และอีก 778 คน ใช้คะแนนจากแบบสอบชุดที่ 2 ส่วนคะแนนที่เหลือใช้เป็นคะแนนตรวจทานผลของการเทียบมาตราแต่ละรูปแบบ

ผลการวิจัยพบว่าการเทียบมาตรารูปแบบอิงทฤษฎีการตอบข้อสอบให้ความเพียงพอ ในระดับที่น่าพอใจ ส่วนรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบให้ความเพียงพอในระดับปานกลาง สำหรับการเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตรารูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบมีความคลาดเคลื่อนสูงกว่ารูปแบบอิงทฤษฎีการตอบข้อสอบอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 (t-test) แต่เมื่อพิจารณาในแต่ละช่วงคะแนนพบว่ารูปแบบทั้งสองให้คะแนนสมมูลใกล้เคียงกันในช่วงตรงกลางของการแจกแจงคะแนน (คะแนน 19-23) ส่วนช่วงคะแนนต่ำและคะแนนสูง พบว่าวิธีทั้งสองให้ผลแตกต่างในทางตรงกันข้าม กล่าวคือในช่วงคะแนนต่ำ (คะแนน 1-18) นั้น รูปแบบอิงทฤษฎีการตอบข้อสอบมีคะแนนสมมูลสูงกว่ารูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ ส่วนช่วงคะแนนสูง (คะแนน 24-35) รูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบมีคะแนนสมมูลสูงกว่ารูปแบบการใช้ทฤษฎีการตอบข้อสอบ

อาภรณ์ กาญจนกิจโสภณ (2531: 66-67) ได้ศึกษาการสร้างแบบสอบและตารางเทียบมาตราคะแนนตามแนวตั้งในวิชาคณิตศาสตร์ เรื่องอสมการและสมการ สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้นโดยใช้วิธีการเทียบมาตรารูปแบบราสช์ กลุ่มตัวอย่างที่ใช้เทียบมาตราเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 จำนวนระดับชั้นละ 490, 482 และ 415 คน ตามลำดับ ผู้วิจัยได้นำแบบสอบที่สร้างขึ้นระดับชั้นละ 90 ข้อ ไปทดลองกับกลุ่มตัวอย่างอื่นแล้วคัดเลือกข้อสอบที่มีคุณภาพเรียงตามค่าความยากที่ต้องการระดับชั้นละ 40, 30 และ 30 ข้อ โดยแต่ละชั้นจะมีข้อสอบร่วมระหว่างชั้นอยู่ 10 ข้อ หลังจากนั้นนำไปทดสอบกับกลุ่มตัวอย่างที่ใช้เทียบมาตรา และสร้างตารางเทียบมาตราเพื่อใช้ศึกษาพัฒนาการของนักเรียนกลุ่มเดิมที่ผ่านขึ้นไปเรียนในชั้นสูงขึ้นว่า นักเรียนมีความก้าวหน้าเป็นไปตามปกติหรือเบี่ยงเบนไปจากปกติ

สุรัตน์ ขวัญบุญจันทร์ (2531: 77-82) ได้ศึกษาการสร้างแบบสอบและตารางเทียบมาตราคะแนนตามแนวนอนในวิชาคณิตศาสตร์ เรื่องพลังงานและการเปลี่ยนแปลงสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 โดยใช้วิธีการเทียบมาตรารูปแบบราสช์ ผู้วิจัยนำแบบสอบที่สร้างขึ้นไปทดลองกับกลุ่มตัวอย่างแล้วคัดเลือกข้อสอบที่มีคุณภาพจำนวน 110 ข้อ แบ่งเป็นสองฉบับ ๆ ละ 60 ข้อ ซึ่งมีแบบสอบร่วมภายในอยู่ฉบับละ 10 ข้อ นำไปสอบกับกลุ่มตัวอย่างเทียบมาตราจำนวน 819 คน สอบแบบสอบคนละฉบับตามการสุ่มได้ฉบับที่ 1 408 คน ฉบับที่สอง 411 คน และนำไปสอบกับกลุ่มทานผลจำนวน 166 คน ให้ทำแบบสอบทั้งสองฉบับ สร้างตารางเทียบมาตราและวิเคราะห์ค่าดัชนีความเพียงพอของการเทียบมาตราเท่ากับ .1019 ซึ่งอยู่ในระดับที่น่าพอใจมาก

สุจินดา ผ่องอักษร (2533: 143-157) ได้ศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะ (วิชาคณิตศาสตร์และภาษาไทย) ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบมาตรารูปแบบราสช์ ใช้กลุ่มตัวอย่างทั้งสิ้นจำนวน 1,454 คน จำนวน 4 อำเภอ จาก 7 อำเภอ ในจังหวัดปทุมธานี และใช้กลุ่มตัวอย่างเพื่อศึกษาค่าคงที่ในการเทียบมาตรา สุ่มอย่างง่ายจากนักเรียนในอำเภอธัญบุรี จำนวน 581 คน ผู้วิจัยได้ออกแบบเพื่อทำการเทียบมาตรา โดยหาค่าคงที่เพื่อใช้ในสมการเชิงเส้นตรง สำหรับปรับคะแนนที่ได้จากแบบสอบต่างชุดในแต่ละกลุ่มวิชาให้อยู่ในสเกลเดียวกัน โดยใช้สเกลคะแนนผลการสอบในปีการศึกษา 2529 เป็นหลัก และเทียบมาตราคะแนนในปี 2530 และ 2531 ไปสู่สเกลเดียวกันสเกลคะแนนในปี 2529 หลังจากนั้นจึงนำคะแนนที่สมมูลกันระหว่างแบบสอบต่างชุดมาเปรียบเทียบกันโดยตรง

ผลการศึกษาพบว่า หลังจากเทียบมาตราแล้วจุดตัดคะแนนในแต่ละปีการศึกษามีค่าแตกต่างกันในกลุ่มวิชาทักษะภาษาไทย ตั้งแต่ปีการศึกษา 2529 ถึง 2531 มีค่าเท่ากับ 15, 18 และ 14 ตามลำดับ กลุ่มทักษะวิชาคณิตศาสตร์มีค่าเท่ากับ 19, 24 และ 18 ตามลำดับ สำหรับผลการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียน พบว่าในกลุ่มทักษะวิชาภาษาไทยนักเรียนมีผลสัมฤทธิ์ทางการเรียนสูงกว่าระดับที่น่าพอใจ และมีผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิม ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์ พบว่ามีผลสัมฤทธิ์ทางการเรียนต่ำกว่าระดับที่น่าพอใจ และมีผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิมเฉพาะในปีการศึกษา 2530 แต่ในปีการศึกษา 2531 พบว่ามีผลสัมฤทธิ์ลดลงกว่าเดิม แต่ยังสูงกว่าผลสัมฤทธิ์ทางการเรียนเมื่อปีการศึกษา 2529 และพบความแตกต่างที่เพิ่มขึ้นและลดลงนี้ อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ พบว่ามีความแตกต่างกัน คือ ในปี 2530 พบว่ากลุ่มทักษะวิชาภาษาไทยมีนักเรียนร้อยละ 10 และในกลุ่มทักษะวิชาคณิตศาสตร์ มีนักเรียนร้อยละ 20 ที่ควรเป็นผู้สอบผ่าน แต่ได้รับการตัดสินให้เป็นผู้สอบไม่ผ่าน ส่วนในปี 2531 พบว่าในกลุ่มทักษะวิชาภาษาไทยมีนักเรียนร้อยละ 7 และกลุ่มทักษะวิชาคณิตศาสตร์พบว่ามีนักเรียนร้อยละ 5 ที่ควรเป็นผู้สอบไม่ผ่านแต่ได้รับการตัดสินให้ผ่าน