

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาครั้งนี้ ผู้วิจัยได้ศึกษาเพื่อใช้เป็นแนวทางในการดำเนินการวิจัยและพัฒนาการวิจัยนี้ให้มีคุณภาพและบรรลุผลสำเร็จด้วยดี จึงได้ทำการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังนี้ การศึกษาค้นคว้าเกี่ยวกับความหมายของการเทียบมาตรา ประเภทของการเทียบมาตรา รูปแบบการเทียบมาตราเชิงเส้นตรง การเทียบมาตรารูปแบบอิกวิเปอร์เซนไทล์ การเทียบมาตรารูปแบบทฤษฎีการตอบสนองข้อสอบ จำนวนข้อของแบบสอบ การประเมินคุณภาพของการเทียบมาตรา และงานวิจัยที่เกี่ยวข้องกับการเทียบมาตรา มีเนื้อหาสาระสำคัญดังต่อไปนี้

ความหมายของการเทียบมาตรา

กุลลิกสัน (Gulliksen, 1950 อ้างถึงใน ศิริชัย กาญจนวาสี, 2545 : 139) ได้ให้ความหมายของการเทียบมาตราไว้ว่า เป็นวิธีการปรับคะแนนที่ได้จากแบบสอบสองฉบับที่วัดเนื้อหาเดียวกัน ให้เป็นคะแนนสมมูล (equivalent scores) ที่เปรียบเทียบกันได้โดยตรง กุลลิกสันได้เสนอวิธีการอย่างง่าย ๆ โดยให้ผู้สอบกลุ่มเดียวกันทำแบบสอบทั้งสองฉบับที่พยายามสร้างให้คู่ขนานกัน แล้วคำนวณคะแนนมาตรฐานของแต่ละฉบับ คะแนนแปลงที่ตรงกับคะแนนมาตรฐานเดียวกัน ถือว่าเป็นคะแนนสมมูลกันสามารถนำมาเทียบกันได้โดยตรง

ปีเตอร์เซน โคลน และ ฮูเวอร์ (Petersen Kolen and Hoover, 1993 : 242) ได้ให้นิยามการเทียบมาตราว่า เป็นกระบวนการเชิงประจักษ์สำหรับสร้างระบบความสัมพันธ์ระหว่างคะแนนของแบบสอบสองฉบับให้สามารถแปลงคะแนนของแบบสอบฉบับหนึ่งไปสู่คะแนนของอีกฉบับหนึ่งได้

คร็อกเกอร์และอัลจินา (Crocker and Algina, 1986 : 457) กล่าวว่า การเทียบมาตราเป็นกระบวนการที่ทำให้คะแนนจากแบบสอบสองชุดมีความสมมูลกันสามารถนำมาเปรียบเทียบกันได้

สต็อกกิง (Stocking, 1999 : 61) กล่าวถึงการเทียบมาตราว่า เป็นวิธีการทางการวัดผลที่ได้รับ ความสนใจจากผู้ออกข้อสอบที่สร้างแบบสอบหลายฉบับซึ่งแต่ละฉบับมีความแตกต่างกัน แต่ต้องการรายงานผลด้วยสเกลเดียวกัน โดยการเทียบมาตราเป็นกระบวนการในการค้นหาคะแนนที่สอดคล้องกันจากแบบสอบต่างชุดกัน

ฟลานาแกน (Flanagan, 1951 อ้างถึงใน สุนิสา จุ้ยม่วงศรี, 2537 : 12) ได้ให้นิยามการเทียบมาตราว่า เป็นวิธีการในการทำคะแนนจากแบบสอบต่างชุดให้สามารถนำมาเปรียบเทียบกันได้ คำว่า

“ความสามารถในการเปรียบเทียบกันได้” มีความหมายเฉพาะที่ว่า เมื่อกำหนดประชากรให้ถ้าการแจกแจงของคะแนนจริงจากแบบสอบทั้งสองชุดซึ่งสอบกับกลุ่มตัวอย่างที่เลือกมาขนาดใหญ่ ๆ มีลักษณะเหมือนกันแล้วคะแนนดิบจากแบบสอบทั้งสองชุดจึงจะสามารถเปรียบเทียบกันได้ การแปลงระบบหน่วยการวัดของแบบสอบชุดหนึ่งไปสู่ระบบหน่วยการวัดของแบบสอบอีกชุดหนึ่ง ซึ่งหลังจากการแปลงแล้วก็จะสามารถเทียบหาคะแนนสมมูลระหว่างแบบสอบต่างชุดได้

แองกอฟฟ์ (Angoff, 1971 อ้างถึงใน รุ่งนภา จันทรา, 2539 : 8) ได้ให้ความหมายของการเทียบมาตราไว้ว่าเป็นวิธีการแปลงระบบหน่วยคะแนนของแบบสอบชุดหนึ่ง ไปสู่ระบบหน่วยคะแนนของแบบสอบอีกชุดหนึ่ง คะแนนที่ผ่านการแปลงแล้ว จะให้ความหมายของการสมมูลกันโดยตรง

ศิริชัย กาญจนวาสี (2545 : 140) กล่าวว่า การเทียบมาตรา หมายถึง การแปลงระบบคะแนนของแบบสอบต่างฉบับที่วัดเนื้อหาเดียวกัน โดยแปลงคะแนนของแบบสอบฉบับหนึ่งให้อยู่บนมาตรหรือสเกลของแบบสอบอีกฉบับหนึ่ง ที่ถือเป็นคะแนนที่สมมูลกัน (equivalent scores) เพื่อให้คะแนนจากแบบสอบต่างฉบับกันสามารถเปรียบเทียบกันได้โดยตรง

สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ (2542 : 16) ได้ให้ความหมายของการเทียบมาตราไว้ว่า เป็นกระบวนการทางสถิติที่ใช้ในการแปลงคะแนนที่อยู่ในระบบหน่วยการวัดของแบบสอบฉบับหนึ่งไปอยู่ในระบบของหน่วยการวัดของแบบสอบอีกฉบับหนึ่ง เพื่อทำการเปรียบเทียบคะแนนโดยแบบสอบทั้งสองฉบับวัดในคุณลักษณะหรือเนื้อหาเดียวกัน

สงบ ลักษณะ (อ้างถึงใน วรเนตร พิชิตกริกพล, 2537 : 13) ได้กล่าวว่า คะแนนจากแบบสอบสองฉบับที่วัดสิ่งเดียวกันแต่ไม่จำเป็นต้องเป็นข้อสอบที่คู่ขนาน จะถือว่าเทียบเท่ากันได้ถ้าคะแนนจากแบบสอบทั้งสองฉบับนั้นมาจากคะแนนจริง (True score) หรือความสามารถแท้ (True ability) ที่เท่ากัน

ชูศักดิ์ ชัมภลิจิต (2529 : 144) ได้สรุปเกี่ยวกับการเทียบมาตราว่า เป็นกระบวนการที่เกี่ยวข้องกับกิจกรรม 2 ประการ คือ

1. กระบวนการที่ทำให้แบบสอบสองฉบับใด ๆ มีความเท่าเทียมกันหรือเท่ากันในเชิงโครงสร้าง
2. การใช้วิธีการทางสถิติเพื่อปรับ (Adjust) คะแนนที่ได้จากแบบทดสอบแต่ละฉบับให้อยู่ในมาตราเดียวกันและเทียบกันได้

ประเภทของการเทียบมาตรา

ในการสอบที่ต้องการเทียบคะแนนจากแบบสอบฉบับหนึ่งไปสู่มาตรหรือสเกลคะแนนของฉบับอื่น ๆ สามารถจำแนกได้เป็น 2 สถานการณ์ (Hambleton & Swaminathan, 1987 : 197) คือ

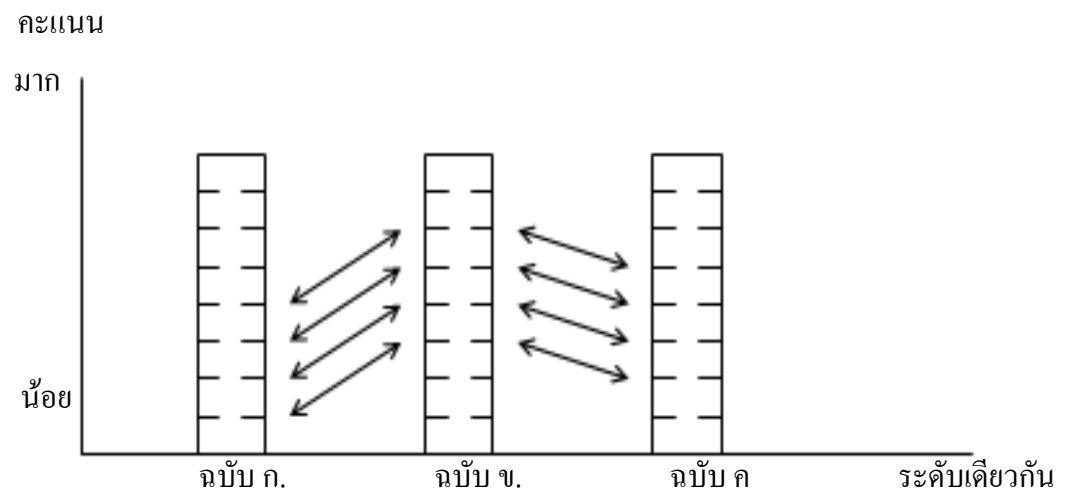
1. การเทียบมาตรฐานเมื่อแบบสอบมีระดับความยากใกล้เคียงกันและใช้สอบกับกลุ่มผู้สอบที่มีระดับความสามารถใกล้เคียงกัน ซึ่งเรียกว่า การเทียบมาตรฐานในแนวนอน (Horizontal Equating)

2. การเทียบมาตรฐานเมื่อแบบสอบมีระดับความยากแตกต่างกันและใช้สอบกับกลุ่มผู้สอบที่มีระดับความสามารถแตกต่างกัน ซึ่งเรียกว่า การเทียบมาตรฐานในแนวตั้ง (Vertical Equating)

ซึ่งศิริชัย กาญจนวาสิ (2545 : 142-143) ได้อธิบายรายละเอียดของการเทียบมาตรฐานทั้ง 2 ประเภทไว้ดังนี้

1. การเทียบมาตรฐานในแนวนอน (Horizontal Equating)

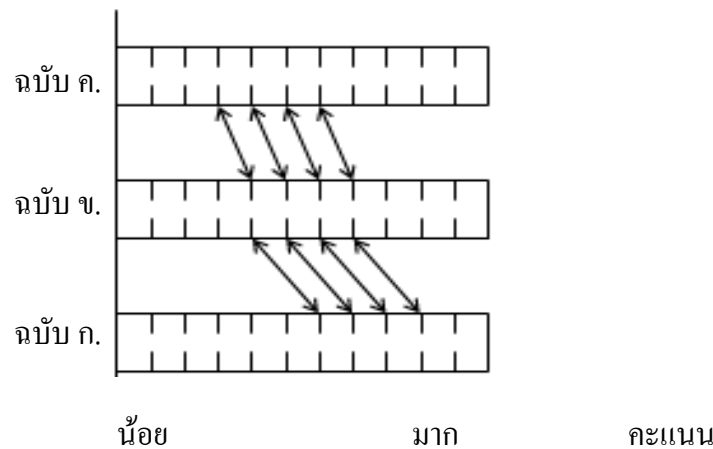
เป็นการเทียบมาตรฐานระหว่างแบบสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน มีระดับความยากใกล้เคียงกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ในประชากรเดียวกัน หรือมีความสามารถใกล้เคียงกัน การเทียบมาตรฐานในแนวนอนเป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบสอบเนื้อหาเดียวกันขึ้นมาหลาย ๆ ฉบับ ด้วยเหตุผลของการนำไปใช้ในการทดสอบให้เกิดความยุติธรรม และป้องกันความลับของข้อสอบเมื่อใช้ต่างเวลากันสำหรับผู้สอบกลุ่มขนาดใหญ่ เพื่อเปรียบเทียบว่าคะแนนที่ได้จากฉบับหนึ่งเป็นเท่าไรของอีกฉบับหนึ่ง ซึ่งวัดในระดับเดียวกัน จึงเป็นการเทียบมาตรฐานระหว่างแบบสอบต่างฉบับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบระดับชั้นเดียวกัน เช่น การเทียบมาตรฐานระหว่างแบบสอบวิชาคณิตศาสตร์ 3 ฉบับสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 เป็นต้น แบบสอบต่างฉบับที่ต้องการนำมาเทียบมาตรฐานกันตามแนวนอนนั้น ถึงแม้แบบสอบเหล่านั้นจะมีความเท่าเทียมกันในทุกด้าน แต่ก็ควรเป็นแบบสอบที่พยายามสร้างหรือมุ่งให้คู่ขนานกัน (Alternate forms) นอกจากนี้การแจกแจงความสามารถของผู้สอบที่ทำการสอบด้วยแบบสอบแต่ละฉบับควรมีการแจกแจงที่ประมาณได้ว่ามีระดับความสามารถที่ใกล้เคียงกัน



ภาพประกอบ 1 การเทียบมาตรฐานตามแนวนอน (Horizontal Equating)

2. การเทียบมาตรฐานแนวตั้ง (Vertical Equating)

เป็นการเทียบมาตรฐานระหว่างแบบสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน แต่มีระดับความยากแตกต่างกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ต่างประชากรกัน หรือมีความสามารถแตกต่างกัน การเทียบมาตรฐานแนวตั้งเป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็น ต้องสร้างแบบสอบเนื้อหาเดียวกัน แต่ต่างฉบับต่างมุ่งวัดความสามารถของผู้สอบที่ต่างระดับกันเพื่อ เปรียบเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่งเทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน จึงเป็นการ เทียบมาตรฐานระหว่างแบบสอบต่างระดับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบต่างระดับชั้นกัน เช่น การ เทียบมาตรฐานระหว่างแบบสอบวิชาคณิตศาสตร์ ชั้น ม.1 ม.2 และ ม.3 เป็นต้น แบบสอบต่างฉบับที่วัดเนื้อ เรื่องเดียวกันแต่ต่างระดับกันที่ต้องการนำมาเทียบมาตรฐานนั้น มีระดับความยากแตกต่างกันไป รวมทั้ง กลุ่มผู้สอบแต่ละฉบับมีการแจกแจงความสามารถที่อยู่ในระดับที่ต่างกัน ดังนั้นการเทียบมาตรฐานแนวตั้ง จึงมีความสลับซับซ้อนมากกว่าการเทียบมาตรฐานแนวนอนทั้งในแง่ทฤษฎีและการปฏิบัติ



ภาพประกอบ 2 การเทียบมาตรฐานแนวตั้ง (Vertical Equating)

รูปแบบการเทียบมาตรฐาน

รูปแบบการเทียบมาตรฐาน เมื่อพิจารณาจากรูปแบบการวัดแล้ว สามารถจำแนกได้ 2 รูปแบบคือ การเทียบมาตรฐานแบบดั้งเดิม และการเทียบมาตรฐานแบบอิงทฤษฎีการตอบสนองข้อสอบ ซึ่งทั้ง 2 รูปแบบมีหลักการในการเทียบมาตรฐานที่เหมือนกัน เพียงแต่มีวิธีการในการเทียบมาตรฐานต่างกัน ผู้วิจัยขอเสนอ รูปแบบการเทียบมาตรฐานที่มีผู้นิยมใช้ดังนี้

1. การเทียบมาตรฐานเชิงเส้นตรง (Linear Equating)

การเทียบมาตรฐานเชิงเส้นตรงมีนิยามว่า สำหรับกลุ่มผู้สอบที่กำหนดคะแนนจากแบบสอบสองฉบับจะถือว่าเท่าเทียมกันถ้าต่างก็ตรงกับคะแนนมาตรฐานเดียวกัน (Angoff, 1971 ; Petersen and Other, 1982 อ้างถึงใน สุนิสา จุ้ยม่วงศรี, 2537 : 13) การเทียบมาตรฐานเชิงเส้นตรงเป็นวิธีการเทียบมาตรฐานที่ง่ายและสะดวกที่สุดในการนำไปใช้ (Hills, Subhiyah and Hirsch, 1988 : 229)

แองกอฟฟ์ (Angoff, 1984 อ้างถึงใน สุนิสา จุ้ยม่วงศรี, 2537 : 13) ได้เสนอแบบแผนการรวบรวมและการจัดกระทำข้อมูลในทางสถิติของการเทียบมาตรฐานเชิงเส้นตรงไว้ 6 รูปแบบ ซึ่งแต่ละรูปแบบจะมีวิธีการประมาณค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของแบบสอบแตกต่างกันออกไป ตามเงื่อนไขของรูปแบบการรวบรวมข้อมูล แต่ทุกรูปแบบจะตัดสินคะแนนสมมูลจากค่าคะแนนมาตรฐานเดียวกันคือ

$$\frac{Y - M_y}{S_y} = \frac{X - M_x}{S_x} \quad (1)$$

เมื่อ	X	คือ	คะแนนจากแบบสอบฉบับ X
	Y	คือ	คะแนนจากแบบสอบฉบับ Y
	M_x	คือ	ค่าเฉลี่ยจากคะแนนแบบสอบฉบับ X
	M_y	คือ	ค่าเฉลี่ยจากคะแนนแบบสอบฉบับ Y
	S_x	คือ	ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนแบบสอบฉบับ X
	S_y	คือ	ค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนแบบสอบฉบับ Y

จากสมการ (1) สามารถเขียนใหม่ได้ดังนี้

$$\begin{aligned} (Y - M_y) \frac{S_x}{S_y} &= X - M_x \\ X &= M_x + (Y - M_y) \frac{S_x}{S_y} \\ &= M_x + Y \frac{S_x}{S_y} - M_y \frac{S_x}{S_y} \\ &= \frac{S_x}{S_y} Y + \left[M_x - M_y \frac{S_x}{S_y} \right] \end{aligned}$$

จะได้สมการเทียบมาตรฐานเชิงเส้นตรง

$$X = AY + B$$

$$\text{เมื่อ } A = \frac{S_x}{S_y}$$

$$B = M_x - AM_y$$

ในแต่ละรูปแบบการรวบรวมข้อมูลของแองกอฟฟ์ ได้เสนอรายละเอียดแยกแยะระหว่างกรณีที่เป็นแบบสอบเทียบมาตรฐานมีความเที่ยงเท่ากัน และเมื่อแบบสอบเทียบมาตรฐานมีความเที่ยงไม่เท่ากัน โดยแต่ละรูปแบบมีรายละเอียดดังนี้

รูปแบบที่ 1 กลุ่มผู้สอบสองกลุ่มที่มาจากประชากรเดียวกัน แต่ละกลุ่มทำแบบสอบเพียงหนึ่งชุด กลุ่มทั้งสองควรมีความคล้ายคลึงกันทางความรู้ ความสามารถ เพื่อให้คะแนนสอบที่ได้ไม่ได้เกิดจากผลของความแตกต่างของกลุ่ม ข้อดีของรูปแบบนี้คือ แต่ละกลุ่มทำแบบสอบฉบับเดียวจึงสามารถหลีกเลี่ยงปัญหาเรื่องการเรียนรู้ การฝึกฝน และความเมื่อยล้าจากการทำแบบสอบ แต่ปัญหาที่ตามมาคือ การใช้กลุ่มที่ไม่เหมือนกันอาจมีการแจกแจงความสามารถที่แตกต่างกัน และไม่มีข้อมูลที่นำมาใช้ปรับความแตกต่างของกลุ่ม ความแตกต่างที่เกิดขึ้นแม้เพียงเล็กน้อยย่อมส่งผลกระทบต่อความลำเอียงในการเทียบมาตรฐานได้ วิธีการที่จะช่วยลดความแตกต่างระหว่างกลุ่มทำได้โดยการใช้กลุ่มตัวอย่างสุ่มขนาดใหญ่

รูปแบบที่ 2 กลุ่มผู้สอบสองกลุ่มที่ถูกสุ่มมาจากประชากรเดียวกัน โดยการสุ่มผู้สอบแยกเป็นสองกลุ่มย่อย แต่ละกลุ่มย่อยได้รับการทดสอบจากแบบสอบทั้ง 2 ฉบับ ในลักษณะที่กลุ่มย่อยหนึ่งให้ทำแบบสอบฉบับที่ 1 แล้วตามด้วยแบบสอบฉบับที่ 2 ส่วนอีกกลุ่มย่อยหนึ่งให้ทำแบบสอบฉบับที่ 2 แล้วตามด้วยแบบสอบฉบับที่ 1 สลับกันเพื่อให้เกิดความสมดุล การใช้รูปแบบนี้กลุ่มผู้สอบจึงได้รับอิทธิพลจากลำดับการสอบ การเรียนรู้ การฝึกฝน และความเมื่อยล้าที่สมดุลกันในทั้งสองกลุ่มย่อย

ทั้งนี้ โคลเลน (Kolen, 1999 : 167) ได้แสดงทัศนะว่ารูปแบบผู้สอบกลุ่มเดียวต้องใช้เวลามากกว่ารูปแบบอื่น ๆ เนื่องจากผู้สอบแต่ละคนต้องทำแบบสอบถึง 2 ฉบับ ซึ่งในทางปฏิบัติมักจะเป็นไปไม่ได้ยาก

รูปแบบที่ 3 กลุ่มผู้สอบสองกลุ่มที่ถูกสุ่มมาจากประชากรเดียวกัน แต่ละกลุ่มทำแบบสอบเพียงฉบับเดียว และทำแบบสอบร่วมที่เหมือนกันอีกส่วนหนึ่ง กล่าวคือกลุ่ม α ทำแบบสอบฉบับ X และแบบ

สอบร่วม (U) ส่วนกลุ่ม β ทำแบบสอบฉบับ Y และแบบสอบร่วม (U) โดยแบบสอบร่วมอาจจะรวมอยู่ในแบบสอบทั้งสองชุดซึ่งเรียกว่า แบบสอบร่วมภายใน (Internal Anchor test) หรืออาจจะจัดเป็นฉบับที่แยกจากแบบสอบฉบับ X และฉบับ Y เรียกว่า แบบสอบร่วมภายนอก (External Anchor test) ซึ่งแบบสอบร่วมภายนอกกลุ่มผู้สอบจะต้องทำและจับเวลาแยกออกจากตัวแบบสอบที่ต้องการนำมาเทียบมาตราแบบสอบร่วมจะต้องถูกจัดให้สอบในลำดับที่เหมือนกันทั้งสองกลุ่ม โคลเลน (Kolen, 1999 : 168) กล่าวว่าแบบสอบร่วมควรจะคู่ขนานหรือคล้ายคลึงกับแบบสอบทั้งสองฉบับให้มากที่สุด และควรมีจำนวนข้อสอบร่วมไม่น้อยกว่า 20 % ของข้อสอบทั้งฉบับ รูปแบบที่ 3 จะมีข้อได้เปรียบรูปแบบที่หนึ่งเพราะแบบสอบร่วมจะทำหน้าที่กำหนดสเกลร่วมของคะแนนจากแบบสอบแต่ละชุดที่สอบโดยผู้สอบทั้งสองกลุ่มเพื่อนำไปปรับค่าความยากของข้อสอบหรือค่าความสามารถของผู้สอบจากแบบสอบต่างชุดให้อยู่ในสเกลร่วมกัน แบบสอบร่วมจะลดความแปรปรวนของความคลาดเคลื่อนในการเทียบมาตรา

รูปแบบที่ 4 กลุ่มที่ไม่ได้มาจากการสุ่มสองกลุ่มแต่ละกลุ่มทำแบบสอบเพียงฉบับเดียว และทำแบบสอบร่วมที่เหมือนกันอีกส่วนหนึ่ง กลุ่มผู้สอบที่ไม่ได้มาจากการสุ่มเป็นกลุ่มที่เกิดขึ้นในสถานการณ์จริง ซึ่งจะต้องสอบแบบสอบต่างชุดในเวลาต่างกัน ผู้สอบจึงไม่ได้ถูกสุ่มมาจากประชากรเดียวกัน เช่นการจัดสอบด้วยแบบสอบฉบับใหม่ในปี พ.ศ. 2546 แต่ต้องการเทียบคะแนนไปยังแบบสอบฉบับเก่าที่สอบในปี พ.ศ. 2545 แบบสอบร่วมจึงต้องมีความคล้ายคลึงกับแบบสอบเทียบมาตราทั้งสองฉบับให้มากที่สุดเพื่อลดอคติของการเทียบมาตราจากความแตกต่างในความสามารถระหว่างกลุ่ม

รูปแบบที่ 5 เป็นวิธีการเทียบมาตราโดยใช้แบบสอบร่วมที่เกี่ยวข้องกับคะแนน มีลักษณะการเทียบเป็นลักษณะเฉพาะ คือ

5.1 แบบสอบฉบับ X และฉบับ Y เทียบไปสู่แบบสอบร่วม เป็นการบริหารแบบสอบร่วมซึ่งอาจจะสอบแบบสอบร่วมก่อน แล้วตามด้วยแบบสอบฉบับ X และฉบับ Y หรือสอบแบบสอบฉบับ X และฉบับ Y ก่อนแล้วตามด้วยแบบสอบร่วม การเทียบมาตราโดยวิธีนี้คือการเทียบมาตราจากแบบสอบฉบับ X ไปยังแบบสอบร่วม และเทียบจากฉบับ Y ไปยังแบบสอบร่วมซึ่งคะแนนที่เทียบแล้วของทั้งสองแบบสอบที่ระดับคะแนนของแบบสอบร่วมเดียวกันถือว่าเท่าเทียมกัน โดยแบบสอบร่วม U จะต้องมียุทธศาสตร์คู่ขนานกับแบบสอบฉบับ X และฉบับ Y แต่ถ้าแบบสอบร่วม U ไม่มียุทธศาสตร์คู่ขนานกับแบบสอบฉบับ X และฉบับ Y แล้วกลุ่ม α และกลุ่ม β ต้องมาจากการสุ่มจากประชากรเดียวกัน

5.2 แบบสอบร่วมเป็นตัวทำนายแบบสอบฉบับ X และ ฉบับ Y ซึ่งจะมีสมการแปลงคะแนนแตกต่างจากข้อ 5.1

5.3 แบบสอบฉบับ X และ ฉบับ Y เป็นตัวทำนายแบบสอบร่วม ซึ่งจะมีสมการแปลงคะแนนแตกต่างจากข้อ 5.1 และ 5.2

รูปแบบที่ 6 เป็นวิธีการเทียบมาตราที่ขึ้นกับลักษณะข้อสอบ ได้แก่ วิธีการของเทอร์สโตนและแฟน (Thorstone, 1925 and Fan, 1957 อ้างถึงในสุนิสา จุ้ยม่วงศรี, 2537 : 16) โดยให้แบบสอบเทียบมาตราทั้งสองฉบับ มีข้อสอบชุดหนึ่งซึ่งมีลักษณะร่วมกันคือค่าความยาก

ในการศึกษาครั้งนี้ผู้วิจัยได้ใช้แบบแผนการรวบรวมข้อมูลรูปแบบที่ 3 ซึ่งเป็นกลุ่มสุ่มสองกลุ่ม แต่ละกลุ่มทำแบบสอบฉบับ X หรือ ฉบับ Y เพียงฉบับเดียวและทั้งสองกลุ่มทำแบบสอบรวม (U) โดยมีสถิติที่ใช้ในการจัดกระทำข้อมูลที่รวบรวมจากรูปแบบที่ 3 ดังต่อไปนี้

สถิติที่ใช้ในการเทียบมาตรารูปแบบที่ 3 สามารถแบ่งได้เป็น 2 กรณี คือ

1) กรณีแบบสอบมีความเที่ยงเท่ากัน

ลอร์ด (Lord, 1955 อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 39) ได้พัฒนาสมการสำหรับรูปแบบนี้ โดยใช้การประมาณค่าความเป็นไปได้สูงสุด (Maximum likelihood estimates) ประมาณค่าเฉลี่ย และค่าความแปรปรวนของแบบสอบฉบับ X และฉบับ Y โดยสมการเหล่านี้ คือ

$$\hat{\mu}_X = M_{x\alpha} + b_{xU\alpha} (\hat{\mu}_U - M_{U\alpha}) \quad (2)$$

$$\hat{\mu}_Y = M_{y\beta} + b_{yU\beta} (\hat{\mu}_U - M_{U\beta}) \quad (3)$$

$$\hat{\sigma}_X^2 = S_{x\alpha}^2 + b_{xU\alpha}^2 (\hat{\sigma}_U^2 - S_{U\alpha}^2) \quad (4)$$

$$\hat{\sigma}_Y^2 = S_{y\beta}^2 + b_{yU\beta}^2 (\hat{\sigma}_U^2 - S_{U\beta}^2) \quad (5)$$

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

$\hat{\mu}_X, \hat{\sigma}_X^2$ คือ ค่าประมาณคะแนนเฉลี่ย และความแปรปรวนของกลุ่มตัวอย่างที่ทำแบบสอบฉบับ X ที่ปรับแล้ว

$\hat{\mu}_Y, \hat{\sigma}_Y^2$ คือ ค่าประมาณคะแนนเฉลี่ย และความแปรปรวนของกลุ่มตัวอย่างที่ทำแบบสอบฉบับ Y ที่ปรับแล้ว

$M_{x\alpha}, S_{x\alpha}^2$ คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนฉบับ X

$M_{y\beta}, S_{y\beta}^2$ คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนฉบับ Y

$M_{U\alpha}, S_{U\alpha}^2$ คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนแบบสอบรวม U จากกลุ่ม α

$M_{U\beta}$, $S_{U\beta}^2$	คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนแบบสอบรวม U จากกลุ่ม β
M_{Ut} , S_{Ut}^2	คือ ค่าเฉลี่ยและความแปรปรวนของคะแนนแบบสอบรวม U จากกลุ่มรวม t
$b_{XU\alpha}$	คือ สัมประสิทธิ์ถดถอย X จาก U ในกลุ่มตัวอย่าง α
$b_{YU\beta}$	คือ สัมประสิทธิ์ถดถอย Y จาก U ในกลุ่มตัวอย่าง β
α	คือ กลุ่มตัวอย่างซึ่งทำแบบสอบฉบับ X และแบบสอบรวม U
β	คือ กลุ่มตัวอย่างซึ่งทำแบบสอบฉบับ Y และแบบสอบรวม U

ซึ่งการประมาณค่าเหล่านี้ใช้กับสมการ $X = AY + B$

$$\text{เมื่อ } A = \frac{\hat{\sigma}_X}{\hat{\sigma}_Y}$$

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

1) กรณีแบบสอบมีความเที่ยงไม่เท่ากัน

เลวิน (Levine, 1955 อ้างถึงใน วรเนตร พิชิตเกริกพล, 2537 : 26) ได้เสนอการเทียบมาตราเมื่อแบบสอบฉบับ X และ ฉบับ Y มีความเที่ยงไม่เท่ากัน ซึ่งเหมาะสมในการแปลงโดยคะแนนจริงมากกว่าคะแนนที่สอบได้ สถิติที่ใช้ในการจัดกระทำข้อมูลในรูปแบบนี้จะต่างกันเมื่อใช้แบบสอบรวมภายนอกและแบบสอบรวมภายในดังนี้

เมื่อเก็บรวบรวมข้อมูลโดยใช้แบบสอบรวมภายนอก (External Anchor test) ความชัน (Slope) และจุดตัดแกน (Intercept) ของสมการ $X = AY + B$ คือ

$$A = \frac{b_{XU\alpha}}{b_{YU\beta}} \quad ; \quad B = \hat{\mu}_X - A\hat{\mu}_Y$$

เมื่อ $b_{XU\alpha}$ และ $b_{YU\beta}$ เป็นสัมประสิทธิ์ถดถอย (Regression Coefficient) ในกลุ่ม α และ β สำหรับการทำนาย X จาก U และ Y จาก U ส่วน $\hat{\mu}_X$ และ $\hat{\mu}_Y$ คำนวณเช่นเดียวกับสมการ (2) และ (3)

เมื่อเก็บรวบรวมข้อมูลโดยใช้แบบสอบร่วมภายใน (Internal Anchor test) จะใช้ค่า

$$A = \frac{b_{y\beta} \hat{\sigma}_x^2}{b_{x\alpha} \hat{\sigma}_y^2} \quad ; \quad B = \hat{\mu}_x - A\hat{\mu}_y$$

ค่าของ $\hat{\mu}_x$, $\hat{\mu}_y$, $\hat{\sigma}_x^2$ และ $\hat{\sigma}_y^2$ คำนวณเช่นเดียวกับสมการ (2) ถึง (5)

โคเลน (Kolen, 1999 : 170) กล่าวว่า การเทียบมาตรฐานตามแนวคิดของเลวิน (Levine) มักจะถูกนำมาใช้มากเมื่อแบบสอบเทียบมาตรฐานมีความยาวแตกต่างกัน ในการศึกษาครั้งนี้ผู้วิจัยจึงจัดกระทำข้อมูลทางสถิติโดยใช้กรณีแบบสอบมีความเที่ยงไม่เท่ากัน

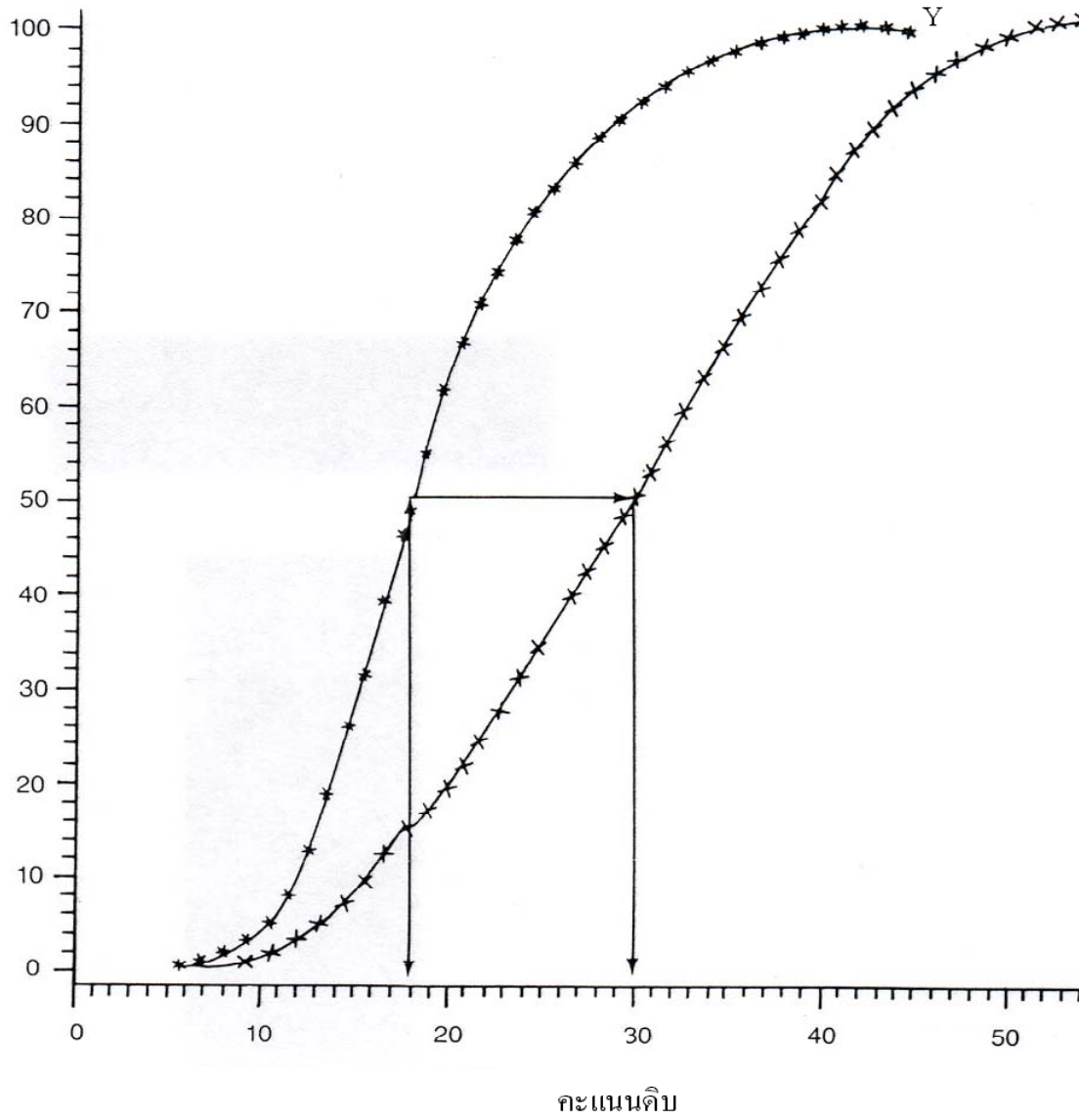
2. การเทียบมาตรฐานรูปแบบอิกวิเปอร์เซนไทล์ (Equipercentile Equating)

การเทียบมาตรฐานรูปแบบอิกวิเปอร์เซนไทล์มีนิยามว่า คะแนนจากแบบสอบสองฉบับจะถือว่าเป็นคะแนนที่สมมูลกัน เมื่อคะแนนของแต่ละฉบับมีตำแหน่งเปอร์เซนไทล์เดียวกัน (Kolen, 1999 : 169) การเทียบมาตรฐานรูปแบบอิกวิเปอร์เซนไทล์ประกอบด้วย 2 ขั้นตอน (Petersen, Kolen and Hoover, 1993 : 247-248) ดังนี้

ขั้นตอนที่หนึ่ง การแจกแจงความถี่สะสมสัมพัทธ์สำหรับแบบสอบที่จะนำมาเทียบมาตรฐาน โดยมีขั้นตอนย่อย ๆ คือ

- 1) นำคะแนนของกลุ่มผู้สอบที่มีความสามารถทั้งเก่ง ปานกลาง และอ่อน ซึ่งสุ่มเป็นสองกลุ่มย่อย ให้กลุ่มหนึ่งทำแบบสอบ X และอีกกลุ่มทำแบบสอบ Y มาทำการแจกแจงคะแนน X และ Y
- 2) คำนวณหาจุดกลางเปอร์เซนไทล์ของแต่ละการแจกแจง
- 3) อ่านและทำเครื่องหมายสำหรับค่าคะแนนของแบบสอบฉบับ X และฉบับ Y ของการแจกแจงที่สมมูลกันบนกระดาษกราฟ โดยแกนนอนเป็นคะแนนดิบ และแกนตั้งเป็นตำแหน่งเปอร์เซนไทล์ประมาณ 30 จุด แล้วลากเส้นเชื่อมต่อกันเป็นกราฟ

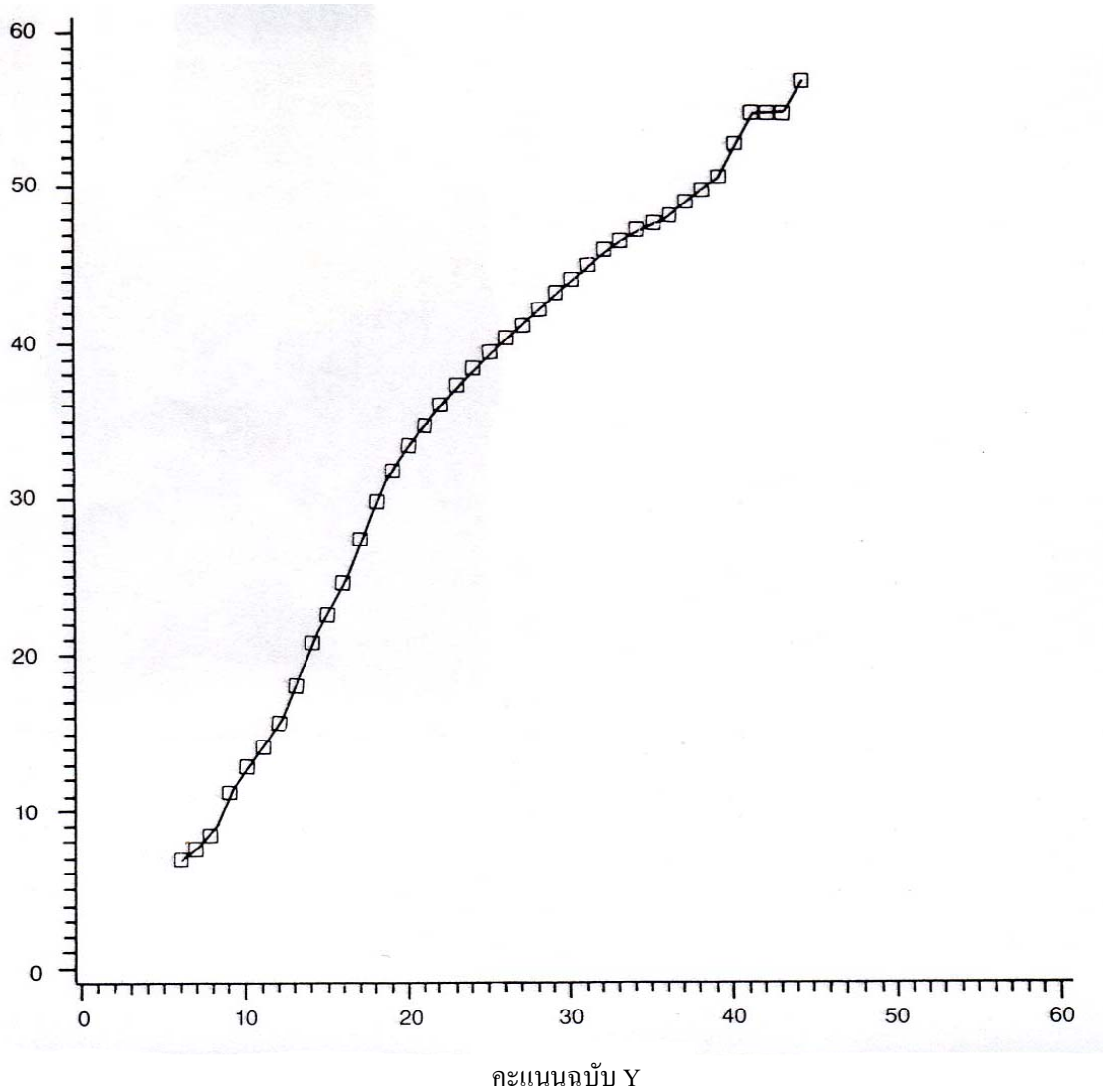
เปอร์เซ็นต์ไทล์



ภาพประกอบ 3 กระบวนการเทียบมาตรฐานรูปแบบอิกวิเปอร์เซ็นต์ไทล์

ขั้นตอนที่สอง เทียบมาตราคะแนนโดยนำคะแนนจากกราฟในขั้นตอนที่หนึ่งมาเขียนกราฟใหม่ลงบนกระดาษกราฟ โดยแกนอนเป็นคะแนนฉบับ Y แกนตั้งเป็นคะแนนฉบับ X ทำการปรับเส้นกราฟให้เรียบ เส้นกราฟที่ได้จะใช้อ่านค่า X ที่สมมูลกับ Y จากนั้นสร้างตารางสำเร็จเพื่ออ่านค่าคะแนนแปลงจากกราฟ ดังภาพ

คะแนนฉบับ X



ภาพประกอบ 4 การแปลงคะแนนของการเทียบมาตรฐานรูปแบบอิกวิเปอร์เซนไทล์

การเทียบมาตรฐานรูปแบบอิกวิเปอร์เซนไทล์โดยใช้แบบสอบร่วม

ลอร์ดและเลวิน (Lord and Levine อ้างถึงใน วรเนตร พิชิตเกริกพล, 2537 : 17) ได้เสนอขั้นตอนการเทียบมาตรฐานวิธีอิกวิเปอร์เซนไทล์โดยใช้แบบสอบร่วมสำหรับกลุ่มที่ไม่ได้สุ่มซึ่งมีความสามารถไม่แตกต่างกันมากนัก โดยที่กลุ่ม α สอบแบบสอบฉบับ X และฉบับ U ส่วนกลุ่ม β สอบแบบสอบฉบับ Y และฉบับ U ขั้นตอนในการเทียบมาตรฐาน คือ ประเมินค่าความถี่ในแบบสอบฉบับ X และ Y สำหรับกลุ่ม $t(\alpha + \beta)$ มีขั้นตอนดังนี้

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

ในการเทียบมาตรฐานรูปแบบอิกวิเปอร์เซนไทล์มีข้อจำกัดที่ควรคำนึงอยู่หลายประการดังนี้ (Angoff, 1971 ; Potthoff, 1982 อ้างถึงใน พรพิมล นาคเวช, 2537 : 14-15)

- 1) การสร้างกราฟเปอร์เซนไทล์เพื่อเทียบมาตรฐานระหว่างแบบสอบ ควรสร้างด้วยความระมัดระวังและไม่ลำเอียง
- 2) รูปแบบอิกวิเปอร์เซนไทล์มีความไวต่อความแปรปรวนเชิงสุ่มมาก โดยเฉพาะกลุ่มตัวอย่างที่มีขนาดเล็ก จึงควรใช้กลุ่มผู้สอบที่มีขนาดใหญ่และมีความสามารถค่อนข้างกระจาย
- 3) แบบสอบที่นำมาเทียบมาตรฐานควรมีความเที่ยงสูงพอกัน ถ้าความเที่ยงต่างกันมากผลการเทียบมาตรฐานจะขาดความคงที่
- 4) การเทียบมาตรฐานระหว่างแบบสอบสองชุดทำได้เฉพาะในช่วงพิสัยของคะแนนที่มีความถี่ของคะแนนสังเกตเพียงพอ ส่วนช่วงที่มีความถี่ของคะแนนน้อยจะมีความคลาดเคลื่อนสูงมาก

3. การเทียบมาตรฐานแบบทฤษฎีการตอบสนองข้อสอบ

ลอร์ด (Lord, 1980 อ้างถึงใน สุนิสา จุ้ยม่วงศรี, 2537 : 22) ได้กล่าวถึงการเทียบมาตรฐานโดยใช้ทฤษฎีการตอบสนองข้อสอบไว้ว่า แบบสอบสองฉบับใด ๆ ที่ใช้ในการเทียบมาตรฐานจะต้องเป็นแบบสอบ

ที่วัดคุณลักษณะเดียวกันและเทียบมาตราคะแนนที่สมมูลกันที่ระดับความสามารถเดียวกัน ซึ่งมีข้อกำหนด 3 ประการดังนี้

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

2. ความไม่ผันแปรตามกลุ่ม (Invariance Across Groups) หมายถึง คะแนนที่แปลงแล้วจะคงที่โดยไม่แปรเปลี่ยนไปตามกลุ่มประชากรที่นำมาเทียบ จึงสามารถนำผลไปใช้กับกลุ่มอื่น ๆ ได้

3. ความสมมาตร (Symmetry) หมายถึง คะแนนแปลงที่ได้จากการเทียบมาตราจะต้องเหมือนกัน ไม่ว่าจะเป็นการเทียบมาตราจากแบบสอบฉบับ X ไปยังแบบสอบฉบับ Y หรือจากแบบสอบฉบับ Y ไปยังแบบสอบฉบับ X

การเทียบมาตรารูปแบบทฤษฎีการตอบสนองข้อสอบสามารถแบ่งได้เป็น 2 วิธีการใหญ่ ๆ ดังนี้ (Lord, 1980 อ้างถึงใน สุนิสา จุ้ยม่วงศรี, 2537 : 23-24)

1) การเทียบมาตราโดยใช้คะแนนจริง (True- Score Equating)

แบบสอบที่มีความยากต่างกัน ถึงแม้ว่าจะนำไปสอบกับประชากรที่มีความสามารถเท่ากันหรือคุณลักษณะเดียวกัน แต่คะแนนที่ได้จะมีการแจกแจงแตกต่างกัน ดังนั้นถ้าสามารถหาคะแนนจริงได้ ก็จะสามารเทียบมาตราคะแนนระหว่างความสามารถ (θ) กับคะแนนจริงได้จากความสัมพันธ์เชิงคณิตศาสตร์ดังนี้

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

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

เมื่อ	ξ	คือ	คะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับ X
	η	คือ	คะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับ Y
	m	คือ	จำนวนข้อของแบบสอบฉบับ X
	n	คือ	จำนวนข้อของแบบสอบฉบับ Y
	θ_x	คือ	ค่าความสามารถของผู้สอบที่ได้จากแบบสอบฉบับ X
	θ_y	คือ	ค่าความสามารถของผู้สอบที่ได้จากแบบสอบฉบับ Y

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

การหาค่า $P_i(\theta)$ และ $P_j(\theta)$ ได้จากการประมาณค่าพารามิเตอร์ของข้อสอบโดยเลือกว่าจะใช้ชนิดหนึ่งพารามิเตอร์, สองพารามิเตอร์ หรือ สามพารามิเตอร์ จากนั้นนำคะแนนจริงของแบบสอบฉบับ X และฉบับ Y มาหาความสัมพันธ์กันโดยใช้ค่าความสามารถ (θ) ที่ระดับเดียวกัน

2) การเทียบมาตราโดยใช้คะแนนสังเกต (Observed- Score Equating)

ปัญหาของการเทียบมาตราโดยใช้คะแนนจริงคือไม่สามารถทราบคะแนนจริงของแต่ละคนได้ นอกจากใช้วิธีประมาณจากผลการสอบ ซึ่งคะแนนที่ได้เป็นเพียงค่าประมาณเท่านั้น ยังไม่มีคุณสมบัติเป็นคะแนนจริง และไม่สามารถทำการเทียบมาตราคะแนนที่ต่ำกว่าคะแนนการเดาได้ (ศิริชัย กาญจนวาสี, 2541 : 30) ลอร์ด (Lord, 1980 อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 43) จึงเสนอการเทียบมาตราโดยใช้คะแนนสังเกต โดยอาศัยข้อมูลจากการตอบแบบสอบร่วม (Anchor test)

การแจกแจงทางทฤษฎีของคะแนนที่สังเกตได้ของผู้สอบที่มีความสามารถ θ จากการทำแบบสอบฉบับที่หนึ่ง สามารถแสดงได้ดังนี้ (Hambleton & Swaminathan, 1985 : 215)

$$\sum_{r=0}^n f(r|\theta) t^r = \prod_{i=1}^n [Q_i(\theta) + W_i(\theta)]$$

ขั้นตอนในการเทียบมาตราคะแนนที่สังเกตได้โดยใช้ทฤษฎีการตอบสนองข้อสอบ สามารถสรุปเป็นขั้นตอนได้ดังนี้ (ศิริชัย กาญจนวาสี, 2545 : 164)

1) จากกลุ่มผู้สอบแบบสอบฉบับ X นำข้อมูลมาวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบ แล้วนำค่า θ_x ของผู้สอบแต่ละคนมาคำนวณความถี่สัมพัทธ์อย่างมีเงื่อนไขทางทฤษฎีของการแจกแจงคะแนนดิบ $f_x(r|\theta_x)$

2) แปลงค่า θ_x ให้สมมูลกับ θ_y ด้วยสูตร $T_{(\theta_x)} = \alpha\theta_x + \beta$ แล้วนำ $T_{(\theta_x)}$ ของกลุ่มผู้สอบมาคำนวณความถี่ของคะแนนดิบทางทฤษฎี $f_x(r)$ จะได้ตารางการแจกแจงความถี่ของคะแนนดิบทางทฤษฎีของกลุ่มผู้สอบ ฉบับ X

3) จากกลุ่มผู้สอบแบบสอบฉบับ Y นำข้อมูลมาวิเคราะห์เพื่อประมาณค่าพารามิเตอร์ของผู้สอบและผู้สอบ แล้วนำค่า θ_y ของผู้สอบแต่ละคนมาคำนวณความถี่สัมพัทธ์อย่างมีเงื่อนไขทางทฤษฎีของการแจกแจงคะแนนดิบ $f_y(r|\theta_y)$

4) นำค่า θ_y ของกลุ่มผู้สอบมาคำนวณความถี่ของคะแนนดิบทางทฤษฎี $f_y(r)$ จะได้ตารางการแจกแจงความถี่ของคะแนนดิบทางทฤษฎีของกลุ่มผู้สอบ ฉบับ Y

5) เทียบมาตราระหว่างคะแนนดิบของแบบสอบฉบับ X และ Y (จากข้อ 2 และ 4) โดยใช้วิธีอิกวิเปอร์เซนไทล์

จำนวนข้อของแบบสอบ

เสริมศักดิ์ วิชาลาภรณ์ และเอนกกุล กรีแสง (2519 : 191) กล่าวว่า แบบสอบควรจะต้องสั้นพอที่จะมีประสิทธิภาพในการปฏิบัติ และยาวพอที่จะมีความเที่ยงในการวัด จำนวนข้อสอบในแบบสอบมีอิทธิพลต่อค่าความตรง อำนาจจำแนกและค่าความเที่ยงของแบบสอบ การเพิ่มจำนวนข้อสอบจะทำให้แบบสอบมีค่าความตรง อำนาจจำแนกและค่าความเที่ยงเพิ่มขึ้น ดังที่ สุพัฒน์ สุขมลสันต์ (2538 : 38) ได้กล่าวไว้ว่าจำนวนข้อของแบบสอบมีผลต่อความผิดพลาดของค่าอำนาจจำแนก ข้อสอบยังมีจำนวนน้อยความผิดพลาดของค่าอำนาจจำแนกยังมีมากขึ้น ส่วนพร้อมพรรณ อุดมสิน (2538 : 119-120) กล่าวถึงจำนวนข้อของแบบสอบที่มีอิทธิพลต่อค่าความตรงไว้ว่า ค่าความตรงจะเปลี่ยนไปตามจำนวนข้อของแบบสอบ คือ ถ้าแบบสอบมีจำนวนข้อมากขึ้นจะทำให้ค่าความแปรปรวนของคะแนนจริง (True Variance) สูงขึ้น และค่าความแปรปรวนของคะแนนผิดพลาดในการวัด (Error Variance) ต่ำลง ปรากฏการณ์เช่นนี้จะทำให้ค่าความแปรปรวนรวมกัน (Common Factor Variance) เพิ่มขึ้น ซึ่งเป็นการเพิ่มความตรงของแบบสอบนั่นเอง ส่วนอิทธิพลของจำนวนข้อสอบที่มีผลต่อค่าความเที่ยงของแบบสอบมีผู้กล่าวไว้หลายท่าน เช่น เขา วดี วิบูลย์ศรี (2545 : 119) พิชิต ฤทธิ์จรูญ (2545 : 160) ศิริชัย กาญจนวาสิ (2544 : 62-63) ซึ่งกล่าวไว้สอดคล้องกันว่าการเพิ่มจำนวนข้อสอบจะทำให้แบบสอบมีค่าความเที่ยงสูงขึ้น เนื่องจากข้อคำถามที่มีมากพอจะทำให้วัดได้ครอบคลุมในเรื่องที่จะศึกษา และช่วยลดการเดาของผู้ตอบรวมทั้งช่วยลดความคลาดเคลื่อนจากแหล่งอื่น ๆ ให้น้อยลง

กรอนลันด์ (Gronlund, 1976 : 117) ได้อธิบายไว้ว่า การเพิ่มจำนวนข้อสอบจะทำให้ค่าความเที่ยงสูงขึ้น ทำให้เห็นได้ง่ายโดยการแบ่งครึ่งข้อสอบฉบับหนึ่งออกเป็น 2 ชุด ตามวิธีแบ่งครึ่งข้อสอบ (Split-half) หากค่าสหสัมพันธ์ระหว่างคะแนน 2 ชุดนี้ เช่น ได้ค่าสหสัมพันธ์ .60 แล้วใช้สูตรปรับขยายของ สเปียร์แมน-บราวน์ หากค่าความเที่ยงจากฉบับเต็ม ได้ค่า .75 ซึ่งก็คือค่าความเที่ยงที่เพิ่มจำนวนข้อสอบเป็นสองเท่าของฉบับเดิมนั่นเอง

ส่วนอดัมส์ (Adams, 1964 : 97 อ้างถึงใน สุรศักดิ์ เมตตาวิมล, 2542 : 39) ได้กล่าวว่าการเพิ่มจำนวนข้อสอบเพื่อให้แบบสอบมีค่าความเที่ยงสูงขึ้นจะต้องอยู่ในกรณีดังต่อไปนี้

1) จำนวนข้อสอบที่เพิ่มขึ้นจะต้องเป็นข้อสอบที่มีคุณภาพเหมือนข้อสอบเดิม

2) การเพิ่มข้อสอบจะต้องให้เหมาะสมกับเวลาที่ใช้สอบ

การประเมินคุณภาพของการเทียบมาตรา

การเทียบมาตราไม่ว่าจะใช้รูปแบบใดก็ตาม จะมีคุณภาพดีที่สุดเมื่อคะแนนที่ได้จากแบบสอบที่นำมาเทียบมาตราเป็นไปตามเงื่อนไขต่าง ๆ ของการเทียบมาตราที่กำหนดไว้ในแต่ละรูปแบบ แต่ในความเป็นจริงอาจมีข้อจำกัดบางประการทำให้ไม่สามารถได้ข้อมูลตรงตามเงื่อนไขที่กำหนดได้ ดังนั้นการเทียบมาตราจึงจำเป็นต้องมีการตรวจสอบคุณภาพของการเทียบมาตรา ซึ่งมีวิธีการตรวจสอบคุณภาพของการเทียบมาตราดังนี้

1) ความเพียงพอในการเทียบมาตรา

ปีเตอร์เซนและคณะ (Petersen et.al. , 1982 อ้างถึงใน พรพิมล นาคเวช, 2537 : 65) ได้เสนอดัชนีความแตกต่างระหว่างคะแนนเกณฑ์กับคะแนนที่ได้จากการเทียบมาตรา โดยถ่วงน้ำหนักด้วยค่าความแปรปรวน เพื่อให้ได้ค่าที่เป็นมาตรฐานโดยมีสูตรดังนี้

$$C = \frac{\sum (x_i - x_i^*)^2}{ns_x^2}$$

เมื่อ	C	คือ	ค่าดัชนีความแตกต่าง
	X_i	คือ	คะแนนของคนที่ i ที่ได้จากการสอบ
	X_i^*	คือ	คะแนนของคนที่ i ที่ได้จากการเทียบมาตรา
	n	คือ	จำนวนคนในกลุ่มสอบทานผล
	S_x^2	คือ	ค่าความแปรปรวนของคะแนนจากแบบสอบฉบับ X

เกณฑ์ที่ใช้ในการประเมินค่าดัชนีความแตกต่างใช้เกณฑ์ของปีเตอร์เซนและคณะ (Petersen and Other, 1982 : 93-94 อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 102-103) โดยกำหนดระดับการยอมรับดังนี้

ระดับน่าพอใจอย่างยิ่ง	เมื่อ	$C \leq (0.05 SD_x)^2$
ระดับน่าพอใจ	เมื่อ	$(0.05 SD_x)^2 < C \leq (0.10 SD_x)^2$
ระดับปานกลาง	เมื่อ	$(0.10 SD_x)^2 < C \leq (0.15 SD_x)^2$
ระดับไม่น่าพอใจ	เมื่อ	$(0.15 SD_x)^2 < C \leq (0.20 SD_x)^2$
ระดับไม่น่าพอใจอย่างยิ่ง	เมื่อ	$C > (0.20 SD_x)^2$

2) ความคลาดเคลื่อนของการเทียบมาตรา

การใช้วิธีการเทียบมาตราทุกวิธี กลุ่มตัวอย่างที่ได้จากการสุ่มจากประชากรเดียวกันหรือหลายประชากร จะมีความผันแปรเชิงสุ่มเกิดขึ้นกับกลุ่มตัวอย่างที่ทำการศึกษา ดังนั้นสามารถวิเคราะห์ความคลาดเคลื่อนที่เกิดขึ้นในเชิงทฤษฎี สำหรับการเทียบมาตราระหว่างแบบสอบจากกลุ่มทวนผลได้ดังนี้ ให้ E เป็นความคลาดเคลื่อนของการเทียบมาตรา

$$E = |X_i - X_i^*|$$

เมื่อ X_i คือ คะแนนของคนที่ i ที่ได้จากการสอบแบบสอบ X

X_i^* คือ คะแนนของคนที่ i ที่ได้จากการตารางเทียบมาตรา

งานวิจัยที่เกี่ยวข้อง

งานวิจัยภายในประเทศ

ชูชีพ พงษ์สมบุญ (2528 : บทคัดย่อ) ได้ศึกษาเปรียบเทียบรูปแบบการเทียบมาตรา 3 วิธี คือ การเทียบมาตราเชิงเส้นตรง การเทียบแบบอิกวิเปอร์เซนไทล์ และการเทียบโดยใช้โค้งลักษณะข้อสอบ จากรูปแบบที่ใช้แบบสอบร่วมและใช้ผู้สอบร่วม เพื่อต้องการเปรียบเทียบประสิทธิภาพและความคงที่ของการเทียบมาตราทั้ง 3 วิธี ใช้กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศสุ่มมาจำนวน 3,721 คน ที่เข้าสอบโครงการตรวจคุณภาพการศึกษาของกรมวิชาการ กระทรวงศึกษาธิการ แบบสอบมีข้อสอบฉบับละ 38 ข้อ มีแบบสอบร่วมภายใน 20 ข้อ ผลการวิจัยพบว่า ประสิทธิภาพของการเทียบมาตราระหว่างรูปแบบข้อสอบร่วมและผู้สอบร่วมในแต่ละวิธีมีประสิทธิภาพไม่ต่างกัน ในแบบแผนที่ใช้ผู้สอบร่วมวิธีอิกวิเปอร์เซนไทล์มีความคงที่ใกล้เคียงกับวิธีโค้งลักษณะข้อสอบ แต่มีความคงที่มากกว่าวิธีเชิงเส้นตรง และความคงที่ของการเทียบมาตราในรูปแบบที่ใช้ข้อสอบร่วมไม่ต่างกับรูปแบบที่ใช้ผู้สอบร่วม

ภาวิณี ศรีสุขวัฒนานันท์ (2529 : 155-166) ได้เปรียบเทียบผลของการใช้วิธีการเทียบมาตรา 3 วิธี คือวิธีอิกวิเปอร์เซนไทล์ วิธีเชิงเส้นตรง และวิธีทฤษฎีการตอบสนองข้อสอบที่มี 3 พารามิเตอร์ โดยใช้แบบสอบร่วมภายในที่ต่างกัน 3 ขนาด คือร้อยละ 20 (7 ข้อ) ร้อยละ 40 (14 ข้อ) และร้อยละ 60 (21 ข้อ) โดยใช้แบบสอบเทียบมาตราสองฉบับ ฉบับละ 35 ข้อ กลุ่มตัวอย่างแยกเป็นสองกรณี คือ กรณีแบบสอบคัดเลือกและกรณีแบบสอบวัดผลสัมฤทธิ์ ผลการวิจัยพบว่า การใช้แบบสอบร่วมที่ยาวกว่าให้ประสิทธิภาพของการเทียบมาตราสูงกว่าในทั้งสองกรณี ส่วนการเปรียบเทียบระหว่างวิธีการเทียบมาตราพบว่าในกรณีแบบสอบคัดเลือกนั้นวิธีที่ให้ความเพียงพอมากที่สุด คือวิธีอิกวิเปอร์เซนไทล์ รองลงมาคือวิธีทฤษฎีการ

ตอบสนองข้อสอบ ซึ่งให้ผลที่แตกต่างกับกรณีแบบสอบวัดผลสัมฤทธิ์ที่พบว่า วิธีที่ให้ความเพียงพอมากที่สุดคือวิธีเชิงเส้นตรง รองลงมาคือวิธีอควิเปอร์เซนไทล์

สุจินดา ผ่องอักษร (2533 : บทคัดย่อ) ได้ทำการศึกษาความก้าวหน้าของผลสัมฤทธิ์ทางการเรียนในกลุ่มทักษะ ของนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เรียนจบตามหลักสูตรประถมศึกษา พุทธศักราช 2521 ในช่วงระยะเวลา 3 ปีการศึกษา (ปีการศึกษา 2529 ถึง 2531) โดยใช้การเทียบมาตรฐานแบบราสช์ ทำการเทียบมาตรฐานโดยหาค่าคงที่ เพื่อใช้ในสมการเชิงเส้นตรง สำหรับปรับคะแนนที่ได้จากแบบสอบ ต่างชุดในแต่ละกลุ่มวิชาให้อยู่ในสเกลเดียวกันทั้งหมด โดยใช้สเกลคะแนนผลการสอบเมื่อปีการศึกษา 2529 เป็นหลัก หลังจากนั้น จึงนำคะแนนที่สมมูลกันระหว่างแบบสอบต่างชุดมาเปรียบเทียบกันโดยตรง พบว่าหลังจากการเทียบมาตรฐานแล้ว คะแนนดิบที่เท่ากัน ระหว่างแบบสอบต่างชุดมีค่าแตกต่างกัน โดยในกลุ่มทักษะวิชา ภาษาไทยมีความแตกต่างกันสูงสุด 3.8 คะแนน ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์มีความแตกต่างกันสูงสุด 6.3 คะแนน จุดตัดในแต่ละปีการศึกษาของทั้งสามปีการศึกษามีค่าแตกต่างกัน คือจุดตัดในกลุ่มทักษะวิชาภาษาไทยมีค่าเท่ากับ 15, 18 และ 14 ตามลำดับ ส่วนจุดตัดในกลุ่มทักษะวิชาคณิตศาสตร์มีค่าเท่ากับ 19, 24 และ 18 ตามลำดับ ดังนั้นความก้าวหน้าของผลสัมฤทธิ์ ทางการเรียนในแต่ละปีการศึกษา และการตัดสินใจผลการสอบในแต่ละปีการศึกษาเมื่อก่อนเทียบมาตรฐานและหลังการเทียบมาตรฐานจึงให้ผลแตกต่างกัน ในกลุ่มทักษะวิชาภาษาไทยส่วนใหญ่ นักเรียนมีผลสัมฤทธิ์ทางการเรียนสูงกว่าระดับที่น่าพอใจ และมีผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิมตามลำดับ ส่วนในกลุ่มทักษะวิชาคณิตศาสตร์พบว่าส่วนใหญ่มีผลสัมฤทธิ์ทางการเรียนต่ำกว่าระดับที่น่าพอใจ และมีผลสัมฤทธิ์ที่ก้าวหน้าขึ้นกว่าเดิมเฉพาะในปีการศึกษา 2530 แต่ในปีการศึกษา 2531 พบว่ามีผลสัมฤทธิ์ลดลงจากเดิม ผลการเปรียบเทียบสัดส่วนจำนวนนักเรียนที่ผ่านเกณฑ์ที่น่าพอใจ ระหว่างก่อนเทียบมาตรฐานและหลังจากการเทียบมาตรฐาน พบว่ามีความแตกต่างกัน

พรพิมล นาคเวช (2537 : 99-102) ได้ทำการศึกษาคุณภาพของการเทียบมาตรฐานในแนวตั้งโดยใช้ทฤษฎีการตอบสนองข้อสอบ สำหรับผู้ที่มีความสามารถ 3 ระดับ คือนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ผลการศึกษาพบว่า ความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานในแนวตั้งมีค่าอยู่ระหว่าง 0.031– 0.051 ซึ่งเป็นความคลาดเคลื่อนมาตรฐานที่ค่อนข้างต่ำ และเมื่อวิเคราะห์ความเพียงพอของการเทียบมาตรฐานพบว่า ค่าดัชนีความแตกต่างของการเทียบมาตรฐานอยู่ในระดับที่น่าพอใจ

วรเนตร พิชิตกริกพล (2537 : 94-100) ได้เปรียบเทียบคุณภาพของวิธีการเทียบมาตรฐานระหว่างแบบอควิเปอร์เซนไทล์กับแบบเชิงเส้นตรง โดยศึกษากรณีแบบสอบที่มีความยากใกล้เคียงกันและแตกต่างกัน สอบกับกลุ่มตัวอย่างที่มีระดับความสามารถใกล้เคียงกันและแตกต่างกัน ผลการวิจัยพบว่า ในกรณีแบบสอบมีความยากใกล้เคียงกันและกลุ่มตัวอย่างที่มีความสามารถใกล้เคียงกันวิธีเชิงเส้นตรงให้คุณภาพดีกว่าวิธีอควิเปอร์เซนไทล์ แต่กับกลุ่มตัวอย่างที่มีระดับความสามารถแตกต่างกันวิธีอควิเปอร์เซนไทล์ให้

คุณภาพดีกว่าวิธีเชิงเส้นตรง ส่วนในกรณีที่แบบสอบมีความยากแตกต่างกันกลุ่มตัวอย่างมีความสามารถใกล้เคียงกันวิธีอิกวิเปอร์เซนไทล์ให้คุณภาพดีกว่าวิธีเชิงเส้นตรง แต่เมื่อใช้กับกลุ่มตัวอย่างที่มีระดับความสามารถแตกต่างกันพบว่า การเทียบมาตราทั้งสองวิธีมีความเพียงพออยู่ในระดับไม่น่าพอใจ

สุนิสา จุ้ยม่วงศรี (2537 : 88-97) ได้ศึกษาวิจัยผลของความยาวของแบบสอบร่วมที่มีต่อคุณภาพของการเทียบมาตราเชิงเส้นตรง โดยใช้แบบสอบร่วมภายในที่มีขนาดความยาวต่างกัน 4 ขนาด คือ 10 15 20 และ 25 ข้อ โดยการเปรียบเทียบความคลาดเคลื่อนมาตรฐานของการเทียบมาตราและดัชนีความแตกต่างจากการวิเคราะห์กลุ่มสอบทานผล ผลการวิจัยพบว่า ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตราเชิงเส้นตรงในแบบสอบร่วมที่ยาวกว่าให้ค่าน้อยกว่าและค่าดัชนีความแตกต่างของการเทียบมาตราให้ผลอยู่ในระดับที่ยอมรับได้ทั้งสิ้น

พิชัย ละแมนชัย (2538 : บทคัดย่อ) ได้ทำการศึกษาเกี่ยวกับขนาดกลุ่มตัวอย่างขั้นต่ำสำหรับการเทียบมาตราในแนวนอนตามทฤษฎีการตอบสนองข้อสอบ ระหว่างแบบสอบที่มีแบบแผนการเทียบมาตราและความยาวแบบสอบแตกต่างกัน โดยใช้ข้อมูลจำลองจากโปรแกรม IRTDATA ผลการวิจัยพบว่า รูปแบบทฤษฎีการตอบสนองข้อสอบที่มี 1 พารามิเตอร์ และ 3 พารามิเตอร์ ทั้งแบบแผนกลุ่มผู้สมมุติและแบบแผนข้อสอบร่วม มีขนาดกลุ่มตัวอย่างขั้นต่ำในแต่ละความยาวแบบสอบแตกต่างกัน โดยรูปแบบ 1 พารามิเตอร์ใช้ขนาดกลุ่มตัวอย่างขั้นต่ำที่น้อยกว่า และเมื่อเปรียบเทียบระหว่างแบบแผนกลุ่มผู้สมมุติและแบบแผนข้อสอบร่วมพบว่า แบบแผนข้อสอบร่วมใช้ขนาดกลุ่มตัวอย่างขั้นต่ำที่น้อยกว่า

รุ่งนภา จันทรา (2540 : 71-75) ได้การเปรียบเทียบความคลาดเคลื่อนและความเพียงพอของการเทียบมาตราระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์ โดยเป็นการเทียบมาตราในแนวนอน ผลการศึกษาพบว่า ค่าดัชนีความเพียงพอของการเทียบมาตรารูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์อยู่ในระดับไม่น่าพอใจ ส่วนรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์มีค่าดัชนีความเพียงพอของการเทียบมาตราอยู่ในระดับน่าพอใจ และความคลาดเคลื่อนของการเทียบมาตรารูปแบบทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์น้อยกว่าความคลาดเคลื่อนของการเทียบมาตรารูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์อย่างมีนัยสำคัญทางสถิติที่ระดับ .01

กนิษฐา แสนแก้ว (2541 : บทคัดย่อ) ได้ศึกษาการปรับแก้เฉลยสะสมระดับมัธยมศึกษาตอนปลายด้วยคะแนนความถนัดทางการเรียน โดยเปรียบเทียบระหว่างวิธีอิกวิเปอร์เซนไทล์ วิธีเชิงเส้นตรง และวิธีไออาร์ที ผลการวิจัยพบว่า เฉลยสะสมระดับมัธยมศึกษาตอนปลายที่ปรับแก้ทั้งสามวิธีมีความสัมพันธ์กับคะแนนสอบเข้ามหาวิทยาลัยและผลการเรียนเฉลี่ยระดับอุดมศึกษาสูงกว่าเฉลยสะสมระดับมัธยมศึกษาตอนปลายปกติ และเมื่อใช้ผลการเรียนเฉลี่ยระดับอุดมศึกษาภาคเรียนที่ 1 เป็น

เกณฑ์วิธีไออาร์ที่เป็นวิธีที่มีคุณภาพมากที่สุด รองลงมาคือวิธีเชิงเส้นตรงและวิธีอิกวิเปอร์เซนไทล์ตามลำดับ แต่เมื่อใช้คะแนนสอบเข้ามหาวิทยาลัยเป็นเกณฑ์วิธีอิกวิเปอร์เซนไทล์กลับเป็นวิธีที่มีคุณภาพมากที่สุด รองลงมา คือ วิธีไออาร์ที่ และวิธีเชิงเส้นตรงตามลำดับ

ปราณี ศรีทองแก้ว (2541 : 64-67) ได้เปรียบเทียบคุณภาพของการเทียบมาตราในแนวตั้งระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์ โดยเปรียบเทียบความคลาดเคลื่อนและตรวจสอบความเพียงพอของการเทียบมาตรา ซึ่งผลการศึกษาพบว่า ความคลาดเคลื่อนของการเทียบมาตราในแนวตั้งระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์ ไม่แตกต่างกัน และความเพียงพอของการเทียบมาตราในแนวตั้งทั้งรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์และสามพารามิเตอร์ อยู่ในระดับที่น่าพอใจอย่างยิ่ง

วิรัช วรรณรัตน์ (2541 : 43-52) ได้ศึกษาวิจัยผลการปรับระดับผลการเรียน โดยใช้วิธีการเทียบมาตราและวิธีการเปอร์เซนไทล์ โดยวิธีการเทียบมาตราที่ใช้เป็นวิธีการเทียบมาตราจาก GPA ที่ผู้วิจัยได้พัฒนาขึ้นเพื่อปรับระดับผลการเรียนเป็นคะแนนการสอบส่วนวิธีการเปอร์เซนไทล์เป็นวิธีการแปลงระดับผลการเรียนโดยใช้ตำแหน่งเปอร์เซนไทล์ของ GPA ของผู้สอบเป็นเกณฑ์โดยพิจารณาเป็นสองแบบคือแบบ 100 ช่วง และแบบ 20 ช่วง ผลการวิจัยพบว่า ค่าเฉลี่ยของคะแนนการสอบที่เป็นผลจากการปรับระดับผลการเรียน โดยวิธีการเทียบมาตรามีค่าสูงกว่าจากวิธีการเปอร์เซนไทล์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนคะแนนการสอบจากวิธีการเปอร์เซนไทล์ทั้งสองช่วงนั้นมีค่าเฉลี่ยไม่แตกต่างกัน และค่าสหสัมพันธ์อันดับที่ระหว่างคะแนนการสอบที่ปรับได้จากวิธีการเทียบมาตราที่รวมกับคะแนนการสอบข้อเขียน มีความสัมพันธ์กับคะแนนการสอบข้อเขียนในทิศทางบวก ($p=0.9476$) สูงกว่าจากวิธีการเปอร์เซนไทล์ 20 ช่วง และ 100 ช่วง ที่รวมกับคะแนนการสอบข้อเขียน ($p=0.7666$) และ ($p=0.7647$) ตามลำดับ

สุภาภรณ์ คงทวี (2541 : บทคัดย่อ) ได้ศึกษาเกี่ยวกับการปรับแก้เฉลี่ยสะสมระดับมัธยมศึกษาตอนปลายด้วยคะแนนสอบเข้ามหาวิทยาลัยโดยทำการเปรียบเทียบระหว่างวิธีการเทียบมาตราแบบอิกวิเปอร์เซนไทล์ เชิงเส้นตรง และวิธีไออาร์ที่ กลุ่มตัวอย่างเป็นนักศึกษาชั้นปีที่ 1 ปีการศึกษา 2540 ในมหาวิทยาลัยของรัฐ 5 แห่ง ผลการศึกษาพบว่า เฉลี่ยสะสมที่ปรับแก้ด้วยวิธีอิกวิเปอร์เซนไทล์ วิธีเชิงเส้นตรงและค่าพารามิเตอร์ความสามารถที่ปรับแก้ด้วยวิธีไออาร์ที่ มีความสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนชั้นปีที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .001 และสามารถทำนายผลสัมฤทธิ์ทางการเรียนได้อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนผลการเปรียบเทียบประสิทธิภาพของวิธีเทียบมาตราระหว่างวิธีอิกวิเปอร์เซนไทล์ วิธีเชิงเส้นตรง และไออาร์ที่ ผลปรากฏว่าในสาขาวิชาวิทยาศาสตร์วิธีเชิงเส้นตรงมีประสิทธิภาพดีกว่าวิธีอิกวิเปอร์เซนไทล์ สาขาวิชาวิศวกรรมศาสตร์วิธีอิกวิเปอร์เซนไทล์ และวิธีไออาร์ที่

มีประสิทธิภาพดีกว่าวิธีเชิงเส้นตรง ส่วนในกลุ่มสาขาวิชาอื่น ๆ วิธีการเปรียบเทียบทั้ง 3 วิธีให้ผลที่ไม่แตกต่างกัน

สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ (2542ก : 103-107) ได้ทำการเทียบมาตรฐานความสามารถทางคณิตศาสตร์ของนักเรียนชั้น ป.1 - ป.6 โดยใช้รูปแบบการเก็บข้อมูลแบบผู้สอบสองกลุ่มไม่เท่าเทียมกัน โดยมีแบบสอบร่วมภายใน ใช้วิธีการเทียบมาตรฐานในแนวตั้งตามทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์และวิธีการเทียบมาตรฐานตามทฤษฎีการสอบแบบดั้งเดิมโดยวิธีของเซอร์สโตน ผลการศึกษาพบว่า เมื่อใช้วิธีการเทียบมาตรฐานตามทฤษฎีการสอบแบบดั้งเดิมโดยวิธีของเซอร์สโตนผลการเทียบมาตรฐานแสดงให้เห็นถึงการมีพัฒนาการของนักเรียนในระดับชั้นที่สูงขึ้นอย่างต่อเนื่อง โดยชั้น ป.3 - ป.4 มีพัฒนาการในอัตราที่สูงกว่าชั้นอื่น ๆ เมื่อเปรียบเทียบระหว่างสมรรถภาพพบว่าทักษะการแก้โจทย์ปัญหามีพัฒนาการสูงกว่าสมรรถภาพอื่น ๆ และเมื่อใช้การเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบก็พบว่าให้ผลเหมือนกับวิธีของเซอร์สโตนเพียงแต่เมื่อเปรียบเทียบระหว่างสมรรถภาพพบว่าด้านความเข้าใจมีพัฒนาการสูงกว่าสมรรถภาพอื่น ๆ

กนกทอง มาทรัพย์ (2543 : บทคัดย่อ) ได้ศึกษาประเภทของแบบสอบร่วมที่มีผลต่อคุณภาพของวิธีการเทียบมาตรฐานเชิงเส้นตรง โดยเปรียบเทียบระหว่างแบบสอบร่วมภายในและแบบสอบร่วมภายนอก กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2542 สังกัดกรมสามัญศึกษา จังหวัดเพชรบูรณ์ โดยแบบสอบที่ใช้เทียบมาตรฐานเป็นแบบสอบวัดผลกลางภาค วิชาคณิตศาสตร์ ผลการวิจัยพบว่า ในกรณีที่ผลการสอบรายบุคคลมีคะแนนรวม 40 คะแนน การใช้แบบสอบร่วมภายในนอกกับการใช้แบบสอบร่วมภายใน มีความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานเชิงเส้นตรง (SEE) ไม่แตกต่างกัน แต่ในกรณีที่ผลการสอบรายบุคคลมีคะแนนรวม 60 คะแนน การใช้แบบสอบร่วมภายนอก มีความคลาดเคลื่อนมาตรฐานมากกว่าการใช้แบบสอบร่วมภายใน ส่วนการเปรียบเทียบดัชนีความแตกต่าง (C) ของการเทียบมาตรฐานเชิงเส้นตรง พบว่า ในกรณีที่ผลการสอบรายบุคคลมีคะแนนรวม 40 คะแนน การใช้แบบสอบร่วมภายนอก มีค่าดัชนีความแตกต่างมากกว่าการใช้แบบสอบร่วมภายใน แต่ในกรณีที่ผลการสอบรายบุคคลมีคะแนนรวม 60 คะแนน ค่าดัชนีความแตกต่างระหว่างการใช้แบบสอบร่วมภายนอกกับการใช้แบบสอบร่วมภายใน ไม่แตกต่างกัน

น้ำฝน กันมา (2543 : บทคัดย่อ) ได้ศึกษาวิธีการแปลงคะแนนสอบคัดเลือกเข้าศึกษาในสถาบันอุดมศึกษาโดยวิธีการของทบวงมหาวิทยาลัย วิธีการแปลงคะแนนมาตรฐานที่ (T-Score) และวิธีการเทียบมาตรฐานโดยใช้ข้อสอบร่วม กลุ่มตัวอย่างที่ใช้คือนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2542 สังกัดกรมสามัญศึกษา จำนวน 1,642 คน โดยจำแนกเป็นนักเรียนจากโรงเรียนขนาดใหญ่พิเศษจำนวน 1,062 คน

โรงเรียนขนาดใหญ่จำนวน 399 คน โรงเรียนขนาดกลางจำนวน 140 คน และโรงเรียนขนาดเล็กจำนวน 77 คน ผลการศึกษาพบว่า

- 1) โรงเรียนที่มีขนาดต่างกันมีคะแนนแปลงผลการเรียนสะสมโดยวิธีของทบวงมหาวิทยาลัยไม่แตกต่างกัน
 - 2) โรงเรียนที่มีขนาดต่างกันมีคะแนนแปลงผลการเรียนสะสมโดยวิธีการแปลงคะแนนมาตรฐานที่ของนักเรียนโปรแกรมวิทย์ไม่ต่างกัน แต่นักเรียนโปรแกรมศิลป์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 และเมื่อทดสอบความแตกต่างเป็นรายคู่ พบว่า โรงเรียนขนาดใหญ่พิเศษมีค่าเฉลี่ยของคะแนนแปลงผลการเรียนสะสมสูงกว่าโรงเรียนขนาดเล็ก
 - 3) โรงเรียนที่มีขนาดต่างกันมีคะแนนแปลงผลการเรียนสะสมโดยวิธีการเทียบมาตราโดยใช้ข้อสอบร่วมของนักเรียนโปรแกรมวิทย์ไม่แตกต่างกันแต่นักเรียนโปรแกรมศิลป์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 และเมื่อทดสอบความแตกต่างเป็นรายคู่ พบว่า โรงเรียนขนาดใหญ่พิเศษมีค่าเฉลี่ยของคะแนนแปลงผลการเรียนสะสมสูงกว่าโรงเรียนขนาดเล็ก
 - 4) คะแนนแปลงผลการเรียนสะสมโดยวิธีของทบวงมหาวิทยาลัย วิธีการแปลงคะแนนมาตรฐานที่และวิธีการเทียบมาตราโดยใช้ข้อสอบร่วมมีความสัมพันธ์กันในทางบวกอย่างมีนัยสำคัญทางสถิติ
- ปทุมทริกา คิววษ์ (2544 : บทคัดย่อ) ได้ศึกษา คุณ ภาพ ของ การ เทียบ มา ตรา แบบ ทด สอบ คู่ ข น น าน ดั ว ย ร ู ป แบบ ก าร เทียบ มา ตรา เ ชิง เ ลื น ต ร ง กลุ่ม ตัวอย่าง ที่ ใช้ เป็น นักเรียน ชั้น มัธยมศึกษา ปี ที่ 5 ภาค เรียน ที่ 1 ปี การ ศึกษา 2541 ใน จั ง ห ว ัด ม หา ส าร ค าม เครื่อง มือ ที่ ใช้ เป็น แบบ สอบ คู่ ข น น าน วิ ช า ฟิสิกส์ ว 022 เรื่อง สมดุล กล จำนวน 2 ฉบับ ฉบับ ละ 40 ข้อ ซึ่ง แต่ละ ฉบับ จะ มี แบบ สอบ ร ่วม ก าย ใน (Internal Anchor Test) จำนวน 10 ข้อ ผล การ ศึกษา พบ ว่า ความ คลาด เคลื่อน มา ตรา ฐาน ของ การ เทียบ มา ตรา เฉลี่ย เท่า กับ 0.5974 ซึ่งเป็น ค่า ที่ ค ่อ น ข ้าง ต ่ำ และ ค่า ดั ช นี ความ แตก ต่าง ของ การ เทียบ มา ตรา (C) มี ค่า เท่า กับ 0.197 ซึ่ง อยู่ ใน ระ ด บ ที่ น ำ พ อ ใจ ซึ่ง ผู้ วิ จ ย แ ส ด ง ความ คิด เห็น ว่า การ เทียบ มา ตรา เ ชิง เ ลื น ต ร ง มี ความ เ หมาะ สม และ มี คุณ ภาพ เพียง พอ ที่ จะ ใช้ ใน การ เทียบ มา ตรา สำหรับ แบบ สอบ คู่ ข น น าน

งานวิจัยต่างประเทศ

สไลด์และลิน (Slind and Linn, 1977, 1978, 1979 อ้างถึงใน ภาวิณี ศรีสุขวัฒนานันท์, 2529 : 68) ได้ทำการศึกษาปัญหาของการเทียบมาตราในแนวตั้ง (Vertical equating) ของแบบสอบสองชุดที่สร้างขึ้นเพื่อใช้กับกลุ่มตัวอย่างที่มีระดับความสามารถต่างกัน จากการศึกษาได้ให้ข้อเสนอแนะว่าการใช้วิธีการเทียบมาตรา 3 รูปแบบ คือ การเทียบมาตราเชิงเส้นตรง รูปแบบอิกวิเปอร์เซนไทล์ และรูปแบบทฤษฎีการตอบสนองข้อสอบชนิดหนึ่งพารามิเตอร์รูปแบบโลจิสอาจมีข้อจำกัดในการเทียบมาตราในแนว

ตั้ง โดยเฉพาะอย่างยิ่งเมื่อทำการเทียบมาตราโดยแบบสองสองชุดที่มีความแตกต่างกันมาก และกลุ่มตัวอย่างสองกลุ่มที่มีความแตกต่างกันมาก จากการศึกษาให้ข้อสังเกตว่า ถ้าใช้รูปแบบทฤษฎีการตอบสนองข้อสอบชนิดสามพารามิเตอร์อาจให้ผลการเทียบมาตราที่ดีกว่าในสถานการณ์เช่นนั้น

คูก์ ดันบาร์ และไอเนอร์ (Cook, Dunbar and Eignor, 1981 อ้างถึงใน วรเนตร พิชิตเกริกพล, 2537 : 43-44) ได้ศึกษาผลของการเทียบมาตราโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์เปรียบเทียบกับวิธีการเทียบมาตราแบบดั้งเดิม 2 วิธี คือ การเทียบมาตราเชิงเส้นตรงและการเทียบมาตราแบบอิกวิเปอร์เซนไทล์ โดยศึกษาทั้งกรณีมีแบบสอบร่วมและไม่มีแบบสอบร่วม ผลการวิจัยพบว่าวิธีการเทียบมาตราแบบดั้งเดิมกับวิธีการเทียบมาตราแบบทฤษฎีการตอบสนองข้อสอบสอดคล้องกันมาก เนื่องจากการแจกแจงคะแนนดิบของแบบสอบ 2 ชุดมีรูปร่างคล้ายคลึงกันมาก อย่างไรก็ตามจากความแตกต่างที่เกิดขึ้นที่ส่วนปลายของวิธีการเทียบมาตราแบบดั้งเดิมทุกวิธีที่ต่างจากการเทียบมาตราแบบทฤษฎีการตอบสนองข้อสอบเป็นผลจากข้อมูลในส่วนนี้มีน้อยและหายากมาก ในทางทฤษฎีแนะนำให้ใช้การเทียบแบบทฤษฎีการตอบสนองข้อสอบเพราะไม่มีผลเนื่องจากการขาดแคลนข้อมูลในส่วนปลายการแจกแจง แต่การเทียบแบบทฤษฎีการตอบสนองข้อสอบก็ไม่สามารถเทียบมาตราในส่วนที่อยู่ปลายด้านล่างที่ต่ำกว่าระดับการเดาได้ ซึ่งจำเป็นต้องใช้วิธีอื่นประมาณค่าเพิ่มเติม นอกจากนี้ผลการวิจัยยังพบว่า การเทียบมาตราแบบทฤษฎีการตอบสนองข้อสอบโดยตรงเมื่อไม่มีแบบสอบร่วมและเมื่อกลุ่มตัวอย่างไม่ได้เป็นกลุ่มที่ทำเทียมกัน โดยสุ่มนั้นให้ผลแตกต่างจากการเทียบเมื่อมีแบบสอบร่วมเพียงเล็กน้อย

โคเลน และวิทนีย์ (Kolen and Whitney, 1982 : 279-293) ได้เปรียบเทียบคุณภาพของการเทียบมาตรา 4 วิธี คือ วิธีอิกวิเปอร์เซนไทล์ วิธีเชิงเส้นตรง วิธีทฤษฎีการตอบสนองข้อสอบชนิดหนึ่งพารามิเตอร์และ สามพารามิเตอร์ โดยใช้แบบสอบมาตรฐานวัดผลสัมฤทธิ์ทางการเรียนใช้กลุ่มตัวอย่างประมาณกลุ่มละ 200 คน ผลการวิจัยพบว่า รูปแบบเชิงเส้นตรงให้ผลการเทียบมาตราที่มีความเพียงพอมากกว่ารูปแบบอิกวิเปอร์เซนไทล์และรูปแบบทฤษฎีการตอบสนองข้อสอบ ขณะเดียวกันรูปแบบทฤษฎีการตอบสนองข้อสอบชนิดหนึ่งพารามิเตอร์ให้ผลการเทียบมาตราที่เพียงพอกว่ารูปแบบสามพารามิเตอร์

ปีเตอร์เซนและคณะ (Petersen and other, 1983 : 137-156 อ้างถึงใน ปุณทริกา ศิววงศ์, 2544 : 74) ได้ศึกษาวิธีการเทียบมาตรา 3 วิธี คือ รูปแบบเชิงเส้นตรง รูปแบบอิกวิเปอร์เซนไทล์ และ รูปแบบทฤษฎีการตอบสนองข้อสอบ โดยใช้คะแนนจากแบบสอบวัดความถนัดทางการเรียน ผลการวิจัยพบว่ารูปแบบทฤษฎีการตอบสนองข้อสอบได้ผลมีความคงที่ที่สุด

ฮิลล์ สับฮิยาห์ และเฮิร์ช (Hills, Subhiyah and Hirsch, 1988 : 221-231) ได้ศึกษาเปรียบเทียบวิธีการเทียบมาตรา 5 วิธี คือ วิธีเทียบมาตราเชิงเส้นตรง และวิธีไออาร์ที 4 วิธี ได้แก่วิธี ไออาร์ที 3 พารามิเตอร์ คือ Concurrent Method (IRTCO) Fixed-parameter Method (IRTFIX) และ Formula

Method (IRTFOR) และวิธีการราสซ์โมเดล และศึกษาขนาดความยาวของแบบสอบรวม 6 ขนาด คือ ขนาด 30 25 20 15 10 และ 5 ข้อโดยใช้การเทียบมาตรฐานแบบ Concurrent IRT ผลการประเมินวิธีการเทียบมาตรฐานทั้ง 5 วิธีไม่มีการสรุปอย่างเป็นแบบแผน เพราะได้นำผลการเทียบมาตรฐานเชิงเส้นตรงมาเป็นเกณฑ์ในการเปรียบเทียบ นั่นคือเมื่อเทียบมาตรฐานในแต่ละวิธีแล้วนำมาหาความแตกต่างของคะแนนในแต่ละวิธีกับคะแนนที่ได้จากการเทียบมาตรฐานเชิงเส้นตรง แต่ผู้วิจัยได้อภิปรายผลไว้ว่า ในโปรแกรมการสอบทั่ว ๆ ไปวิธีการไออาร์ทีไม่ใช่วิธีเดียวที่จะใช้ในการพัฒนาแบบสอบและประเมินคุณภาพของข้อสอบ แต่วิธีการเชิงเส้นตรงก็เป็นวิธีการที่ดีที่ควรนำมาใช้และเป็นวิธีการที่รู้จักกันอย่างกว้างขวาง สำหรับจำนวนข้อคำถามของแบบสอบรวมที่ใช้ในการเทียบมาตรฐาน พบว่าข้อสอบรวมตั้งแต่ 10 ข้อขึ้นไปก็มีประสิทธิภาพเพียงพอในการเทียบมาตรฐานเท่ากับแบบสอบรวมจำนวน 30 ข้อ แต่แบบสอบรวมขนาด 5 ข้อ ไม่มีประสิทธิภาพเท่า ซึ่งผู้วิจัยกล่าวว่าผลที่ได้นี้ต่างจากผลการวิจัยของ Wingersky และ Lord (1984) , Raju, Edwards และ Osberg (1983) และ Raju, Bode, Larsen และ Steinhaus (1986) ที่ชี้ให้เห็นว่าเมื่อใช้การเทียบมาตรฐานรูปแบบไออาร์ทีชนิด 3 พารามิเตอร์ การใช้ข้อสอบรวมจำนวน 5 หรือ 6 ข้อก็ให้ประสิทธิภาพเพียงพอ

ทริสคาริ (Triscari, 1990 อ้างถึงใน สำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ, 2542 ก : 22) ได้ทำการวิจัยเปรียบเทียบวิธีการเทียบมาตรฐาน 5 วิธี แบบเชิงเส้นตรง แบบอิกวิเปอร์เซนไทล์ และแบบทฤษฎีการตอบสนองข้อสอบ 1, 2 และ 3 พารามิเตอร์ โดยใช้รูปแบบการเก็บข้อมูลแบบกลุ่มตัวอย่างไม่เท่าเทียมกันและใช้ข้อสอบรวม พบว่า การเทียบมาตรฐานตามแนวทฤษฎีการตอบสนองข้อสอบให้ความคลาดเคลื่อนน้อยที่สุดและการเทียบมาตรฐานแบบเส้นตรงให้ความคลาดเคลื่อนมากที่สุด

ฮาน, โคลเลน และ พอลมันน์ (Han, Kolen, and Pohlmann, 1997 : บทคัดย่อ) ได้เปรียบเทียบคุณภาพของการเทียบมาตรฐานระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่ใช้คะแนนจริง คะแนนสังเกต และการเทียบมาตรฐานแบบอิกวิเปอร์เซนไทล์ โดยใช้แบบสอบที่มีความยากต่างกันสองฉบับ ผลการศึกษาพบว่า การเทียบมาตรฐานโดยใช้คะแนนจริงจะให้ผลที่คงที่มากที่สุด รองลงมาคือการเทียบมาตรฐานโดยใช้คะแนนสังเกต ซึ่งดัชนีความแตกต่างของการเทียบมาตรฐานโดยใช้ทฤษฎีการตอบสนองข้อสอบทั้งสองวิธีมีน้อยมาก ส่วนการเทียบมาตรฐานโดยวิธีอิกวิเปอร์เซนไทล์ให้ความคงที่น้อยที่สุด

จากงานวิจัยที่เกี่ยวข้องพอจะสรุปได้ว่า ไม่มีวิธีการเทียบมาตรฐานวิธีใดที่ให้ผลดีที่สุดในทุกสถานการณ์ แต่สามารถสรุปแนวทางในการตัดสินใจเลือกวิธีการในการเทียบมาตรฐานที่เหมาะสมได้ดังนี้

1. เมื่อแบบสอบเทียบมาตรฐานมีความยากที่ใกล้เคียงกันหรือเป็นแบบสอบคู่ขนาน และกลุ่ม

ตัวอย่างเป็นกลุ่มที่มีระดับความสามารถที่ใกล้เคียงกันแล้ว ควรใช้วิธีการเทียบมาตรฐานเชิงเส้นตรงเนื่องจากเป็นวิธีที่สามารถทำได้สะดวกและให้ผลการเทียบมาตรฐานอยู่ในระดับที่น่าพอใจ

2. เมื่อแบบสอบเทียบมาตรฐานมีความยากแตกต่างกันมาก และกลุ่มตัวอย่างทั้งสองกลุ่มเป็นกลุ่มที่มีระดับความสามารถแตกต่างกันมาก หรือเป็นการเทียบมาตรฐานในแนวตั้ง การเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบชนิดสามพารามิเตอร์จะให้ผลเป็นที่น่าพอใจมากกว่าวิธีอื่น ๆ

3. เมื่อกลุ่มตัวอย่างมีขนาดเล็กและแบบสอบเทียบมาตรฐานวัดในสิ่งเดียวกันแล้ว การเทียบมาตรฐานเชิงเส้นตรงและการเทียบมาตรฐานรูปแบบทฤษฎีการตอบสนองข้อสอบชนิดหนึ่งพารามิเตอร์ให้ผลที่น่าพอใจกว่าวิธีอื่น ๆ

4. เมื่อแบบสอบมีความยากไม่ต่างกัน แต่กลุ่มตัวอย่างเป็นกลุ่มที่มีระดับความสามารถต่างกัน วิธีอิกวิเปอร์เซนไทล์จะมีคุณภาพดีกว่าวิธีเทียบมาตรฐานเชิงเส้นตรง