

การปรับเทียบคะแนนแบบทดสอบด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์เพื่อหาคุณภาพ ของแบบทดสอบต่างฉบับ

ผู้ช่วยศาสตราจารย์ ดร.วินิตา แก้วเกื้อ

บทคัดย่อ

การเปรียบเทียบคุณภาพของแบบทดสอบต่างฉบับสามารถทำได้หลายวิธี ซึ่งหนึ่งในวิธีการเปรียบเทียบคุณภาพของแบบทดสอบต่างฉบับ คือ การปรับเทียบคะแนนของแบบทดสอบ ทั้งนี้การปรับเทียบคะแนนสามารถแบ่งออกเป็น 2 ประเภท คือ การปรับเทียบคะแนนตามแนวนอน (Horizontal Equating) และ การปรับเทียบคะแนนตามแนวตั้ง (Vertical Equating) และวิธีการที่ใช้ในการปรับเทียบคะแนนมีหลายวิธีเช่นกัน แต่สามารถจำแนกตามแนวคิดพื้นฐานของการปรับเทียบได้เป็นรูปแบบหลัก 2 รูปแบบ ได้แก่ รูปแบบการปรับเทียบคะแนนโดยใช้ทฤษฎีดั้งเดิม (Classical Models of Equating) และรูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Equating through Item Response Theory) ในส่วนของบทความนี้มีวัตถุประสงค์เพื่อนำเสนอ การปรับเทียบคะแนนแบบทดสอบด้วยวิธีอิกวิเปอร์เซ็นต์ไทล์ ซึ่งเป็นรูปแบบการปรับเทียบคะแนนโดยใช้ทฤษฎีดั้งเดิม โดยนำเสนอใน 3 หัวข้อ คือ 1.ความหมายของการปรับเทียบคะแนน 2. รูปแบบวิธีการปรับเทียบคะแนน และ 3. การประยุกต์ใช้โปรแกรม jamovi ในการปรับเทียบคะแนนแบบอิกวิเปอร์เซ็นต์ไทล์

บทนำ

วิธีการเทียบมาตรฐานเป็นรากฐานของการทดสอบทางการศึกษา ช่วยให้สามารถจัดคะแนนจากการทดสอบให้สอดคล้องกันในรูปแบบการทดสอบหรือการบริหารการทดสอบที่แตกต่างกันบนมาตราการวัดทั่วไป (common scale) เพื่อให้เกิดความยุติธรรมและสามารถเปรียบเทียบได้ (Kolen & Brennan, 2014) ในบรรดาวิธีการเทียบมาตรฐานที่หลากหลาย มีแนวทางที่ถูกนำมาใช้สองแนวทางที่โดดเด่นคือ การปรับเทียบเชิงเส้น (linear equating) และการปรับเทียบแบบอิกวิเปอร์เซ็นต์ไทล์ตามทฤษฎีการทดสอบแบบคลาสสิก (Equipercentile Classical Test Theory Equating) (Daramola et al., 2018)

การปรับเทียบแบบอิกวิเปอร์เซ็นต์ไทล์ตามทฤษฎีการทดสอบแบบคลาสสิก เป็นทางเลือกในการเทียบคะแนนที่แตกต่างไปจากหลักการเทียบคะแนนแบบเส้นตรง (Daramola et al., 2018) โดยการเทียบคะแนนแบบเส้นตรงถือว่ามีความสัมพันธ์เชิงเส้นตรงระหว่างคะแนนในแบบทดสอบต่าง ๆ ส่วนการเทียบคะแนนแบบอิกวิเปอร์เซ็นต์ไทล์ จะเน้นที่การจัดลำดับเปอร์เซ็นต์ให้ตรงกันในแบบทดสอบต่าง ๆ เพื่ออธิบายความแตกต่างของระดับความยาก (Oguoma, 2017) ซึ่งทั้งสองวิธีมีเป้าหมายเดียวกันคือการรักษาระดับความสัมพันธ์ (relative) ของผู้เข้าสอบภายในกลุ่มของตน โดยต้องแน่ใจว่าบุคคลที่มีระดับความสามารถที่ใกล้เคียงกันจะได้รับอันดับเปอร์เซ็นต์ไทล์ที่เทียบเคียงได้ไม่ว่าจะทำแบบทดสอบใดก็ตาม (Rossiter et al., 2023)

แต่หลักการเทียบคะแนนแบบเส้นตรงนี้ยอมรับว่าแบบทดสอบอาจมีระดับความยากที่แตกต่างกัน ซึ่งอาจทำให้คะแนนดิบของผู้เข้าสอบมีการกระจายไม่เท่ากัน เพื่อแก้ไขปัญหานี้ การปรับเทียบแบบ อิกวิเปอร์เซ็นไทล์ตามทฤษฎีการทดสอบแบบคลาสสิก จะปรับคะแนนตามลำดับเปอร์เซ็นต์ไทล์จึงทำให้ คะแนนมีความยุติธรรมและสามารถเปรียบเทียบกันได้มากขึ้น (Wiberg, 2021)

สำหรับบทความนี้เลือกนำเสนอวิธีการปรับเทียบคะแนนตามแนวนอน เนื่องจากการปรับเทียบ ของชุดวิชาเดียวกันที่มีความยากใกล้เคียงกัน โดยเป็นแบบทดสอบต่างฉบับกันที่มีการจัดทำแบบทดสอบแบ่ง ตามแผนการศึกษา ก1 ก2 และ ก3 ตามบริบทของมหาวิทยาลัยสุโขทัยธรรมมาธิราช ดังนั้น จึงเป็นการจัดสอบ ในรอบการสอบเดียวกันและกลุ่มผู้สอบมีความสามารถใกล้เคียงกัน คืออยู่ในระดับชั้นเดียวกัน ในที่นี้คือเป็น กลุ่มผู้สอบที่เป็นนักศึกษามหาวิทยาลัยระดับปริญญาตรีเหมือนกัน แต่เป็นกลุ่มนักศึกษาผู้เข้าสอบคนละกลุ่ม คือ 1) กลุ่มที่สอบในชุดวิชาที่เปิดสอบในแผนการศึกษา ก1 และ ก2 (เป็นนักศึกษากลุ่มเดียวกัน) และ 2) กลุ่ม ที่สอบในชุดวิชาที่เปิดสอบในแผนการศึกษา ก1 และ ก3 (เป็นนักศึกษากลุ่มเดียวกัน) ทั้งนี้หมายความว่า นักศึกษากลุ่มที่สอบในชุดวิชาชุดเดียวกัน เช่น ชุดวิชา B ที่เปิดสอบในแผนการศึกษา ก1 และ ก2 จะสอบด้วย แบบทดสอบฉบับเดียวกัน แต่มีการให้คะแนนต่างกัน เนื่องจากแผนการศึกษา ก1 มีการคิดคะแนนจากการทำ แบบทดสอบที่มีเนื้อหาจากเอกสารการสอนในหน่วยที่ 1-15 มาออกสอบ จึงเป็นการให้คะแนนจาก แบบทดสอบเพียงอย่างเดียวหรือ 100 เปอร์เซ็นต์ แต่แผนการศึกษา ก2 มีคะแนนจากกิจกรรมการสอนเสริม 40 คะแนน ส่วนอีก 60 คะแนนจะเป็นคะแนนจากการสอบปลายภาค ส่วนแบบทดสอบจากแผนการศึกษา ก 1 และ ก3 จะมีการใช้ข้อสอบร่วมกันในการสอบปลายภาค (หน่วยที่ 8-15) แต่ข้อสอบจะแตกต่างกันในส่วน ของแบบทดสอบกลางภาคของแผนการศึกษา ก3 ที่ออกข้อสอบใหม่เป็นคนละฉบับกับแผนการศึกษา ก1 ใน หน่วยที่ 1-7

1. ความหมายของการปรับเทียบคะแนน

การปรับเทียบคะแนนมีนักวิชาการหลายท่านได้ให้นิยามไว้ดังนี้

ปีเตอร์เซนและคณะ (Petersen & other, 1982: 242) ให้นิยามของการปรับเทียบคะแนนระหว่าง แบบทดสอบว่าเป็นกระบวนการเชิงประจักษ์ที่ใช้ในการแปลงคะแนนจากแบบทดสอบฉบับหนึ่งไปยัง แบบทดสอบอีกฉบับหนึ่ง โดยที่แบบทดสอบทั้งสองนั้นต้องวัดคุณลักษณะเดียวกัน และสอดคล้องกับศิริชัย กาญจนวาสี (2541: 2) ซึ่งได้ให้นิยามว่า การปรับเทียบคะแนนระหว่างแบบทดสอบเป็นกระบวนการแปลง คะแนนของแบบทดสอบต่างฉบับที่มุ่งวัดคุณลักษณะเดียวกันให้เป็นคะแนนที่สมมูลกันเพื่อให้คะแนน สามารถเปรียบเทียบกันได้โดยตรง ส่วนลอร์ด (Lord, 1980 : 195) ได้กล่าวถึงการปรับเทียบคะแนน ระหว่างแบบทดสอบว่าเป็นการแปลงคะแนนจากแบบทดสอบต่างฉบับให้มีความหมายสับเปลี่ยนกันได้ และเพื่อความ เสมอภาคของบุคคลผู้รับการทดสอบ ในขณะที่แองกอฟ (Angoff, 1984: 85) ให้ความหมายของการปรับเทียบ คะแนนระหว่างแบบทดสอบว่า หมายถึง กระบวนการแปลงระบบคะแนนของแบบทดสอบฉบับหนึ่งไปสู่ระบบ คะแนนของแบบทดสอบ อีกฉบับหนึ่งซึ่งวัดสิ่งเดียวกัน เพื่อให้คะแนนแปลงจากแบบทดสอบทั้งสองฉบับนั้น สามารถเทียบเท่ากัน และใช้แทนกันได้

จากแนวคิดและการให้ความหมายดังกล่าวข้างต้นพอจะสรุปได้ว่า การปรับเทียบคะแนน เป็นวิธีการทางสถิติในการนำคะแนนจากแบบทดสอบต่างฉบับกันแต่วัดในสิ่งเดียวกันไปแปลงให้มีหน่วยวัดเดียวกันเพื่อให้คะแนนจากแบบสอบต่างฉบับกันสามารถเปรียบเทียบกันได้โดยตรง

2.รูปแบบวิธีการปรับเทียบคะแนน (Designs for Equating Method)

วิธีการที่ใช้ในการปรับเทียบคะแนนมีหลายวิธี แต่สามารถจำแนกตามแนวคิดพื้นฐานของการปรับเทียบได้เป็นรูปแบบหลัก 2 รูปแบบ (ศิริชัย กาญจนวาสี, 2541) ได้แก่

- 1) รูปแบบการปรับเทียบคะแนนโดยใช้ทฤษฎีดั้งเดิม (Classical Models of Equating)
- 2) รูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Equating through Item Response Theory)

ในที่นี้ผู้เขียนขอกล่าวถึงเฉพาะรูปแบบและวิธีการที่ผู้เขียนได้นำมาใช้ในการทำวิจัย คือ อิกวิเปอร์เซ็นต์ไทล์ (Equipercentile Equating) ซึ่งเป็นวิธีการที่อยู่ในรูปแบบการปรับเทียบคะแนนโดยใช้ทฤษฎีดั้งเดิม ดังรายละเอียดต่อไปนี้

2.1 การปรับเทียบคะแนนอิกวิเปอร์เซ็นต์ไทล์

ปรับเทียบคะแนนแบบอิกวิเปอร์เซ็นต์ไทล์ หรืออิกวิเปอร์เซ็นต์ไทล์ (Equipercentile equating) เป็นวิธีการปรับเทียบที่ยึดหลักการว่า การแจกแจงของคะแนนจากแบบสอบ X และแบบสอบ Y มีลักษณะคล้ายกัน การเทียบหาคะแนนสมมูลทำได้โดยใช้คะแนน ณ ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกันของคะแนน 2 ชุดนั้น ผลการปรับเทียบคะแนนแสดงได้ด้วยกราฟ ถ้าหากความยากของแบบสอบทั้ง 2 ชุดนั้นใกล้เคียงกัน เส้นกราฟจะใกล้เคียงกับเส้นตรง แต่ถ้าหากแบบสอบ 2 ชุดนั้น มีความยากแตกต่างกันเส้นกราฟจะเป็นเส้นโค้ง การนำวิธีการปรับเทียบวิธีนี้ไปใช้มีข้อแนะนำ ดังนี้

1. ควรใช้กลุ่มผู้สอบขนาดใหญ่ ควรเป็นกลุ่มที่มีความสามารถค่อนข้างกระจายและกระจายพอ ๆ กัน ถ้าใช้กลุ่มตัวอย่างขนาดเล็กจะทำให้คะแนนมีความไวต่อความแปรปรวน ผลการปรับเทียบคะแนนสามารถแปรผันไปตามกลุ่มผู้สอบได้

2. การสร้างกราฟเปอร์เซ็นต์ไทล์เพื่อปรับเทียบคะแนนระหว่างแบบสอบควรสร้างด้วยความระมัดระวังและไม่ลำเอียง

3. แบบสอบที่นำมาปรับเทียบคะแนนกัน ควรมีความเที่ยงใกล้เคียงกัน ถ้าต่างกันมากจะทำให้การปรับเทียบขาดความคงที่

4. การปรับเทียบคะแนนควรอยู่ในช่วงพิสัยของคะแนนสังเกต การปรับเทียบคะแนนที่อยู่นอกพิสัยของคะแนนสังเกตจะมีความคลาดเคลื่อนสูง (Angoff, 1984)

การปรับเทียบคะแนนอิกวิเปอร์เซ็นต์ไทล์ ถือว่าคะแนนมีความเท่าเทียมกัน เมื่อคะแนนของแต่ละฉบับมีตำแหน่งเปอร์เซ็นต์ไทล์เท่ากัน (Franagan, 1951) จากสูตร

$$X_{PR_k} = L_0 + \left[\frac{\frac{kn}{100} - F_b}{f} \right] (c)$$

โดยที่	X_{PR_k}	แทน คะแนนจากแบบทดสอบฉบับ X ตรงตำแหน่งเปอร์เซ็นต์ไทล์ที่ k
	L_0	แทน ขีดจำกัดล่างที่แท้จริงของคะแนนในชั้นเปอร์เซ็นต์ไทล์ที่ k
	n	แทน จำนวนสมาชิกในกลุ่มตัวอย่าง
	f	แทน ความถี่ของคะแนนในชั้นเปอร์เซ็นต์ไทล์ที่ k
	F_b	แทน ความถี่สะสมก่อนถึงคะแนนในชั้นเปอร์เซ็นต์ไทล์ที่ k
	c	แทน ความกว้างของชั้น

2.2 คุณภาพของการปรับเทียบคะแนน

คุณภาพของการปรับเทียบคะแนน เป็นการตรวจสอบความคลาดเคลื่อนที่เกิดขึ้นของการปรับเทียบคะแนนระหว่างแบบสอบ โดยแหล่งของความคลาดเคลื่อนมาจาก 2 แหล่ง คือ 1) มาจากความคลาดเคลื่อนอย่างสุ่มของการปรับเทียบ (Random equating error) มาจากการเลือกตัวอย่างจากประชากรมาใช้ในการประมาณค่าพารามิเตอร์ (เช่น ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน การจัดอันดับเปอร์เซ็นต์ไทล์ เป็นต้น) ซึ่งความคลาดเคลื่อนของการปรับเทียบคะแนน (Standard Error of Equating : SEE) เป็นดัชนีชี้ความคลาดเคลื่อนอย่างสุ่มของการปรับเทียบ เมื่อขนาดของกลุ่มตัวอย่างใหญ่ขึ้นความคลาดเคลื่อนของการปรับเทียบคะแนนก็จะน้อยลง จึงสามารถลดความคลาดเคลื่อนอย่างสุ่มของการปรับเทียบด้วยการกำหนดขนาดตัวอย่างให้ใหญ่ขึ้น และด้วยการใช้วิธีการออกแบบการปรับเทียบที่เหมาะสม 2) มาจากความคลาดเคลื่อนอย่างเป็นระบบของการปรับเทียบ (Systematic equating error) เป็นผลมาจากการหลีกเลี่ยงข้อดกลางเบื้องต้นหรือเงื่อนไขของการปรับเทียบ เช่น ในกรณีการออกแบบการปรับเทียบใช้กลุ่มสุ่ม (random groups design) แต่ปรากฏว่ากลุ่มตัวอย่างที่ได้มีความสามารถไม่เท่าเทียมกัน หรือการใช้ข้อสอบรวมที่ไม่ได้เป็นตัวแทนของแบบสอบทั้งฉบับ เป็นต้น (Kolen & Brennan, 2004)

การศึกษาคุณภาพของการปรับเทียบคะแนน จึงใช้การประเมินคุณภาพของการปรับเทียบ โดยการหาค่าความคลาดเคลื่อนของการปรับเทียบคะแนน หรือใช้การตรวจสอบความเพียงพอของการปรับเทียบคะแนน วิธีการตรวจสอบคุณภาพสามารถทำได้หลายวิธี ดังนี้

2.3 ความคลาดเคลื่อนของการปรับเทียบคะแนน

การประเมินความคลาดเคลื่อนของการปรับเทียบคะแนน เป็นการวิเคราะห์แหล่งความลำเอียง (bias) และความแปรปรวน (variability) ซึ่งในการปรับเทียบคะแนนทุกวิธี ไม่ว่าจะใช้วิธีการปรับคะแนนใดก็ตามจะมีความคลาดเคลื่อนมาตรฐานของการปรับเทียบ อยู่ด้วยเสมอ ซึ่งมีนักวิชาการหลายท่านได้นำเสนอสูตรตัวอย่างเช่น

ลอร์ด (Lord 1950 cited in Angoff 1984) สร้างสูตรการคำนวณความคลาดเคลื่อนมาตรฐานของการปรับเทียบในรูปแบบเชิงเส้นตรง เมื่อสุ่มกลุ่มตัวอย่างทำแบบสอบกลุ่มละชุด

$$SEE_{x^*}^2 = \frac{2S_x^2}{N_t} (Z_y^2 + 2)$$

เมื่อ S_x^2 คือ ความแปรปรวนของคะแนนในกลุ่มที่ทำแบบสอบ X

N_t คือ จำนวนคนที่ทำแบบสอบทั้ง 2 ฉบับ

$$Z_y = (Y - \mu_y) / \sigma_y$$

หรือจะหาจากสูตร

$$SD_E = \sqrt{\frac{\sum E_i^2}{n} - \left(\frac{\sum E_i}{n}\right)^2}$$

เมื่อ $E_i = |y_i - y_i^*|$ โดยที่

y_i = คะแนนคนที่ i จากแบบสอบ y

y_i^* = คะแนนคนที่ i จากการปรับเทียบคะแนนแบบสอบ x บนสเกลของแบบสอบ y

n = ขนาดกลุ่มตัวอย่างในการสอบทานผล

$$SEE = \frac{SD_E}{\sqrt{n}}$$

2.4 ความเพียงพอของการปรับเทียบ

เป็นการประเมินประสิทธิภาพของการปรับเทียบคะแนนซึ่งมีอยู่หลายวิธีเช่น วิธีดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ที่ใช้กลุ่มสอบทานในการหาคุณภาพการปรับเทียบคะแนน ดัชนีตรวจสอบความเพียงพอของเจเกอร์ ดัชนีความแตกต่าง (Discrepancy Index) (อุทัยวรรณ พงศ์อร่าม, 2545: 82 – 85) รายละเอียดดังนี้

ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ (The Percentile Comparison index) เป็นการวัดความไม่สอดคล้องระหว่างการแจกแจงของคะแนนในแบบสอบชุด X กับแบบสอบชุด Y ที่ได้แปลงไปอยู่ในมาตรคะแนนของ X แล้ว ดัชนีเปรียบเทียบเปอร์เซ็นต์ไทล์ คือ ค่าเฉลี่ยกำลังสองของความแตกต่าง ที่ได้จากการแจกแจงของคะแนนต่าง ๆ ของคะแนน X กับ คะแนนแปลง X ที่แปลงมาจาก Y ณ ตำแหน่งเปอร์เซ็นต์ไทล์เดียวกันที่ใช้ข้อมูลจากกลุ่มตัวอย่างสอบทานผล สูตรที่ได้จากการคำนวณคือ

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

เมื่อ n คือ จำนวนคนของกลุ่มสอบทานผล

K คือ จำนวนข้อสอบรวม

X คือ คะแนนจากแบบสอบชุด X ซึ่งเป็นคะแนนเกณฑ์

X^* คือ คะแนนจากแบบสอบชุด Y ที่แปลงให้อยู่ในมาตรคะแนนของ X ณ ตำแหน่ง เปอร์เซ็นต์ไทล์เดียวกันดัชนีความแตกต่าง (Discrepancy Index) ของปีเตอร์สัน มาร์โค และสตีเวอร์ท (Pertersen, Marco and Stevart, 1982 อ้างถึงใน ศิริชัย กาญจนวาสี, 2556) เป็นค่าความคลาดเคลื่อนรวมในการปรับเทียบคะแนน โดยถ่วงน้ำหนักด้วยค่าความแปรปรวน จากสูตร

$$C = \frac{\sum (y_i - y_i^*)^2}{nS_y^2} \text{ เมื่อ}$$

y_i = คะแนนจากแบบสอบชุด Y ของคนที่ i ซึ่งเป็นคะแนนเกณฑ์

y_i^* = คะแนนปรับเทียบจากแบบสอบชุด X ที่อยู่บนสเกลของ Y สำหรับคนที่ i

n = จำนวนคนสอบทานผล

S_y^2 = ความแปรปรวนของคะแนนเกณฑ์

3. การประยุกต์ใช้โปรแกรม jamovi ในการปรับเทียบคะแนนแบบอิควิเปอร์เซ็นต์ไทล์

3.1 ผู้วิจัยได้รับข้อมูลคะแนนกิจกรรม คะแนนสอบและผลการศึกษาของนักศึกษาที่เข้าสอบในชุดวิชาที่ทำการวิจัย ในภาคเรียนที่ 1 ปีการศึกษา 2565 ของแผนการศึกษา ก1 ก2 และ ก3 โดยจำแนกตามชุดวิชาที่เปิดสอบตามแผนการศึกษา ก1 และ ก2 กับ ชุดวิชาที่เปิดสอบตามแผนการศึกษา ก1 และ ก3 แล้วจึงทำการตรวจสอบว่าข้อมูลที่ได้รับมีความครบถ้วน ถูกต้อง สมบูรณ์ หรือไม่

3.2 ผู้วิจัยทำการพิจารณาคัดเลือกข้อมูลคะแนนกิจกรรม คะแนนสอบและผลการศึกษาของนักศึกษาที่เข้าสอบในชุดวิชาที่ทำการวิจัย ในภาคเรียนที่ 1 ปีการศึกษา 2565 ของแผนการศึกษา ก1 ก2 และ ก3 โดยจำแนกตามชุดวิชาที่เปิดสอบตามแผนการศึกษา ก1 และ ก2 กับ ชุดวิชาที่เปิดสอบตามแผนการศึกษา ก1 และ ก3 โดยพิจารณาเฉพาะชุดวิชาตามเกณฑ์

3.3 ผู้วิจัยนำข้อมูลที่ได้จากข้อ 3.2 มาจัดเตรียมข้อมูลให้อยู่ในรูปแบบไฟล์ Excel เพื่อใช้ในการวิเคราะห์ปรับเทียบคะแนนรวมทั้งหมดของแผนการศึกษา ก1 และ ก2 กับ ก1 และ ก3 ด้วยสถิติการปรับเทียบคะแนนอิควิเปอร์เซ็นต์ไทล์ (Equipercentile Equating) โดยใช้โปรแกรม jamovi

3.4 ผู้วิจัยวิเคราะห์คะแนนรวมทั้งหมดที่ได้จากคะแนนสอบและคะแนนกิจกรรมทั้งหมดของแผนการศึกษา ก1 ก2 และ ก3 โดยวิเคราะห์แยกเป็นรายชุดวิชาที่เปิดสอบพร้อมกันตามแผนการศึกษา ก1 และ ก2 กับ ก1 และ ก3 โดยเลือกชุดวิชาที่มีจำนวนนักศึกษาเข้าสอบมากที่สุดแผนการศึกษาละ 3 ชุดวิชา รวม 6 ชุดวิชา และพิจารณาเฉพาะชุดวิชาที่ผ่านตามเกณฑ์ที่ผู้วิจัยกำหนด จากนั้นจึงวิเคราะห์ด้วยสถิติ Equipercentile Equating โดยใช้โปรแกรม Jamovi

4. ตัวอย่างที่ได้จากผลการวิเคราะห์ด้วยโปรแกรม Jamovi

4.1 ผลการวิเคราะห์การปรับเทียบคะแนนที่ได้มาจากคะแนนการสอบและคะแนนที่ได้จากการทำกิจกรรมทั้งหมดของแผนการศึกษา ก1 และ ก2

ผลการปรับเทียบคะแนนรวมทั้งหมดที่ได้มาจากคะแนนการสอบและคะแนนที่ได้จากการทำกิจกรรมทั้งหมดของแผนการศึกษา ก1 และ ก2 ของชุดวิชา A สาขาวิชาศิลปศาสตร์ ด้วยสถิติอิควิเปอร์เซ็นต์ไทล์ (Equipercentile Equating) โดยในคอลัมน์ที่ 1 และ คอลัมน์ที่ 3 เป็นคะแนนก่อนการปรับเทียบจากแผนการศึกษา ก2 ไปยังคะแนนที่ได้จากแผนการศึกษา ก1 พบว่า คะแนนในช่วงระดับที่ต่ำสุด 3 อันดับแรก คือ ค่าคะแนนที่ได้ในแผนการศึกษา ก2 เท่ากับ 21 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 (ในคอลัมน์ที่ 2 ของตาราง) จะได้คะแนนเท่ากับ 16 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 16

คะแนนเท่ากัน, คะแนนที่ได้ในแผนการศึกษา ก2 เท่ากับ 38 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 จะได้คะแนนเท่ากับ 18 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 18 คะแนนเท่ากัน และคะแนนที่ได้ในแผนการศึกษา ก2 เท่ากับ 52 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 จะได้คะแนนเท่ากับ 29.2 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 29 คะแนน เป็นต้น

ส่วนคะแนนใน 3 อันดับสูงสุด คือ ค่าคะแนนที่ได้ในแผนการศึกษา ก2 เท่ากับ 71 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 (ในคอลัมน์ที่ 2 ของตาราง) จะได้คะแนนเท่ากับ 46.7 ซึ่งเทียบกับคะแนนที่ได้

จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 47 คะแนน คะแนนที่ได้ในแผนการศึกษา ก2 เท่ากับ 75 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 จะได้คะแนนเท่ากับ 52 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 52 คะแนนและคะแนนที่ได้ในแผนการศึกษา ก2 เท่ากับ 82 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 จะได้คะแนนเท่ากับ 57.3 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 57 คะแนน เป็นต้น

จากผลการปรับเทียบคะแนน เห็นได้ว่าให้ผลคะแนนที่สอดคล้องกับลักษณะของแบบทดสอบจากแผนการศึกษา ก1 และ ก2 ที่มหาวิทยาลัยสุโขทัยธรรมาธิราช ได้กำหนดไว้ว่า แบบทดสอบจากแผนการศึกษา ก1 และ ก2 เป็นแบบทดสอบฉบับเดียวกัน แต่มีการให้คะแนนต่างกัน โดยเมื่อคิดคะแนนเป็นรายข้อจากแผนการศึกษา ก2 จะมีการให้คะแนนต่อข้อมากกว่า จึงทำให้ผลคะแนนที่มีค่าต่ำ เมื่อนำมาเปรียบเทียบกับผลคะแนนที่มีค่าสูงกว่าแล้ว ผลที่ได้จะมีการปรับค่าคะแนนให้ใกล้เคียงกัน หรือแตกต่างกันไม่มาก และผลคะแนนที่มีค่าสูงมาก เมื่อนำมาปรับเทียบกับคะแนนจากแผนการศึกษา ก1 แล้วจะมีคะแนนไม่แตกต่างกันมาก ผลการปรับเทียบคะแนน ได้คะแนนสมมูลดังตารางที่ 4.1 (วินิตา แก้วเกื้อ และ สฤษฎ์เทพ สุขแก้ว, 2567)

ตารางที่ 4.1 ผลการปรับเทียบคะแนนรวมที่ได้มาจากคะแนนการสอบและคะแนนที่ได้จากการทำกิจกรรมทั้งหมดของชุดวิชา A สาขาวิชาศิลปศาสตร์ ของแผนการศึกษา ก1 และ ก2

ความสอดคล้องของคะแนน (Concordance)			ความสอดคล้องของคะแนน (Concordance)		
คะแนนจาก แผนการ ศึกษา ก2	คะแนนจาก แผนการศึกษา ก2	คะแนนจาก แผนการศึกษา ก2 ที่ปรับเทียบกับ คะแนนจาก แผนการศึกษา ก1	คะแนนจาก แผนการศึกษา ก1	คะแนนจาก แผนการศึกษา ก2	คะแนนจาก แผนการศึกษา ก2 ที่ปรับเทียบกับ คะแนนจาก แผนการศึกษา ก1
21	16	16	62	40.3	39
38	18	18	62	40.7	40
52	29.2	29	64	42	41
53	30	30	65	43.7	42
58	30.6	32	70	45.5	45
59	31.7	33	71	45.7	46
60	34.6	34	71	46.7	47
61	35	35	75	52	52
			82	57.3	57

4.2 ผลการวิเคราะห์การปรับเทียบคะแนนรวมที่ได้มาจากคะแนนการสอบและคะแนนที่ได้จากการทำกิจกรรมทั้งหมดของชุดวิชา F สาขาวิชารัฐศาสตร์ ของแผนการศึกษา ก1 และ ก3

ผลการวิเคราะห์การปรับเทียบคะแนนรวมที่ได้มาจากคะแนนการสอบและคะแนนที่ได้จากการทำกิจกรรมทั้งหมดของชุดวิชา F สาขาวิชารัฐศาสตร์ ในแผนการศึกษา ก1 และ ก3 โดยในคอลัมน์ที่ 1 และ

คอลัมน์ที่ 3 เป็นคะแนนก่อนการปรับเทียบจากแผนการศึกษา ก3 ไปยังคะแนนที่ได้จากแผนการศึกษา ก1 พบว่า คะแนนในช่วงระดับที่ต่ำสุด 3 อันดับแรก คือ ค่าคะแนนที่ได้ในแผนการศึกษา ก3 เท่ากับ 15 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 (ในคอลัมน์ที่ 2 ของตาราง) จะได้คะแนนเท่ากับ 27.5 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 21 คะแนน คะแนนที่ได้ในแผนการศึกษา ก3 เท่ากับ 18 (ลำดับที่ 1) เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 จะได้คะแนนเท่ากับ 28.3 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 23 คะแนน และคะแนนที่ได้ในแผนการศึกษา ก3 เท่ากับ 29 (ลำดับที่ 2) เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 จะได้คะแนนเท่ากับ 30 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 26 คะแนน เป็นต้น

ส่วนคะแนนใน 3 อันดับสูงสุด คือ ค่าคะแนนที่ได้ในแผนการศึกษา ก3 เท่ากับ 76 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 (ในคอลัมน์ที่ 2 ของตาราง) จะได้คะแนนเท่ากับ 54.2 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 54 คะแนน ถือว่ามีค่าใกล้เคียงกัน คะแนนที่ได้ในแผนการศึกษา ก3 เท่ากับ 78 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 จะได้คะแนนเท่ากับ 55 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 55 คะแนนเท่ากัน และคะแนนที่ได้ในแผนการศึกษา ก3 เท่ากับ 85 เมื่อนำมาปรับเทียบคะแนนกับแผนการศึกษา ก1 จะได้คะแนนเท่ากับ 76.7 ซึ่งเทียบกับคะแนนที่ได้จากแผนการศึกษา ก1 (ในคอลัมน์ที่ 3) ได้ 77 คะแนน ถือว่ามีค่าใกล้เคียงกัน เป็นต้น

จากผลการปรับเทียบคะแนนจะเห็นได้ว่าคะแนนจากแบบทดสอบแผนการศึกษา ก1 และ ก3 มีลักษณะแตกต่างกันค่อนข้างชัดเจน ในบางช่วงของการเปรียบเทียบ และมีคะแนนบางช่วงที่เมื่อทำการปรับเทียบแล้วทำให้คะแนนมีความใกล้เคียงกัน เนื่องจากแบบทดสอบจากแผนการศึกษา ก1 และ ก3 จะมีการใช้ข้อสอบร่วมกันในการสอบปลายภาค(หน่วยที่ 8-15) แต่ข้อสอบจะแตกต่างกันในส่วนองแบบทดสอบกลางภาคของแผนการศึกษา ก3 ที่ออกข้อสอบจากหน่วยที่ 1-7 แต่แบบทดสอบของแผนการศึกษา ก1 ในหน่วยที่ 1-7 จะเป็นข้อสอบอีกชุดหนึ่งที่ใช้ในการสอบปลายภาคร่วมกับหน่วยที่ 8-15 รวมเป็น 15 หน่วย ผลการปรับเทียบคะแนน ได้คะแนนสมมูลดังตารางที่ 4.2 (วินิตา แก้วเกื้อ และ สฤกษ์เทพสุขแก้ว, 2567)

ตารางที่ 4.2 ผลการวิเคราะห์เปรียบเทียบคะแนนรวมที่ได้มาจากคะแนนการสอบและคะแนนที่ได้จากการทำกิจกรรมทั้งหมดของชุดวิชา F สาขาวิชารัฐศาสตร์ ของแผนการศึกษา ก1 และ ก3

ความสอดคล้องของคะแนน (Concordance)			ความสอดคล้องของคะแนน (Concordance)		
คะแนนจาก แผนการ ศึกษา ก3	คะแนนจาก แผนการศึกษา ก3	คะแนนจาก แผนการศึกษา ก3 ที่ปรับเทียบกับ คะแนนจาก แผนการศึกษา ก1	คะแนนจาก แผนการศึกษา ก1	คะแนนจาก แผนการศึกษา ก3	คะแนนจาก แผนการศึกษา ก3 ที่ปรับเทียบกับ คะแนนจาก แผนการศึกษา ก1
15	27.5	21	62	36.3	36
18	28.3	23	66	36.7	37
29	30	26	67	39.2	39
33	30.4	28	68	42.5	43
37	30.8	28	68	43.3	43
42	31.3	28	69	44.2	45
47	31.7	30	70	45	45
54	32.1	31	71	45.4	46
56	33.1	31	73	46.7	47
57	33.6	32	76	54.2	54
57	34.2	32	78	55	55
59	34.6	35	85	76.7	77
61	35	35	66	36.7	37
			67	39.2	39

สรุป

คนส่วนใหญ่อาจมีความคุ้นเคยกับการใช้อันดับเปอร์เซ็นต์ไทล์ที่มักใช้ในการเปรียบเทียบประสิทธิภาพของบุคคลกับกลุ่มบรรทัดฐาน (norm group) โดยให้อันดับที่สามารถนำมาเปรียบเทียบที่ดีความได้ง่าย (Kolen & Brennan, 2014) อย่างไรก็ตาม สิ่งสำคัญคือต้องพิจารณาข้อจำกัดของวิธีการเหล่านี้ เช่น ในการนำคะแนนเฉลี่ยและค่าเบี่ยงเบนมาตรฐานมาใช้ในการเปรียบเทียบ สารสนเทศที่จะได้จากข้อมูลเหล่านี้คือ แนวโน้มเข้าสู่ส่วนกลาง (Central Tendency) และการกระจายของคะแนน (dispersion of scores) แต่ค่าเหล่านี้ไม่ได้คำนึงถึงความเบ้หรือความเบ้ที่อาจเกิดขึ้นในการกระจายคะแนน (Alordiah, 2015; Thompson & Levitov, 1985) นอกจากนี้ อันดับเปอร์เซ็นต์ไทล์ แม้จะมีประโยชน์ในการเปรียบเทียบ แต่ก็ไม่ได้ให้ข้อมูลเกี่ยวกับระดับสัมบูรณ์ของประสิทธิภาพ (Temitope, 2021; Wright & Masters, 1982)

ผู้กำหนดนโยบายและนักการศึกษาอาจชอบการเทียบ CTT เปอร์เซ็นต์ไทล์หรืออควิเปอร์เซ็นต์ไทล์ เท่ากันมากกว่าเนื่องจากมีความโปร่งใสและมีความสัมพันธ์โดยตรงกับอันดับเปอร์เซ็นต์ไทล์ ซึ่งมักใช้ในกระบวนการตัดสินใจทางการศึกษา ทั้งนี้วิธีการเทียบเคียงที่เชื่อถือได้ควรให้คะแนนที่สอดคล้องกันในการบริหารการทดสอบที่แตกต่างกัน (Oladele, 2021; Crocker & Algina, 1986)

บรรณานุกรม

- วินิตา แก้วแก้ว และ สฤณีเทพ สุขแก้ว. (2567). การศึกษาเปรียบเทียบคุณภาพของรูปแบบการวัดผลและประเมินผลตามแผนการศึกษาของมหาวิทยาลัยสุโขทัยธรรมมาธิราช. รายงานการวิจัย. นนทบุรี: มหาวิทยาลัยสุโขทัยธรรมมาธิราช.
- ศิริชัย กาญจนวาสี. (2541). *การปรับเทียบคะแนนระหว่างแบบสอบ (Test Equating)*. กรุงเทพฯ: ศูนย์ตำราและเอกสารทางวิชาการ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2556). *ทฤษฎีการทดสอบแบบดั้งเดิม*. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- อุทัยวรรณ พงศ์อร่าม.(2545).*การศึกษาขนาดกลุ่มตัวอย่างที่เหมาะสมสำหรับการปรับเทียบคะแนนด้วยวิธีอิควิเปอร์เซ็นไทล์ และวิธีเชิงเส้นตรงตามแบบจำลองคะแนนจริงสัมพันธ์ ที่มีแบบแผนการปรับเทียบและความยาวของแบบสอบแตกต่างกัน*.(ปริญญาณิพนธ์คุชฎิบัณฑิตไม่ได้ตีพิมพ์).มหาวิทยาลัยศรีนครินทรวิโรฒประสานมิตร, กรุงเทพมหานคร.
- Alordiah, C. (2015). Comparison of index of Differential Item functioning under the methods of Item Response theory and classical testtheory in Mathematics. *An unpublished Ph. D thesis of Delta State University, Abraka, Delta State, Nigeria.*
- Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service.
- Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Holt,Rinehart and Winston.
- Daramola, D. S., Owolabi, H. O., & Olutola, A., (2018). Equipercetile Equating of WAEC, NECO And NABTEB Senior School Certificate Multiple Choice Test Items In Economics. *Journal of Education in Developing Areas*, 26(2), 658-668.
- Flanagan, T.C (1951). Unit, Scores and Normsin E.F. Lindquist (Ed.) Educational Measurement. Washington D.C. : American Council on Education, P. 695-763
- Kolen, M. J., & Brennan, R. L. (2014). Test equating, scaling, and linking: Methods and practices. Springer.
- Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Oladele, B. K. (2021). Comparison Of Secondary School Students'ability In Mathematics Constructed-Response Items Under Classical Test And Item Response Measurement Theories In The Ibadan Metropolis, Nigeria (Doctoral dissertation).
- Oguoma, C.C.(2017).*Assessment Of The Equivalence Of Waec And Neco Mathematics Multiplechoice Tests Using Item Response Theory.*

- Petersen, N.S., Marco, C.L., and Stewart, E.E. (1982). *A test of the adequacy of linear Score Problems*. Hillssdale, N.J. : Erlbaun.
- Rossiter, J., Abreh, M. K., Ali, A., & Sandefur, J. (2023). The High Stakes of Bad Exams. *Journal of Human Resources*.
- Temitope, B. (2021). Determination Of The Equivalence Of Waec And Neco Ssce Chemistry Items Using Linear Equating Approaches Of Classical Test Theory And Item Response Theory. *Bulgarian Journal Of Science & Education Policy*, 15(1).
- Thompson, B., & Levitov, J. E. (1985). Using microcomputers to score and interpret psychological tests. *Journal of Counseling Psychology*, 32(3), 471-477.
- Wiberg, M.(2021).On the use of different linkage plans with different observed-score equipercentile equating methods. *Practical Assessment, Research, and Evaluation*, 26(1), 23.
- Wright, B. D., & Masters, G. N.(1982). *Rating scale analysis*. Mesa Press.

