

ความนำ

การปรับเทียบคะแนนระหว่างแบบสอบ เป็นกระบวนการแปลงคะแนนของแบบสอบต่างฉบับ ที่มุ่งวัดคุณลักษณะเดียวกันให้เป็นคะแนนที่สมมูลกัน เพื่อให้คะแนนสามารถเปรียบเทียบกันได้โดยตรง จำแนกได้เป็น 2 รูปแบบ ได้แก่ (ศิริชัย กาญจนวาสี, 2545)

1. การปรับเทียบคะแนนตามแนวนอนหรือแนวนอน (Horizontal Equating) เป็นการปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน มีระดับความยากใกล้เคียงกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ในประชากรเดียวกัน หรือมีความสามารถใกล้เคียงกัน เป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบทดสอบเนื้อหาเดียวกันขึ้นมาหลายฉบับ ด้วยเหตุผลของการนำไปใช้ในการทดสอบเพื่อให้เกิดความยุติธรรม และป้องกันความลับของข้อสอบเมื่อใช้เวลาต่างกัน สำหรับผู้สอบกลุ่มขนาดใหญ่ เพื่อปรับเทียบว่าคะแนนที่ได้จากฉบับหนึ่งเทียบเป็นเท่าไรของอีกฉบับหนึ่ง ซึ่งวัดในระดับเดียวกัน จึงเป็นการปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบระดับชั้นเดียวกัน

2. การปรับเทียบคะแนนตามแนวตั้งหรือแนวตั้ง (Vertical Equating) เป็นการปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน แต่มีระดับความยากแตกต่างกันและกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ต่างประชากรกัน หรือมีความสามารถแตกต่างกัน เป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบสอบเนื้อหาเดียวกัน แต่ต่างฉบับต่างมุ่งวัดความสามารถของผู้สอบที่ต่างระดับกัน เพื่อปรับเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่งเทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน จึงเป็นการปรับเทียบคะแนนระหว่างแบบสอบต่างระดับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบต่างระดับชั้นกัน แบบสอบต่างฉบับที่วัดเนื้อเรื่องเดียวกัน แต่ต่างระดับกันที่ต้องการนำมาปรับเทียบคะแนนกันนั้น แบบสอบแต่ละฉบับจึงมีความยากแตกต่างกันไป รวมทั้งกลุ่มผู้สอบแต่ละฉบับมีการแจกแจงความสามารถอยู่ต่างประชากร หรือมีการแจกแจงความสามารถที่อยู่ในระดับที่แตกต่างกัน ดังนั้นการปรับเทียบคะแนนตามแนวตั้ง จึงมีความสลับซับซ้อนกว่าการปรับเทียบคะแนนตามแนวนอนทั้งในแง่ทฤษฎีและการปฏิบัติ

การปรับเทียบคะแนนไม่ว่าจะเป็นการปรับเทียบคะแนนตามแนวนอนหรือตามแนวตั้งเป็นการกำหนดเงื่อนไขและสร้างกฎการแปลงคะแนนโดยยึดทฤษฎีของการวัดผล ดังนั้นถ้าแบ่งการปรับเทียบคะแนนตามทฤษฎีที่สนับสนุนแล้วจะแบ่งได้ 2 กลุ่มใหญ่ ๆ คือ (Flanagan, 1951)

1. การปรับเทียบคะแนนโดยยึดตามทฤษฎีคลาสสิกอลโมเดล (Classical Methods of Equating) ซึ่ง Angoff (1971) ได้กล่าวว่ามี 3 รูปแบบ ได้แก่ การปรับเทียบคะแนนรูปแบบเชิงเส้นตรง (Linear Equating) การปรับเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นไทล์ (Equipercentile Equating) และการปรับเทียบคะแนนโดยใช้สมการถดถอย (Regression Method)

2. การปรับเทียบคะแนนโดยยึดตามทฤษฎีตอบสนองข้อสอบ (Equating through Item Response Theory: IRT) ซึ่งนิยมกัน 2 รูปแบบ คือ การปรับเทียบคะแนนรูปแบบหนึ่งพารามิเตอร์ และการปรับเทียบคะแนนรูปแบบสามพารามิเตอร์

จากการศึกษาแนวคิดเชิงทฤษฎีและนิยามของการปรับเทียบคะแนนพอสรุปได้ว่า การปรับเทียบคะแนนที่ใช้ทฤษฎีคลาสสิกอลโมเดลที่ยังมีการใช้กันอยู่ คือ รูปแบบเชิงเส้นตรงและรูปแบบอิกวิเปอร์เซ็นไทล์ ซึ่งทั้งสองรูปแบบถูกพัฒนาขึ้นมาพร้อม ๆ กัน การปรับเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นไทล์เป็นรูปแบบที่ให้ความสะดวกในการหาคะแนนแปลง และมีความเหมาะสมในการแปลงคะแนนอย่างมาก เมื่อการแจกแจงคะแนนของแบบทดสอบต่างฉบับ

ที่นำมาเปรียบเทียบคะแนนนั้นรูปร่างเหมือนกัน ถ้าหากรูปร่างของการแจกแจงคะแนนจากแบบทดสอบต่างฉบับคล้ายคลึงกัน ควรใช้การเปรียบเทียบคะแนนรูปแบบเชิงเส้นตรง ซึ่งจะให้ผลดีกว่าการเปรียบเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นไทล์ เพราะเป็นปรนัยมากกว่าและไม่มี ความคลาดเคลื่อนในการปรับหรือเกลากกราฟให้เรียบตามวิธีการเปรียบเทียบคะแนนรูปแบบอิกวิเปอร์เซ็นไทล์ซึ่งจะมีความคลาดเคลื่อนมากในช่วงที่มีข้อมูลน้อย (Angoff, 1971) ส่วนการเปรียบเทียบคะแนนโดยอิงทฤษฎีตอบสนองข้อสอบนั้น เป็นรูปแบบใหม่ล่าสุดที่พัฒนาตามทฤษฎีการวัดความสามารถแฝง (Latent Trait Theory) ซึ่งมีคุณสมบัติของความไม่แปรเปลี่ยนค่าพารามิเตอร์ของข้อสอบเมื่อเปลี่ยนกลุ่มของผู้สอบ ดังนั้นจึงไม่มีนิยามเกี่ยวกับการแจกแจงคะแนนที่ได้จากแบบทดสอบ การเปรียบเทียบคะแนนที่ยึดตามทฤษฎีนี้ จึงมีผลดีแม้ว่าการแจกแจงคะแนนจากแบบทดสอบต่างฉบับจะแตกต่างกันก็ตาม (Kolen & Whitney, 1982) นอกจากนี้แล้วการเปรียบเทียบคะแนนที่ยึดตามทฤษฎีตอบสนองข้อสอบจะไม่ได้รับ ความกระทบกระเทือนจากแบบทดสอบที่มีความยากต่างกันและระดับความสามารถของกลุ่มตัวอย่างที่ต่างกัน (Lord, 1980) สำหรับการเปรียบเทียบคะแนนโดยยึดทฤษฎีตอบสนองข้อสอบที่นิยม มี 2 รูปแบบ ได้แก่ รูปแบบหนึ่งพารามิเตอร์หรือโมเดลของราสช์ (Rasch Model) เหมาะกับกลุ่มตัวอย่างที่มีขนาดเล็ก ใช้เวลาคำนวณน้อย และสามารถวิเคราะห์ได้ด้วยตนเองโดยไม่ต้องอาศัยเครื่องคอมพิวเตอร์ ส่วนการเปรียบเทียบคะแนนรูปแบบสามพารามิเตอร์นั้นได้เพิ่มค่าพารามิเตอร์การเดาข้อสอบเข้าไป ย่อมทำให้เกิดความแม่นยำในการเปรียบเทียบคะแนนมากขึ้น

จากสภาพปัญหาการที่ไม่สามารถเปรียบเทียบคะแนนหรือเปรียบเทียบระดับความสามารถระหว่างบุคคลจากคะแนนที่ได้จากการวัดโดยใช้แบบทดสอบที่เทียบเคียงกันหลาย ๆ ฉบับในสถานการณ์การสอบเพื่อตัดสินผลการเรียน การสอบแข่งขันเพื่อเข้าศึกษาต่อในสถาบันการศึกษาต่าง ๆ หรือสอบแข่งขันเพื่อเข้าทำงาน ตลอดจนปัญหาในการเปรียบเทียบคะแนนผลสัมฤทธิ์ทางการเรียนของนักเรียนในปีการศึกษาต่าง ๆ ซึ่งต้องประสบปัญหาความยุ่งยากในการสร้างแบบทดสอบคู่ขนานที่มีข้อจำกัดมากในทางปฏิบัติดังที่ได้กล่าวมาแล้วข้างต้น แนวทางแก้ปัญหาที่ง่ายและสะดวก สามารถทำได้โดยการใช้สถิติมาปรับเทียบคะแนนให้เกิดการเท่าเทียมกันคือ การปรับเทียบคะแนนตามแนวตั้ง (Vertical Equating) ซึ่งมีหลายรูปแบบตามทฤษฎีที่สนับสนุน จึงเกิดปัญหาว่าการปรับเทียบคะแนนแบบทดสอบรูปแบบใดที่ให้ผลได้ถูกต้องและแม่นยำที่สุดหรือโอกาสที่จะเกิดความคลาดเคลื่อนมีน้อยที่สุด ดังนั้นถ้าหากมีงานวิจัยที่มีข้อสรุปที่เชื่อถือ ได้ว่า การปรับเทียบคะแนนรูปแบบใดมีความคลาดเคลื่อนของการปรับเทียบคะแนนต่ำและมีความถูกต้องแม่นยำสูง ก็จะเป็นประโยชน์ต่อการนำวิธีการปรับเทียบคะแนนไปใช้ในการปฏิบัติ จากปัญหานี้ผู้วิจัยจึงมีความสนใจที่จะศึกษาประสิทธิภาพของการปรับเทียบคะแนนในเชิงเปรียบเทียบผลการปรับเทียบคะแนนตามแนวตั้ง ระหว่างรูปแบบที่ยึดตามทฤษฎีคลาสสิกโมเดล ได้แก่ รูปแบบเชิงเส้นตรงกับการปรับเทียบคะแนนที่ยึดตามทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ ทั้งนี้ผลที่ได้จากการวิจัยจะเป็นข้อมูลในการพิจารณาตัดสินใจเลือกวิธีการปรับเทียบคะแนนที่มีความถูกต้องและเหมาะสมที่สุดในการนำไปใช้เพื่อให้เกิดความเท่าเทียมกันและความยุติธรรมแก่ผู้สอบ ซึ่งจะเป็นการเพิ่มทางเลือกสำหรับผู้ที่เกี่ยวข้องกับงานทางด้านการวัดและประเมินผลการศึกษา อีกทั้งจะได้สารสนเทศที่เกี่ยวกับการพัฒนาองค์ความรู้ทางการปรับเทียบคะแนนต่อไป

วัตถุประสงค์ของการวิจัย

การศึกษานี้มีจุดมุ่งหมายเพื่อเปรียบเทียบคุณภาพของการปรับเทียบคะแนนแนวตั้งโดยใช้แบบทดสอบร่วมภายในเป็นการเปรียบเทียบระหว่างวิธีการปรับเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ ดังนี้

1. เพื่อเปรียบเทียบค่าความคลาดเคลื่อนของการปรับเทียบคะแนนแนวตั้งโดยใช้แบบทดสอบร่วมภายในระหว่างวิธีการปรับเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์

2. เพื่อเปรียบเทียบค่าความเพียงพอของการปรับเทียบคะแนนแนวตั้งโดยใช้แบบทดสอบร่วมภายใน ระหว่างวิธีการปรับเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์

สมมติฐานการวิจัย

1. ค่าความคลาดเคลื่อนของการปรับเทียบคะแนนแนวตั้ง โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าน้อยกว่าวิธีการปรับเทียบเชิงเส้นตรง

2. ค่าความเพียงพอของการปรับเทียบคะแนนแนวตั้งสำหรับผู้สอบที่มีความสามารถต่างกัน โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าต่ำกว่าวิธีการปรับเทียบเชิงเส้นตรง

นิยามศัพท์เฉพาะ

ค่าความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard Error of Equating) หมายถึง ความคลาดเคลื่อนมาตรฐานของค่าเฉลี่ยของความแตกต่างระหว่างคะแนนสมมูลซึ่งเป็นคะแนนจากแบบสอบฉบับ X ที่แปลงคะแนนจากตารางปรับเทียบคะแนนไปเป็นคะแนนจากแบบสอบฉบับ Y ของนักเรียนคนที่ i กับคะแนนเกณฑ์ซึ่งเป็นคะแนนจากแบบสอบฉบับ Y ที่นักเรียนคนที่ i สอบได้จริงของนักเรียนกลุ่มสอบทานผล

ค่าความเพียงพอของการปรับเทียบคะแนน (Adequacy of Equating) หมายถึง ค่าที่แสดงความแตกต่างของการปรับเทียบคะแนนจากแบบทดสอบฉบับหนึ่งไปสู่แบบทดสอบอีกฉบับหนึ่งตามวิธีการปรับเทียบคะแนน การตัดสินใจว่าค่าความเพียงพอของการปรับเทียบคะแนนใช้การเปรียบเทียบกับดัชนีความแตกต่าง

วิธีดำเนินการวิจัย

กลุ่มตัวอย่าง

กลุ่มตัวอย่าง เป็นนักเรียนระดับช่วงชั้นที่ 3 ปีการศึกษา 2552 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษา กำแพงเพชร เขต 1 จำนวน 1,080 คน ซึ่งได้โดยการสุ่มแบบหลายขั้นตอน (Multi-stage sampling) แบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่ม คือ กลุ่มที่ 1 กลุ่มตัวอย่างปรับเทียบคะแนน Kolen and Brennan (1987) แนะนำว่ากลุ่มตัวอย่างที่ใช้ในการปรับเทียบคะแนนไม่ควรต่ำกว่า 400 คน เพื่อให้งานวิจัยมีความแข็งแกร่งของข้อมูลและลดความคลาดเคลื่อนที่เกิดขึ้น ผู้วิจัยจึงใช้กลุ่มตัวอย่างในการปรับเทียบคะแนนระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 300 คน ระดับชั้นมัธยมศึกษาปีที่ 2 จำนวน 300 คน และระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 300 คน รวมกลุ่มตัวอย่างที่ใช้ในการปรับเทียบคะแนนจำนวน 900 คน กลุ่มที่ 2 กลุ่มตัวอย่างสอบทานผล ซึ่ง Kolen and Whitney (1982) ได้เสนอว่า กลุ่มสอบทานผลควรมีขนาดไม่น้อยกว่า 10 เปอร์เซนต์ ของกลุ่มตัวอย่างปรับเทียบคะแนน ดังนั้นเพื่อให้งานวิจัยมีความแข็งแกร่งของข้อมูลและลดความคลาดเคลื่อนที่เกิดขึ้น ผู้วิจัยจึงใช้กลุ่มสอบทานผลประมาณ 20 เปอร์เซนต์ ของกลุ่มปรับเทียบคะแนน ทำให้ได้กลุ่มสอบทานผลในระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 60 คน ระดับชั้นมัธยมศึกษาปีที่ 2 จำนวน 60 คน และระดับชั้นมัธยมศึกษาปีที่ 3 จำนวน 60 คน รวมกลุ่มตัวอย่างที่ใช้ในการสอบทานผลจำนวน 180 คน

แบบแผนการวิจัย

การปรับเทียบคะแนนครั้งนี้ออกแบบรวบรวมข้อมูลโดยใช้แบบทดสอบร่วมภายในกับกลุ่มผู้สอบที่มีความสามารถต่างกัน 3 กลุ่ม ได้แก่ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 , 2 และ 3 โดยกลุ่มตัวอย่างปรับเทียบคะแนนทำแบบทดสอบในระดับชั้นของตนเองเพียง 1 ฉบับ ซึ่งในแบบทดสอบมีข้อสอบร่วมระหว่างชั้น อย่างละ 5 ข้อ แสดงดังตารางที่ 1

ตารางที่ 1 ผลแสดงแบบทดสอบและแบบทดสอบร่วมระหว่างระดับชั้นของกลุ่มตัวอย่างเปรียบเทียบคะแนน

กลุ่มตัวอย่าง เปรียบเทียบ คะแนน	แบบทดสอบ				
	แบบทดสอบ ระดับ ม.1	แบบทดสอบร่วม ระหว่าง ม.1 กับ ม.2	แบบทดสอบ ระดับ ม.2	แบบทดสอบร่วม ระหว่าง ม.2 กับ ม.3	แบบทดสอบ ระดับ ม.3
ม.1	25 ข้อ	5 ข้อ	-	-	-
ม.2	-	5 ข้อ	20 ข้อ	5 ข้อ	-
ม.3	-	-	-	5 ข้อ	25 ข้อ

สำหรับในกลุ่มตัวอย่างสอบทานผล ต้องทำแบบทดสอบทั้ง 3 ฉบับ ซึ่งผู้วิจัยนำมารวมไว้เป็นฉบับเดียวกัน รวมข้อสอบที่ต้องทำทั้งหมด 80 ข้อ แสดงดังตารางที่ 2

ตารางที่ 2 แบบทดสอบและแบบทดสอบร่วมระหว่างระดับชั้นของกลุ่มตัวอย่างสอบทานผล

กลุ่มตัวอย่าง สอบทานผล	แบบทดสอบ				
	แบบทดสอบ ระดับ ม.1	แบบทดสอบ ร่วมระหว่าง ม.1 กับ ม.2	แบบทดสอบ ระดับ ม.2	แบบทดสอบ ร่วมระหว่าง ม.2 กับ ม.3	แบบทดสอบ ระดับ ม.3
ม.1	25 ข้อ	5 ข้อ	20 ข้อ	5 ข้อ	25 ข้อ
ม.2	25 ข้อ	5 ข้อ	20 ข้อ	5 ข้อ	25 ข้อ
ม.3	25 ข้อ	5 ข้อ	20 ข้อ	5 ข้อ	25 ข้อ

เครื่องมือการวิจัย

เครื่องมือที่ใช้ในการวิจัยนี้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่อง สมการ ระดับช่วงชั้นที่ 3 จำนวน 3 ฉบับ เนื่องจากเนื้อหาเรื่อง สมการ เป็นเรื่องที่มีการเรียนในทั้ง 3 ชั้น ทำให้ไม่เกิดความได้เปรียบเสียเปรียบกันในแต่ละระดับชั้น ซึ่งผู้วิจัยจะดำเนินการสร้างแบบทดสอบตามระดับชั้น ๆ ละ 30 ข้อ เป็นแบบทดสอบเฉพาะแต่ละระดับชั้นจำนวน 25 ข้อ และแบบทดสอบร่วมภายใน (Internal Anchor Items) ซึ่ง Angoff (1984) ได้เสนอแนะว่า ควรมีไม่ต่ำกว่าร้อยละ 20 ของจำนวนข้อสอบในแบบทดสอบ ผู้วิจัยจะดำเนินการสร้างข้อสอบร่วมภายในจำนวน 5 ข้อ รวมข้อสอบทั้งหมดในแต่ละชั้นจำนวน 30 ข้อ

1. แบบทดสอบที่ใช้สอบกับกลุ่มตัวอย่างเปรียบเทียบคะแนน มีดังนี้

1.1 แบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 1 มีข้อสอบเฉพาะจำนวน 25 ข้อ และข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 2 จำนวน 5 ข้อ รวมเป็น 30 ข้อ

1.2 แบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 2 มีข้อสอบเฉพาะจำนวน 20 ข้อ และข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 1 จำนวน 5 ข้อ และข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 3 จำนวน 5 ข้อ รวมเป็น 30 ข้อ

1.3 แบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 3 มีข้อสอบเฉพาะจำนวน 25 ข้อ และข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 2 จำนวน 5 ข้อ รวมเป็น 30 ข้อ

2. แบบทดสอบที่ใช้สอบกับกลุ่มตัวอย่างสอบทานผล เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนของชั้นมัธยมศึกษาปีที่ 1 , 2 และ 3 ทั้งสามฉบับมีข้อสอบร่วมภายในระหว่างชั้นแต่ละชั้น รวม 80 ข้อ

ผู้วิจัยได้รวมข้อสอบทั้ง 3 ฉบับมาเป็นฉบับเดียวกัน ประกอบด้วยข้อสอบเฉพาะสำหรับชั้นมัธยมศึกษาปีที่ 1 จำนวน 25 ข้อ ข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 1 กับชั้นมัธยมศึกษาปีที่ 2 จำนวน 5 ข้อ ข้อสอบเฉพาะสำหรับชั้นมัธยมศึกษาปีที่ 2 จำนวน 20 ข้อ ข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 2 กับชั้นมัธยมศึกษาปีที่ 3 จำนวน 5 ข้อ และข้อสอบเฉพาะสำหรับชั้นมัธยมศึกษาปีที่ 3 จำนวน 25 ข้อ รวมทั้งสิ้นจำนวน 80 ข้อ

การเก็บรวบรวมข้อมูล

ผู้วิจัยขอความร่วมมือจากอาจารย์ผู้สอนให้ช่วยดำเนินการสอบ โดยได้ข้อมูลครบร้อยละเปอร์เซ็นต์

การวิเคราะห์ข้อมูล

1. วิเคราะห์ค่าสถิติพื้นฐานของแบบสอบฉบับต่าง ๆ ได้แก่ คะแนนเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด ค่าต่ำสุด โดยใช้โปรแกรมสำเร็จรูป SPSS

2. การปรับเทียบโดยวิธีเชิงเส้นตรง

2.1 จากคะแนนเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของแบบทดสอบแต่ละฉบับนำมาคำนวณหาค่าความชัน (Slope: Ayx) และจุดตัดแกน (Intercept: Byx) ของสมการเส้นตรง เพื่อปรับเทียบคะแนนจากแบบสอบแต่ละฉบับให้เป็นคะแนนที่สมมูลกัน โดยใช้สูตร (ศิริชัย กาญจนวาสี, 2541)

$$T(x_i) = Byx + Ayx X_i$$

2.2 สร้างตารางเทียบคะแนนระหว่างแบบสอบ X และ Y โดยแทนค่า X_i ลงในสมการเพื่อหาคะแนนสมมูล Y_i

3. การปรับเทียบโดยวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์

3.1 ประมวลค่าพารามิเตอร์ของข้อสอบในแบบสอบร่วมและประมวลค่าความสามารถของผู้สอบแต่ละคน ด้วยโปรแกรม BILOG

3.2 ปรับค่าความสามารถของผู้สอบให้สมมูลกัน โดยใช้หลักการปรับเทียบเชิงเส้นตรง ดังนี้ (Hambleton & Swanminathan, 1985 อ้างถึงใน สุมนา ศรีอภัย และคณะ, 2550)

$$\theta_{ij} = \frac{a_i}{a_j} \theta_i + b_j - \frac{a_i}{a_j} b_i$$

3.3 นำค่าพารามิเตอร์ที่ผ่านการแปลงให้อยู่บนสเกลเดียวกันแล้ว มาคำนวณโอกาสในการตอบถูกของทั้งสองกลุ่ม จากสูตร (ศิริชัย กาญจนวาสี, 2545)

$$P_i(\theta) = c_i + \frac{(1 - c_i)}{1 + e^{-Da_i(\theta - b_i)}}$$

3.4 คำนวณหาค่าคะแนนจริงของผู้สอบแบบสอบแต่ละชุดในแต่ละระดับชั้น โดยใช้สูตรในการคำนวณ ดังนี้ (ศิริชัย กาญจนวาสี, 2545)

$$\xi_X = \sum_{i=1}^n P_i(\theta_X) \quad ; \quad \xi_Y = \sum_{i=1}^n P_i(\theta_Y)$$

- 3.5 สร้างตารางเปรียบเทียบคะแนนทั้งสองฉบับ โดยยึดค่าความสามารถระดับเดียวกันเป็นหลัก
4. การหาประสิทธิภาพของการเปรียบเทียบคะแนน

4.1 หาค่าความคลาดเคลื่อนของการเปรียบเทียบคะแนน โดยใช้สูตร

$$E_i = \left| x_i - x_i^* \right|$$

4.2 การหาค่าความเพี้ยนของการเปรียบเทียบคะแนน โดยตัดสินจากดัชนีความแตกต่างที่ได้จากกลุ่มตัวอย่างสอบทานผล โดยใช้สูตรของปีเตอร์เซน และคณะ (ศิริชัย กาญจนวาสี, 2545; อ้างอิงจาก Petersen, et al. 1982) ดังนี้

$$C = \frac{\sum_{k=1}^n (x_{ik} - x_{jik})^2}{ns_i^2}$$

4.3 ประเมินค่าดัชนีความแตกต่าง ตามหลักการประเมินของปีเตอร์เซนและคณะ (ศิริชัย กาญจนวาสี, 2545; อ้างอิงจาก Petersen, et al. 1982) ซึ่งมีเกณฑ์การยอมรับ ดังนี้

ระดับน่าพอใจอย่างยิ่ง	หมายถึง	$C \leq (.05 S_x)^2$
ระดับน่าพอใจ	หมายถึง	$(.05 S_x)^2 \leq C \leq (.10 S_x)^2$
ระดับปานกลาง	หมายถึง	$(.10 S_x)^2 \leq C \leq (.15 S_x)^2$
ระดับไม่น่าพอใจ	หมายถึง	$(.15 S_x)^2 \leq C \leq (.20 S_x)^2$
ระดับไม่น่าพอใจอย่างยิ่ง	หมายถึง	$(.20 S_x)^2 < C$

4.4 ทดสอบความแตกต่างของค่าความคลาดเคลื่อนและค่าความเพี้ยนของการเปรียบเทียบคะแนน ระหว่างวิธีการเปรียบเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ โดยใช้สถิติทดสอบที (t - test) ดังนี้

$$t = \frac{\sum D}{\sqrt{\frac{N \sum D^2}{N-1} - \frac{(\sum D)^2}{N}}}$$

ผลการวิจัย

1. ผลการเปรียบเทียบค่าความคลาดเคลื่อนของการเปรียบเทียบคะแนน

ผู้วิจัยได้หาค่าความคลาดเคลื่อนของการเปรียบเทียบคะแนน หาค่าเฉลี่ยของค่าความคลาดเคลื่อน และส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนของแบบทดสอบแล้วนำมาเปรียบเทียบค่าความคลาดเคลื่อนของการเปรียบเทียบคะแนนระหว่างวิธีการเปรียบเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ โดยใช้สถิติทดสอบที (t - test) ดังมีรายละเอียดตามตารางที่ 3

ตารางที่ 3 การเปรียบเทียบค่าความคลาดเคลื่อนของการปรับเทียบคะแนนระหว่างวิธีการปรับเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์

วิธีการปรับเทียบ	จำนวน	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	t – test
วิธีเชิงเส้นตรง	60	1.97	0.88	5.02 **
วิธีทฤษฎีการตอบสนองข้อสอบ	60	1.81	0.80	

หมายเหตุ. ** $p < .05$

จากตารางที่ 3 ค่าความคลาดเคลื่อนของการปรับเทียบคะแนนวิธีเชิงเส้นตรงมีค่าเฉลี่ยเท่ากับ 1.97 และวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าเฉลี่ยเท่ากับ 1.81 เมื่อเปรียบเทียบทั้งสองวิธี พบว่าค่าความคลาดเคลื่อนของการปรับเทียบคะแนนวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าน้อยกว่าวิธีเชิงเส้นตรง อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2. ผลการวิเคราะห์และการเปรียบเทียบค่าความเพียงพอของการปรับเทียบคะแนน

ผู้วิจัยได้หาค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนวิธีเชิงเส้นตรงและวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ ในแต่ละระดับชั้น แล้วนำมาเปรียบเทียบค่าดัชนีความแตกต่าง (C) เพื่อประเมินความแตกต่างของความเพียงพอของการปรับเทียบคะแนนโดยใช้สถิติทดสอบที (t - test) รายละเอียดตามตารางที่ 4

ตารางที่ 4 การเปรียบเทียบค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนระหว่างวิธีการปรับเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์

วิธีการปรับเทียบ	จำนวน C	ค่าเฉลี่ย	ส่วนเบี่ยงเบนมาตรฐาน	t – test
วิธีเชิงเส้นตรง	6	0.08	0.04	2.20**
วิธีทฤษฎีการตอบสนองข้อสอบ	6	0.03	0.01	

หมายเหตุ. ** $p < .05$

จากตารางที่ 4 ค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนวิธีเชิงเส้นตรงมีค่าเฉลี่ยเท่ากับ 0.08 และค่าดัชนีความแตกต่าง (C) ของการปรับเทียบคะแนนวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าเฉลี่ยเท่ากับ 0.03 แสดงว่า ความเพียงพอของวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าต่ำกว่าวิธีเชิงเส้นตรงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

สรุปผลการวิจัยได้ดังนี้

1. ผลการวิเคราะห์ความคลาดเคลื่อนของวิธีปรับเทียบเชิงเส้นตรงมีความคลาดเคลื่อนเฉลี่ยเท่ากับ 1.97 และวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีความคลาดเคลื่อนเฉลี่ยเท่ากับ 1.81 เมื่อเปรียบเทียบความแตกต่างของความคลาดเคลื่อนของการปรับเทียบคะแนนตามแนวทางวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ ปรากฏว่า ความคลาดเคลื่อนของวิธีปรับเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยที่ความคลาดเคลื่อนของวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์น้อยกว่าวิธีการปรับเทียบเชิงเส้นตรง

2. ผลการวิเคราะห์ความเพียงพอของการปรับเทียบคะแนนตามแนวตั้ง โดยพิจารณาจากค่าดัชนีความแตกต่าง (C) พบว่า วิธีการปรับเทียบเชิงเส้นตรงมีค่าดัชนีความแตกต่างเฉลี่ยเท่ากับ 0.08 และวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีค่าดัชนีความแตกต่างเฉลี่ยเท่ากับ 0.03 เมื่อเปรียบเทียบความแตกต่างของค่าดัชนีความแตกต่างระหว่างวิธีการปรับเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ ปรากฏว่าค่าดัชนีความแตกต่างของวิธีการปรับเทียบเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 โดยที่ค่าดัชนีความแตกต่างของวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์น้อยกว่าวิธีการปรับเทียบเชิงเส้นตรง แสดงว่า วิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีความเพียงพอของการปรับเทียบคะแนนอยู่ในระดับที่น่าพอใจมากกว่าวิธีการปรับเทียบเชิงเส้นตรง

การปรับเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายใน วิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีคุณภาพดีกว่าวิธีการปรับเทียบเชิงเส้นตรง

อภิปรายผลการวิจัย

ผลการวิจัยในครั้งนี้ มีประเด็นสำคัญที่นำมาอภิปราย ดังต่อไปนี้

1. การเปรียบเทียบความคลาดเคลื่อนของวิธีการปรับเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายในวิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์เป็นวิธีที่มีความคลาดเคลื่อนน้อยกว่าวิธีเชิงเส้นตรง ซึ่งสอดคล้องกับสมมติฐานที่กำหนดไว้ เนื่องจากการปรับเทียบโดยใช้วิธีเชิงเส้นตรงเป็นวิธีการปรับเทียบแบบดั้งเดิมและไม่ได้ใช้การเดาเข้ามาเกี่ยวข้องซึ่งในทางปฏิบัติจริงแบบทดสอบที่นำไปสอบกับผู้สอบนั้นย่อมมีค่าการเดาเกิดขึ้น จึงอาจเป็นสาเหตุทำให้การปรับเทียบโดยใช้วิธีเชิงเส้นตรงมีความคลาดเคลื่อนมาก นอกจากนี้ ศิริชัย กาญจนวาสี (2541), Kolen (1981) ยังได้กล่าวไว้อีกว่า การปรับเทียบเชิงเส้นตรงเป็นวิธีที่ไม่ค่อยเหมาะสมสำหรับการปรับเทียบคะแนนตามแนวตั้ง (Vertical Equating) เนื่องจากการปรับเทียบดังกล่าวมีกลุ่มตัวอย่างที่ระดับความสามารถไม่เท่ากัน อีกทั้งการแจกแจงคะแนนอาจมีรูปทรงที่แตกต่างกันได้ทั้งในแง่ความเบ้ และความโด่ง ซึ่งสอดคล้องกับคำกล่าวของ Slide and Linn (1979) ที่กล่าวว่า การปรับเทียบในแนวตั้ง ควรใช้โมเดล สองพารามิเตอร์หรือ สามพารามิเตอร์ จะมีความเหมาะสมมากกว่าโมเดลหนึ่งพารามิเตอร์ เนื่องจากลักษณะข้อสอบอาจไม่สอดคล้องกับโมเดล

2. การเปรียบเทียบความเพียงพอของการปรับเทียบคะแนนตามแนวตั้ง โดยพิจารณาจากค่าดัชนีความแตกต่าง (C) วิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์เป็นวิธีที่มีค่าดัชนีความแตกต่าง (C) น้อยกว่าวิธีเชิงเส้นตรงทำให้วิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์มีความเพียงพอของการปรับเทียบคะแนนอยู่ในระดับที่น่าพอใจมากกว่าวิธีเชิงเส้นตรง ซึ่งสอดคล้องกับสมมติฐานและสอดคล้องกับผลการศึกษาของ พรพิมล นาคเวช (2537) ที่พบว่าการปรับเทียบคะแนนตามแนวตั้งโดยใช้ทฤษฎีการตอบสนองข้อสอบ สำหรับผู้ที่มีความสามารถต่างกัน 3 ระดับ คือนักเรียนชั้น ม.1 , ม.2 และ ม.3 มีค่าความเพียงพอของการเทียบคะแนนอยู่ในระดับที่น่าพอใจ นอกจากนี้ยังสอดคล้องกับ รุ่งภา จันทรา (2540) และ เกตรินทร์ หมดอะดัม (2545) พบว่า การเปรียบเทียบค่าความเพียงพอระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์ โดยที่ค่าความเพียงพอของการเทียบคะแนนรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีสามพารามิเตอร์น้อยกว่าค่าความเพียงพอของการเทียบคะแนนรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์ และอยู่ในระดับที่น่าพอใจกว่าแบบหนึ่งพารามิเตอร์

ข้อเสนอแนะ

การนำผลการวิจัยไปใช้

ในการปรับเทียบคะแนนตามแนวตั้ง และกลุ่มตัวอย่างมีความสามารถต่างกัน ควรใช้วิธีทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์ในการปรับเทียบคะแนน เนื่องจากเป็นวิธีที่มีค่าความคลาดเคลื่อนต่ำและมีค่าความเพียงพอของการปรับเทียบคะแนนอยู่ในระดับที่น่าพอใจ อีกทั้งยังเป็นวิธีที่เหมาะสมและถูกต้องตามสถานการณ์ในปัจจุบัน เพราะมีการหาโอกาสในการเดาเข้ามาเกี่ยวข้องด้วย

การทำวิจัยต่อไป

1. ควรมีการศึกษาเปรียบเทียบคุณภาพของการปรับเทียบคะแนนตามแนวตั้งในวิธีอื่น ๆ เช่น วิธีการใช้สมการถดถอย เป็นต้น
2. ควรมีการศึกษาเกี่ยวกับผลของความยาวของแบบทดสอบร่วมที่มีผลต่อการปรับเทียบคะแนนตามแนวตั้ง โดยใช้ทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์
3. ควรมีการเปรียบเทียบคุณภาพของการปรับเทียบคะแนนตามแนวตั้งและแนวนอนโดยใช้ทฤษฎีการตอบสนองข้อสอบแบบสามพารามิเตอร์

เอกสารอ้างอิง

- เกตรินทร์ หมดอดัม. (2545). *การเปรียบเทียบคุณภาพของการเทียบมาตรฐานตามแนวตั้งระหว่างรูปแบบทฤษฎีการตอบสนองที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์*. วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยทักษิณ.
- พรพิมล นาคเวช. (2537). *การศึกษาคุณภาพของการเทียบมาตรฐานในแนวตั้งโดยใช้ทฤษฎีการตอบสนองข้อสอบ*. วิทยานิพนธ์ปริญญาครุศาสตรมหาบัณฑิต, สาขาวิชาการวัดและประเมินผลการศึกษา, บัณฑิตวิทยาลัย, จุฬาลงกรณ์มหาวิทยาลัย.
- รุ่งนภา จันทรา. (2540). *การเปรียบเทียบความคลาดเคลื่อนและความเพียงพอของการเทียบมาตรฐานระหว่างรูปแบบทฤษฎีการตอบสนองข้อสอบที่มีหนึ่งพารามิเตอร์กับสามพารามิเตอร์*. วิทยานิพนธ์ปริญญาการศึกษามหาบัณฑิต, สาขาวิชาการวัดผลและวิจัยการศึกษา, บัณฑิตวิทยาลัย, มหาวิทยาลัยสงขลานครินทร์.
- สุนนา ศรีอบเชย, เสรี ชัดแฉ่ม และ สมโภชน์ อเนกสุข (2550). การปรับเทียบคะแนนตามแนวตั้งโดยใช้แบบทดสอบร่วมภายใน: การเปรียบเทียบคุณภาพระหว่างวิธีเชิงเส้นตรงกับวิธีทฤษฎีการตอบสนองข้อสอบ. *วารสารวิจัยและวัดผลการศึกษา มหาวิทยาลัยบูรพา*. 5(2), 49-60.
- ศิริชัย กาญจนวาสี. (2541). *การปรับเทียบมาตรฐานระหว่างแบบทดสอบ*. กรุงเทพมหานคร : ศูนย์ตำราและเอกสารทางวิชาการ. จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2545). *ทฤษฎีการทดสอบแนวใหม่ (Modern Test Theories)*. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร : จุฬาลงกรณ์มหาวิทยาลัย.
- Angoff, W. H. (1971). *Scales, norms, and equivalent scores in R.L. Thorndike (Ed.) Educational Measurement*. 2 nd ed. Washington, D.C.: American on Education.

- Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, N. J.: Educational Testing Service.
- Flanagan, T. C. (1951). *Units, scores and norms*. In F.F. Linguist (Ed). Educational measurement. Washington D.C.: American Council on Education.
- Kolen, M.J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 18(spring 1981), 1-11.
- Kolen, M.J., & Brennan, R.L. (1987). A reply to Angoff. *Applied Psychological Measurement*, 11(3), 301-306.
- Kolen, M.J., & Whitney, D.R. (1982). Comparison of four procedures for equating tests of general educational development. *Journal of Educational Measurement*, 19 (Winter 1982), 279-292.
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc.