



สรุปผลการวิจัย อภิปรายผลและข้อเสนอแนะ

การเทียบมาตรฐานในปัจจุบันนี้มีหลายวิธีการและหลายรูปแบบ ซึ่งการนำไปใช้ควรมีการศึกษาและวิจัยเกี่ยวกับประสิทธิภาพของการเทียบมาตรฐานเสียก่อน ดังนั้นในการวิจัยครั้งนี้จึงมีวัตถุประสงค์ดังต่อไปนี้

1. เปรียบเทียบประสิทธิภาพของการเทียบมาตรฐานระหว่างรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบรวม ด้วยวิธี
 - 1.1 การเทียบเชิงเส้นตรง
 - 1.2 การเทียบโดยใช้เปอร์เซ็นต์ไทล์
 - 1.3 การเทียบโดยใช้โค้งลักษณะข้อสอบ
2. เปรียบเทียบความคงที่ของวิธีการเทียบมาตรฐานโดยรูปแบบที่ใช้ผู้สอบร่วมระหว่างวิธี
 - 2.1 การเทียบเชิงเส้นตรงกับการเทียบโดยใช้เปอร์เซ็นต์ไทล์
 - 2.2 การเทียบเชิงเส้นตรงกับการเทียบโดยใช้โค้งลักษณะข้อสอบ
 - 2.3 การเทียบโดยใช้เปอร์เซ็นต์ไทล์กับการเทียบโดยใช้โค้งลักษณะข้อสอบ
3. เปรียบเทียบความคงที่ของวิธีการเทียบมาตรฐานโดยรูปแบบที่ใช้แบบสอบร่วมระหว่างวิธี
 - 3.1 การเทียบเชิงเส้นตรงกับการเทียบโดยใช้เปอร์เซ็นต์ไทล์
 - 3.2 การเทียบเชิงเส้นตรงกับการเทียบโดยใช้โค้งลักษณะข้อสอบ
 - 3.3 การเทียบโดยใช้เปอร์เซ็นต์ไทล์กับการเทียบโดยใช้โค้งลักษณะข้อสอบ

สมมติฐานของการวิจัย

1. การเปรียบเทียบรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบรวม ทั้ง 3 วิธี ผู้วิจัยได้ตั้งสมมติฐานไว้ คือ

1.1 การเทียบเชิงเส้นตรงรูปแบบที่ใช้แบบสอบถามน่าจะมีประสิทธิภาพดีกว่ารูปแบบที่ใช้ผู้สอบรวม

1.2 การเทียบโดยใช่เปอร์เซ็นต์ รูปแบบที่ใช้แบบสอบถามน่าจะมีประสิทธิภาพดีกว่ารูปแบบที่ใช้ผู้สอบรวม

1.3 การเทียบโดยใช่โค้งลักษณะข้อสอบ รูปแบบที่ใช้แบบสอบถามน่าจะมีประสิทธิภาพดีกว่ารูปแบบที่ใช้ผู้สอบรวม

2. การเปรียบเทียบวิธีการเทียบมาตราทั้ง 3 วิธี โดยรูปแบบที่ใช้ผู้สอบรวม ผู้วิจัยได้ตั้งสมมุติฐานไว้ คือ

2.1 การเทียบเชิงเส้นตรงและการเทียบโดยใช่เปอร์เซ็นต์น่าจะมี ความไม่คงที่พอๆกัน

2.2 การเทียบโดยใช่โค้งลักษณะข้อสอบน่าจะมี ความคงที่มากกว่าการเทียบเชิงเส้นตรง

2.3 การเทียบโดยใช่โค้งลักษณะข้อสอบน่าจะมี ความคงที่มากกว่า การเทียบโดยใช่เปอร์เซ็นต์

3. การเปรียบเทียบวิธีการเทียบมาตราทั้ง 3 วิธี โดยรูปแบบที่ใช้แบบสอบถาม ผู้วิจัยได้ตั้งสมมุติฐานไว้ คือ

3.1 การเทียบเชิงเส้นตรงและการเทียบโดยใช่เปอร์เซ็นต์น่าจะมี ความไม่คงที่พอๆกัน

3.2 การเทียบโดยใช่โค้งลักษณะข้อสอบน่าจะมี ความคงที่มากกว่าการเทียบเชิงเส้นตรง

3.3 การเทียบโดยใช่โค้งลักษณะข้อสอบน่าจะมี ความคงที่มากกว่า การเทียบโดยใช่เปอร์เซ็นต์

วิธีดำเนินการวิจัย

ประชากรที่ใช้ในการวิจัยเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศ จำนวน 146,355 คน ที่เข้าสอบตามโครงการตรวจสอบคุณภาพการศึกษา ระหว่างวันที่ 13-14 ธันวาคม พ.ศ. 2526

กลุ่มตัวอย่างได้จากการสุ่ม 3 ครั้ง ครั้งแรกใช้วิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) ตามเขตการศึกษาทั้ง 13 เขต ได้จำนวน 3,721 คน

ครั้งที่สองใช้วิธีการสุ่มอย่างมีระบบ (Systematic sampling) แยกเป็น 3 กลุ่ม คือกลุ่มที่ 1 จำนวน 1,241 คน กลุ่มที่ 2 จำนวน 1,240 คน และกลุ่มที่ 3 จำนวน 1,240 คน ครั้งที่สามใช้วิธีการสุ่มอย่างมีระบบแยกทั้ง 3 กลุ่ม เป็น 6 กลุ่มย่อยไม่ซ้ำกัน

เครื่องมือที่ใช้ในการวิจัย เป็นแบบสอบถามชนิดเลือกตอบ 60 ข้อ ผู้วิจัยได้แบ่งแบบสอบถามออกเป็น 2 พอร์ม พอร์มละ 38 ข้อ มีแบบสอบถามชนิดแบบสอบถามภายใน (Internal anchor test) จำนวน 20 ข้อ (ผู้วิจัยได้ใช้หลักเกณฑ์ที่แองกอฟฟ์ เสนอไว้ว่าไม่ควรน้อยกว่า 20 ข้อซึ่งมากกว่า 20 % ของจำนวนข้อในแต่ละพอร์ม)

ข้อมูลที่ใช้ในการวิจัยนี้ได้จากบันทึกผลการสอบของนักเรียนชั้น ม.6 ทั่วประเทศ ที่เข้าสอบตามโครงการตรวจสอบคุณภาพการศึกษาระหว่าง วันที่ 13-14 ธันวาคม 2526 จากสำนักทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ

การวิเคราะห์ข้อมูล

1. ใช้โปรแกรมคอมพิวเตอร์ SPSSX คำนวณค่าสถิติพื้นฐานของกลุ่มตัวอย่าง ทั้ง 6 กลุ่ม คือค่าเฉลี่ย (mean) และส่วนเบี่ยงเบนมาตรฐาน (Standard deviation)
2. การวิเคราะห์รายข้อ
 - 2.1 ใช้โปรแกรม LOGIST ประมาณค่าพารามิเตอร์ของข้อกระทงและประมาณค่าความสามารถของผู้สอบ
 - 2.2 ใช้โปรแกรม ITEM ANALYSIS วิเคราะห์ค่าความยากและอำนาจจำแนก
3. การเทียบมาตรฐานด้วยวิธีการเทียบเชิงเส้นตรงและการเทียบโดยใช้เปอร์เซ็นต์ ใช้วิธีการที่เสนอโดยแองกอฟฟ์ (Angoff 1971) ส่วนวิธีการเทียบโดยใช้โค้งลักษณะข้อสอบใช้วิธีที่เสนอโดย วอร์ม (Warm 1978) และลอร์ด (Lord 1980)
4. การเปรียบเทียบประสิทธิภาพของการเทียบมาตรฐานระหว่างรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบถามใช้ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐาน (Standard error of equating) เป็นสิ่งเปรียบเทียบ
5. การเปรียบเทียบความคงที่ของวิธีการเทียบมาตรฐาน (ความไม่แปรผันตามกลุ่มสมมาตร และอัสคีร์) ใช้สถิติโคโมโกรอฟ-สมิรโนฟ (Kolmogorow-Smirnov two sample test)

ผลการวิจัย

1. ผลการเปรียบเทียบประสิทธิภาพของการเทียบมาตรฐานระหว่างรูปแบบที่ให้ผู้สอบร่วมกับรูปแบบที่ไร้แบบสอบรวม ด้วยวิธีแตกต่างกัน 3 วิธี ปรากฏผลดังนี้

1.1 วิธีการเทียบเชิงเส้นตรงรูปแบบที่ให้ผู้สอบร่วมมีประสิทธิภาพไม่แตกต่างไปจากรูปแบบที่ไร้แบบสอบรวม (ตารางที่ 6) ซึ่งไม่เป็นไปตามสมมุติฐานที่ 1.1

1.2 วิธีการเทียบโดยไร้เปอร์เซ็นต์รูปแบบที่ให้ผู้สอบร่วมมีประสิทธิภาพไม่แตกต่างไปจากรูปแบบที่ไร้แบบสอบรวม (ตารางที่ 7) ซึ่งไม่เป็นไปตามสมมุติฐานที่ 1.2

1.3 วิธีการเทียบโดยไร้เครื่องหมายข้อสอบรูปแบบที่ให้ผู้สอบร่วมมีประสิทธิภาพไม่แตกต่างไปจากรูปแบบที่ไร้แบบสอบรวม (ตารางที่ 8) ซึ่งไม่เป็นไปตามสมมุติฐานที่ 1.3

2. ผลการเปรียบเทียบความคงที่ของวิธีการเทียบมาตรฐานโดยรูปแบบที่ให้ผู้สอบร่วมทั้ง 3 วิธี ปรากฏผลดังนี้

2.1 การเทียบโดยไร้เปอร์เซ็นต์มีความคงที่มากกว่าการเทียบเชิงเส้นตรง (ตารางที่ 20) ซึ่งไม่เป็นไปตามสมมุติฐานที่ 2.1

2.2 การเทียบโดยไร้เครื่องหมายข้อสอบมีความคงที่มากกว่าการเทียบเชิงเส้นตรง (ตารางที่ 20) ซึ่งเป็นไปตามสมมุติฐานที่ 2.2

2.3 การเทียบโดยไร้เปอร์เซ็นต์มีความคงที่พอๆกับการเทียบโดยไร้เครื่องหมายข้อสอบ (ตารางที่ 20) ซึ่งไม่เป็นไปตามสมมุติฐานที่ 2.3

3. ผลการเปรียบเทียบความคงที่ของวิธีการเทียบมาตรฐานโดยรูปแบบที่ไร้แบบสอบรวมทั้ง 3 วิธี ปรากฏผลดังนี้

3.1 การเทียบโดยไร้เปอร์เซ็นต์มีความคงที่มากกว่าการเทียบเชิงเส้นตรง (ตารางที่ 32) ซึ่งไม่เป็นไปตามสมมุติฐานที่ 3.1

3.2 การเทียบโดยไร้เครื่องหมายข้อสอบมีความคงที่มากกว่าการเทียบเชิงเส้นตรง (ตารางที่ 32) ซึ่งเป็นไปตามสมมุติฐานที่ 3.2

3.3 การเทียบโดยไร้เปอร์เซ็นต์มีความคงที่พอๆกับการเทียบโดยไร้เครื่องหมายข้อสอบ (ตารางที่ 32) ซึ่งไม่เป็นไปตามสมมุติฐานที่ 3.3

การอภิปรายผลการวิจัย

ในการเปรียบเทียบประสิทธิภาพของการเทียบมาตราโดยรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบรวม ทั้ง 3 วิธี คือการเทียบเชิงเส้นตรง , การเทียบโดยใช้เปอร์เซ็นต์ และ การเทียบโดยใช้โค้งลักษณะข้อสอบ ปรากฏผลว่ารูปแบบที่ใช้ผู้สอบรวมมีประสิทธิภาพไม่แตกต่างไปจากรูปแบบที่ใช้แบบสอบรวมทั้ง 3 วิธี ซึ่งไม่เป็นไปตามสมมุติฐานที่ 1.1 , 1.2 และ 1.3 นั้น ทั้งนี้เป็นผลมาจากการสุ่มตัวอย่างที่เป็นนักเรียนทั่วประเทศโดยใช้โปรแกรมคอมพิวเตอร์ทำให้ได้กลุ่มตัวอย่างที่เป็นตัวแทนประชากรที่ดี และเมื่อสุ่มอย่างมีระบบในชั้นที่สองแยกเป็นกลุ่มอัลฟา (α) และกลุ่มเบตา (β) ก็ได้กลุ่มตัวอย่างที่มีความสามารถใกล้เคียงกันมาก ซึ่งทำให้งานวิจัยนี้มีลักษณะเป็นการออกแบบที่ดี (well design) จึงเป็นผลทำให้คะแนนฟอร์ม X จากการสอบของกลุ่มอัลฟา (α) และคะแนนฟอร์ม Y จากการสอบของกลุ่มเบตา (β) มีสเกลคะแนนใกล้เคียงกัน ดังนั้นจึงเป็นสาเหตุทำให้เทคนิคการปรับคะแนนหรือการเทียบมาตราระหว่างการเทียบโดยใช้รูปแบบที่ใช้ผู้สอบรวมกับรูปแบบที่ใช้แบบสอบรวมได้ผลใกล้เคียงกันทั้งสองรูปแบบโดยที่วิธีการทั้งสามวิธีได้ผลไม่แตกต่างกัน

ถ้าพิจารณาแต่ละวิธีการจะพบว่า วิธีการเทียบเชิงเส้นตรง (ตารางที่ 6) รูปแบบที่ใช้แบบสอบรวมมีค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตราสูงกว่ารูปแบบที่ใช้ผู้สอบรวมเหมือนกันตลอดช่วงคะแนน (กราฟในภาพที่ 4) ซึ่งแตกต่างจากวิธีการเทียบโดยใช้เปอร์เซ็นต์ (ตารางที่ 7) รูปแบบที่ใช้ผู้สอบรวมและรูปแบบที่ใช้แบบสอบรวมมีค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตราสูงค่าสลับกัน บางช่วงก็เท่ากัน (กราฟในภาพที่ 5) ส่วนวิธีการเทียบโดยใช้โค้งลักษณะข้อสอบ (ตารางที่ 8) ได้คะแนนที่ผ่านการเทียบมาตราทั้งสองแบบเท่ากันทุกช่วงคะแนน สาเหตุที่เป็นเช่นนี้เพราะการปรับสเกลคะแนนทั้งสองสเกลโดยรูปแบบที่ใช้ผู้สอบรวมใช้ค่าพารามิเตอร์ความสามารถ (θ) ของผู้สอบเป็นตัวปรับ ส่วนรูปแบบที่ใช้แบบสอบรวมใช้ค่าพารามิเตอร์ความยากของข้อกระทง (b) เป็นตัวปรับ แต่ทั้ง θ และ b ต่างก็อยู่ในสเกลเดียวกัน และการวิจัยครั้งนี้ใช้กลุ่มตัวอย่างทั้งสองกลุ่มที่มีความสามารถใกล้เคียงกัน ($\theta_1 \approx \theta_2$) แบบสอบทั้งสองฟอร์มมีความยากใกล้เคียงกัน ($b_1 \approx b_2$) จึงทำให้การปรับสเกลคะแนนทั้งสองรูปแบบได้ผลเท่ากัน ซึ่งถ้าพิจารณาผลการวิจัยแยกเปรียบเทียบแต่ละวิธีการจะพบว่าแตกต่างกันเล็กน้อย แต่โดยสรุปแล้วทั้งสามวิธีก็ได้ผลตรงกันคือรูปแบบที่ใช้แบบสอบรวมและรูปแบบที่ใช้ผู้สอบรวมมีประสิทธิภาพ

ภาพไม่แตกต่างกัน และเป็นที่สังเกตได้ว่า ค่าความคลาดเคลื่อนมาตรฐานของการเทียบ
 มาตรฐานทั้งวิธีการเทียบเชิงเส้นตรงและวิธีการเทียบโดยใช้เปอร์เซนไทล์มีลักษณะใกล้เคียง
 กันคือ ช่วงคะแนนต่ำและช่วงคะแนนสูง จะได้ค่าความคลาดเคลื่อนมาตรฐานของการเทียบ
 มาตรฐานสูงกว่าช่วงคะแนนตรงกลาง (ภาพที่ 4 และภาพที่ 5) นั่นคือ ถ้ากลุ่มตัวอย่างมีความ
 สามารถต่ำจะได้ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานสูง และถ้ากลุ่มตัวอย่าง
 มีความสามารถสูงจะได้ค่าความคลาดเคลื่อนมาตรฐานของการเทียบมาตรฐานสูง เช่นเดียวกัน
 แต่กลุ่มที่มีความสามารถปานกลางถ้าเทียบมาตรฐานแล้วจะได้ค่าความคลาดเคลื่อนมาตรฐานต่ำ

สำหรับการเปรียบเทียบความคงที่ของวิธีการเทียบมาตรฐานโดยรูปแบบที่ใช้ผู้สอบ
 รวมทั้ง 3 วิธี คือการเทียบเชิงเส้นตรง การเทียบโดยใช้เปอร์เซนไทล์ และการเทียบ
 โดยใช้โค้งลักษณะข้อสอบได้ผลคือ วิธีการเทียบโดยใช้เปอร์เซนไทล์มีความคงที่มากกว่า
 การเทียบเชิงเส้นตรงซึ่งไม่เป็นไปตามสมมุติฐานที่ 2.1 การเทียบโดยใช้โค้งลักษณะ
 ข้อสอบมีความคงที่มากกว่าการเทียบเชิงเส้นตรง เป็นไปตามสมมุติฐานที่ 2.2 และการ
 เทียบโดยใช้เปอร์เซนไทล์มีความคงที่เหมือนกับการเทียบโดยใช้โค้งลักษณะข้อสอบซึ่งไม่
 เป็นไปตามสมมุติฐานที่ 2.3

ส่วนการเปรียบเทียบความคงที่ของวิธีการเทียบมาตรฐานโดยรูปแบบที่ใช้แบบสอบ
 รวมทั้ง 3 วิธีได้ผลเช่นเดียวกับรูปแบบที่ใช้ผู้สอบร่วมกันนั่นคือ วิธีการเทียบโดยใช้เปอร์-
 เซนไทล์มีความคงที่มากกว่าการเทียบเชิงเส้นตรงซึ่งไม่เป็นไปตามสมมุติฐานที่ 3.1
 การเทียบโดยใช้โค้งลักษณะข้อสอบมีความคงที่มากกว่าการเทียบเชิงเส้นตรง เป็นไปตาม
 สมมุติฐานที่ 3.2 และการเทียบโดยใช้เปอร์เซนไทล์มีความคงที่เหมือนกับการเทียบโดย
 ใช้โค้งลักษณะข้อสอบ ซึ่งไม่เป็นไปตามสมมุติฐานที่ 3.3

ทั้งนี้อาจเนื่องมาจากความสัมพันธ์ระหว่างคะแนนของแบบสอบทั้งสองฉบับมีลักษณะ
 ไม่เป็นเส้นตรงตลอดช่วงคะแนนทำให้วิธีการเทียบเชิงเส้นตรงไม่มีคุณสมบัติอิควิตี้ (Equity)
 และจากเหตุผลอีกประการหนึ่งที่ปีเตอร์เซน มาร์โคและสจิวท (Petercen Marco &
 Steward 1982 : 72) กล่าวไว้ในงานวิจัยของเขาคือ ปัญหาของการเทียบมาตรฐานด้วย
 วิธีการเทียบเชิงเส้นตรงมักพบเมื่อเทียบจากคะแนนสอบฟอร์ม x ไปสู่ฟอร์ม y จะได้ผล
 ไม่เหมือนกันกับเมื่อเทียบจากคะแนนฟอร์ม y ไปสู่ฟอร์ม x โดยเขาให้เหตุผลว่า
 รีเกรสชันของ x โดย y (Regression of X on Y) ไม่เหมือนกับรีเกรสชันของ
 y โดย x (Regression of Y on X) แต่ทั้งสองวิธีการคือ การเทียบโดยใช้

เปอร์เซนไทล์ และวิธีการเทียบโดยใช้โค้งลักษณะข้อสอบมีคุณสมบัติคือ (Equity) ดังนั้นจึงถือว่าวิธีการเทียบโดยใช้เปอร์เซนไทล์และวิธีการเทียบโดยใช้โค้งลักษณะข้อสอบมีความคงที่ไม่แตกต่างกันและทั้งสองวิธีมีความคงที่มากกว่าการเทียบเชิงเส้นตรง จากผลการวิจัยนี้สอดคล้องกับผลการวิจัยของ ปีเตอร์เซน มาร์โก และสจิวท (Petersen Marco & Steward 1982 : 71-135) ที่พบว่า การเทียบโดยใช้เปอร์เซนไทล์ดีกว่าการเทียบเชิงเส้นตรง และในปี ค.ศ. 1983 กลุ่มเดียวกันนี้ได้ทำการวิจัยพบว่า การเทียบโดยใช้โค้งลักษณะข้อสอบมีความคงที่มากกว่าการเทียบเชิงเส้นตรง

ข้อเสนอแนะ

1. สามารถเลือกใช้รูปแบบใดรูปแบบหนึ่งจากรูปแบบที่ใช้ผู้สอบรวมหรือรูปแบบที่ใช้แบบสอบรวม กับการเทียบมาตราที่กลุ่มตัวอย่างมีความสามารถใกล้เคียงกัน
2. สามารถเลือกใช้เทคนิคการเทียบมาตราวิธีใดวิธีหนึ่งได้จากวิธีการเทียบโดยใช้เปอร์เซนไทล์หรือวิธีการเทียบโดยใช้โค้งลักษณะข้อสอบ กับการเทียบคะแนนวิชาคณิตศาสตร์
3. ปัจจุบันได้มีการพัฒนาเทคนิคการเทียบมาตราใหม่ๆหลายวิธี เช่น การนำเทคนิคการวิเคราะห์องค์ประกอบ (Factor analysis) มาใช้ในการเทียบมาตรา ดังนั้นจึงควรจะมีการทำการศึกษาวิจัยต่อไปถึงประสิทธิภาพหรือคุณภาพว่าจะดีมากน้อยเพียงใด
4. ควรจะได้มีการศึกษาวิจัยถึงประสิทธิภาพของรูปแบบต่างๆทั้ง 6 รูปแบบที่แองกอล์ฟเสนอไว้เมื่อ ค.ศ. 1971
5. ควรจะได้มีการศึกษาประสิทธิภาพของการเทียบมาตราโดยใช้คะแนนจากวิชาอื่นๆที่ไม่ใช่วิชาคณิตศาสตร์
6. ในกรณีของการเทียบมาตราที่ใช้แบบสอบรวม (Anchor test) น่าจะมีการศึกษาวิจัยถึงประสิทธิภาพของการเทียบมาตราที่ใช้แบบสอบรวมว่า ควรจะใช้แบบสอบรวมสั้นหรือยาวเพียงใด จึงจะได้ผลการเทียบมาตราที่มีประสิทธิภาพสูงสุด