วิธีปรับเทียบแบบใหม่ในกลุ่มตัวอย่างขนาดเล็ก
New Equating Methods in Small Samples

น้ำทิพย์ องอาจวาณิชย์[1]
Namthip Ongardwanich[1]

บทคัดย่อ

การปรับเทียบคะแนนระหว่างแบบสอบ เป็นกระบวนการทางสถิติที่ใช้ในการแปลงคะแนนจากแบบสอบต่างฉบับที่วัดเนื้อหาเดียวกัน ให้อยู่บนมาตรหรือสเกลของแบบสอบอีกฉบับที่ถือเป็นคะแนนที่สมมูลกัน เพื่อทำให้คะแนนจากแบบสอบต่างฉบับสามารถเปรียบเทียบกันได้โดยตรง โดยที่ขนาดกลุ่มตัวอย่างมีผลต่อการปรับเทียบคะแนน โดยขนาดกลุ่มตัวอย่างที่ใหญ่ประมาณ 3,000 ให้ผลการปรับเทียบที่มีประสิทธิภาพซึ่งในทางปฏิบัติสำหรับการสอบทั่วไปจะไม่ค่อยพบกลุ่มตัวอย่างขนาดใหญ่ จึงมีการพัฒนาวิธีที่ใช้ในการปรับเทียบใหม่ ๆ เพื่อให้เหมาะสมกับการปรับเทียบกลุ่มตัวอย่างขนาดเล็กโดยให้มีความคลาดเคลื่อนน้อยที่สุด คือ วิธี Circle-arc equating ของ Livingston และ Kim (2009) และวิธี Synthetic Linking Function ของ Kim, von Davier และ Haberman (2011)

คำสำคัญ: วิธีปรับเทียบแบบใหม่ กลุ่มตัวอย่างขนาดเล็ก

Abstract

Test Score Equating is a statistical process used to transform test scores from different test forms that measure the same subject matter onto the scale of another test form to have equivalent scores. This process can compare test scores from different test forms directly. The sample affects test score equating. The 3,000-sample size gives efficient equating. However, this sample size does not appear in general testing. Therefore, new test score equating methods are developed to be applied with small sample size with least error. These method are called Circle-arc equating invented by Livingston and Kim (2009) and Synthetic Linking Function invented by Kim, von Davier and Haberman (2011).

Keywords: new equating methods, small sample size

[1] คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร e-mail: namthipo@nu.ac.th

[1] Faculty of Education, Naresuan University, e-mail: namthipo@nu.ac.th

บทนำ

ในการสอบที่ใช้แบบสอบเดียวกันในผู้สอบต่างกลุ่มกันและต่างเวลาทำให้ผู้สอบกลุ่มหลังอาจจะได้เปรียบผู้สอบกลุ่มแรก เกิดจากที่ผู้สอบครั้งแรกจะนำข้อสอบไปเผยแพร่กับคนที่สอบในครั้งหลังทราบและมีการศึกษาค้นคว้าเพิ่มเติม ทำให้ผลการสอบไม่ตรงกับความสามารถของผู้สอบ เพื่อให้ผลการสอบมีประสิทธิภาพ มีความยุติธรรมกับผู้สอบ จึงต้องมีการปรับเทียบคะแนนก่อนนำผลไปใช้ร่วมกัน ซึ่งการปรับเทียบ (Equating) เป็นกระบวนการที่ใช้กำหนดความสัมพันธ์ระหว่างคะแนนจากแบบสอบต่างฉบับกัน ทดสอบคนต่างกลุ่ม และทำการทดสอบต่างเวลากันก็สามารถใช้คะแนนจากแบบสอบฉบับหนึ่งปรับเข้าสู่สเกลของแบบสอบฉบับอื่นได้ ทำให้คะแนนจากการสอบสามารถปรับเทียบกันได้อย่างมีความหมาย ดังนั้นจึงเกิดการปรับเทียบคะแนน (Equating methods) (Kolen and Brennan, 2014) เพื่อให้คะแนนสามารถใช้สับเปลี่ยน (Interchangeably) กันได้

แนวคิดเชิงทฤษฎีของการปรับเทียบ

การปรับเทียบคะแนน เป็นกระบวนการเชิงประจักษ์สำหรับสร้างความสัมพันธ์ระหว่างคะแนนใน 2 แบบสอบเพื่อให้สามารถแปลงคะแนนของแบบสอบฉบับหนึ่งไปสู่คะแนนของแบบสอบอีกฉบับหนึ่ง (Petersen, Kolen and Hoover, 1989 ; Livingston and Kim, 2010) การปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกันมีระดับความยากใกล้เคียงกัน และกลุ่มผู้สอบมีการแจกแจงความสามารถ อยู่ในประชากรเดียวกันหรือมีความสามารถใกล้เคียงกัน เป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบสอบเนื้อหาเดียวกันขึ้นมาหลาย ๆ ฉบับ เพื่อนำไปใช้ในการทดสอบให้เกิดความยุติธรรมและป้องกันความลับของข้อสอบเมื่อใช้ต่างเวลากัน สำหรับกลุ่มผู้สอบขนาดใหญ่ เพื่อปรับเทียบว่าคะแนนที่ได้จากฉบับหนึ่งเทียบเป็นเท่าไรอีกฉบับหนึ่ง ซึ่งวัดในระดับเดียวกัน จึงเป็นการปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับของวิชาเดียวกัน สำหรับกลุ่มผู้สอบระดับชั้นเดียวกัน เรียกว่า การปรับเทียบคะแนนตามแนวนอน (Horizontal equating) การปรับเทียบคะแนนระหว่างแบบสอบต่างฉบับกัน เมื่อแต่ละฉบับมุ่งวัดคุณลักษณะเดียวกัน แต่มีระดับความยากแตกต่างกัน และกลุ่มผู้สอบมีการแจกแจงความสามารถอยู่ต่างประชากรกันหรือมีความสามารถแตกต่างกันซึ่งเป็นเทคนิคที่เหมาะสมในสถานการณ์ที่มีความจำเป็นต้องสร้างแบบสอบเนื้อหาเดียวกัน แต่ต่างฉบับต่างมุ่งวัดความสามารถของผู้สอบที่ต่างระดับกัน เพื่อปรับเทียบว่าคะแนนที่สอบได้จากฉบับหนึ่ง เทียบเป็นเท่าไรของฉบับอื่นที่วัดต่างระดับกัน จึงเป็นการปรับเทียบคะแนนระหว่างแบบสอบวิชาเดียวกัน สำหรับกลุ่มผู้สอบต่างระดับชั้นกัน เรียกว่า การปรับเทียบคะแนนตามแนวตั้ง (vertical equating) (Hambleton and Swamminatan, 1985 ; ศิริชัย กาญจนวาสี, 2550) นอกจากนี้ได้กำหนดเงื่อนไขของการปรับเทียบคะแนนไว้ 4 ประการดังนี้ (Hambleton and Swamminatan, 1985 ; Kolen, 1999 ; Peterson, 2007)

1. ความเสมอภาค (Equity) คือ เมื่อทุกกลุ่มมีความสามารถเท่าเทียมกัน การแจกแจงคะแนนของแบบสอบ y หลังจากที่มีการปรับคะแนนแล้วจะมีการแจกแจงเหมือนกับการแจกแจงของคะแนนจากแบบสอบ x จะต้องไม่มีความแตกต่างที่ทุก ๆ ระดับความสามารถไม่ว่าจะสอบด้วยแบบสอบฉบับ x หรือ y แบบสอบที่วัดคุณลักษณะ (Traits) หรือความสามารถ (Abilities) ต้องไม่ต่างกัน มีความคู่ขนานกัน

2. ความไม่แปรเปลี่ยนตามกลุ่ม (Invariance across groups) คือการปรับคะแนนต้องเป็นไปในลักษณะเดียวกัน ไม่ว่าคะแนนจะมาจากกลุ่มตัวอย่างใดก็ตาม

3. ความสมมาตร (Symmetry) คือ การปรับคะแนนสามารถเปลี่ยนกลับได้เช่น การแปลงคะแนนจากฉบับ x ไปยังฉบับ y มีผลเช่นเดียวกันกับฉบับ y ไปสู่ฉบับ x

4. ความเป็นเอกมิติของแบบสอบ (Unidimentionality of the tests) คือแบบสอบทั้งสองฉบับวัดในคุณลักษณะเดียวกัน อาจเป็นคุณลักษณะแฝง ความสามารถหรือทักษะ อย่างใดอย่างหนึ่งก็ได้

ในการสอบแต่ละครั้งอาจเกิดปัญหาโดยมีผู้สอบจำนวนน้อยซึ่งต้องนำมาใช้ในการปรับเทียบ Harris and Crouse (1993) พบว่าเมื่อกลุ่มตัวอย่างที่เพิ่มขึ้น ความถูกต้องในการปรับเทียบจะมากขึ้น ขนาดกลุ่มตัวอย่างมีผลต่อการปรับเทียบคะแนน (Kim, von Davier and Haberman, 2008 ; Parshall, Houghton and Kromrey, 1995) และตามหลักนี้กลุ่มตัวอย่างที่ใหญ่คือประมาณ 3,000 ให้ผลการปรับเทียบที่มีประสิทธิภาพซึ่งในทางปฏิบัติสำหรับการสอบทั่วไปมักหากลุ่มตัวอย่างขนาดใหญ่เช่นนี้ได้ลำบาก (พิชัย ละแมนชัย, 2538) จึงมีวิธีที่ใช้ในการปรับเทียบตามแนวคิดพื้นฐานได้ 2 รูปแบบคือ รูปแบบการปรับเทียบโดยใช้ทฤษฎีดั้งเดิม (Classical models of equating) และรูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ (Item response theory) แต่รูปแบบการปรับเทียบโดยใช้ทฤษฎีการตอบสนองข้อสอบ จะไม่เหมาะในกรณีขนาดตัวอย่างน้อย (Skaggs, 2005)

ดังนั้น ในปัจจุบันมีงานวิจัยที่พัฒนาวิธีการปรับเทียบเพิ่มมากขึ้นมีทั้งที่พัฒนามาจากวิธีดั้งเดิม (Chen and Holland, 2010 ; Wang and Brennan, 2009) และพัฒนาขึ้นใหม่เพื่อให้เหมาะสมกับการกลุ่มตัวอย่างขนาดเล็ก (Livingston and Kim, 2009, 2010 ; Kim, von Davier and Haberman, 2011) ดังรายละเอียดต่อไปนี้

1. วิธี Circle-arc equating method (Livingston and Kim, 2009, 2010) เป็นการออกแบบการปรับเทียบที่เหมาะในสถานการณ์กลุ่มตัวอย่างขนาดเล็กคล้ายวิธีการปรับเทียบค่าเฉลี่ย (mean equating)

วิธีปรับเทียบค่าเฉลี่ย (Mean equating)

การปรับเทียบค่าเฉลี่ย ใช้สำหรับรูปแบบการเก็บรวบรวมข้อมูลแบบกลุ่มสุ่ม (Random group) เมื่อมีการปรับเทียบคะแนนสังเกตได้ (Observed score equating) เป็นการพิจารณาค่าความยากของแบบสอบชุดที่ 1 ต่างจากชุดที่ 2 คงที่ตลอดตั้งแต่คะแนนน้อยที่สุด จนถึงคะแนนมากที่สุด เช่น คะแนนจากแบบสอบชุดที่ 1 ง่ายกว่าแบบสอบชุดที่ 2 อยู่ 2 คะแนนในกลุ่มผู้ได้คะแนนสูง และง่ายกว่ากลุ่มต่ำ 2 คะแนนเช่นเดียวกัน มีกระบวนการปรับเทียบคะแนน ดังนี้

กำหนดแบบสอบ X เป็นแบบสอบใหม่ที่ต้องการเปรียบเทียบ X เป็นคะแนนที่ได้จากแบบสอบ X ทั้งหมด และ X แทน คะแนนเฉพาะที่ต้องการปรับเทียบ และแบบสอบ Y เป็นแบบสอบชุดเดิม y แทนคะแนนเฉพาะที่ได้จากแบบสอบ Y ทั้งหมด และ y แทนคะแนนเฉพาะแบบสอบ Y

ให้

$\mu(x)$ เป็นค่าเฉลี่ยของคะแนนที่ได้จากแบบสอบ x

$\mu(y)$ เป็นค่าเฉลี่ยของคะแนนที่ได้จากแบบสอบ y

ความหมายของการปรับเทียบคะแนนคือคะแนนจาก 2 แบบสอบมีความแตกต่างจากค่าเฉลี่ยของแต่ละกลุ่มเท่ากัน กำหนดโดย

$x\text{-}\mu(x) = y\text{-}\mu(y)$

จะได้ $m_y(x) = y = x\text{-}\mu(x)+\mu(y)$ (1)

เมื่อ $m_y(x)$ แทน คะแนน x จากแบบสอบ X แปลงไปสู่สเกลของแบบสอบ Y โดยใช้วิธีการปรับเทียบค่าเฉลี่ย

วิธี Circle-arc equating method ใช้การประมาณการเปลี่ยนรูปการปรับเทียบทั้งหมดจากการกำหนดจุดบนเส้นโค้ง การปรับเทียบเส้นโค้งคือโครงสร้างผ่านจุด การเชื่อม 2 จุดปลายเป็นการกำหนดปราศจากการอ้างอิงข้อมูล กำหนดจุดที่ตัดกันของคะแนนเฉลี่ยบนแบบสอบที่ปรับเทียบจุดเหนือกว่า (Upper end-point) คือการตัดกันของคะแนนที่เป็นไปได้สูงสุด จุดต่ำกว่า (Lower end-point) คือการตัดกันของคะแนนที่มีความหมายต่ำสุด การปรับเทียบวิธีนี้แบ่งได้เป็น 2 ประเภท คือ symmetric circle-arc และ simplified circle-arc ทั้งสองรูปแบบจะแตกต่างกันในวิธีการที่สอดคล้องกับแนวโค้งใน 3 จุดเชื่อมโยง และ simplified circle-arc มีโมเดลที่ซับซ้อนกว่า symmetric circle-arc โดยในการปรับเทียบ symmetric circle-arc เส้นโค้งการปรับเทียบคือ ส่วนโค้งของวงกลมที่ประกอบด้วย 2 จุดปลาย (end-points) และเป็นการกำหนดจุดกลาง (middle point) ซึ่งนำ $(x_1, y_1), (x_2, y_2)$ และ (x_3, y_3) แสดงจุดต่ำกว่าจุดกลาง จุดเหนือกว่า ตามลำดับ ของเส้นโค้งเป็นการประมาณเส้นโค้งวงกลม นำ (x_c, y_c) แทนศูนย์กลางของวงกลม และนำ r แทนรัศมีสมการของวงกลมคือ

$$(X-x_c)^2+(Y-y_c)^2=r^2$$

หรือเทียบเท่ากับ

$$|Y-y_c|=\sqrt{r^2-(X-x_c)^2} \quad (2)$$

ถ้าแบบสอบใหม่ (New form) ยากกว่าแบบสอบอ้างอิง (Reference form) จุดกลางจะวางเหนือเส้นเชื่อม จุดต่ำกว่า และจุดเหนือกว่า ดังนั้น ศูนย์กลางของวงกลมจะต่ำกว่าส่วนโค้ง สำหรับจุดทั้งหมด (x, y) บนส่วนโค้ง $Y > y_c$ ดังนั้น $|Y-y_c| = Y-y_c$ และสูตรสำหรับเส้นโค้ง คือ

$$Y = y_c + \sqrt{r^2-(X-x_c)^2} \quad (3)$$

ถ้าแบบสอบใหม่ง่ายกว่า แบบสอบอ้างอิงจุดกลางจะวางใต้เส้นเชื่อมจุดต่ำกว่าและจุดเหนือกว่า ดังนั้น ศูนย์กลางของวงกลมจะเหนือส่วนโค้ง สำหรับจุดทั้งหมด (x,y) บนส่วนโค้ง $Y < y_c$ ดังนั้น $|Y-y_c| = y_c-Y$ และสูตรสำหรับเส้นโค้ง คือ

$$Y = y_c - \sqrt{r^2-(X-x_c)^2} \quad (4)$$

หลักการตัดสินใจในการใช้สูตร (3) ถ้า $y_2 > y_c$ และสูตร (4) ถ้า $y_2 < y_c$ สูตรสำหรับ x_c และ y_c

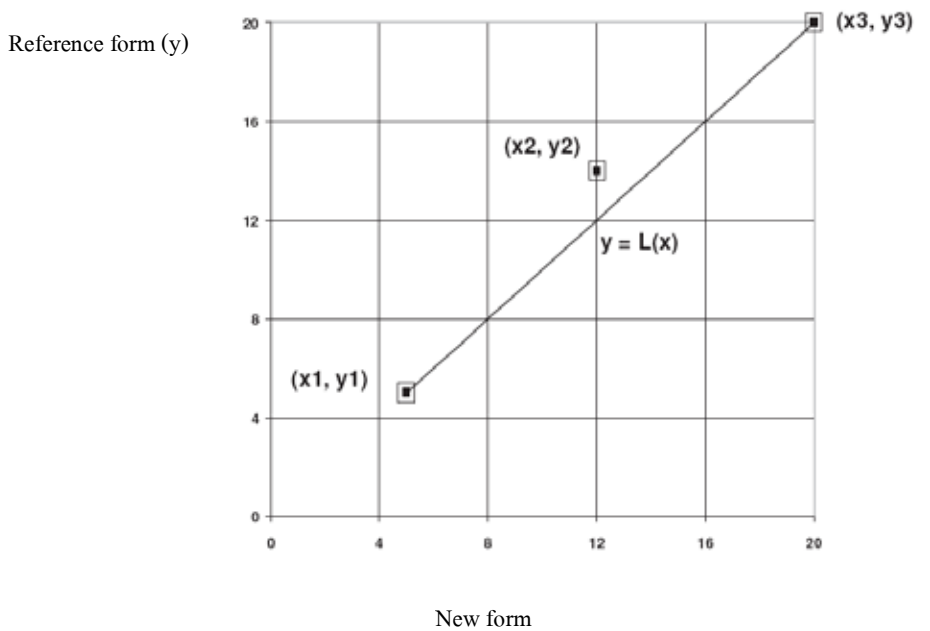
$$x_c = \frac{(x_1^2+y_1^2)(y_3-y_2)+(x_2^2+y_2^2)(y_1-y_3)+(x_3^2+y_3^2)(y_2-y_1)}{2[x_1(y_3-y_2)+x_2(y_1-y_3)+x_3(y_2-y_1)]} \quad (5)$$

$$y_c = \frac{(x_1^2+y_1^2)(x_3-x_2)+(x_2^2+y_2^2)(x_1-x_3)+(x_3^2+y_3^2)(x_2-x_1)}{2[y_1(x_3-x_2)+y_2(x_1-x_3)+y_3(x_2-x_1)]} \quad (6)$$

สูตรสำหรับ r^2

$$r^2 = (x_1 - x_c)^2 + (y_1 - y_c)^2 \qquad (7)$$

ในการปรับเทียบ simplified circle-arc เส้นโค้งการปรับเทียบคือ การเปลี่ยน 3 จุดโดยสลายฟังก์ชั่นการปรับเทียบ (Equating function) ไปยังส่วนประกอบเส้นตรง (Linear component) และส่วนประกอบเส้นโค้ง (Curvilinear component) ส่วนประกอบเส้นตรงคือ เส้นที่เชื่อมสองจุดปลาย ส่วนประกอบเส้นโค้งคือ ความเบี่ยงเบนตั้งฉากของโค้งการปรับเทียบจากเส้น ถูกประมาณโดยสอดคล้องกับส่วนโค้งวงกลม จาก 3 จุดที่เปลี่ยนรูป แนวคิดพื้นฐานของวิธีนี้มาจากการนำเสนอโดย Divig (1987) แนวคิดของเขาจำกัดการประมาณเส้นโค้งการปรับเทียบผ่าน 2 จุดปลาย และกำหนดจุดกลาง ในวิธีของ Divig จุดปลายกำหนดโดยคะแนนที่เป็นได้สูงสุดและต่ำสุด บนแบบสอบในการปรับเทียบ จุดกลาง กำหนดโดยคะแนนเฉลี่ยการประมาณการแปลงการปรับเทียบประกอบด้วยลูกบาศก์ผ่าน 3 จุดกับความชันของส่วนโค้งที่จุดกลาง กำหนดโดยสัดส่วนของส่วนเบี่ยงเบนมาตรฐาน (Standard divition) สำหรับรูปภาพ 1-3 แสดงให้เห็นวิธีกับตัวอย่างสมมุติ ในแบบสอบใหม่ของแบบสอบ 20 ข้อ 4 ตัวเลือก ในการปรับเทียบไปยัง แบบสอบอ้างอิงประกอบด้วย 20 ข้อ 4 ตัวเลือก 3 จุดในการกำหนดเส้นโค้งวงกลมบ่งชี้โดย กล่องสี่เหลี่ยม จุดเหนือกว่าของส่วนโค้งกำหนดโดยคะแนนเป็นไปได้สูงสุด คือ (20,20) จุดต่ำกว่าของเส้นโค้งกำหนดโดยคะแนนเป็นไปได้ คือ (5, 5) จุดกลาง กำหนดจากข้อมูลเป็น (12, 14) สรุปได้ว่า แบบสอบใหม่ยากกว่าแบบสอบอ้างอิง

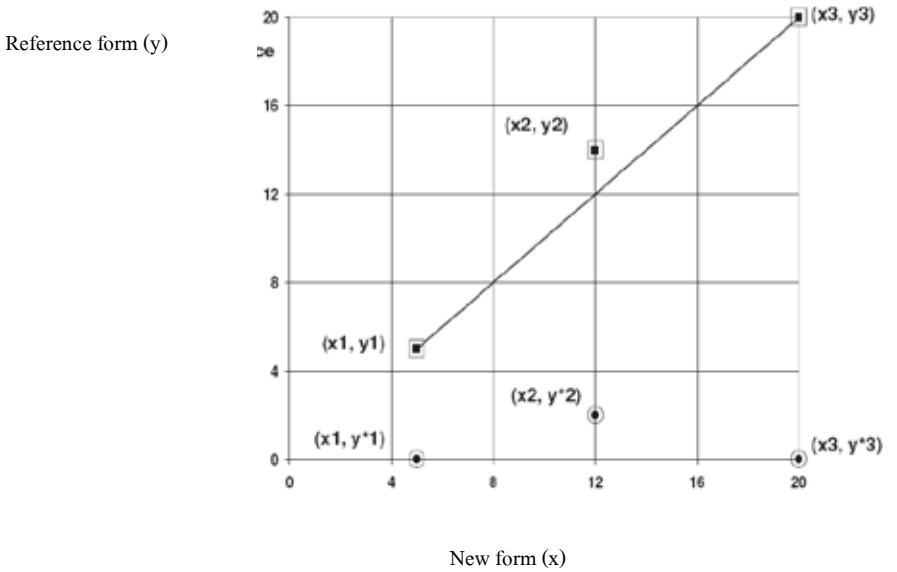


ภาพที่ 1 เส้นตรงของฟังก์ชั่นการปรับเทียบ
ที่มา: Livingston and Kim (2009)

ภาพที่ 1 แสดง 3 จุดกำหนดเส้นโค้งการปรับเทียบ หมายถึง จุดต่ำกว่า(x_1, y_1), จุดเหนือกว่า (x_3, y_3) และจุดกลาง กำหนดโดยข้อมูล (x_2, y_2) แสดงการเชื่อมเส้นตรง 2 จุดปลาย เส้นนี้เป็นส่วนประกอบเส้นตรงของฟังก์ชั่นการปรับเทียบ เรียกว่า L(x)

$$L(x) = y_1 + \frac{y_3 - y_1}{x_3 - x_1}(x - x_1) \qquad (8)$$

ถ้าแบบสอบใหม่และแบบสอบอ้างอิงเหมือนกันในความยาวแบบสอบ รูปแบบ และระบบการให้คะแนน จะเป็นเส้นเอกลักษณ์

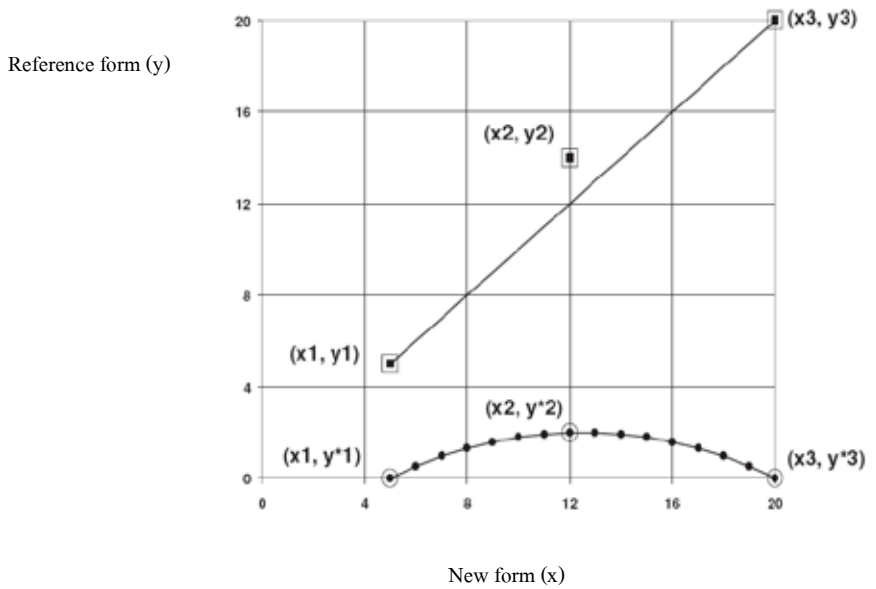


ภาพที่ 2 การแปลง 3 จุด
ที่มา: Livingston and Kim (2009)

ภาพที่ 2 3 จุดเป็นการแปลงโดยลบความสูงของเส้น L(x) จะใช้เครื่องหมาย y^* ตัวแทนความสูงของคะแนนแปลง

$$y^* = y - L(x) \qquad (9)$$

เนื่องจาก 2 จุดปลายของเส้นโค้งบนเส้น L(x) การแปลงขั้นนี้ ค่า y เป็น 0 ดังนั้น $y_1^* = y_3^* = 0$ ความสูงของการแปลง จุดกลาง (x_2, y_2^*) ใกล้พื้นล่างของภาพที่ 2 คือเท่าเทียมกับระยะตั้งฉากของจุดกลางดั้งเดิม (x_2, y_2) บนการเชื่อมเส้น 2 จุดปลาย ระยะขึ้นกับตัวบ่งชี้การปรับเทียบกลุ่มตัวอย่างขนาดเล็ก เกี่ยวกับความยากของแบบสอบใหม่ และแบบสอบอ้างอิง ถ้าตัวบ่งชี้การปรับเทียบกลุ่มตัวอย่างขนาดเล็กแบบสอบใหม่ ยากกว่าแบบสอบอ้างอิงจุดกลางจะอยู่เหนือ เชื่อมเส้นจุดปลาย และ y_2^* จะเป็นบวก ถ้าตัวบ่งชี้การปรับเทียบกลุ่มตัวอย่างขนาดเล็กแบบสอบใหม่ง่ายกว่าแบบสอบอ้างอิง จุดกลางจะอยู่ใต้ การเชื่อมเส้นจุดปลาย และ y_2^* จะเป็นลบ ในตัวอย่างนี้ 3 จุดที่แปลงคือ (5, 0), (12, 2) และ (20, 0)



ภาพที่ 3 กำหนดเส้นโค้งวงกลม
ที่มา: Livingston and Kim (2009)

ภาพที่ 3 สามจุดที่แปลงใช้กำหนดเส้นโค้งวงกลมและกำหนดส่วนโค้งค่าของ y^* ที่แต่ละคะแนนดิบแบบสอบใหม่ ถ้า r แสดงรัศมีของวงกลมและความเสมอกันของศูนย์กลางคือ (x_c, y_c) สมการของวงกลมคือ

$$(\mathrm{X} - x_c)^2 + (\mathrm{Y} - y_c)^2 = r^2 \qquad (10)$$

ถ้า y_2^* คือ ค่าบวก สูตรสำหรับเส้นโค้งคือ

$$y^* = y_c + \sqrt{r^2 - (\mathrm{X} - x_c)^2} \qquad (11)$$

ถ้า y_2^* คือ ค่าลบ สูตรสำหรับเส้นโค้งคือ

$$y^* = y_c - \sqrt{r^2 - (\mathrm{X} - x_c)^2} \qquad (12)$$

ในระดับเดียวกันของจุดศูนย์กลางคือ

$$x_c = \frac{\left(x_3^2 - x_1^2\right)}{2(x_3 - x_1)} \qquad (13)$$

$$y_c = \frac{\left(x_1^2\right)(x_3 - x_2) - \left(x_2^2 + \left(y_2^*\right)^2\right)(x_3 - x_1) + \left(x_3^2\right)(x_2 - x_1)}{2\left[y_2^*(x_1 - x_3)\right]} \qquad (14)$$

และรัศมี คือ

$$r = \sqrt{(x_c - x_1)^2 + (y_c)^2} \qquad (15)$$

2. วิธี Synthetic Linking Function (Kim, von Davier and Haberman, 2011) เป็นวิธีการเชื่อมโยงแบบสอบกับกลุ่มตัวอย่างขนาดเล็ก ในกลุ่มผู้สอบที่มีความสามารถแตกต่างกัน การใช้ฟังก์ชั่นเอกลักษณ์ เป็นทางเลือกในการปรับเทียบกลุ่มตัวอย่างขนาดเล็ก ที่เหมาะสมสำหรับการเชื่อมโยงเมื่อแบบสอบคู่ขนานอย่างสมบูรณ์ การใช้ฟังก์ชั่นเอกลักษณ์ คือ

$$ID_y(X) = X \qquad (16)$$

เมื่อ x คือคะแนนดิบของแบบสอบใหม่ (new form) x เป็นการวางบนสเกลคะแนนดิบของแบบสอบเดิม (old form) y ในลักษณะเส้นตรง ความคลาดเคลื่อนในการปรับเทียบอย่างสุ่ม คือ 0 สำหรับการเชื่อมโยงเอกลักษณ์ เพราะการปรับเทียบคะแนนได้มาโดยผ่านวิธีการจำลองเชิงกำหนด อย่างไรก็ตาม การใช้เอกลักษณ์สามารถเพิ่มความคลาดเคลื่อนเชิงระบบ เช่น ความลำเอียงทั่วไปฟังก์ชั่นเอกลักษณ์ คือความเหมาะสมเมื่อลักษณะเฉพาะแบบสอบถูกอธิบายได้ดี และสองแบบสอบคู่ขนานกันในความยากและเนื้อหา นำเสนอวิธีทางเลือกที่เรียกว่า synthetic function method สำหรับแสดงการปรับเทียบกับกลุ่มตัวอย่างขนาดเล็ก synthetic function คือความจำเป็นในการยอมรับระหว่างการใช้ฟังก์ชั่นการปรับเทียบตัวอย่าง และการใช้โดยไม่ปรับเทียบ (ผ่านการใช้ฟังก์ชั่นเอกลักษณ์) โดยรวมการใช้ระบบน้ำหนักที่กำหนด

$$Synthetic_y(x) = w \times e_y(x) + (1-w) \times ID_y(x) \qquad (17)$$

W คือ น้ำหนักระหว่าง 0 และ 1

X คือ คะแนนดิบในแบบสอบ x

e_y คือ ฟังก์ชั่นการปรับเทียบตัวอย่าง เช่น chain linear, Tucker หรือ Levine method

ID_y คือ ฟังก์ชั่นเอกลักษณ์

ขณะที่การอภิปรายก่อนหน้า การปรับเทียบกับตัวอย่างขนาดเล็กจะนำไปสู่ความคลาดเคลื่อนอย่างสุ่ม(Sampling error) สะท้อนในความคลาดเคลื่อนมาตรฐานของการปรับเทียบ (Standard error of equating) การให้ ความคลาดเคลื่อนมาตรฐานของการปรับเทียบสำหรับการเชื่อมโยงเอกลักษณ์ คือ 0 ความคลาดเคลื่อนมาตรฐานของการปรับเทียบสำหรับ synthetic linking สามารถลดความสำคัญในการเปรียบทียบฟังก์ชั่นการปรับเทียบกลุ่มตัวอย่าง การสาธิตความคลาดเคลื่อนในการปรับเทียบสามารถลดเมื่อใช้ synthetic function

$$\begin{aligned} \text{Var}\left(Syn_y(x)\right) &= w^2 Var\left[e_y(x)\right] + (1-w) \times ID_y(x) \\ &\quad + 2w(1-w)Cov\left[e_y(x), ID_y(x)\right] \\ &= w^2 Var\left[e_y(x)\right] \end{aligned} \qquad (18)$$

ดังนั้น

$$\text{SEE}\left[Syn_y(x)\right] = w \times SEE\left[e_y(x)\right] \qquad (19)$$

จากสมการ (19) เราจะเห็นว่าเมื่อน้ำหนักเท่ากันให้เอกลักษณ์และฟังก์ชั่นการปรับเทียบตัวอย่าง ความคลาดเคลื่อนในการปรับเทียบลดลงโดยครึ่งหนึ่ง คล้ายสมการ (20) แสดงความลำเอียงของ synthetic function แนะนำโดยการใช้ฟังก์ชั่นเอกลักษณ์สามารถลดภายใต้ข้อตกลงเบื้องต้น ฟังก์ชั่นการปรับเทียบตัวอย่างไม่ลำเอียงหรือความลำเอียงน้อยเมื่อเปรียบเทียบฟังก์ชั่นเอกลักษณ์ สำหรับข้อจำกัดของวิธี synthetic function คือ แบบสอบทั้งสองควรมีจำนวนข้อสอบที่เท่ากันและคู่ขนานกัน

$$\mu[Synthetic_y(x)] = w \times Mean[e_y(x)] + (1-w) \times Mean[ID_y(x)] \quad (20)$$

วิธีการปรับเทียบที่พัฒนาขึ้นใหม่มีนักวิจัยได้ทำการตรวจสอบคุณภาพของการปรับเทียบกับวิธีที่เดิมที่ใช้ในการปรับเทียบกับกลุ่มตัวอย่างขนาดเล็กโดยมีการตรวจสอบคุณภาพการปรับเทียบโดยใช้สถิติ ตามที่ Kim and Livingston (2010) ศึกษาเกี่ยวกับการปรับเทียบคะแนนในกลุ่มตัวอย่างขนาดเล็กของผู้เข้าสอบโดยใช้วิธี chained equipercentile equating of smoothed distributions, chained linear equating, chained mean equating, the symmetric circle-arc method, and the simplified circle-arc method จากผู้เข้าสอบ 100, 50, 25 และ 10 ความถูกต้องได้ถูกอธิบายในความแตกต่างของ รากของความแตกต่างกำลังสองเฉลี่ย (root mean squared) จากเกณฑ์การปรับเทียบ โดยภาพรวมผลการปรับเทียบค่าเฉลี่ย (chained mean equating) ให้ความถูกต้องสูงสำหรับคะแนนต่ำ แต่สำหรับวิธี circle-arc ทั้งสองให้ผลความถูกต้องสูง ส่วนหนึ่งสูงกว่าครึ่งจากการแจกแจงของคะแนน ความแตกต่างในการปรับเทียบที่ถูกต้องระหว่าง สองวิธีของ circle-arc methods แตกต่างกันเล็กน้อย และ Livingston and Kim (2010) ศึกษาเกี่ยวกับการปรับเทียบ 5 วิธีคือ equipercentile equating of smoothed distributions, linear equating, mean equating, symmetric circle-arc equating, และ simplified circle-arc equating ในการออกแบบกลุ่มสุ่ม (random groups design) กับกลุ่มตัวอย่างของผู้ทำแบบสอบ 400, 200, 100 และ 50 คน ใช้ 6 แบบสอบคือ Social studies, Elementary education, English language, Speech-language pathology, Educational leadership และ Fundamental subjects ความถูกต้องในการปรับเทียบถูกบ่งชี้โดยรากของความแตกต่างกำลังสองเฉลี่ย (Root mean squared) พบว่าวิธี circle-arc ให้ผลลัพธ์ที่ถูกต้องที่สุดสำหรับการสำรวจขนาดตัวอย่างทั้งหมด โดยเฉพาะในการเพิ่มของการกระจายคะแนน ในความถูกต้องการปรับเทียบระหว่างวิธี circle-arc ทั้ง 2 แตกต่างเล็กน้อย

ขั้นตอนการปรับเทียบคะแนน

จากวิธีการปรับเทียบคะแนนข้างต้นนำไปสู่ประโยชน์กับการจัดการศึกษา โดยมีขั้นตอนในการดำเนินการทั่วไปเพื่อนำวิธีการปรับเทียบแบบใหม่ไปประยุกต์ใช้ดังต่อไปนี้

1. กำหนดจุดมุ่งหมายในการปรับเทียบคะแนนว่าจะนำผลจากการปรับเทียบคะแนนเพื่อดูพัฒนาการเปลี่ยนแปลงทางการศึกษา หรือเพื่อเทียบความสามารถผู้สอบเป็นข้อมูลใช้ตัดสินผลร่วมกัน ใช้แทนกันได้

2. สร้างแบบสอบหลายฉบับ แต่ละฉบับวัดเนื้อหาเดียวกัน และสร้างตามแบบแผนการออกข้อสอบเดียวกัน (item specification) เป็นแบบสอบที่มีลักษณะของความเป็นคู่ขนานในด้านเนื้อ

เรื่อง โครงสร้าง รูปแบบ ชนิดของข้อสอบ และเวลาที่ใช้

3. เลือกวิธีเก็บรวบรวมข้อมูล มีรูปแบบกลุ่มสุ่ม (Random group design) โดยสุ่มกลุ่มตัวอย่างจากประชากรกลุ่มเดียวกัน แต่ละกลุ่มทำแบบสอบคนละชุด รูปแบบผู้สอบกลุ่มเดียว (Single Group Design) ผู้สอบกลุ่มเดียวทำแบบสอบทั้งสองชุด รูปแบบกลุ่มเดียวที่ได้รับการจัดให้สมดุล (Single group design with counterbalancing) แบ่งผู้สอบแต่ละกลุ่ม แบบสอบแต่ละชุดเป็น 2 ส่วน ให้ผู้สอบกลุ่มย่อยแรกสอบแบบสอบชุดที่ 1 ตอนแรก ตามด้วยแบบสอบชุดที 2 ตอนหลัง และผู้สอบกลุ่มย่อยที่สอง ทำแบบสอบชุดที่ 2 ตอนแรก ตามด้วยแบบสอบชุดที่ 1ตอนหลัง และรูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้แบบสอบร่วม (Common-item non-equivalent groups design) ผู้สอบต่างกลุ่มประชากรทำแบบสอบคนละชุด และผู้สอบทุกคนทำแบบสอบร่วมที่อาจจะแทรกภายในแบบสอบทั้ง 2 ฉบับ (Internal common-item) หรือแยกออกจากแบบสอบ (external common-item)

4. เก็บรวบรวมข้อมูลตามรูปแบบที่กำหนดไว้ จากนั้นเลือกนิยามเชิงปฏิบัติการของการปรับเทียบคะแนน เพื่อตัดสินใจว่าจะใช้วิธีการปรับเทียบคะแนนเชิงเส้นตรง (Linear equating methods) หรือวิธีการปรับเทียบที่ไม่เป็นเชิงเส้นตรง (Nonlinear equating methods)

5. เลือกวิธีประมาณค่าสถิติที่ใช้วิเคราะห์เลือกให้สอดคล้องกับนิยามเชิงปฏิบัติการที่กำหนด ประเมินผลการปรับเทียบคะแนน มีเกณฑ์ที่ใช้ดังนี้ ความเสมอภาค (Weak Equity) ของ Divgi และ Yen ที่พิจารณาจากความเท่าเทียมกันของการแจกแจงตามเงื่อนไขของคะแนนที่ได้จากแบบสอบต่างฉบับหลังจากปรับเทียบแล้ว ดัชนีสำหรับการเปลี่ยนแปลงคะแนนของ Angoff ความคลาดเคลื่อนมาตรฐานของ Angoff เป็นการวิเคราะห์เพื่อประมาณความคลาดเคลื่อนของการปรับเทียบจากการสุ่มตัวอย่าง เป็นต้น

สรุป

การปรับเทียบคะแนนแบบสอบถ้าตัวอย่างขนาดใหญ่ความสัมพันธ์การปรับเทียบในตัวอย่างจะเป็นตัวแทนที่ถูกต้องในความสัมพันธ์การปรับเทียบในประชากร เมื่อขนาดตัวอย่างเล็กฟังก์ชั่นการปรับเทียบที่คำนวณจะแตกต่างจากฟังก์ชั่นปรับเทียบของประชากร ทั้งความคลาดเคลื่อนในการสุ่มและความลำเอียง มีอิทธิพลต่อคุณภาพการปรับเทียบ การวิจัยที่ผ่านมามีการหาวิธีที่เหมาะสม (Livingston, 1993 ; Kim, Davier and Haberman, 2006 ; von Davier, 2007) คือ วิธี Circle-arc equating method และวิธี Synthetic Linking Function เป็นวิธีที่ทำให้เกิดความคลาดเคลื่อนน้อยที่สุดเมื่อเทียบกับวิธีอื่นๆ

เอกสารอ้างอิง

พิชัย ละแมนชัย. (2538). *ขนาดกลุ่มตัวอย่างขั้นต่ำสำหรับการปรับเทียบคะแนนในแนวระดับตามแนวทฤษฎีการตอบสนองรายข้อ ระหว่างแบบสอบที่มีแบบแผนกาปรับเทียบและความยาวแบบสอบแตกต่างกัน*. วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต สาขาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่*. พิมพ์ครั้งที่ 3. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Chen, H. and Holland, P. (2010). New equating methods and their relationships with Levine Observed Score Linear Equating under the Kernel Equating Framework. *psychometrika, 75*(3): 542-557.

Hambleton, R.K. and Swaminatan, H. (1985). *Item response theory: Principles and application*. 2nd ed. Boston: Kluwer Nijhoff Publishing.

Harris, D.J. and Crouse (1993) A study of criteria used in equating. *Applied Measurement in Education, 6*(3): 195-240.

Kim, S., von Davier, A. A. and Haberman, S. (2008). Small sample equating using a synthetic Linking Function. *Journal of Educational Measurement, 45*(4). 325-342.

Kim, S., von Davier, A. A. and Haberman, S. (2011). Practical Application of a Synthetic linking function on small-sample equating. *Applied Measurement in Education, 24*(2): 95-114.

Kolen, M. J. (1999). *Equating of tests. Advances in Measurement in Educational Research and Assessment*. Edited by Masters, G. N. and Keeves, J. P. p.165-175. Oxford: Elsevier Science.

Kolen, M.J. and Brennan, R.J. (2014). *Test equating, scaling and linking: Methods and practices*. New York: Springer-Verlag.

Livingston, S. A. and Kim S. (2010). Random-groups equating with samples of 50 to 400 test takers. *Journal of Educational Measurement, 47*(2): 175-185.

Livingston, S. A. and Kim, S. (2009). The Circle-arc method for equating in small samples. *Journal of Educational Measurement, 46*(3): 330-343.

Parshall, C. G., Houghton, D. B. P., and Kromrey, J. D. (1995). Equating error and statistical bias in small sample linear equating. *Journal of Educational Measurement, 32*(1): 37-54.

Petersen, N. S. (2007). *Equating: Best practices and challenges to best practices*. In N. J. Dorans, M. Pommerich and P. W. Holland (Eds), Linking and aligning scores and scales. New York: Springer.

Petersen, N.S., Kolen, M.J. and Hoover, H.D. (1989). *Scaling, norming and equating. republication*. In R.L. Linn(Ed.), Educational Measurement. Washington, D.C.: American Council on Education.

Skaggs, G. (2005). Accuracy of random groups equating with very small samples. *Journal of Educational Measurement, 42*(4): 309-330.

Wang, Y. and Brennan, R. L. (2009). A modified frequency estimation equating method for the common-item nonequivalent groups design. *Applied Psychological Measurement, 33*(2): 118-132.