

แนวคิดและการประยุกต์ใช้กระบวนการปรับเทียบคะแนนตามทฤษฎี
การตอบสนองข้อสอบแบบพหุมิติ
CONCEPT AND APPLICATION OF MULTIDIMENSIONAL IRT EQUATING PROCEDURE

พนิดา พานิชวัฒน์¹
PANIDA PANIDVADTANA¹

บทคัดย่อ

การปรับเทียบคะแนน (Equating) เป็นกระบวนการสถิติที่ใช้ในการแปลงคะแนนบนสเกลของแบบสอบฉบับหนึ่งไปยังสเกลของแบบสอบอีกฉบับ เพื่อให้สามารถเปรียบเทียบคะแนนระหว่างแบบสอบได้อย่างถูกต้อง วิธีการปรับเทียบคะแนนแบ่งเป็น 2 ประเภทหลัก คือ วิธีการปรับเทียบคะแนนแบบดั้งเดิม (Classical Equating Methods) และวิธีการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Equating Methods: IRT Equating Methods) ซึ่งวิธีการปรับเทียบทั้ง 2 ประเภทถูกพัฒนาในเชิงทฤษฎีและประยุกต์ใช้กับการทดสอบจริงและข้อมูลที่จำลองขึ้นอย่างต่อเนื่อง เมื่อพิจารณาการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบที่นิยมนำมาใช้อย่างแพร่หลายนั้น เริ่มต้นจากการพัฒนาวิธีการปรับเทียบตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional IRT Equating Methods: UIRT Equating Methods) เพื่อแปลงคะแนนระหว่างแบบสอบที่มุ่งวัดคุณลักษณะเดียว ซึ่งประสิทธิภาพของวิธีการดังกล่าวขึ้นอยู่กับความเป็นไปตามข้อตกลงเบื้องต้นตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ ภายหลังมีการพัฒนาแบบสอบพหุมิติและนิยมนำมาใช้อย่างแพร่หลาย นักวิชาการจึงพัฒนากระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Multidimensional IRT Equating Procedures :MIRT Equating Procedures) ที่สอดคล้องกับโครงสร้างของแบบสอบที่มุ่งวัดหลายคุณลักษณะ อีกทั้งยังมีการวิจัยและประยุกต์ใช้วิธีการดังกล่าวกับการทดสอบจริงและข้อมูลที่จำลองขึ้น สำหรับข้อสอบที่ตรวจให้คะแนนสองค่าและหลายค่า รวมถึงแบบสอบรูปแบบผสม

คำสำคัญ : 1. การปรับเทียบคะแนน 2. ทฤษฎีการตอบสนองข้อสอบ 3.แบบสอบพหุมิติ 4. การแปลงคะแนน

ABSTRACT

Equating is a statistical procedure that transforms scores on scale of a test to scale of another test in order to compare score between the tests correctly. The two main categories of equating methods are classical equating methods and item response theory

¹ ภาควิชาการประเมินและการวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยรามคำแหง

equating methods. The two method are continually developed in theoretical aspect and applied with real testing and simulation data. When considering the item response theory equating methods that are widely used, Unidimensional IRT equating methods (UIRT equating methods) were developed at the beginning to transform score between tests measuring one attribute. Efficiency of UIRT equating methods depends on concordance with assumptions of UIRT theory. Later, multidimensional tests were developed and widely applied, scholars developed multidimensional IRT equating procedures (MIRT Equating Procedures) that conform to the structure of tests measuring multi-attribute. Moreover, the procedures were used in research and applied real testing and simulation data for dichotomous and polytomous items including mixed-format tests

Keywords : 1. Equating 2. Item Response Theory 3. Multidimensional tests
4. Score Transformation

ความนำ

การปรับเทียบคะแนนเป็นกระบวนการทางสถิติที่แปลงคะแนนบนสเกลของแบบสอบฉบับหนึ่งไปยังสเกลของแบบสอบอีกฉบับที่มุ่งวัดโครงสร้างหรือเนื้อหาเดียวกัน จึงทำให้คะแนนที่ผ่านการปรับเทียบคะแนนสามารถเปรียบเทียบกันได้อย่างถูกต้อง อีกทั้งยังสามารถใช้การปรับเทียบคะแนนในการพัฒนาค้างข้อสอบ โดยก่อนการเพิ่มข้อสอบเข้าไปในคลังข้อสอบต้องเปรียบเทียบค่าพารามิเตอร์ของข้อสอบที่จะเพิ่มเข้าไปให้อยู่สเกลเดียวกับค่าในคลังข้อสอบด้วยการใช้ข้อสอบร่วม นอกจากนี้ยังสามารถใช้การปรับเทียบคะแนนเพื่อศึกษาพัฒนาการของผู้เรียนระหว่างระดับชั้นต่าง ๆ ได้อีกด้วย โดยการปรับเทียบคะแนนที่มีประสิทธิภาพต้องอยู่ภายใต้เงื่อนไขเหล่านี้ (1) โครงสร้างเท่าเทียมกัน (Equal structure) (2) ความเที่ยงเท่าเทียมกัน (Equal reliability) (3) ความสมมาตร (Symmetry) กล่าวคือ ฟังก์ชันการปรับเทียบคะแนนแบบสอบ Y ไปยังแบบสอบ X ผกผันกับฟังก์ชันการปรับเทียบคะแนนแบบสอบ X ไปยังแบบสอบ Y (4) ความเสมอภาค (Equity) กล่าวคือ คะแนนที่ถูกแปลง (Converted score) จากแบบสอบ X และแบบสอบ Y มีค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน และลักษณะการแจกแจงเหมือนกัน และ (5) ความไม่แปรปรวนของประชากร (Population invariance) กล่าวคือ ฟังก์ชันการปรับเทียบคะแนนไม่แปรผันไปตามกลุ่มประชากร (Kolen & Brennan, 2004; Holland, 2007; ศิริชัย กาญจนวาสี, 2555)

การปรับเทียบคะแนนที่มีประสิทธิภาพจำเป็นต้องพิจารณาการออกแบบการเก็บรวบรวมข้อมูล และวิธีการปรับเทียบคะแนน การออกแบบการเก็บรวบรวมข้อมูลเป็นการพิจารณาลักษณะของแบบสอบ ผู้สอบ และการจัดการทดสอบ ซึ่งแบ่งออกเป็น 4 รูปแบบ (Kolen & Brennan, 2004) ได้แก่ (1) รูปแบบผู้สอบกลุ่มสุ่ม (Random group) ผู้สอบแต่ละคนทำแบบสอบฉบับเดียวในช่วงเวลาเดียวกัน ซึ่งผู้สอบจะถูกสุ่มให้แบบสอบแต่ละรูป เช่น ผู้สอบคนที่ 1 ทำแบบฉบับ A และผู้สอบคนที่ 2 ทำแบบสอบฉบับที่ 2 สลับไปเรื่อย ๆ โดยเทียบเคียงกับรูปแบบผู้สอบกลุ่มเท่าเทียมกันอย่างสุ่ม (Randomly equivalent groups) (2) รูปแบบผู้สอบกลุ่มเดียว (single group) ผู้สอบแต่ละคนทำแบบสอบสองฉบับ

ในเวลาเดียวกัน เช่น ผู้สอบทำแบบสอบฉบับ ก ก่อนแล้วจึงทำแบบสอบฉบับ ข ซึ่งผู้สอบอาจเกิดความเมื่อยล้าและอาจเกิดผลการเรียงลำดับของแบบสอบ (order effect) (3) รูปแบบผู้สอบกลุ่มเดียวที่ได้รับการจัดให้สมดุล (Single group with counterbalancing design) ผู้สอบจะถูกแบ่งเป็นกลุ่มย่อย โดยผู้สอบแต่ละกลุ่มทำแบบสอบ 2 ฉบับที่มีการจัดลำดับให้สมดุล เช่น ผู้สอบกลุ่ม 1 ทำแบบสอบฉบับ ก ก่อนแล้วทำแบบสอบฉบับ ข ส่วนผู้สอบกลุ่ม 2 ทำแบบสอบฉบับ ข ก่อนแล้วทำแบบสอบฉบับ ก และ (4) รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม (Non-equivalent group with anchor test: NEAT) ผู้สอบแต่ละคน/แต่ละกลุ่มทำแบบสอบฉบับเดียว ซึ่งผู้สอบแต่ละกลุ่มไม่จำเป็นต้องทำการทดสอบพร้อมกันหรือวันเดียวกัน เช่น ปีแรกผู้สอบกลุ่ม 1 ทำแบบสอบฉบับ ก ส่วนปีที่สองผู้สอบกลุ่ม 2 ทำแบบสอบฉบับ ข โดยปรับเทียบความแตกต่างระหว่างผู้สอบโดยใช้ข้อสอบร่วม (Common items)

วิธีการปรับเทียบคะแนนแบ่งออกเป็น 2 ประเภทหลัก คือ วิธีการปรับเทียบคะแนนแบบดั้งเดิม (Classical equating methods) และวิธีการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory Equating Methods: IRT Equating Methods) ซึ่งการปรับเทียบคะแนนทั้งสองวิธีได้รับการพัฒนาเชิงทฤษฎีและการประยุกต์ใช้กับบริบทต่าง ๆ อย่างต่อเนื่อง โดยในช่วงแรกวิธีการปรับเทียบคะแนนด้วยโมเดล IRT ถูกพัฒนาสำหรับแบบสอบเอกมิติ เรียกว่า วิธีการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (Unidimensional IRT equating methods: UIRT equating methods) ซึ่งนิยมนำใช้ในการปรับเทียบคะแนนระหว่างแบบสอบที่มุ่งวัดคุณลักษณะเดียว เนื่องจากแบบสอบที่นำมาปรับเทียบคะแนนสามารถมีความยากแตกต่างกันได้มากกว่าวิธีการปรับเทียบคะแนนเชิงเส้น และสามารถแก้จุดอ่อนของการปรับเทียบคะแนนแบบดั้งเดิมทั้งด้านความเสมอภาค ความสมมาตร และความไม่ปรวนของประชากร อีกทั้งวิธีการปรับเทียบดังกล่าวยังสามารถปรับเทียบได้ทั้งคะแนนสังเกตได้ (Observed score) และคะแนนจริง (True score) (Kolen & Brennan, 2004; ศิริชัย กาญจนวาสี, 2555)

ภายหลังมีการพัฒนาแบบสอบพหุมิติที่มุ่งวัดหลายคุณลักษณะ (Attribute) หรือความสามารถ (Ability) และถูกนำมาใช้อย่างกว้างขวางในการทดสอบ การปรับเทียบคะแนนแบบสอบพหุมิติด้วยโมเดล UIRT ย่อมเป็นการละเมิดข้อตกลงเบื้องต้นสำคัญของโมเดลดังกล่าวที่ระบุว่าแบบสอบต้องมุ่งวัดคุณลักษณะเดียว ซึ่งส่งผลให้เกิดความคลาดเคลื่อนในการปรับเทียบคะแนน จึงมีการพัฒนากระบวนการปรับเทียบคะแนนตามทฤษฎี การตอบสนองข้อสอบแบบพหุมิติ (Full MIRT Score Equating Procedure: FMIRT) ซึ่งประกอบด้วย (1) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูปแบบ (Full MIRT observed score equating procedure: MOSE) และ (2) กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (Unidimensional Approximation to MIRT Equating Procedure) ซึ่งประกอบด้วยกระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (Unidimensional Approximation to MIRT Observed Score Equating Procedure: AOSE)

และกระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (Unidimensional Approximation to MIRT True Score Equating Procedure: ATSE)

การเปรียบเทียบประสิทธิภาพของการปรับเทียบคะแนนแบบสอบพหุมิติด้วยโมเดล UIRT และ MIRT สำหรับข้อสอบที่ตรวจให้คะแนนสองค่าภายใต้รูปแบบผู้สอบกลุ่มเท่าเทียม และแบบสอบรูปแบบผสมภายใต้รูปแบบผู้สอบกลุ่มสุ่ม (Brossman, 2010; Lee & Brossman, 2012) พบว่า การปรับเทียบคะแนนด้วยโมเดล MIRT มีประสิทธิภาพมากกว่าการปรับเทียบคะแนนด้วยโมเดล UIRT ดังนั้นกระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (FMIRT) จึงเป็นวิธีการที่ทันสมัยและมีความสำคัญต่อการปรับเทียบคะแนนระหว่างแบบสอบที่มุ่งวัดหลายคุณลักษณะ ผู้เขียนจึงขอเสนอแนวคิดและการประยุกต์ใช้การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติใน 2 ประเด็นหลัก คือ มโนทัศน์ของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ และความก้าวหน้าของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติทั้งทางทฤษฎีและทางการวิจัย ซึ่งนักวิจัยยังมีการพัฒนาการปรับเทียบคะแนนทั้ง 2 ส่วนเพื่อให้สอดคล้องกับลักษณะของการทดสอบ แบบสอบและผู้สอบ ด้วยเหตุนี้การพิจารณาความก้าวหน้าดังกล่าวย่อมทำให้เห็นช่องว่างในเชิงทฤษฎี อันจะนำไปสู่การขยายแนวคิด/สมการของการปรับเทียบคะแนนด้วยโมเดล MIRT ไปยังบริบทต่าง ๆ นอกจากนี้ยังทำให้เห็นช่องว่างในการวิจัยกับสถานการณ์การทดสอบหรือตัวแปรที่ยังไม่มีการศึกษา ซึ่งผู้วิจัยสามารถประยุกต์ใช้วิธีการปรับเทียบด้วยโมเดล MIRT ไปยังบริบทการทดสอบที่แตกต่างกัน อันจะนำไปสู่ข้อสรุปที่สะท้อนคุณภาพของการปรับเทียบคะแนนด้วยโมเดล MIRT

มโนทัศน์ของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

การปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (MIRT Equating) เป็นกระบวนการแปลงคะแนนบนสเกลของแบบสอบฉบับหนึ่งไปยังสเกลของแบบสอบอีกฉบับ ซึ่งแบบสอบแต่ละฉบับที่มุ่งวัดหลายคุณลักษณะหรือหลายความสามารถ โดยการปรับเทียบคะแนนด้วยโมเดล MIRT แตกต่างจากการปรับเทียบคะแนนด้วยโมเดล UIRT ตรงที่ค่าความสามารถตามโมเดล IRT เป็นสเกลาร์ (θ) ส่วนค่าความสามารถด้วยโมเดล MIRT เป็นเวกเตอร์ที่ประกอบด้วยความสามารถในมิติต่าง ๆ $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots, \theta_m]$

กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Full MIRT Score Equating Procedure: FMIRT) ได้รับการพัฒนาเพื่อไปปรับเทียบคะแนนแบบสอบพหุมิติ ซึ่งประกอบด้วย 3 กระบวนการ ได้แก่ (1) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูปแบบ (MOSE) (2) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (AOSE) และ (3) กระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (ATSE)

กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (Full MIRT Observed Score Equating Procedure: MOSE)

กระบวนการ MOSE เป็นการปรับเทียบคะแนนที่สังเกตได้ระหว่างแบบสอบพหุมิติ โดยคำนวณการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขในแต่ละกลุ่มระดับความสามารถ (Conditional observed score distributions: $(f_r(x|\theta))$) โดยอาศัยสมการวนซ้ำ (Recursion formula) เนื่องจากค่าความสามารถของผู้สอบตามโมเดล MIRT มีลักษณะแบบเวกเตอร์ที่ประกอบด้วยความสามารถในมิติต่าง ๆ ($\theta = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m]$) แตกต่างจากการปรับเทียบคะแนนด้วยโมเดล UIRT ที่การแจกแจงคะแนนสังเกตได้จะพิจารณาในค่าความสามารถนั้น ๆ กระบวนการ MOSE ถูกประยุกต์ใช้กับข้อสอบที่มีการตรวจให้คะแนนสองค่า (Dichotomous items) ข้อสอบที่มีการตรวจให้คะแนนหลายค่า (Polytomous items) และแบบสอบรูปแบบผสม (Mixed-format test) ซึ่งมีขั้นตอนหลัก 3 ขั้นตอนเหมือนกัน คือ (1) คำนวณการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไข (2) คำนวณการแจกแจงคะแนนสังเกตได้รวม และ (3) ปรับเทียบคะแนนระหว่างแบบสอบด้วยวิธีอีควิเปอร์เซ็นไทล์ (Brossman & Lee, 2013; Peterson 2014) มีรายละเอียดดังนี้

1. คำนวณการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขในแต่ละกลุ่มของระดับความสามารถ $(f_r(x|\theta))$ เมื่อ θ คือ เวกเตอร์ของความสามารถ (Ability vector) $\theta = [\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m]$ และ r คือ ดัชนีข้อสอบ/ข้อสอบข้อที่ (Item index) การคำนวณการแจกแจงดังกล่าวสำหรับข้อสอบที่มีการตรวจให้คะแนนสองค่าใช้อัลกอริทึมของ Lord-Wingersky และข้อสอบที่มีการตรวจให้คะแนนหลายค่าใช้สูตรของ Hanson (1994) และ Thissen และคณะ (1995 อ้างอิงใน Peterson, 2014)

- คำนวณการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขสำหรับข้อสอบที่มีการตรวจให้คะแนนสองค่า

การแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไข คือ ความน่าจะเป็นของผู้สอบที่จะได้คะแนน x เมื่อทำข้อสอบถึงข้อที่ r บนเงื่อนไขของเวกเตอร์ความสามารถ (θ_j) ดังสมการในตาราง 1

ตาราง 1 สมการคำนวณการแจกแจงคะแนนสังเกตได้ สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า

สมการคำนวณการแจกแจงคะแนนสังเกตได้ สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า	
สมการการแจกแจง เมื่อทำข้อสอบข้อที่ 1 ($r=1$)	1. ความน่าจะเป็นที่จะได้ 0 คะแนนจากข้อสอบข้อ 1 คือ $f_1(x=0 \theta_j) = 1 - P_1 \quad (1)$ 2. ความน่าจะเป็นที่จะได้ 1 คะแนนจากข้อสอบข้อ 1 คือ $f_1(x=1 \theta_j) = P_1 \quad (2)$ เมื่อ P_1 คือ โอกาสที่จะตอบข้อสอบข้อที่ 1 ถูกต้อง $1 - P_1$ คือ โอกาสที่จะตอบข้อสอบข้อที่ 1 ผิด
สมการการแจกแจง เมื่อทำข้อสอบตั้งแต่	1. โอกาสที่จะได้ 0 คะแนน เมื่อทำข้อสอบถึงข้อที่ r คือ

ข้อที่ 2 เป็นต้นไป ($r > 1$)	$f_r(x \theta_j) = f_{r-1}(x \theta_j)(1 - P_r)$ (3)
เมื่อ x คือ คะแนน	2. โอกาสที่จะได้คะแนน $0 < x < r$ เมื่อทำข้อสอบถึงข้อที่ r คือ $f_r(x \theta_j) = f_{r-1}(x \theta_j)(1 - P_r) + f_{r-1}(x - 1 \theta_j)P_r$ (4)
	3. โอกาสที่จะได้คะแนน $x = r$ เมื่อทำข้อสอบถึงข้อที่ r คือ $f_r(x \theta_j) = f_{r-1}(x - 1 \theta_j)P_r$ (5)

- *คำนวณการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขสำหรับข้อสอบที่มีการตรวจให้คะแนนหลายค่า*

การแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไข คือ ความน่าจะเป็นของผู้สอบที่จะได้คะแนน x เมื่อทำข้อสอบถึงข้อที่ r บนเงื่อนไขของเวกเตอร์ความสามารถ (θ_j) ดังสมการในตาราง 2

ตาราง 2 สมการสำหรับการคำนวณการแจกแจงคะแนนสังเกตได้ สำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า

การคำนวณการแจกแจงคะแนนสังเกตได้ สำหรับข้อสอบที่ตรวจให้คะแนนหลายค่า	
สมการการแจกแจง เมื่อทำข้อสอบข้อที่ 1 ($r=1$)	$f_r(U_i = W_{rk} \theta_j) = P_{ik}(\theta_j)$ (6) เมื่อ W_{rk} คือ การให้คะแนนของลำดับขั้นที่ k สำหรับข้อสอบข้อที่ r
สมการการแจกแจง เมื่อทำข้อสอบตั้งแต่ ข้อที่ 2 เป็นต้นไป ($r > 1$) เมื่อ x คือ คะแนน	โอกาสที่จะได้คะแนน $0 < x < r$ เมื่อทำข้อสอบถึงข้อที่ r คือ $f_r(x \theta_j) = \sum_{k=1}^{K_j} f_{r-1}(x - W_{rk}) P_{ik}(\theta_j)$ เมื่อ $\min_r < x < \max_r$ (7) เมื่อ $\min_r$ คือ คะแนนต่ำสุดที่เป็นไปได้หลังจากเพิ่มของข้อสอบข้อที่ r $\max_r$ คือคะแนนสูงสุดที่เป็นไปได้หลังการเพิ่มของข้อสอบข้อที่ r ถ้า $x - W_{rk} < \min_r$ หรือ $x - W_{rk} > \max_r$ จะได้ $f_{r-1}(x - W_{rk}) = 0$

- *คำนวณการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขสำหรับแบบสอบรูปแบบผสม*

การแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขสำหรับแบบสอบรูปแบบผสม คือ ผลรวมของสมการวนซ้ำของข้อสอบที่ตรวจให้คะแนนหลายสองค่าและหลายค่าบนเวกเตอร์เงื่อนไขความสามารถ ดังตาราง 1 และ 2 ตามลำดับ (θ_j)

2. *คำนวณการแจกแจงคะแนนสังเกตได้รวม (Marginal observed score distributions: $f(x)$)* ด้วยการนำการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขคูณด้วยความหนาแน่นความสามารถ

แบบหลายตัวแปร (Multivariate ability density: $\psi(\theta)$) โดยการแจกแจงคะแนนสังเกตได้รวมสำหรับข้อสอบที่ตรวจให้คะแนนสองค่าและข้อสอบที่ตรวจให้คะแนนหลายค่าเหมือนกัน ดังสมการ (8)

$$f(x) = \int_{\theta_1} \int_{\theta_2} \cdots \int_{\theta_m} f(x|\theta) \psi(\theta) d(\theta) \quad \text{หรือ} \quad f(x) = \sum_{\theta_1} \sum_{\theta_2} \cdots \sum_{\theta_m} f(x|\theta) \psi(\theta) \quad (8)$$

ส่วนการแจกแจงดังกล่าวสำหรับแบบสอบรูปแบบผสมจะทำการแจกแจงปกติแบบ 3 ตัวแปร (trivariate normal distribution) ดังสมการ (9)

$$f(x) = \iiint_{-\infty}^{\infty} f(x|\theta_G, \theta_M, \theta_F) g(\theta_G, \theta_M, \theta_F) d\theta_G d\theta_M d\theta_F = \sum_{\theta_G} \sum_{\theta_M} \sum_{\theta_F} f(x|\theta_G, \theta_M, \theta_F) q(\theta_G, \theta_M, \theta_F) \quad (9)$$

เมื่อ $q(\theta_G, \theta_M, \theta_F)$ คือ ฟังก์ชันความหนาแน่นของการแจกแจงปกติของตัวแปร 3 ตัวสำหรับคุณลักษณะแฝงทั่วไป (θ_1) คุณลักษณะแฝงเฉพาะข้อสอบแบบหลายตัวเลือก (θ_M) และคุณลักษณะแฝงเฉพาะข้อสอบแบบสร้างคำตอบ (θ_F)

3. ปรับเทียบคะแนนแบบสอบทั้งสองฉบับด้วยวิธีอีควิเปอร์เซ็นต์ไทล์ (equipercentile method)

กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (Unidimensional Approximation to MIRT Equating Procedure)

กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติเป็นการปรับเทียบระหว่างแบบสอบพหุมิติ โดยประมาณค่าพารามิเตอร์และค่าความสามารถด้วยการประมาณแบบเอกมิติตามแนวคิดของ Zhang และ Stout (1999) ซึ่งจะได้ค่าพารามิเตอร์แบบเอกมิติ (Unidimensional item parameters) และค่าความสามารถแบบเอกมิติ (Unidimensional abilities) ซึ่งแตกต่างกับกระบวนการ MOSE ที่ประมาณค่าพารามิเตอร์และค่าความสามารถด้วยการประมาณแบบพหุมิติ โดยกระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ ประกอบด้วย กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (AOSE) และกระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (ATSE) ซึ่งในขณะนี้มีการนำเสนอสมการของกระบวนการทั้งสองแบบสำหรับข้อสอบที่มีการตรวจให้คะแนนแบบสองค่าเท่านั้น (Zhang, 2012; Brossman & Lee, 2013)

1. กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณเอกมิติ (Unidimensional Approximation to MIRT Observed Score Equating Procedure: AOSE) เป็นการปรับเทียบคะแนนสังเกตได้ระหว่างแบบสอบพหุมิติ ซึ่งขั้นตอนหลักของกระบวนการ AOSE หลังจากการประมาณค่าแบบเอกมิติ มี 3 ขั้นตอนเหมือนกับกระบวนการ MOSE ดังนี้

1.1 คำนวณการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไข (Conditional observed score distributions: $f(x|\theta_\alpha)$) ในแต่ละระดับคอมโพสิทความสามารถ (Composite ability level: θ_α) โดยใช้สูตรสมการวนซ้ำของ Lord-Wingersky

1.2 คำนวณการแจกแจงคะแนนสังเกตได้รวม (Marginal observed score distributions: $f(x)$) จากการนำการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขคูณกับการแจกแจงความสามารถแบบเอกมิติที่ประมาณได้ ดังสมการ 10; $f(x) = \sum_{\theta_\alpha} f(x|\theta_\alpha) \psi(\theta_{\alpha i})$ หรือ $\int_{\theta_1} f(x|\theta_\alpha)(\theta_{\alpha i})d\theta_\alpha$ (10)

1.3 ปรับเทียบคะแนนแบบสอบทั้งสองฉบับด้วยวิธีอีควิเปอร์เซ็นไทล์

2. กระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (Unidimensional Approximation to MIRT True Score Equating Procedure: ATSE) เป็นการปรับเทียบคะแนนจริงระหว่างแบบสอบพหุมิติ โดยหลังจากการประมาณค่าแบบเอกมิติแล้วจะใช้กระบวนการ ATSE มีขั้นตอนดังนี้

2.1 ระบุคะแนนจริงคอมโพสิทของแบบสอบฉบับที่ถูกปรับเทียบ ($T_{A\alpha}$) ที่สอดคล้องกับคะแนนจริงคอมโพสิทของแบบสอบฐาน

2.2 ระบุระดับความสามารถคอมโพสิทแบบเอกมิติ (Unidimensional composite ability level: θ_α) ที่สอดคล้องกับคะแนนจริงของแบบสอบฉบับที่ถูกปรับเทียบด้วยวิธีการวนซ้ำ ที่ทำให้ฟังก์ชันคอมโพสิทมีความแตกต่างกันน้อยที่สุดดังสมการ 11;

$$func(\theta_{\alpha i}) = \tau_{\alpha A} - \sum_{j:B} p_{ij}(\theta_{\alpha i}; a_{\alpha j}, b_{\alpha j}, c_j) \quad (11)$$

2.3 ระบุคะแนนจริงคอมโพสิทของแบบสอบฐาน ($\tau_{\alpha B}(\theta_{\alpha i})$) ที่สอดคล้องกับคะแนนจริงคอมโพสิทของแบบสอบที่ถูกปรับเทียบ $\tau_{\alpha E}(\theta_{\alpha i})$ โดยใช้กรอบแนวคิดของโมเดล UIRT ดังสมการ 12

$$\tau_{\alpha B}(\theta_{\alpha i}) = \sum_{j:B} p_{ij}(\theta_{\alpha i}; a_{\alpha j}, b_{\alpha j}, c_j) \quad (12)$$

ความก้าวหน้าในเชิงทฤษฎีของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

จากแนวคิดของกระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ (Full MIRT Observed Score Equating Procedure: FMIRT) ซึ่งประกอบด้วย กระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติแบบเต็มรูป (MOSE) กระบวนการปรับเทียบกระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (AOSE) และกระบวนการปรับเทียบกระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (ATSE) สะท้อนให้เห็นว่า นักวิชาการ/นักวิจัยพยายามขยายองค์ความรู้ของการปรับเทียบคะแนนให้เหมาะสมกับลักษณะของแบบสอบที่มุ่งวัดหลายคุณลักษณะโดยอาศัยองค์ความรู้ทางสถิติที่มีอยู่ อย่างเช่น ในกระบวนการ

MOSE ที่ประยุกต์ใช้สมการวนซ้ำ (Recursion formula) ที่นักวิจัยก่อนหน้านี้ได้ระบุไว้ เพื่อคำนวณการแจกแจงคะแนนสังเกตได้อย่างมีเงื่อนไขสำหรับข้อสอบที่มีการตรวจให้คะแนนสองค่าและข้อสอบที่ตรวจให้คะแนนหลายค่า อย่างไรก็ตามในกระบวนการปรับเทียบคะแนนด้วยโมเดล MIRT แบบเต็มรูปยังไม่มี การนำเสนอสมการสำหรับการปรับเทียบคะแนนจริง และเมื่อพิจารณากระบวนการปรับเทียบคะแนนแบบสอบพหุมิติที่อาศัยการประมาณค่าแบบเอกมิติ พบว่า มีการขยายแนวคิดทั้งการปรับเทียบคะแนนสังเกตได้ซึ่งก็คือ กระบวนการ AOSE และการปรับเทียบคะแนนจริง ซึ่งก็คือ กระบวนการ ATSE ไปยังข้อสอบที่มีการตรวจให้คะแนนสองค่าและหลายค่า อย่างไรก็ตามยังไม่มี การนำเสนอสมการของกระบวนการ AOSE และกระบวนการ ATSE สำหรับแบบสอบรูปแบบผสม จากช่องว่างในเชิงทฤษฎีดังกล่าวสะท้อนให้เห็นว่า นักวิชาการยังสามารถขยายทฤษฎี/องค์ความรู้ของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติไปยังบริบทดังกล่าวได้

ความก้าวหน้าทางการวิจัยของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

งานวิจัยเกี่ยวกับการปรับเทียบคะแนนแบบสอบพหุมิติส่วนใหญ่เป็นการปรับเทียบประสิทธิภาพระหว่างกระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติแบบเต็มรูป (FMIRT) ทั้ง 3 แบบกับวิธีการอื่น ๆ ไม่ว่าจะเป็นการเปรียบเทียบประสิทธิภาพระหว่างกระบวนการ FMIRT, กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ และวิธีอีคิวเปอร์เซนไทล์ ซึ่งมีทั้งการเปรียบเทียบประสิทธิภาพสำหรับข้อสอบที่มีการตรวจให้คะแนนสองค่าภายใต้รูปแบบผู้สอบกลุ่มเท่าเทียมกัน (Brossman, 2010), การเปรียบเทียบประสิทธิภาพสำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม (Lee & Brossman, 2012) และการเปรียบเทียบประสิทธิภาพสำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม (Lee, 2013) นอกจากนี้ยังมีการเปรียบเทียบประสิทธิภาพระหว่างกระบวนการ FMIRT, กระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ, วิธีการปรับเทียบคะแนนแบบสององค์ประกอบ (Bi-factor equating) และวิธีอีคิวเปอร์เซนไทล์ สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม (Peterson, 2014) ซึ่งงานวิจัยเหล่านี้พบว่า กระบวนการ FMIRT มีประสิทธิภาพในการปรับเทียบคะแนนมากกว่าการปรับเทียบด้วยโมเดล IRT และเมื่อเปรียบเทียบประสิทธิภาพระหว่างกระบวนการ FMIRT ทั้ง 3 แบบ พบว่า กระบวนการ MOSE มีประสิทธิภาพมากที่สุด สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่มสุ่ม และกระบวนการ AOSE มีประสิทธิภาพมากที่สุด สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบ NEAT

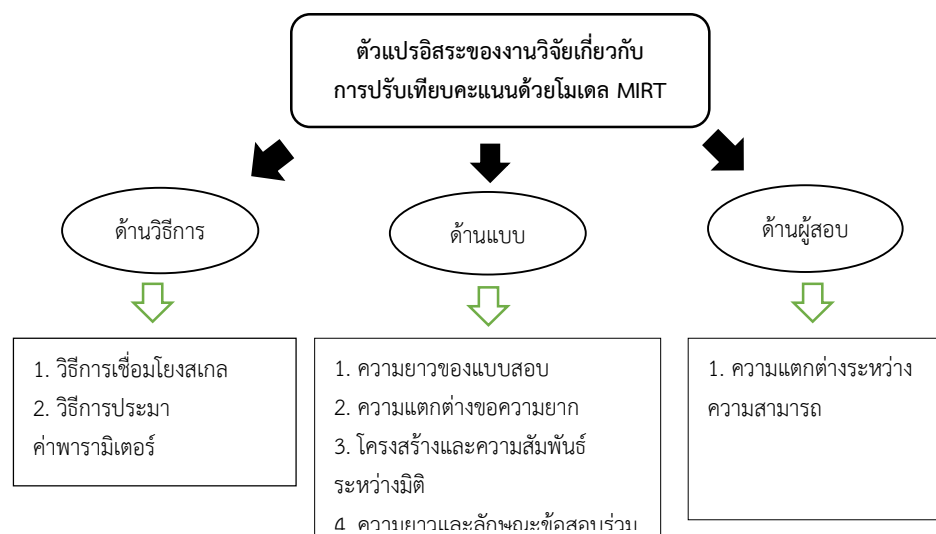
งานวิจัยบางส่วนเปรียบเทียบประสิทธิภาพระหว่างกระบวนการ FMIRT 3 แบบ คือ กระบวนการ MOSE กระบวนการ AOSE และกระบวนการ ATSE เมื่อมีวิธีการเชื่อมโยงสเกลด้วยโมเดล MIRT แตกต่างกัน สำหรับข้อสอบที่ตรวจให้คะแนนสองค่า ภายใต้รูปแบบผู้สอบกลุ่มไม่เท่าเทียมกันโดยใช้ข้อสอบร่วม (NEAT) ซึ่งพบว่ากระบวนการ ATSE มีประสิทธิภาพมากที่สุด (Zhang, 2012) และงานวิจัยบางส่วนศึกษาเฉพาะประสิทธิภาพกระบวนการ MOSE โดยเปรียบเทียบประสิทธิภาพของกระบวนการ MOSE เมื่อมีการประมาณค่าพารามิเตอร์ต่างกัน 2 แบบ คือ กระบวนการ MOSE ที่มีการประมาณค่าพารามิเตอร์พร้อมกัน (Concurrent Calibrated MOSE: CMOSE) และกระบวนการ

MOSE ที่มีการประมาณค่าพารามิเตอร์แยกกัน โดยเชื่อมโยงสเกลด้วยวิธี TCF (Separated Calibrated MOSE with Test characteristic Function Procedure: SMOSE) สำหรับแบบสอบรูปแบบผสม ภายใต้รูปแบบ NEAT ซึ่งพบว่า กระบวนการ MOSE ที่มีการประมาณค่าพารามิเตอร์พร้อมกันมีประสิทธิภาพมากกว่ากระบวนการ MOSE ที่มีการประมาณค่าพารามิเตอร์แยกกัน โดยเชื่อมโยงสเกลด้วยวิธี TCF (พนิดา พานิชวัฒน์, 2560)

งานวิจัยที่กล่าวมาข้างต้นมีวัตถุประสงค์หลักอยู่ที่การเปรียบเทียบประสิทธิภาพระหว่างวิธีการปรับเทียบคะแนน อีกทั้งยังมีการศึกษาตัวแปรอิสระอื่น ๆ ร่วมด้วย ไม่ว่าจะเป็น (1) ด้านวิธีการ ได้แก่ วิธีการเชื่อมโยงสเกล และวิธีการประมาณค่าพารามิเตอร์ (2) ด้านแบบสอบ ได้แก่ ความยาวของแบบสอบ โครงสร้างและความสัมพันธ์ระหว่างมิติความสามารถ ความแตกต่างของความยากระหว่างแบบสอบ และความยาวและลักษณะของข้อสอบรวม และ (3) ด้านผู้สอบ คือ ความแตกต่างระหว่างความสามารถของผู้สอบ (Lee & Brossman, 2012; Lee, 2013; พนิดา พานิชวัฒน์, 2560) โดยผู้เขียนได้สรุปตัวแปรอิสระที่มีการศึกษาในการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติไว้ดังภาพ 1

นอกจากนี้เมื่อพิจารณาบริบทของงานวิจัยเกี่ยวกับการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ พบว่า กลุ่มตัวอย่างมีขนาดใหญ่ตั้งแต่กลุ่มละ 2,000 คนขึ้นไป และมีการปรับเทียบคะแนนทั้งข้อมูลจริงและข้อมูลจำลอง ในกรณีของข้อมูลจริงจำเป็นต้องใช้ข้อมูลจากการทดสอบขนาดใหญ่ ซึ่งเป็นข้อจำกัดในการประยุกต์ใช้องค์ความรู้ดังกล่าวกับการทดสอบจริงที่มีขนาดเล็ก

ดังนั้นหากนักวิจัยที่ต้องการประยุกต์ใช้การปรับเทียบคะแนนแบบสอบพหุมิตีกับการทดสอบจริงที่มีกลุ่มตัวอย่างขนาดเล็ก อาจใช้การจำลองข้อมูลเป็นกรณีศึกษานำร่องเพื่อหาขนาดกลุ่มตัวอย่างขนาดเล็กที่เหมาะสมสำหรับการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ หลังจากนั้นจึงประยุกต์ใช้ขนาดกลุ่มตัวอย่างดังกล่าวกับกระบวนการปรับเทียบเทียบด้วยโมเดล MIRT ในสถานการณ์การทดสอบจริง



ภาพ 1 สรุปตัวแปรอิสระในงานวิจัยเกี่ยวกับการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ

บทสรุป

การปรับเทียบคะแนนเป็นองค์ความรู้ทางการวัดและประเมินผลที่ช่วยให้ระบบการทดสอบมีประสิทธิภาพและมีมาตรฐานยิ่งขึ้น โดยการปรับเทียบคะแนนทำให้คะแนนระหว่างแบบสอบสามารถเปรียบเทียบหรือแลกเปลี่ยนกันได้อย่างถูกต้อง และกระบวนการดังกล่าวยังช่วยให้การพัฒนาค้างข้อสอบมีคุณภาพมากยิ่งขึ้น เพราะสามารถแปลงค่าพารามิเตอร์ของข้อสอบที่จะเพิ่มเข้าคลังให้อยู่ในสเกลเดียวกับข้อสอบข้ออื่น ๆ ในคลังข้อสอบ จากความสำคัญของการปรับเทียบคะแนนดังกล่าวทำให้นักวิชาการและนักวิจัยพยายามอย่างยิ่งที่จะพัฒนาระบบการดังกล่าวให้สอดคล้องกับลักษณะของแบบสอบและผู้สอบในบริบทต่าง ๆ ซึ่งหนึ่งในความพยายามนั้นก็คือ การพัฒนาระบบการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติแบบเต็มรูป (FMIRT) ทั้ง 3 กระบวนการ ได้แก่ (1) กระบวนการปรับเทียบคะแนนที่สังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป (MOSE) และ (2) กระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (AOSE) และ (3) กระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติโดยใช้การประมาณค่าแบบเอกมิติ (ATSE) และงานวิจัยที่ผ่านมาสะท้อนให้เห็นว่ากระบวนการ FMIRT มีประสิทธิภาพในการปรับเทียบคะแนนแบบสอบพหุมิติมากกว่ากระบวนการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบแบบเอกมิติ (UIRT equating procedure) อย่างไรก็ตามยังมีช่องว่างขององค์ความรู้เชิงทฤษฎีสำหรับกระบวนการปรับเทียบคะแนนจริงตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเต็มรูป รวมถึงช่องว่างในการขยายองค์ความรู้ของการปรับเทียบคะแนนดังกล่าวไปยังสถานการณ์ทดสอบจริงที่มีกลุ่มตัวอย่างขนาดเล็ก

บรรณานุกรม / เอกสารอ้างอิง

- พนิดา พานิชวัฒน์. (2560). *การเปรียบเทียบประสิทธิภาพของกระบวนการปรับเทียบคะแนนสังเกตได้ตามทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ สำหรับแบบสอบรูปแบบผสม*. (ครุศาสตรมหาบัณฑิต), จุฬาลงกรณ์มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่*. พิมพ์ครั้งที่ 2. กรุงเทพมหานคร: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- Brossman, B. G. (2010). *Observed score and true score equating procedures for multidimensional item response theory*. (Doctor of Philosophy), University of Iowa. Retrieved from <http://ir.uiowa.edu/etd/469>
- Brossman, B. G., & Lee, W.-C. (2013). Observed score and true score equating procedures for multidimensional item response theory. *Applied Measurement in Education*, 27(3), 159-172. doi: 10.1080/08957347.2014.905785

- Holland, P. W. (2007). A framework and history for score linking. In N. J. Dorans, M. Pommerich & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 5-30). New York: Springer Science+Business Media, LLC.
- Kolen, M. J., & Brennan, R. L. (2004). *Test Equating, scaling, and linking: methods and practices*. In S. E. Fienberg & W. J. v. d. Linden (Series Eds.), S. E. Fienberg & W. J. v. d.
- Lee, E. (2013). *Equating multidimensional tests under a random groups design: a comparison of various equating procedures*. (Doctor of Philosophy), University of Iowa. Retrieved from <http://ir.uiowa.edu/etd/5011>
- Lee, W.-C., & Brossman, B. G. (2012). Observed score equating for mixed-format tests using a simple-structure multidimensional IRT framework. *Center for Advanced Studies in Measurement and Assessment (CASMA)*, 2(2.2), 115-142.
- Peterson, J. L. (2014). *Multidimensional item response theory observed score equating methods for mixed-format tests*. (Doctor of Philosophy), The University of Iowa.
- Zhang, J., & Stout, W. (1999). Conditional covariance structure of generalized compensatory multidimensional items. *Psychometrika*, 64(2), 129-152.
- Zhang, O. (2012). *Observed score and true score equating for multidimensional item response theory under nonequivalent group anchor test design*. (The Degree of Doctor of Philosophy), University of Florida.