

การวิเคราะห์พหุมิติ (Multidimensional Analysis)

ดร.ชัยวิชิต เขียรชนะ*

คำถามในเรื่องของมิติโครงสร้างลักษณะทางพฤติกรรมศาสตร์ยังเป็นที่สนใจของผู้ศึกษาค้นคว้าเป็นอย่างมาก เนื่องจากความสนใจที่ต้องการทราบว่าโครงสร้างของคุณลักษณะที่ทำการวัดอาจจะมีลักษณะของความเป็นพหุมิติ (multidimensionality) นั่นคือ อาจจะมีลักษณะมิติที่มากกว่าหนึ่งมิติ แนวคิดพื้นฐานการวิเคราะห์พหุมิติที่จะได้นำเสนอต่อไปนี้จะมียุทธศาสตร์ในการทำความเข้าใจโครงสร้างคุณลักษณะที่วัดว่า มีความเหมาะสมในการวัดหรือมีความเหมาะสมในการแปลความหมาย ลักษณะพหุมิติหรือลักษณะเอกมิติ? บทความนี้จะนำเสนอสาระสำคัญเกี่ยวกับการวิเคราะห์ พหุมิติ โดยแบ่งออกเป็น 5 ตอน คือ แนวคิดพื้นฐานของลักษณะเอกมิติและพหุมิติ การตรวจสอบลักษณะความเป็นเอกมิติและพหุมิติเบื้องต้น ไม่เดลการตอบสนองข้อสอบแบบพหุมิติ (MRCMLM) การประยุกต์ใช้ และบทส่งท้าย

แนวคิดพื้นฐานของลักษณะเอกมิติและพหุมิติ

ส่วนนี้จะขอนำเสนอ 2 ประเด็น คือ 1) แนวคิดพื้นฐานของลักษณะเอกมิติและลักษณะพหุมิติ และ 2) การแปลความหมายตามลักษณะเอกมิติและลักษณะพหุมิติ มีรายละเอียดดังนี้

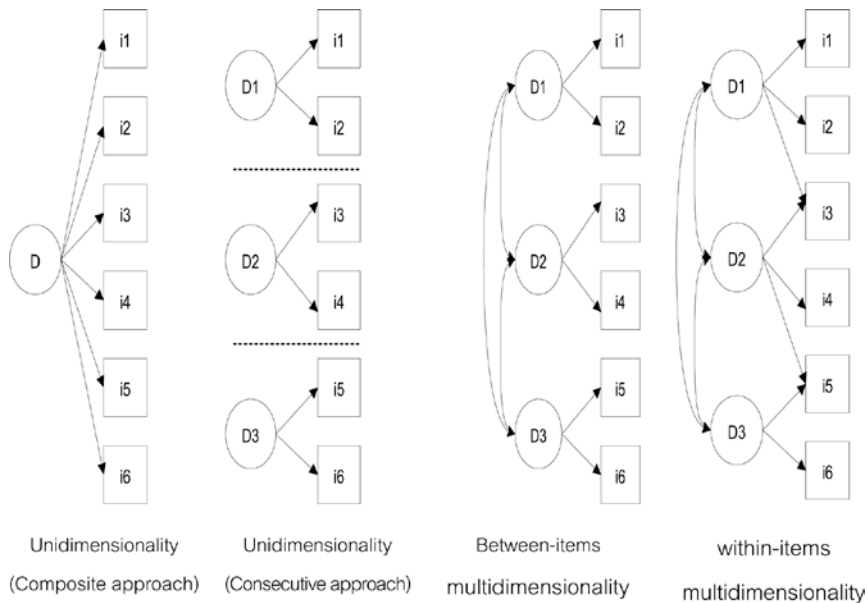
แนวคิดพื้นฐานของลักษณะเอกมิติ โดยลักษณะไม่เดลพื้นฐานของความเป็นเอกมิติมีข้อตกลงเบื้องต้นที่ว่า คุณลักษณะแฝงที่จะทำการวัดไม่มีความสัมพันธ์กัน ซึ่งมีอยู่ 2 ลักษณะคือ 1) ความเป็นเอกมิติรวม (Composite approach) มีลักษณะของข้อคำถามวัดคุณลักษณะแฝงเดียวและมีคุณลักษณะแฝงเดียว และ 2) ความเป็นเอกมิติแยกตามมิติ (Consecutive approach) มีลักษณะของ

ข้อคำถามวัดคุณลักษณะแฝงเดียว แต่มีหลายคุณลักษณะแฝงทั้งนี้แต่ละคุณลักษณะแฝงไม่มีความสัมพันธ์กัน

ส่วนลักษณะความเป็นพหุมิติมีข้อตกลงเบื้องต้นที่ว่า คุณลักษณะที่จะทำการวัดมีความสัมพันธ์กัน ซึ่งมีอยู่ 2 ลักษณะคือ 1) ความเป็นพหุมิตระหว่างข้อคำถาม (Between-items multidimensionality) มีลักษณะของข้อคำถามวัดคุณลักษณะแฝงเดียว มีหลายคุณลักษณะแฝงทั้งนี้แต่ละคุณลักษณะแฝงมีความสัมพันธ์กัน และ 2) ความเป็นพหุมิติภายในข้อคำถาม (within-items multidimensionality) มีลักษณะของข้อคำถามวัดหลายคุณลักษณะแฝง มีหลายคุณลักษณะแฝง ทั้งนี้แต่ละคุณลักษณะแฝงมีความสัมพันธ์กัน ซึ่งแนวคิดพื้นฐานของลักษณะเอกมิติและลักษณะพหุมิติ แสดงดังภาพที่ 1

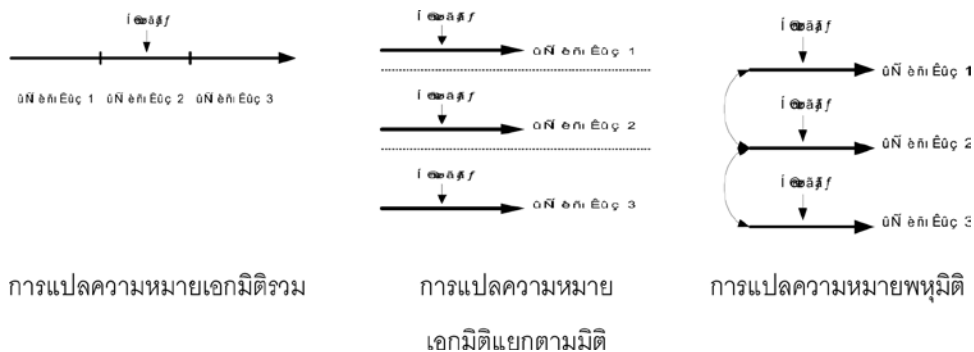
สำหรับแนวคิดการแปลความหมาย สามารถอธิบายตามความเหมาะสมคุณลักษณะที่มุ่งวัด เช่น คุณลักษณะที่มุ่งวัดคือ กลยุทธ์การเรียนรู้ ประกอบด้วย 3 องค์ประกอบคือ กลยุทธ์การรู้คิด กลยุทธ์จิตพิสัย และกลยุทธ์ทักษะการเรียนรู้ หากมีการอธิบายตามความเหมาะสมแบบเอกมิติรวมอธิบายโดยรวมคุณลักษณะกลยุทธ์ทั้ง 3 องค์ประกอบเข้าด้วยกัน หากอธิบายตามความเหมาะสมแบบเอกมิติแยกตามมิติ ต้องแยกอธิบายแต่ละคุณลักษณะ โดยมีความเชื่อพื้นฐานที่ว่า แต่ละองค์ประกอบไม่มีความสัมพันธ์กัน และหากอธิบายตามความเหมาะสมแบบพหุมิติ แยกอธิบายแต่ละคุณลักษณะโดยมีความเชื่อพื้นฐานที่ว่า แต่ละองค์ประกอบมีความสัมพันธ์กัน ซึ่งแนวคิดการแปลความหมาย แสดงดังภาพที่ 2

*อาจารย์ประจำภาควิชาบริหารเทคนิคการศึกษา คณะครุศาสตร์อุตสาหกรรม มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ
Email: chianchanachch@hotmail.com, chawichitc@kmutnb.ac.th



หมายเหตุ: D หมายถึง มิติ หรือ องค์ประกอบ, i หมายถึง ข้อสอบ หรือ ข้อคำถาม

ภาพที่ 1 แนวคิดเอกมิติและพหุมิติ



ภาพที่ 2 การแปลความหมายเอกมิติและพหุมิติ

การตรวจสอบลักษณะความเป็นเอกมิติและพหุมิติเบื้องต้น

การตรวจสอบลักษณะความเป็นเอกมิติและพหุมิติเบื้องต้น (เชิงสำรวจ) มีวิธีการดังนี้ 1) การตรวจสอบแบบแนวนอน ตามที่ Reckase (1979) เสนอให้พิจารณาจากค่าความแปรปรวนที่ตัวประกอบแรกสามารถอธิบายความแปรปรวนทั้งหมดได้อย่างน้อย 20% จะบ่งบอกถึงความเป็นเอกมิติที่มีลักษณะองค์ประกอบ

หลักตัวแรกเด่นกว่าองค์ประกอบหลักตัวอื่นๆ (essential unidimensionality) และ 2) การตรวจสอบแบบแนวตั้ง Morizot, Ainsworth และ Reise (2007) ได้เสนอให้พิจารณาความเป็นเอกมิติ (unidimensionality) จากค่าอัตราส่วนระหว่างค่า Eigen องค์ประกอบแรกต่อค่า Eigen องค์ประกอบที่สอง หากมีค่ามากกว่าหรือเท่ากับ 3.00 จะบ่งบอกถึงความเป็นเอกมิติ

นอกจากนี้การตรวจสอบโครงสร้างลักษณะเชิง
ยืนยัน เป็นลักษณะของการตรวจสอบความตรงของ
โมเดลว่ามีความสอดคล้องกับข้อมูลลักษณะเอกมิติ
หรือพหุมิติ ซึ่งจะได้นำเสนอในประเด็นการตรวจสอบ
ความตรง

โมเดลการตอบสนองข้อสอบแบบพหุมิติ (MRCMLM)

โมเดลการตอบสนองข้อสอบแบบพหุมิตินั้นได้
พัฒนาโดยสามารถตรวจสอบผลจากลักษณะข้อมูลแบบ
พหุมิติเหมือนกับว่าข้อมูลนั้นเป็นแบบมิติเดียวและทฤษฎี
การตอบสนองข้อสอบแบบพหุมิติ (MIRT) เป็นโมเดลที่
รวมจุดเด่นของเทคนิควิธีวิเคราะห์ตัวประกอบและทฤษฎี
การตอบสนองข้อสอบ เข้าไว้ด้วยกัน ซึ่งวิธีวิเคราะห์
ตัวประกอบเป็นเทคนิคที่ใช้ในการสำรวจและยืนยัน
ตัวประกอบ และทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎี
ที่มีหลักการสำคัญ 2 ประการ คือ ประการแรก ผลการตอบ
ข้อคำถามของผู้ตอบสามารถอธิบายได้ด้วยความสามารถ
ที่มีอยู่ภายในของผู้ตอบ และประการที่สอง ความสัมพันธ์
ระหว่างผลการตอบข้อคำถามกับความสามารถที่มีอยู่
ภายในสามารถอธิบายได้ด้วยฟังก์ชันลักษณะข้อคำถาม
แม้ว่าการพัฒนารรณกรรมทางวิธีวิทยาการประยุกต์ใช้
โมเดลการตอบสนองข้อสอบแบบพหุมิติในทางพฤติกรรม
ศาสตร์ยังมีข้อจำกัด อันมีสาเหตุเนื่องมาจากปัญหาทาง
สถิติที่เกี่ยวกับความเหมาะสมของโมเดลและความยาก
ในการตีความหมายโมเดลการตอบสนองข้อสอบแบบ
พหุมิติ แต่ทั้งนี้ได้มีการแก้ปัญหาทางสถิติของโมเดล
การตอบสนองข้อสอบแบบพหุมิติโดยได้พัฒนา
multidimensional random coefficients multinomial
logit model (MRCMLM) ขึ้นมาเพื่อแก้ปัญหาทางสถิติ
ซึ่งโมเดลนี้อยู่ในตระกูลของราสซ์โมเดล โดยโมเดลนี้
มีการพัฒนาเพื่อให้เกิดความยืดหยุ่นในการตอบสนอง
ข้อสอบให้มีลักษณะที่แผ่ขยายออกไปสามารถประยุกต์
ใช้ในมาตรวัดทั้งการตอบที่เป็นลักษณะใช่/ไม่ใช่ (yes/no)
หรือการตอบในมาตราประมาณค่าแบบลิเคิร์ท นอกจากนี้
ยังสามารถประยุกต์ใช้ในมาตรวัดที่มีข้อสอบเกิดจาก
สถานการณ์ที่ซับซ้อน (complex situations) ใช้ในการ
ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (differential

item functioning) และยังประยุกต์ใช้กับการให้คะแนน
โดยผู้ตรวจคะแนน (raters) ได้อีกด้วย (Allen และ
Wilson, 2006)

สำหรับ multidimensional random
coefficients multinomial logit model (MRCMLM)
เป็นโมเดลของ Adams, Wilson และ Wang พัฒนา
ขึ้นมาในปี ค.ศ. 1997 ซึ่งประโยชน์ของการใช้โมเดล
MRCMLM คือ 1) เป็นโมเดลที่ได้รับการยอมรับจาก
นักพัฒนาแบบวัดในเรื่องการตัดสินความเหมาะสมของ
โครงสร้างมิติที่ทำการวัด 2) เป็นโมเดลที่ประมาณค่า
โดยตรง (direct estimates) ของความสัมพันธ์ระหว่าง
มิติแฝง และ 3) เป็นโมเดลที่มีความแกร่งของความสัมพันธ์
ระหว่างมิติแฝงกับการประมาณค่าพารามิเตอร์ที่จะให้ผล
อย่างถูกต้อง (Wilson และ Hoskens, 2005) นอกจากนี้
โมเดล MRCMLM ยังสามารถนำมาใช้ประโยชน์ในสาระ
อื่นๆ ได้อีก ดังที่ Wolfe และคนอื่นๆ (2007) ได้กล่าว
ถึงประโยชน์ในการนำโมเดล MRCMLM มาใช้ มีดังนี้
1) หากจำนวนคุณลักษณะแฝงยังขาดความชัดเจน
สามารถเปรียบเทียบความเหมาะสมของโมเดล การวัด
เพื่อช่วยในการตัดสินความเหมาะสมของโครงสร้างมิติ
ที่ทำการวัด 2) ใช้ในการวินิจฉัยจำแนกคุณลักษณะ ตัวแปร
แฝง (latent trait) ที่ได้จากการวัดที่หลากหลายซึ่งยัง
ขาดความสอดคล้องทั้งในเชิงสถิติและเชิงแนวคิด และ
3) โมเดลการวัดแบบพหุมิติจะมีความเที่ยง (reliability)
สูง เนื่องจากโมเดลนี้มีความสัมพันธ์ระหว่างคุณลักษณะ
แฝง สำหรับโมเดล MRCMLM สามารถทำการวิเคราะห์
ด้วยโปรแกรม ConQuest ซึ่งโมเดล MRCMLM เป็น
โมเดลที่สร้างบนพื้นฐานของบล็อก (block) หากข้อสอบ
มีเป้าหมายในการวัดคุณลักษณะแฝงเด่นเพียงมิติเดียว
จะเรียกโมเดลนี้ว่า โมเดลการวัดความสามารถหลายมิติ
ระหว่างข้อสอบ (multidimensional between items) และ
ถ้าหากข้อสอบมีเป้าหมายในการวัดคุณลักษณะแฝงสอง
มิติหรือมากกว่าจะเรียกโมเดลนี้ว่าโมเดลการวัดความ
สามารถหลายมิติภายในข้อสอบ (multidimensional
within items) (Allen และ Wilson, 2006) การวิเคราะห์
โมเดลนี้มีความเหมาะสมในการวิเคราะห์เชิงยืนยัน
มากกว่าเชิงสำรวจ ซึ่งโมเดลนี้จะอธิบายความสัมพันธ์
ระหว่างข้อสอบและการตอบของผู้ตอบที่มีต่อมิติ (d)



อธิบายดังสมการที่ (1) จะเห็นได้ว่าโอกาสในการตอบรายการคำตอบ K ของข้อที่ i ของผู้ตอบ (P_{ik}) ค่อนข้างจะมากกว่ารายการคำตอบ k-1 (P_{ik-1}) ซึ่งสัมพันธ์กับระดับเจตคติที่มีต่อมิติ (θ_d) และความยากสัมพัทธ์ (relative difficulty) ของรายการคำตอบ k (δ_{ik})

$$\text{Log}(P_{ik} / P_{ik-1}) = \theta_d - \delta_{ik} \quad \text{..... (1)}$$

โมเดลมีสมมุติฐานในเซตของคุณลักษณะแฝง (D) ที่ถูกกำหนดให้ทำแบบสอบระดับบุคคลในมิติคุณลักษณะแฝง D จาก $\theta'_n = (\theta_{n-1}, \dots, \theta_{n-D})$ ในอันที่จะมีการพิจารณาความเป็นตัวแทนของการสุ่มตัวอย่างจากประชากรด้วย multivariate density function ($g(\theta'_n; \alpha)$) โดยที่ α เป็นเวกเตอร์พารามิเตอร์ลักษณะการกระจาย ถ้า g ถูกกำหนดให้เป็นปกติ ดังนั้น $\alpha = (\mu, \Sigma)$ ภายใต้โมเดล MRCMLM โอกาสที่จะตอบรายการคำตอบ k ของข้อสอบข้อที่ i สำหรับคนที่ n แสดงได้ดังโมเดล (2)

$$f(X_{nik} = 1; \xi | \theta_n) = \frac{\exp(b'_{ik} \theta_n + a'_{ik} \xi)}{\sum_{u=1}^{K_i} \exp(b'_{iu} \theta_n + a'_{iu} \xi)} \quad \text{.....(2)}$$

เมื่อ

$X_{nik} = 1$ ถ้าการตอบสนองของข้อสอบข้อที่ i ของบุคคล n เป็นรายการคำตอบ k และ 0 เป็นอื่นๆ
 K_i = จำนวนรายการคำตอบในข้อสอบข้อที่ i
 ξ = เวกเตอร์พารามิเตอร์ความยากของข้อสอบ

b_{ik} = เวกเตอร์คะแนนที่เป็นรายการคำตอบ k ของข้อสอบข้อที่ i ตามคุณลักษณะแฝง D รวบรวมจากเมตริกซ์คะแนน B

a_{ik} = เวกเตอร์แบบแผนของรายการคำตอบ k ของข้อสอบข้อที่ i ที่อธิบายความสัมพันธ์เชิงเส้นตรงระหว่าง element ของ ξ โดยรวบรวมจากข้อสอบในเมตริกซ์แบบแผน A

สมการที่ (2) สามารถเขียนได้อีกรูปแบบ คือ

$$\ln \left[\frac{f(X_{nik} = 1; \xi | \theta_n)}{f(X_{ni,k-1} = 1; \xi | \theta_n)} \right] = (b'_{ik} - b'_{i,k-1}) \theta_n + (a'_{ik} - a'_{i,k-1}) \xi \quad \text{.....(3)}$$

จากสมการที่ (3) สามารถเขียนได้อีกรูปแบบคือ (wang, 2004)

$$\log \eta_{ik} = \ln \left(\frac{P_{ik}}{P_{i(k-1)}} \right) = \ln \left(\frac{P_i(\theta_n)}{P_{i(k-1)}(\theta_n)} \right) = (b'_{ik} - b'_{i(k-1)}) \theta_n + (a'_{ik} - a'_{i(k-1)}) \xi \\ \equiv b'_{ik} \theta_n + a'_{ik} \xi, \quad \text{.....(4)}$$

จากสมการที่ (4) สามารถขยายไปสู่โมเดลที่จะนำเสนอคือ Simple logistic model, Rating scale model, Partial credit model มีรายละเอียดแต่ละโมเดล ดังนี้

1. Simple logistic model

1.1 โมเดลการตอบสนองข้อสอบแบบเอกมิติ Simple logistic model

$$\ln \left(\frac{P_{i1}(\theta_n)}{P_{i0}(\theta_n)} \right) = \theta_n - \delta_i, \quad \text{.....(5)}$$

$P_{i1}(\theta_n)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง(θ) คนที่ n จะตอบข้อคำถามข้อที่ i ได้ถูกต้อง (1 คะแนน),

$P_{i0}(\theta_n)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง(θ) คนที่ n จะตอบข้อคำถามข้อที่ i ผิด (0 คะแนน),

θ_n คือ คุณลักษณะแฝงคนที่ n,

δ_i คือ ค่าความยากทั้งหมด (overall difficulty) ของข้อคำถามข้อที่ i,

1.2 โมเดลการตอบสนองข้อสอบแบบพหุมิติ Simple logistic model

$$\ln \left(\frac{P_{i1d}(\theta_{nd})}{P_{i0d}(\theta_{nd})} \right) = \theta_{nd} - \delta_{id}, \quad \text{.....(6)}$$

$P_{i1d}(\theta_{nd})$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง(θ) คนที่ n จะตอบข้อคำถามข้อที่ i ได้ถูกต้อง (1 คะแนน) ในมาตรวัดย่อยหรือมิติ d,

$P_{i(k-1)}(\theta_n)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง(θ) คนที่ n จะตอบข้อคำถามข้อที่ i ผิด (0 คะแนน) ในมาตรวัดย่อยหรือมิติ d,

θ_{nd} คือ คุณลักษณะแฝงคนที่ n ในมาตรวัดย่อยหรือมิติ d,

δ_{id} คือ ค่าความยากทั้งหมด (overall difficulty) ของข้อคำถามข้อที่ i ในมาตรวัดย่อยหรือมิติ d,

2. Rating scale model

2.1 โมเดลการตอบสนองข้อสอบแบบ
เอกมิติ Rating scale model

$$\ln\left(\frac{P_{ik}(\theta_n)}{P_{i(k-1)}(\theta_n)}\right) = \theta_n - \delta_i - \tau_k, \quad \dots(7)$$

$P_{ik}(\theta_n)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง (θ) คนที่ n จะตอบข้อคำถามข้อที่ i ด้วยการเลือกการตอบ k ,

$P_{i(k-1)}(\theta_n)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง (θ) คนที่ n จะตอบข้อคำถามข้อที่ i ด้วยการเลือกการตอบ $k-1$,

θ_n คือ คุณลักษณะแฝงคนที่ n ,

δ_i คือ ค่าความยากทั้งหมด (overall difficulty) ของข้อคำถามข้อที่ i ,

τ_k คือ ค่าพารามิเตอร์ threshold สำหรับทุกข้อคำถาม

2.2 โมเดลการตอบสนองข้อสอบแบบ
พหุมิติ Rating scale model

$$\ln\left(\frac{P_{ikd}(\theta_{nd})}{P_{i(k-1)d}(\theta_{nd})}\right) = \theta_{nd} - \delta_{id} - \tau_{kd}, \quad \dots(8)$$

$P_{ikd}(\theta_{nd})$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง (θ) คนที่ n จะตอบข้อคำถามข้อที่ i ด้วยการเลือกการตอบ k ในมาตรวัดย่อยหรือมิติ d ,

$P_{i(k-1)d}(\theta_{nd})$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง (θ) คนที่ n จะตอบข้อคำถามข้อที่ i ด้วยการเลือกการตอบ $k-1$ ในมาตรวัดย่อยหรือมิติ d ,

θ_{nd} คือ คุณลักษณะแฝงคนที่ n ในมาตรวัดย่อยหรือมิติ d ,

δ_{id} คือ ค่าความยากทั้งหมด (overall difficulty) ของข้อคำถามข้อที่ i ในมาตรวัดย่อยหรือมิติ d ,

τ_{kd} คือ ค่าพารามิเตอร์ threshold สำหรับทุก ข้อคำถาม ในมาตรวัดย่อยหรือมิติ d

3. Partial credit model

3.1 โมเดลการตอบสนองข้อสอบแบบ
เอกมิติ Partial credit model

$$\ln\left(\frac{P_{ik}(\theta_n)}{P_{i(k-1)}(\theta_n)}\right) = \theta_n - \delta_i - \tau_{ik}, \quad \dots(9)$$

$P_{ik}(\theta_n)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง (θ) คนที่ n จะตอบข้อคำถามข้อที่ i ด้วยการเลือกการตอบ k ,

$P_{i(k-1)}(\theta_n)$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง (θ) คนที่ n จะตอบข้อคำถามข้อที่ i ด้วยการเลือกการตอบ $k-1$,

θ_n คือ คุณลักษณะแฝงคนที่ n ,

δ_i คือ ค่าความยากทั้งหมด (overall difficulty) ของข้อคำถามข้อที่ i ,

τ_{ik} คือ ค่าพารามิเตอร์ threshold สำหรับ ข้อคำถามข้อที่ i

3.2 โมเดลการตอบสนองข้อสอบแบบพหุ
มิติ Partial credit model

$$\ln\left(\frac{P_{ikd}(\theta_{nd})}{P_{i(k-1)d}(\theta_{nd})}\right) = \theta_{nd} - \delta_{id} - \tau_{ikd}, \quad \dots(10)$$

$P_{ikd}(\theta_{nd})$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง (θ) คนที่ n จะตอบข้อคำถามข้อที่ i ด้วยการเลือกการตอบ k ในมาตรวัดย่อยหรือมิติ d ,

$P_{i(k-1)d}(\theta_{nd})$ คือ ความน่าจะเป็นที่ผู้ตอบที่มีคุณลักษณะแฝง (θ) คนที่ n จะตอบข้อคำถามข้อที่ i ด้วยการเลือกการตอบ $k-1$ ในมาตรวัดย่อยหรือมิติ d ,

θ_{nd} คือ คุณลักษณะแฝงคนที่ n ในมาตรวัดย่อยหรือมิติ d ,

δ_{id} คือ ค่าความยากทั้งหมด (overall difficulty) ของข้อคำถามข้อที่ i ในมาตรวัดย่อยหรือมิติ d ,

τ_{ikd} คือ ค่าพารามิเตอร์ threshold สำหรับข้อคำถามข้อที่ i ในมาตรวัดย่อยหรือมิติ d



การประยุกต์ใช้

1. การตรวจสอบความเที่ยง ซึ่งเป็นการตรวจสอบความคงเส้นคงวาของการวัดด้วยวิธีการวิเคราะห์พหุมิติ เรียกความเที่ยงนี้ว่า EAP reliability ด้วยการประมาณค่าแบบมาร์จิ้นแนลแมกซ์ลิคไลฮูด (marginal maximum-likelihood; MML) ดังสูตรของ mislevy และคนอื่นๆ (1992) ที่เสนอไว้ดังนี้

$$\rho_{MML} = \frac{\sigma_{EAP}^2}{\sigma^2}$$

เมื่อ

σ_{EAP}^2 แทน ความแปรปรวนของการประมาณค่าแบบ expected a posterior (EAP)

σ^2 แทน ความแปรปรวนของคุณลักษณะแฝง (latent trait)

ความเที่ยงแบบ EAP reliability เป็นการประยุกต์ค่าความเที่ยงของแบบสอบตามทฤษฎีการตอบสนองข้อสอบ (Item response theory) ให้มีรูปแบบที่คล้ายกับความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical test theory) ดังข้อสรุปของ Adams (2005) สรุปว่า ความเป็นจริงแล้วค่าความเที่ยงทั้งตามทฤษฎีการทดสอบแบบดั้งเดิมเป็นค่าที่เหมือนกับทฤษฎีการตอบสนองข้อสอบที่ใช้โมเดลมาร์จิ้นแนล (marginal model) โดยสามารถใช้ค่าความเที่ยงในการแสดงถึงคุณภาพของการออกแบบการวัด (measurement design) ดังนั้นในการพิจารณาค่าความเที่ยงจึงพิจารณาตามเกณฑ์เดียวกันกับความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม โดย Nunnally และ Bernstein (1994); Hair และคนอื่นๆ (2006) เสนอเกณฑ์การพิจารณาค่าความเที่ยงด้วยสูตรสัมประสิทธิ์แอลฟาไวส์อดคล้องกัน โดยเสนอให้พิจารณายอมรับค่า .700 ขึ้นไป Cohen และ Swerdlik (2005) เสนอเกณฑ์การพิจารณาค่าความเที่ยงด้วยสูตรสัมประสิทธิ์แอลฟาพิจารณายอมรับค่า .800 ขึ้นไป และ Peterson (1994) ได้ทำการวิเคราะห์ห่อภิมาณ (Meta-analysis) ค่าความเที่ยงที่ใช้สูตรสัมประสิทธิ์แอลฟา จำนวนทั้งสิ้น 4,286 ค่า พบว่า ค่ามัธยฐานของค่าความเที่ยงที่ใช้สูตรสัมประสิทธิ์แอลฟามีค่าเท่ากับ .790

2. การตรวจสอบความตรงเชิงโครงสร้าง ซึ่งเป็นการตรวจสอบความถูกต้องแม่นยำของการวัด

โครงสร้างที่มีลักษณะพหุมิติ ตรวจสอบด้วยวิธีการวิเคราะห์พหุมิติ

ในการตรวจสอบความตรงเชิงโครงสร้างด้วยวิธีการของความเป็นพหุมิตินั้น มีวิธีการแสดงหลักฐานความเป็นพหุมิติ โดยเป็นการประเมินพิจารณาเมื่อการวัดมีลักษณะที่มุ่งวัดเป็นแบบพหุมิติหรือการวิเคราะห์ที่มุ่งจะชี้วัดความเป็นเอกมิติไม่มีความเหมาะสม จึงต้องมีการใช้โมเดลการวัดแบบพหุมิติช่วยยืนยันโครงสร้างให้มีความเหมาะสมยิ่งขึ้น (Wolfe และ Smith, 2007) ตัวชี้วัดที่จะแสดงถึงความตรงเชิงโครงสร้างนั้นพิจารณาโดยหากโมเดลซ้อนสัมพันธ์กัน (nested) หรือสัมพันธ์กันลักษณะเชิงลดหลั่น (hierarchically related) นั่นก็คือ โมเดลแบบพหุมิติ (multidimensional approach) กับ โมเดลแบบเอกมิติรวม (composite approach) (Briggs และ Wilson, 2003; Allen และ Wilson, 2006) ประเมินเปรียบเทียบโมเดลจากค่าสถิติดีวีเนียนซ์ (Deviance Statistic; G^2) ซึ่งเป็นสถิติทดสอบอัตราส่วนไลค์ลิฮูด (likelihood-ratio-test) หรือเรียกอีกอย่างหนึ่งว่า สถิติทดสอบไคสแควร์อัตราส่วนไลค์ลิฮูด (likelihood ratio chi-squared statistic) โดยจะเป็นสถิติที่ใช้ในการประเมินบ่งชี้ถึงความเหมาะสมของโมเดล (Hoskens และ Boeck, 2001) และหากโมเดลไม่ซ้อนสัมพันธ์กัน (non-nested) หรือไม่สัมพันธ์กันลักษณะเชิงลดหลั่น (non hierarchically related) ดังเช่น โมเดลแบบพหุมิติ (multidimensional approach) กับ โมเดลแบบเอกมิติแยกตามมิติ (consecutive approach) ประเมินเปรียบเทียบจากเกณฑ์สารสนเทศไอเคคิ (Akaike Information Criterion; AIC) (Briggs และ Wilson, 2003; Allen และ Wilson, 2006) โดยจะเป็นค่าสถิติที่ใช้ในการประเมินบ่งชี้ถึงความเหมาะสมของโมเดล เช่นกัน (Hoskens และ Boeck, 2001)

ในการแสดงหลักฐานความตรงเชิงโครงสร้างด้วยวิธีการของความเป็นพหุมิติจะเป็นการแสดงหลักฐานจากการประเมินเปรียบเทียบโมเดล (competing models) ระหว่างโมเดลโครงสร้างแบบพหุมิตีกับโมเดลโครงสร้างเอกมิติ โดยพิจารณาความสอดคล้องกลมกลืนของโมเดลคุณลักษณะแฝงกับข้อมูลว่าโมเดลโครงสร้างคุณลักษณะแฝงใดจะมีความสอดคล้องกับข้อมูลมากกว่ากัน ประเมินโดยใช้สถิติดีวีเนียนซ์ (Deviance Statistic; G^2) ซึ่งเป็นดัชนี

บ่งชี้ถึงความเหมาะสมของโมเดล (Allen และ Wilson, 2006; Liu, Wilson และ Paek, 2008) และเป็นสถิติที่ใช้ในการเปรียบเทียบโมเดลที่มีการซ้อนสัมพันธ์กัน (nested models) คือ การเปรียบเทียบระหว่างโมเดลแบบพหุมิติ (multidimensional approach) กับโมเดลแบบเอกมิตินรวม (composite approach) คำนวณได้จากสูตร $G^2 = -2(L_M - L_S)$ โดย L_M คือ ค่า maximum log-likelihood สำหรับโมเดลที่ศึกษา และ L_S คือ ค่า maximum log-likelihood สำหรับโมเดลอิ่มตัว (saturated model) (Agresti, 1996; Wolfe และ Smith, 2007) การเปรียบเทียบความเหมาะสมเชิงสัมพัทธ์ (relative fit) ของสองโมเดลคือโมเดลเต็มรูป (fuller model) กับโมเดลซ้อนสัมพันธ์ (nested model) สามารถประเมินโดยการเปรียบเทียบจากค่าสถิติดีเวียนซ์ (Deviance Statistic) ของสองโมเดล ด้วยวิธีการทางสถิติที่เรียกว่า สถิติทดสอบไคสแควร์อัตราส่วนไคสแควร์ (likelihood ratio chi-squared statistic) ดังสูตร $G^2(M_{\text{Nested}} | M_{\text{Fuller}}) = \chi^2_{LR} = G^2_N - G^2_F$ โดย χ^2_{LR} คือ การแจกแจงที่มีลักษณะเหมือนสถิติไคสแควร์กับองศาความเป็นอิสระ (degree of freedom) ซึ่งองศาความเป็นอิสระก็คือความแตกต่างระหว่างจำนวนของการประมาณค่าพารามิเตอร์ของสองโมเดล (McCullagh และ Nelder, 1990 อ้างถึงใน Wolfe และ Smith, 2007; Allen และ Wilson, 2006; Liu, Wilson และ Paek, 2008) สำหรับค่าสถิติดีเวียนซ์ (Deviance Statistic) จะลดลงตามการเพิ่มขึ้นของพารามิเตอร์โมเดล (Allen และ Wilson, 2006) ทั้งนี้ดัชนีการเปรียบเทียบโมเดลคือ สถิติดีเวียนซ์ (Deviance Statistic) ที่กล่าวมา มีความไวต่อขนาดกลุ่มตัวอย่างในการตีความหมายอาจจะประยุกต์ใช้ค่าสัดส่วนคงที่ (proportionality constant; PC) สามารถคำนวณจากสูตร $PC = G^2/df$ (ในกรณีศึกษาค่าสัดส่วนคงที่ (PC) จากสูตรดังกล่าว $df = \text{จำนวนผู้ตอบ} - \text{จำนวนพารามิเตอร์}$) (Agresti, 1996; Wolfe และคนอื่นๆ, 2007) ซึ่งค่าสัดส่วนคงที่เป็นดัชนีที่บ่งชี้ถึงค่าเฉลี่ยรวมของความแปรปรวนคะแนนสังเกตที่อธิบายได้ด้วยพารามิเตอร์ของโมเดล

สำหรับการประเมินโมเดลไม่ซ้อนสัมพันธ์กัน (non-nested) พิจารณาจากเกณฑ์สารสนเทศเอไค (Akaike information criterion; AIC) คำนวณมาจาก

สูตร $AIC = (-2) \log \text{maximum likelihood} + 2$ (จำนวนพารามิเตอร์) (Akaike, 1987) หรือกล่าวอีกนัยหนึ่งนั่นคือสูตร $AIC = G^2 + 2p$ (เมื่อ p คือ จำนวนพารามิเตอร์) นั่นเอง (Yao และ Schwarz, 2006) ซึ่งเป็นค่าสถิติที่ใช้ในการเปรียบเทียบความเหมาะสมเชิงสัมพัทธ์ (relative fit) ของสองโมเดลที่ไม่มีการซ้อนสัมพันธ์กัน (nonnested models) และยังเป็นสถิติที่สามารถที่จะอธิบายความแตกต่างจำนวนพารามิเตอร์ระหว่างสองโมเดลได้ โดยเป็นการเปรียบเทียบระหว่างโมเดลแบบพหุมิติกับโมเดลแบบเอกมิติแยกตามมิติ ซึ่งโมเดลแบบเอกมิติแยกตามมิติ จะเป็นโมเดลที่มีโครงสร้างเอกมิติที่อิสระจากกัน ประมาณค่าความสามารถคุณลักษณะแฝงในแต่ละมิติอย่างอิสระจากกัน ซึ่งในแต่ละมิติสามารถทำการวิเคราะห์โดยการรวมคะแนนจากข้อคำถามในแต่ละมิติเพื่อทำการวิเคราะห์ เมื่อต้องทำการเปรียบเทียบโมเดลเพื่อสะท้อนโมเดลที่มีลักษณะเอกมิติแยกตามมิติในภาพรวม สามารถคำนวณค่าสถิติของโมเดลได้จากการรวมค่าสถิติที่ได้ในแต่ละมิติเข้าด้วยกัน (Agresti, 1984; Briggs และ Wilson, 2003; Allen และ Wilson, 2006) เช่น โมเดลเอกมิติแยกตามมิติ มี 3 มิติ หากต้องคำนวณค่าสถิติดีเวียนซ์ (Deviance Statistic; G^2) และจำนวนพารามิเตอร์ (p) คำนวณได้จากการรวมค่าสถิติทุกมิติเข้าด้วยกัน (consecutive approach $G^2 = G^2_{d1} + G^2_{d1} + G^2_{d3}$ และจำนวนพารามิเตอร์ $= P_{d1} + P_{d2} + P_{d3}$, เมื่อ d คือ มิติ) แต่ทั้งนี้สถิติ AIC ยังเป็นสถิติที่ยากต่อการตีความหมายของข้อมูล Wolfe และคนอื่นๆ (2007) จึงเสนอแนะให้ประยุกต์ใช้สถิติ PC_{AIC} ซึ่งเป็นสัดส่วนคงที่ของค่า AIC สามารถคำนวณจากสูตร $PC_{AIC} = AIC/df$ ซึ่งมีความคล้ายคลึงกับค่าสัดส่วนคงที่ (proportionality constant; PC)

ทั้งนี้ในการตีความหมายของค่าสถิติดีเวียนซ์ (Deviance Statistic; G^2) และเกณฑ์สารสนเทศเอไค (Akaike Information Criterion; AIC) พิจารณาจากโมเดลใดมีค่าน้อยจะแสดงถึงโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์มากกว่า และสถิติทั้งสองยังเป็นตัวบ่งชี้ถึงโมเดลที่คาดว่าจะดีที่สุดจากการพิจารณาจากความแตกต่างระหว่างโมเดล ซึ่งจะนำโมเดลไปใช้ในการตีความหมายอย่างมีนัยสำคัญทางสถิติ (Briggs และ Wilson, 2003; Allen และ Wilson, 2006)



3. การตรวจสอบความเหมาะสมรายข้อ

ค่าสถิติความเหมาะสมรายข้อตามทฤษฎีการตอบสนองข้อสอบ (IRT item fit statistics) เป็นค่าสถิติที่ระบุว่าข้อมูลรายข้อนั้นสอดคล้องกับโมเดล IRT ที่กำหนดไว้หรือไม่ ข้อมูลรายข้อที่สอดคล้องกับโมเดล IRT แสดงว่าข้อนั้นมีรูปแบบการตอบที่สอดคล้องกับโมเดลที่มีลักษณะเป็นรูปโค้งความถึ่สะสม กล่าวคือคนที่มีความสามารถในการสังเกตสูงมีโอกาสในการตอบข้อสอบได้ดีกว่าคนที่มีความสามารถในการสังเกตต่ำ

ดัชนีชี้วัดความเหมาะสมของข้อสอบรายข้อมี 2 ค่าคือ OUTFIT MNSQ (OUTFIT Mean Square หรือ unweighted Mean Square) และ INFIT MNSQ (INFIT Mean Square หรือ weighted Mean Square) โดยทั้งสองค่านี้อยู่บนฐานแนวคิดของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนมาตรฐานระหว่างผลการตอบข้อสอบและโมเดลการวัดนั้น (DeMars, 2004)

โดย Wright และ Masters (1982 อ้างถึงใน ชัยวิชิต เขียรชนะ, 2552) เสนอให้ใช้ค่า OUTFIT MNSQ และ INFIT MNSQ ที่มีค่าอยู่ระหว่าง .75 ถึง 1.33 Lunz, Wright, และ Linacre (1990 อ้างถึงใน ชัยวิชิต เขียรชนะ, 2552) เสนอให้ใช้ค่าที่อยู่ระหว่าง .60 ถึง 1.50 และ Roiser (1995); Mok และคนอื่นๆ (2006 อ้างถึงใน ชัยวิชิต เขียรชนะ, 2552) เสนอแนะให้ใช้ค่าที่อยู่ระหว่าง .70 ถึง 1.30 จึงจะแสดงถึงโมเดลมีความเหมาะสมกับข้อมูลรายข้อ นอกจากนี้แล้ว Wright และคนอื่นๆ (1994 อ้างถึงใน ชัยวิชิต เขียรชนะ, 2552) ยังได้เสนอแนะให้ใช้ค่า OUTFIT MNSQ และ INFIT MNSQ ให้มีความเหมาะสมกับประเภทการสอบวัดต่างๆ โดยมีรายละเอียดดังตารางที่ 1 และการตีความหมายจากค่าสถิติ OUTFIT MNSQ และ INFIT MNSQ ดังตารางที่ 2

ตารางที่ 1 ช่วงการพิจารณาค่าสถิติความเหมาะสมรายข้อตามประเภทการสอบ

ประเภทการสอบวัด	ช่วงของการพิจารณาค่า OUTFIT MNSQ และ INFIT MNSQ
การวัดที่มีผลกระทบสูงๆ (high stakes)	.80 ถึง 1.20
การวัดทั่วไป	.70 ถึง 1.30
มาตรประเมินค่า (rating scale) (เชิงสำรวจ)	.60 ถึง 1.40
การวัดเชิงคลินิก (clinical observation)	.50 ถึง 1.70
การวัดที่ต้องลงความเห็น (judged)	.40 ถึง 1.20

ตารางที่ 2 การตีความหมายของระดับค่าสถิติความเหมาะสมรายข้อ

ระดับของค่า OUTFIT MNSQ และ INFIT MNSQ	การตีความหมาย
> 2.00	ข้อสอบผิดเพี้ยนออกไปจากโครงสร้างการวัด (distorts or degrading)
1.50 ถึง 2.00	ข้อสอบไม่มีความเหมาะสมสำหรับโครงสร้างการวัด แต่ข้อสอบไม่ได้ผิดเพี้ยนออกไปจากโครงสร้างการวัด (not degrading)
0.50 ถึง 1.50	ข้อสอบมีความเหมาะสมสำหรับโครงสร้างการวัด
< .50	ข้อสอบมีความเหมาะสมน้อยสำหรับโครงสร้างการวัด แต่ข้อสอบไม่ได้ผิดเพี้ยนออกไปจากโครงสร้างการวัด (not degrading) ทั้งนี้อาจจะมีผลทำให้ค่าความเที่ยง (reliabilities) ผิดเพี้ยนไป

บทส่งท้าย

ธรรมชาติโครงสร้างคุณลักษณะที่มุ่งวัดมีทั้งแบบเอกมิติและพหุมิติ ซึ่งมีความจำเป็นอย่างยิ่งที่ต้องตรวจสอบคุณลักษณะที่มุ่งวัดว่ามีลักษณะเอกมิติหรือพหุมิติ การตรวจสอบเหล่านี้ไม่เดลการตอบสนองข้อสอบแบบพหุมิติมีความสามารถที่จะสะท้อนถึงความเหมาะสมว่า คุณลักษณะที่มุ่งวัดเหมาะสมที่จะอธิบายด้วยลักษณะใด การวิจัยเพื่อตรวจสอบคุณภาพเครื่องมือที่ประยุกต์ใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติในต่างประเทศได้เริ่มมีการแสดงหลักฐานบ้าง ส่วนงาน

วิจัยในประเทศไทยมีเพียงงานวิจัยของผู้เขียนบทความที่ได้ศึกษาในปี พ.ศ. 2552 เรื่อง การพัฒนาแบบวัดกลยุทธ์การเรียนรู้แบบพหุมิติสำหรับนักเรียนมัธยมศึกษาตอนปลาย ทั้งนี้ นอกจากการตรวจสอบลักษณะโครงสร้างคุณลักษณะที่วัดด้วยการประยุกต์ใช้โมเดลการตอบสนองข้อสอบ (MRCMLM) แล้ว โมเดลความสัมพันธ์โครงสร้างเชิงเส้น (LISREL) ยังสามารถประยุกต์ใช้ในการตรวจสอบได้เช่นกัน จึงมีความท้าทายอย่างยิ่งในการศึกษาเรื่องคุณลักษณะที่มุ่งวัดว่ามีความเหมาะสมที่จะทำการวัดและการแปลความหมายในลักษณะเอกมิติหรือลักษณะพหุมิติ

เอกสารอ้างอิง

ชัยวิชิต เขียรชนะ. (2552). การพัฒนาแบบวัดกลยุทธ์การเรียนรู้แบบพหุมิติสำหรับนักเรียนมัธยมศึกษา ตอนปลาย.

วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต. สาขาวิชาการวัดและประเมินผลการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation*, 31, 162-172.

Adams, R., Wilson, M., and Wang, W. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21, 1-23.

Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.

_____. (1996). *An introduction to categorical data analysis*. New York: Wiley.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, 52(3), 317-332.

Allen, D.D., and Wilson, M. (2006). Introducing multidimensional item response modeling in the behavioral sciences. *Health Education Research Theory & Practice* [Online]. Available from: <http://her.oxfordjournals.org/cgi/reprint/cyl086v2> [2007, October 12].

Briggs, D.C., and Wilson, M. (2003). An introduction to multidimensional measurement using rasch model. *Journal of applied measurement*, 4(1), 87-100.

Cohen, R.J., and Swerdlik, M.E. (2005). *Psychological testing and assessment: An introduction to tests and measurement*. 6th ed. Boston: McGraw-Hill.

DeMars, C. (2004). Measuring higher education outcomes with a multidimensional rasch model. *Journal of Applied Measurement*, 5(4), 350-361.

Hair, J.F., Black, W.C., Babin, B.J., Anderson, R.E., and Tatham, R.L. (2006). *Multivariate Data Analysis*. 6th ed. New Jersey: Pearson Education.

Hoskens, M., and Boeck, P.D. (2001). Multidimensional componential item response theory models for polytomous items. *Applied Psychological Measurement*, 25(1), 19-37.



- Liu, O.L., Wilson, M., and Paek, I. (2008). A multidimensional rasch analysis of gender differences in PISA mathematics. **Journal of Applied Measurement**, 9(1), 18-35.
- Mislevy, R.J., Beaton, A.E., Kaplan, B., and Sheehan, K.M. (1992). Estimating population characteristics from sparse matrix samples of item responses. **Journal of Educational Measurement**, 29, 133-136.
- Morizot, J., Ainsworth, A.T., and Reise, S. (2007). Toward modern psychometrics: Application of item response theory models. In R.W. Robins, R.C. Fraley, and R.F. Krueger (eds.), pp. 407-423. **Handbook of research methods in personality psychology**. New York: Guilford Press.
- Nunnally, J.C., and Bernstein, I.H. (1994). **Psychometric theory**. 3rd ed. New York: McGraw-Hill.
- Peterson, R.A. (1994). A meta-analysis of cronbach's coefficient alpha. **Journal of consumer research**, 21, 381-391.
- Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. **Journal of Educational Statistics**, 4(3), 207-230.
- Wang, W-C. (2004). Direct estimation of correlation as a measure of association strength using multidimensional item response models. **Educational and Psychological Measurement**, 64(6), 937-955.
- Wilson, M., and Hoskens, M. (2005). Multidimensional item response: Multimethod/ Multitrait perspective. In S. Alagumalai, D.D. Curtis, and N. Hungi (eds.), pp.287-307. **Applied rasch measurement: A book of exemplars papers in honour of John P.Keeves**. Netherlands: Springer.
- Wolfe, E.W., and Smith, E.V. (2007). Instrument tools and activities for measure validation using rasch models: Part II – Validation activities. **Journal of Applied Measurement**, 8(2), 204-234.
- Wolfe, E. V., Viger, S.G., Jarvinen, D.W., and Linksman, J. (2007). Validation of scores from a measure of teachers'efficacy toward standards-aligned classroom assessment. **Educational and Psychological Measurement**, 67(3), 460-474.
- Yao, L., and Schwarz, R.D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. **Applied Psychological Measurement**, 30(6), 469-492.