

การใช้โมเดลเชิงเส้นระดับชั้นในการอธิบายความแปรปรวนในการให้ข้อมูลป้อนกลับแบบ 360 องศา
 (The use of hierarchical linear model for explaining variance in 360 degree feedback)

สันต์ทัด พรประเสริฐมานิต จูติธิดา สุขทอง จุฑาทิพย์ วิวัฒนาพันธุ์วงศ์
 คณะจิตวิทยา จุฬาลงกรณ์มหาวิทยาลัย

การวัดความเที่ยงในทางจิตวิทยา เป็นการวัดว่าในการประเมินแต่ละครั้งมีความคงเส้นคงวามากน้อยเพียงใด แหล่งที่ทำให้ความเที่ยงของการวัดลดลงมีได้หลายรูปแบบ เช่น ความผิดพลาดที่เกิดจากผู้ประเมิน (Rater effect) ความผิดพลาดที่เกิดจากเวลาที่ประเมิน (Occasion effect) ความผิดพลาดที่เกิดจากความยากง่ายของชุดของแบบวัดที่แตกต่างกัน (Item effect)

ทฤษฎีการวัดแบบดั้งเดิม (Classical test theory) เชื่อว่าคะแนนของวัตถุที่ได้รับการวัด (Object) เช่น บุคคล องค์กร ฯลฯ หรือเรียกว่าคะแนนที่สังเกตได้ (Observed score) เกิดจากคะแนน 2 ส่วนรวมกัน คือ คะแนนที่แท้จริงของวัตถุนั้น (True score) คะแนนที่แท้จริงจะคงที่ในแต่ละวัตถุ ค่านี้จะคะแนนของแต่ละบุคคลที่สะท้อนถึงวัตถุประสงค์ในการวัด แต่สาเหตุที่ทำให้คะแนนที่วัดแต่ละครั้งในคนเดียวกันไม่คงที่ เกิดจากคะแนนอีกส่วนหนึ่ง คือ คะแนนที่เกิดจากความผิดพลาด (Error score) เช่น เวลาในการวัด มาตราวัด เป็นต้น คะแนนที่เกิดจากความผิดพลาดเหล่านี้จะเกิดขึ้นในแต่ละครั้งของการวัด และจะเกิดขึ้นโดยสุ่ม (Random) สามารถเขียนได้ดังสมการ

$$O = T + E$$

O คือ คะแนนที่สังเกตได้

T คือ คะแนนจริงของบุคคลนั้น

E คือ คะแนนคาดเคลื่อนโดยสุ่ม

เนื่องจากคะแนนที่เกิดจากความผิดพลาดเกิดขึ้นโดยสุ่ม ทำให้เมื่อวัดจนได้คะแนนหลายๆ ค่า เช่น ด้วยวิธีการวัดหลายๆ ครั้ง วัดด้วยหลายๆ ข้อกระทง ประเมินด้วยผู้ประเมินหลายๆ คน จะทำให้คะแนนที่เกิดจากความผิดพลาดถูกหักล้างกันไป จนอาจพูดได้ว่าถ้าทำให้ได้คะแนนจนถึงอนันต์ครั้ง จะทำให้คะแนนที่วัดได้เท่ากับคะแนนที่แท้จริงของวัตถุนั้น

$$\lim_{n \rightarrow \infty} O = T$$

วิธีการวัดจนถึงอนันต์ครั้งไม่สามารถทำได้ในความเป็นจริง ในความเป็นจริงที่สามารถทำได้ คือ พยายามทำให้ได้คะแนนมาหลายๆ คะแนน แล้วบอกว่าคะแนนหลายๆ ค่านี้สามารถเชื่อถือได้มากน้อยเพียงใด ค่าตัวเลขที่จะบอกนี้เรียกว่า ความเที่ยง (Reliability)

เนื่องจากคะแนนที่ต้องการคือคะแนนที่แท้จริงของวัตถุนั้น และไม่ต้องการคะแนนที่เกิดจากความผิดพลาด จึงอาจพูดได้ว่าถ้าคะแนนที่สังเกตได้เกิดจากคะแนนจริงเปลี่ยนแปลงไปมาก จะเป็นสิ่งที่น่าพึงปรารถนา แต่ถ้าคะแนนที่เปลี่ยนแปลงไปเกิดจากคะแนนความผิดพลาด จะเป็นสิ่งไม่น่าพึงปรารถนาในการวัด ในที่นี้ขอยกตัวอย่างประกอบกรอธิบายในการหาค่าความเที่ยง

บันทึกจากผู้เขียน

รายงานฉบับนี้ถูกเขียนขึ้นเมื่อ มีนาคม 2006 ประกอบรายวิชา การประเมินผลในองค์กร คณะจิตวิทยา
 จุฬาลงกรณ์มหาวิทยาลัย

ติดต่อผู้เขียน: Sunthud Pornprasertmanit, Email: psunthud@gmail.com

ตัวอย่างในที่นี้เป็นการประเมินผลงานของตนเองจำนวน 5 คน ด้วยข้อกระทงจำนวน 5 ข้อ ถ้านำคะแนนทุกข้อกระทงมาหาค่าเฉลี่ย จะได้คะแนนจริงในการประเมินผลงานของบุคคล คะแนนแต่ละข้อกระทงเป็นคะแนนที่สังเกตได้ คะแนนแห่งความผิดพลาดในการประเมินนี้คือความผิดพลาดจากข้อกระทงที่วัดออกมาได้คะแนนแตกต่างกัน (ความผิดพลาดอื่นๆ ที่ไม่ได้สำรวจในการประเมินครั้งนี้) จากสมการการวัดข้างต้นสามารถเขียนได้ดังสมการ

$$O_{ij} = T_{0j} + e_{ij}$$

O_{ij} เป็นคะแนนที่สังเกตได้ในการวัดข้อกระทง i ในคนที่ j

T_{0j} เป็นคะแนนที่แท้จริงของแต่ละบุคคลที่ j เป็นคะแนนที่สะท้อนถึงวัตถุประสงค์ของการวัด

e_{ij} เป็นคะแนนที่เปลี่ยนแปลงไปเนื่องจากข้อกระทงที่ i ในบุคคลที่ j

คะแนนที่สังเกตได้เกิดจากคะแนนที่แท้จริงของแต่ละบุคคล รวมกับคะแนนความผิดพลาดที่เกิดจากข้อกระทง ถ้านำคะแนนของแต่ละคนมาหาค่าเฉลี่ย จะได้ค่าประมาณของประชากรโดยทั่วไป และเกิดค่าเบี่ยงเบนที่เกิดจากความแตกต่างระหว่างบุคคล เห็นได้ดังสมการ

$$T_{0j} = G_{00} + u_{0j}$$

T_{0j} เป็นคะแนนที่แท้จริงของแต่ละบุคคลที่ j เป็นคะแนนที่สะท้อนถึงวัตถุประสงค์ของการวัด

G_{00} เป็นค่าเฉลี่ยของผู้ทดสอบทั้งหมด ซึ่งเป็นค่าประมาณค่าเฉลี่ยของประชากร

u_{0j} คือค่าความที่อธิบายได้จากบุคคลที่ j

ถ้านำทั้งสองสมการมารวมกันจะได้สมการดังนี้

$$O_{ij} = G_{00} + u_{0j} + e_{ij}$$

จะเห็นว่าสมการล่าสุดจะคล้ายกับโมเดลเชิงเส้นแบบทั่วไป (General linear model) ในการวิเคราะห์ความแปรปรวน

$$X = \text{grand mean} + \text{group deviation} + \text{within-group deviation}$$

จากความรู้เรื่องการวิเคราะห์ความแปรปรวนแบบสุ่ม (Random variable ANOVA) จะเห็นว่าค่าที่คงที่ในสมการนี้มีเพียงค่าประมาณประชากรเท่านั้น ส่วนค่าความแตกต่างระหว่างบุคคลและค่าความผิดพลาดเป็นสิ่งที่เกิดขึ้นโดยสุ่ม ดังนั้นความเบี่ยงเบนของคะแนนที่สังเกตได้ เกิดจากการเปลี่ยนแปลงจากความแตกต่างระหว่างบุคคลและค่าความผิดพลาด และเนื่องจากค่าเบี่ยงเบนทั้งสองไม่สัมพันธ์กัน ดังนั้นความแปรปรวนของคะแนนที่สังเกตได้เกิดจากความแปรปรวนของความแตกต่างระหว่างบุคคลและความแปรปรวนของค่าความผิดพลาด ดังสมการ

$$SS_{\text{Observed}} = SS_{\text{Individual}} + SS_{\text{Error from items}}$$

ดังนั้นอาจเปลี่ยนรูปสมการได้เป็น

$$1 = \frac{SS_{\text{individual}}}{SS_{\text{observed}}} + \frac{SS_{\text{error}}}{SS_{\text{observed}}}$$

อัตราส่วนระหว่างความแปรปรวนที่เกิดจากความแตกต่างระหว่างบุคคลต่อความแปรปรวนที่เกิดจากคะแนนที่สังเกตได้ทั้งหมดก็คือความเที่ยงของการวัด ด้วยเหตุนี้ ค่าความเที่ยงจึงมีสูตรดังนี้

$$R_{xx} = 1 - \frac{SS_{\text{error}}}{SS_{\text{observed}}}$$

จากสูตรข้างบนจะคล้ายกับสูตรค่าสัมประสิทธิ์ความเที่ยงแบบอัลฟาของครอนบาค

$$R_{xx} = \left(\frac{n}{n-1}\right)\left(1 - \frac{\sum SS_{each\ item}}{SS_{observed}}\right)$$

ถ้าความเที่ยงของการวัดสูงแสดงว่าความแปรปรวนที่เกิดจาก item มีน้อย การวัดจึงคงเส้นคงวา แต่ถ้าความเที่ยงน้อย ความแปรปรวนจากความผิดพลาดมาก การวัดไม่คงเส้นคงวา

ความผิดพลาดจากสูตรข้างต้นสามารถประยุกต์ใช้ได้กับการหาความผิดพลาดได้หลายรูปแบบ เช่น ความผิดพลาดจากผู้ประเมิน ความผิดพลาดจากเวลาในการประเมิน แต่ว่าวิธีนี้มีข้อด้อยคือว่าไม่สามารถแบ่งแยกระหว่างความผิดพลาดที่สนใจ กับความผิดพลาดอื่นๆ ที่ไม่ได้สนใจ (Residual variance) กล่าวคือ ความแปรปรวนที่เกิดจากข้อกระทง อาจจะมีค่าความแปรปรวนที่เกิดจากความผิดพลาดอื่นๆ เข้ามาด้วย ถ้าไม่ต้องการให้มีความผิดพลาดเหล่านี้ จะต้องควบคุมให้เป็นมาตรฐาน เช่น ผู้ทดสอบจะต้องมีลักษณะปฏิสัมพันธ์แบบเดียวกันกับทุกคน เป็นต้น แต่ในบางครั้งไม่สามารถควบคุมได้ เช่น ผลที่เกิดจากเวลา ทำให้อาจจะมีค่าความแปรปรวนที่เกิดจากความผิดพลาดอื่นเข้ามาปะปน

อย่างไรก็ตาม ในกรณีการหาค่าความเที่ยงโดยปกติ การหาความเที่ยงโดยวิธีนี้ก็ยังสามารถใช้ได้ เพราะไม่ได้สนใจว่าความผิดพลาดนั้นเกิดจากแหล่งความแปรปรวนใด แต่ถ้ากรณีที่ต้องการรู้สาเหตุแห่งผิดพลาดว่าเกิดจากเหตุใดบ้าง ทฤษฎีความเที่ยงแบบดั้งเดิมยังทำไม่ได้

การหาความเที่ยงด้วยการวิเคราะห์ความแปรปรวนด้วยวิธีของ Hoyt

ข้อด้อยข้างต้นเป็นจุดอ่อนของทฤษฎีการวัดแบบดั้งเดิม อาจแก้ไขได้โดยการปรับปรุงการคำนวณโดยมองคะแนนจริงแทนที่จะเป็นคะแนนจริงในแต่ละคน มองให้เป็นคะแนนจริงที่แต่ละคนที่ได้ในแต่ละข้อแทน จะคำนวณได้ดังสมการ

Observed score of each person and each item = True score + Error score

$$O_{jk} = T_{jk} + e_{jk}$$

O_{jk} คือ ค่าคะแนนที่สังเกตได้ของแต่ละบุคคลที่ j ในข้อกระทงที่ k

T_{jk} คือ ค่าคะแนนที่แท้จริงของบุคคลที่ j ในข้อกระทงที่ k

e_{jk} คือ ค่าความผิดพลาดอื่นๆ ของบุคคลที่ j ในข้อกระทงที่ k

แต่ว่าไม่สามารถหาคะแนนจริงได้โดยการหาค่าเฉลี่ย เพราะว่าค่าที่สังเกตได้มีเพียงแค่ค่าเดียวในการประมาณค่า แต่ค่าคะแนนจริงสามารถประมาณค่าได้จากอีกสมการหนึ่ง คือ

$$T_{jk} = G_{00} + u_{j0} + v_{0k} + w_{jk}$$

T_{jk} คือ ค่าคะแนนที่แท้จริงของบุคคลที่ j ในข้อกระทงที่ k

G_{00} คือ ค่าคะแนนเฉลี่ยของทุกคนและทุกข้อ

u_{j0} คือ ค่าคะแนนที่สามารถอธิบายได้จากบุคคลที่ j

v_{0k} คือ ค่าคะแนนที่สามารถอธิบายได้จากข้อกระทงที่ k

w_{jk} คือ ค่าคะแนนที่สามารถอธิบายได้จากปฏิสัมพันธ์ระหว่างบุคคลที่ j และข้อกระทงที่ k

แต่ว่าผลปฏิสัมพันธ์ระหว่างบุคคลและข้อกระทงไม่สามารถวัดได้ เนื่องจากแต่ละบุคคลตอบแต่ละข้อกระทงเพียงแค่ครั้งเดียว แต่จริงๆ แล้วผลจากสิ่งนี้ยังคงมีอยู่ ทำให้ผลปฏิสัมพันธ์นี้ไปรวมอยู่กับค่าความผิดพลาดอื่นๆ ทำให้สมการรวมได้เป็น

$$O_{jk} = G_{00} + u_{j0} + v_{0k} + e_{jk}$$

ค่าความผิดพลาดในที่นี่จะรวมค่าปฏิสัมพันธ์ระหว่างบุคคลและข้อกระทง และค่าผิดพลาดอื่นๆ ไปด้วยกัน จากสมการข้างต้น ค่าเฉลี่ยจะเป็นค่าเดียวที่คงที่ ส่วนค่าอีก 3 ค่าที่เหลือ จะเป็นค่าที่เปลี่ยนแปลงได้โดยสุ่ม (Random effect) ดังนั้นความแปรปรวนของค่าที่สังเกตได้จะเกิดจากความแปรปรวน 3 ส่วนด้วยกัน ดังสมการ

$$SS_o = SS_p + SS_i + SS_e$$

ในการวิเคราะห์ความแปรปรวนตามแนวคิดของ Hoyt จะคำนวณด้วยสูตร

$$R_{xx} = 1 - \frac{MS_i}{MS_p}$$

โดย $MS_i = \frac{SS_i}{n_i - 1}$ และ $MS_p = \frac{SS_p}{n_p - 1}$

อย่างไรก็ตาม แนวคิดของ Hoyt จะเป็นแนวคิดในยุคเดิม ที่ต้องการนำความแปรปรวนที่อธิบายได้ด้วยข้อกระทง ออกจากคะแนนจริง ทำให้ไม่ได้สนใจความผิดพลาดที่เหลืออยู่ทั้งที่ความแปรปรวนที่เหลือเหล่านี้เป็นสิ่งที่น่าสนใจในการนำมาวิเคราะห์ต่อไป

ทฤษฎีการอ้างอิงโดยสรุป (Generalizability theory)

ทฤษฎีการวัดแบบดั้งเดิมเป็นการหาค่าความเที่ยง ทำให้รู้ความคงเส้นคงวาของการวัด แต่ไม่รู้แหล่งแห่งความผิดพลาดอย่างชัดเจน ถ้าจะทำให้รู้แหล่งแห่งความผิดพลาดชัดเจนจะสามารถทำได้โดยควบคุมแหล่งแห่งความผิดพลาดอื่นทั้งหมด (แต่ในทางปฏิบัติแล้วทำไม่ได้) แล้วหาแหล่งแห่งความผิดพลาดที่เกิดจากแหล่งดังกล่าว เช่น การหาความสัมพันธ์จากการวัดครั้งที่ 1 และการวัดครั้งที่ 2 จะพบว่าข้อกระทง บุคคลถูกควบคุมให้คงที่ แต่ให้เวลาแตกต่างกัน จะได้ค่าความเที่ยงในการวัดซ้ำ

ส่วนแนวคิดการวิเคราะห์ความแปรปรวนตามแนวคิดของ Hoyt ยังมีปัญหาที่ไม่ได้สนใจแหล่งความผิดพลาดอื่นๆ เลย ด้วยข้อผิดพลาดเหล่านี้ Lee Cronbach เสนอแนวคิดไว้ในหนังสือเรื่อง the dependability of behavioral measurements: Theory of generalizability of scores and profiles เพื่อใช้ในการปรับแก้ปัญหาที่ไม่สามารถระบุแหล่งแห่งความแปรปรวนได้ ทฤษฎีนี้จะใช้วิธีการระบุความแปรปรวนของตัวแปรที่สังเกตได้ให้มากที่สุด แล้วระบุความเที่ยงตามวัตถุประสงค์ของการวัด

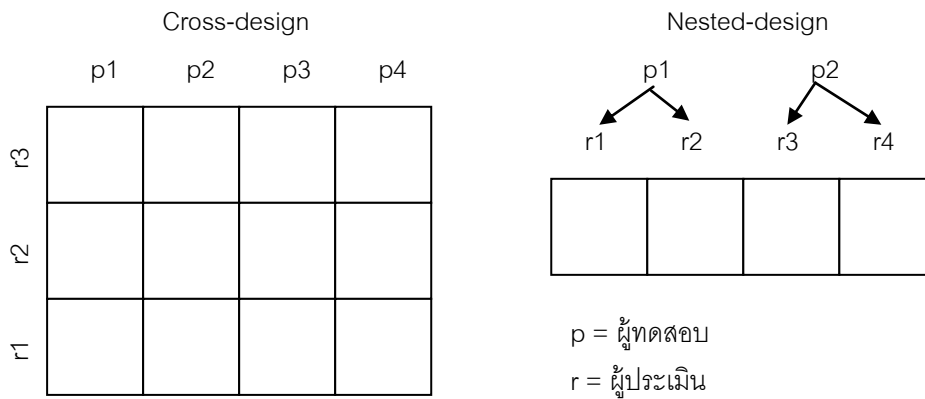
วัตถุประสงค์ของการวัดในที่นี่แบ่งออกเป็น 2 ประเภทคือ

- 1) การตัดสินใจเชิงสัมพัทธ์ (Relative decision) เป็นการตัดสินใจด้วยตำแหน่งของแต่ละบุคคลเมื่อเปรียบเทียบกับคนอื่นแล้ว (การตัดสินใจกลุ่ม)
- 2) การตัดสินใจเชิงสัมบูรณ์ (Absolute decision) คือ เป็นการตัดสินใจจากคะแนนดิบของแต่ละบุคคล เช่น ถ้าสอบไม่ได้ถึง 50 คะแนนจะได้เกรด F (การตัดสินใจเกณฑ์)

การตัดสินใจที่แตกต่างกัน 2 ประเภทนี้ ทำให้การระบุแหล่งแห่งความผิดพลาดแตกต่างกัน เช่น คะแนนความยากของข้อกระทงที่แตกต่างกัน 2 ชุด หรือคะแนนที่ผู้ประเมินปล่อยคะแนน หรือกดคะแนน ถ้าทำให้ลำดับของบุคคลไม่เปลี่ยนแปลง ทำให้ไม่ถือว่าเป็นแหล่งแห่งความผิดพลาดในการตัดสินใจเชิงสัมพัทธ์ แต่เป็นข้อผิดพลาดในการตัดสินใจเชิงสัมบูรณ์

ทฤษฎีการอ้างอิงโดยสรุปมีความสามารถในการทดสอบความแปรปรวนกับข้อมูลได้หลายรูปแบบ ไม่ว่าจะเป็นข้อมูลแบบข้าม (Cross-design) เช่น ผู้ทดสอบ (Persons) ต้องตอบคำตอบทุกข้อ (Items) 2 ครั้ง

(Occasion) ข้อมูลนี้จะเป็นกรณี ผู้ทดสอบข้ามกับข้อกระทงและข้ามกับครั้งในการทดสอบ (Person x Item X Occasion: $p \times i \times o$) และจะทดสอบว่าความผิดพลาดที่เกิดจากข้อกระทง หรือเหตุการณ์จะมีความแปรปรวนมากเพียงใดในประชากร การออกแบบอีกรูปแบบหนึ่งคือแบบรังนก (Nested design) เช่น ผู้ทดสอบ (Persons) 2 คน ได้รับการประเมินจากผู้ประเมิน (Raters) คนละ 2 คน โดยผู้ประเมินในผู้ทดสอบทั้ง 2 คน ไม่เหมือนกัน ข้อมูลนี้จะเป็นในรูปแบบรังนกโดยผู้ประเมินซ้อนอยู่ในผู้ทดสอบ (Item nested in person; $r:p$) เปรียบเทียบได้ดังรูป



รูปที่ 1 วิธีการออกแบบข้อมูล เปรียบเทียบระหว่างข้อมูลแบบข้าม และข้อมูลแบบรังนก

นอกจากนี้ยังมีการออกแบบข้อมูลที่ผสมระหว่างข้อมูลแบบข้ามและข้อมูลแบบรังนกด้วย ซึ่งจะกล่าวในภายหลัง การออกแบบข้อมูลที่ง่ายที่สุด คือ ข้อมูลแบบข้ามในกรณีที่ระบุแหล่งแห่งความแปรปรวนเพียงแหล่งเดียว เช่น ผู้ทดสอบทั้งหมดจำนวน 10 คน ได้ทำข้อสอบจำนวน 30 ข้อเหมือนกันทุกคน ด้วยเหตุนี้แหล่งความแปรปรวนที่ต้องการคือ ความแปรปรวนที่ได้รับจากผู้ทดสอบ แหล่งแห่งความคาดเคลื่อนคือข้อกระทง และความคาดเคลื่อนที่เกิดร่วมกันระหว่างผู้ทดสอบและข้อกระทง รวมถึงแหล่งความผิดพลาดอื่นๆ สามารถเขียนสมการได้ดังนี้

$$O_{pi} = G_{00} + u_{p0} + v_{0i} + w_{pi} + e_{pi}$$

แต่เนื่องจากข้อมูลในผู้ประเมินแต่ละคน ในข้อกระทงแต่ละข้อมีเพียงข้อมูลเดียว ทำให้ไม่สามารถหาปฏิสัมพันธ์ได้ แม้ว่าในความเป็นจริงจะมีก็ตาม ทำให้สมการลดเหลือเพียง

$$O_{pi} = G_{00} + u_{p0} + v_{0i} + e_{pi,e}$$

เนื่องจากมีเพียง grand mean ที่ไม่เปลี่ยนแปลง ดังนั้นแหล่งความแปรปรวนของคะแนนผู้ประเมินในแต่ละข้อกระทงเกิดจากความแปรปรวนที่เกิดจากบุคคล ที่เกิดจากข้อกระทง และที่เกิดจากปฏิสัมพันธ์รวมถึงความผิดพลาดอื่นๆ ดังสมการ

$$SS_o = SS_p + SS_i + SS_{p,i,e}$$

จะเห็นว่าสมการเหมือนกับการแบ่งแยกความแปรปรวนในสูตรการหาความเที่ยงของ Hoyt แต่ในทฤษฎีการอ้างอิงโดยสรุปมีวิธีการประมาณค่าความแปรปรวนเฉลี่ย (Mean of square) ทั้งหมด 2 ประเภท คือ วิธีความคาดเคลื่อนกำลังสองต่ำสุด (Ordinary least squares; OLS) ซึ่งเป็นวิธีดั้งเดิมที่ Hoyt ใช้ และวิธีการประมาณค่าเป็นไปได้อย่างสูงสุดแบบจำกัด (Restricted maximum likelihood; RMR) ซึ่งประมาณค่าได้ดีกว่าการ

ประมาณค่าเป็นไปได้อย่างสูงสุด (Maximum likelihood) ปกติในกรณีของกลุ่มตัวอย่างน้อย และสามารถช่วยลดข้อตกลงเบื้องต้นในการวัดของวิธีประมาณค่าความคาดเคลื่อนกำลังสองต่ำสุด คือข้อตกลงเบื้องต้นเรื่องตัวแปรตามต้องมีการกระจายเป็นโค้งปกติ (Normality)

เมื่อได้ค่าความแปรปรวนเฉลี่ยแล้ว ในที่นี้มีทั้งหมด 3 ค่า คือ ความแปรปรวนเฉลี่ยที่เกิดจากผู้ทดสอบ (E(MSp)) ที่เกิดจากข้อกระทง (E(MSi)) และที่เกิดจากผลปฏิสัมพันธ์และข้อผิดพลาด (E(MSpi,e)) แล้ว จะนำค่าเหล่านี้มาหาค่าความแปรปรวนในกรณีแต่ละองค์ประกอบเป็น 1 หน่วยคือถ้าผู้ประเมิน 1 คน ทำข้อกระทง 1 ข้อ จะมีความแปรปรวนเป็นเท่าไร จากแนวคิดเรื่องการวิเคราะห์ความแปรปรวน (Analysis of variance) ค่าความแปรปรวนเฉลี่ยจะประมาณค่าพารามิเตอร์ในประชากรดังนี้

$$E(MS_p) = \sigma_{pi,e}^2 + n_i \sigma_p^2$$

$$E(MS_i) = \sigma_{pi,e}^2 + n_p \sigma_i^2$$

$$E(MS_{pi,e}) = \sigma_{pi,e}^2$$

จากสูตรเหล่านี้ทำให้ประมาณค่าความแปรปรวนในแต่ละองค์ประกอบในประชากรได้ แล้วนำค่าเหล่านี้ไปหาค่าความแปรปรวนที่เกิดจากความผิดพลาดในการวัด แล้วค่อยนำค่าเหล่านี้ไปหาค่าความเที่ยงได้ อย่างไรก็ตาม ความผิดพลาดในการวัดในการประเมินเชิงสัมพัทธ์ และเชิงสัมบูรณ์แตกต่างกัน ค่าความแปรปรวนจากข้อกระทงจะเป็นค่าความผิดพลาดในการวัดแบบสัมบูรณ์ แต่ไม่เป็นความผิดพลาดในการวัดแบบสัมพัทธ์ สามารถหาได้ดังสูตร

$$\sigma_{Rel}^2 = \frac{\sigma_{pi,e}^2}{n_i}$$

$$\sigma_{Abs}^2 = \frac{\sigma_{pi,e}^2}{n_i} + \frac{\sigma_i^2}{n_i}$$

อย่างไรก็ตาม n_i ไม่จำเป็นจะต้องเท่ากับจำนวนข้อกระทงตอนเริ่มต้น อาจคำนวณในกรณีที่มีจำนวนข้อกระทงแตกต่างกันได้ ถ้าเพิ่มข้อกระทงขึ้น จะทำให้ค่าความผิดพลาดน้อยลง แต่ถ้าลดข้อกระทงจะทำให้ค่าผิดพลาดมากขึ้นตามสูตร เมื่อได้ค่าความแปรปรวนที่เกิดจากความผิดพลาดในการวัดแล้ว สามารถนำมาหาค่าความเที่ยงได้ดังนี้

$$Rel r_{xx} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{Rel}^2}$$

$$Abs r_{xx} = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_{Abs}^2}$$

ความเที่ยงเชิงสัมพัทธ์อาจเรียกว่า ค่าสัมประสิทธิ์การอ้างอิง (Generalizability coefficient) และค่าความเที่ยงเชิงสัมบูรณ์อาจเรียกว่า ดัชนีการพึ่งพิง (Index of dependability) สูตรทั้งสองนี้สามารถใช้ได้ในทุกรูปแบบของข้อมูล ถ้ารู้ความผิดพลาดเชิงสัมพัทธ์และเชิงสัมบูรณ์

กรณีต่อไป เป็นข้อมูลแบบข้ามเช่นเดียวกัน แต่ระบุแหล่งแห่งความผิดพลาด 2 แหล่ง ยกตัวอย่างเช่น มีผู้ประเมิน 3 คน ประเมินผู้ปฏิบัติงานจำนวน 4 คน โดยผู้ประเมินทั้ง 3 คน ประเมินผู้ปฏิบัติงานครบทุกคน ใช้ 2 ครั้งในการประเมิน การออกแบบจะเป็นในรูป ผู้ปฏิบัติงาน x ผู้ประเมิน x ครั้งที่ประเมิน ($p \times r \times c$) ค่ะแผนที่

แท้จริงในผู้ปฏิบัติงานแต่ละคน ผู้ประเมินแต่ละคน และครั้งที่ประเมินแต่ละครั้ง สามารถประมาณค่าได้ 2 สมการ คือ

$$X_{prc} = T_{prc} + e_{prc}$$

$$T_{prc} = G_{000} + u_{p00} + v_{0r0} + w_{00c} + k_{pr0} + m_{0rc} + n_{p0c} + q_{prc}$$

แต่ปฏิสัมพันธ์ระหว่างผู้ปฏิบัติงาน ผู้ประเมิน และเวลา ไม่สามารถทดสอบได้ แม้ว่าในความเป็นจริงจะมีอยู่ก็ตาม เพราะว่าผู้ประเมิน ประเมินผู้ปฏิบัติงานในระยะเวลาแต่ละครั้งเพียงครั้งเดียว ทำให้เหลือสมการด้านล่างเพียง

$$T_{prc} = G_{000} + u_{p00} + v_{0r0} + w_{00c} + k_{pr0} + m_{0rc} + n_{p0c}$$

เมื่อรวมสมการทั้งสองแล้ว ทำให้ได้สมการเป็น

$$X_{prc} = G_{000} + u_{p00} + v_{0r0} + w_{00c} + k_{pr0} + m_{0rc} + n_{p0c} + e_{prc}$$

แต่ค่าความผิดพลาดโดยสุ่ม (e) จะรวมปฏิสัมพันธ์ระหว่างผู้ประเมิน ผู้ปฏิบัติงาน และช่วงเวลาเข้าไปด้วย จะเห็นว่า Grandmean เป็นเพียงค่าเดียวที่คงที่ ส่วนอีก 7 ตัวที่เหลือสามารถเบี่ยงเบนไปได้ ทำให้ค่าความแปรปรวนของ X_{prc} สามารถแบ่งได้เป็น 7 ส่วนด้วยกัน คือ

$$SS_x = SS_p + SS_r + SS_c + SS_{pr} + SS_{rc} + SS_{pc} + SS_e$$

แต่อย่างทีกล่าวไปข้างต้น ในทฤษฎีการอ้างอิงโดยสรุป อาจใช้การประมาณค่าแบบอื่นได้นอกเหนือจาก OLS ซึ่งพบว่ามักจะใช้ RMR ทำให้ได้ค่า $E(MS)$ ในแต่ละส่วนมา แล้วนำค่าเหล่านี้มาหาค่าความแปรปรวนในแต่ละส่วน ได้ดังนี้

$$E(MS_p) = \sigma_{prc,e}^2 + n_r \sigma_{pc}^2 + n_c \sigma_{pr}^2 + n_r n_c \sigma_p^2$$

$$E(MS_r) = \sigma_{prc,e}^2 + n_p \sigma_{rc}^2 + n_c \sigma_{pr}^2 + n_p n_c \sigma_r^2$$

$$E(MS_c) = \sigma_{prc,e}^2 + n_p \sigma_{rc}^2 + n_r \sigma_{pc}^2 + n_p n_r \sigma_c^2$$

$$E(MS_{pr}) = \sigma_{prc,e}^2 + n_c \sigma_{pr}^2$$

$$E(MS_{pc}) = \sigma_{prc,e}^2 + n_r \sigma_{pc}^2$$

$$E(MS_{rc}) = \sigma_{prc,e}^2 + n_p \sigma_{rc}^2$$

$$E(MS_{prc,e}) = \sigma_{prc,e}^2$$

เมื่อได้ค่าความแปรปรวนในแต่ละส่วน จะทำให้สามารถหาความแปรปรวนที่เกิดจากความผิดพลาดในการวัด กรณีการวัดเชิงสัมพัทธ์และการวัดเชิงสัมบูรณ์ได้ดังนี้

$$\sigma_{Rel}^2 = \frac{\sigma_{prc,e}^2}{n_c n_r} + \frac{\sigma_{pr}^2}{n_r} + \frac{\sigma_{pc}^2}{n_c}$$

$$\sigma_{Abs}^2 = \frac{\sigma_{prc,e}^2}{n_c n_r} + \frac{\sigma_{pr}^2}{n_r} + \frac{\sigma_{pc}^2}{n_c} + \frac{\sigma_{rc}^2}{n_r n_c} + \frac{\sigma_r^2}{n_r} + \frac{\sigma_c^2}{n_c}$$

เมื่อหาความแปรปรวนที่เกิดจากความผิดพลาดในการวัดได้แล้ว สามารถหาความเที่ยงเชิงสัมพัทธ์และความเที่ยงเชิงสัมบูรณ์ได้ ตามสูตรที่ได้กล่าวไว้แล้ว

รูปแบบข้อมูลที่จะกล่าวถึงต่อไป คือ ข้อมูลแบบเรียงน ในกรณีที่มีแหล่งความผิดพลาดแหล่งเดียว ตัวอย่าง เช่น มีผู้ประเมิน 6 คน ผู้ประเมิน 3 คนแรก ประเมินผู้ปฏิบัติงานคนแรก และผู้ประเมินอีก 3 คน ประเมินผู้ปฏิบัติงานคนที่สอง จะเห็นว่าผู้ประเมินทุกคนจะต้องประเมินผู้ปฏิบัติงานเพียงคนเดียว และผู้ปฏิบัติงานจะ

ได้รับการประเมินจากผู้ประเมินตั้งแต่ 2 คนขึ้นไป การออกแบบการวัดนี้ จะเรียกว่าการออกแบบระดับชั้น โดยผู้ประเมินซ่อนอยู่ในผู้ปฏิบัติงาน (r:p) คะแนนจริงของผู้ประเมินคนหนึ่งในการประเมินผู้ปฏิบัติงานคนหนึ่ง จะเขียนได้สมการดังนี้

$$\begin{aligned} X_{pr} &= T_{pr} + e_{pr} \\ T_{pr} &= S_{0p} + u_{pr} \\ S_{0p} &= G_{00} + v_{0p} \end{aligned}$$

แต่เนื่องจากข้อมูลการประเมินของผู้ประเมินคนหนึ่ง มีเพียงค่าเดียว ทำให้ไม่สามารถแบ่งแยกค่าความผิดพลาดที่เกิดโดยผู้ประเมิน และค่าความผิดพลาดที่เกิดจากผู้ประเมินแต่ละคนได้ ทำให้รวม e_{pr} และ u_{pr} เข้าด้วยกันเป็น e_{pr} เพียงตัวเดียว แล้วยุบสมการเหลือเพียง 2 ระดับชั้น คือ เป็น

$$\begin{aligned} X_{pr} &= S_{0p} + e_{pr} \\ S_{0p} &= G_{00} + v_{0p} \end{aligned}$$

รวมสมการได้เป็น

$$X_{pr} = G_{00} + v_{0p} + e_{pr}$$

จะเห็นว่าค่า Grandmean เป็นค่าเดียวที่คงที่ จะเห็นว่า X_{pr} จะเปลี่ยนแปลงได้ด้วยค่า v_{0p} และ e_{pr} แสดงให้เห็นว่าค่าความแปรปรวนของ X_{pr} เกิดจากความแปรปรวน 2 ค่าคือความแปรปรวนจากผู้ปฏิบัติงาน และความแปรปรวนจากผู้ประเมิน ปฏิสัมพันธ์ระหว่างผู้ประเมินและผู้ปฏิบัติงาน และค่าผิดพลาดที่เหลือรวมกัน เขียนได้ดังสมการ

$$SS_x = SS_p + SS_{r,pr,e}$$

ถ้าใช้การประมาณค่าด้วย OLS หรือ RMR จะได้ค่า $E(MS)$ มา สามารถหาค่าความแปรปรวนของแต่ละส่วนได้โดย

$$E(MS_p) = \sigma_{r,pr,e}^2 + n_r \sigma_p^2$$

$$E(MS_{r,pr,e}) = \sigma_{r,pr,e}^2$$

เมื่อได้ค่าความแปรปรวนแล้ว สามารถหาค่าความคาดเคลื่อนทั้งแบบสัมบูรณ์และสัมพัทธ์ได้ แต่ว่าค่าความคาดเคลื่อนทั้งสองไม่แตกต่างกัน เพราะไม่สามารถแบ่ง σ_r^2 ออกมาได้ ไม่เหมือนวิธีการแบบข้ามที่ สามารถหาข้อมูลส่วน σ_r^2 ออกมาได้ ทำให้วิธีการแบบข้ามมีประโยชน์มากกว่าวิธีการแบบระดับชั้นในการอธิบายความแปรปรวน วิธีการหาความคาดเคลื่อนทั้งแบบสัมพัทธ์และแบบสัมบูรณ์หาได้โดย

$$\sigma_{Rel}^2 \text{ or } \sigma_{Abs}^2 = \frac{\sigma_{r,pr,e}^2}{n_r}$$

จากนั้นนำค่าความแปรปรวนที่เกิดจากความคาดเคลื่อนในการวัดไปหาค่าความเที่ยงได้ ค่าความเที่ยงเชิงสัมพัทธ์และสัมบูรณ์จะเท่ากัน

รูปแบบข้อมูลที่จะอธิบายเป็นรูปแบบสุดท้าย คือ รูปแบบการวัดแบบผสมระหว่างแบบข้ามและแบบเรียง ตัวอย่างเช่น ผู้ประเมินจำนวน 15 คน ประเมินผู้ปฏิบัติงานจำนวน 5 คน โดยผู้ประเมิน 3 คน ประเมินผู้ปฏิบัติงาน 1 คน โดยไม่มีผู้ประเมินคนไหนประเมินผู้ปฏิบัติงานมากกว่า 1 คน การประเมินมีทั้งหมด 2 ครั้ง จะเป็นการออกแบบในรูปแบบ ผู้ประเมินซ่อนในผู้ปฏิบัติงาน x เวลาในการประเมิน (r:p x c) สามารถเขียนสมการได้ดังนี้

$$\begin{aligned}
 X_{prc} &= T_{prc} + e_{prc} \\
 T_{prc} &= S_{p0c} + u_{p0c} + v_{pr0} + w_{prc} \\
 S_{p0c} &= G_{000} + z_{p00} + t_{00c}
 \end{aligned}$$

แต่เนื่องจากข้อมูลการประเมินของผู้ประเมินคนหนึ่ง ในเวลาหนึ่ง มีเพียงค่าเดียว ทำให้ไม่สามารถแบ่งแยกค่าความผิดพลาดที่เกิดโดยสุ่ม และความผิดพลาดที่เกิดจากผู้ประเมินแต่ละคนในเวลาหนึ่งได้ ทำให้รวม cr และ eprc เข้าด้วยกันเป็น eprc เพียงตัวเดียว แล้วยุบสมการเหลือเพียง 2 ระดับขึ้น คือ

$$\begin{aligned}
 X_{prc} &= S_{p0c} + u_{p0c} + v_{pr0} + e_{prc} \\
 S_{p0c} &= G_{000} + z_{p00} + t_{00c}
 \end{aligned}$$

สามารถรวมสมการได้เป็น

$$X_{prc} = G_{000} + z_{p00} + t_{00c} + u_{p0c} + v_{pr0} + e_{prc}$$

จากสมการพบว่ามีค่า Grandmean เพียงตัวเดียวที่คงที่ ดังนั้นค่าคะแนนที่ได้ เกิดจากการเปลี่ยนแปลงใน 5 องค์ประกอบ 1) ผลงานจริงของผู้ปฏิบัติงาน 2) การเปลี่ยนแปลงในเวลา 3) ผลปฏิสัมพันธ์ระหว่างค่าจริงของแต่ละบุคคลและเวลา 4) การเปลี่ยนแปลงในตัวผู้ประเมินและผลปฏิสัมพันธ์ระหว่างผู้ประเมินและผู้ปฏิบัติงาน และ 5) ผลปฏิสัมพันธ์ระหว่าง rc, prc และผลจากค่าความผิดพลาดอื่นๆ ทำให้ค่าความแปรปรวนสามารถแบ่งได้เป็น 5 ส่วนด้วยกัน ดังสมการ

$$SS_x = SS_p + SS_c + SS_{pc} + SS_{r,pr} + SS_{rc,prc,e}$$

เมื่อใช้การประมาณค่าแบบ CLS หรือ RMR จะได้ค่า E(MS) ทั้ง 5 ค่า ทำให้สามารถหาความแปรปรวนในแต่ละส่วนได้ ดังนี้

$$\begin{aligned}
 E(MS_p) &= \sigma_{rc,prc,e}^2 + n_c \sigma_{r,pr}^2 + n_r n_c \sigma_p^2 \\
 E(MS_c) &= \sigma_{rc,prc,e}^2 + n_r \sigma_{pc}^2 + n_p n_r \sigma_c^2 \\
 E(MS_{pc}) &= \sigma_{rc,prc,e}^2 + n_r \sigma_{pc}^2 \\
 E(MS_{r,pr}) &= \sigma_{rc,prc,e}^2 + n_c \sigma_{r,pr}^2 \\
 E(MS_{rc,prc,e}) &= \sigma_{rc,prc,e}^2
 \end{aligned}$$

เมื่อหาค่าความแปรปรวนในแต่ละส่วนได้แล้ว ทำให้หาค่าความคาดเคลื่อนเชิงสัมพัทธ์และความเที่ยงเชิงสัมบูรณ์ได้ดังนี้

$$\begin{aligned}
 \sigma_{Rel}^2 &= \frac{\sigma_{rc,prc,e}^2}{n_r n_c} + \frac{\sigma_{r,pr}^2}{n_r} + \frac{\sigma_{pc}^2}{n_c} \\
 \sigma_{Abs}^2 &= \frac{\sigma_{rc,prc,e}^2}{n_r n_c} + \frac{\sigma_{r,pr}^2}{n_r} + \frac{\sigma_{pc}^2}{n_c} + \frac{\sigma_c^2}{n_c}
 \end{aligned}$$

เมื่อได้ค่าเหล่านี้แล้ว ก็สามารถหาค่าความเที่ยงทั้งแบบสัมพัทธ์และแบบสัมบูรณ์ได้

จากตัวอย่างข้างต้น จะเห็นถึงประโยชน์ของการนำทฤษฎีการอ้างอิงโดยสรุปมาประยุกต์ใช้ ทำให้สามารถแบ่งความแปรปรวนออกเป็นความแปรปรวนของคะแนนจริง และความแปรปรวนจากค่าความผิดพลาด ทั้งที่แบบกำหนดได้ และกำหนดไม่ได้ (Residual) จะพบว่าวิธีแบบบังนิก จะมีประโยชน์ในการอธิบายความแปรปรวนน้อยกว่าการใช้วิธีแบบข้าม นอกจากนี้ทฤษฎีนี้ยังสามารถยังไปใช้ในการประยุกต์เปรียบเทียบจำนวนของแหล่งแห่งความผิดพลาด (เช่น จำนวนข้อกระทง จำนวนครั้งที่ประเมิน จำนวนผู้ประเมิน) โดยการ

ปรับค่าจำนวนของแหล่งแห่งความผิดพลาดในสูตรของความแปรปรวนผิดพลาดในการวัดทั้งในรูปแบบสัมพัทธ์ และสัมบูรณ์ และยังเปรียบเทียบระหว่างโมเดล เช่น เปรียบเทียบโมเดลที่ใช้วิธีการแบบข้าม และโมเดลที่ใช้วิธีการแบบลำดับขั้น ว่าวิธีไหนทำให้ความเที่ยงออกมามากที่สุดได้ รายละเอียดสามารถหาได้ใน Shavelson & Webb (1991)

การทดสอบความเที่ยงในการป้อนกลับแบบ 360 องศา

งานวิจัยของ Greguras & Robie (1998) ได้ใช้ทฤษฎีการอ้างอิงโดยสรุป มาใช้ในการทดสอบความแปรปรวน โมเดลในการทดสอบความเที่ยงในการประเมินผู้จัดการจำนวน 153 คน ซึ่งทั้งหมดนี้ต้องมีผู้ประเมินระดับสูงกว่าตั้งแต่ 2 คนขึ้นไปเท่านั้น แล้วออกแบบการตรวจสอบค่าความแปรปรวน โดยให้ผู้ประเมินซ่อนอยู่ในผู้ได้รับการประเมินหรือผู้จัดการ และผู้ประเมินทุกคนจะประเมินผู้จัดการด้วยแบบวัดที่มีจำนวนองค์ประกอบ (Scale) เท่ากันคือ 15 องค์ประกอบ ทำให้การออกแบบอยู่ในรูป $rater:person \times scale (r:p \times s)$ โดยวิเคราะห์แยกกัน 3 ครั้ง คือ วิเคราะห์เฉพาะผู้ประเมินอยู่ในระดับสูงกว่า วิเคราะห์เฉพาะผู้ประเมินอยู่ในระดับเดียวกัน และวิเคราะห์เฉพาะผู้ประเมินอยู่ในระดับต่ำกว่า

ผลการวิจัยพบว่า ผลจากองค์ประกอบมีน้อยที่สุด รองลงมาคือผลร่วมกันระหว่างผู้ถูกประเมินและองค์ประกอบ ความแปรปรวนทั้งสองส่วนถือว่าน้อยมาก เมื่อเทียบกับความแปรปรวนอีก 3 ส่วนที่เหลือ คือ ผลความแปรปรวนจากคะแนนจริงของผู้ถูกประเมิน ผลความแปรปรวนจากผู้ประเมินและผลร่วมกันระหว่างผู้ประเมินและผู้ถูกประเมินรวมกัน และส่วนที่สูงที่สุดคือความแปรปรวนระหว่างผู้ประเมินกับข้อกระทง ผู้ประเมินร่วมกับข้อกระทงร่วมกับผู้ถูกประเมิน และความแปรปรวนจากความผิดพลาดอื่นๆ รวมกัน ซึ่งส่วนสุดท้ายยังถือว่าสูงอยู่มาก (ประมาณ 40 %)

การเพิ่มจำนวนองค์ประกอบไม่ได้ช่วยเพิ่มความเที่ยงในการทดสอบมากนัก แต่ส่วนที่เพิ่มมากคือการเพิ่มจำนวนผู้ประเมิน พบว่าในสถานการณ์ที่มีผู้ประเมินไม่จำกัด ผู้ประเมินเป็นผู้ที่อยู่ระดับสูงกว่าจะประเมินได้เที่ยงกว่า คือ แค่ 4 คนก็ถึงระดับค่าความเที่ยงเท่ากับ .70 แต่ระดับเดียวกันต้องใช้ 6 คน และระดับต่ำกว่าต้องใช้ถึง 7 คนด้วยกัน แต่เมื่อเปรียบเทียบกับจำนวนในความเป็นจริงแล้ว (ในงานวิจัยใช้ค่าเฉลี่ยของจำนวนผู้ประเมินในแต่ละแหล่ง) พบว่าผู้ประเมินอยู่ในระดับเดียวกันจำนวน 4 คนดีที่สุด

ข้อด้อยของงานวิจัยนี้ คือ การคัดเลือกกลุ่มตัวอย่าง ให้มีผู้ประเมินเป็นระดับที่สูงกว่าจำนวน 2 คนขึ้นไป เท่านั้น ในสภาพความเป็นจริงน้อยคนที่จะมีผู้ประเมินเป็นผู้ที่อยู่ในระดับสูงกว่าจำนวน 2 คน หัวหน้าของตนเองมักจะมีเพียงคนเดียว ที่ใกล้ชิด อาจทำให้มีปัญหาในการคัดเลือกกลุ่มตัวอย่าง และการนำกลุ่มตัวอย่างที่มีผู้ประเมินระดับสูงกว่า 2 คน มาเผยแพร่ไปยังกลุ่มตัวอย่างที่มีผู้ประเมินระดับสูงกว่าเพียงคนเดียวอาจไม่ถูกต้องนัก นอกจากนี้การไม่ได้ทดสอบความแตกต่างระหว่างผู้ประเมินระดับสูงกว่า ต่ำกว่า หรือระดับเดียวกันว่าแตกต่างกันอย่างมีนัยสำคัญหรือไม่ และมีผลปฏิสัมพันธ์กับตัวแปรอื่นหรือไม่ น่าจะเป็นข้อจำกัดของทฤษฎีการอ้างอิงโดยสรุป ทำให้ต่อไปจะกล่าวถึงโมเดลเชิงเส้นแบบระดับขั้น และการปรับปรุงโมเดลเชิงเส้นระดับขั้นให้สามารถตรวจสอบสิ่งที่กล่าวมาข้างต้นได้

โมเดลเชิงเส้นระดับขั้น

โมเดลเชิงเส้นระดับชั้น (Hierarchical linear model) เป็นโมเดลความสามารถในการวิเคราะห์สูงขึ้นไปกว่าโมเดลเชิงเส้นโดยทั่วไป (General linear model) คือสามารถวิเคราะห์ในข้อมูลแบบลำดับชั้นได้ และสามารถใส่ตัวแปรในการวิเคราะห์ในแต่ละลำดับชั้นได้ด้วย ซึ่งคุณสมบัติด้านที่สุด ถือว่าเป็นประโยชน์มาก เช่น ในการทำนายคะแนนสอบเอนทรานซ์ของนักเรียนชั้น ม.6 มีตัวแปรระดับบุคคล เช่น ความฉลาดทางปัญญา สถานะทางเศรษฐกิจสังคม เป็นต้น และตัวแปรระดับโรงเรียน เช่น ขนาดโรงเรียน อัตราส่วนระหว่างครูต่อนักเรียน เป็นต้น โดยเฉพาะอย่างยิ่ง ตัวแปรระดับโรงเรียนจะช่วยให้การอธิบายความแตกต่างระหว่างโรงเรียน ว่าความแตกต่างดังกล่าวสามารถอธิบายได้ด้วยตัวแปรอะไรบ้าง

นอกจากนี้ โมเดลเชิงเส้นระดับชั้น จะสามารถดูผลการเป็นตัวแปรกำกับของความแตกต่างระหว่างโรงเรียน และตัวแปรระดับโรงเรียนได้ด้วย เช่น สถานะทางเศรษฐกิจสังคมส่งผลต่อผลการสอบแตกต่างกันหรือไม่ในระหว่างโรงเรียน (ค่าสัมประสิทธิ์ถดถอยแตกต่างกันหรือไม่) และตัวแปรในระหว่างโรงเรียนอย่างเช่น ขนาดของโรงเรียน ส่งผลเป็นตัวแปรกำกับในความสัมพันธ์ระหว่างสถานะทางเศรษฐกิจสังคม และผลการสอบหรือไม่

ระดับของตัวแปรจะแบ่งเป็น 2 ระดับด้วยกัน คือ ระดับจุลภาค (Micro level) และระดับมหภาค (Macro level) สมาชิกของในระดับจุลภาค จะต้องอยู่ระดับมหภาคเพียงประเภทเดียวเท่านั้น เช่น นักเรียน (ระดับจุลภาค) คนหนึ่งจะต้องอยู่ในโรงเรียน (ระดับมหภาค) เพียงโรงเรียนเดียวเท่านั้น มิฉะนั้นจะใช้วิธีการวิเคราะห์แบบระดับชั้นเชิงเส้นไม่ได้

เริ่มต้นที่โมเดลอย่างง่าย คือ โมเดลผู้ประเมินซ้อนอยู่ในผู้ถูกประเมิน จะสามารถเขียนสมการได้ดังนี้

$$\text{ระดับผู้ประเมิน (Micro level)} \quad Y_{ij} = B_{0j} + r_{ij}$$

$$\text{ระดับผู้ถูกประเมิน (Macro level)} \quad B_{0j} = G_{00} + u_{0j}$$

$$\text{รวมสมการ} \quad Y_{ij} = G_{00} + u_{0j} + r_{ij}$$

จะเห็นว่าสมการคล้ายกับโมเดลที่นำเสนอไว้ในทฤษฎีการอ้างอิงโดยสรุป โปรแกรมวิเคราะห์โมเดลเชิงเส้นแบบระดับชั้นจะใช้การประมาณค่าแบบ Bayesian estimation ซึ่ง Raudenbush & Bryk (2002) ได้สรุปไว้ในบทที่ 4 ว่าการประมาณค่าแบบนี้ดีกว่า RMR และ OLS ในกรณีที่กลุ่มตัวอย่างมีจำนวนน้อย ทำให้ทราบค่าความแปรปรวนที่อธิบายได้ด้วย u_{0j} และ r_{ij}

โปรแกรม HLM จะสามารถประเมินค่าความเชื่อถือในการประเมินค่าพารามิเตอร์ที่ประมาณได้ในระดับจุลภาค เช่น ประเมินค่าเฉลี่ยของผู้ประเมินทุกคนในการประเมินผู้ถูกประเมิน ว่าน่าเชื่อถือได้มากน้อยเพียงใด เริ่มต้นด้วยการนำค่าเฉลี่ยที่เกิดจากผู้ประเมิน (r_{ij}) มาหารจำนวนผู้ประเมินในผู้ถูกประเมินแต่ละคน (n_j) ทำให้ได้ค่าความแปรปรวนที่เกิดจากความผิดพลาดของผู้ถูกประเมินแต่ละคน (V_j) และค่าความเชื่อมั่นหาได้โดยสูตร

$$\text{ความเชื่อมั่น} (\lambda_j) = \frac{\tau_{00}}{\tau_{00} + V_j}$$

ถ้าค่าความเชื่อมั่นต่ำ จะทำให้ค่าเฉลี่ยของผู้ประเมินในผู้ถูกประเมินแต่ละคนประมาณค่าผลงานที่แท้จริงโดยทั่วไปไม่ถูกต้องนัก ในหนังสือของ Raudenbush & Bryk (2002) แนะนำให้ใช้ค่าผลงานเฉลี่ยทุกคนในการประเมิน รวมถึงถ้ามีตัวแปรในระดับมหภาคด้วย จะใช้ตัวแปรเหล่านี้ช่วยทำนายด้วย

ถ้านำค่าความเชื่อมั่นในผู้ถูกประเมินแต่ละคนมาหาค่าเฉลี่ยทุกคนแล้ว จะทำให้ได้ค่าความเที่ยง ดังสมการ

$$\bar{\lambda} = \frac{\sum \lambda_j}{n_j} = \frac{\sum \left(\frac{\tau_{00}}{\tau_{00} + V_j} \right)}{n_j} = \frac{\sum_{j=1}^p \left(\frac{\tau_{00}}{\tau_{00} + \left(\sum_{i=1}^r r_{ij} / n_i \right)} \right)}{n_j}$$

เปรียบเทียบใน G theory

$$r_{xx} = \frac{\sigma_p^2}{\left(\sigma_p^2 + \frac{\sigma_{r,pr,e}^2}{n_r} \right)}$$

จะเห็นว่าสูตรทั้งสองเหมือนกันเกือบทั้งหมด เพียงแต่ว่าความเที่ยงใน HLM จะมีรายละเอียดมากกว่า เพราะสามารถหาความเที่ยงในแต่ละผู้ถูกประเมินได้

ต่อไปจะใส่ตัวแปรในการทำนาย เริ่มต้นคือการใส่ตัวแปรระดับจุดภาค คือ สถานะของผู้ประเมิน (สูงกว่า, พอๆ กัน และต่ำกว่า) นำสถานะนี้มาแปลงค่าเป็นตัวแปรดัมมี่ 2 ตัว คือการอยู่ในระดับสูงกว่า (1) หรือต่ำกว่า (0) และการอยู่ในระดับพอๆ กัน (1) หรือต่ำกว่า (0) ทำให้เขียนสมการได้เป็น

ระดับผู้ประเมิน $Y_{ij} = B_{0j} + B_{1j}X_{1ij} + B_{2j}X_{2ij} + r_{ij}$

ระดับผู้ถูกประเมิน $B_{0j} = G_{00} + u_{0j}$

$B_{1j} = G_{10} + u_{1j}$

$B_{2j} = G_{20} + u_{2j}$

จะเห็นว่า ค่า B_{0j} หรือค่าเฉลี่ยจากการประเมินผู้ปฏิบัติงานแต่ละคน เปลี่ยนแปลงไปได้ตามค่า X_{1ij} และ X_{2ij} อย่างไรก็ตาม เมื่อดูระดับมหภาคแล้ว จะพบว่า การเปลี่ยนแปลง X_{1ij} และ X_{2ij} ต่อ Y_{ij} ในผู้ถูกประเมินแต่ละคนไม่เท่ากัน เนื่องจากค่า u_{1j} และ u_{2j} แตกต่างกันในแต่ละผู้ถูกประเมิน เมื่อรวมสมการทั้ง 4 สมการเป็นดังนี้

$$Y_{ij} = G_{00} + (G_{10} + u_{1j})X_{1ij} + (G_{20} + u_{2j})X_{2ij} + u_{0j} + r_{ij}$$

จากสมการจะเห็นประโยชน์ของการใช้ HLM ในการวิเคราะห์ที่เพิ่มเติมมากกว่า G theory และยังได้ Between-rater reliability ผิดกับงานวิจัยของ Greguras & Robie (1998) ที่ตรวจสอบได้เฉพาะ within-rater reliability ที่จำเป็นต้องหา Between-rater reliability เพราะว่าเป็นสภาพความเป็นจริงในการให้ข้อมูลป้อนกลับแบบ 360 องศา

ในโมเดลแบบข้าม อาจจะไม่จำเป็นในการวิเคราะห์ด้วย HLM แต่อาจนำความรู้การใส่ตัวแปร HLM เช่น การประเมิน ratee จำนวน 20 คน ด้วยผู้ประเมิน 20 คน โดยผู้ประเมินทุกคนจะต้องประเมิน ratee ทุกคน ทำให้การวิจัยเป็นรูปแบบ p x r เขียนสมการได้ดังนี้

$$X_{pr} = T_{pr} + r_{pr}$$

$$T_{pr} = G_{00} + u_{p0} + v_{0r} + z_{pr}$$

แต่เนื่องจากไม่สามารถแยกแยะระหว่าง z_{pr} กับ r_{pr} ได้ จึงขอรวมว่าเป็น e_{pr} แล้วทำให้รวมสมการได้ดังนี้

$$T_{pr} = G_{00} + P_{p0} + R_{0r} + e_{pr}$$

อาจใส่สมการได้ดังนี้

$$X_{pr} = G_{00} + P_{0p0} + (G_{10} + P_{10r})W_{1p0} + (G_{20} + P_{20r})W_{2p0} + R_{0r} + e_{pr}$$

G_{00} , G_{10} และ G_{20} เป็นผลคงที่ P_{0p0} เป็นค่าที่เปลี่ยนแปลงตาม p และ R_{0r} เป็นส่วนที่เปลี่ยนแปลงตามค่า r ส่วน P_{10r} และ P_{20r} เป็น slope ที่เปลี่ยนแปลงไปตามค่าของ R_{0r} ส่วน W_{1p0} และ W_{2p0} เป็นตัวแปรต้น และสุดท้ายเป็น e_{pr} ซึ่งเป็นผลความผิดพลาดปฏิสัมพันธ์ระหว่าง z และความผิดพลาดโดยสุ่ม

จากจุดนี้ ถ้าทดสอบค่าด้วย HLM สมการ

$$X_{pri} = T_{pr0} + r_{pri}$$

$$T_{pr0} = G_{000} + P_{0p00} + (G_{100} + P_{10r0})W_{1p00} + (G_{200} + P_{20r0})W_{2p00} + R_{0r0}$$

แล้วทดสอบด้วย Cross-level effect ใน HLM ได้ ค่าความเชื่อมั่นของ rater แต่ละคนในโมเดลเปล่าไม่มีตัวแปรต้นสามารถหาได้โดย

$$r_{xx} = \frac{\sigma^2(P_{p00})}{\sigma^2(P_{p00}) + V_{jk}}$$

เนื่องจาก V_{jk} เท่ากัน และแต่ละ p มีจำนวน r เท่ากัน ทำให้ค่าความเที่ยงของทุกๆ p เท่ากัน ถ้าในกรณีที่มีตัวแปรต้น มีแนวโน้มที่จะทำให้ $\sigma^2(P_{p00})$ น้อยลง ทำให้ r_{xx} น้อยลง แสดงว่าเป็นไปได้ที่ r และ G_{000} รวมถึง W_{1p00} และ W_{2p00} ร่วมกันอธิบายได้ดีมากขึ้น

ส่วนโมเดลที่ซ้อนกันระหว่างแบบลำดับชั้น และแบบข้าม จะกล่าวถึงในหัวข้อ “เหนือกว่าโปรแกรม HLM”

ข้อดีของโปรแกรม HLM

1. การหาค่า reliability ของค่า T_{p00} และค่า T_{p00} รวม
2. การใช้ Bayesian estimation จะทำให้ประมาณค่าได้ถูกต้องมากกว่า RMR และ OLS
3. ถึงแม้ว่า $\sigma^2_{r,pr}$ จะ contaminated กัน แต่ยังพอเอาตัวแปรต้นมาอธิบายได้ เช่น ถ้าตัวแปรต้นระดับจุลภาค สามารถทำนายได้ แต่พบว่าค่า slope ไม่แตกต่างกันในระดับ macro ทำให้น่าจะเป็นผลจากระดับ rater อย่างเดียว แต่ถ้าค่า slope แตกต่างกัน แสดงว่าเกิด interaction แต่ยังไม่สามารถแบ่งค่าความแปรปรวนได้อย่างชัดเจนนัก

เหนือกว่าโปรแกรม HLM

สุดท้ายนำ Model มาอธิบายใน mixed model (r:p x c) จากตัวอย่างเดิมในทฤษฎีการอ้างอิงเชิงสรุป สามารถเขียนสมการได้ 3 ระดับดังนี้

$$X_{prc} = T_{prc} + e_{prc}$$

$$T_{prc} = S_{p0c} + u_{p0c} + v_{pr0} + w_{prc}$$

$$S_{p0c} = G_{000} + Z_{p00} + t_{00c}$$

แต่เนื่องจากข้อมูลการประเมินของผู้ประเมินคนหนึ่ง ในเวลาหนึ่ง มีเพียงค่าเดียว ทำให้ไม่สามารถแบ่งแยกค่าความผิดพลาดที่เกิดโดยสุ่ม และความผิดพลาดที่เกิดจากผู้ประเมินแต่ละคนในเวลาหนึ่งได้ ทำให้รวม or และ epro เข้าด้วยกันเป็น epro เพียงตัวเดียว แล้วยุบสมการเหลือเพียง 2 ระดับชั้น คือ

$$X_{prc} = S_{p0c} + u_{p0c} + v_{pr0} + e_{prc}$$

$$S_{p0c} = G_{000} + Z_{p00} + t_{00c}$$

สามารถรวมสมการได้เป็น

$$X_{prc} = G_{000} + Z_{p00} + t_{00c} + u_{p0c} + v_{pr0} + e_{prc}$$

จะเห็นว่าตัวแปรที่ใส่ไม่ได้เป็น fixed effect แต่เป็น random effect ในระดับ micro level โมเดลแบบนี้ในโปรแกรม HLM ปัจจุบันยังไม่มี ซึ่งจุดนี้ยังอาจเป็นช่องว่างในงานวิจัยที่สามารถไปค้นคว้าและศึกษาเพิ่มเติมได้ การใส่ตัวแปรคล้ายกับการใส่ตัวแปรแบบข้อมูลข้าม

ข้อเสนอแนะในการวิจัยต่อไป

- 1) ตัวแปรตามอาจมีหลายตัว เช่น ผลงานทั่วไป (Task performance) และผลงานเพิ่มเติม (Contextual performance) อาจใช้แนวคิดทางด้าน Multivariate Hierarchical Linear Model ในการวิเคราะห์ความแปรปรวนได้ แต่โมเดลนี้จะมีประโยชน์ก็ต่อเมื่อมีตัวแปรต้นในการวิจัย ถ้าต้องการอธิบายความแปรปรวนของตัวแปรตามหลายตัวนั้น อาจไม่จำเป็น เพราะว่าการอ่านความหมายตัวแปรระดับพหุ (variate) ทำได้ลำบาก
- 2) จากแนวคิดข้างต้น สามารถหาตัวแปรภายในผู้ประเมิน ผู้ถูกประเมิน และผลร่วมกันระหว่างผู้ประเมินและผู้ถูกประเมินได้
- 3) ผลจากการเปลี่ยนแปลงภายในบุคคล อาจสามารถตรวจสอบได้ในโมเดล HLM ทำให้อาจใช้ในการวิจัยบางด้านได้ เช่น การพัฒนาภายในตัวบุคคล ผลของความคุ้นเคยในการใช้การให้ข้อมูลป้อนกลับแบบ 360 องศา ผลจากการเปลี่ยนงาน เป็นต้น
- 4) นอกจากนี้ยังใส่ตัวแปรระดับแผนก ระดับทีม หรือระดับองค์การ เพื่อตรวจหาผลที่กระทบต่อการให้ข้อมูลป้อนกลับแบบ 360 องศา

หนังสืออ้างอิง

- ชุมพร ยงกิตติกุล. (2547). *การวัดและการทดสอบทางจิตวิทยา*. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing (7th ed.)*. Upper Saddle River: Prentice-Hall.
- Greguras, G. L., & Robie, C. (1998). A new look at within-source interrater reliability of 360-degree feedback ratings. *Journal of Applied Psychology, 83*, 960-968.
- Kirk, R. E. (1995). *Experimental design: Procedure for the behavioral sciences (3rd ed.)*. Pacific Grove: Brooks/Cole.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Application and data analysis methods (2nd ed.)*. Thousand Oaks: Sage.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Thousand Oaks: Sage.
- Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist, 44*, 922-932.