

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ความหมาย

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability theory: G theory) หมายถึงการศึกษาวิเคราะห์ความน่าเชื่อถือของผลการวัดด้วยการให้ความสำคัญกับการออกแบบการวัด (measurement design) ที่คำนึงถึงแหล่งความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง ระดับความน่าเชื่อถือของผลการวัดที่ได้มีชื่อเรียกเฉพาะว่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (generalizability coefficient)

ความเป็นมา

ครอนบาค และคณะ (Cronbach et al. 1972; อ้างอิงจาก ศิริชัย กาญจนวาสี. 2555: 12) ได้เสนอทฤษฎีสำหรับการวิเคราะห์ความน่าเชื่อถือของผลการวัดพฤติกรรม (dependability of behavioral measurement) สำหรับสถานการณ์ของการวัดผลลักษณะต่างๆ ทั้งด้านจิตวิทยาและการศึกษาซึ่งต่อมารู้จักกันในชื่อของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability theory) หรือ G theory

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดเป็นทฤษฎีที่ขยายแนวคิดจากทฤษฎีการทดสอบแบบดั้งเดิม (classical test theory) ซึ่งเป็นทฤษฎีที่ให้ความสำคัญกับการตรวจสอบและระบุถึงปริมาณของแหล่งความคลาดเคลื่อนจากกระบวนการวัดที่ส่งผลให้เกิดความไม่คงที่ของคะแนนสังเกตได้ (inconsistencies in observed scores) (Brennan. 2011: 2)

จากโมเดลการวัดตามแนวคิดทฤษฎีการทดสอบแบบดั้งเดิม คือ

$$X = T + E$$

เมื่อ X แทน คะแนนสังเกตได้

T แทน คะแนนจริง

E แทน คะแนนคลาดเคลื่อนแบบสุ่ม

โมเดลการวัดดังกล่าวมีข้อควรพิจารณา 2 ประการ คือ

1. คะแนนจริงและคะแนนคลาดเคลื่อนแบบสุ่มเป็นตัวแปรที่ไม่สามารถสังเกตได้โดยตรง หากต้องการใช้โมเดลนี้จำเป็นต้องมีข้อตกลงเบื้องต้นเพิ่มเติม คือ

1.1 หากระบุค่าคะแนนจริงเสมือนกับเป็นค่าคาดหวังของคะแนนสังเกตได้ จะเป็นเงื่อนไขให้ค่าคาดหวังของคะแนนคลาดเคลื่อนเป็นศูนย์

1.2 หากระบุค่าคาดหวังของคะแนนคลาดเคลื่อนเสมือนกับมีค่าเป็นศูนย์ จะเป็นเงื่อนไขให้คะแนนจริงเป็นค่าคาดหวังของคะแนนสังเกตได้

ข้อสังเกตจากข้อตกลงเบื้องต้นเพิ่มเติม 1.1 และ 1.2 คือ ผลที่เกิดขึ้นจากข้อตกลงเบื้องต้นเพิ่มเติมทั้ง 2 ข้อได้ผลเหมือนกัน คะแนนจริงและคะแนนคลาดเคลื่อนมีความเกี่ยวข้องกัน

2. คะแนนจริงเป็นคะแนนที่ไม่สามารถทราบค่าที่แท้จริงได้

ศิริชัย กาญจนวาสี (2555: 14) ได้อธิบายว่าทฤษฎีการทดสอบแบบดั้งเดิมมุ่งที่จะอธิบายความไม่คงเส้นคงวาของการวัดด้วยการพิจารณาจากแหล่ง

ความคลาดเคลื่อนเพียงแหล่งเดียวสำหรับการทดสอบในแต่ละครั้ง เช่น

1) ความคลาดเคลื่อนมาจากการทิ้งช่วงระยะเวลาของการสอบซ้ำจะประมาณค่าความเชื่อมั่นด้วยวิธีทดสอบซ้ำ (test-retest method)

2) ความคลาดเคลื่อนมาจากชุดของแบบทดสอบคู่ขนานจะประมาณค่าความเชื่อมั่นด้วยวิธีแบบทดสอบสมมูล (equivalence forms method) และ

3) ความคลาดเคลื่อนมาจากความยาวของแบบทดสอบจะประมาณค่าความเชื่อมั่นด้วยวิธีแบ่งครึ่งข้อสอบและปรับเป็นทั้งฉบับด้วยสูตรของสเปียร์แมน (Spearman formula) เป็นต้น ในขณะที่ G theory ได้พิจารณาถึงแหล่งความคลาดเคลื่อนพหุ (multiple sources of error) ที่สามารถวิเคราะห์ความคลาดเคลื่อนจากหลายแหล่งพร้อมกันไปได้โดยใช้เทคนิคการวิเคราะห์ความแปรปรวน (ANOVA)

หลักการของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

สำหรับหลักการของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ประกอบด้วย ความคลาดเคลื่อนมาตรฐานของการวัดตามแนวคิดของลอร์ด แนวคิดสำคัญและข้อตกลงเบื้องต้น และแหล่งความคลาดเคลื่อนที่สามารถเกิดขึ้นในกระบวนการวัด ดังรายละเอียด

ความคลาดเคลื่อนมาตรฐานของการวัดตามแนวคิดของลอร์ด (Lord's Standard Error of Measurement)

การวิเคราะห์ตาม G theory สามารถวิเคราะห์ความคลาดเคลื่อนจากหลายแหล่งพร้อมกันไปได้ นั้น Brennan (2011: 4-5) กล่าวว่าหลักการของการ

วิเคราะห์ของ G theory ประยุกต์ใช้หลักการของการประมาณค่าความคลาดเคลื่อนมาตรฐานของการวัดตามแนวคิดของลอร์ด

ค.ศ. 1955 และ 1957 ลอร์ดได้เสนอแนวคิดในการประมาณค่าความคลาดเคลื่อนมาตรฐานของการวัดด้วยการพิจารณาจากความคลาดเคลื่อนของคะแนนสังเกตเฉลี่ยของผู้สอบโดยที่คะแนนสังเกตนั้นเกิดจากการที่ผู้สอบทำข้อสอบ จำนวน k ข้อที่สุ่มแบบเป็นไปได้อย่างหมดจากเอกภพของข้อสอบ (universe of items)

$$\sigma(E^*) = \sqrt{\frac{\tau_p(1-\tau_p)}{k}}$$

$$\text{หรือ } \hat{\sigma}(E^*) = \sqrt{\frac{\bar{X}_p(1-\bar{X}_p)}{k-1}} \quad \text{นำไปใช้ทางปฏิบัติ}$$

เมื่อ $\hat{\sigma}(E^*)$ แทน ความคลาดเคลื่อนมาตรฐานของการวัดตามแนวคิดของลอร์ด

$\hat{\sigma}(E^*)$ แทน ค่าที่ได้จากการประมาณค่าความคลาดเคลื่อนมาตรฐานของการวัดตามแนวคิดของลอร์ด

τ_p แทน คะแนนจริงของผู้สอบที่ได้จากคะแนนเฉลี่ยของผู้สอบ

k แทน จำนวนข้อสอบที่สุ่มแบบเป็นไปได้อย่างหมดจากเอกภพของข้อสอบ (universe of items)

\bar{X}_p แทน คะแนนสังเกตเฉลี่ยของผู้สอบ

จะเห็นได้ว่าแนวคิดความคลาดเคลื่อนมาตรฐานของการวัดของลอร์ดเป็นที่มาของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด สามารถสรุปที่มาของแนวคิดได้ 2 ประเด็น คือ

1. ลอร์ดใช้กระบวนการสุ่มเพื่อประมาณค่าความแปรปรวนของความคลาดเคลื่อนมากกว่าการใช้ความ

เป็นคู่ขนานของแบบทดสอบตามแนวคิดของสเปียร์แมนดังสูตร

$$\sigma(E) = \sigma(X) \sqrt{1 - \rho^2(X,T)}$$

เมื่อ $\sigma(E)$ แทน ความคลาดเคลื่อนมาตรฐานของการวัด

$\sigma(X)$ แทน ค่าเบี่ยงเบนมาตรฐาน ของคะแนนสังเกตได้

$\rho^2(X,T)$ แทน ความเชื่อมั่น (reliability) หรือร้อยละของความสามารถของการอธิบายหรือทำนายคะแนนจริงจากคะแนนสังเกตได้

2. ลอร์ดใช้การออกแบบการวัดที่คำนึงถึงความแปรปรวนภายในบุคคล

จากสาระสำคัญดังกล่าวแสดงให้เห็นว่าทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดใช้แนวคิดของลอร์ดที่เน้นการประมาณค่าความแปรปรวนของความคลาดเคลื่อน ส่วนทฤษฎีทดสอบแบบดั้งเดิมเน้นที่ความเป็นคู่ขนานของแบบทดสอบ (parallelism assumption) ดังนั้น G theory จึงมีการออกแบบการวัดที่ต่างจากทฤษฎีทดสอบแบบดั้งเดิม กล่าวคือ การออกแบบการวัดตามทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ประกอบด้วย 1) ระบุแหล่งความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง ที่มีเรียกชื่อเฉพาะว่า เจอนไซของการวัด (facet) จำแนกเป็น เจอนไซของการวัด 1 เจอนไซ (one facet design) และมากกว่า 1 เจอนไซ (multifaceted designs) 2) นำเจอนไซการวัดที่ระบุไว้มาออกแบบการวัดให้สอดคล้องกับวัตถุประสงค์ของการศึกษาเชิงสรุปอ้างอิง (gen-

eralizability study) โดยที่การศึกษาเชิงสรุปอ้างอิงเป็นการศึกษาและวิเคราะห์ความแปรปรวนของคะแนนจริงและคะแนนคลาดเคลื่อน จำแนกเป็น แบบไขว้ (crossed design) แบบแฝงเต็มรูป (fully nested design) และแบบแฝงบางส่วน (partially nested design) ดังนั้นผลการวิเคราะห์ที่ได้จากการออกแบบการวัดขั้นต้น คือ ผลการศึกษาเชิงสรุปอ้างอิงซึ่งผู้วิจัยสามารถนำไปใช้สำหรับวางแผนเพื่อตัดสินใจว่าสถานการณ์ของการวัดในลักษณะใดควรนำไปใช้เพื่อให้เกิดความน่าเชื่อถือของผลการวัด ที่เรียกว่า การศึกษาเชิงตัดสินใจ (decision study)

แนวคิดสำคัญและข้อตกลงเบื้องต้น

ศิริชัย กาญจนวาสี (2555: 13-15) กล่าวว่าการศึกษาความน่าเชื่อถือของผลการวัดตามแนว G theory อยู่บนพื้นฐานของข้อตกลงเบื้องต้น ดังนี้

1. คุณลักษณะที่มุ่งวัดของบุคคลไม่ว่าจะเป็นความรู้ ทักษะ หรือคุณลักษณะอื่นๆ ซึ่งเป็นเป้าหมายของการวัด เป็นค่าที่อยู่ในสภาวะคงที่ (steady state)
2. ผู้สอบคนเดียวได้คะแนนต่างกันจากการวัดในแต่ละสถานการณ์หรือเจอนไซของการวัด เนื่องจากความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง โดยองค์ประกอบด้านวุฒิภาวะและการเรียนรู้ระหว่างการวัดไม่เป็นแหล่งความคลาดเคลื่อนของคะแนนที่ได้จากการวัด
3. เมื่อพิจารณาผู้สอบทั้งกลุ่ม ความแปรปรวนของคะแนนสังเกตได้ ประกอบด้วยความแปรปรวนของคะแนนจริง ซึ่งเป็นความแตกต่างที่แท้จริงระหว่างบุคคล ความแปรปรวนของคะแนนความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่งและความแปรปรวนของคะแนนความคลาดเคลื่อนสุ่ม

แหล่งความคลาดเคลื่อนที่สามารถเกิดขึ้นในกระบวนการวัด เช่น

1. การทิ้งช่วงระยะเวลาของการทดสอบซ้ำ ซึ่งความคลาดเคลื่อนจากระยะเวลาอาจส่งผลต่อความไม่คงที่ของผลการวัด
2. จำนวนข้อของแบบทดสอบหรือแบบวัดพฤติกรรม ซึ่งการลดความยาวของเครื่องมือวัดมีผลต่อการลดลงของค่าความเชื่อมั่น
3. จำนวนผู้ตรวจสำหรับการให้คะแนนของแบบทดสอบอัตนัย แบบประเมินคุณลักษณะทางจิตวิทยา หรือการประเมินเพื่อคัดเลือกบุคคลจำเป็นต้องอาศัยผู้ตรวจหรือผู้ประเมินที่มีความเชี่ยวชาญในการประเมินคุณลักษณะนั้นๆ การออกแบบการวัดที่มีจำนวนผู้ตรวจหรือผู้ประเมินที่ไม่เหมาะสมจะทำให้เกิดผลต่อความน่าเชื่อถือของผลการวัด
4. อิทธิพลของอันดับในการจัดเรียงข้อสอบตามค่าความยาก หรือการจัดเรียงอันดับของข้อคำถามตามกระบวนการคิดที่นำมาใช้ในการแก้โจทย์ปัญหา หากการจัดเรียงข้อคำถามที่ไม่เหมาะสมจะทำให้เกิดผลต่อความน่าเชื่อถือของผลการวัด
5. คุณลักษณะของผู้ตรวจให้คะแนน เช่น ความเข้มงวด ความใจดีของผู้ตรวจ หากลดอิทธิพลของผู้ให้คะแนนแบบเข้มงวดหรือใจดีด้วยการกำหนดจำนวนผู้ตรวจให้มีคุณลักษณะที่เป็นกลางหรือระบุจำนวนของผู้ตรวจที่ละคุณลักษณะที่เหมาะสมจะทำให้เกิดความน่าเชื่อถือของผลการวัด

ตามแนวคิดของ G theory คะแนนที่ได้จากการวัดภายใต้สถานการณ์หรือเงื่อนไขของการวัดเดียวด้วยแบบทดสอบชุดเดียว และทำการสอบครั้งเดียว คะแนนที่ได้ไม่สามารถเชื่อถือได้อย่างเต็มที่เพราะคะแนนที่ได้ไม่เป็นตัวแทนที่ดีของคะแนนจริงซึ่งเป็นคะแนนเฉลี่ย

ที่ได้จากการทดสอบในหลายๆสถานการณ์ด้วยแบบทดสอบหลายชุด และทำการสอบหลายครั้ง

จากแนวคิดดังกล่าวสามารถเสนอโมเดลของ G theory (Brennan. 2001: 22; Brennan. 2011: 7) คือ

$$X = \mu_p + E_1 + E_2 + \dots + E_H$$

เมื่อ X แทน คะแนนสังเกตได้

μ_p แทน คะแนนเอกภพของบุคคล (person's universe score)

ตาม G theory ถือว่าเป็นคะแนนจริงคำนวณจากค่าคาดหวังของคะแนนเฉลี่ย

E_H แทน แหล่งความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง

ดังนั้นจากโมเดลของ G theory จึงต้องดำเนินการออกแบบการวัดเพื่อทำการศึกษาและวิเคราะห์ความแปรปรวนของคะแนนจริงและคะแนนคลาดเคลื่อน ซึ่งผลการวิเคราะห์สามารถนำไปใช้เพื่อการวางแผนสำหรับการศึกษาค่าสัมประสิทธิ์การสรุปร่างอิงความน่าเชื่อถือของผลการวัด (generalizability coefficient) ซึ่งคล้ายกับค่าสัมประสิทธิ์ความเชื่อมั่น (reliability coefficient) ของแบบทดสอบภายใต้เงื่อนไขของการวัดที่ผู้วิจัยสนใจ และตัดสินใจเลือกเงื่อนไขของการวัดที่ได้ค่าความเชื่อมั่นตามระดับที่ผู้วิจัยต้องการ

การออกแบบการวัดสำหรับการศึกษาเชิงสรุปอ้างอิง (measurement design for generalizability study)

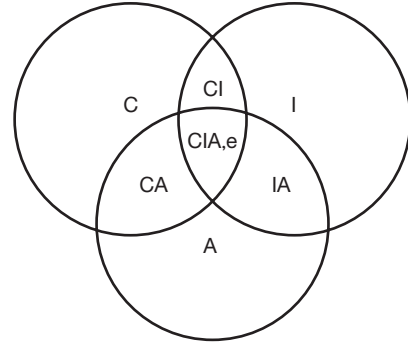
สำหรับการออกแบบการวัดการศึกษาเชิงสรุปอ้างอิง (generalizability study) เป็นการศึกษาและวิเคราะห์ความแปรปรวนของคะแนนจริงและคะแนนคลาดเคลื่อน จำแนกเป็น แบบไขว้ (crossed design) แบบแฝงเต็มรูป (fully nested design) และแบบแฝงบางส่วน (partially nested design) สำหรับเงื่อนไขของการวัดที่ใช้ประกอบการอธิบายการออกแบบการวัดในครั้งนี้ คือ 2 เงื่อนไข ซึ่งเป็นเงื่อนไขของการวัดหรือแหล่งความคลาดเคลื่อนที่เป็นระบบซึ่งเกี่ยวข้องกับกระบวนการประเมินบุคคล

1. การออกแบบการวัดแบบไขว้ (crossed design)

วัตถุประสงค์ของการศึกษาเชิงสรุปอ้างอิง คือ ผู้วิจัยต้องการศึกษาค่าสัมประสิทธิ์สรุปอ้างอิงของผลการประเมินจากกระบวนการประเมินที่ควบคุมแหล่งความคลาดเคลื่อนจากกระบวนการประเมิน 2 แหล่ง หรือ เงื่อนไขของการวัด 2 เงื่อนไข คือ ผู้ประเมินซึ่งเป็นนักจิตวิทยา (I) และคุณลักษณะที่ใช้ประเมิน (A) กล่าวคือ ผลของการให้คะแนนการประเมินของผู้จัดการแต่ละคนโดยทีมนักจิตวิทยาที่ประเมินจากคุณลักษณะ 4 ด้านมีความน่าเชื่อถือเพียงใด

การออกแบบการวัดแบบไขว้ คือ กระบวนการประเมินเพื่อคัดเลือกผู้จัดการครั้งนี้ดำเนินการโดยกำหนดให้ผู้จัดการ (C) ทุกคนได้รับการประเมินจากนักจิตวิทยา (I) ทุกคน ด้วยการประเมินคุณลักษณะ (A) จำนวน 4 ด้าน คือ บุคลิกภาพ ภาวะผู้นำ แรงจูงใจใฝ่สัมฤทธิ์ และการปรับตัวทางสังคม ดังนั้นการออกแบบดังกล่าวเป็นการออกแบบการวัดแบบไขว้ ลักษณะ

ที่ใช้ คือ $C \times I \times A$ สามารถแสดงด้วยแผนภาพเวนน (Venn diagram)

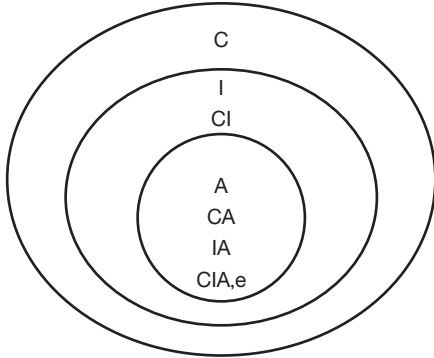


2. การออกแบบการวัดแบบแฝงเต็มรูป (fully nested design)

วัตถุประสงค์ของการศึกษาเชิงสรุปอ้างอิง คือ ผู้วิจัยต้องการศึกษาค่าสัมประสิทธิ์สรุปอ้างอิงของผลการประเมินจากกระบวนการประเมินที่ควบคุมแหล่งความคลาดเคลื่อนจากกระบวนการประเมิน 2 แหล่ง หรือ เงื่อนไขของการวัด 2 เงื่อนไข คือ ชุดของผู้ประเมินซึ่งเป็นนักจิตวิทยา (I) และชุดของรายการของแบบประเมินคุณลักษณะ (A) เพื่อลดอิทธิพลของความเมื่อยล้าของผู้ประเมินและจำนวนของรายการประเมิน กล่าวคือ ผลของการให้คะแนนการประเมินของผู้จัดการแต่ละคนโดยทีมนักจิตวิทยาต่างชุดที่ประเมินจากชุดของรายการประเมินที่ต่างกัน มีความน่าเชื่อถือเพียงใด

การออกแบบการวัดแบบแฝงเต็มรูป คือ กระบวนการประเมินเพื่อคัดเลือกผู้จัดการครั้งนี้ดำเนินการโดยกำหนดให้ผู้จัดการแต่ละคน (C) ได้รับการสัมภาษณ์โดยชุดของนักจิตวิทยา (I) ที่ต่างกัน และนักจิตวิทยาแต่ละคนใช้ชุดของรายการของแบบประเมินคุณลักษณะ (A) ที่ต่างกัน ดังนั้นการออกแบบดังกล่าวเป็นการออกแบบการวัดแบบแฝงเต็มรูป ลักษณะ

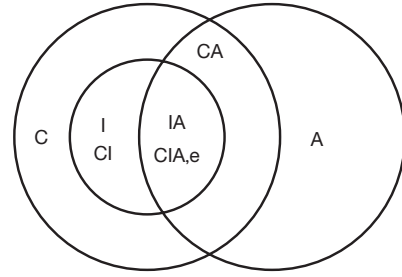
ใช้ คือ (A : I : C) สามารถแสดงด้วยแผนภาพเวนน์ (Venn diagram)



3. การออกแบบการวัดแบบแฝงบางส่วน (partially nested design)

วัตถุประสงค์ของการศึกษาเชิงสรุปอ้างอิง คือ ผู้วิจัยต้องการศึกษาค่าสัมประสิทธิ์สรุปอ้างอิงของผลการประเมินจากกระบวนการประเมินที่ควบคุมแหล่งความคลาดเคลื่อนจากกระบวนการประเมิน 2 แหล่ง หรือ 2 เงื่อนไขการวัด คือ ชุดของผู้ประเมินซึ่งเป็นนักจิตวิทยา (I) และคุณลักษณะที่ใช้ประเมิน (A) เพื่อลดอิทธิพลของความเอนเอียงของผู้ประเมิน กล่าวคือ ผลของการให้คะแนนการประเมินของผู้จัดการแต่ละคน โดยทีมนักจิตวิทยาต่างชุดที่ประเมินจากคุณลักษณะ 4 ด้านมีความน่าเชื่อถือเพียงใด

การออกแบบการวัดแบบแฝงบางส่วน คือ กระบวนการประเมินเพื่อคัดเลือกผู้จัดการครั้งนี้ดำเนินการโดยกำหนดให้ผู้จัดการแต่ละคน (C) ได้รับการสัมภาษณ์ด้วยแบบประเมินคุณลักษณะ (A) จำนวน 4 ด้านโดยชุดของนักจิตวิทยา (I) ที่แตกต่างกัน ดังนั้นการออกแบบดังกล่าวเป็นการออกแบบการวัดแบบแฝงบางส่วน สัญลักษณ์ที่ใช้ คือ (I:C) x A สามารถแสดงด้วยแผนภาพเวนน์ (Venn diagram)



โปรแกรมที่ใช้ในการวิเคราะห์สัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability Coefficient)

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดสามารถวิเคราะห์จากโปรแกรมสำเร็จรูป EduG พัฒนาโดยกลุ่มนักการศึกษาของแคนาดา (Canadian company Educac) ที่ได้รับการสนับสนุนทุนจากสถาบันทางวิชาการของสวิตเซอร์แลนด์และแคนาดา ร่วมกับสถาบันวิจัย (Cardinet; Johnson; & Pini. 2010: 9) เป็นโปรแกรมที่สามารถดาวน์โหลดจาก <http://www.irdp.ch/edumetrie/englishprogram.htm>

ขั้นตอนการวิเคราะห์ EduG (Cardinet; Johnson; & Pini. 2010: 38-60) ประกอบด้วย

1. การออกแบบการเก็บข้อมูล การประมาณค่า และการวัดตามเงื่อนไขที่สนใจ จำแนกเป็น

1.1 การออกแบบการเก็บข้อมูลและการประมาณค่า (observation and estimation designs)

1.2 การออกแบบการวัด (measurement design)

2. การศึกษาเชิงการสรุปอ้างอิง (G study) เป็นการบรรยายความแปรปรวนของแหล่งความคลาดเคลื่อนต่างๆ เพื่อสรุปอ้างอิงไปยังเอกภพของการวัด

3. การศึกษาการตัดสินใจเงื่อนไขการวัดที่ได้ค่าความเชื่อมั่นตามระดับที่ผู้วิจัยต้องการ (D study)

การวิจัยทางการศึกษากับการใช้ทฤษฎีการสรุป อ้างอิงความน่าเชื่อถือของผลการวัด

จากการรวบรวมงานวิจัยที่เกี่ยวข้องกับการใช้
ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด
มีดังนี้

ดวงใจ สีเขียว (2549: 207 – 214) ดำเนินการ
พัฒนาระบบการประเมินนิสิตนักศึกษาฝึกประสบการณ์
วิชาชีพตามแนวคิดการประเมินแบบ 360 องศา โดย
การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือ
ของผลการวัด โดยมีวัตถุประสงค์การวิจัยเพื่อ
พัฒนาและประเมินคุณภาพของระบบการประเมิน
นิสิตนักศึกษาฝึกประสบการณ์วิชาชีพตามแนวคิดการ
ประเมินแบบ 360 องศา ผลการศึกษาค่าสัมประสิทธิ์
การสรุปอ้างอิงของการประเมินความก้าวหน้าหรือ
ระหว่างการศึกษาฝึกประสบการณ์วิชาชีพครู จำนวน 4 ครั้ง
ด้วยผู้ประเมิน 4 แหล่ง แหล่งละ 2 คน พบว่ามีค่า
สัมประสิทธิ์การสรุปอ้างอิงเท่ากับ 0.7082 ส่วนการ
ประเมินสรุปรวมหรือสอบสอนที่มีการประเมิน จำนวน
2 ครั้งด้วยผู้ประเมิน 2 แหล่ง แหล่งละ 2 คน พบว่ามี
ค่าสัมประสิทธิ์การสรุปอ้างอิงเท่ากับ 0.6123

จากสาระสำคัญของผลงานวิจัยจะเห็นได้ว่า
นักวิจัยมีวัตถุประสงค์ของการศึกษาเชิงสรุปอ้างอิง
(generalizability study) การออกแบบการวัด และ
นำไปใช้เพื่อการวางแผนสำหรับการพัฒนาระบบการ
ประเมินนิสิตนักศึกษาฝึกประสบการณ์วิชาชีพ ดังนี้

1. เพื่อศึกษาค่าสัมประสิทธิ์สรุปอ้างอิงของผล
การประเมินนิสิตเมื่อจำนวนแหล่งผู้ประเมิน จำนวนผู้
ประเมินในแต่ละแหล่ง และจำนวนครั้งในการประเมิน
ที่แตกต่างกัน

2. เพื่อเปรียบเทียบค่าสัมประสิทธิ์สรุปอ้างอิง
ของผลการประเมินนิสิตเมื่อจำนวนแหล่งผู้ประเมิน

จำนวนผู้ประเมินในแต่ละแหล่ง และจำนวนครั้งในการ
ประเมินที่แตกต่างกัน

การออกแบบการวัดแบบแบ่งบางส่วนโดยมี
เงื่อนไขของการวัด 3 เงื่อนไข สัญลักษณ์ที่ใช้ คือ
 $P \times O \times (R:S)$ กล่าวคือนิสิตชั้นปีที่ 4 คณะครุศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย (P) ทุกคนได้รับการประเมิน
(O) ตั้งแต่ 1, 2, 3 และ 4 ครั้ง จากแหล่งผู้ประเมิน(S)
จำนวน 5 แหล่ง คือ อาจารย์นิเทศก์สถาบัน อาจารย์
นิเทศก์ประจำโรงเรียน นิสิตประเมินตนเอง เพื่อนนิสิต
และนักเรียน โดยจำนวนผู้ประเมิน (R) แต่ละแหล่ง คือ
ตั้งแต่ 1, 2, 3 และ 4 คนที่แตกต่างกัน

ข้อเสนอแนะในการนำผลการวิจัยไปใช้เพื่อให้
เกิดความน่าเชื่อถือของผลการวัด คือ การประเมิน
ความก้าวหน้าหรือระหว่างการศึกษาฝึกประสบการณ์วิชาชีพ
ครูควรทำการประเมินด้วยผู้ประเมิน 4 แหล่ง แหล่งละ
2 คน และประเมิน 4 ครั้ง ส่วนการประเมินผลแบบสรุป
รวมหรือสอบสอนของนิสิตฝึกประสบการณ์วิชาชีพครู
ควรทำการประเมินด้วยผู้ประเมิน 2 แหล่ง แหล่งละ
2 คน และประเมิน 2 ครั้ง

น้ำผึ้ง อินทะเนตร; งามอาจ นัยพัฒน์; ผจงจิต
อินทสุวรรณ; และสุทธิวรรณ พิรศักดิ์โสภณ (2554:
131) ทำการศึกษาเกี่ยวกับคุณลักษณะของคะแนน
สอบจากแบบทดสอบปลายเปิดวิชาคณิตศาสตร์เมื่อ
จำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่าง
กัน โดยใช้โมเดลการสรุปอ้างอิงและโมเดลหลายองค์
ประกอบของราล์ซ โดยมีวัตถุประสงค์การวิจัยเพื่อ
ศึกษาคุณลักษณะของคะแนนแบบทดสอบปลายเปิด
วิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจและรูปแบบการ
ตรวจให้คะแนนต่างกันโดยใช้โมเดลการสรุปอ้างอิง
และโมเดลหลายองค์ประกอบของราล์ซภายใต้เงื่อนไข
จำนวนผู้ตรวจต่างกันสามลักษณะ คือ 2 คน 3 คน

และ 4 คน และรูปแบบการตรวจให้คะแนนต่างกัน สามลักษณะ คือ รูปแบบที่ 1 ผู้ตรวจตรวจสอบข้อสอบทุกข้อของผู้สอบทุกคน รูปแบบที่ 2 ผู้ตรวจตรวจสอบข้อสอบทุกข้อของผู้สอบบางคน และรูปแบบที่ 3 ผู้ตรวจตรวจสอบข้อสอบบางข้อของผู้สอบทุกคน เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบปลายเปิด จำนวน 12 ข้อ กลุ่มตัวอย่าง คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 จากโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาจังหวัดน่าน ปีการศึกษา 2552 จำนวน 180 คน ซึ่งได้จากการสุ่มแบบสองขั้นตอน ผลการวิจัยพบว่าสำหรับการวิเคราะห์ด้วยโมเดลการสรุปอ้างอิงเมื่อใช้รูปแบบการตรวจให้คะแนนเดียวกันในทุกเงื่อนไขจำนวนผู้ตรวจ ความแปรปรวนขององค์ประกอบเดียวกันมีค่าใกล้เคียงกัน และค่าสัมประสิทธิ์การสรุปอ้างอิงในรูปแบบการตรวจที่ 2 มีค่าสูงสุด รองลงมาคือรูปแบบการตรวจที่ 1 และรูปแบบการตรวจที่ 3 ตามลำดับ ค่าสัมประสิทธิ์การสรุปอ้างอิงในรูปแบบการตรวจที่ 1 มีค่าสูงขึ้นเมื่อจำนวนผู้ตรวจเพิ่มขึ้น คะแนนในทุกเงื่อนไขที่ต่างกันมีความเที่ยงตรงตามสภาพสูง และแตกต่างกันไม่มีนัยสำคัญทางสถิติ

จากสาระสำคัญของผลงานวิจัยจะเห็นได้ว่านักวิจัยมีวัตถุประสงค์ของการศึกษาเชิงสรุปอ้างอิง (generalizability study) การออกแบบการวัด และนำไปใช้เพื่อการวางแผนสำหรับการทดสอบด้วยแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ที่มีอิทธิพลของจำนวนผู้ตรวจและรูปแบบการตรวจให้มีความน่าเชื่อถือ ดังนี้

1. เพื่อศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนสอบจากแบบทดสอบปลายเปิดวิชาคณิตศาสตร์เมื่อจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่างกัน
2. เพื่อเปรียบเทียบค่าสัมประสิทธิ์การสรุป

อ้างอิงของคะแนนสอบจากแบบทดสอบปลายเปิดวิชาคณิตศาสตร์เมื่อจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่างกัน

การออกแบบการวัดโดยมีเงื่อนไขของการวัด 2 เงื่อนไข ดำเนินการโดยนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 (P) ได้รับการทดสอบด้วยแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจ (R) คือ ตั้งแต่ 2, 3 และ 4 คน และรูปแบบการตรวจให้คะแนน (I) ต่างกัน ดังนั้นการออกแบบการวัด จำแนกเป็น 3 ลักษณะ คือ 1) การออกแบบการวัดแบบไขว้ คือ $P \times I \times R$ กล่าวคือผู้ตรวจตรวจสอบข้อสอบทุกข้อของผู้สอบทุกคน 2) การออกแบบการวัดแบบแฝงบางส่วน คือ $(P:R) \times I$ กล่าวคือผู้ตรวจตรวจสอบข้อสอบทุกข้อของผู้สอบบางคน และ 3) การออกแบบการวัดแบบแฝงบางส่วน คือ $(I:R) \times P$ กล่าวคือผู้ตรวจตรวจสอบข้อสอบบางข้อของผู้สอบทุกคน

ข้อเสนอแนะในการนำผลการวิจัยไปใช้เพื่อให้เกิดความน่าเชื่อถือของผลการวัด คือ ผลการวิจัยครั้งนี้ชี้ให้เห็นว่าความแปรปรวนของผู้ตรวจมีค่าค่อนข้างน้อย ค่าสัมประสิทธิ์การสรุปอ้างอิงและความเชื่อมั่นแยกส่วนของผู้สอบมีค่าสูงและมีความเที่ยงตรงตามสภาพของคะแนนในทุกเงื่อนไข ดังนั้นจึงไม่จำเป็นที่ผู้ตรวจต้องตรวจสอบข้อสอบทุกข้อของผู้สอบทุกคน สามารถเลือกรูปแบบการตรวจอื่นที่มีความสะดวกและประหยัดได้ จำนวนผู้ตรวจ 2-3 คน น่าจะเป็นจำนวนที่เหมาะสมและมีความเป็นไปได้ในปฏิบัติได้จริง

สาวิตรี จัยทอง; สุวิมล กฤษชฤทธานันท์; และ นิภา ศรีไพโรจน์ (2554: 110 - 111) ทำการศึกษาเพื่อสร้างชุดเครื่องมือการประเมินผลการเรียนรู้ตามสภาพจริงกลุ่มสาระการเรียนรู้คณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 และตรวจสอบคุณภาพชุดเครื่องมือการประเมินผลการเรียนรู้ตามสภาพจริง

โดยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด กลุ่มตัวอย่างเป็นนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 ของโรงเรียนสังกัดสำนักงาน เขตพื้นที่การศึกษามัธยมศึกษาเขต 1 จำนวน 7 โรงเรียน รวม 10 ห้องเรียน จำนวนนักเรียน 413 คน ซึ่งได้มาโดยวิธีการสุ่มแบบหลายขั้นตอน (multi - stage sampling) เครื่องมือที่ใช้ในการวิจัย คือ ชุดเครื่องมือการประเมินผลการเรียนรู้ตามสภาพจริง คือ แบบทดสอบวัดความสามารถจริง (authentic test) แบบสังเกตพฤติกรรมการแสดงออกในการเรียนรู้ (behavior - observation) แบบประเมินแฟ้มสะสมงาน (portfolio) แบบวัดการรับรู้ความสามารถของตนเอง (self - efficacy) และแบบวัดการกำกับตนเอง (self - regulation) สถิติที่ใช้ในการวิเคราะห์ คือ ค่าสัมประสิทธิ์แอลฟา (alpha coefficient) ของครอนบาค (Cronbach) สหสัมพันธ์แบบไบซีเรียล (biserial correlation) และค่าสัมประสิทธิ์การสรุปอ้างอิง (G - Coefficient) ผลการวิจัยพบว่าชุดเครื่องมือการประเมินผลการเรียนรู้ตามสภาพจริง รูปแบบที่ 1 ซึ่งประกอบด้วยแบบประเมินจำนวน 3 ฉบับ มีค่าสัมประสิทธิ์การสรุปอ้างอิง เท่ากับ 0.89 รูปแบบที่ 2 ซึ่งประกอบด้วยแบบประเมิน จำนวน 4 ฉบับ มีค่าสัมประสิทธิ์การสรุปอ้างอิง เท่ากับ 0.92 และรูปแบบที่ 3 ซึ่งประกอบด้วยแบบประเมิน จำนวน 5 ฉบับ มีค่าสัมประสิทธิ์การสรุปอ้างอิง เท่ากับ 0.93

จากสาระสำคัญของผลงานวิจัยจะเห็นได้ว่านักวิจัยมีวัตถุประสงค์ของการศึกษาเชิงสรุปอ้างอิง (generalizability study) การออกแบบการวัด และนำไปใช้เพื่อการวางแผนสำหรับการประเมินตามสภาพจริงกลุ่มสาระการเรียนรู้คณิตศาสตร์ด้วยเครื่องมือการประเมินที่หลากหลาย ดังนี้

1. เพื่อศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนเมื่อใช้จำนวนแบบประเมินตามสภาพจริงของกลุ่มสาระการเรียนรู้คณิตศาสตร์ที่ต่างกัน

2. เพื่อเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนเมื่อใช้จำนวนแบบประเมินตามสภาพจริงของกลุ่มสาระการเรียนรู้คณิตศาสตร์ที่ต่างกัน

การออกแบบการวัดแบบไขว้โดยมีเงื่อนไขของการวัด 1 เงื่อนไข สัญลักษณ์ที่ใช้ คือ $P \times T$ กล่าวคือ นักเรียนชั้นมัธยมศึกษาปีที่ 1 (P) ได้รับการประเมินโดยเครื่องมือการประเมินการเรียนรู้ตามสภาพจริง (T) จำนวน 5 ฉบับ คือ 1) แบบทดสอบวัดความสามารถจริง (authentic Test) 2) แบบสังเกตพฤติกรรมการแสดงออกในการเรียนรู้ โดยครูผู้สอนเป็นผู้สังเกตในระหว่างการจัดกิจกรรมการเรียนการสอน เรื่อง ระบบจำนวนเต็ม 3) แบบประเมินแฟ้มสะสมงาน (4) แบบวัดการรับรู้ความสามารถของตนเอง และ 5) แบบวัดการกำกับตนเอง

ข้อเสนอแนะในการนำผลการวิจัยไปใช้เพื่อให้เกิดความน่าเชื่อถือของผลการวัด คือ การประเมินผลการเรียนรู้ตามสภาพจริงกลุ่มสาระการเรียนรู้คณิตศาสตร์ จะต้องมีความสมดุลระหว่างความรู้ความเข้าใจในเนื้อหาสาระคณิตศาสตร์ มีทักษะกระบวนการทางคณิตศาสตร์ มีเจตคติที่ดีต่อคณิตศาสตร์ สามารถนำความรู้ทางคณิตศาสตร์ไปพัฒนาคุณภาพชีวิต ตลอดจนสามารถนำความรู้ทางคณิตศาสตร์ไปเป็นเครื่องมือในการเรียนรู้สิ่งต่างๆ และเป็นพื้นฐานในการศึกษาในระดับที่สูงขึ้น จากผลการวิจัยแสดงให้เห็นว่าควรใช้เครื่องมือการประเมินการเรียนรู้ตามสภาพจริง จำนวน 3 ฉบับ ได้แก่ แบบทดสอบวัดความสามารถจริง แบบสังเกตพฤติกรรมการแสดงออกในการเรียนรู้ (Behavior - Observation)

และแบบประเมินแฟ้มสะสมงาน จึงจะทำให้เกิดความน่าเชื่อถือของผลการวัดและสามารถดำเนินการได้ในทางปฏิบัติจริง

อกนิษฐ์ บุญสะอาด; สุวิมล กฤษฎคทาสน์; และ อธิธิฤทธิ์ พงษ์ปิยะรัตน์ (2557: 1906) ทำการศึกษาเกี่ยวกับการพัฒนาแบบประเมินความคิดรวบยอดวิชาฟิสิกส์โดยใช้แผนผังมโนทัศน์ด้วยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด โดยมีวัตถุประสงค์การวิจัยเพื่อพัฒนาและตรวจสอบคุณภาพของแบบประเมินความคิดรวบยอดโดยใช้แผนผังมโนทัศน์วิชาฟิสิกส์เรื่องงานและพลังงานของนักเรียนชั้นมัธยมศึกษาปีที่ 4 และเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบประเมินความคิดรวบยอดโดยใช้แผนผังมโนทัศน์แบบการสร้างผังคำและคำเชื่อม และแบบการเลือกคำเชื่อมที่มีลำดับของการสร้างแผนผังมโนทัศน์ และจำนวนครั้งในการวัดต่างกัน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 จำนวน 68 คน แผนการเรียนวิทยาศาสตร์ ผลการศึกษาพบว่าแบบประเมินความคิดรวบยอดวิชาฟิสิกส์โดยใช้แผนผังมโนทัศน์แบบการเลือกคำเชื่อม จำนวน 4 ฉบับมีค่าความเชื่อมั่นระหว่าง 0.65 ถึง 0.70 และแบบประเมินความคิดรวบยอดวิชาฟิสิกส์โดยใช้แผนผังมโนทัศน์แบบการสร้างผังคำและคำเชื่อม จำนวน 4 ฉบับมีค่าความเชื่อมั่นระหว่าง 0.66 ถึง 0.84 ค่าสัมประสิทธิ์การสรุปอ้างอิงของลำดับของการสร้างแผนผังมโนทัศน์ที่เริ่มจากแบบสร้างผังคำและคำเชื่อมตามด้วยแบบการเลือกคำเชื่อมมีค่าสูงกว่าของการสร้างแผนผังมโนทัศน์ที่เริ่มจากแบบการเลือกคำเชื่อมตามด้วยการสร้างผังคำและคำเชื่อมอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบประเมินความคิดรวบยอดโดยใช้แผนผังมโนทัศน์แบบ

การสร้างผังคำและคำเชื่อมมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในการวัด 1 ครั้ง และ 2 ครั้ง

จากสาระสำคัญของผลงานวิจัยจะเห็นได้ว่า นักวิจัยมีวัตถุประสงค์ของการศึกษาเชิงสรุปอ้างอิง (generalizability study) การออกแบบการวัด และนำไปใช้เพื่อการวางแผนสำหรับการประเมินความคิดรวบยอดวิชาฟิสิกส์โดยใช้แผนผังมโนทัศน์ ดังนี้

1. เพื่อศึกษาค่าสัมประสิทธิ์สรุปอ้างอิงของคะแนนจากแบบประเมินความคิดรวบยอดวิชาฟิสิกส์เมื่อนักเรียนใช้ลำดับของการสร้างผังมโนทัศน์และจำนวนครั้งของการประเมินต่างกัน
2. เพื่อเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนจากแบบประเมินความคิดรวบยอดวิชาฟิสิกส์เมื่อนักเรียนใช้ลำดับของการสร้างผังมโนทัศน์และจำนวนครั้งของการประเมินต่างกัน

การออกแบบการวัดโดยมีเงื่อนไขของการวัด 2 เงื่อนไข สามารถจำแนกเป็น 2 ลักษณะ คือ 1) การออกแบบการวัดแบบแบ่งบางส่วน คือ $(O:F) \times P$ เพื่อศึกษาความคิดรวบยอดที่ได้จากนักเรียนใช้วิธีการสร้างแผนผังมโนทัศน์ที่มีลำดับการสร้างต่างกัน ดำเนินการโดยให้นักเรียนชั้นมัธยมศึกษาปีที่ 4 (P) ได้รับการประเมินด้วยวิธีการสร้างแผนผังมโนทัศน์ (F) ที่แตกต่างกัน คือ กลุ่มที่ 1 เริ่มต้นจากการสร้างแผนผังมโนทัศน์ด้วยการสร้างผังคำและคำเชื่อมแล้วตามด้วยการเลือกคำเชื่อม กลุ่มที่ 2 เริ่มต้นจากการเลือกคำเชื่อมแล้วตามด้วยการสร้างแผนผังมโนทัศน์ด้วยการสร้างผังคำและคำเชื่อม ซึ่งผู้เรียนแต่ละกลุ่มได้รับการประเมิน (O) เพียง 1 ครั้ง และ 2) การออกแบบการวัดแบบไขว้ คือ $P \times F \times O$ เพื่อศึกษาคะแนนที่ได้จากจำนวนครั้งของการประเมินที่ต่างกัน กล่าวคือ นักเรียน

ชั้นมัธยมศึกษาปีที่ 4 (P) สร้างแผนผังมโนทัศน์ด้วยการสร้างผังคำและคำเชื่อม (F) ด้วยการประเมินซ้ำ (O) จำนวน 2 ครั้ง และนักเรียนชั้นมัธยมศึกษาปีที่ 4 (P) สร้างแผนผังมโนทัศน์ที่สร้างแผนผังมโนทัศน์ด้วยการเลือกคำเชื่อม ด้วยการประเมินซ้ำ (O) จำนวน 2 ครั้ง

ข้อเสนอแนะในการนำผลการวิจัยไปใช้เพื่อให้เกิดความน่าเชื่อถือของผลการวัด คือ การประเมินความคิดรวบยอดในรายวิชาฟิสิกส์ควรทำการประเมินด้วยการให้นักเรียนเริ่มต้นจากการสร้างแผนผังมโนทัศน์ด้วยการสร้างผังคำและคำเชื่อมด้วยตนเอง และครั้งที่ 2 ให้สร้างแผนผังมโนทัศน์ด้วยวิธีการเลือกคำเชื่อมที่ผู้สอนกำหนดให้ ควรทำการประเมิน 2 ครั้ง เพื่อให้ผลการประเมินความคิดรวบยอดมีความน่าเชื่อถือของผลการวัด

สรุป

ทฤษฎีการสร้างรูปร่างอิงความน่าเชื่อถือของผลการวัดเน้นการประมาณค่าความแปรปรวนของความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่งที่เกิดจากกระบวนการวัดและส่งผลให้เกิดความไม่คงที่ของคะแนนสังเกตได้ สำหรับแหล่งความคลาดเคลื่อนที่สามารถเกิดขึ้นในกระบวนการวัดอาจเกิดจากการที่ช่วงเวลาของการทดสอบซ้ำ จำนวนข้อของเครื่องมือวัด จำนวนผู้ตรวจ อิทธิพลของอันดับของค่าความยากของข้อสอบ และคุณลักษณะของผู้ตรวจ เป็นต้น เมื่อนักวิจัยทราบถึงแหล่งความคลาดเคลื่อนจึงสามารถออกแบบการวัดเพื่อควบคุมแหล่งความคลาดเคลื่อนได้ตรงประเด็นและทำให้นักวิจัยทราบถึงระดับความน่าเชื่อถือของผลการวัด เรียกว่าสัมประสิทธิ์การสร้างรูปร่างอิงความน่าเชื่อถือของผลการวัด (generalizability coefficient) คล้ายกับค่าสัมประสิทธิ์ความเชื่อมั่น (reliability coefficient) ที่สามารถบ่งบอกถึงความแม่นยำของการสร้างรูปร่างอิงจากคะแนนสังเกตได้ของผู้สอบแต่ละคนไปยังคะแนนเฉลี่ยของผู้สอบที่ควรจะได้จากการทดสอบภายใต้สถานการณ์หรือเงื่อนไขของการวัดที่ยอมรับได้ทั้งหมด

สุวิมล กฤษศยสาสน์

บรรณานุกรม

- ดวงใจ สีเขียว. (2549). **การพัฒนากระบวนการประเมินนิสิตนักศึกษาฝึกประสบการณ์วิชาชีพตามแนวคิดการประเมินแบบ 360 องศา โดยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด.** วิทยานิพนธ์ ค.ด.(การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- น้ำผึ้ง อินทเนตร; องอาจ นัยพัฒน์; ผจงจิต อินทสุวรรณ; และสุทธิวรรณ พิศศักดิ์โสภณ. (2555, กรกฎาคม). “การศึกษาคุณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์เมื่อจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่างกันโดยใช้โมเดลการสรุปอ้างอิงและโมเดลหลายองค์ประกอบของราล์ช,” **วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม**. 17(1): 131.
- ศิริชัย กาญจนวาสี. (2555). **ทฤษฎีการทดสอบแนวใหม่**. พิมพ์ครั้งที่ 4. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- สาวิตรี จ้อยทอง; สุวิมล กฤษศยาสา; และนิภา ศรีไพโรจน์. (2554-2555, ตุลาคม– มกราคม). “การตรวจสอบคุณภาพรูปแบบของชุดเครื่องมือการประเมินผลการเรียนรู้ตามสภาพจริง กลุ่มสาระการเรียนรู้คณิตศาสตร์ ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 โดยการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด,” **วารสารศึกษาศาสตร์**. 23(1): 110-121.
- อกนิษฐ์ บุญสะอาด; สุวิมล กฤษศยาสา; และอิทธิฤทธิ์ พงษ์ปิยะรัตน์. (2557, มกราคม). “การพัฒนาแบบประเมินความคิดรวบยอดวิชาฟิสิกส์โดยใช้แผนผังมโนทัศน์:การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด,” **วารสารพฤติกรรมศาสตร์เพื่อการพัฒนา**. 6 (1): 190.
- Brennan, Robert L. (2001). **Generalizability theory**. New York: Springer-Verlag.
- (2011). “Generalizability theory and classical test theory,” **Applied measurement in Education**. 24: 1–21
- Shavelson, Richard J.; & Webb, Noreen M. (1991). **Generalizability theory: A primer**. London: Sage Publications.
- Cardinet, Jean; Johnson, Sandra; & Pini, Gianreto (2010). **Applying generalizability theory using EduG**. New York: Routledge.