บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในบทนี้กล่าวถึงลักษณะของทฤษฎีการอ้างอิงสรุปและงานวิจัยที่เกี่ยวข้อง โดยแบ่งออกเป็น 3 ตอน คือ ตอนที่ 1 ลักษณะของทฤษฎีการอ้างอิงสรุป ตอนที่ 2 การพัฒนาแนวคิดเกี่ยวกับการวิเคราะห์ความตรงสู่เข้าตามทฤษฎีการอ้างอิงสรุป และตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 ลักษณะของทฤษฎีการอ้างอิงสรุป

ทฤษฎีการอ้างอิงสรุปมีลักษณะที่แตกต่างจากทฤษฎีการวัดแบบดั้งเดิม ในลักษณะต่อไปนี้

1. ด้านโมเดล

โมเดลของทฤษฎีการวัดแบบดั้งเดิม เป็นโมเดลเชิงเส้นตรงที่ไม่เน้นเรื่องการอ้างอิงสรุป แต่ทฤษฎีการอ้างอิงสรุปเป็นโมเดลเชิงสุ่ม ที่ถือว่าเงื่อนไขของการวัดที่นำมาศึกษานั้น ได้จากการสุ่มแบบง่าย หรือการสุ่มแบบชั้น จึงเป็นทฤษฎีที่ให้สารสนเทศเพื่อนำไปอ้างอิงสรุปยังเงื่อนไขการวัดอื่นๆ ในเอกภพเดียวกันได้ และการประยุกต์ใช้ของทฤษฎีนี้ เริ่มจากการใช้โมเดลการวิเคราะห์ความแปรปรวน ที่เหมาะสมต่อรูปแบบของการวัด (Schroeder and Hakstian, 1990: 249) โดยทำการประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ ด้วยการใช้กระบวนการวิเคราะห์ความแปรปรวน [Analysis of Variance (ANOVA) procedure] จึงกล่าวได้ว่าเป็น ANOVA of Measurement Model

2. ข้อตกลงเบื้องต้น

การประมาณค่าความเที่ยงด้วยทฤษฎีการวัดแบบดั้งเดิม ถือว่า การวัดนั้นมีความเท่าเทียมกัน ในลักษณะของความเป็นคู่ขนาน ซึ่งในทางปฏิบัติมักประสบปัญหา ในขณะที่ทฤษฎีการอ้างอิงสรุปไม่มีข้อตกลงดังกล่าวนี้ ทฤษฎีนี้ไม่ได้ขึ้นตรงกับข้อตกลงเบื้องต้นอย่างเข้มงวดของการวัดที่เป็นแบบคู่ขนาน เป็นทฤษฎีที่ยินยอมให้นักวิจัยทำการออกแบบวิธีการเก็บรวบรวมข้อมูลให้เหมาะสมสำหรับปัญหาของการวัดผล ภายใต้การพิจารณาต่าง ๆ กันได้ จึงสามารถประยุกต์ใช้ได้กว้างขวางกว่า

3. แหล่งของความคลาดเคลื่อนหรือแหล่งของความแปรปรวน

ทฤษฎีการวัดแบบดั้งเดิม ถือว่า แหล่งความคลาดเคลื่อนมีเพียงแหล่งเดียว จึงสามารถประมาณความคลาดเคลื่อนได้จากแหล่งความคลาดเคลื่อนเพียงแหล่งเดียว ในขณะที่ทฤษฎี

การอ้างอิงสรุปมีจุดแกร่งกว่า ในลักษณะของการประมาณความคลาดเคลื่อนในการวัดได้หลาย ๆ แหล่ง โดยจะประมาณค่าความแปรปรวนได้จากการวิเคราะห์ครั้งเดียว (single analysis) มุ่งประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ หลายแหล่ง ที่เป็นไปได้ทั้งหมด ทั้งจากผลหลัก (Main Effect) และผลของผลร่วมหรือปฏิสัมพันธ์ (Interaction Effect) ระหว่างองค์ประกอบด้วย ช่วยให้มีความถูกต้อง (Accuracy) แม่นยำ (Precision) และได้สารสนเทศที่เพียงพอ สำหรับการตัดสินใจเลือกแบบการวัดที่มีประสิทธิภาพ นอกจากนี้ ยังเป็นทฤษฎีที่ยินยอมให้ผู้ตัดสินใจเลือกใช้การแปลความที่แตกต่างกันได้ ทั้งในลักษณะของความแปรปรวนของความคลาดเคลื่อนสัมพัทธ์ และความแปรปรวนของความคลาดเคลื่อนสัมบูรณ์

4. จำนวนฟาเซ็ตหรือองค์ประกอบที่ศึกษา

ในการประมาณความเที่ยงตามทฤษฎีการวัดแบบดั้งเดิมนั้น ในการวิเคราะห์แต่ละครั้ง จะทำการวิเคราะห์เพียงฟาเซ็ตเดียว (Single Facet) เช่น ข้อสอบ (Items) เป็นต้น ดังนั้น ในการปรับรูปแบบเพื่อให้ได้ค่าความเที่ยงเพิ่มขึ้น จะกระทำได้เพียงแต่การเพิ่มจำนวนของข้อสอบเหล่านั้น นอกจากนี้ ถ้าใช้สูตรสเปียร์แมนบราวน์ (Spearman-Brown Formula) คำนวณค่าสัมประสิทธิ์ความเที่ยงของแบบสอบ โดยเพิ่มจำนวนข้อสอบ ในกรณีที่ศึกษาตั้งแต่สองฟาเซ็ตขึ้นไป จะได้ค่าประมาณความเที่ยงที่สูงกว่าที่ควรจะเป็น เนื่องจากทฤษฎีนี้ไม่มีการจำแนกแหล่งความแปรปรวนออกเป็นหลายแหล่ง ในขณะที่ทฤษฎีการอ้างอิงสรุปสามารถ ศึกษาได้หลายฟาเซ็ต (Multifacet) พร้อมกัน รวมทั้งจำแนกแหล่งความแปรปรวนที่เป็นไปได้ทั้งหมด จึงได้ค่าความแปรปรวนของความคลาดเคลื่อนที่ถูกต้อง ทำให้ค่าประมาณความเที่ยงที่ได้ถูกต้องและแม่นยำกว่า และสามารถปรับรูปแบบการวัด เพื่อให้ได้ค่าความเที่ยงที่ดีขึ้นได้หลายวิธี โดยได้จากการเพิ่ม-ลดในฟาเซ็ตต่าง ๆ อาทิ จำนวนข้อสอบ จำนวนผู้ประเมิน จำนวนครั้งที่ใช้สอบ จำนวนแบบสอบย่อย เป็นต้น

ทฤษฎีการอ้างอิงสรุปสามารถที่จะช่วยให้ผู้ตัดสินใจ ได้ทำการพิจารณาว่า จะจัดกระทำกับข้อมูลในการวัดผลอย่างไร เป็นต้นว่าเพื่อให้มีโอกาส รูปแบบของแบบสอบ และการบริหารด้วยจำนวนข้อสอบเท่าใด เพื่อต้องการให้ได้รับคะแนนที่เชื่อถือได้

5. ความหมายของคำที่ใช้

ประชากร

ประชากร (Population) ตามทฤษฎีนี้หมายถึงสิ่งที่ต้องการวัด (Object of Measurement) หรือสิ่งที่มุ่งวัดทั้งหมด เช่น ผู้สอบ บุคคล นักเรียน หรือ นักเรียนทั้งชั้น (Brennan, 1983: 3) ในสถานการณ์การสอบทั่วไป สิ่งที่มุ่งวัดมักได้แก่ บุคคล หรือผู้ทำการสอบ

ฟาเซ็ตของการวัด

ฟาเซ็ต (Facet) ของการวัด เป็นชุดของเงื่อนไขของการวัด ที่มีลักษณะคล้ายกัน หรือกลุ่มเงื่อนไขของการวัด ซึ่งเป็นองค์ประกอบที่คาดว่ามีผลต่อการวัดความคลาดเคลื่อน เช่น ความยาวของแบบสอบ รูปแบบของข้อสอบ จำนวนครั้งของการสอบ จำนวนผู้ตรวจให้คะแนน โดยที่ฟาเซ็ตที่ต้องการศึกษา อาจเป็นองค์ประกอบสุ่มหรือองค์ประกอบที่เจาะจง ถ้าเงื่อนไขการวัดถูกเลือกมาอย่างเจาะจงจากองค์ประกอบที่ศึกษา แสดงว่า ผู้ศึกษาสามารถทำการสรุปอ้างอิงไปยังองค์ประกอบเฉพาะในระดับของเงื่อนไขที่เลือกมาศึกษาเท่านั้น แต่ถ้าเงื่อนไขการวัดได้รับการสุ่มเพื่อเป็นตัวแทนองค์ประกอบที่ศึกษา แสดงว่าผู้ศึกษาสามารถทำการสรุปอ้างอิงไปยังระดับต่าง ๆ ขององค์ประกอบที่ศึกษาได้ ฟาเซ็ตของการวัดที่ใช้มากในการสอบ ได้แก่ ฟาเซ็ตข้อสอบ (item facet) ฟาเซ็ตผู้ตรวจข้อสอบ (rater facet)

เงื่อนไขของการวัด

เงื่อนไขของการวัด (condition of measurement) เป็นส่วนต่าง ๆ ของฟาเซ็ตหรือระดับของฟาเซ็ตของการวัด เช่น ฟาเซ็ตของจำนวนผู้ตรวจ อาจกำหนดจำนวนระดับเป็น 1, 2 และ 3 คน ฟาเซ็ตของความยาวแบบสอบ อาจกำหนดระดับความยาวของแบบสอบเป็น 10, 20 และ 30 ข้อ เป็นต้น ดังนั้นข้อกระทงหรือข้อสอบแต่ละข้อ จึงเป็นเงื่อนไขการวัดหนึ่ง ๆ ของฟาเซ็ตข้อกระทง ผู้ตรวจข้อสอบแต่ละคนเป็นเงื่อนไขการวัดหนึ่ง ๆ ของฟาเซ็ตผู้ตรวจข้อสอบ

เอกภพ

เอกภพ (Universe) หมายถึง เงื่อนไขของการวัดทั้งหมดของแต่ละฟาเซ็ต หรือกล่าวได้ว่า เป็นเงื่อนไขของการวัดที่สนใจทั้งหมด เช่น จำนวนข้อกระทงทั้งหมด จำนวนผู้ตรวจทั้งหมด

เอกภพของการสังเกตที่ยอมรับได้ หรือ เอกภพของค่าที่ได้จากการสังเกตทั้งหมด (Universe of admissible observation) เป็นกลุ่มเงื่อนไขของการวัดที่เป็นไปได้ ซึ่งสามารถวัดหรือสังเกตได้ในแต่ละฟาเซ็ต เช่น ประกอบด้วยฟาเซ็ตข้อสอบ และฟาเซ็ตของผู้ตรวจ

เอกภพของการอ้างอิง (Universe of generalization) เป็นเงื่อนไขการวัดทั้งหมดที่เป็นเป้าหมายของการสรุปอ้างอิง กล่าวได้ว่า เป็นการวัดที่ครอบคลุมเงื่อนไขที่สนใจทั้งหมด หรือเป็นเงื่อนไขในเอกภพของการสังเกตที่ยอมรับได้ทั้งหมด ซึ่งอาจประกอบด้วยเซ็ตย่อยของเงื่อนไขในเอกภพของการสังเกตที่ยอมรับได้

การศึกษา G (G-Study) และการศึกษา D (D-Study)

ทฤษฎีการอ้างอิงสรุป ประกอบด้วยขั้นตอนของการศึกษาที่สำคัญ 2 ขั้นตอน คือ การศึกษาเพื่อการอ้างอิงสรุป หรือการศึกษา G (Generalizability Study or G-Study) และการศึกษาเพื่อการตัดสินใจ หรือการศึกษา D (Decision Study or D-Study)

การศึกษา G

การศึกษา G (G Study) เป็นการสรุปอ้างอิงผลที่ได้จากการศึกษาตัวอย่างการวัดตามเงื่อนไขที่สนใจ บรรยายความแปรปรวนของความคลาดเคลื่อน จากแหล่งความคลาดเคลื่อนต่าง ๆ เพื่อสรุปอ้างอิงไปยังเอกภพของการวัด

การศึกษา D

การศึกษา D (D-Study) เป็นการใช้ข้อมูลจากการศึกษา G ที่สอดคล้องกับจุดประสงค์เฉพาะของการวัด ตัดสินใจเลือกใช้แบบสอบหรือวิธีวัด ในสถานการณ์ต่าง ๆ ของการวัด

จุดประสงค์ของการศึกษา G คือ ความต้องการประมาณค่าความแปรปรวนของคะแนนจริง และความแปรปรวนของคะแนนความคลาดเคลื่อนจากแหล่งความคลาดเคลื่อนต่าง ๆ ที่สนใจ และใช้เป็นข้อมูลสำหรับวางแผนเพื่อตัดสินในใจ ในการศึกษา D จะเกี่ยวกับความเที่ยงของแบบสอบในสถานการณ์ของการวัดต่าง ๆ ดังนั้น การออกแบบ G-Study จึงควรครอบคลุมเงื่อนไขของการวัดที่ต้องการตัดสินใจนำแบบสอบไปใช้ใน D-Study

ความแปรปรวนของคะแนน

ความแปรปรวนที่สำคัญของทฤษฎีการอ้างอิงสรุป ได้แก่ ความแปรปรวน 2 ประเภท คือ ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ และความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ (Absolute Error Variance and Relative Error Variance)

ตามทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory)นั้น คะแนนจริงของผู้สอบ (true score : T_p) คือค่าเฉลี่ยของคะแนนจากการสอบซ้ำ ๆ ด้วยแบบสอบคู่ขนาน จึงเป็นความแปรปรวนของค่าเฉลี่ยของการสอบซ้ำ และความแปรปรวนของคะแนนที่สังเกตได้ เป็นผลรวมของความแปรปรวนของคะแนนจริงกับความแปรปรวนของความคลาดเคลื่อน ดังนี้

$$X = T_{pt} + E_{pt}$$

$$\sigma^2_x = \sigma^2_T + \sigma^2_E$$

สำหรับทฤษฎีการอ้างอิงสรุปนั้น คะแนนเอกภพ (Universe score : μ_p) คือค่าเฉลี่ยของคะแนนการวัดซ้ำหลาย ๆ ครั้งตามเงื่อนไขการวัดในการสรุปอ้างอิง สำหรับความคลาดเคลื่อนของการวัด ถูกจำแนกออกเป็น ความคลาดเคลื่อนจากฟาเซ็ตหรือกลุ่มเงื่อนไขของการวัด และความคลาดเคลื่อนจากแหล่งที่เหลืออื่นๆ ซึ่งการวัดแต่ละครั้งไม่จำเป็นต้องใช้แบบสอบคู่ขนานเหมือนทฤษฎีการวัดแบบดั้งเดิม ส่วนความแปรปรวนของค่าคาดหมายของคะแนนที่สังเกตได้ เป็นผลรวมของความแปรปรวนของคะแนนเอกภพ (σ^2_p) กับความแปรปรวนของคะแนนความคลาดเคลื่อนจากฟาเซ็ตหรือองค์ประกอบต่าง ๆ ของการวัด และความแปรปรวนของความคลาดเคลื่อนจากแหล่งอื่น ๆ

ความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่าง ๆ ของการวัด แบ่งออกเป็น 2 ประเภท ได้แก่

ความแปรปรวนของคะแนนความคลาดเคลื่อนแบบสัมบูรณ์ (Absolute error variance : σ^2_{Abs} หรือ $\sigma^2(\Delta)$ หรือ ความแปรปรวนของ$(\mu_p - X_p)$) เป็นความแปรปรวนของผลต่างระหว่างคะแนนที่สังเกตได้และคะแนนเอกภพ ซึ่งมีค่าเท่ากับผลรวมของความแปรปรวนที่ประมาณได้ทั้งหมด ยกเว้นความแปรปรวนจากแหล่งบุคคล (σ^2_p) เท่านั้น ดังนั้นจึงคำนวณได้จาก ผลรวมของความแปรปรวนของคะแนนจากแหล่งต่าง ๆ ยกเว้น σ^2_p

ความแปรปรวนของคะแนนความคลาดเคลื่อนแบบสัมพัทธ์ (Relative error variance : σ^2_{Rel} หรือ $\sigma^2(\delta)$) เป็นความแปรปรวนของ $\mu_p - X_p$ นั่นคือ เป็นความแปรปรวนของผลต่างระหว่างส่วนเบี่ยงเบนคะแนนที่สังเกตได้ (ความแตกต่างของคะแนนที่สังเกตได้จากคะแนนเฉลี่ยของประชากรของคะแนนที่สังเกตได้) และส่วนเบี่ยงเบนคะแนนเอกภพ (ความแตกต่างของคะแนนเอกภพจากคะแนนเฉลี่ยของประชากรของคะแนนเอกภพ)

ความแปรปรวนของคะแนนความคลาดเคลื่อนแบบสัมพัทธ์ จึงคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่าง ๆ ที่มีปฏิสัมพันธ์กับผู้สอบ

ค่าสัมประสิทธิ์การอ้างอิงสรุป

ค่าสัมประสิทธิ์การอ้างอิงสรุป (Generalizability coefficient : $E\rho^2$) เป็นอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพ และความแปรปรวนของคะแนนสังเกตที่คาดหวัง ซึ่งค่าสัมประสิทธิ์การอ้างอิงสรุป สามารถประมาณได้จากกำลังสองของค่าสหสัมพันธ์ระหว่างคะแนนที่สังเกตได้และคะแนนเอกภพ

หรือกล่าวได้อีกนัยหนึ่งว่า ตามทฤษฎีการอ้างอิงสรุปนั้น สัมประสิทธิ์การอ้างอิงสรุปเป็นสัดส่วนระหว่างความแปรปรวนของคะแนนเอกภพ กับความแปรปรวนของค่าคาดหมายของคะแนนที่สังเกตได้

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

สัมประสิทธิ์การอ้างอิงสรุป จึงมีคุณสมบัติของความเป็นดัชนี ที่อธิบายความแม่นยำของการวัด เช่นเดียวกับค่าสัมประสิทธิ์ความเที่ยงแบบดั้งเดิม สามารถใช้คำนวณช่วงความเชื่อมั่นของคะแนนเอกภพ หรือใช้ในสมการถดถอยในการประมาณค่าคะแนนเอกภพ และใช้ในการปรับแก้ค่าสหสัมพันธ์ที่ลดลง อันเนื่องจากความคลาดเคลื่อน

สัมประสิทธิ์การอ้างอิงสรุป ใช้เป็นค่าประมาณของค่าเฉลี่ยของสหสัมพันธ์ ระหว่างค่าการวัดที่สุ่มมาจากเอกภพรายคู่ (Cronbach, 1972: 157) เช่น ค่าสัมประสิทธิ์การอ้างอิงสรุปเมื่ออ้างอิงไปยังชุดข้อสอบ (แบบสอบ) ซึ่งประกอบด้วยข้อสอบ 20 ข้อ มีค่าเป็น 0.83 นั้นหมายความว่า ถ้าเราสุ่มนักเรียนมาจากประชากรหนึ่ง สมมติเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ของอำเภอหนึ่งมาทำการทดสอบ สุ่มแบบสอบมาทีละฉบับ ๆ ละ 20 ข้อที่ไม่ซ้ำกัน ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบสอบที่สุ่มมาจะมีค่าเป็น .83

สัมประสิทธิ์การอ้างอิงสรุป ในลักษณะที่เป็นค่ากำลังสองของค่าสหสัมพันธ์ ระหว่างคะแนนเอกภพกับคะแนนสังเกต (Brennan and Kane, 1979: 40) ได้แก่ ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบคณิตศาสตร์ "เรื่องสมการ" ของชั้นประถมศึกษาปีที่ 6 จำนวน 20 ข้อ มีค่าเป็น 0.90 แสดงว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนเอกภพเรื่องสมการ กับคะแนนสังเกตของนักเรียนประถมศึกษาปีที่ 6 ยกกำลังสองมีค่าเป็น 0.90 ถ้าถอดรากที่สอง จะได้ค่าสหสัมพันธ์ระหว่างคะแนนเอกภพเรื่องสมการกับคะแนนสังเกต

ค่าสัมประสิทธิ์การอ้างอิงสรุป สามารถอธิบายในรูปอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพกับคะแนนสังเกต (Cronbach, et al. 1972) เช่น สัมประสิทธิ์การอ้างอิงสรุปเป็น .90 แสดงว่าความแตกต่างที่วัดได้ ร้อยละ 90 เป็นความแตกต่าง อันเนื่องมาจากคะแนนเอกภพ อีกเพียงร้อยละ 10 เป็นความแตกต่างอันเนื่องมาจากความคลาดเคลื่อน

เนื่องจากความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่างๆ ของการวัด มี 2 ประเภท จึงทำให้สัมประสิทธิ์การอ้างอิงสรุปมี 2 ประเภท ดังนี้

1. สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจสัมบูรณ์

เมื่อคะแนนความคลาดเคลื่อน เป็นความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ สัมประสิทธิ์นี้บ่งบอกความเที่ยงของแบบสอบ ในสถานการณ์ของการตัดสินใจ ที่ขึ้นกับคะแนนของผู้สอบตามลำพัง ไม่มีการเปรียบเทียบภายในกลุ่มหรือระหว่างกลุ่ม เช่น แบบสอบอิงเกณฑ์ ซึ่งใช้ตรวจสอบความสามารถของผู้สอบเทียบกับเกณฑ์ที่กำหนดไว้ล่วงหน้า

2. สัมประสิทธิ์การอ้างอิงสรุปสำหรับการตัดสินใจสัมพัทธ์

เมื่อคะแนนความคลาดเคลื่อน เป็นความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ สัมประสิทธิ์นี้บ่งบอกความเที่ยงของแบบสอบ ในสถานการณ์ของการตัดสินใจ ที่มีการเปรียบเทียบคะแนนระหว่างผู้สอบ เช่น แบบสอบอิงกลุ่มซึ่งใช้เปรียบเทียบคะแนนระหว่างผู้สอบ

ด้วยการเข้าโค้งการแจกแจงคะแนน และตัดเกรดหรือให้ระดับผลการเรียน หรือการสอบแข่งขันเพื่อหาผู้ที่ได้คะแนนสูงเข้าศึกษา

ในขณะที่ตามทฤษฎีการวัดแบบดั้งเดิมนั้น สัมประสิทธิ์ความเที่ยงของแบบสอบจะเป็นสัดส่วนระหว่างความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่สังเกตได้

ถึงแม้ว่าค่าสัมประสิทธิ์การอ้างอิงสรุป จะมีความหมายที่ใช้เช่นเดียวกับความเที่ยงของทฤษฎีการวัดแบบดั้งเดิม แต่ก็มีประเด็นที่แตกต่างกันดังนี้

1. การวัดแต่ละครั้งมีค่าสัมประสิทธิ์การอ้างอิงสรุปได้มากกว่า 1 ค่า

2. การอ้างอิงไปยังเอกภพใด จะต้องระบุและอธิบายเอกภพนั้นให้ชัดเจน และต้องสุ่มเงื่อนไขนั้นมาศึกษาด้วย

3. ค่าสัมประสิทธิ์การอ้างอิงสรุป สามารถบอกถึงความเป็นเอกพันธ์ของเอกภพได้ ถ้าข้อสอบที่นำมาศึกษาเป็นตัวอย่างสุ่มจากเอกภพ ข้อสอบที่มีความเป็นเอกพันธ์ จะสามารถใช้คะแนนสังเกตแทนคะแนนเอกภพได้อย่างมั่นใจ

Signal-Noise Ratio

Signal-Noise Ratio เป็นดัชนีที่ใช้ในการแปลความความแปรปรวนของคะแนนความคลาดเคลื่อน (Brennan, 1983: 17) ที่นิยมใช้ต่อการอ้างอิงดัชนีของคุณภาพในกระบวนการวัด โดยใช้ในการเปรียบเทียบขนาดความแปรปรวนของคะแนนความคลาดเคลื่อน กับความแปรปรวนของคะแนนเอกภพ ในลักษณะของการเปรียบเทียบความแกร่งของกระบวนการส่งผ่าน (the strength of the transmission) กับความแกร่งของการเข้ามาเกี่ยวพัน (the strength of the interference) (Brennan, 1983: 17)

Signal-Noise Ratio เป็นอัตราส่วนที่เกิดจากส่วนสำคัญ 2 ส่วน คือ Signal และ Noise โดยที่ Signal มุ่งที่จะระบุขนาดความเข้มของการจำแนกในสิ่งที่ต้องการ และ Noise มุ่งที่จะระบุผลของตัวแปรแทรกซ้อนที่เข้ามาเกี่ยวพันกับการจำแนก ในลักษณะของการสะท้อนขนาดของความคลาดเคลื่อนที่เกิดจากกระบวนการวัด เป็นสิ่งที่สะท้อนความแม่นยำในกระบวนการวัด ถ้าค่า Signal มีขนาดใหญ่ เมื่อเทียบกับ Noise แล้ว แสดงว่าการจำแนกในสิ่งที่ตั้งใจไว้ก็ทำได้ง่ายขึ้น แต่ถ้าค่า Signal อ่อนลงมากเมื่อเทียบกับ Noise แล้ว แสดงให้เห็นว่าไม่สามารถทำการจำแนกในสิ่งที่ตั้งใจไว้ได้ ทั้งนี้ค่า Signal-Noise Ratio ได้จากอัตราส่วนของ Noise Power กับ Signal Power โดยที่ Signal Power เป็นค่าเฉลี่ยกำลังสองของความแตกต่างในประชากร นั่นคือเป็นความแปรปรวนของคะแนนเอกภพ (σ^2_p) สำหรับค่าของ Noise เป็นการพิจารณาว่าคะแนนสังเกตของแตกต่างจากคะแนนเอกภพอย่างไร

Signal-Noise Ratio ได้นำมาใช้ในทฤษฎีการอ้างอิงสรุป โดยจำแนกเป็น 2 ประเภทตามลักษณะของประเภทความแปรปรวนของคะแนนความคลาดเคลื่อน ในลักษณะของความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ และความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ โดยมีสูตรที่ใช้ในการคำนวณ Signal-Noise Ratio ดังนี้

1. ในกรณีของ Relative Error Variance

ค่า Signal-Noise Ratio จะให้ดัชนีของความแม่นยำสัมพัทธ์ของกระบวนการวัด สำหรับการแปลความเปรียบเทียบคะแนนของผู้เข้าสอบ (Brennan, 1983: 19) ดังนั้น

$$\lambda_{\delta} = \sigma^2_{p} / \sigma^2(\delta)$$

โดย λ_{δ} เป็น Signal-Noise Ratio สำหรับความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์

σ^2_{p} เป็นความแปรปรวนของคะแนนเอกภพ

$\sigma^2(\delta)$ เป็นความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์

2. ในกรณีของ Absolute Error Variance

ค่า Signal-Noise Ratio จะให้ดัชนีของความแม่นยำสัมพัทธ์ของกระบวนการวัดสำหรับการแปลความโดเมนของคะแนน (Brennan, 1983: 19) ดังนั้น

$$\lambda_{\Delta} = \sigma^2_{p} / \sigma^2(\Delta)$$

โดย λ_{Δ} เป็น Signal-Noise Ratio สำหรับความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์

σ^2_{p} เป็นความแปรปรวนของคะแนนเอกภพ

$\sigma^2(\Delta)$ เป็นความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์

6. **ขั้นตอนของทฤษฎีการอ้างอิงสรุป**

เบรนเนน (Brennan) ได้เสนอกรอบของทฤษฎีการอ้างอิงสรุป โดยเสนอขั้นตอนของการประมาณค่าความเที่ยงออกเป็น 2 ขั้นตอน ขั้นตอนแรกเป็นการศึกษาเพื่อการอ้างอิงสรุป หรือการศึกษา G (Generalizability Study: G Study) ขั้นตอนที่สองเป็นการศึกษาเพื่อการตัดสินใจ หรือการศึกษา D (Decision Study: D Study) สำหรับหลักการของการดำเนินงานในแต่ละขั้นตอน มีดังต่อไปนี้

1) **การศึกษาเพื่อการอ้างอิงสรุป**

การศึกษาเพื่อการอ้างอิงสรุป เป็นกระบวนการที่มุ่งประมาณความแปรปรวนจากแหล่งต่าง ๆ ภายใต้เอกภพของคะแนนสังเกตที่ยอมรับได้ โดยมีขั้นตอนดังนี้

1. กำหนดสิ่งที่จะวัด เช่น กำหนดสิ่งที่ต้องการวัด คือ ความสามารถด้านการเขียนของนักเรียน

2. กำหนดฟาเซ็ตของการวัด เช่นกำหนดให้มี 2 ฟาเซ็ต คือ ข้อสอบ i) และผู้ตรวจให้คะแนน (r)

3. กำหนดเอกภพของการสังเกตที่ยอมรับได้ ในที่นี้ประกอบด้วยฟาเซ็ตข้อสอบฟาเซ็ตของผู้ตรวจให้คะแนน

4. กำหนดความสัมพันธ์ระหว่างฟาเซ็ต ว่าจะเป็นแบบ Crossed ซึ่งแทนด้วย X หรือ nested แทนด้วย : ถ้าต้องการศึกษาแบบให้ผู้ตรวจทุกคนตรวจข้อสอบทุกข้อ ก็เป็นแบบ crossed คือ i x r แต่ถ้าให้ผู้ตรวจแต่ละคนตรวจชุดข้อสอบต่างกัน ก็ถือว่าเป็นแบบ nested คือ i : r ทั้งนี้แล้วแต่ความสนใจของผู้ศึกษา ซึ่งโดยหลักการแล้ว ถ้าเป็น แบบ crossed design จะทำให้ได้จำนวนแหล่งของความแปรปรวน (variance component) มากกว่าแบบ nested design ซึ่งมีประโยชน์ที่จะทำให้ทราบว่า ความแปรปรวนแหล่งใด มีผลต่อสัมประสิทธิ์การอ้างอิงสรุป และช่วยให้สามารถเลือกรูปแบบ ที่เหมาะสมในขั้น D Study ตลอดจนได้ค่าสัมประสิทธิ์การอ้างอิงสรุปที่มีค่ามากกว่า nested design ในกรณีของ 1-facet design แต่ในแง่ปฏิบัตินั้น nested design จะสามารถใช้ได้ง่ายกว่า กล่าวคือ การให้ผู้ตรวจแต่ละคนตรวจข้อสอบต่างข้อกัน โดยสรุปแล้วในขั้น G Study นั้น ควรใช้ crossed design เพราะขั้นนี้มุ่งประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ อย่างเต็มที่ สำหรับในขั้น D Study การเลือกใช้รูปแบบใดขึ้นอยู่กับความสนใจที่ต้องการศึกษา

5. การเก็บรวบรวมข้อมูล โดยการสุ่มตัวอย่างมาศึกษา เช่น ในที่นี้สุ่มตัวอย่างนักเรียน n_p คน ตัวอย่างข้อสอบมา n_i ข้อ และตัวอย่างผู้ตรวจ n_r คน จะเป็นการศึกษาแบบ G Study p x i x r design ในกรณีนี้จะประกอบด้วย แหล่งความแปรปรวน 7 แหล่ง เป็นผลหลัก (main effect) 3 แหล่ง คือ ผลจากนักเรียน ผลจากข้อสอบและผลจากผู้ตรวจข้อสอบ

และเป็นผลของปฏิสัมพันธ์อีก 4 แหล่ง ได้แก่ ปฏิสัมพันธ์ระหว่างนักเรียนกับข้อสอบ นักเรียนกับผู้ตรวจข้อสอบ ข้อสอบกับผู้ตรวจ และนักเรียนกับข้อสอบและผู้ตรวจ ในกรณีนี้จะมีผลจากแหล่งต่าง ๆ ดังนี้

Design	Main Effects	Interaction Effects
p x i x r	p, i, r	pi, pr, ir, pir
p x (i : r)	p, r, i : r	pr, pi : r
(i : p) x r	p, r, i : p	pr, ir : p
i : (p x r)	p, r, i : pr	pr
(i x r) : p	p, i : p, r : p	ir : p
i : r : p	p, r : p, i : r : p	

6. การวิเคราะห์ข้อมูล หลังจากผ่านขั้นตอนในการพิจารณารูปแบบการวัด และได้นำไปสู่การเก็บรวบรวมข้อมูลจากรูปแบบที่กำหนดแล้ว ข้อมูลที่ได้จะนำไปสู่การวิเคราะห์ทางสถิติเพื่อประมาณขนาดของความคลาดเคลื่อนในการวัดจากแหล่งต่าง ๆ กระบวนการทางสถิติที่ใช้ในทฤษฎีการอ้างอิงสรุปก็คือ กระบวนการวิเคราะห์ความแปรปรวน (Analysis of Variance (ANOVA) Procedure) โดยเฉพาะอย่างยิ่งหลังจากที่ได้เก็บรวบรวมข้อมูลแล้ว ก็จะทำการวิเคราะห์ผ่านกระบวนการ ANOVA ในขั้นตอนของการวิเคราะห์ในการศึกษา G (G-study)

ในการวิเคราะห์ความแปรปรวนที่นำไปใช้โดยทั่วไปนั้น จะใช้ในการทดสอบนัยสำคัญทางสถิติของสถิติอ้างอิง โดยมีจุดมุ่งหมายเพื่อประมาณค่า F-ratio หรือทดสอบความมีนัยสำคัญของสัดส่วนความแปรปรวน โดยการใช้สถิติทดสอบ F (F-test) และใช้ F-ratio นี้ในการตัดสินใจว่าจะคงไว้หรือจะปฏิเสธสมมุติฐานศูนย์นั้น(retain or reject null hypothesis) แต่ในการประยุกต์ใช้ ANOVA กับทฤษฎีการอ้างอิงสรุปนั้น F-ratio เป็นสิ่งที่ไม่มีความจำเป็นใด ๆ แต่ได้ใช้ค่าประมาณเฉลี่ยกำลังสอง (Mean Squares - MS) จากการวิเคราะห์ความแปรปรวน เพื่อใช้เป็นฐานในการประมาณองค์ประกอบความแปรปรวนต่อไป โดยมีจุดมุ่งหมายที่นำ ANOVA มาใช้ก็เพื่อประมาณค่าความแปรปรวน ที่เนื่องมาจากสิ่งที่ถูกวัด (Object of Measurement) ฟาเซ็ตของการวัดอื่น ๆ และปฏิสัมพันธ์ หรือผลร่วมระหว่างสิ่งที่ถูกวัดกับฟาเซ็ตอื่น ๆ

ดังนั้น การวิเคราะห์ความแปรปรวน ในที่นี้เป็นเพียงกระบวนการเพื่อการประมาณค่าความแปรปรวนใน G Study ซึ่งไม่ใช่เป็นการทดสอบสมมติฐานด้วย ANOVA เหมือนที่ใช้กันทั่วไปในทางสถิติ สำหรับความแปรปรวนที่ประมาณได้ใน G Study (Estimated G-Study variance component) เป็นการประมาณความแปรปรวนที่แท้จริง (Actual variance or parameter) กับเงื่อนไข (ข้อสอบและผู้ตรวจ) ในเอกภพของการสังเกตที่ยอมรับได้และกับบุคคลเดียว (single person) ในประชากรตัวอย่าง เช่น $\hat{\sigma}^2_{(p)}$ เป็นค่าประมาณของ $\sigma^2_{(p)}$ ซึ่งในการตีความ จะพิจารณานักเรียนแต่ละคนในประชากรนักเรียน โดยนักเรียนแต่ละคนจะมีคะแนนที่ได้จากการทดสอบทุกข้อ N_i ซึ่งถูกตรวจโดยผู้ตรวจ N_r คน ในเอกภพของการสังเกตที่ยอมรับได้ ดังนั้นนักเรียนแต่ละคนจะได้คะแนนเฉลี่ย (μ_p) และความแปรปรวนของคะแนนเฉลี่ยดังกล่าวของนักเรียนคือ $\sigma^2(\mu_p)$ ซึ่งก็คือ σ^2_p สำหรับความแปรปรวนของผลหลักอื่น ๆ ในฟาเซ็ตข้อกระทงและฟาเซ็ตผู้ตรวจก็ตีความในทำนองเดียวกัน

อนึ่ง estimate variance component ที่ได้ใน G Study จะมีความคงที่มาก ถ้าวิเคราะห์จากตัวอย่างเงื่อนไขการวัดขนาดใหญ่จากฟาเซ็ต ซึ่งผลที่ได้จากกการประมาณในขั้นนี้จะเป็นสารสนเทศสำหรับการตัดสินใจเกี่ยวกับสิ่งที่จะวัดใน D Study

2) การศึกษาเพื่อการตัดสินใจ

การศึกษาเพื่อการตัดสินใจ (D Study) เป็นขั้นตอนของการนำผลที่ได้จากการประมาณความแปรปรวนใน G Study เพื่อตัดสินใจเลือกวิธีการวัดที่เฉพาะเจาะจง และเหมาะสม เพื่อให้ได้ค่าประมาณความเที่ยงที่ดี ใน D Study มีขั้นตอนดังนี้

1. กำหนดเอกภพของการอ้างอิงสรุป (Universe of generalization) ที่ผู้ตัดสินใจต้องการสรุปอ้างอิง ใน D Study ซึ่งเอกภพของการอ้างอิงสรุป อาจประกอบด้วยเงื่อนไขทั้งหมดหรือเป็นเพียงเงื่อนไขย่อยของเงื่อนไขในเอกภพของการสังเกตที่ยอมรับได้ ในขั้นนี้ผู้วิจัยต้องเลือกว่า จะใช้โมเดลใด จะเป็นโมเดลสุ่ม (Random model) โมเดลกำหนด (fixed model) หรือโมเดลผสม (mixed model) ซึ่งโดยหลักการแล้ว ควรใช้โมเดลแบบสุ่มดีที่สุด เพราะจะทำให้การอ้างอิงสรุปได้กว้างขวางกว่าโมเดลแบบผสม และโมเดลแบบกำหนดตามลำดับ

2. การกำหนดขนาดตัวอย่าง (Sample size) จำนวนเงื่อนไขของตัวอย่างฟาเซ็ตใน G Study ไม่จำเป็นต้องเป็นตัวอย่างใน D Study ทั้งนี้การกำหนดขนาดตัวอย่างขึ้นอยู่กับความสนใจของผู้วิจัย

3. โครงสร้างของแบบที่ศึกษา (Design structure) ใน D Study อาจใช้โครงสร้างของแบบที่ศึกษาเหมือนหรือต่างจากใน G Study ก็ได้ อาทิเช่น ถ้าใน D Study ตัดสินใจให้นักเรียนทุกคนทำข้อสอบ n_i ข้อ เหมือนกัน จะมีโครงสร้างเป็น D Study ในลักษณะ

P x I x R design เหมือนโครงสร้างใน G Study ที่กล่าวมาแล้ว แต่ถ้าตัดสินใจให้นักเรียนทุกคนทำข้อสอบทุกข้อ โดยผู้ตรวจแต่ละคนทำการตรวจชุดของข้อสอบต่างกัน ในกรณีนี้จะได้โครงสร้างเป็น P x (I : R)

4. การประมาณค่าความแปรปรวน (Estimated D Study variance component) นั้น มีขั้นตอนดังนี้

4.1 ระบุ Design และขนาดตัวอย่าง

4.2 คำนวณหา universe score variance และ error score variance แล้วคำนวณค่าสัมประสิทธิ์การอ้างอิงสรุป

จะเห็นได้ว่า ทฤษฎีการอ้างอิงสรุปมีความแตกต่างจากทฤษฎีแบบดั้งเดิม ในหลาย ๆ ด้าน ด้านที่จะน่าจะแตกต่างกันมากก็คือมุมมอง โดยเฉพาะอย่างยิ่งภายใต้ทฤษฎีการอ้างอิงสรุปนั้น ความเที่ยงได้รับการพิจารณาภายในปริบทของสถานการณ์การทดสอบ การวัดผลได้รับการพิจารณาภายในปริบทเฉพาะของสถานการณ์การสอบ ในขณะที่ทฤษฎีการวัดแบบดั้งเดิมนั้น ความคลาดเคลื่อนของการวัด ได้อธิบายไว้ภายใต้แนวคิดเกี่ยวกับความคลาดเคลื่อนเชิงสุ่ม โดยปราศจากปริบทเฉพาะใด ๆ

ทฤษฎีการอ้างอิงสรุปมีจุดเด่นกว่าทฤษฎีการวัดแบบดั้งเดิมอย่างน้อย 2 ประการ ดังนี้ ประการแรกก็คือ ใน G-Theory นั้นได้บ่งชี้ว่า ข้อตกลงที่เข้มงวดเกี่ยวกับความเป็นแบบสอบคู่ขนานนั้น (restrictive parallel tests assumptions) ได้รับการยืดหยุ่นมากขึ้น ข้อตกลงเบื้องต้นที่อ่อนลงกว่า ซึ่งมาแทนที่นี้ก็คือ ข้อตกลงของความเป็นแบบสอบคู่ขนานโดยการสุ่ม (randomly parallel test assumption) (Suen, 1990: 41) ด้วยข้อตกลงนี้เองทำให้สามารถยอมรับได้ว่า ข้อกระทงในแบบสอบ 2 ฉบับ เป็นกลุ่มตัวอย่างสุ่มที่มาจากกลุ่มเดียวกันของข้อกระทงที่เป็นไปได้ทั้งหมด ความเด่นประการที่ 2 ของทฤษฎีการอ้างอิงสรุปก็คือ ความสามารถที่จะประมาณความคลาดเคลื่อนการวัดจากหลาย ๆ แหล่ง ทั้งนี้ได้ใช้เทคนิคการวิเคราะห์ความแปรปรวนในการประเมินความคลาดเคลื่อนการวัด ในขณะที่ในทฤษฎีดั้งเดิมนั้น จะมีค่าสัมประสิทธิ์ความเที่ยงเพียงค่าเดียว แต่การใช้การวิเคราะห์ความแปรปรวน (ANOVA) ในทฤษฎีการอ้างอิงสรุปสามารถทำให้เกิดสัมประสิทธิ์ความเที่ยงได้มากมาย เป็นต้นว่า การใช้ข้อมูลการวัดผลกลุ่มเดียวกันในทั้ง 2 ทฤษฎีนั้น ทฤษฎีการวัดแบบดั้งเดิมจะไม่สามารถก่อกำเนิดสัมประสิทธิ์ความเที่ยงมากกว่า 1 ค่าได้ กล่าวคือจะให้สัมประสิทธิ์ความเที่ยงได้เพียงค่าเดียวเท่านั้น แต่การอ้างอิงสรุปในทฤษฎีการอ้างอิงสรุปนั้น สามารถนำไปสู่สัมประสิทธิ์ความเที่ยงได้หลาย ๆ ค่า ตามจำนวนคำถามความเที่ยงที่ตั้งขึ้นไว้ นอกจากนี้ทฤษฎีการอ้างอิงสรุปยังให้สารสนเทศเกี่ยวกับจำนวนและแหล่งของความคลาดเคลื่อนในการวัดได้เด่นชัดด้วย

ตอนที่ 2 การพัฒนาแนวคิดเกี่ยวกับการวิเคราะห์ความตรงสู่เข้าตามทฤษฎีการอ้างอิงสรุป

แม้ว่าจะมีแนวคิดที่ปรากฏที่ระบุถึงระหว่างความแตกต่างของความเที่ยงและความตรงก็ตาม แต่เมื่อมีการใช้หลักฐานทางสถิติทั่วไป เพื่อใช้เป็นสื่อในการสนับสนุนแนวคิดของความเที่ยงและความตรงนั้น ก็ยังพบว่า ความแตกต่างระหว่างความเที่ยงและความตรง กลับดูเหมือนกับไม่ชัดเจนนัก มีลักษณที่คลุมเครือ สถานการณ์ตัวอย่างที่เสนอให้พิจารณา เช่น ในสถานการณ์ที่มีการใช้แบบสอบ 2 ฉบับ คือแบบสอบฉบับใหม่และแบบสอบฉบับเดิม คะแนนของการสอบทั้ง 2 ครั้งได้นำมาใช้ และได้ประมาณค่าสหสัมพันธ์ระหว่างกลุ่มของคะแนนทั้ง 2 กลุ่มนี้ โดยมีจุดมุ่งหมายที่ต้องการตรวจสอบว่า แบบสอบฉบับใหม่มีความเหมาะสม พอที่จะไปแทนที่ฉบับเดิมได้หรือไม่ ซึ่งในลักษณะนี้ จะพิจารณาค่าสหสัมพันธ์ ในฐานะที่เป็นหลักฐานของความตรงเชิงเกณฑ์สัมพันธ์ (criterion related validity) โดยที่แบบสอบฉบับเดิมใช้เป็นเกณฑ์ (criterion) แต่อย่างไรก็ตาม ก็ยังมีทางเลือกอีกทางหนึ่ง โดยได้จากการพิจารณาสถานการณ์ที่ว่า แบบสอบใหม่ที่ประกอบด้วยข้อกระทงที่มาจากเอกภพเดียวกันและแบบสอบทั้ง 2 ฉบับนั้น เป็นแบบสอบคู่ขนาน สถานการณ์ในลักษณะนี้จะพบว่า ค่าสหสัมพันธ์ที่ได้ จะเป็นตัวชี้ในลักษณะของความเที่ยงมากกว่าความตรง ในขณะที่สถานการณ์แรกเป็นการบ่งชี้ความตรง ดังนั้นความแตกต่างระหว่างความเที่ยงและความตรงจึงไม่ได้แยกจากกันอย่างชัดแจ้งเสมอไป (Suen, 1990: 150-151)

เมื่อพิจารณาจากแนวคิดของความเที่ยงตามทฤษฎีการวัดแบบดั้งเดิม ซึ่ง Lord และ Novick (1968) ได้ชี้ว่า ความเที่ยงของคะแนนสอบ (reliability of a test score) เป็นเพียงความตรงที่เกี่ยวกับแบบสอบคู่ขนาน ทำนองเดียวกับที่ Campbell and Fiske (1959) ได้เสนอว่า ความเที่ยงเป็นสหสัมพันธ์ระหว่างการวัด 2 อย่างที่คล้ายคลึงกันที่สุด และความตรงเป็นสหสัมพันธ์ระหว่างการวัด 2 อย่างที่มีความแตกต่างกันมากที่สุด นั่นก็คือ ความเที่ยงและความตรงมีจุดแตกต่างกันง่าย ๆ เกี่ยวกับความต่อเนื่องของสหสัมพันธ์ระหว่างการวัด

โมเดลเชิงสุ่มสำหรับความตรง (A Sampling Model for Validity)

Kane (1982) ได้เสนอโมเดลเชิงสุ่มสำหรับความตรง โดยใช้แนวคิดจากทฤษฎีการอ้างอิงสรุปเพื่อมุ่งไปสู่การแปลความของความเที่ยงและความตรง ซึ่งเน้นประเด็นทางด้านแนวคิด โดยเสนอผลในลักษณะขององค์ประกอบความแปรปรวน อันมีสาระสรุปได้ดังนี้

เมื่อกล่าวถึงสิ่งที่ถูกวัด (Object of Measurement) นั้น ค่าของคุณลักษณะของสิ่งที่ถูกวัด (value of an attribute for an object of measurement) เป็นค่าคาดหวังที่

ได้จากการสังเกตทั้งหมดในเอกภพ เพื่อบ่งชี้คุณลักษณะของสิ่งที่ถูกวัด ในลักษณะของคะแนนเอกภพ ถ้าคะแนนเอกภพเป็นสิ่งที่หาได้จริง จะทำให้การวัดคุณลักษณะนั้นมีความแม่นยำอย่างสมบูรณ์ แต่อย่างไรก็ตาม จะพบว่า ในทางปฏิบัตินั้น คะแนนเอกภพเป็นคะแนนที่ไม่สามารถหาได้ ดังนั้นจึงต้องใช้ข้อมูลจากตัวอย่างที่ได้จากการสังเกตเพื่อการประมาณค่าคุณลักษณะนั้นแทน ซึ่งประเด็นนี้นำไปสู่นิยามของความตรงของคุณลักษณะที่ว่า

"กระบวนการวัดที่กล่าวได้ว่ามีความตรงสำหรับคุณลักษณะนั้น ก็ต่อเมื่อให้ค่าประมาณของค่าคาดหวังที่ครอบคลุมเอกภพของการสังเกตในคุณลักษณะได้อย่างแม่นยำ"

ความตรงจะสะท้อนความแม่นยำของการอ้างอิงจากคะแนนสังเกตไปสู่ค่าของคุณลักษณะ โดยที่ค่าความแม่นยำนี้ พิจารณาได้จาก ค่าคาดหวังของความคลาดเคลื่อนกำลังสอง ในการประมาณค่า (The expected squared error in estimation) ความตรงนี้เป็นสิ่งที่บ่งชี้ในลักษณะของระดับของความมากน้อย มากกว่าที่จะบ่งว่า "มี" หรือ "ไม่มี" (all or none) และยังขึ้นอยู่กับรูปแบบของกระบวนการวัดและการแปลความของคุณลักษณะด้วย

โมเดลเชิงสุ่มที่ใช้เป็นฐานในการศึกษานิยามของความตรงนี้ เป็นโมเดลแบบหลายฟาเซ็ต (multifacet) ในลักษณะที่ว่า เอกภพได้กำหนดคุณลักษณะที่อาจเกี่ยวข้องกับการสังเกต ซึ่งแปรเปลี่ยนไปตามจำนวนมิติหรือจำนวนฟาเซ็ตทั้งหลาย เป็นโมเดลเชิงสุ่มหลายฟาเซ็ตที่เกี่ยวข้องกับกระบวนการวัด โมเดลเชิงสุ่มนั้น เป็นเสมือนกับแบบของความตรงเชิงโครงสร้าง ในขณะที่ความตรงเชิงเกณฑ์สัมพันธ์ไม่นับรวมอยู่ในโมเดลนี้ ทั้งนี้เนื่องจากความตรงเชิงเกณฑ์สัมพันธ์จะเข้ามาภายหลังจากที่แสวงหาเกณฑ์ที่มีความตรงแล้ว โดยโมเดลเชิงสุ่มจะเสนอกลไกในการพัฒนาเกณฑ์ของคุณลักษณะที่ศึกษา

จากการที่โมเดลเชิงสุ่มมีรากฐานมาจากทฤษฎีการอ้างอิงสรุป โดยค่าที่แท้จริงของคุณลักษณะก็คือ คะแนนเอกภพ ทั้งนี้กระบวนการวัดที่บอกความตรงของคุณลักษณะ ถือว่าคะแนนสังเกตมาจากกระบวนการที่ประมาณค่าของคะแนนเอกภพ หลักเบื้องต้นของโมเดลถือว่า การวัดเกี่ยวข้องกับการอ้างอิงจากคะแนนสังเกต ซึ่งมาจากตัวอย่างของการสังเกตไปสู่ค่าเฉลี่ยของเอกภพที่ตัวอย่างนั้นอยู่ ค่าความตรงสามารถประมาณค่าโดยใช้ตัวอย่างจากการสังเกต นั่นคือการสุ่มตัวอย่างอย่างง่ายจากเอกภพของการอ้างอิง เพื่อประมาณค่าสัมประสิทธิ์การอ้างอิงสรุป และสะท้อนถึงความเป็นอิสระของการอ้างอิงจากคะแนนสังเกตไปสู่คะแนนเอกภพ

ทฤษฎีการอ้างอิงสรุป ยอมให้มีการใช้โมเดลเชิงเส้นตรงที่หลากหลายในการออกแบบและการตีความในการศึกษาเพื่อการอ้างอิงสรุป หรือการศึกษา G (G Study) และการศึกษาเพื่อการตัดสินใจ หรือการศึกษา D (D Study) ซึ่งเอกภพของการอ้างอิงสรุป เกี่ยวข้องกับจำนวนฟาเซ็ต และโมเดลเชิงเส้นตรงสำหรับคะแนนสังเกต สามารถแสดงถึงฟาเซ็ตที่ศึกษาได้ สำหรับตัวอย่างการนำเสนอในที่นี้ เป็นการนำเสนอในลักษณะของโมเดลจำนวนหนึ่งฟาเซ็ตที่มี

การซ้ำ เพื่อใช้เป็นพื้นฐานในการวิพากษ์ เป็นโมเดลอย่างง่ายที่มีเพียงหนึ่งฟาเซ็ต ที่ต้องการพิจารณา ในขณะที่ฟาเซ็ตอื่น ๆ ถือว่าเป็นตัวอย่างสุ่มและเป็นอิสระต่อกันและอยู่ภายใต้ฟาเซ็ตซ้ำ ดังรายละเอียดของโมเดลที่แสดงคะแนนสังเกตในลักษณะโมเดลเชิงเส้นตรง ดังนี้

$$X_{oir} = \mu + \alpha_o + \alpha_i + \alpha_{oi} + \alpha_r \quad \text{.....(1)}$$

(Kane, 1982: 136)

โดยที่

μ เป็น ค่าเฉลี่ยรวม (grand mean)

α_o เป็น ผลหลักของสิ่งที่ต้องการวัด (main effect for the object of the measurement, o)

α_i เป็น ผลหลักของฟาเซ็ต i (main effect for the i facet)

α_{oi} เป็น ผลร่วมระหว่างฟาเซ็ต i และฟาเซ็ต o (oi interaction)

α_r เป็น ผลที่เกิดจากการซ้ำ (replication effect)

โมเดลเชิงเส้นตรงที่เสนอในสมการนี้ แสดงถึงคะแนนสังเกตในเอกภพของการอ้างอิงโดยไม่ได้มุ่งเสนอที่จะเป็นตัวแทนของรูปแบบการสุ่ม สำหรับการศึกษาเพื่อการอ้างอิงสรุป (การศึกษา G) หรือการศึกษาเพื่อการตัดสินใจ (การศึกษา D) ในสมการนี้คะแนนสังเกตจะประกอบด้วยแหล่งความแปรปรวน 4 แหล่ง และสามารถที่จะขยายไปสู่ฟาเซ็ตอื่นๆ เพิ่มขึ้นมาอีกได้ และถ้าเทียบกับโมเดลแบบดั้งเดิม (Classical Model)แล้ว ในคะแนนสังเกตจะมีแหล่งความแปรปรวนเพียง 2 แหล่งเท่านั้น

โมเดลในสมการที่ 1 ประกอบด้วย 2 ฟาเซ็ต คือฟาเซ็ต i และฟาเซ็ต r และภายในฟาเซ็ตเหล่านี้ จะประกอบด้วยเอกภพของเงื่อนไข (universe of condition) ซึ่งจะเป็นเงื่อนไขที่มุ่งศึกษาเป็นการเฉพาะ ที่ต้องการนำมาพิจารณาต่อไป สำหรับการศึกษา G ซึ่งฟาเซ็ต i ได้ crossed กับสิ่งที่ถูกวัด และการซ้ำได้ nested อยู่ในผลรวมของ oi ดังนั้นจะมี 4 องค์ประกอบของความแปรปรวนที่สามารถประมาณค่าได้อย่างเป็นอิสระกัน ประกอบด้วย $\sigma^2(o)$, $\sigma^2(i)$, $\sigma^2(oi)$ และ $\sigma^2(r)$ และถ้าเป็นการศึกษาเพื่อการตัดสินใจ (การศึกษา D) แล้ว องค์ประกอบของความแปรปรวนสำหรับผลที่เป็นการสุ่มนั้นจากสมการที่ 1 จะเป็นดังนี้

$$\sigma^2_{(I)} = \sigma^2_{(i)}/n_i \quad \text{.....2(a)}$$

$$\sigma^2_{(oI)} = \sigma^2_{(oi)}/n_i \quad \text{.....2(b)}$$

$$\sigma^2_{(R)} = \sigma^2_{(r)}/n_i n_r \quad \text{.....2(c)}$$

โดยที่ n_i คือ เงื่อนไขของฟาเซ็ต i

n_r คือ จำนวนการซ้ำ

การวัดที่มีพื้นฐานมาจากการสุ่มจากเอกภพของการอ้างอิง

(Measurement Based on Random Sampling from the Universe of Generalization)

ในการศึกษาเพื่อการตัดสินใจ D-Study [ซึ่ง i nested ใน o นั้น (i nested within o)] คะแนนสังเกตมีดังนี้

$$X_{oIR} = \mu + \alpha_o + \alpha_I + \alpha_{oI} + \alpha_R \quad \text{.....(3)}$$

โดยที่ o แทน สิ่งที่ถูกวัด

I แทน จำนวนตัวอย่างของเงื่อนไขในฟาเซ็ต i ซึ่งมีจำนวน n_i

R แทน จำนวนตัวอย่างของการซ้ำสำหรับแต่ละเงื่อนไขของฟาเซ็ต i

เนื่องจากผลในสมการที่ 3 ไม่มีความสัมพันธ์กัน ดังนั้น ค่าคาดหวังความแปรปรวนของคะแนนสังเกตที่ครอบคลุมประชากรและเอกภพของการอ้างอิง คือ

$$\sigma^2(X) = \sigma^2(o) + \sigma^2(I) + \sigma^2(oI) + \sigma^2(R) \quad \text{.....(4)}$$

คะแนนเอกภพของสิ่งที่ถูกวัด o (universe score for the object of measurement o) คือ

$$\mu_{o} = \underset{IR}{EE}\,(X_{oIR}) = \mu + \alpha_{o} \qquad \text{.....(5)}$$

และความแปรปรวนของคะแนนเอกภพ (universe score variance) มีค่าดังสมการ

$$E\,(\mu_{o} - \mu)^{2}_{o} = \sigma^{2}(o) \qquad \text{.....(6)}$$

เมื่อการสังเกตเป็นการสุ่มอย่างง่ายจากเอกภพของการอ้างอิง สำหรับสิ่งที่ถูกวัดแล้ว ค่าคาดหวังของคะแนนสังเกตที่ครอบคลุมการวัดซ้ำ มีค่าเท่ากับคะแนนเอกภพ และคะแนนสังเกตเป็นค่าประมาณที่ไม่ลำเอียงของคะแนนเอกภพ

สำหรับการวิเคราะห์ความคลาดเคลื่อนของการวัดนั้น ครอนบัคและคณะ (Cronbach et. al, 1972 :76) ได้บ่งชี้ความแตกต่างระหว่างความคลาดเคลื่อน Δ และความคลาดเคลื่อน δ โดยที่ Δ เป็นความคลาดเคลื่อน ในการประมาณค่าจุดของคะแนนเอกภพ (error in point estimates of universe scores) และ δ เป็นความคลาดเคลื่อนในการประมาณค่าของคะแนนเอกภพ ที่แสดงโดยการเบี่ยงเบนจากค่าเฉลี่ยรวม μ (grand mean μ)

ความคลาดเคลื่อนของการวัด สำหรับการประมาณค่าจุดของ μ_{o} ที่ขึ้นอยู่กับ X_{oIR} คือ

$$\Delta_{o:R} = X_{oIR} - \mu_{o}$$

$$= \alpha_{I} + \alpha_{oI} + \alpha_{R} \qquad \text{.....(7)}$$

เนื่องจาก I และ R เป็นตัวอย่างสุ่มของการวัด ค่าคาดหวังของ Δ_{oIR} ที่ครอบคลุมการวัดซ้ำ มีค่าเป็นศูนย์ และ คะแนนสังเกตเป็นค่าประมาณของคะแนนเอกภพ o ที่ไม่ลำเอียง ดังนั้น ค่าคาดหวังของความคลาดเคลื่อนกำลังสองในการประมาณค่าจุด ตลอด I และ R (expected value of the squared error in point estimate over I and R) เป็นดังนี้ (Kane, 1982: 137)

$$\underset{IR}{EE}\,(\Delta^{2}_{oIR}) = \sigma^{2}(I) + \sigma^{2}(oI) + \sigma^{2}(R) \qquad \text{.....(8)}$$

ซึ่งเป็นตัวแทนของความแปรปรวนของความคลาดเคลื่อน สำหรับการประมาณค่าจุดของคะแนนเอกภพ

ถ้าเงื่อนไขของฟาเซ็ต i เป็นตัวอย่างที่เป็นอิสระจากแต่ละการสังเกตแล้ว ค่าคาดหวังของ X_{oiR} ที่ครอบคลุมประชากร จะมีค่าเท่ากับค่าเฉลี่ยรวม μ และค่าความคาดเคลื่อนในการประมาณคะแนนเบี่ยงเบนเอกภพ (error in estimating universe deviation scores) ก็คือ

$$\delta_{oiR} = (X_{oiR} - \mu) - (\mu_o - \mu)$$

$$\delta_{oiR} = \alpha_i + \alpha_{oi} + \alpha_R \qquad \text{.....(9)}$$

เนื่องจาก I และ R เป็นตัวอย่างที่เป็นอิสระสำหรับการสังเกตสิ่งที่ถูกวัด ค่าคาดหวังของ δ_{oiR} ตลอดเอกภพ จึงมีค่าเป็นศูนย์ด้วย และคะแนนเบี่ยงเบนของการสังเกต เป็นค่าประมาณที่ไม่ลำเอียงของคะแนนเบี่ยงเบนเอกภพ (universe deviation score) ความแปรปรวนใน δ_{oiR} จึงเท่ากับความแปรปรวนใน Δ_{oiR} ดังที่แสดงไว้ในสมการที่ 8

ดังนั้น ความแปรปรวนร่วมที่ครอบคลุมประชากรของความคลาดเคลื่อน Δ_{oiR} ในการบริหารการวัด 2 ครั้ง เป็นดังนี้

$$COV(\Delta_{oiR}, \Delta_{oiR}) = E_o(\alpha_i + \alpha_{oi} + \alpha_R)(\alpha_i + \alpha_{oi} + \alpha_R) \qquad \text{.....(10)}$$

เนื่องจากฟาเซ็ต i และฟาเซ็ต r เป็นตัวอย่างที่เป็นอิสระจากการสังเกต ดังนั้นค่าคาดหวังที่ครอบคลุมผลคูณในสมการที่ 10 จึงมีค่าเป็น 0 และความคลาดเคลื่อน Δ_{oiR} นั้นไม่มีความสัมพันธ์กันในลักษณะที่คล้ายคลึงกัน ค่าความแปรปรวนร่วมของความคลาดเคลื่อน δ_{oiR} สำหรับการบริหารกระบวนการวัด 2 ครั้ง จึงมีค่าเป็น 0 ด้วย

ในทฤษฎีการวัดแบบดั้งเดิมนั้น ความคลาดเคลื่อนของการวัด มีค่าคาดหวังเป็นศูนย์และไม่มีความสัมพันธ์ระหว่างคะแนนสังเกตแต่ละคู่ ซึ่งความคลาดเคลื่อนของการวัดในลักษณะนี้จะเรียกว่า ความคลาดเคลื่อนแบบสุ่ม (random error) ดังนั้นในกระบวนการวัดที่ขึ้นอยู่กับการสุ่มตัวอย่างที่เป็นอิสระต่อกันจากเอกภพของการอ้างอิงแล้ว ความคลาดเคลื่อนของการวัดทั้งหมดเป็นความคลาดเคลื่อนแบบสุ่ม

ครอนบัคและคณะ (Cronbach et. al,1972:) ได้นิยามสัมประสิทธิ์การอ้างอิงสรุปไว้ว่า เป็นอัตราส่วนระหว่าง ความแปรปรวนของคะแนนเอกภพ กับ ความแปรปรวนของคะแนนสังเกตที่คาดหวัง (ratio of universe score variance to expected observed score variance) ดังนั้นจากสมการที่ 4 และ 6 จะได้ค่าสัมประสิทธิ์การอ้างอิงสรุป ดังนี้

$$E\rho^2 \ (X_{oIR} \ , \ \mu_o) = \frac{\sigma^2(O)}{\sigma^2(o) + \sigma^2(oI) + \sigma^2(I) + \sigma^2(R)} \quad .....(11)$$

โดยที่สัญลักษณ์ $E\rho^2$ เน้นถึงการตีความของค่าสัมประสิทธิ์ ในฐานะดัชนีของกำลังสองของค่าสหสัมพันธ์ระหว่างคะแนนสังเกตและคะแนนเอกภพ (squared correlation between observed scores and universe scores) โดยค่าสัมประสิทธิ์การอ้างอิงสรุปในสมการที่ 11 เกี่ยวข้องกับการทดลองกฎความไม่แปรเปลี่ยน 2 กฎ โดยกฎหนึ่งเป็นของฟาเซ็ต i และอีกกฎหนึ่งเป็นของฟาเซ็ตที่เหลือ

ค่าของสัมประสิทธิ์การอ้างอิงสรุป ขึ้นอยู่กับกระบวนการวัดที่เป็นตัวอย่างของเอกภพการอ้างอิง ซึ่งพิจารณาได้จากรูปแบบของกระบวนการวัด และจากนิยามของคุณลักษณะที่ต้องการวัด หากเอกภพของการอ้างอิงได้นิยามไว้ไม่กว้างมากแล้ว ก็จะยิ่งทำให้เกิดความเชื่อถือได้ของการวัดเพิ่มมากขึ้น

ดังนั้น ถ้าใช้สมการที่ 2 ร่วมกับสมการที่ 11 แล้ว ค่าสัมประสิทธิ์การอ้างอิงสรุปสามารถได้รับการประมาณสำหรับจำนวนใด ๆ ของเงื่อนไขของฟาเซ็ต i และจำนวนใด ๆ ของการซ้ำ การเพิ่มขนาดตัวอย่างสำหรับฟาเซ็ตต่าง ๆ จะช่วยทำให้เกิดแนวทางในการพัฒนาความเชื่อถือได้ของการวัด แต่ในทางปฏิบัติจะมีข้อจำกัดในเรื่องขนาดของตัวอย่าง

กระบวนการวัดจะมีความตรง ก็ต่อเมื่อประมาณค่าคะแนนเอกภพได้แม่นยำ สำหรับกระบวนการวัดที่ประกอบด้วยการสุ่มอย่างง่ายจากเอกภพของการอ้างอิงนั้น คะแนนสังเกตเป็นค่าประมาณที่ไม่ลำเอียงของคะแนนเอกภพ และความคลาดเคลื่อนเชิงสุ่มก็เป็นความคลาดเคลื่อนของการวัดเท่านั้น แต่เนื่องจากค่าสัมประสิทธิ์การอ้างอิงสรุปในสมการที่ 11 แสดงให้เห็นว่าคะแนนเอกภพจะสามารถอ้างอิงจากคะแนนสังเกตได้อย่างไร ดังนั้น จึงสามารถที่จะแปลความของสัมประสิทธิ์การอ้างอิงสรุปดังกล่าว ในลักษณะสัมประสิทธิ์ความตรง (validity coefficient) ได้

ถึงแม้ว่าสัมประสิทธิ์การอ้างอิงสรุป สามารถใช้เป็นดัชนีของความตรง (index of validity) ได้ แต่ก็มีค่าประมาณของสัมประสิทธิ์การอ้างอิงสรุปเป็นจำนวนมาก ที่ไม่ใช่สัมประสิทธิ์ความตรง ในกรณีของสมการที่ 11 ที่เป็นสมการที่แสดงถึงสัมประสิทธิ์ความตรง ก็ขึ้นอยู่กับข้อตกลงเบื้องต้นของการสุ่มที่แกร่งว่า คะแนนสังเกตต้องมาจากตัวอย่างที่ได้จากการสุ่มในเอกภพของการอ้างอิงเท่านั้น

ความตรงใน G-Theory มีลักษณะคล้ายกับความตรงเชิงเกณฑ์ในทฤษฎีการวัดดั้งเดิม โดยที่คะแนนเอกภพทำหน้าที่เหมือนเป็นเกณฑ์ แต่สิ่งที่ไม่เหมือนกันก็คือ คะแนนเอกภพเป็นสิ่งที่เป็นนามธรรม ไม่สามารถสังเกตได้โดยตรง จึงเป็นไปไม่ได้ที่จะประมาณความตรง โดยใช้สหสัมพันธ์ระหว่างคะแนนสังเกตและคะแนนเอกภพ

ถึงแม้ว่าคะแนนเอกภพเป็นสิ่งที่เป็นนามธรรม ซึ่งไม่สามารถสังเกตได้โดยตรงก็ตาม แต่ก็ยังสามารถขยายต่อไปได้ว่า ความแม่นยำที่ได้รับจากการประมาณค่าคะแนนเอกภพ สามารถนำมาใช้ประมาณค่าสัมประสิทธิ์การอ้างอิงสรุป ดังนั้นความตรงของการวัดสามารถนำเสนอได้ในลักษณะของสัมประสิทธิ์การอ้างอิงสรุป

โมเดลเชิงสุ่มของความตรงเป็นแนวคิดที่มีประโยชน์ ที่ได้มาจากการขยายกรอบแนวคิดของลักษณะวิธีหลากวิธีหลาย (multitrait-multimethod paradigm) และความเชื่อมโยงของความเที่ยงกับความตรง (reliability validity continuum)

ถ้าคะแนนการสอบไม่แปรเปลี่ยนระหว่างวิธีการวัดที่แตกต่างกันแล้ว คะแนนการสอบจากกลุ่มเงื่อนไขการวัดหนึ่งก็จะสามารถที่จะใช้สนับสนุนในการอ้างอิง ภายใต้กลุ่มเงื่อนไขการวัดอื่น ๆ (Messick, 1989 cited by Suen, 1990) นั่นก็คือ คะแนนสอบนั้นไม่เพียงแต่ จะเป็นตัวบ่งชี้ความเชื่อถือได้ของคะแนนจริงภายในกลุ่มหนึ่งของเงื่อนไขการวัดแล้ว แต่ยังเป็นตัวบ่งชี้ของคะแนนเอกภพระหว่างเงื่อนไขการวัดที่แตกต่างออกไปด้วย ซึ่งเทียบเคียงได้กับความคงเส้นคงวาระหว่างวิธีวัดที่แตกต่างกัน นั่นก็คือ การอ้างอิงสรุประหว่างวิธีของการวัดที่แตกต่างกันจะเป็นตัวบ่งชี้ของความตรงของคะแนนสอบ

การอ้างอิงสรุปภายในแต่ละวิธีวัด ซึ่งใช้เป็นตัวชี้ว่าคะแนนจริงนั้นจะอยู่ภายใต้เงื่อนไขการวัดที่เฉพาะเจาะจงนั้น จะสามารถใช้คะแนนสังเกตเป็นตัวแบบได้ ลักษณะนี้นี้ก็เทียบเคียงได้กับความคงเส้นคงวา (stability) ระหว่างวิธีการวัดที่คล้ายคลึงกัน นั่นก็คือจะบ่งชี้ความเที่ยงของแต่ละวิธีวัด

I 16998431

เมื่อใช้วิธีวัดเป็นฟาเซตหนึ่งของการศึกษา นั้น จะมีสถิติที่ใช้ในการศึกษาสัมประสิทธิ์ความตรง (Kane, 1982 cited by Suen, 1990) ดังต่อไปนี้

$$E\rho^2 = \frac{\sigma^2_\tau}{\sigma^2_m + \sigma^2_e}$$

เมื่อ σ^2_τ คือ ความแปรปรวนของคะแนนแท้จริง (ค่าที่แท้จริง)

σ^2_m คือ ความแปรปรวนของ facet ของวิธีวัด หรือกล่าวได้ว่าเป็นผลหลักขององค์ประกอบความแปรปรวน ที่เนื่องมาจากฟาเซตของวิธีวัด

σ^2_e คือ ความแปรปรวนที่มาจากผลรวมขององค์ประกอบความแปรปรวนอื่น ๆ

$E\rho^2$ คือ สัมประสิทธิ์การอ้างอิงสรุป ที่เรียกว่า สัมประสิทธิ์ความตรง (validity coefficient) ซึ่งได้จากการใช้ทฤษฎีการอ้างอิงสรุป กับวิธีการวัดหลายวิธีในเนื้อหาเดียวกัน ซึ่งเป็นการชี้ให้เห็นความตรงในการวัด

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

3.1 ต่างประเทศ

ในปี ค.ศ.1972 หน่วยงานของสหรัฐอเมริกา คือ The US Department of Labour (อ้างใน Shavelson and Webb : 1981) ได้ประยุกต์ทฤษฎีการอ้างอิงสรุปในการพัฒนามาตรวัดพัฒนาการศึกษาทั่วไป (General Education Development : GED) ที่ใช้ในการประเมินความสามารถด้านเหตุผล คณิตศาสตร์ และภาษาที่จำเป็นต่องานอาชีพต่าง ๆ เพื่อประมาณเวลาที่จะใช้ในการเรียนรู้งาน การเทียบงาน และการสร้างโปรแกรมการฝึกงาน โดยใช้รูปแบบการศึกษาแบบไม่สมดุล คือให้ผู้ประเมิน (r) Nested อยู่ในศูนย์ (c) ต่าง ๆ ของหน่วยงาน และ Crossed กับงานอาชีพ (j) และจำนวนครั้งที่ประเมิน (o) เขียนในรูปสัญลักษณ์ว่า (r : c) x j x o ความสามารถแต่ละด้าน จะได้รับการประเมินโดยผู้ประเมิน 71 คน จากศูนย์อาชีพ 11 แห่ง พบว่า ความเชื่อมั่นของการประเมิน มีค่าเพิ่มขึ้นตามการเพิ่มจำนวนผู้ประเมิน ข้อค้นพบนี้แสดงให้เห็นว่า จำนวนผู้ประเมินมีผลต่อความเชื่อมั่นในการประเมิน

สมิทช์ (Smith: 1978) ได้ศึกษาถึงความคลาดเคลื่อนในการสุ่มตัวอย่าง ของการศึกษาอ้างอิงสรุปชนิดหลายองค์ประกอบ (multifacet study) ที่ใช้กลุ่มตัวอย่างจำนวนน้อย ๆ และเสนอวิธีลดความคลาดเคลื่อนเชิงสุ่มไว้ 3 วิธี คือ 1) สุ่มระดับของฟาเซ็ตแต่ละตัวในแบบจำลองการวิเคราะห์ความแปรปรวน ให้ได้จำนวนมากที่สุดเท่าที่จะทำได้ 2) สุ่มระดับของฟาเซ็ตแต่ละตัว ให้เป็นสัดส่วนกับขนาดของเอกภพของฟาเซ็ต และ 3) เปลี่ยนรูปแบบความสัมพันธ์ระหว่างฟาเซ็ต สำหรับข้อ 2) และ 3) ควรใช้เมื่อไม่สามารถสุ่มตัวอย่างงให้ได้จำนวนมากๆ

แมคเกรดี (Macready, 1983: 149-157) ได้ประยุกต์ทฤษฎีการอ้างอิงสรุป ในการประเมินความยากและความเป็นเอกพันธ์ (homogeneity) ของข้อสอบอิงโดเมน ที่ใช้ในการวินิจฉัย โดยมีฟาเซ็ตที่ศึกษาประกอบด้วย โดเมนการคูณจำนวนเต็ม (d) ห้องเรียน (c) จำนวนหลักของตัวคูณ (n) ข้อสอบซึ่งแฝงอยู่ในโดเมนและจำนวนหลัก (i : (d x n)) และนักเรียนซึ่งแฝงอยู่ในห้องเรียน (s : c) จากแบบการวิเคราะห์ G Study ที่อยู่ในลักษณะ (s : c) x (i :(d x n)) โดยให้ ห้องเรียน นักเรียน และข้อสอบเป็นฟาเซ็ตสุ่ม ให้โดเมนและจำนวนหลักของตัวคูณเป็นฟาเซ็ตคงที่ พบว่า (s:c) x (i:(dxn)),e เป็นแหล่งความแปรปรวนที่มากที่สุดคือ 47% อีก 4 แหล่งที่มีค่ารองลงไป ได้แก่ ห้องเรียน นักเรียน ซึ่งแฝงอยู่ในห้องเรียน โดเมน และผลร่วมระหว่างนักเรียนกับโดเมน (s:c) x d โดยทั้ง 4 แหล่งมีค่ารวมกันถึง 51% เฉพาะแหล่งความแปรปรวนสุดท้าย หมายถึง ความยากของข้อสอบในแต่ละโดเมน สำหรับนักเรียนแต่ละคนมีค่าไม่เท่ากัน แหล่งความแปรปรวนอื่น ๆ ที่เหลือมีค่าน้อยมาก โดยเฉพาะจำนวนหลักของตัวคูณ Macready สรุปได้ว่า ไม่ควรจะใช้ตัวคูณ 4 ตำแหน่ง ใช้เพียง 3 ตำแหน่งก็พอ เพราะให้ค่าความยากไม่แตกต่างกัน อีกแหล่งหนึ่งมีค่าน้อย คือ i:(dxn) แสดงว่าความยากของข้อสอบในทุกโดเมน ที่มีตำแหน่งตัวคูณเท่ากัน มีค่าพอ ๆ กัน เมื่อตรวจดูความยากรายโดเมน พบว่าเกือบทุกโดเมนมีข้อสอบที่มีความยากเท่าเทียมกัน มีเพียงโดเมนที่ 15 ที่ค่อนข้างแตกต่างกัน Macready เสนอว่าควรจะแยกให้เป็นโดเมนย่อย หรือนำไปรวมกับโดเมนอื่น จากการประมาณค่าสัมประสิทธิ์การอ้างอิงสรุป ของข้อสอบแต่ละโดเมน โดยใช้รูปแบบการวิเคราะห์ 2 รูปแบบคือ s x i และ s x r เมื่อ r หมายถึงการสอบซ้ำ พบว่าค่าสัมประสิทธิ์การอ้างอิงสรุปของข้อสอบข้อหนึ่ง มีค่าอยู่ระหว่าง .338 ถึง .606

อิบราฮิม (Ibrahim : 1984) ได้ประยุกต์ทฤษฎีการอ้างอิงสรุป เพื่อประมาณค่าความแปรปรวน ที่มีผลต่อการประเมินวัตถุประสงค์ทางการศึกษา (The Rating of Evaluational Goals) โดยสุ่มตัวอย่างครู 80 คน และนักศึกษา 80 คน ในประเทศซูดาน ประเมินวัตถุประสงค์ทางการศึกษา 2 ชนิด คือ วัตถุประสงค์ที่สำคัญจริง ๆ และวัตถุประสงค์ที่คาดหวัง

องค์ประกอบในการศึกษา คือ ผู้ประเมิน กลุ่มผู้ประเมิน จำนวนครั้งของการประเมิน ถิ่นที่อยู่ของผู้ประเมิน ชนิดของวัตถุประสงค์ สถานที่ทำงานของผู้ประเมิน และเพศของผู้ประเมิน พบว่าองค์ประกอบที่มีผลต่อการประเมินมากที่สุด ได้แก่ ผู้ประเมิน และกลุ่มผู้ประเมิน ส่วนสถานที่ทำงานของผู้ประเมิน มีผลเล็กน้อย ส่วนองค์ประกอบอื่นที่เหลือ ไม่มีผลต่อการประเมินเลย

โอเบิร์น (O'Brien ,1986) ใช้ทฤษฎีการอ้างอิงสรุปในการประมาณความเที่ยงของตัวแปรระดับโรงเรียน 16 ตัว ซึ่งเป็นค่าเฉลี่ยหรือร้อยละของกลุ่มตัวอย่างนักเรียน ตัวแปรแบ่งเป็น 5 กลุ่ม คือ 1) ค่าเฉลี่ยของสถานภาพทางสังคม และเศรษฐกิจของครอบครัว 2) ค่าเฉลี่ยของผลสัมฤทธิ์ทางการเรียน 3) ร้อยละของนักเรียนที่มีบิดาหรือมารดาอาศัยอยู่ด้วย 4) ร้อยละของคุณภาพของห้องสมุดและการเรียนการสอนของโรงเรียน และ 5) ร้อยละของนักเรียนที่ยอมรับกฎระเบียบของโรงเรียน กลุ่มตัวอย่าง คือโรงเรียนจำนวน 1,122 โรง และนักเรียนชั้นปีที่ 2 โรงเรียนละ 36 คน ดำเนินการวิจัย โดยให้กลุ่มตัวอย่างตอบคำถามของตัวแปรแต่ละตัว จากการศึกษาพบว่า ความเที่ยงในการตอบแบบสอบถาม เพิ่มขึ้นตามการเพิ่มจำนวนผู้ตอบคำถามจากแต่ละโรงเรียน จำนวนข้อคำถาม และจำนวนโรงเรียน ข้อค้นพบนี้แสดงว่า จำนวนผู้ประเมิน จำนวนข้อคำถามของเครื่องมือประเมิน และจำนวนสถานที่ทำงานของผู้ประเมินมีผลต่อความเที่ยงของการประเมิน

เวบบ์ และคณะ (Webb, Herman and Cabello, 1987: 130) ได้ประยุกต์ทฤษฎีการอ้างอิงสรุป ในการวัดแบบอิงโดเมน เพื่อการสอบวินิจฉัยของการวัดภาษาเรื่องสรรพนาม ได้เสนอวิธีการพัฒนาและวิเคราะห์การใช้ประโยชน์ของแบบสอบวินิจฉัยสำหรับครู โดยการเชื่อมโยงระหว่างการวัดผลแบบอิงโดเมนและทฤษฎีการอ้างอิงสรุป เพื่อหาว่าความสามารถด้านเนื้อหากลุ่มใด ควรนำเสนอในเส้นภาพคะแนน (profile) ของนักเรียน การศึกษาแบ่งเป็น 4 ขั้น คือ 1) กำหนดลักษณะเฉพาะของโดเมนและการสร้างแบบสอบ 2) เลือกกลุ่มเนื้อหาที่ควรเสนอในเส้นภาพคะแนน 3) กำหนดจำนวนข้อสอบที่จำเป็นต่อความเที่ยงของการวัด และ 4) คำนวณค่าความแม่นยำของเส้นภาพคะแนน ในขั้นที่ 2 เป็นขั้นที่เริ่มใช้ทฤษฎีการอ้างอิงสรุป รูปแบบที่ใช้วิเคราะห์ใน G Study คือ s x i และใช้คะแนนความคลาดเคลื่อนแบบสัมบูรณ์ คำนวณค่าสัมประสิทธิ์การอ้างอิงสรุป ทั้งแบบ univariate และ multivariate ฟาเซ็ตที่ใช้ศึกษาประกอบด้วยกลุ่มเนื้อหาย่อยของเรื่องสรรพนาม ดังนี้ 1) rule ได้แก่ nominative, direct object และ indirect object of preposition 2) form ได้แก่ relative และ nonrelative 3) number ได้แก่ singular และ plural 4) embeddedness ได้แก่ sentence และ paragraph แต่ละฟาเซ็ต crossed กัน แต่ข้อสอบ (i) แฝงอยู่ในฟาเซ็ตอื่น

สิ่งที่ถูกวัดคือนักเรียน ซึ่งสุ่มมาจากนักเรียนเกรด 6 จำนวน 128 คน ให้เนื้อหาเป็นฟาเซ็ตคงที่ แต่ข้อสอบเป็นฟาเซ็ตสุ่ม ผลการวิเคราะห์ความแปรปรวน พบว่ามี 3 ฟาเซ็ตที่มีผลต่อคะแนนมากที่สุดคือ form, embeddedness และ rule มีเพียงบางฟาเซ็ต มีความสัมพันธ์กับความแตกต่างระหว่างนักเรียน ฟาเซ็ตใดที่ไม่มีปฏิสัมพันธ์กับนักเรียน แสดงว่ามีผลคงที่สำหรับนักเรียนทุก ๆ คน ได้แก่ ผลหลักของ form ผลร่วมกันระหว่าง form- embeddedness และ rule-form ผลร่วมระหว่าง form- embeddedness แสดงว่าแต่ละ Form ของ pronoun ในแต่ละ embeddedness นั้น ข้อสอบมีความยากไม่เท่ากัน แต่ยังถือว่าผลดังกล่าวมีความคงเส้นคงวาในนักเรียนที่ทำการทดสอบทุกคน

ชอย (Choi, 1989: 3091-A) ได้ใช้ทฤษฎีการอ้างอิงสรุป ในการวิเคราะห์แหล่งของความแปรปรวน ที่ได้จากการประเมินพฤติกรรมของครูในระบบชั้นเรียน โดยใช้ทฤษฎีนี้ในการเปรียบเทียบการประมาณค่าองค์ประกอบความแปรปรวนของความแปรปรวนคลาดเคลื่อน ที่เนื่องมาจากผู้ประเมิน (rater) โอกาส (occasion) ครู (teacher) และ ปฏิสัมพันธ์ ผลจากการศึกษาพบว่า องค์ประกอบความแปรปรวนสัมพัทธ์ขนาดใหญ่เกิดเนื่องจากฟาเซ็ตโอกาส ในขณะที่องค์ประกอบความแปรปรวนของผู้ประเมินมีค่าน้อยมาก ในส่วนของปฏิสัมพันธ์นั้น องค์ประกอบความแปรปรวนของปฏิสัมพันธ์ระหว่างโอกาสและครู มีค่ามากกว่าปฏิสัมพันธ์ของครูและผู้ประเมิน นอกจากนี้ ความแปรปรวนที่เนื่องจากโอกาส มีขนาดใหญ่กว่าความแปรปรวนที่มาจากผู้ประเมิน ดังนั้น การเพิ่มจำนวนโอกาสจะทำให้เกิดค่าสัมประสิทธิ์การอ้างอิงสรุป มีค่าสูงกว่าสัมประสิทธิ์การอ้างอิงสรุปที่มาจากการเพิ่มจำนวนผู้ประเมิน

ซุน และคณะ (Suen, H.K., and others, 1989: 136) ได้ศึกษาการประเมินทฤษฎีการอ้างอิงสรุปของข้อมูลการสังเกตโดยตรง (Generalizability Assessment of Autocorrelated Direct Observation Data) โดยศึกษาถึงความสามารถในการประยุกต์ของ The Tiao-Tan Method and Alternative ความสามารถในการประยุกต์ได้ถูกค้นพบโดยการวิเคราะห์ด้วย The Bayesian random - effect analysis variance (ANOVA) model ซึ่งพัฒนาโดย G.C Tiao และ W. Y. Tan โดยวิธีการได้รับคำแนะนำจาก H.K. Suen และ P.S. Lee เพื่อที่จะใช้การอ้างอิงสรุป ของข้อมูล autocorrelated (the generallizability analysis of autocorrelated data)

ข้อมูลเกี่ยวกับองค์ประกอบของ autocorrelated สามารถที่จะตัดไปยังกระบวนการ Box-Jenkins procedures และข้อมูลที่เหลือ (residual) สามารถวิเคราะห์โดย ANOVA ซึ่งทำให้การประะมาณค่าความแปรปรวนไม่มีความลำเอียง

ข้อดีและข้อจำกัดของวิธีการทั้ง 2 วิธี ได้ถูกกำหนดและมุ่งเน้นที่ autoregressive integrated moving average ผลจากการประยุกต์ของวิธีการที่ใช้ข้อมูลมาก ๆ แสดงให้เห็นว่าความแปรปรวนที่มีระบบ (the systematic variance) และผลจากการประยุกต์ของวิธีการที่มีข้อมูลจำนวนมาก ๆ แสดงให้เห็นว่า autocorrelation ไม่มีผลต่อความแปรปรวนที่เป็นระบบ ระหว่างกลุ่มตัวอย่าง วิธีการของ Suen-Lee จะดีกว่าวิธีการ Tiao-Tan

ครอฟอร์ด (Crawford, 1990: 354-A) ได้ศึกษาความเที่ยงและความตรงของการประเมินครูในรัฐเท็กซัส ข้อมูลในการประเมิน ได้รวบรวมจากโรงเรียนชานเมืองแห่งหนึ่งในรัฐเท็กซัส ซึ่งมีจำนวนนักเรียนประมาณ 6,000 คน ครู 386 คน และใช้ผู้ประเมิน 21 คน โดยใช้ทฤษฎีการอ้างอิงสรุป ในการประเมินองค์ประกอบของความแปรปรวนจากตัวแปรต่อไปนี้ เวลา (time) ผู้ประเมิน (rater) และ ข้อกระทง (item)

ผลการศึกษาโดยใช้ทฤษฎีการอ้างอิงสรุป พบว่า ความแปรปรวนในคะแนนของครูมีมากกว่าแหล่งความแปรปรวนอื่น ๆ สำหรับผลการศึกษาความตรงต่าง ๆ ได้แก่ ความตรงเชิงเนื้อหา ซึ่งได้จากการตรวจสอบ วรรณกรรมของการประเมินครูและงานวิจัยของการสอน ที่มีประสิทธิผล ได้แสดงให้เห็นว่า การสอนที่มีประสิทธิผลเป็นสิ่งที่มีผลต่อการปรับปรุงคะแนนการสอบของนักเรียน พฤติกรรมของครูที่แตกต่างกันก็จะมีส่วนสำคัญในการแก้ปัญหานักเรียน สำหรับความตรงเชิงโครงสร้าง ได้ตรวจสอบจากการใช้การวิเคราะห์ความแปรปรวน โดยมีตัวแปรที่ใช้ในการเปรียบเทียบในการประเมินครู คือ วิทยาเขต ระดับ สถานภาพด้านคุณธรรม เชื้อชาติ เพศ และจำนวนปีที่มีประสบการณ์ ผลการศึกษาพบว่า วิทยาเขต เพศ และจำนวนปีที่มีประสบการณ์ มีนัยสำคัญในความแปรเปลี่ยนของคะแนนครู นอกจากนี้ได้ศึกษาความตรงเชิงเกณฑ์สัมพันธ์ โดยทำการเปรียบเทียบคะแนนของครูในกลุ่ม 5 เปอร์เซ็นต์ คือกลุ่มที่ได้รับการบ่งชี้จากครูใหญ่ว่าเป็นผู้ที่มีคุณธรรมสูงกับคะแนนของครูในกลุ่มที่เหลือ ผลการศึกษาพบว่าความตรงเชิงเกณฑ์สัมพันธ์ได้รับการสนับสนุนจากกลุ่มครูที่มีคุณธรรม มากกว่ากลุ่มครูที่มีคุณธรรมน้อยกว่า อย่างมีนัยสำคัญที่ระดับ 0.05

กูดวิน และ กูดวิน (Goodwin and Goodwin, 1991: 193-204) ได้ศึกษาและเปรียบเทียบวิธีการประเมินค่าความเที่ยงระหว่างผู้ประเมิน (interater reliability) 4 วิธี ในงานวิจัยทางด้านการศึกษาพิเศษของวัยเด็กตอนต้น โดยประกอบด้วย วิธีสหสัมพันธ์ เปรียบเทียบค่าเฉลี่ย เปอร์เซ็นต์ของความสอดคล้อง และ ทฤษฎีการอ้างอิงสรุป จากการศึกษาพบว่า ทฤษฎีการอ้างอิงสรุปเป็นวิธีที่เหมาะสมที่สุด โดยเป็นวิธีการประมาณขนาดของความแปรปรวนของคุณลักษณะ จากแหล่งความคลาดเคลื่อนหลายแหล่งภายในการศึกษาเพียงครั้งเดียว

สน็อดการ์ด (Snodgrass, 1991: 1-274) ได้ศึกษาความสัมพันธ์ระหว่างการประเมินค่าประสิทธิภาพของครูและระดับชั้นเรียน ด้วยการใช้ทฤษฎีการอ้างอิงสรุปในการวิเคราะห์ ข้อมูลในการศึกษาได้จากมหาวิทยาลัย 4 แห่ง คือ มหาวิทยาลัยเท็กซัส (The University of Texas at Austin) มหาวิทยาลัยวอชิงตัน (The University of Washington) มหาวิทยาลัยมิสซูรี (The University of Missouri at Columbia) มหาวิทยาลัยโคโลราโด (The University of Colorado)

คำถามที่ต้องการตรวจสอบก็คือ

1. มีนัยสำคัญระหว่างการประเมินครูระหว่างนักศึกษาในระดับปริญญาตรี และระดับบัณฑิตศึกษาหรือไม่
2. ถ้ามีนัยสำคัญแล้วเป็นผลเนื่องจากศาสตร์หรือไม่

การศึกษาครั้งนี้ใช้ห้องเรียนเป็นหน่วยของการวิเคราะห์ โดยประกอบด้วย 5,110 ห้องเรียน จำนวนนักศึกษา 117,290 คน โดย 3,953 ห้องเรียนเป็นของระดับปริญญาตรี และ 1,157 ห้องเรียนเป็นของระดับบัณฑิตศึกษา ผลการศึกษาพบว่า การประเมินคุณภาพของการสอนขึ้นอยู่กับระดับชั้นเรียน นอกจากนี้ความสัมพันธ์ระหว่างระดับชั้นเรียนและการประเมินผลนักศึกษาขึ้นอยู่กับศาสตร์ด้วย

เบรนเนน (Brennan, 1992: 27-34) ได้ศึกษาขอบข่ายและกระบวนการของทฤษฎีการอ้างอิงสรุป โดยนำเสนอในลักษณะของโมดุลการสอนที่เรียกว่า An NCME Instruction Module ซึ่งเป็นการเสนอเรื่องที่เกี่ยวกับประสิทธิภาพการเขียน ทั้งนี้ก็เพราะการวิเคราะห์อ้างอิงสรุปเป็นสิ่งที่มีประโยชน์ ต่อการทำให้เกิดความเข้าใจ ในความสัมพันธ์ที่สำคัญของแหล่งความแปรปรวนของความคลาดเคลื่อน อันจะนำไปใช้ประโยชน์ในการออกแบบกระบวนการวัดที่มีประสิทธิภาพ

ไนเซอร์ (Naizer, 1992: 13-20) ได้เสนอแนวคิดพื้นฐานของทฤษฎีการอ้างอิงสรุปโดยเห็นว่า ทฤษฎีนี้เป็นวิถีทางที่มีพลังในการประเมินความเที่ยง และทฤษฎีนี้มีความสัมพันธ์อย่างใกล้ชิดกับทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) เขาเสนอว่า ทฤษฎีการอ้างอิงสรุปเป็นทฤษฎีที่เกี่ยวกับความเชื่อถือได้ของการวัดพฤติกรรม ซึ่งยอมให้มีการประมาณค่าความแปรปรวนของความคลาดเคลื่อนจากหลาย ๆ แหล่งได้พร้อม ๆ กัน ในทฤษฎีนี้ได้แสดงความแตกต่างระหว่างการตัดสินใจแบบสัมบูรณ์และแบบสัมพัทธ์ และเป็นทฤษฎีที่มีกลไกในการประมาณความแปรปรวนของความคลาดเคลื่อนในขั้นการศึกษาเพื่อการตัดสินใจ (D-Studies) อันเป็นสิ่งที่ช่วยนักวิจัยในการพัฒนาการวัดที่จะลดความคลาดเคลื่อน นอกจากนั้น เขาได้เสนอตัวอย่างของ

การใช้ทฤษฎีการอ้างอิงสรุปจากการศึกษาตัวอย่าง 20 คน โดยให้ทำงานด้านปฏิบัติ จำนวน 3 ครั้ง และ ประเมินโดยผู้ประเมินจำนวน 2 คน

แบกซ์เตอร์ (Baxter, 1992: 283A) ได้ศึกษาความสามารถของการแลกเปลี่ยนในการประเมินการปฏิบัติงานทางด้านวิทยาศาสตร์ โดยศึกษาจากวิธีวัดหลาย ๆ วิธี ในการวัดผลสัมฤทธิ์ทางการเรียนในระดับประถมศึกษา ด้วยการศึกษานักเรียนที่มาจาก 2 โรงเรียน ซึ่งมีความแตกต่างกันทางด้านการสอน โดยโรงเรียนหนึ่งยึดการใช้หนังสือเป็นหลัก และอีกโรงเรียนหนึ่งใช้การสืบเสาะหาความรู้เป็นหลัก

นักเรียนทั้ง 2 โรงเรียน ได้รับการประเมินจากแบตเตอรี่แบบสอบ ซึ่งประกอบด้วยแบบสอบผลสัมฤทธิ์ทางการเรียน และ แบบวัดความถนัด ในลักษณะของแบบเลือกตอบ จากกระบวนการวัด 4 วิธี ได้แก่ การสังเกตการปฏิบัติงาน การตรวจสอบจากสมุดงาน การจำลองสถานการณ์โดยใช้คอมพิวเตอร์ และ การตอบปัญหาสั้น ๆ

ผลการศึกษาพบว่า การบ่งชี้ผลสัมฤทธิ์ทางการเรียนทางด้านวิทยาศาสตร์ว่าจะขึ้นอยู่กับวิธีวัดใดนั้นยังไม่ค่อยชัดเจน

เกา (Gao,1993: 1331-A) ได้ใช้ทฤษฎีการอ้างอิงสรุป ในการวัดภาคปฏิบัติด้านวิทยาศาสตร์ ในการประเมินความถูกต้องและความคงที่ของการประมาณค่าความแปรปรวนเชิงสุ่ม ทั้งนี้ข้อมูลที่ใช้ในการประเมินภาคปฏิบัติ ได้จากนักเรียนระดับประถมศึกษาปีที่ 6 จำนวน 600 คน ซึ่งได้จากการสุ่มใน 40 โรงเรียน สำหรับแบบสอบภาคปฏิบัติทางวิทยาศาสตร์ ประกอบด้วย 5 งานจากโดเมนเนื้อหาที่ต่างกัน ประกอบด้วยผู้ประเมิน 2 คน โดยมุ่งวิเคราะห์ข้อมูลในลักษณะการศึกษา G (generalizability studies) และการศึกษา D (decision studies) สำหรับการศึกษา G นั้น มุ่งประเมินความแปรปรวนเชิงสุ่ม และการอ้างอิงสรุปของแบบสอบในระดับบุคคลและระดับโรงเรียน ในขณะที่การศึกษา D มุ่งหาจำนวนงาน ที่ทำให้ค่าสัมประสิทธิ์การอ้างอิงสรุปมีค่า .80

ผลจากการศึกษา ได้ชี้ให้เห็นว่า ความแปรปรวนเชิงสุ่มของงาน เป็นแหล่งความคลาดเคลื่อนการวัดที่มีค่ามาก ในขณะที่ความแปรปรวนเชิงสุ่มของผู้ประเมินมีค่าน้อยมาก และเมื่อจำนวนงานเพิ่มขึ้น ก็จะทำให้การอ้างอิงสรุปของการวัดเพิ่มมากขึ้น นอกจากนั้นความแปรปรวนของการปฏิบัติ ระหว่างนักเรียนภายในโรงเรียน ยังมีค่ามากกว่าความแปรปรวนระหว่างโรงเรียน นั่นคือ การเพิ่มจำนวนนักเรียนในแต่ละโรงเรียน จะทำให้การอ้างอิงสรุปมีค่าเพิ่มขึ้น

เชพเวลสัน และ คณะ (Shavelson and others, 1993: 215-232) ได้ศึกษาการประเมินการปฏิบัติงานโดยใช้ทฤษฎีการอ้างอิงสรุป ศึกษาจากองค์ประกอบด้านงาน (task) โอกาส (occasion) ผู้ประเมิน (rater) และวิธีวัด (method) เพื่อแสดงถึงความสามารถในการอ้างอิงในลักษณะของความเที่ยง และ ความตรงลู่เข้าของการประเมินการปฏิบัติงาน โดยศึกษาทั้งในระดับบุคคลและระดับโรงเรียน

ผลการศึกษาพบว่า ความแปรเปลี่ยนเชิงลุ่ม อันเนื่องจากผู้ประเมินไม่ใช่ประเด็นหลัก ทั้งนี้เพราะสามารถปรับปรุงได้ โดยการฝึกฝนเพื่อใช้ในการตัดสินการปฏิบัติงาน ในขณะที่ความแปรเปลี่ยนเชิงลุ่มอันเนื่องจากงานนับได้ว่า เป็นความคลาดเคลื่อนในการวัดที่มีขนาดใหญ่ จำนวนงานที่มากเป็นสิ่งจำเป็น ในการบ่งชี้ความเที่ยงของการวัดผลสัมฤทธิ์ทางการเรียน และ เมื่อพิจารณาความตรงลู่เข้า จากผลการวิจัยได้บ่งชี้ว่า ไม่มีความตรงลู่เข้าหากัน

ไนเซอร์ (Naizer, 1993: 2121-A) ได้ศึกษาการประเมินภาคปฏิบัติของครูในด้านความสามารถในการวางแผน การออกแบบ การบริหารและการประเมินเกี่ยวกับปัญหาการสอบ โดยใช้การวิเคราะห์จำแนก ทฤษฎีการอ้างอิงสรุป และสหสัมพันธ์แบบเพียร์สัน ทั้งนี้ได้ใช้ทฤษฎีการอ้างอิงสรุปในการพิจารณาแหล่งของความแปรปรวนของคะแนน

รุซไพร์มู และ คณะ (Ruiz Primo and others, 1993: 41-53) ได้ศึกษาความคงที่ของการประเมินการปฏิบัติงานจาก 2 แบบ คือ การตรวจสอบโดยการสังเกต และ การตรวจสอบจากสมุดบันทึก เป็นการศึกษาจากนักเรียนระดับ 6 จำนวน 29 คน ใน 2 โอกาส ผลจากการศึกษาบ่งชี้ว่า การปฏิบัติงานของนักเรียน และกระบวนการมีการเปลี่ยนแปลง ในขณะที่ความสามารถในการอ้างอิงไปสู่จำนวนโอกาสมีค่าในระดับปานกลาง

กรีนและเจอร์เรลล์ (Green and Jerrell, 1994: 141-51) ได้ศึกษาการอ้างอิงสรุปของการประเมินค่าตามหน้าที่ทางจิตวิทยาสังคม โดยใช้การอ้างอิงสรุปแบบพหุตัวแปร ได้ศึกษาจากตัวอย่าง 396 คน ซึ่งเป็นคนไข้หนักของสถาบันสุขภาพจิตของรัฐ ผลการศึกษาปรากฏว่า การอ้างอิงสรุปของการใช้สเกลย่อยให้ค่าต่ำกว่าการใช้การรวมของสเกลย่อย

คิม (Kim, 1994: 54-A) ได้ศึกษาการใช้โมเดล 2 พารามิเตอร์ของทฤษฎีการตอบสนองข้อสอบของซามิจิมา (Samejima) และทฤษฎีการอ้างอิงสรุป ในการพัฒนาการตรวจสอบความตรงของแบบสอบภาคปฏิบัติทางคณิตศาสตร์ การศึกษานี้ได้มีการใช้โปรแกรม GENOVA ในการศึกษาองค์ประกอบความแปรปรวน (Variance Component) สัมประสิทธิ์สหสัมพันธ์การ

อ้างอิงสรุปและสัมประสิทธิ์ความเชื่อถือได้ (Generalizability and Dependability Coefficient) ผลจากการศึกษานี้ทำให้เกิดสารสนเทศ ที่เป็นประโยชน์ต่อการใช้ทฤษฎีการอ้างอิงสรุปและทฤษฎีการตอบสนองข้อสอบ

มาร์คูไลด์ (Marcoulides, 1994: 3-7) ได้ศึกษาการคัดเลือกกรอบน้ำหนักของการศึกษาการอ้างอิงสรุปแบบพหุตัวแปร จากผลของน้ำหนักที่แตกต่างกัน ในการคัดเลือกจำนวนลังเกตที่เหมาะสมในรูปแบบการอ้างอิงสรุป ได้รับการศึกษาโดยการเปรียบเทียบจาก 4 กรอบ ซึ่งผ่านการจำลองสถานการณ์นั้น ผลการศึกษาได้บ่งชี้ว่าทั้ง 4 กรอบ ได้ให้คุณค่าที่เหมาะสมคล้ายคลึงกัน ซึ่งความเที่ยงก็ควรจะคล้ายคลึงกันด้วย

ลิน (Lin, 1994 : 5094-A.) ได้พัฒนาการสำรวจความกดดันของผู้ใหญ่ชาวไต้หวัน โดยทำการวิจัยเป็น 3 ระยะ ระยะแรกได้ใช้ตัวอย่างจำนวน 33 คน เพื่อสัมภาษณ์ประสบการณ์ที่เกี่ยวกับความกดดัน และใช้การวิเคราะห์เนื้อหาเพื่อวิเคราะห์สารสนเทศที่ได้จากการสัมภาษณ์ ในระยะที่ 2 ได้ใช้ผู้เชี่ยวชาญจำนวน 7 คน เพื่อพิจารณาตัดสินว่าข้อกระทงที่สร้างขึ้นจากการสัมภาษณ์นั้น เกี่ยวข้องกับความกดดันหรือไม่ โดยใช้ดัชนีความตรงเชิงเนื้อหาในการพิจารณาความตรงเชิงเนื้อหา สำหรับระยะที่ 3 ได้ใช้กลุ่มตัวอย่างจำนวน 351 คน เพื่อใช้กับเครื่องมือที่ได้ทำการพัฒนาขึ้นจากการศึกษาใน 2 ระยะแรก การวิจัยนี้ ได้ใช้ทฤษฎีการอ้างอิงสรุปในการพิจารณาความคงที่ และความสอดคล้องภายในโดยใช้สัมประสิทธิ์การอ้างอิงสรุป ทั้งนี้ในลักษณะของความคงที่นั้น สัมประสิทธิ์การอ้างอิงสรุปสำหรับการตัดสินใจสัมพัทธ์ มีค่า 0.899 และสัมประสิทธิ์การอ้างอิงสรุปสำหรับการตัดสินใจสัมบูรณ์มีค่า 0.897 และเมื่อพิจารณาในลักษณะของความสอดคล้องภายในแล้ว สัมประสิทธิ์การอ้างอิงสรุปมีค่าเป็น 0.719 และ 0.717 ในลักษณะของการตัดสินแบบสัมพัทธ์และการตัดสินแบบสัมบูรณ์ตามลำดับ

3.2 ประเทศไทย

ในประเทศไทย ยังมีการศึกษาทฤษฎีการอ้างอิงสรุปไม่มากนัก ในปี พ.ศ. 2529 จักรกฤษณ์ สำราญใจ ได้เขียนบทความเรื่อง Generalizability Theory โดยเนื้อหาส่วนใหญ่ มาจากบทความของคาร์ดิเนทและคณะ (Cardinet et al, 1976: 119-135) ศิริชัย กาญจนวาสี (2535, 109-118) ได้เขียนเรื่องทฤษฎีการอ้างอิงสรุป ไว้ในเอกสารการสอนชุดวิชาการพัฒนาแบบทดสอบผลสัมฤทธิ์ทางการเรียน เนื้อหาของทฤษฎีการอ้างอิงสรุปประกอบด้วยแนวคิดของทฤษฎีการอ้างอิงสรุป คำศัพท์ที่เกี่ยวข้องกับทฤษฎีการอ้างอิงสรุป และการประมาณค่าสัมประสิทธิ์การอ้างอิงสรุปสำหรับการออกแบบการวัดที่มีฟาเซ็ตเดียว ในลักษณะ

ของความเที่ยง และอุทุมพร จามรมาน (2538, 142-154) เขียนเรื่องทฤษฎีการอ้างอิงสรุปไว้ในหนังสือทฤษฎีการวัดทางจิตวิทยา เนื้อหาของทฤษฎีการอ้างอิงสรุปนี้ ประกอบด้วยความเป็นมาและแนวคิดของทฤษฎี แนวคิดในการวิเคราะห์ข้อมูล แบบการวิเคราะห์ข้อมูล และทฤษฎีการสรุปอ้างอิงกับความเที่ยงและความตรง

สำหรับงานวิจัย ที่มีการประยุกต์ทฤษฎีการอ้างอิงสรุปในการวัดผลนั้น ได้แก่ การประยุกต์ทฤษฎีการอ้างอิงสรุป ในการหาความเชื่อมั่นของการประเมินความตรงเชิงเนื้อหา (แดง กลางท่าได่, 2531) และการศึกษาสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง (ไพรัตน์ วงษ์นาม, 2533) ซึ่งมีรายละเอียดโดยสรุป ดังนี้

แดง กลางท่าได่ (2531) ได้ประยุกต์ทฤษฎีการอ้างอิงสรุป ในการหาความเชื่อมั่นของการประเมินความตรงเชิงเนื้อหา มีวัตถุประสงค์เพื่อค้นหาแหล่งความแปรปรวน ที่มีอิทธิพลต่อการหาความเชื่อมั่น ของการประเมินความตรงเชิงเนื้อหา จากแบบสอบวัดผลสัมฤทธิ์ทางการเรียนเรื่องเซ็ต โดยผู้เชี่ยวชาญ และเพื่อหารูปแบบการวัดที่ให้ความเชื่อมั่นของการประเมินสูงตลอดจน เพื่อหาขนาดของกลุ่มตัวอย่าง ที่เป็นตัวแทนประชากรขององค์ประกอบที่ศึกษา 3 องค์ประกอบ คือ ข้อสอบ ผู้เชี่ยวชาญ และ โรงเรียน

รูปแบบที่ใช้ในการวิจัย คือ ให้ผู้เชี่ยวชาญทุกคนประเมินข้อสอบทุกข้อ โดยสุ่มผู้เชี่ยวชาญมาจากแต่ละโรงเรียนที่ศึกษามีจำนวนเท่า ๆ กัน กลุ่มตัวอย่างที่ศึกษา เป็นตัวอย่างเชิงสุ่มจากเอกภพการสังเกตที่ยอมรับได้ขององค์ประกอบแต่ละตัว ซึ่งประกอบด้วยข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์เรื่องเซ็ตจำนวน 30 ข้อ โรงเรียนระดับมัธยมศึกษา สังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการจำนวน 9 โรงเรียน และผู้เชี่ยวชาญจำนวน 45 คน ในการเก็บข้อมูล ใช้แบบสอบ และแบบประเมินความตรงเชิงเนื้อหาชนิดมาตรประมาณค่า เป็นเครื่องมือ แล้วทำการวิเคราะห์ข้อมูลตามกระบวนการวิเคราะห์การอ้างอิงสรุป ปรากฏผลดังนี้

1. แหล่งความแปรปรวนที่มีอิทธิพล ต่อการประมาณค่าความเชื่อมั่นของการประเมินความตรงเชิงเนื้อหา ได้แก่ ข้อสอบ ผู้เชี่ยวชาญ ซึ่งอยู่ในโรงเรียน และปฏิสัมพันธ์ระหว่างข้อสอบกับโรงเรียน

2. รูปแบบการวัดที่ให้ค่าความเชื่อมั่นของการประเมินความตรงเชิงเนื้อหา โดยผู้เชี่ยวชาญสูงสุดคือ M_3 (I/-/S/R)

3. การที่จะให้ได้ค่าสัมประสิทธิ์การอ้างอิงสรุปอย่างน้อยมีค่าเท่ากับ 0.80 ต้องใช้ข้อสอบอย่างน้อย 9 ข้อ ผู้เชี่ยวชาญไม่เกิน 9 คน ต่อโรงเรียน โดยสุ่มจากโรงเรียนอย่างน้อย 7 โรงเรียน

4. การเพิ่มขนาดกลุ่มตัวอย่างขององค์ประกอบแต่ละองค์ประกอบ ทำให้ความแปรปรวนของความคลาดเคลื่อนมีค่าลดลง และสัมประสิทธิ์การอ้างอิงสรุปมีค่าเพิ่มขึ้น โดยเฉพาะการเพิ่มขนาดกลุ่มตัวอย่างขององค์ประกอบที่ต้องการสรุปอ้างอิง จะทำให้สัมประสิทธิ์การอ้างอิงสรุปมีค่าเพิ่มขึ้น มากกว่าการเพิ่มขนาดตัวอย่างขององค์ประกอบอื่น และการสรุปอ้างอิงผลการวัดไปยังเอกภพจำกัด จะให้ค่าสัมประสิทธิ์การอ้างอิงสรุป สูงกว่าการสรุปอ้างอิงผลการวัดไปยังเอกภพไม่จำกัด

ไพรัตน์ วงษ์นาม (2533) ได้ศึกษาสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง โดยมีวัตถุประสงค์เพื่อศึกษาและเปรียบเทียบผลของวิธีตรวจ การชี้แจง และการรู้ผลการเรียนของผู้ตอบ ที่มีต่อค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง วัดความสามารถในการแสดงความคิดเห็นเกี่ยวกับข่าวและเหตุการณ์ของนักเรียนชั้นประถมศึกษาปีที่ 6 พร้อมทั้งเลือกวิธีตรวจ จำนวนผู้ตรวจและจำนวนข้อสอบที่ให้ค่าสัมประสิทธิ์การอ้างอิงสรุปอย่างต่ำ 0.50 ในการศึกษา ได้ศึกษาจากตัวอย่างนักเรียน จำนวน 30 คน ข้อสอบแบบความเรียงที่ผู้วิจัยสร้างขึ้น 5 ข้อ และครูจำนวน 20 คน ทำการสุ่มโดยแบ่งผู้ตรวจออกเป็น 2 กลุ่ม กลุ่มละ 10 คน โดยให้กลุ่มหนึ่งตรวจแบบประเมินรวม อีกกลุ่มหนึ่งตรวจแบบวิเคราะห์ ซึ่งการตรวจแต่ละกลุ่มให้ทำ 3 ครั้ง โดยครั้งที่ 1 ให้ตรวจโดยใช้ประสบการณ์เดิมของผู้ตรวจ ครั้งที่ 2 ตรวจตามวิธีที่ได้รับคำชี้แจง โดยไม่รู้ผลการเรียนผู้สอบ และครั้งที่ 3 ตรวจตามที่ได้รับการชี้แจง และรู้ผลการเรียนผู้สอบ ซึ่งผลการวิจัย มีดังนี้

1. เมื่ออ้างอิงไปยังเอกภพของข้อสอบและผู้ตรวจพร้อมกัน (M1) ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงที่ตรวจโดยวิธีประเมินรวม มีค่าอยู่ระหว่าง 0.3328-0.4782 และตรวจโดยวิธีวิเคราะห์ มีค่าสัมประสิทธิ์การอ้างอิงสรุปอยู่ระหว่าง 0.3348-0.5895

2. เมื่ออ้างอิงไปยังเอกภพของผู้ตรวจอย่างเดียว (M2) จะได้ค่าสัมประสิทธิ์การอ้างอิงสรุปของวิธีประเมินรวม มีค่าระหว่าง 0.4743-0.6865 สำหรับวิธีวิเคราะห์ มีค่าระหว่าง 0.5985-0.7761

3. วิธีตรวจ การชี้แจง และการรู้ผลของผู้สอบต่างกัน ไม่มีผลต่อความแตกต่างของค่าสัมประสิทธิ์การอ้างอิงสรุปสำหรับ M1 แต่ใน M2 พบว่า ค่าสัมประสิทธิ์การอ้างอิงสรุป ของผลการตรวจ ตามที่ได้รับการชี้แจง มีค่าสูงกว่าการตรวจโดยใช้ประสบการณ์เดิม เฉพาะกรณีที่ผู้ตรวจที่ได้รับการฝึกอบรมไม่รู้ผลการเรียนของผู้สอบเท่านั้น ถ้ารู้ผลการเรียนของผู้สอบ ค่าสัมประสิทธิ์การอ้างอิงสรุปจะไม่แตกต่างกัน ค่าสัมประสิทธิ์การอ้างอิงสรุปของการตรวจ ที่ผู้ตรวจไม่รู้ประวัติการเรียนของผู้ตอบ มีค่าสูงกว่าการตรวจที่ผู้ตอบรับรู้ประวัติของผู้สอบ

4. วิธีตรวจ จำนวนข้อสอบ และจำนวนผู้ตรวจ ที่ให้ค่าสัมประสิทธิ์การอ้างอิงสรุปไม่น้อยกว่า 0.50 คือ วิธีตรวจวิเคราะห์ตามที่ได้รับคำชี้แจง และผู้ตรวจไม่รู้ผลการเรียนของผู้สอบ ใช้ข้อสอบ 6 ข้อ ผู้ตรวจอย่างน้อย 5 คน สำหรับ M1 และสำหรับ M2 ต้องใช้ผู้ตรวจ 5 คน และ ข้อสอบ 5 ข้อ

จากการศึกษาค้นคว้าในทฤษฎีการอ้างอิงสรุป ทั้งจากเอกสารและงานวิจัยที่เกี่ยวข้อง ทำให้ผู้วิจัยเกิดแนวในการทำวิจัย ที่นำทฤษฎีการอ้างอิงสรุปมาประยุกต์ใช้ในการศึกษาความตรงของการวัด ในการวิเคราะห์ความตรงลู่เข้าของการวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ โดยใช้แนวคิดด้านโมเดลเชิงสุ่มในการศึกษาความตรงลู่เข้าด้วยทฤษฎีการอ้างอิงสรุป