Generalizability Theory

จักรกฤษณ์ สำราญใจ

แนวคิดพื้นฐาน

ในทฤษฎีการวัดดั้งเดิม (Classical Test Theory) นั้น เชื่อว่าคะแนนที่เป็นผลการวัดที่สังเกตได้ (X) นั้น ส่วนหนึ่งเป็นคะแนนที่เป็นความสามารถที่แท้จริง (T) และอีกส่วนหนึ่งเป็นความคลาดเคลื่อน (E) เขียนเป็นแบบจำลองได้ว่า

$$X = T + E$$

และถือว่าส่วนของความคลาดเคลื่อนนี้มีเพียงส่วนเดียว (single error component) ไม่สามารถจะแยกแหล่งของความคลาดเคลื่อนออกไปได้ ส่วนคะแนนความสามารถที่แท้จริง (true score) ก็หมายถึงค่าที่คาดหวังของค่าสังเกตจากการวัดที่มีความเสมอเหมือนกัน (equivalent) ที่เป็นไปได้ทั้งหมด จากข้อสมมุติที่ว่า คะแนนความสามารถที่แท้จริงกับคะแนนความคลาดเคลื่อนไม่สัมพันธ์กัน ทำให้ได้ว่า

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

และนิยามค่าสัมประสิทธิ์ความเที่ยง (reliabilily coefficient) ว่าคือ อัตราส่วนของความแปรปรวนของคะแนนจริงกับความแปรปรวนของค่าที่สังเกตได้ นั่นคือ

$$\rho_{XT}^2 = \frac{\sigma_T^2}{\sigma_X^2} = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

การประมาณค่าอัตราส่วนความแปรปรวนหรือค่าสัมประสิทธิ์ความเที่ยง ก็ใช้ค่าสหสัมพันธ์ระหว่างแบบวัดที่คู่ขนานกัน หรือใช้การวิเคราะห์ความแปรปรวนแบบทางเดียว (one way ANOVA)

การถือว่าความแปรปรวนของผลการวัดมีเพียงแหล่งเดียวและความแปรปรวนเนื่องจากความคลาดเคลื่อนมีเพียงค่าเดียวเช่นนี้ทำให้การตีความหมายผลที่ได้จากแบบจำลองนี้มีความสับสนคลุมเครือเป็นอย่างมาก เมื่อมีการประมาณค่าสัมประสิทธิ์ความเที่ยงได้แตกต่างกัน

มีหลักฐานเป็นที่แน่ชัดว่าความแปรปรวนของผลการวัดเกิดขึ้นได้จากแหล่งต่างๆ หลายแหล่งด้วยกัน ทฤษฎี Generalizability ยอมรับในเรื่องนี้ โดยพยายามประมาณค่าของความแปรปรวนจากแหล่งต่างๆ โดยอาศัยเทคนิควิธีวิเคราะห์ความแปรปรวนประมาณค่า การทราบค่าความแปรปรวนที่เกิดจากแหล่งต่าง ๆ นี้ จะช่วยให้เราเข้าใจว่าความแปรปรวนที่ไม่พึงประสงค์เกิดขึ้นได้อย่างไร และสามารถที่จะออกแบบการเก็บรวบรวมข้อมูลเพิ่มเติมให้มีประสิทธิภาพยิ่งขึ้น คุณค่าของการวิเคราะห์ความแปรปรวนจากหลาย ๆ แหล่ง (multifacet analysis of error) Cronbach และคณะได้สรุปไว้ดังนี้

1. การพิจารณาการวัดในหลาย ๆ สถานการณ์อย่างชัดแจ้งจะช่วยขจัดความกำกวมที่มีและซ่อนเร้นอยู่ในแบบจำลองดั้งเดิมได้
2. การวิเคราะห์แบบนี้สามารถตรวจสอบผลของปฏิสัมพันธ์ของสถานการณ์ต่าง ๆ ซึ่งไม่สามารถทำได้โดยใช้แบบจำลองดั้งเดิม จะช่วยให้เราเข้าใจเรื่องของการวัดมากยิ่งขึ้น
3. การวิเคราะห์เพียงครั้งเดียวสามารถตอบคำถามได้หลาย ๆ เรื่อง ซึ่งถ้าใช้แบบจำลองดั้งเดิมจะต้องใช้ข้อมูลหลาย ๆ ชุด
4. ช่วยให้มีการวางแผนเก็บรวบรวมข้อมูลอย่างมีประสิทธิภาพ

อาจสรุปได้ว่า Generalizability Theory (G – Theory) เป็นทฤษฎีที่ได้จากการปรับปรุงแก้ไขทฤษฎีดั้งเดิม (Classical Theory) โดยดำเนินการดังนี้

1. รับว่าสถานการณ์การวัดหลาย ๆ สถานการณ์ไม่จำเป็นต้องคู่ขนานกัน (parallel)
2. สถานการณ์การวัดที่นำมาศึกษาได้รับการเลือกมาจาก universe ของสถานการณ์โดยการสุ่มหรือการเลือกแบบแบ่งชั้นภูมิ (stratified design)
3. ทำการวิเคราะห์ facet ตั้งแต่ 2 facets ขึ้นไปพร้อม ๆ กัน

มโนทัศน์เบื้องต้น (Basic Concept)

การจะศึกษาเรื่องของ G Theory จำเป็นที่ผู้ศึกษาจะต้องทำความเข้าใจกับมโนทัศน์ที่สำคัญของทฤษฎีนี้ก่อน มโนทัศน์เหล่านี้ได้แก่

สถานการณ์ (Condition) หมายถึงเงื่อนไขที่ทำให้เราได้ค่าสังเกตแต่ละค่าขึ้นมา สถานการณ์อาจจะเป็น test forms, สิ่งเร้า, ผู้สังเกต, ผู้ตรวจให้คะแนน, โอกาสที่สังเกต เป็นต้น

Facet หมายถึงชุดของเงื่อนไขที่เป็นชนิดเดียวกัน

ตัวอย่างเช่น นายสนิทได้พัฒนาวิธีวัดความสามารถในการเขียน โดยให้เขียนเรียงความ 2 เรื่อง และให้มีผู้ตรวจให้คะแนน 3 คน ในกรณีนี้การวัดจะมี 2 facets คือ หัวข้อเรียงความและผู้ตรวจให้คะแนนใน facet ที่ 1 คือหัวข้อเรียงความก็จะมี 2 conditions คือ 2 เรื่อง ใน facet ที่ 2 คือผู้ตรวจให้คะแนนจะมี 3 conditions จะเห็นว่ามีเมื่อเทียบกับ ANOVA แล้ว faccet ก็คือ factor และ condition ก็คือ level ของ factor ต่างๆ นั่นเอง

ใน G Theory ตามแนวคิดของ Cronbach นั้น ไม่ถือว่าสิ่งที่ถูกวัด (object of measurement) เป็น facet จึงมีคำสองคำที่ใช้ในความหมายแตกต่างกันคือ universe กับ population

Universe หมายถึง เซทของ conditions ทุก conditions ในแต่ละ facet

Population หมายถึง เซทของสิ่งที่วัดทุกหน่วย

แต่ต่อมา Cardinet และคณะได้พยายามขยายการประยุกต์ใช้ G Theory ให้กว้างขวางขึ้นโดยให้ถือว่า สิ่งที่ถูกวัดก็เป็น facet ได้ และใช้คำว่า universe ในความหมายของ population ตามที่ Cronbach เคยใช้ด้วย

Universe of Admissible Observations หมายถึง เซทของ all possible combination ของ condition ที่จะทำการสังเกตหรือวัดทั้งหมดในกรณีของนายสนิทก็จะประกอบด้วย เรียงความเรื่องที่ 1 กับผู้ตรวจคนที่ 1 เรียงความเรื่องที่ 2 กับผู้ตรวจที่ 1 ....และเรียงความเรื่องที่ 2 กับผู้ตรวจคนที่ 3 รวม 6 เซท (กรณีที่เป็น cross design)

Universe of Generalization หมายถึง universe ที่ผู้ตรวจตัดสินใจใช้ผลการวัดต้องการจะ generalize ผลการวัดให้ครอบคลุมไปถึง universe of generalization อาจมีได้หลาย universe เพราะผู้ตัดสินใจที่ต่างคนกันย่อมสนใจที่จะ generalize ผลการวัดไปยัง universe ที่ต่างกัน

Universe score หมายถึง ค่าเฉลี่ยของค่าสังเกตใน universe of generalization เนื่องจาก universe of generalization มีได้หลาย unliveres ดังนั้น universe score จึงมีได้หลายค่า

ความคลาดเคลื่อนในการวัด (error of measurement) แบ่งเป็น 2 ชนิดคือ

1. ความคลาดเคลื่อนสัมบูรณ์ (absolute error - Δ error) คือ ผลต่างระหว่างคะแนนที่สังเกตหรือวัดได้กับ universe score

2. ความคลาดเคลื่อนสัมพัทธ์ (relative error — δ error) คือ ผลต่างระหว่างส่วนเบี่ยงเบนที่สังเกตได้ (observed deviation score) กับ universe score ในรูปของคะแนนเบี่ยงเบน

Generalizability เครื่องมือวัดจะมีประโยชน์ก็ต่อเมื่อคะแนนที่ได้จากการวัดให้สารสนเทศอะไรบางอย่างแก่เรา และค่าจากการวัดครั้งหนึ่งควรจะให้สารสนเทศเช่นเดียวกับที่เราคาดว่าจะได้จากการวัดครั้งอื่น ๆ ในสถานการณ์ที่เสมอเหมือนกัน (equivalent conditions) ซึ่งใน Classical model หมายถึงเครื่องมือวัดจะต้องมีความเที่ยง (reliability) นั่นเอง

ใน classical model ถือว่าสถานการณ์ที่เสมอเหมือนกันนั้น เป็นสถานการณ์ที่คู่ขนาน (parallel) คะแนนจริงที่ได้จากการวัดทุกสถานการณ์จะเท่ากัน นั่นคือสถานการณ์ต่างๆ ไม่ทำให้เกิดความแปรปรวนที่เป็นระบบ แต่ใน G — Theory ถือว่าสถานการณ์ในการวัดเป็นตัวแทนหรือตัวอย่างของสถานการณ์ทั้งหมดใน universe สถานการณ์ที่เสมอเหมือนกันก็คือ สถานการณ์ที่เป็นสมาชิกของ universe เดียวกันการวัดภายใต้สถานการณ์ต่าง ๆ กัน ย่อมหลีกไม่พ้นที่จะเกิดแหล่งความแปรปรวนที่เป็นระบบขึ้น ดังนั้นการคาดหวังว่า ผลการวัดภายใต้สถานการณ์ชุดอื่น ๆ ใน universe จะให้สารสนเทศเช่นเดียวกับผลการวัดภายใต้สถานการณ์ที่ถูกเลือกมาเป็นตัวอย่าง จึงมีลักษณะเป็นการสรุปอ้างอิง (generalize) ผลการวัดไปยังสถานการณ์ใน universe นั่นเอง

สรุปแล้วใน classical model ใช้คำ reliability เพราะเชื่อว่าสถานการณ์การวัดทุกสถานการณ์เป็นสถานการณ์ที่คู่ขนานกัน ส่วนใน G — Theory ถือว่าสถานการณ์ต่าง ๆ ไม่ได้คู่ขนานกันเพียงเป็นแต่เป็นสมาชิกของ universe เดียวกัน จึงใช้คำว่า generalizability

Generalizability Coefficient ($\varepsilon \rho^2$) หมายถึง อัตราส่วนระหว่างความแปรปรวนของ univnrsn scorne กับความแปรปรวนที่คาดหวังของคะแนมที่สังเกตได้

$$\varepsilon \rho^2 = \frac{\sigma^2 (p)}{\varepsilon \sigma^2 (x)}$$

สำหรับค่า $\varepsilon\sigma^2 (x)$ จะเท่ากับ $\sigma^2 (p) + \sigma^2(\Delta)$ หรือ $\sigma^2(p) + \sigma^2(\delta)$ แล้วแต่กรณีในการหา Generalizability coefficient นั้น เราจะต้องกำหนด universe of generalization ให้แน่นอนแล้วจึงหาส่วนความแปรปรวนของ universe score และส่วนความแปรปรวนของความ

คลาดเคลื่อน ดังนั้น generalizability coefficient จึงมีได้หลายค่าแตกต่างกัน ถ้าเรากำหนด universe of generalization แตกต่างกัน การแปลความหมายก็คล้ายกับ reliability coefficient ใน classical model คือ จะเป็นดัชนีที่บอกถึงว่า ถ้าเราสังเกตหรือวัดภายใต้ conditions อื่น ๆ ใน universe of generalization เดียวกันแล้วผลการวัดจะมีความคงเส้นคงวาหรือแม่นยำเพียงใด ประโยชน์ที่จะได้จากค่า $\varepsilon \mu^2$ ก็คือ ช่วยให้เราสามารถปรับแบบการวัด (measurement design) ให้ได้ระดับของ generalizability ที่ต้องการภายใต้ข้อจำกัดของค่าใช้จ่าย เวลา ทรัพยากรอื่น ๆ เท่าที่มีอยู่ เนื่องจากเราทราบว่า ความแปรปรวนจากแหล่งใดบ้างมีผลกระทบต่อ generalizability coefficient เราจะสามารถปรับ design เพื่อจัดการกับค่าความแปรปรวนจากแหล่งนั้น ๆ ให้ค่าที่เหมาะสมได้ ประโยชน์อีกประการหนึ่งก็คือจะเป็นตัวกำหนดว่า measurement design นั้น ๆ เหมาะสมกับวัตถุประสงค์ของการวัดแล้วหรือยัง

กระบวนการวิเคราะห์ generalizability

Cronbach และคณะได้เสนอกระบวนการวิเคราะห์ generalizability เป็น 2 ตอน คือ Generalizability Study Design (G – Study) กับ Decision Study Design (D – Study)

G – study design เป็นการรวบรวมข้อมูลเพื่อที่จะประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ ในการทำ G – study ผู้ศึกษาจะต้องกำหนด facet ต่าง ๆ ดูความสัมพันธ์ระหว่าง facet กำหนด universe of admissible observations และใช้เทคนิค ANOVA หาค่า mean square ของแหล่งความแปรปรวนต่าง ๆ แล้วประมาณค่าความแปรปรวนของแหล่งต่าง ๆ

D – study design เป็นการรวบรวมข้อมูลเพื่อทำการตัดสินให้หรือลงข้อสรุป ในการทำ D – study ผู้ศึกษาจะต้องกำหนด universe of generalization หาความแปรปรวนของ universe score ความแปรปรวนของส่วนที่เป็นความคลาดเคลื่อนหาค่า generalizability coefficient ปรับแบบแผนการวัดเพื่อให้ได้ค่า generalizability ที่เหมาะสม

ต่อมา Cardinet และคณะได้พยายามขยายแนวคิดของ G Theory ให้กว้างขึ้นโดยใช้หลักของความสมมาตร (principle of symmetry) ที่ว่า แต่ละ factor ของ design สามารถถูกเลือกให้

เป็น object of measurement ได้ทั้งนั้นและได้ขยายการวิเคราะห์ generalizability ออกเป็น 4 ตอน คือ

1. Observation Design งานในขั้นนี้ได้แก่ การเลือก facet ความสัมพันธ์ระหว่าง facet ว่าเป็น crossed หรือ nested หรือ confounded กำหนดจำนวน condition ที่จะสังเกตในแต่ละ facet รวบรวมข้อมูล ใช้ ANOVA หาค่า mean square ของแต่ละแหล่ง
2. Estimation Design งานในขั้นนี้เป็นการพิจารณาการเลือก condition ในแต่ละ facet ว่าเป็น fixed หรือ random แล้วประมาณค่าความแปรปรวนของแต่ละแหล่ง

ขั้นตอนที่ 1 และ 2 นี้เป็นการประยุกต์เทคนิค ANOVA มาใช้ ซึ่งก็คือ G – Study นั่นเอง

3. Measurement Design เป็นขั้นตอนที่เริ่มนำมโนทัศน์ของ G – Theory มาใช้ จะแยก facet ออกเป็น 2 พวก (two faces of measurement) พวกหนึ่งเป็น facet ที่เกี่ยวกับสิ่งที่ถูกวัดเรียกว่า the face of differentiation อีกพวกหนึ่งที่เหลือเป็น facet ที่เกี่ยวกับเครื่องมือวัดเรียกว่า the face of instrument (หรือ conditions of observation ตามที่ Cronbach และคณะใช้นั่นเอง) จากนี้ก็แยกแหล่งความแปรปรวนออกเป็น Differentiation variance (หรือ universe score variance) กับ Generalization variance (หรือ error variance) แล้วคำนวณค่า Generalizability Coefficient โดยการหาร Differentiation variance ด้วยผลบวกของ Differentiation variance กับ Generalizability variance
4. Optimization Design เป็นการปรับแบบแผน (design) เพื่อจะลดความคลาดเคลื่อนลง หรือลดค่าใช้จ่าย แล้วประมาณค่า Generalizability coefficient ใหม่ ซึ่งจะนำไปสู่ design ที่เหมาะสมที่สุดที่จะนำไปใช้ต่อไป

ขั้นที่ 3 และ 4 นี้ก็คือ D – Study ของ Cronbach และคณะนั่นเอง

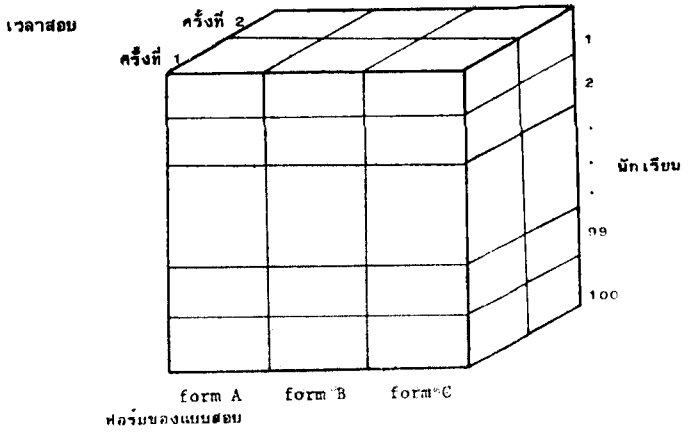
จะเห็นว่าการวิเคราะห์ generalizability นั้นอาศัยเทคนิควิธีของ ANOVA อย่างมาก ดังนั้นถ้าจะศึกษาเรื่องของ G – Theory ให้เข้าใจแจ่มแจ้งแล้ว ผู้ศึกษาจะต้องมีพื้นฐานของ ANOVA

ที่แน่นพอสมควรเพื่อจะได้ประมาณค่าความแปรปรวนได้เหมาะสมกับ design อย่างไรก็ตามการใช้ ANOVA ใน G – Theory มีข้อแตกต่างจากการใช้ ANOVA ในการวิเคราะห์ข้อมูลทั่วไปคือ ในการวิเคราะห์ข้อมูลทั่วไปนั้นเรามุ่งใช้ ANOVA เพื่อทดสอบสมมุติฐานทางสถิติ แต่ใน G – Theory จะไม่สนใจทดสอบสมมุติฐานแต่จะมุ่งประมาณค่าความแปรปรวนของแหล่งต่างๆมากกว่า แม้ว่า ANOVA จะมีบทบาทสำคัญใน G – Theory ก็ตามแต่ ANOVA ก็ไม่ใช่ส่วนสำคัญหรือหัวใจของ G – Theory

ตัวอย่าง

สมมุติว่า เราต้องการสร้างแบบสอบเพื่อวัดความเข้าใจในการอ่าน ได้ออกแบบวิธีการวัดดังนี้ เลือกบทความที่มีความยาวเท่ากัน 3 เรื่อง เขียนโดยผู้เขียน 3 คน มาสร้างเป็นแบบสอบโคลซ์ (close Test) โดยตัดทุก ๆ คำที่ 7 ออกให้ผู้ตอบเติมคำที่ถูกต้องลงไป นำแบบสอบนี้ไปสอบนักเรียน 100 คน 1 ครั้ง ระหว่างที่กำลังเรียนวิชาการอ่าน

จาก design ข้างต้น จะเห็นว่า design นี้มี 3 facets คือ (1) ผู้ตอบหรือนักเรียน (n_s = 100) (2) ฟอร์มของแบบสอบ (n_f = 3) (3) เวลาในการสอบ (n_m = 2) ข้อมูลที่จะได้จากการสังเกตจะมีลักษณะเป็น factorial design มี 3 factors ดังภาพ



ภาพที่ 1 แบบแผนการทดลองใช้แบบสอบโคลซ์วัดความเข้าใจในการอ่าน 3 forms 2 ครั้ง

ทำการวิเคราะห์ความแปรปรวนข้อมูล สมมุติว่าได้ผลการวิเคราะห์ดังตารางที่ 1

ตารางที่ 1

การวิเคราะห์ความแปรปรวนของข้อมลจาก desian ในภาพที่ 1

แหล่งความแปรปรวน	SS	df	MS	E(MS)
นักเรียน (S)	8201	99	83	$\sigma^2_{sfm,e} + n_m\sigma^2_{sj} + n_f\sigma^2_{sm} + n_f n_m \sigma^2_s$
ฟอร์ม (F)	3216	2	1608	$\sigma^2_{sfm,e} + n_m\sigma^2_{sf} + n_s\sigma^2_{mf} + n_s n_m \sigma^2_f$
เวลา (M)	1917	1	1917	$\sigma^2_{sfm,e} + n_s\sigma^2_{mf} + n_f\sigma^2_{sm} + n_s n_f \sigma^2_m$
SxF	1584	198	8	$\sigma^2_{sfm,e} + n_m\sigma^2_{sf}$
SxM	1683	99	17	$\sigma^2_{sfm,e} + n_f\sigma^2_{sm}$
FxM	804	2	402	$\sigma^2_{sfm,e} + n_s\sigma^2_{fm}$
SxFxM	396	198	2	$\sigma^2_{sfm,e}$
	17801	599		

ประมาณค่าความแปรปรวนของแหล่งต่าง ๆ ได้ผลดังตารางที่ 2

ตารางที่ 2

สูตรประมาณค่าความแปรปรวน และความแปรปรวนจากแหล่งต่าง ๆ

แหล่งความแปรปรวน	สูตร	ค่าความแปรปรวน
$\hat{\sigma}^2_{sfm,e}$	MS_{sfm}	2
$\hat{\sigma}^2_{jm}$	$(MS_{fm} - MS_{sfm})/n_s$	4
$\hat{\sigma}^2_{sm}$	$(MS_{sm} - MS_{sfm})/n_f$	5
$\hat{\sigma}^2_{sf}$	$(MS_{sf} - MS_{sfm})/n_m$	3
$\hat{\sigma}^2_{m}$	$(MS_{m} - MS_{sm} - MS_{fm} + MS_{sfm})/n_s n_f$	5
$\hat{\sigma}^2_{f}$	$(MS_{f} - MS_{sf} - MS_{fm} + MS_{sfm})/n_s n_m$	6
$\hat{\sigma}^2_{s}$	$(MS_{s} - MS_{sf} - MS_{sm} + MS_{sfm})/n_f n_m$	10

ในแต่ละบัญหาจะประมาณค่า generalizability coefficient ได้ดังตารางที่ 4

ตารางที่ 4

สูตรสำหรับคำนวณค่า generalizability coefficient และค่าสัมประสิทธิ์สำหรับแต่ละบัญหาในต่างๆ จากตารางที่ 3

รูปแบบที่	face of diff.	face of gener.		สูตร	$\hat{\rho}^2$
		rand	fix		
1	S	F	M	$\frac{\sigma_s^2 + (1/n_m)\sigma_{sm}^2}{\sigma_s^2 + (1/n_m)\sigma_{sm}^2 + (1/n_f)\sigma_{sf}^2 + (1/n_f n_m)\sigma_{sfm,e}^2}$	.904
2	S	M	F	$\frac{\sigma_s^2 + (1/n_f)\sigma_{sf}^2}{\sigma_s^2 + (1/n_m)\sigma_{sm}^2 + (1/n_f)\sigma_{sf}^2 + (1/n_f n_m)\sigma_{sfm,e}^2}$	.792
3	S	F&M	-	$\frac{\sigma_s^2}{\sigma_s^2 + (1/n_m)\sigma_{sm}^2 + (1/n_f)\sigma_{sf}^2 + (1/n_f n_m)\sigma_{sfm,e}^2}$	.720
4	F	S	M	$\frac{\sigma_f^2 + (1/n_m)\sigma_{fm}^2}{\sigma_f^2 + (1/n_m)\sigma_{fm}^2 + (1/n_s)\sigma_{sf}^2 + (1/n_s n_m)\sigma_{sfm,e}^2}$	.995
5	F	M	S	$\frac{\sigma_f^2 + (1/n_s)\sigma_{sf}^2}{\sigma_f^2 + (1/n_m)\sigma_{fm}^2 + (1/n_s)\sigma_{sf}^2 + (1/n_s n_m)\sigma_{sfm,e}^2}$	.750
6	F	S&M	-	$\frac{\sigma_f^2}{\sigma_f^2 + (1/n_m)\sigma_{fm}^2 + (1/n_s)\sigma_{sf}^2 + (1/n_s n_m)\sigma_{sfm,e}^2}$	.746
7	M	S	F	$\frac{\sigma_m^2 + (1/n_f)\sigma_{fm}^2}{\sigma_m^2 + (1/n_f)\sigma_{fm}^2 + (1/n_s)\sigma_{sm}^2 + (1/n_s n_f)\sigma_{sfm,e}^2}$	.992
8	M	F	S	$\frac{\sigma_m^2 + (1/n_s)\sigma_{sm}^2}{\sigma_m^2 + (1/n_f)\sigma_{fm}^2 + (1/n_s)\sigma_{sm}^2 + (1/n_s n_f)\sigma_{sfm,e}^2}$	.791
9	M	S&F	-	$\frac{\sigma_m^2}{\sigma_m^2 + (1/n_f)\sigma_{fm}^2 + (1/n_s)\sigma_{sm}^2 + (1/n_s n_f)\sigma_{sfm,e}^2}$	.783
10	S&F	M	-	$\frac{\sigma_s^2 + \sigma_j^2 + \sigma_{sf}^2}{\sigma_s^2 + \sigma_f^2 + \sigma_{so}^2 + 1/n_m(\sigma_{sm}^2 + \sigma_{fm}^2 + \sigma_{sfm,e}^2)}$	.776
11	S&M	F	-	$\frac{\sigma_s^2 + \sigma_m^2 + \sigma_{sm}^2}{\sigma_s^2 + \sigma_m^2 + \sigma_{sm}^2 + 1/n_f(\sigma_{sf}^2 + \sigma_{mf}^2 + \sigma_{sfm,e}^2)}$	.870
12	F&M	S	-	$\frac{\sigma_f^2 + \sigma_m^2 + \sigma_{fm}^2}{\sigma_f^2 + \sigma_m^2 + \sigma_{fm}^2 + 1/n_s(\sigma_{sf}^2 + \sigma_{sm}^2 + \sigma_{sfm,e}^2)}$	.993

กำหนด universe of generalization หรือ face of generalization และ face of differentiation หรือ object of measurement การกำหนดจะขึ้นอยู่กับปัญหาของผู้ศึกษาสำหรับกรณีตัวอย่างของเราปัญหาที่เป็นไปได้ทั้งหมดสามารถกำหนดได้ 12 รูปแบบ ดังตารางที่ 3

รูปแบบที่	face of diff	face of gen.		ลักษณะปัญหา	Universe of general
		rand	fix		
1	S	F	M	ความเข้าใจในการอ่านของนักเรียนในการสอบ 2 ครั้งนี้ สำหรับแบบสอบฟอร์มใด ๆ	ฟอร์ม
2	S	M	F	ความเข้าใจในการอ่านของนักเรียนในแบบสอบ 3 ฟอร์มนี้ สำหรับการสอบครั้งใด ๆ	เวลาที่สอบ
3	S	F&M	-	ความเข้าใจในการอ่านของนักเรียนในการสอบครั้งใด ๆ และแบบสอบฟอร์มใด ๆ	ฟอร์มและเวลาที่สอบ
4	F	S	M	ความยากของแบบสอบ 3 ฟอร์มในการสอบ 2 ครั้งนี้ สำหรับนักเรียนคนใด ๆ	นักเรียน
5	F	M	S	ความยากของแบบสอบ 3 ฟอร์มนี้ สำหรับนักเรียนกลุ่มนี้กับการสอบครั้งใด ๆ	เวลาที่สอบ
6	F	S&M	-	ความยากของแบบสอบ 3 ฟอร์มนี้ สำหรับนักเรียนคนใด ๆ และการสอบครั้งใด ๆ	นักเรียนและเวลาที่สอบ
7	M	S	F	ผลของการสอนอ่านเกี่ยวกับแบบสอบ 3 ฟอร์มนี้ สำหรับนักเรียนคนใด ๆ	นักเรียน
8	M	F	S	ผลของการสอนอ่านสำหรับนักเรียนกลุ่มนี้ในการเข้าใจในการอ่านบทความใด ๆ	ฟอร์ม
9	M	S&F	-	ผลของการสอนอ่านสำหรับนักเรียนคนใด ๆ ในการเข้าใจในการอ่านบทความใด ๆ	นักเรียนฟอร์ม
10	S&F	M	-	ความเข้าใจในการอ่านบทความแต่ละบทของนักเรียนกลุ่มนี้ในการสอบครั้งใด ๆ	เวลาที่สอบ
11	S&M	F	-	ความเข้าใจในการอ่านของนักเรียนกลุ่มนี้ในการสอบแต่ละครั้งใน 2 ครั้งนี้ ต่อบทความใด ๆ	ฟอร์ม
12	F&M	S	-	ความยากของแบบสอบทั้ง 3 ฟอร์มในการสอบแต่ละครั้ง ใน 2 ครั้งนี้ สำหรับนักเรียนคนใด ๆ	นักเรียน

เมื่อพิจารณาค่า $\hat{\rho}^2$ ของ design ต่าง ๆ แล้ว จะเห็นว่า design ที่ 3 ไม่เหมาะสมที่จะอ้างอิงผลการวัดความเข้าใจในการอ่านของนักเรียนไปยังบทความอื่น ๆ และการสอบครั้งอื่น ๆ ($\hat{\rho}^2$ = .720) ถ้ายังต้องการอ้างอิงผลในลักษณะนี้อยู่ก็คงจะต้องปรับ design ใหม่ โดยอาจจะต้องเพิ่มจำนวนครั้งของการสอนเป็น 4 ครั้ง และใช้แบบสอบประมาณ 5 ฟอร์ม ซึ่งจะทำให้ $\hat{\rho}^2$ เพิ่มขึ้นเป็น .837 อยู่ในระดับที่พอจะรับได้ แต่ถ้าเราจะจำกัดผลการวัดเฉพาะบทความ 3 บทความนี้และอ้างอิงผลไปยังการสอบครั้งใด ๆ การอ้างอิงก็เกือบอยู่ในระดับที่พอจะยอมรับได้ ถ้าเรากลับกันระหว่างฟอร์มกับเวลาที่สอบ โดยให้จำกัดการอ้างอิงเฉพาะการสอบ 2 ครั้งนี้ไปยังแบบสอบฟอร์มใด ๆ (หรือบทความใด ๆ) ค่า $\hat{\rho}^2$ จะสูงถึง .904 นั่นแสดง design นี้เหมาะสำหรับการวัดความเข้าใจในการอ่านบทความใด ๆ ถ้าจำกัดการวัดแค่เพียง 2 ครั้งนี้เท่านั้น

เปรียบเทียบ design ที่ 4, 5, 6 จาก design ที่ 5 ยังไม่สมควรจะสรุปว่าความยากของแบบสอบจะคงที่ถ้าหากไม่จำกัดผลแค่เพียง 2 ครั้งนี้ เราอาจจะสรุปผลได้ถ้าหากว่าเราเพิ่มตัวอย่างจำนวนครั้งของการวัดให้มากขึ้น ค่า $\hat{\rho}^2$ จะสูงเป็น .81 ถ้ามีการวัด 3 ครั้งในช่วงเวลาเดียวกัน และจะสูงเป็น .85 ถ้ามีการวัด 4 ครั้ง

จาก design ที่ 7 ค่า $\hat{\rho}^2$ = .992 นับว่าสูงมาก แสดงว่า เราสามารถสรุปได้ว่าผลการสอบอ่านในวิชาการอ่านจะให้ผลเช่นเดียวกันกับนักเรียนคนใด ๆ ในการอ่านบทความ 3 บทนี้ ในการทดลองอาจจะใช้นักเรียนให้น้อยลงอีกได้เพื่อการประหยัดค่าใช้จ่าย แต่ถ้าจะอ้างอิงไปยังบทความอื่น ๆ แล้ว design นี้อาจจะยังไม่เหมาะสม ($\hat{\rho}^2$ = .791 และ .783)

สรุป

Generalizability Theory เป็นแนวคิดมองความแปรปรวนของการวัดว่าเกิดจากหลายแหล่งและพยายามจะแยกแยะความแปรปรวนที่เกิดขึ้นจากแหล่งต่าง ๆ ให้ชัดเจน โดยอาศัยเทคนิคการวิเคราะห์ความแปรปรวนประมาณค่าความแปรปรวนของแหล่งต่าง ๆ และ พยายามหาว่า จากวิธีการวัดที่กำหนดขึ้นมาเราจะสามารถอ้างอิงผลการวัดไปยังมิติต่าง ๆ ที่สนใจได้เพียงใด โดยมี Generalizability Coefficient เป็นตัวชี้ ซึ่งจะช่วยให้เราเห็นแนวทางที่จะปรับปรุงวิธีการของเราให้ดีขึ้น สามารถอ้างอิงผลการวัดไปยังมิติที่ต้องการได้อย่างมีประสิทธิภาพมากขึ้น

บรรณานุกรม

Brennan, R.L., and Kane, M.T. "Generalizability Theory : A Review." New Direction for Testing and Measurement 4 (1979) : 33 – 51.

Brennan, R.L. Elements of Generalizability Theory. Iowa : The American College Testing Program, 1983.

Cardinet, J, Tourneur, Y, and Allal, L. "The Symmetry of Generalizability Theory : Applications to Educational Measurement." Journal of Educational Measurement 13 (Summer 1976) : 119 – 135.

_____ . "Extension of Generalizability Theory and Its Applications in Educational Measurement." Journal of Educational Measurement 18 (Winter 1981) : 183 – 204.

Cronbach, L.J., Rajaratnam, N., and Gleser, G.C. "Theory of Generalizability : A Liberalization of Reliability Theory." The British Journal of Statistical Psychology XVI (November 1963) : 137 – 163.

Cronbach, L.J. et al. The Dependability of Behavioral Measurements : Kheory of Generalizability for Scores and Profiles. New York : John Wiley & Sons, Inc., 1972.

Kane, M.T., and Brennan, R.L. "The Generalizability of Class Means." Review of Educational Research 47 (Winter 1977) : 297 – 262.