วารสารมหาวิทยาลัยราชภัฏยะลา
JYRU
JOURNAL OF YALA RAJABHAT UNIVERSITY
Print ISSN: 1905-2383 e-ISSN: 2651-0863

บทความวิชาการ

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด : แนวคิดและการประยุกต์ใช้การวิจัยทางการศึกษา
Generalizability Theory: Principles and Application in Educational Research

รุจิรา คงประพันธ์[1*] และน้ำทิพย์ องอาจวาณิชย์[2]
Ruchira Kongprapan[1*] and Namthip Ongrardwanich [2]

[1]สำนักงานเขตพื้นที่การศึกษาประถมศึกษาพัทลุง เขต 2 ตำบลแม่ขรี อำเภอตะโหมด จังหวัดพัทลุง 93160
[2]คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร อำเภอเมืองพิษณุโลก จังหวัดพิษณุโลก 65000
[1]Phatthalung Primary Educational Service Area Office 2, Mae khri, Tamot, Phatthalung 93160. Thailand
[2]Faculty of Education, Naresuan University, Mueng Phitsanulok, Phitsanulok 65000. Thailand
*Corresponding Author, E-mail: ruchira7311@gmail.com
(Received: Oct 14, 2022; Revised: Feb 8, 2023; Accepted: Mar 1, 2023)

บทคัดย่อ

บทความวิชาการนี้มีวัตถุประสงค์เพื่ออธิบายแนวคิด หลักการทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดและอธิบายการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดทางวิจัยการศึกษา โดยนำเสนอเนื้อหาหลักออกเป็น 3 ส่วน คือ แนวคิดทฤษฎี การวิเคราะห์สัมประสิทธิ์การสรุปอ้างอิง และการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดกับการวิจัยทางการศึกษา ซึ่งทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด เป็นทฤษฎีที่ใช้วิเคราะห์ความน่าเชื่อถือของผลการวัดในสถานการณ์ของการวัดลักษณะต่าง ๆ ที่เป็นเป้าหมายของการนำเครื่องมือไปใช้ ซึ่งเป็นทั้งวิธีการประเมินที่น่าเชื่อถือหรือความเที่ยงของเครื่องมือและกลยุทธ์ของการออกแบบการวัดให้ผลของการวัดมีความน่าเชื่อถือ หรือมีความเที่ยงสูงถึงระดับที่ต้องการ เพื่อนำผลไปใช้เป็นสารสนเทศสำหรับการตัดสินใจอย่างมีประสิทธิภาพ

คำสำคัญ : ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด แนวคิดทฤษฎี การประยุกต์ใช้

Abstract

This academic article aims to explain the concept, the principle of the generalizability theory. And explain the application of This academic article aims to explain the concept, the principle of the generalizability theory in educational research. The main content is presented in 3 parts ; Theoretical concepts, Analysis of coefficients and the application of the generalizability theory with educational research. The generalizability theory is a theory used to analyze the reliability of measurement results in situations of measurement of various characteristics that are the target of the instrument's use. It is either a reliable assessment method or the validity of the instrument and the measurement design strategy to ensure that the results of the measurement are reliable. or has a high precision to the desired level to use the results as information for effective decision making.

Keywords : Generalizability theory, Principles, Application

บทนำ

แนวคิดทฤษฎีการทดสอบแบบดั้งเดิม ใช้สำหรับประมาณค่าความเที่ยงของแบบสอบที่ใช้กับเฉพาะกลุ่มบุคคลภายใต้เงื่อนไขของการทดสอบที่เฉพาะ ได้แก่ การทดสอบที่ผู้สอบทุกคนต้องทำข้อสอบเหมือนกันทุกข้อ ผลการตอบได้รับการให้คะแนนโดยผู้ตรวจคนเดียว และแบบสอบที่ใช้ตั้งอยู่บนข้อตกลงเบื้องต้นของความเป็นคู่ขนานระหว่างแบบสอบ แต่ถ้าสถานการณ์ของการทดสอบแตกต่างไปจากที่กล่าวมา จะไม่สามารถนำมาประยุกต์ใช้ได้ เช่น กรณีของการทดสอบที่มีผู้ตรวจหลายคนให้คะแนนข้อสอบความเรียงของผู้สอบทุกคน หรือกรณีการทดสอบที่ใช้ผู้ตรวจหลายคนให้คะแนนการตอบแบบสอบที่มีความยาวต่างกัน หรือกรณีที่ผู้ตรวจหลายคนให้คะแนนผลการตอบแบบสอบที่มีรูปแบบความยาวและจำนวนครั้งของการทดสอบแตกต่างกัน เป็นต้น (Kanjanawasee, 2020)

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด หรือ G–Theory จึงเป็นทฤษฎีที่ขยายแนวคิดมาจากทฤษฎีการทดสอบแบบดั้งเดิม ที่สามารถใช้วิเคราะห์ความน่าเชื่อถือของผลการวัดในสถานการณ์ของการวัดลักษณะต่าง ๆ ที่เป็นเป้าหมายของการนำเครื่องมือไปใช้ สามารถประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ ได้จากการวัดบริบทเดียวกันและจากการวิเคราะห์เพียงครั้งเดียว สามารถประมาณค่าสัมประสิทธิ์ความเที่ยงโดยเน้นไปที่องค์ประกอบความแปรปรวนของแต่ละแหล่งความคลาดเคลื่อนของการวัด สามารถแบ่งตัดสินใจออกเป็น 2 ลักษณะคือ การตัดสินใจเชิงสัมพันธ์ และการตัดสินใจเชิงสมบูรณ์ สามารถประมาณน้ำหนักของแหล่งความคลาดเคลื่อนของการวัดที่สามารถเป็นไปได้ทั้งหมด การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดทางวิจัยการศึกษา ทำให้ทราบว่าแหล่งความคลาดเคลื่อนที่เกิดขึ้นในกระบวนการวัดเกิดจากจำนวนผู้ตรวจให้คะแนน จำนวนข้อสอบ ความยาวของแบบสอบ รูปแบบของข้อสอบ จำนวนครั้งของการสอบ หรือคุณลักษณะของผู้ตรวจ เป็นต้น เมื่อนักวิจัยทราบแหล่งความคลาดเคลื่อนจะสามารถออกแบบการวัดซึ่งทำได้ 3 รูปแบบ ได้แก่ การออกแบบการวัดแบบไขว้ (Crossed design) การออกแบบการวัดแบบแฝง (Nested design) และการออกแบบการวัดแบบผสม (Confounded design) เพื่อควบคุมแหล่งความคลาดเคลื่อนได้ตรงประเด็น นักวิจัยยังทราบถึงระดับความน่าเชื่อถือของผลการวัด เรียกว่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability coefficient) ที่สามารถบ่งบอกถึงความแม่นยำของการสรุปอ้างอิง จากคะแนนสังเกตได้ของผู้สอบแต่ละคนไปยังคะแนนเฉลี่ยของผู้สอบที่ควรจะได้จากการทดสอบภายใต้สถานการณ์หรือเงื่อนไขของการวัดที่ยอมรับ ได้ทั้งหมด นักวิจัยสามารถนำข้อมูลที่ได้มาใช้ในการออกแบบการวัดในครั้งต่อไป เพื่อเป็นการลดความคลาดเคลื่อนของการวัดให้น้อยลงหรือมีความเที่ยงสูงถึงระดับที่ต้องการ และนำผลไปใช้เป็นสารสนเทศสำหรับการตัดสินใจอย่างมีประสิทธิภาพ

บทความนี้ได้สรุปองค์ความรู้ของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด เพื่อให้ผู้สนใจที่จะนำแนวคิดทฤษฎีนี้ประยุกต์ใช้ หรือผู้ที่สนใจศึกษา เข้าใจเนื้อหา สามารถนำไปประยุกต์ได้ง่ายขึ้น จะประกอบด้วยประเด็นความรู้เกี่ยวกับแนวคิดทฤษฎี การวิเคราะห์สัมประสิทธิ์การสรุปอ้างอิง และการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดกับการวิจัยทางการศึกษา

วัตถุประสงค์

1. เพื่ออธิบายแนวคิด หลักการทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด
2. เพื่ออธิบายการวิเคราะห์สัมประสิทธิ์การสรุปอ้างอิง
3. เพื่ออธิบายการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดทางวิจัยการศึกษา

แนวคิดทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด (Generalizability theory: G-Theory) หมายถึงการศึกษาวิเคราะห์ความน่าเชื่อถือของผลการวัดด้วยการให้ความสำคัญกับการออกแบบการวัด (Measurement design) ที่คำนึงถึงแหล่งความคลาดเคลื่อน ที่เป็นระบบอย่างน้อย 1 แหล่ง ระดับความน่าเชื่อถือของผลการวัดที่ได้มีชื่อเรียกเฉพาะว่า “สัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด”

ความเป็นมาของทฤษฎี

ในอดีตปัญหาที่เกิดขึ้นกับการวัดทางจิตวิทยาและการศึกษา คือ การศึกษาความเที่ยงและวิธีการประเมินค่าความเที่ยงของแบบสอบ ซึ่งได้รับการแก้ไขปัญหานั้นโดยใช้หลักการของทฤษฎีการทดสอบแบบดั้งเดิม (Classical test theory) คือ คะแนนของการวัด (X) ได้มาจากคะแนนจริง (T) รวมกับคะแนนจากแหล่งความคลาดเคลื่อน (E) ซึ่งเขียนสมการได้ดังนี้ X = T+E โดยแหล่งความคลาดเคลื่อนถือเป็นการรวมทุกแหล่งความคลาดเคลื่อนเข้าไว้ด้วยกัน ทำให้ไม่สามารถทำการวิเคราะห์ได้ว่าความคลาดเคลื่อนที่เกิดขึ้นจากการวัดมาจากแหล่งใด จึงมีนักทฤษฎีหลายคนพยายามศึกษาแหล่งความคลาดเคลื่อนและวิเคราะห์เพื่อจำแนกแหล่งความคลาดเคลื่อนของการวัด โดยการนำเทคนิคการวิเคราะห์ความแปรปรวน (ANOVA) มาใช้ ซึ่งเป็นการวิเคราะห์ความแปรปรวนทางสถิติที่เป็นการประมาณค่าความเที่ยงของแบบสอบ และในเวลาต่อมา Cronbach *et al.* (1963) ได้พัฒนาทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดขึ้นเป็นครั้งแรก ซึ่งเป็นทฤษฎีที่สามารถวิเคราะห์ความแปรปรวนของแหล่งความคลาดเคลื่อนได้อย่างเป็นระบบ ภายใต้เงื่อนไขและสถานการณ์ของการวัด ประโยชน์ที่ได้รับคือสามารถควบคุมแหล่งความคลาดเคลื่อนของการวัดได้ และยังสามารถประมาณค่าความเที่ยงหรือความเที่ยงของแบบสอบในสถานการณ์ต่าง ๆ ที่ต้องการวัดได้

วารสารมหาวิทยาลัยราชภัฏยะลา
JYRU
JOURNAL OF YALA RAJABHAT UNIVERSITY
Print ISSN: 1905-2383 e-ISSN: 2651-0863

แนวคิดพื้นฐานของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด

Cronbach *et al.* (1972) ได้เสนอทฤษฎีสำหรับการวิเคราะห์ความน่าเชื่อถือของผลการวัดพฤติกรรม (The dependability of behavioral measurement) สำหรับสถานการณ์ของการวัดผลลักษณะต่าง ๆ ซึ่งต่อมาเป็นที่รู้จักในชื่อของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ทฤษฎีนี้ได้ขยายแนวคิดของความเที่ยงในทฤษฎีการทดสอบแบบดั้งเดิม ให้สามารถสรุปความเที่ยงในสถานการณ์หรือเงื่อนไขการวัดในลักษณะต่าง ๆ ได้ ทฤษฎีการทดสอบแบบดั้งเดิมไม่ได้ให้ความสนใจต่อสถานการณ์ หรือเงื่อนไขการวัดและปฏิสัมพันธ์ระหว่างผู้สอบกับเงื่อนไขของการวัดที่สามารถส่งผลต่อความน่าเชื่อถือของผลการวัดและเสนอว่ามีความคลาดเคลื่อนจากการวัดแหล่งเดียว แต่ G-Theory ได้เสนอว่ามีความคลาดเคลื่อนมากกว่าหนึ่งแหล่ง (Multiple error source) โดยแยกส่วนความคลาดเคลื่อนของการวัดเป็นสองแหล่งประกอบด้วยความคลาดเคลื่อนอย่างมีระบบ (Systematic source) และความคลาดเคลื่อนสุ่ม (Random source) การศึกษาความน่าเชื่อถือของผลการวัดตามแนวคิดของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดอยู่บนพื้นฐาน ดังนี้

1. โมเดล ข้อตกลงเบื้องต้น แหล่งความคลาดเคลื่อนหรือแหล่งความแปรปรวน และจำนวนฟาเซทหรือองค์ประกอบ ดังนี้

1.1 โมเดล โมเดลของทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดเป็นโมเดลเชิงสุ่ม ถือว่าเงื่อนไขการวัดที่นำมาศึกษานั้นได้มาจากการสุ่มแบบง่ายหรือสุ่มแบบแบ่งชั้น จึงเป็นทฤษฎีที่ให้สารสนเทศที่สามารถนำไปสรุปอ้างอิงไปยังเงื่อนไขการวัดอื่น ๆ ในเอกภพเดียวกันได้ การประยุกต์ใช้ทฤษฎีนี้โดยการวิเคราะห์ความแปรปรวนศึกษารูปแบบที่เหมาะสมต่อการวัด โดยทำการประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ ด้วยการวิเคราะห์ความแปรปรวน (Analysis of variance: ANOVA)

1.2 ข้อตกลงเบื้องต้น การประมาณค่าความเที่ยงด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ไม่มีข้อตกลงว่า การวัดมีความเท่าเทียมกันในลักษณะของความเป็นคู่ขนาน เป็นทฤษฎีที่ยินยอมให้ออกแบบเก็บรวบรวมข้อมูลให้เหมาะสมกับปัญหาของการวัดผล

1.3 แหล่งความคลาดเคลื่อนหรือแหล่งความแปรปรวน ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดประมาณค่าความคลาดเคลื่อนของการวัดได้หลายแหล่ง โดยประมาณค่าความแปรปรวนได้จากการวิเคราะห์ครั้งเดียว จากนั้นประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ ที่เป็นไปได้ทั้งหมดทั้งจากผลหลัก (Main effect) และผลร่วมหรือปฏิสัมพันธ์ (Interaction effect) ระหว่างองค์ประกอบด้วย เพื่อให้ผลการวิเคราะห์มีความถูกต้อง แม่นยำ และได้สารสนเทศที่เพียงพอสำหรับการตัดสินใจเลือกแบบวัดที่มีประสิทธิภาพ

1.4 จำนวนฟาเซทหรือองค์ประกอบ ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดสามารถศึกษาได้หลายฟาเซท พร้อมกัน รวมทั้งสามารถจำแนกแหล่งความแปรปรวนที่เป็นไปได้ทั้งหมด ได้ค่าความแปรปรวนของความคลาดเคลื่อนที่ถูกต้องทำให้ประมาณค่าความเที่ยงได้ถูกต้องแม่นยำ สามารถปรับรูปแบบการวัด เพื่อให้ได้ค่าความเที่ยงที่ดีขึ้นได้ในหลายสถานการณ์วัด โดยการเพิ่มหรือลดจำนวนฟาเซทต่าง ๆ เช่น จำนวนข้อสอบ จำนวนครั้งในการวัด จำนวนผู้ตรวจหรือจำนวนผู้ประเมิน เป็นต้น

2. ความรู้พื้นฐานเกี่ยวกับโครงสร้างของการศึกษาทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ได้เสนอคำศัพท์ที่ใช้ในทฤษฎีการสรุปอ้างอิง ดังนี้

2.1 ประชากร (Population) หมายถึง สิ่งที่เป็นเป้าหมายของการวัดหรือสิ่ง ที่มุ่งวัดทั้งหมด (Object of measurement) ในสถานการณ์ของการทดสอบโดยทั่วไป สิ่งที่มุ่งวัด ได้แก่ ผู้สอบ บุคคล นักเรียนรายบุคคล หรือนักเรียนทั้งชั้น

2.2 องค์ประกอบ (Facet) หมายถึง ชุดของเงื่อนไขของการวัดที่มีลักษณะ คล้ายกันหรือกลุ่มเงื่อนไขของการวัดเป็นองค์ประกอบที่คาดว่ามีผลต่อการวัดความคลาดเคลื่อน เช่น องค์ประกอบของจำนวนข้อสอบ (Item facet) องค์ประกอบของจำนวนครั้งของการสอบ (Occasion facet) องค์ประกอบของผู้ตรวจข้อสอบ (Rater facet)

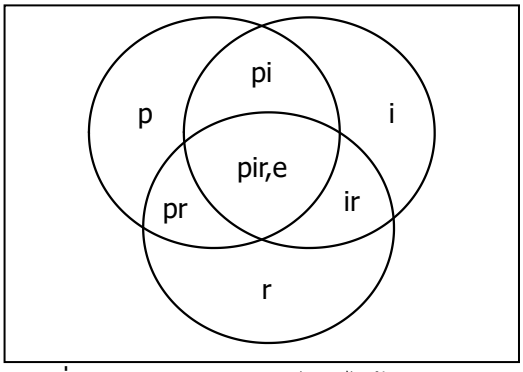
2.3 เงื่อนไขของการวัด (Condition of measurement) หมายถึง ระดับขององค์ประกอบที่ทำให้ได้ค่าสังเกตแต่ละค่าในการวัดครั้งหนึ่ง เช่น องค์ประกอบหรือฟาเซทของจำนวนผู้ตรวจ กำหนดระดับเงื่อนไขการวัดเป็น 1 2 และ 3 คน องค์ประกอบ/ฟาเซทความยาวของแบบสอบ (จำนวนข้อของแบบสอบ) กำหนดระดับเงื่อนไขการวัดเป็น 10 20 และ 30 ข้อ

2.4 รูปแบบความสัมพันธ์ระหว่างองค์ประกอบ หมายถึง องค์ประกอบเป็นสิ่งที่ชี้ให้เห็นแหล่งความแปรปรวนของรูปแบบการวัดในแบบจำลองการวิเคราะห์ความแปรปรวน รูปแบบความสัมพันธ์แบ่งออกเป็น 3 รูปแบบ ดังนี้

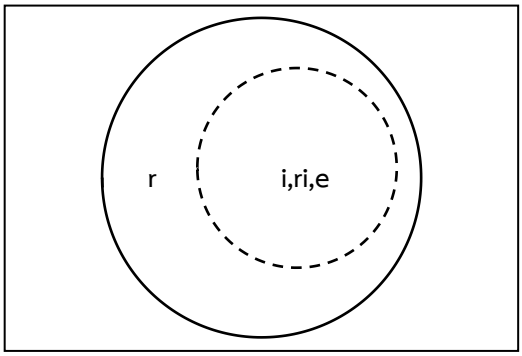
1) ความสัมพันธ์แบบไขว้ (Crossed) คือ ลักษณะความสัมพันธ์ที่แต่ละระดับของสิ่งที่ถูกวัดนั้น ถูกวัดภายใต้เงื่อนไขเดียวกันทั้งหมด ใช้สัญลักษณ์ “x” คือ “Crossed with” เช่น กำหนดรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น p x i x r หมายถึง ผู้สอบ (p) ทำข้อสอบ (i) ทุกข้อ และผู้ตรวจ (r) ตรวจข้อสอบทุกข้อของผู้สอบทุกคน

2) ความสัมพันธ์แบบแฝง (Nested) คือ ลักษณะความสัมพันธ์ที่แต่ละระดับของสิ่งที่ถูกวัดนั้น ถูกวัดภายใต้เงื่อนไขที่แตกต่างกัน ใช้สัญลักษณ์ “:” คือ “Nested within” เช่น กำหนดรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น i : r หมายถึง ผู้ตรวจ (r) ตรวจข้อสอบ (i) ต่างข้อกัน

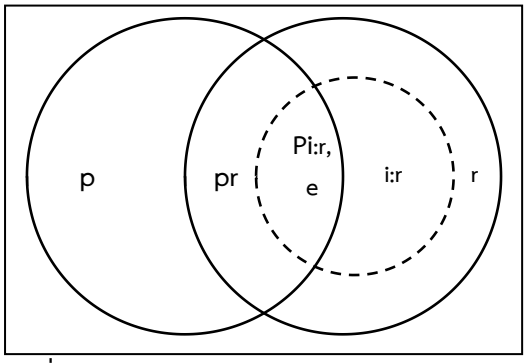
3) ความสัมพันธ์แบบผสม (Confounded) คือ ลักษณะความสัมพันธ์ที่มีทั้งความสัมพันธ์แบบไขว้และความสัมพันธ์แบบแฝงรวมกันอยู่ เช่น กำหนดรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น p x (i : r) หมายถึง ผู้สอบ (p) ทำข้อสอบ (i) บางข้อที่อยู่ในแบบสอบ (r)



ภาพที่ 1 รูปแบบความสัมพันธ์แบบไขว้ (Crossed)
(ที่มา : Kanjanawasee, 2020)



ภาพที่ 2 รูปแบบความสัมพันธ์แบบแฝง (Nested)
(ที่มา : Kanjanawasee, 2020)



ภาพที่ 3 รูปแบบความสัมพันธ์แบบผสม (Confounded)
(ที่มา : Kanjanawasee, 2020)

2.5 เอกภพ (Universe) หมายถึงเงื่อนไขการวัด (Condition of measurement) ทั้งหมด โดยกลุ่มเงื่อนไขการวัดเรียกว่า ฟาเซท (Facet) ซึ่งเป็นองค์ประกอบที่คาดว่า มีผลต่อความคลาดเคลื่อนของการวัด เช่น ความยาวของแบบสอบ รูปแบบของข้อสอบ จำนวนครั้งของการสอบ จำนวนผู้ตรวจ

2.6 คะแนนเอกภพ (Universe score) หมายถึง คะแนนของเงื่อนไขการวัด ทั้งหมดของทุกองค์ประกอบซึ่งคล้ายกับคะแนนจริง (Ideal score) ของสิ่งที่วัดในทฤษฎีการทดสอบแบบดั้งเดิม

2.7 เอกภพของการสังเกตที่ยอมรับได้ (University of admission observation) หมายถึง กลุ่มเงื่อนไขของการวัด ที่เป็นไปได้ซึ่งสามารถวัดหรือสังเกตได้ในแต่ละฟาเซท เช่น ฟาเซท ข้อสอบ และฟาเซทของผู้ตรวจ เป็นต้น

2.8 เอกภพการสรุปอ้างอิง (Universe of generalization) หมายถึง สถานการณ์หรือเงื่อนไขการวัดอันเป็นเป้าหมายของการสรุปอ้างอิง กล่าวได้ว่า เป็นการวัดครอบคลุมเงื่อนไขที่สนใจทั้งหมด หรือเงื่อนไขในเอกภพของการสังเกตที่ยอมรับได้ทั้งหมด ซึ่งอาจประกอบด้วยเซ็ตย่อยของเงื่อนไขในเอกภพของการสังเกตได้

3. การศึกษาเพื่อการสรุปอ้างอิง (Generalizability study : G-study) และการศึกษาเพื่อการตัดสินใจ(Decision study: D-study) ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ประกอบด้วยการศึกษา 2 ขั้นตอนที่สำคัญคือ การศึกษาเพื่อการสรุปอ้างอิง การศึกษาเพื่อการตัดสินใจ มีรายละเอียดดังนี้

วารสารมหาวิทยาลัยราชภัฏยะลา
JYRU
JOURNAL OF YALA RAJABHAT UNIVERSITY
Print ISSN: 1905-2383 e-ISSN: 2651-0863

3.1 การศึกษาเพื่อการสรุปอ้างอิง หมายถึง การวิเคราะห์สรุปอ้างอิง เพื่อประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ ของรูปแบบการวัด โดยเริ่มจากการกำหนดองค์ประกอบต่าง ๆ กำหนดความสัมพันธ์ขององค์ประกอบ และกำหนดเอกภพของการสังเกตที่ยอมรับได้และการประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ

3.2 การศึกษาเพื่อการตัดสินใจ หมายถึง การศึกษาเพื่อประเมินวิธีการวัดที่เหมาะสม และตัดสินใจ ซึ่งเป็นขั้นสุดท้ายของการวิเคราะห์การสรุปอ้างอิง เริ่มจากระบุสิ่งที่ต้องการวัด เอกภพการสรุปอ้างอิง จำนวนเงื่อนไขของการวัด หรือขนาดกลุ่ม ตัวอย่างของแต่ละองค์ประกอบ และรูปแบบของการวัด จากนั้นคำนวณค่าความแปรปรวนจากการศึกษาสัมประสิทธิ์หรือความเที่ยงของแบบสอบ ขนาดกลุ่มตัวอย่างเพื่อการตัดสินใจไม่จำเป็นต้องเท่ากับขนาดกลุ่มตัวอย่างในการศึกษาเพื่อการสรุปอ้างอิง แต่ขึ้นอยู่กับการนำไปใช้เพื่อการตัดสินใจ เกี่ยวกับสิ่งที่วัด ซึ่งกำหนดโดยมีกลุ่มตัวอย่างในการศึกษาเพื่อการสรุปอ้างอิงเป็นฐาน

4. ความแปรปรวนของคะแนนความคลาดเคลื่อน ความแปรปรวนของคะแนนความคลาดเคลื่อนจากองค์ประกอบต่าง ๆ ของการวัด แบ่งออกเป็น 2 ประเภท คือ ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (Absolute error variance: σ^2_{Abs} หรือ σ^2_{Δ}) และความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ (Relative error variance: σ^2_{Rel} หรือ σ^2_{δ}) มีรายละเอียด ดังนี้

4.1 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (Absolute error variance: σ^2_{Abs} หรือ σ^2_{Δ}) หมายถึง ความแปรปรวนของผลต่างระหว่างคะแนนสังเกตได้กับคะแนนเอกภพ ($\mu_p - X_p$) ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่ง ต่างๆ ยกเว้นความแปรปรวนของแหล่งบุคคล ($\sigma^2_{\mu p}$ หรือ σ^2_p)

4.2 ความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ (Relative error variance: σ^2_{Rel}หรือ σ^2_{δ}) หมายถึง ความแปรปรวนของผลต่างระหว่างคะแนนสังเกตได้กับคะแนนเอกภพ ($\mu_p - X_p$) ซึ่งคำนวณจากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่าง ๆ ที่มีปฏิสัมพันธ์ร่วมกับผู้สอบ (p)

5. สัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient: ρ^2) หมายถึง อัตราส่วนของความแปรปรวนของคะแนนเอกภพ และความแปรปรวนของคะแนนที่สังเกตที่คาดหวัง ซึ่งค่าสัมประสิทธิ์การสรุปอ้างอิงคำนวณได้จากกำลังสองของค่าสหสัมพันธ์ระหว่างคะแนนเอกภพกับคะแนนสังเกตได้โดยสัมประสิทธิ์การสรุปอ้างอิงแบ่งเป็น 2 ประเภทคือ สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์กับสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ดังนี้

5.1 สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์หรือการวัดแบบอิงกลุ่ม เป็นความคลาดเคลื่อนในการวัดแบบสัมพัทธ์ค่าสัมประสิทธิ์การสรุปอ้างอิงประมาณค่า ได้จากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพและผลบวกของความแปรปรวนของคะแนนเอกภพกับความแปรปรวนของความคลาดเคลื่อนสัมพัทธ์

5.2 สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์หรือการวัดแบบอิงเกณฑ์ เป็นความคลาดเคลื่อนในการวัดแบบสัมบูรณ์ค่าสัมประสิทธิ์การสรุปอ้างอิงประมาณ ค่าได้จากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพและผลบวกของความแปรปรวนของคะแนนเอกภพกับความแปรปรวนของความคลาดเคลื่อนสัมบูรณ์

การวิเคราะห์สัมประสิทธิ์การสรุปอ้างอิง

ขั้นตอนการวิเคราะห์สัมประสิทธิ์การสรุปอ้างอิงออกเป็น 2 ขั้นตอน หลักการดำเนินงานแต่ละขั้นตอนมีดังต่อไปนี้ (Cronbach *et al.*, 1972)

ขั้นตอนที่หนึ่ง การศึกษาเพื่อการสรุปอ้างอิง (Generalizability study: G-study) เป็นกระบวนการที่มุ่งประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ ภายใต้เอกภพของคะแนนสังเกตที่ยอมรับได้โดยมีขั้นตอนดังนี้

1) กำหนดสิ่งที่จะวัด/กำหนดสิ่งที่ต้องการวัด เช่น ความสามารถด้านการคิดวิเคราะห์ของนักเรียน

2) กำหนดฟาเซทของการวัด เช่น กำหนดให้มี 2 ฟาเซท คือ จำนวนข้อสอบและจำนวนผู้ตรวจให้คะแนน

3) กำหนดเอกภพของการสังเกตที่ยอมรับได้ ประกอบด้วย ฟาเซทข้อสอบ และฟาเซทผู้ตรวจให้คะแนน

4) กำหนดความสัมพันธ์ระหว่างฟาเซทว่าเป็นแบบ Crossed แทนด้วย “x” หรือ Nested แทนด้วย “:” ถ้าต้องการศึกษาให้ผู้ตรวจทุกคนตรวจข้อสอบทุกข้อจะเป็นการศึกษาแบบ Crossed คือ i x r ถ้าให้ผู้ตรวจแต่ละคนตรวจชุดข้อสอบต่างกันจะเป็นการศึกษาแบบ Nested design คือ i : r ทั้งนี้ แล้วแต่ความสนใจของผู้ศึกษา ซึ่งโดยหลักการแล้ว ถ้าเป็นแบบ Crossed design จะทำให้ได้จำนวนแหล่งความแปรปรวนมากกว่าแบบ Nested design ซึ่งจะทำให้ทราบว่า ความแปรปรวน

แหล่งใดมีผลต่อสัมประสิทธิ์การสรุปอ้างอิง และช่วยเลือกรูปแบบที่เหมาะสมในขั้น D-study และได้ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีค่ามากกว่า Nested design ในกรณีของ 1-facet design แต่ในการปฏิบัตินั้นสามารถใช้ได้ง่ายกว่า

5) การเก็บรวบรวมข้อมูล โดยการสุ่มตัวอย่างมาศึกษา เช่น การศึกษาหนึ่ง สุ่มตัวอย่างนักเรียน n_p คน สุ่มตัวอย่างข้อสอบ n_iข้อ และสุ่มตัวอย่างผู้ตรวจ n_r คน จะเป็นการศึกษา แบบ G-study ในกรณีนี้ประกอบด้วยแหล่งความแปรปรวน 7 แหล่ง เป็นผลหลัก (Main effect) 3 แหล่ง คือ ผลจากนักเรียน ผลจากข้อสอบ และผลจากผู้ตรวจ และเป็นผลของปฏิสัมพันธ์ (Interaction effect) 4 แหล่ง คือ ปฏิสัมพันธ์ระหว่างนักเรียนกับข้อสอบ นักเรียนกับผู้ตรวจข้อสอบ ข้อสอบกับข้อสอบ นักเรียนกับแบบสอบและผู้ตรวจ จะเห็นว่ามีผลหลัก (Main effect) อยู่ 3 ค่า คือ ผลของนักเรียน (p) ผลของข้อสอบ (i) และผลของผู้ตรวจ (r) ส่วนผลร่วมหรือปฏิสัมพันธ์ (Interaction effect) 2 ระดับ 3 ค่า คือ ผลร่วมของผู้สอบและข้อสอบ (pi) ผลร่วมของผู้สอบและผู้ตรวจ (pr) ผลร่วมของข้อสอบและผู้ตรวจ (ir) และผลร่วมทั้งหมดของผู้สอบ ข้อสอบ และผู้ตรวจ (pir) ค่าความแปรปรวนของแหล่งต่าง ๆ คำนวณจากค่ากำลังสองเฉลี่ย (Mean square) โดยวิเคราะห์ ความแปรปรวนแบบ factorial design p x i x r ในการประมาณค่าความแปรปรวนจากค่าเฉลี่ยกำลังสองสูตรคำนวณใช้ค่าเฉลี่ยของกำลังสองของความเบี่ยงเบน (Mean square of deviation) หรือ MS(α) โดยคำนวณจาก MS(α) = $\frac{SS(\alpha)}{df(\alpha)}$

เมื่อ $df(\alpha)$ แทน ชั้นแห่งความเป็นอิสระที่สอดคล้องกับผลขององค์ประกอบ

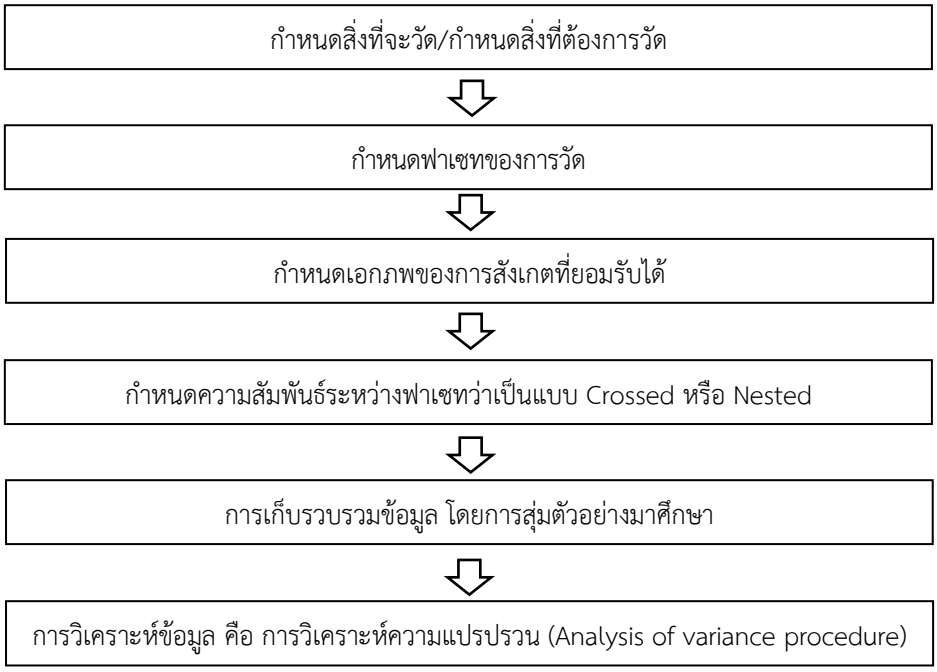
$SS(\alpha)$ แทน ผลขององค์ประกอบที่ต้องการศึกษา

MS(α) แทน ค่าเฉลี่ยกำลังสองของค่าความเบี่ยงเบน

Effect(α) แทน ผลรวมกำลังสองของค่าเบี่ยงเบน

$\sigma^2_{(\alpha)}$ แทน ค่าประมาณความแปรปรวน

6) การวิเคราะห์ข้อมูล หลังจากผ่านขั้นตอนในการพิจารณารูปแบบการวัดและเก็บรวบรวมข้อมูลจากรูปแบบที่กำหนดแล้ว ข้อมูลที่ได้จะนำสู่กระบวนการวิเคราะห์ทางสถิติเพื่อประมาณขนาดของความคลาดเคลื่อนของการวัดจากแหล่งต่าง ๆ กระบวนการทางสถิติที่ใช้ในทฤษฎีการสรุปอ้างอิง คือ การวิเคราะห์ความแปรปรวน (Analysis of variance procedure) โดยวิเคราะห์ในขั้นตอนของการศึกษา G-study



ภาพที่ 4 การศึกษาเพื่อการสรุปอ้างอิง (Generalizability study : G-study)
(ที่มา : Kongprapan, 2022)

วารสารมหาวิทยาลัยราชภัฏยะลา
JYRU
JOURNAL OF YALA RAJABHAT UNIVERSITY
Print ISSN: 1905-2383 e-ISSN: 2651-0863

ขั้นตอนที่สอง การศึกษาเพื่อการตัดสินใจเป็นการนำผลที่ได้จาก G-study มาวิเคราะห์เพื่อการตัดสินใจเลือกสถานการณ์ของการวัด โดยมีขั้นตอน ดังนี้

1) กำหนดเอกภพของการสรุปอ้างอิงที่ผู้ตัดสินใจต้องการสรุปอ้างอิงใน D-study ซึ่งเอกภพของการสรุปอ้างอิงประกอบด้วยเงื่อนไขทั้งหมดหรือเป็นเพียงเงื่อนไขย่อยในเอกภพ การสังเกตที่ยอมรับได้ผู้วิจัยต้องเลือกว่าจะใช้โมเดลรูปแบบใดเป็นโมเดลแบบสุ่ม (Random model) โมเดลแบบกำหนด (Fixed model) หรือโมเดลแบบผสม (Mixed model) ซึ่งโดยหลักการแล้วโมเดลแบบสุ่มจะสามารถสรุปอ้างอิงได้กว้างกว่าโมเดลแบบกำหนดและโมเดลแบบผสม

2) การกำหนดขนาดกลุ่มตัวอย่าง (Sample size) จำนวนเงื่อนไขของตัวอย่าง ฟาเซทใน G-study ไม่จำเป็นต้องเป็นตัวอย่างใน D-study ทั้งนี้การกำหนดขนาดกลุ่มตัวอย่างขึ้นกับความสนใจของผู้วิจัย

3) โครงสร้างของแบบที่ศึกษา (Design structure) ใน D-study อาจใช้โครงสร้างของแบบที่ศึกษาเหมือนกันหรือต่างจากใน G-study ก็ได้เช่น ใน D-study ตัดสินใจให้นักเรียนทุกคนทำข้อสอบ n_iข้อ เหมือนกันจะมีโครงสร้างเป็น D-study ในลักษณะ p x i x r เหมือนโครงสร้างใน G-study ที่กล่าวมาแล้ว แต่ถ้าตัดสินใจให้นักเรียนทุกคนทำข้อสอบทุกข้อโดยผู้ตรวจแต่ละคนทำการตรวจชุดของข้อสอบต่างกัน จะได้โครงสร้างเป็น D-study ในลักษณะ p x (i : r)

4) การประมาณค่าความแปรปรวน (Estimate D-study variance component) ในการศึกษาเพื่อตัดสินใจขึ้นอยู่กับนักวิจัย ดังนั้นต้องประมาณค่าความแปรปรวน โดยอาศัยผลจากการประมาณค่าในขั้นการศึกษาเพื่อการสรุปอ้างอิง ซึ่งสอดคล้องกับแบบแผนและขนาดตัวอย่างที่ต้องการตัดสินใจ

5) ประมาณค่าความแปรปรวนของความคลาดเคลื่อน นอกจากการประมาณค่าองค์ประกอบความแปรปรวน (Variance component) ตามรูปแบบและขนาดกลุ่มตัวอย่างที่ต้องการแล้ว จากนั้นประมาณค่าความแปรปรวนของความคลาดเคลื่อน โดยความแปรปรวนของความคลาดเคลื่อน แบ่งออกเป็น 2 ประเภท คือ

5.1) การประมาณค่าความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ (Absolute error variance: σ^2_{Abs} หรือ σ^2_{Δ}) เป็นความแปรปรวนของผลต่างระหว่างคะแนนสังเกตได้กับคะแนนเอกภพ ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่าง ๆ ยกเว้น ความแปรปรวนของแหล่งบุคคล ($\sigma^2_{\mu p}$ หรือ σ^2_{p}) เช่น การออกแบบ p x i x r design โดยมีสูตร ดังนี้

$$\sigma^2_{Abs} = \sigma^2_{i} + \sigma^2_{r} + \sigma^2_{pi} + \sigma^2_{pr} + \sigma^2_{ir} + \sigma^2_{pir,e}$$

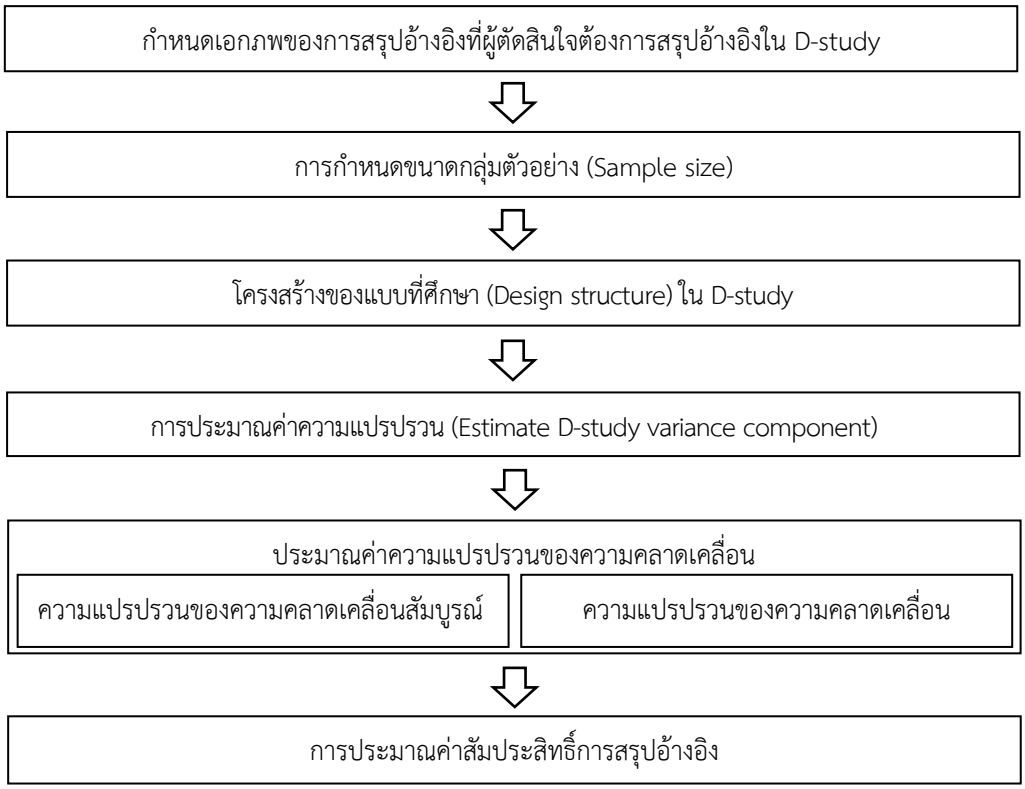
5.2) การประมาณค่าความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ (Relative error variance: σ^2_{Rel} หรือ σ^2_{δ}) หมายถึง ความแปรปรวนของผลต่างระหว่างคะแนนสังเกตได้กับคะแนนเอกภพ ซึ่งคำนวณได้จากผลรวมของความแปรปรวนของคะแนนจากแหล่งต่าง ๆ ที่มีปฏิสัมพัทธ์ร่วมกับผู้สอบ (p) โดยมีสูตรดังนี้

$$\sigma^2_{Rel} = \sigma^2_{pi} + \sigma^2_{pr} + \sigma^2_{pir,e}$$

6) การประมาณค่าสัมประสิทธิ์การสรุปอ้างอิง เป็นขั้นสุดท้ายในการศึกษาเพื่อตัดสินใจ เป็นดัชนีบอกความน่าเชื่อถือของผลการวัด ซึ่งคำนวณจากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพและความแปรปรวนของคะแนนคาดหวังที่สังเกตได้ ซึ่งค่าสัมประสิทธิ์การสรุปอ้างอิง (G-coefficient) สำหรับการตัดสินใจเชิงสัมบูรณ์ (ρ^2_{Abs}หรือρ^2_{Δ}) และสำหรับการตัดสินใจเชิงสัมพัทธ์ (ρ^2_{Rel}หรือρ^2_{δ}) สามารถคำนวณได้จากสูตร ดังนี้

6.1) สูตรคำนวณสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ $\rho^2_{Abs} = \frac{\sigma^2_{\rho}}{\sigma^2_{\rho}+\sigma^2_{Abs}}$

6.2) สูตรคำนวณสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ $\rho^2_{Rel} = \frac{\sigma^2_{\rho}}{\sigma^2_{\rho}+\sigma^2_{Rel}}$



ภาพที่ 5 การศึกษาเพื่อการตัดสินใจ
(ที่มา : Kongprapan, 2022)

การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดกับการวิจัยทางการศึกษา

ปัจจุบันมีผู้วิจัยประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดมากขึ้น จากงานวิจัยปีล่าสุดที่น่าสนใจสามารถแบ่งออกเป็น 4 หัวข้อ ดังนี้

1. บริบทการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด ผู้วิจัยส่วนใหญ่ประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดกับบริบททางการศึกษา แบ่งเป็น 4 ลักษณะ ได้แก่ การพัฒนารูปแบบการประเมิน (Srikaew, 2014) การพัฒนามาตรวัด (Donmuang, 2018) การเปรียบเทียบความเที่ยงของแบบสอบ (Atilla & Ezel, 2010; Gebril, 2010; Nuntapanich, 2013; Umnacil, 2014; Sanguanwai, 2015; Praserthlap, 2015; Phadungphon, 2016) และความสอดคล้องในแนวเดียวกันระหว่างข้อสอบ (Janprasert, 2017)

2. วัตถุประสงค์ของการศึกษาเชิงสรุปอ้างอิง (Generalizability study) แบ่งเป็น 3 ลักษณะ ได้แก่ 1) เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมิน/คะแนนแบบสอบ (Gebril, 2010; Nuntapanich, 2013; Srikaew, 2014; Praserthlap, 2015) 2) เพื่อเปรียบเทียบความเที่ยงของแบบสอบ (Atilla and Ezel, 2010; Umnacil, 2014; Phadungphon, 2016; Donmuang, 2018) และ 3) เพื่อประมาณค่าความแปรปรวนขององค์ประกอบที่มีผลต่อสัมประสิทธิ์การสรุปอ้างอิง (G-coefficient) ของแบบสอบ (Sanguanwai, 2015)

3. การออกแบบการวัด ผู้วิจัยเลือกใช้การออกแบบ 2 ลักษณะ ได้แก่ 1) การออกแบบการวัดแบบไขว้ (Gebril, 2010; Umnacil, 2014; Srikaew, 2014; Donmuang, 2018) 2) การออกแบบการวัดแบบไขว้ และการวัดแบบแฝง (Janprasert, 2017; Nuntapanich, 2013) และจำนวนเงื่อนไข ตั้งแต่ 1–3 แหล่ง (Facets) จำนวน 1 แหล่ง (Facets) (Umnacil, 2014; Kaewprasert, 2015; Donmuang, 2018) จำนวน 2 แหล่ง (Facets) (Srikaew, 2014; Sanguanwai, 2015; Praserthlap, 2015; Janprasert, 2017; Nuntapanich, 2013) และจำนวน 3 แหล่ง (Facets) (Phadungphon, 2016)

4. ข้อค้นพบจากการวิจัย ผู้วิจัยมีข้อค้นพบแต่ละการวิจัยแตกต่างกันที่น่าสนใจ ได้แก่ Gebril (2010) ศึกษาผลการเปรียบเทียบแบบแผนการให้คะแนนของแบบสอบการเขียนภาษาอังกฤษที่ต่างกันระหว่างรูปแบบการตรวจให้คะแนน มีข้อ

วารสารมหาวิทยาลัยราชภัฏยะลา
JYRU
JOURNAL OF YALA RAJABHAT UNIVERSITY
Print ISSN: 1905-2383 e-ISSN: 2651-0863

ค้นพบว่า การให้คะแนนที่มีผู้ตรวจคนเดียวกันและผู้ตรวจต่างกัน ได้ผลในทำนองเดียวกัน คือ เมื่อผู้ตรวจเพิ่มขึ้น จำนวนชิ้นงานเพิ่มขึ้น ค่า G-coefficient และ Phi-coefficient จะเพิ่มขึ้นด้วย

Srikaew (2014) ศึกษาสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของรูปแบบการประเมินทักษะการพูดภาษาอังกฤษของนักเรียนชั้นประถมศึกษาปีที่ 6 โดยใช้แฟ้มสะสมงานด้วยทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือแบบพหุตัวแปรมีข้อค้นพบว่า หากผู้บริหารการประเมินต้องการเพิ่มความเที่ยงในการประเมินทักษะการพูดภาษาอังกฤษของนักเรียน ชั้นประถมศึกษาปีที่ 6 เช่น การออกแบบการประเมินระดับโรงเรียน หรือเพื่อการแข่งขัน หรือสำหรับการประเมินในระดับชาติ ผู้บริหารการประเมินสามารถใช้ภาระงานทั้ง 4 ภาระงาน ได้แก่ การสัมภาษณ์ การนำเสนอปากเปล่า การเล่านิทาน และการบรรยายภาพโดยใช้เกณฑ์การให้คะแนนแบบแยกส่วนที่ประกอบด้วยเกณฑ์การให้คะแนน 6 ด้าน ได้แก่ คำศัพท์ การสร้างประโยค การเชื่อมโยงเนื้อหา การออกเสียง การบรรลุวัตถุประสงค์ และความคล่องแคล่ว เพื่อให้ได้ระดับความเที่ยงที่เชื่อถือได้ ผู้บริหารการประเมินควรใช้วิธีการเพิ่มจำนวนภาระงานมากกว่าการเพิ่มผู้ประเมิน เนื่องจากการเพิ่มภาระงานทำให้มีค่าใช้จ่ายน้อยกว่าการเพิ่มจำนวน ผู้ประเมิน แต่ยังได้ระดับความเที่ยงที่สูงและเชื่อถือได้เช่นกัน ในส่วนของการบริหารการประเมินทักษะการพูดภาษาอังกฤษนั้นเพื่อให้เกิดความสะดวกและเป็นไปได้ในทางปฏิบัติรวมถึงประหยัดค่าใช้จ่าย และแรงงาน ผู้บริหารการประเมินควรใช้รูปแบบการประเมินที่ผู้ประเมิน มีรูปแบบการให้คะแนนที่ต่างกัน กล่าวคือ ในภาระงานที่ต่างกันเมื่อใช้ผู้ประเมินที่ต่างกัน หรือผู้ประเมินไม่จำเป็นต้องตรวจให้คะแนนนักเรียนในทุกภาระงาน ซึ่งพบว่า รูปแบบการประเมินนี้ทำให้เกิดความคลาดเคลื่อนน้อยลงและทำให้มีความเที่ยงในการประเมินสูงขึ้น จึงเป็นรูปแบบการประเมินที่มีความเหมาะสมและสามารถนำมาใช้ได้อย่างมีประสิทธิภาพ

Umnacil (2014) ศึกษาเปรียบเทียบความเที่ยงของแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่ใช้วิธีการตรวจให้คะแนนที่แตกต่างกันภายใต้จำนวนเหตุการณ์ที่ต่างกัน : การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด มีข้อค้นพบว่า กรณีที่ครูต้องการใช้วิธีการตรวจให้คะแนนตามแบบของ Knox ตรวจให้คะแนนแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่มีค่าความเที่ยงตั้งแต่ 0.7 ขึ้นไป สามารถใช้แบบสอบที่มีจำนวนเหตุการณ์ตั้งแต่ 7 เหตุการณ์ขึ้นไป กรณีที่ครูต้องการใช้วิธีการตรวจให้คะแนนตามแบบของ Feletti ตรวจให้คะแนนแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่มีค่าความเที่ยงตั้งแต่ 0.7 ขึ้นไป สามารถใช้แบบสอบที่มีจำนวนเหตุการณ์ตั้งแต่ 13 เหตุการณ์ขึ้นไป และกรณีที่ครูต้องการใช้วิธีการตรวจให้คะแนนแบบวิเคราะห์ย่อยตรวจให้คะแนนแบบสอบอัตนัยประยุกต์วัดความสามารถในการแก้ปัญหาทางวิทยาศาสตร์ที่มีค่าความเที่ยงตั้งแต่ 0.7 ขึ้นไป สามารถใช้แบบสอบที่มีจำนวนเหตุการณ์ตั้งแต่ 5 เหตุการณ์ขึ้นไป

Sanguanwai (2015) ศึกษาเปรียบเทียบความเที่ยงของแบบสอบวัดความสามารถในการแก้ปัญหา อย่างสร้างสรรค์ทางคณิตศาสตร์ : การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด มีข้อค้นพบว่า การตรวจให้คะแนนแบบสอบวัดความสามารถในการแก้ปัญหาอย่างสร้างสรรค์ที่ผู้วิจัยพัฒนาขึ้นควรใช้วิธีการตรวจให้คะแนนแบบผู้ตรวจตรวจแบบสอบเฉพาะข้อของผู้สอบทุกคน (p x (i : r)) เนื่องจากจะส่งผลต่อค่าความเที่ยงสูงกว่าการตรวจให้คะแนนแบบผู้ตรวจตรวจข้อสอบทุกข้อของผู้สอบทุกคน (p x i x r) เนื่องจากการตรวจจำนวนข้อที่น้อยลงจะเพิ่มประสิทธิภาพการตรวจให้คะแนน และเมื่อเพิ่มจำนวนผู้ตรวจและจำนวนข้อสอบจะทำให้มีค่าความเที่ยงเพิ่มขึ้นตามลำดับ ส่วนเลือกวิธีการตรวจให้คะแนนแบบผู้ตรวจตรวจข้อสอบทุกข้อของผู้สอบทุกคน (p x i x r) เพื่อให้มีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (G-coefficient for relative decisions : $\rho^2\delta$) มากกว่า 0.80 ควรเลือกใช้ในสถานการณ์ที่มีจำนวนผู้ตรวจ 1 คน ตรวจให้คะแนนแบบสอบจำนวน 7 ข้อ และสำหรับการเลือกวิธีการตรวจให้คะแนนแบบผู้ตรวจตรวจข้อสอบเฉพาะข้อของผู้สอบทุกคน (p x (i : r)) ควรเลือกใช้ในสถานการณ์ที่มีจำนวนผู้ตรวจ 2 คน ตรวจให้คะแนน จำนวน 5 ข้อ ซึ่งการพิจารณาค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์จะทำให้สามารถพิจารณาความสามารถของผู้สอบได้ว่า ผู้สอบคนใดมีความสามารถในการแก้ปัญหาอย่างสร้างสรรค์อยู่ในระดับใด และการเลือกวิธีการตรวจให้คะแนนแบบผู้ตรวจตรวจข้อสอบทุกข้อของผู้สอบทุกคนเพื่อให้มีค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (G-coefficient for relative decisions : $\rho^2\Delta$) มากกว่า 0.80 ควรเลือกใช้ในสถานการณ์ที่มีจำนวนผู้ตรวจ 3 คน ตรวจให้คะแนน แบบสอบจำนวน 13 ข้อ และสำหรับการเลือกวิธีการตรวจให้คะแนนแบบผู้ตรวจตรวจข้อสอบเฉพาะข้อของผู้สอบทุกคน ควรเลือกใช้ในสถานการณ์ที่มีจำนวนผู้ตรวจ 3 คน ตรวจให้คะแนนแบบสอบ จำนวน 7 ข้อ ซึ่งการพิจารณาค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์จะทำให้สามารถพิจารณาความสามารถของผู้สอบได้ว่า ผู้สอบคนใดมีความสามารถในการแก้ปัญหาอย่างสร้างสรรค์อยู่ในระดับสูงเพื่อใช้เป็นข้อมูลสำหรับการเข้าร่วมแข่งขันทักษะกระบวนการแก้ปัญหาอย่างสร้างสรรค์ทางคณิตศาสตร์ได้

Janprasert (2017) ศึกษาความสอดคล้องในแนวเดียวกันระหว่างข้อสอบในการประเมินระดับชาติกับข้อสอบในการประเมินระดับชั้นเรียน กลุ่มสาระการเรียนรู้วิทยาศาสตร์ : การประยุกต์ใช้โมเดลหลายองค์ประกอบของราส์ช และทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด มีข้อค้นพบว่า องค์กรหรือหน่วยงาน เช่น สพฐ. สทศ. สำนักงานเขตพื้นที่การศึกษา นักวิชาการ/ครู/ผู้บริหาร หรือผู้ที่สนใจศึกษาความสอดคล้องในแนวเดียวกันระหว่างมาตรฐานกับตัวชี้วัดกับข้อสอบในการประเมินชั้นเรียน ควรออกแบบการประเมินโดยใช้ผู้ประเมินข้อสอบในแต่ละข้อ อย่างน้อย 2 คน ใน การประเมินระดับความซับซ้อนทางปัญญา และใช้ผู้ประเมิน อย่างน้อย 3 คน ในการประเมินระดับความสอดคล้องฯ โดยใช้มาตรประมาณค่า ซึ่งจะให้ค่าความเที่ยงในการประเมินตั้งแต่ 0.70 ขึ้นไป อย่างไรก็ตาม หากเป็นการศึกษาความสอดคล้องในแนวเดียวกันของข้อสอบในระดับที่สูงกว่าข้อสอบที่ประเมินระดับชั้นเรียน ผู้วิจัยสามารถเลือกใช้เกณฑ์ความน่าเชื่อถือของการวัดที่มีความเหมาะสมกับงานวิจัย เช่น ข้อสอบประเมินระดับเขตพื้นที่ หรือระดับชาติ อาจกำหนดเกณฑ์ในการตัดสินความน่าเชื่อถือของผลการวัดที่มีความเข้มงวดมากขึ้น เช่น เกณฑ์ค่าความเที่ยง 0.80 (Sattler, 1998) ซึ่งจากผลการวิจัยต้องใช้ผู้ประเมินข้อสอบในแต่ละข้อ อย่างน้อย 5 คน หรือเกณฑ์ค่าความเที่ยง 0.90 (Gay, 1992) ซึ่งจากผลการวิจัยต้องใช้ผู้ประเมินอย่างน้อย 8 คน

Donmuang (2018) ศึกษาการสร้างมาตรวัดการเผชิญปัญหาและฟันฝ่าอุปสรรคตามแนววิถีพุทธของนักเรียนอาชีวศึกษา มีข้อค้นพบว่า ผลของความเที่ยง (Reliability) ของมาตรวัดการเผชิญปัญหาและฟันฝ่าอุปสรรคตามแนววิถีพุทธของนักเรียนอาชีวศึกษา ดำเนินการภายใต้กรอบทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือผลการวัด การศึกษาขั้นตัดสินใจพบว่าเมื่อข้อคำถามมากขึ้นค่าสัมประสิทธิ์การสรุปอ้างอิงจะสูงขึ้นตามไปด้วย

การประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดกับการวิจัยทางการศึกษา เพื่อนำผลการประยุกต์ใช้ไปเป็นสารสนเทศสำหรับการตัดสินใจอย่างมีประสิทธิภาพนั้น ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด มักประยุกต์ใช้กับการบริหารการสอบ ซึ่งแบบสอบเป็นแบบสอบอัตนัย ผู้วิจัยควรคำนึงว่า ความเป็นอัตนัยของการให้คะแนน คะแนนที่ได้จะแปรเปลี่ยนไปตามลักษณะของผู้ตรวจ เพื่อให้การตรวจแบบสอบมีความเที่ยงเพิ่มขึ้น ผู้ตรวจควรปฏิบัติ อย่างเคร่งครัดดังนี้ 1) ใช้วิธีการตรวจที่เหมาะสมสามารถขจัดอคติได้ 2) การตรวจต้องมุ่งสนใจอยู่เฉพาะประเด็น ที่เกี่ยวข้องกับผลการเรียนที่มุ่งวัดเท่านั้น 3) ระมัดระวังอย่าให้นิสัยส่วนตัวของผู้ตรวจมีอิทธิพลต่อคะแนน 4) ต้องใช้เกณฑ์การตรวจกับนักเรียนทุก ๆ คนอย่างคงเส้นคงวา และเมื่อเพิ่มจำนวนผู้ตรวจจะทำให้ความเที่ยงสูงขึ้น และการตรวจให้คะแนนของผู้ตรวจ ควรใช้วิธีการตรวจให้คะแนนแบบผู้ตรวจตรวจข้อสอบเฉพาะข้อของผู้สอบทุกคน เพราะจะส่งผลต่อค่าความเที่ยงสูงกว่าการตรวจให้คะแนนแบบผู้ตรวจตรวจข้อสอบทุกข้อของผู้สอบทุกคน เนื่องจากการตรวจจำนวนข้อที่น้อยลงจะเพิ่มประสิทธิภาพการตรวจให้คะแนน

สรุป

ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด เป็นทฤษฎีที่ขยายแนวคิดมาจากทฤษฎีการทดสอบแบบดั้งเดิมที่วิเคราะห์ความน่าเชื่อถือของผลการวัดด้วยการให้ความสำคัญกับการออกแบบการวัด (Measurement design) ที่คำนึงถึงแหล่งความคลาดเคลื่อนที่เป็นระบบอย่างน้อย 1 แหล่ง ระดับความน่าเชื่อถือของผลการวัดที่ได้มีชื่อเรียกเฉพาะว่า “สัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของผลการวัด” สามารถออกแบบการวัดเพื่อควบคุมแหล่งความคลาดเคลื่อนได้ 3 รูปแบบ ดังนี้ การออกแบบการวัดแบบไขว้ การออกแบบการวัดแบบแฝง และการออกแบบการวัดแบบผสม เพื่อเป็นการลดความคลาดเคลื่อนของการวัดให้น้อยลง หรือมีความเที่ยงสูงถึงระดับที่ต้องการ ซึ่งการวิเคราะห์สัมประสิทธิ์การสรุปอ้างอิง มี 2 ขั้นตอน คือ ขั้นตอนที่หนึ่ง การศึกษาเพื่อการสรุปอ้างอิง เป็นกระบวนการที่มุ่งประมาณค่าความแปรปรวนจากแหล่งต่าง ๆ ภายใต้เอกภพของคะแนนสังเกตที่ยอมรับได้ และขั้นตอนที่สอง การศึกษาเพื่อการตัดสินใจ เป็นการนำผลที่ได้จาก G-study มาวิเคราะห์เพื่อการตัดสินใจเลือกสถานการณ์ของการวัด ปัจจุบันมีการประยุกต์ใช้ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัดทางวิจัยการศึกษา เพื่อพัฒนารูปแบบการประเมินการพัฒนามาตรวัด เปรียบเทียบความเที่ยงของแบบสอบ และความสอดคล้องในแนวเดียวกันระหว่างข้อสอบ ในการออกแบบการวัดครั้งต่อไปสามารถใช้ข้อมูลที่ได้ เพื่อลดความคลาดเคลื่อนของการวัดให้น้อยลง หรือมีความเที่ยงสูงถึงระดับที่ต้องการ

เอกสารอ้างอิง

Atilla, Y. & Ezel, T. (2010). The examination of reliability according to classical test and generalizability on a job performance scale. *Educational Sciences: Theory & Practice*, *10*(3), 1847-1854.

Cronbach, L. J., Gleser, G. C., & Rajaratnam, N. (1963). Theory of generalizability: A liberalization of reliability theory. *British journal of mathematical and statistical psychology*, *16*, 137-173.

Cronbach, L. J., *et al.* (1972). *The dependability of behavioral measurement*. New York: John Wiley.

Gebril, A. (2010). Bringing reading-to-write and writing-only assessment tasks together: a generalizability analysis. *Assessing Writing*, *15*(2), 100-117.

Donmuang, P. (2018). The construction of adversity quotient with buddhist scale for the vocational students. *MBU Education Journal: Faculty of Education Mahamakut Buddhist University*, *6*(2), 210-224. (in Thai)

Janprasert, B. (2017). Alignment between national and classroom science test items: an application of many-facet rasch measurement and generalizability theory. *STOU Education Journal, 11*(2), 70-85. (in Thai)

Kanjanawasee, S. (2020). *Modern test theoryies* (5th ed.). Bangkok : Chulalongkorn University Printing House. (in Thai)

Nuntapanich, R. (2013). A comparison of the mathematics problem-solving diagnostic tests based on meta-cognition knowledge approach and meta-cognition knowledge with experience approach: generalizability theory. *Journal of Education Graduate Studies Research Khon Kaen University, 7*(4), 26-33. (in Thai)

Phadungphon, S. (2016). Comparison of reliability of modified essay question test for measuring the abilities in using scientific method in physics under different numbers of event and rater: an application of generalizability theory. *An Online Journal of Education, 12*(4), 381-393. (in Thai)

Praserthlap, M. (2015). Reliability comparison of likert scales on attitude with different rating scales between classical test theory and generalizability theory. *Silpakorn Educational Research Journal, 7*(2), 249-258. (in Thai)

Sanguanwai, C. (2015). *Comparison of test reliability for measuring mathematics creative problem solving abilities: application of generalizability theory*. *An Online Journal of Education, 11*(4), 743-760. (in Thai)

Srikaew, D. (2014). Development of an English speaking skill assessment model for grade 6 students by using portfolio: an application of multivariate generalizability theory. *Procedia Social and Behavioral Sciences*, *191*(2015), 764-768. (in Thai)

Umnacil, M. (2014). *Comparison of reliability of modified essay question test for measuring scientific problem solving ability using different scoring methods under different numbers of event : an application of generalizability theory*. *An Online Journal of Education, 10*(1), 459-473. (in Thai)