

ดัชนีความสอดคล้องของผู้ประเมิน : RAI

ฉัตรศิริ ปิยะพิมลสิทธิ์*

ในปัจจุบันยุคของการปฏิรูปการศึกษา การวัดผลการประเมินผลได้เปลี่ยนแปลงไปจากเดิม จากการสอบด้วยข้อสอบเลือกตอบ เปลี่ยนมาเป็นการประเมินด้วยการใช้แฟ้มสะสมงาน ใช้การเขียนอัตนัย การประเมินเป็นรายบุคคล การประเมินการปฏิบัติ และวิธีอื่น ๆ ที่เหมาะสม สำหรับการให้คะแนนนั้นไม่ใช่ตอบถูกต้องกับเฉลยให้ 1 ตอบไม่ตรงกับเฉลยให้ 0 อีกต่อไป แต่มีการให้คะแนนหลากหลายโดยไม่มีค่าเฉลยที่ถูกต้องตายตัว และครูที่มาประเมินผลงานของผู้เรียนก็อาจไม่ได้ประเมินเพียงคนเดียวอีกต่อไป

วิธีการวัดผลและประเมินผลเปลี่ยนแปลงไป แต่วิธีการวิเคราะห์เพื่อหาคุณภาพของการประเมินก็ยังคงวางเวียนอยู่แต่ภายในทฤษฎีมาตรฐานเดิมและสัมประสิทธิ์การสรุปอ้างอิง โดยเฉพาะการวิเคราะห์เพื่อหาความเชื่อมั่น (Reliability) ซึ่งหมายถึง สัดส่วนของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนที่สังเกตได้ และดัชนีความเชื่อมั่นที่เป็นที่รู้จักกันมากที่สุด และใช้กันมากที่สุดก็เห็นจะเป็นสูตรสัมประสิทธิ์แอลฟาของครอนบาช (Cronbach alpha) โคเฮน (Cohen) แนะนำการใช้สัมประสิทธิ์ความสอดคล้องสำหรับมาตรวัดนามบัญญัติ การประเมินจะถูกคำนวณโดยใช้ “สัดส่วนของผู้ตัดสินที่เห็นสอดคล้องและสัดส่วนที่คัดหัว” เรียกว่า สัมประสิทธิ์แคปป่า (Kappa) นับว่าเป็นดัชนีหนึ่งที่เหมาะสมในการประเมินความสอดคล้องของผู้สังเกต ซึ่งสัมประสิทธิ์ความสอดคล้องสำหรับข้อมูลระดับนามบัญญัตินี้จะมีข้อจำกัดก็คือจะต้องมีจำนวนของผู้ตัดสินตั้งแต่ 2 คนขึ้นไปสังเกตพฤติกรรมตั้งแต่ 1 พฤติกรรมขึ้นไป

ในอีกวิธีการหนึ่งสำหรับการตรวจสอบความเชื่อมั่นที่ใช้ผู้ประเมิน 2 คน ก็คือการใช้สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson product moment correlation) ซึ่งวิธีนี้จะเป็นเป็นวิธีการหาความเชื่อมั่นแบบดั้งเดิม อย่างไรก็ตามค่าสัมประสิทธิ์สหสัมพันธ์ไม่สามารถอธิบายระดับความสอดคล้องของผู้ประเมินได้

ในการหาความเชื่อมั่นของการประเมินภาคปฏิบัติมักจะใช้ผู้ประเมินหลายคน กลุ่มตัวอย่างหลายคนและประเมินหลายพฤติกรรม อาจใช้กระบวนการวิเคราะห์ความแปรปรวน ซึ่งเรียกว่า Generalizability theory (G-theory) ซึ่งเป็นการคำนวณองค์ประกอบความแปรปรวนที่ต้องมีความรู้ในเชิงสถิติและการวัดผลสำหรับแปลความหมาย สัมประสิทธิ์การตัดสินใจ (D-study) จำเป็นที่จะใช้หลังจากคำนวณ G-Study แล้ว ถ้าผู้ประเมินสามารถประเมินได้สอดคล้องกันอย่างสมบูรณ์ ความแปรปรวนขององค์ประกอบจะเท่ากับ 0.00 เพราะที่ไม่มี ความแปรปรวนระหว่างผู้ประเมิน

* กศ.ม. (การวัดผลการศึกษา) <http://www.watpon.com>

ในบทความนี้จะนำเสนอดัชนีที่เรียกว่า Rater agreement index (RAI) ซึ่งดัชนี RAI นี้ไม่ถือว่าเป็นดัชนีความเชื่อมั่นเพราะไม่ได้คำนวณโดยใช้องค์ประกอบความแปรปรวนของคะแนนจริงและคะแนนที่สังเกตได้ สถิตินี้ไม่ได้ยุ่งยากหรือเข้มงวดเหมือนกับสัมประสิทธิ์ความสอดคล้องภายในอย่างสัมประสิทธิ์แอลฟาของครอนบาค หรือ G-Theory ซึ่งดัชนีความสอดคล้องของผู้ประเมินนี้สามารถใช้กับพฤติกรรมเท่าไรก็ได้, กับกลุ่มตัวอย่างกี่คนก็ได้ และการให้คะแนนจะเป็นแบบไหนก็ได้

ดัชนีความสอดคล้องระหว่างผู้ประเมิน

RAI สามารถใช้คำนวณกับการประเมินที่ให้เป็นคะแนนเป็นช่วงสเกล I สเกล ซึ่งค่า RAI ที่ได้จะมีพิสัยตั้งแต่ 0 – 1 ซึ่งถ้ามีค่าเข้าใกล้ 1 แสดงว่าผู้ประเมินเห็นสอดคล้องกันมาก

1. RAI กรณีมีเป้าพฤติกรรมเดียว, ผู้ประเมิน 2 คน และกลุ่มตัวอย่างคนเดียว

ให้ R_1 และ R_2 แทนผู้ประเมิน 2 คนซึ่งประเมินกลุ่มตัวอย่างคนเดียวและประเมินพฤติกรรมเดียว ประเมินในสเกลที่มีจำนวน I สเกล

$$RAI = 1 - \frac{|R_1 - R_2|}{I-1}$$

ถ้าผู้ประเมินทั้ง 2 คนประเมินให้คะแนนเหมือนกัน RAI จะเท่ากับ 1 ถ้าผู้ประเมิน 2 คนประเมินตรงกันข้ามแล้วค่า RAI จะเท่ากับ 0 ในกรณีของคะแนนที่ให้แบบ 0 กับ 1 แล้ว I จะมีค่าเท่ากับ 2 ดังนั้น RAI จะเขียนใหม่ได้ว่า $RAI = 1 - [R_1 - R_2]$

2. กรณีที่ประเมิน K พฤติกรรม ผู้ประเมิน 2 คน และกลุ่มตัวอย่างคนเดียว

ให้ R_{1k} และ R_{2k} แทนผู้ประเมิน 2 คน ซึ่งประเมินกลุ่มตัวอย่างเพียงคนเดียวในพฤติกรรมที่ k

$$RAI = 1 - \frac{\sum^K |R_{1k} - R_{2k}|}{K(I-1)}$$

RAI มีค่าระหว่าง 0 ถึง 1 ถ้ามีค่าเป็น 1 แสดงว่าผู้ประเมิน 2 คนมีการประเมินสอดคล้องกันในทุกพฤติกรรม

3. กรณีที่ประเมิน K พฤติกรรม ผู้ประเมิน 2 คน และกลุ่มตัวอย่าง N คน

การประเมินนี้จะใช้ผู้ประเมิน 2 คนประเมินกลุ่มตัวอย่าง N คน ซึ่งมีจำนวน K พฤติกรรมที่เหมือนกัน เราจะให้ R_{1kn} และ R_{2kn} แทนการประเมินของผู้ประเมิน 2 คนที่ประเมินกลุ่มตัวอย่างคนที่ n ในพฤติกรรมที่ k

$$RAI = 1 - \frac{\sum_{k=1}^K \sum_{n=1}^N |R_{1kn} - R_{2kn}|}{KN(I-1)}$$

4. กรณีที่ประเมินพฤติกรรมเดียว ผู้ประเมินหลายคน และกลุ่มตัวอย่างคนเดียว

ให้ R_1, R_2, \dots, R_m แทนผู้ประเมินจำนวน M คนซึ่งประเมินกลุ่มตัวอย่างคนเดียวและพฤติกรรมเดียว และให้ \bar{R} เป็นค่าเฉลี่ย เราจะได้สูตรว่า

$$RAI = 1 - \frac{\sum_{m=1}^M |R_m - \bar{R}|}{(M-1)(I-1)}$$

5. กรณีประเมิน K พฤติกรรม ผู้ประเมิน M คน และกลุ่มตัวอย่างคนเดียว

ให้ R_{mk} แทนการประเมินของผู้ประเมินคนที่ m บนพฤติกรรมที่ k ให้ \bar{R}_k แทนค่าเฉลี่ยของการประเมินจำนวน k พฤติกรรม นั่นคือ

$$\bar{R}_k = \frac{1}{M} \sum_{m=1}^M R_{mk}$$

$$RAI = 1 - \frac{\sum_{k=1}^K \sum_{m=1}^M |R_{mk} - \bar{R}_k|}{K(M-1)(I-1)}$$

6. กรณีประเมิน K พฤติกรรม ผู้ประเมิน M คน และกลุ่มตัวอย่าง N คน

ให้ R_{mnk} แทนการประเมินของผู้ประเมินคนที่ m บนพฤติกรรมที่ k ในกลุ่มตัวอย่างคนที่ n ให้ \bar{R}_{nk} เป็นค่าเฉลี่ยของการประเมินจำนวน M คน นั่นคือ

$$\bar{R}_{nk} = \frac{1}{M} \sum_{m=1}^M R_{mnk}$$

$$RAI = 1 - \frac{\sum_{k=1}^K \sum_{n=1}^N \sum_{m=1}^M |R_{mnk} - \bar{R}_{nk}|}{KN(M-1)(I-1)}$$

ตัวอย่างการคำนวณ

สมมติว่ามีผู้ประเมิน 3 คน ประเมินกลุ่มตัวอย่าง 3 คนและกลุ่มตัวอย่างแต่ละคนจะวัดใน 4 พฤติกรรม โดยใช้แบบประเมินที่มี 5 สเกล

พฤติกรรม	ผู้ประเมินคนที่ 1			ผู้ประเมินคนที่ 2			ผู้ประเมินคนที่ 3		
	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 1	คนที่ 2	คนที่ 3
1	1	3	2	1	1	1	3	2	3
2	4	5	4	3	3	2	3	5	4
3	4	5	5	2	3	2	4	5	5
4	3	5	4	2	4	3	4	5	5

สร้างเมตริกของค่าเฉลี่ยที่จำเป็นในการคำนวณ RAI ด้วยสมการ $\bar{R}_{nk} = \frac{1}{M} \sum R_{mk}$

ได้ดังนี้

1.67	2	2
3.33	4.33	3.33
3.33	4.33	4
3	4.67	4

เราแทนค่าในสูตรคำนวณจะได้

$$\begin{aligned}
 RAI &= 1 - \frac{\sum \sum \sum |R_{mnk} - \bar{R}_{nk}|}{KN(M-1)(I-1)} \\
 &= 1 - \frac{28}{4 \cdot 3 \cdot (3-1)(5-1)} \\
 &= .708
 \end{aligned}$$

ค่า RAI คำนวณได้ .708 บ่งชี้ถึงความสอดคล้องกันสูงระหว่างผู้ประเมิน แม้ว่าจะมี ความแปรปรวนระหว่างผู้ประเมิน แต่โดยทั่วไปการประเมินจะสูงในคู่ของผู้ประเมินที่มีการ ประเมินให้คะแนนปานกลางและสูง โดยเฉพาะในตัวอย่างนี้ที่มีอยู่ 2 คนที่ประเมินให้คะแนนสูง และมีอยู่ 1 คนที่ประเมินให้คะแนนต่ำ ซึ่งมีความสอดคล้องกันสูงระหว่างผู้ประเมิน ดังนั้น RAI ก็ ควรจะมีค่าสูงตามไปด้วย

บรรณานุกรม

Burly-Stock, Judith A. and Other. (1996). "Rater Agreement Indexes for Performance Assessment," *Educational and Psychological Measurement*. Vol.56 No.2, April 1996 : 251 – 262.