วารสารศึกษาศาสตร์ ปีที่ ๗ ฉบับที่ ๑ มกราคม - ธันวาคม ๒๕๕๐
Journal of Education, Volume 7, Number 1 January - December 2007

มโนทัศน์ของความเที่ยงและความตรงของแบบทดสอบจากเบื้องหลังทางทฤษฎีและความสัมพันธ์ของการวิเคราะห์

Concepts of Reliability and Validity of the Tests from Theory Behind and Relationship of Analysis

ชัยลิขิต สร้อยเพชรเกษม*

บทนำ

เราทราบกันโดยทั่วไปว่าการวัดนั้นเป็นการกำหนดปริมาณของคุณลักษณะใดๆ(Traits)ในสิ่งที่ต้องการวัด เพื่อตอบปัญหาหรือคำถามว่า ปริมาณที่กล่าวนี้เป็นปริมาณ**เท่าไร** และไม่ว่าศาสตร์ใดก็ตามหากมีวิธีการวัดที่ตอบคำถามนี้ได้แล้วย่อมยกฐานะศาสตร์ของตนจากความหมายของกลุ่มสาขาวิชามาเป็นศาสตร์ในความหมายของวิทยาศาสตร์ได้อีกด้วย คำว่าศาสตร์ในความหมายแรก หมายถึง ตำรา หรือวิชา เช่น ศิลปศาสตร์ คือตำราว่าด้วยวิชาความรู้ 18 ประการ อาทิ การคำนวณ(สังขยา) การยิงธนู(ธนุพเพธา) ดาราศาสตร์(โชติ) การพูด(เกตุ)... ส่วนความหมายที่สอง หมายถึงวิทยาการที่มีวัตถุประสงค์ตรงความจริง เช่น วิทยาศาสตร์ (พระธรรมปิฎก.2543: 291-292) คำว่า วิทยาศาสตร์ไม่มีความหมายอื่นใดนอกเหนือไปจากวิธีการทางวิทยาศาสตร์ (Scientific method) ดังนั้นวิชาใดที่อาศัยวิธีการทางวิทยาศาสตร์เป็นเครื่องมือสำหรับการแสวงหาความรู้ของตน จึงได้ชื่อว่า ศาสตร์(Science) และการวัดนั้นเป็นสิ่งจำเป็นสำหรับวิธีการทางวิทยาศาสตร์(Measurement is sin qua non of science) (Kline.1986: 1) ในส่วนของการวัดทางจิตวิทยานั้นสามารถตอบคำถามนี้ได้แล้ว แต่ก็เกิดคำถามตามมาอีกคือ การวัดปริมาณได้เท่าไรนี้ทำได้ ชัดเจนและเป็นจริงเพียงใด คำถามนี้ทำให้นักจิตวิทยาได้พัฒนาระบบการวัดของตนเองขึ้น โดยเฉพาะเรื่องการพัฒนาเครื่องมือในการวัดให้มีความเชื่อถือได้ และเป็นจริง เมื่อนำเครื่องมือลักษณะดังกล่าวนี้ไปทำการวัดแล้วย่อมตอบคำถามที่ตามมานี้ได้ในที่สุด ทั้งหมดที่กล่าวมานี้มีความเกี่ยวข้องกับคุณสมบัติของเครื่องมือวัดทางจิตวิทยาที่เรียกว่า ความเที่ยง (Reliability) และความตรง(Validity) ของเครื่องมือวัด ในบทความนี้จึงมุ่งเสนอมโนทัศน์ของคุณสมบัติของเครื่องมือทั้งสองประการนี้ รวมถึงเบื้องหลังทางทฤษฎีตลอดจนความสัมพันธ์ระหว่างการวิเคราะห์อันเป็นที่มาของการตรวจสอบคุณสมบัติของเครื่องมือวัดดังกล่าว เพื่อให้เป็นไปตามคติที่ว่า **ทดสอบเพื่อค้นและพัฒนาสมรรถภาพมนุษย์** (ชวาล แพรัตกุล2518: 34)

---

* อาจารย์ประจำภาควิชาการประเมินผลและวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยทักษิณ

ความเที่ยง (Reliability)

ความเที่ยงของการวัดด้วยแบบทดสอบ หมายถึง ความคงที่(Consistency) ของคะแนนการวัด ซึ่งเป็นมโนทัศน์ทางสถิติที่ตั้งอยู่บนพื้นฐานของความสัมพันธ์ระหว่างคะแนนการวัดจากแบบทดสอบคู่ขนาน (Parallel form)จากกลุ่มผู้ตอบกลุ่มเดียวกัน ค่าความสัมพันธ์นี้เรียกว่า ค่าสัมประสิทธิ์ของความเที่ยง (Reliability coefficients) มีนัยเป็นค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficients) ขนาดของความสัมพันธ์มีค่าระหว่าง 0 ถึง 1 ค่าสัมประสิทธิ์ของความเที่ยงที่มีค่าเป็น 0 หมายความว่า ไม่มีความเที่ยง ถ้าค่าสัมประสิทธิ์ของความเที่ยงมากกว่า 0 ถึง 1 หมายความว่า มีความเที่ยงต่ำ-สูง ตามขนาด (Magnitude)ของความสัมพันธ์นั้น มีข้อตกลงทางทฤษฎี ดังนี้

ข้อตกลงของทฤษฎีการทดสอบคะแนนจริงดั้งเดิม (Classical True – Score Theory)

(Allen and Yen. 1979 : 57)

1. $X = T + E$ (คะแนนที่สังเกตหรือวัดได้เท่ากับคะแนนจริงรวมกับคะแนนความคลาดเคลื่อน)
2. $\varepsilon(X) = T$ (ค่าคาดหมาย หรือค่าเฉลี่ยของประชากรของคะแนนที่สังเกตเท่ากับคะแนนจริง)
3. $\rho_{ET} = 0$ (คะแนนความคลาดเคลื่อนกับคะแนนจริงของประชากรผู้สอบไม่สัมพันธ์กัน)
4. $\rho_{E_1E_2} = 0$ (คะแนนความคลาดเคลื่อนของแบบทดสอบทั้งสองฉบับของประชากรผู้สอบไม่สัมพันธ์กัน)
5. $\rho_{E_1T_2} = 0$ (คะแนนความคลาดเคลื่อนของแบบทดสอบฉบับหนึ่ง กับคะแนนจริงของแบบทดสอบฉบับอื่นของประชากรผู้สอบไม่สัมพันธ์กัน)
6. ถ้าแบบทดสอบ 2 ฉบับ ให้ค่าคะแนนที่สังเกตได้คือ X และ X' ถ้ายอมรับข้อตกลงข้อ 1 – 5 และถ้ายอมรับได้ทุกประชากรผู้สอบแล้ว $T = T'$ และ $\sigma_E^2 = \sigma_{E'}^2$ แบบทดสอบ 2 ฉบับนี้ เรียกว่า แบบทดสอบคู่ขนาน
7. ถ้าแบบทดสอบ 2 ฉบับ ให้ค่าคะแนนที่สังเกตได้คือ X_1 และ X_2 ถ้ายอมรับข้อตกลงข้อที่ 1 – 5 และถ้ายอมรับได้ทุกประชากรผู้สอบแล้ว $T_1 = T_2 + C_{12}$ เมื่อ C_{12} เป็นค่าคงที่แบบทดสอบ 2 ฉบับนี้ เรียกว่า แบบทดสอบคู่ขนานเทียม (Essentially τ Equivalent Test)

นิยามและการแปลความหมายค่าสัมประสิทธิ์ความเที่ยงจากแบบทดสอบคู่ขนาน(Allen and Yen. 1979 : 72-75)

นิยามและการแปลความหมายของความเที่ยงมีหลายประการ ตัวอย่างเช่น แบบทดสอบมีความเที่ยงถ้าคะแนนที่สังเกตได้ของผู้สอบทุกคน มีสหสัมพันธ์สูงกับคะแนนจริงที่เขาได้รับ กำลังสองของ

สหสัมพันธ์ระหว่างคะแนนที่สังเกตกับคะแนนจริง (ρ^2_{XT}) เรียกว่า ค่าสัมประสิทธิ์ความเที่ยง (Reliability coefficient) ของแบบทดสอบ หรือความเที่ยงแสดงให้เห็นได้จากสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้เมื่อสอบด้วยแบบทดสอบคู่ขนาน ถ้าแบบทดสอบคู่ขนาน 2 ฉบับ สอบโดยผู้สอบกลุ่มประชากรเดียวกัน และผลของคะแนนที่สังเกตได้มีสหสัมพันธ์กัน (สหสัมพันธ์นี้ใช้สัญลักษณ์ $\rho_{xx'}$ คือค่าสัมประสิทธิ์สหสัมพันธ์) ในทุกกรณีค่าคะแนนจริงจะไม่สามารถรู้ได้ และแบบทดสอบแบบคู่ขนานก็หาได้ยาก ดังนั้น ความเที่ยงจึงจำเป็นต้องประมาณค่าด้วยวิธีการอื่น ๆ การพิจารณาแนวทางการนิยามและแปลความหมายของค่าสัมประสิทธิ์สหสัมพันธ์ที่จะกล่าวต่อไปนี้ 6 แนวทาง ซึ่งจะให้สัญลักษณ์ $\rho_{xx'}$ ว่าเป็นค่าสัมประสิทธิ์ความเที่ยง แม้ว่าการใช้ข้อสอบสอบแบบคู่ขนานจะไม่สามารถใช้สอบตามนิยามความเที่ยงได้ก็ตาม

1. $\rho_{xx'}$ = สหสัมพันธ์ระหว่างคะแนนที่สังเกตได้ 2 ชุดจากแบบทดสอบคู่ขนาน
2. $\rho^2_{xx'}$ = สัดส่วนความแปรปรวนของ X ที่อธิบายได้ด้วยความสัมพันธ์เชิงเส้นตรงกับ X'
3. $\rho_{xx'} = \sigma^2_T / \sigma^2_X$
4. $\rho_{xx'} = \rho^2_{XT}$
5. $\rho_{xx'} = 1 - \rho^2_{XE}$
6. $\rho_{xx'} = 1 - \sigma^2_E / \sigma^2_X$

ข้อ 1 หมายความว่า ความเที่ยงของแบบทดสอบเท่ากับสหสัมพันธ์ระหว่างคะแนนที่สังเกตได้สองชุดที่ได้จากการสอบด้วยแบบทดสอบคู่ขนาน ถ้าผู้สอบแต่ละคนได้คะแนนที่สังเกตได้เหมือนกัน และความแปรปรวนของคะแนนเท่ากัน ค่าความเที่ยงจะมีค่าสมบูรณ์ ($\rho_{xx'}$ =1.00) ถ้าไม่มีสหสัมพันธ์กันแล้ว ($\rho_{xx'}$ = 0.00) แบบทดสอบจะไม่มีความเที่ยง

ข้อ 2 เป็นการแปลความหมายตามมาตรฐานของค่าสัมประสิทธิ์สหสัมพันธ์เพียร์สัน ค่าสัมประสิทธิ์สหสัมพันธ์กำลังสองหมายถึง สัดส่วนของความแปรปรวนของตัวแปรตัวหนึ่งซึ่ง ถูกอธิบายด้วยตัวแปรอีกตัวหนึ่งในรูปความสัมพันธ์เชิงเส้นตรง ดังนั้น $\rho^2_{xx'}$ จะแสดงให้เห็นสัดส่วนของความแปรปรวนของคะแนนที่สังเกตได้ชุดหนึ่งกับคะแนนอีกชุดหนึ่งในรูปความสัมพันธ์เชิงเส้นตรง เมื่อคะแนนสองชุดนั้นได้มาจากแบบทดสอบคู่ขนานกัน

ข้อ 3 เป็นนิยามของค่าสัมประสิทธิ์ความเที่ยง หมายถึง ค่าอัตราส่วนที่ได้จากความแปรปรวนของคะแนนที่สังเกตได้ในส่วนที่เป็นคะแนนจริงต่อคะแนนที่สังเกตได้ทั้งหมด สำหรับแบบทดสอบที่มีความเที่ยงสมบูรณ์เมื่อ $\rho^2_{xx'} = 1$ แล้ว σ^2_T / σ^2_X จะเท่ากับ 1 แสดงให้เห็นว่า ความแปรปรวนของคะแนนที่สังเกตได้ทั้งหมดเท่ากับความแปรปรวนของคะแนนจริง ความแตกต่างระหว่างคะแนนที่สังเกตได้ของผู้สอบ

แต่ละคน จะสะท้อนให้เห็นความแตกต่างในคะแนนจริงของผู้สอบแต่ละคนด้วย ถ้า $\sigma_X^2 = \sigma_T^2$ แล้ว σ_E^2 จะเท่ากับ 0 การประมาณค่าความคลาดเคลื่อนแบบสุ่มทั้งหมดจะเป็น 0 และเมื่อ $\sigma_E^2 = 0$ ดังนั้น $\rho_{xx'} = 1$ หมายถึง การสอบวัดจะไม่มีความคลาดเคลื่อนเลย กรณีเมื่อ $\rho_{xx'} < 1$ แล้ว ความคลาดเคลื่อนจะยังคงมีอยู่ในการสอบนั้น และเมื่อ $\rho_{xx'} = 0$ แสดงว่า $\sigma_X^2 = \sigma_E^2$ หมายถึง คะแนนที่สังเกตได้ทั้งหมดเป็นคะแนนความคลาดเคลื่อน ดังนั้น ความแตกต่างระหว่างคะแนนที่สังเกตได้ของผู้สอบแต่ละคนจะสะท้อนให้เห็นคะแนนความคลาดเคลื่อนแบบสุ่มที่มีมากกว่าคะแนนจริง ถ้าความเที่ยงของแบบทดสอบเพิ่มขึ้น ความแปรปรวนของความคลาดเคลื่อนจะลดลง และเมื่อความคลาดเคลื่อนของคะแนนลดลง ความแตกต่างของคะแนนที่สังเกตได้ของผู้สอบแต่ละคนจะเข้าใกล้คะแนนจริงของตนเอง อย่างไรก็ตามคะแนนที่สังเกตได้จะประมาณค่าคะแนนจริงได้ไม่ดีเมื่อ ความแปรปรวนของคะแนนความคลาดเคลื่อนมีมาก

ข้อ 4 สมการ $\rho_{xx'} = \rho_{XT}^2$ แสดงค่าสัมประสิทธิ์ความเที่ยง เท่ากับกำลังสองของสหสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนจริง เช่น ถ้า $\rho_{xx'} = 0.81$ แล้ว $\rho_{XT} = 0.90$ เป็นต้น เมื่อใดก็ตาม ถ้า $0 < \rho_{xx'} < 1.00$ แล้ว เราจะเห็นว่า $\rho_{XT} > \rho_{xx'}$

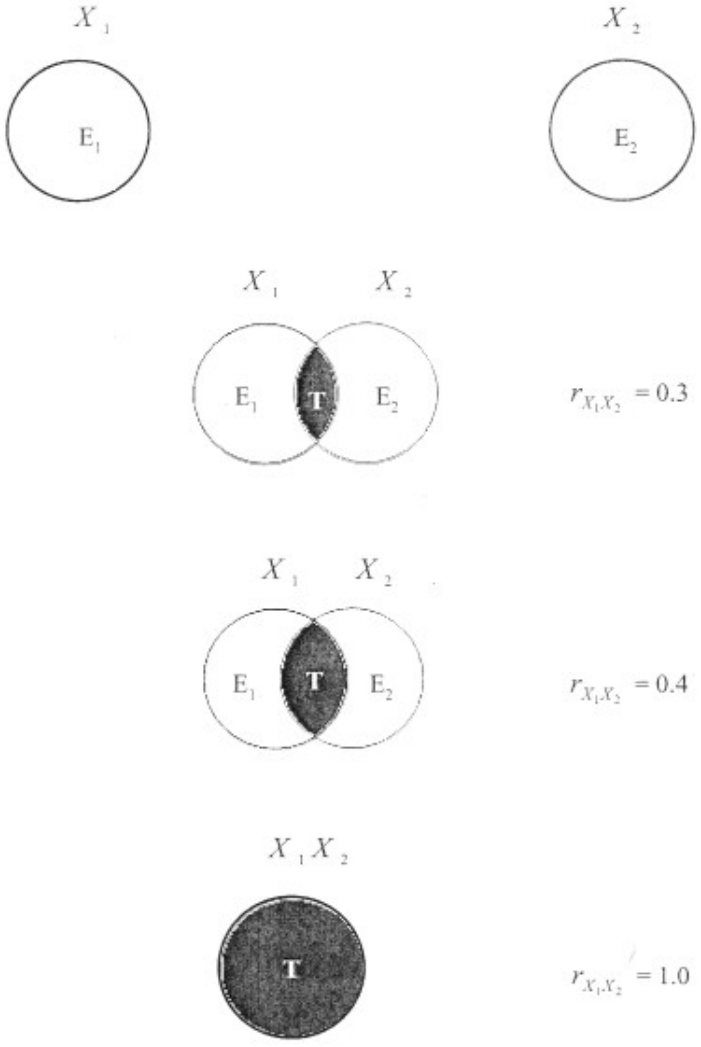
ข้อ 5 สมการ $\rho_{xx'} = 1 - \rho^2_{XE}$ แสดงว่า ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 1 ลบด้วยกำลังสองของสหสัมพันธ์ระหว่างคะแนนที่สังเกตได้กับคะแนนความคลาดเคลื่อน โดยปกติแล้ว ρ_{XE} ควรเท่ากับ 0 แต่จะอยู่ภายใต้เงื่อนไขที่ $\rho_{XE} = 0$ เมื่อ $\rho_{xx'} = 1$

ข้อ 6 สมการ $\rho_{xx'} = 1 - \sigma_E^2 / \sigma_X^2$ ค่าความเที่ยงจากสมการนี้เกี่ยวข้องกับความแปรปรวนของคะแนนความคลาดเคลื่อน และความแปรปรวนของคะแนนที่สังเกตได้ อธิบายว่า เมื่อ $\rho_{xx'} = 1$, $\sigma_E^2 = 0$ และเมื่อ $\rho_{xx'} = 0$, $\sigma_E^2 = \sigma_X^2$ ระดับความเป็นวิวิธพันธ์ของคะแนนที่สังเกตได้จากการสอบของกลุ่มผู้สอบกลุ่มใด ๆ ก็ตาม จะมีผลกระทบต่อความเที่ยงอย่างแน่นอน

นิยามทั้งหมดนี้เป็นนิยามความเที่ยงโดยใช้แบบทดสอบคู่ขนาน ซึ่งเป็นข้อสมมติทางทฤษฎีที่เป็นแนวความคิดตามอุดมคติ เพราะเชื่อว่า ถ้าใช้แบบทดสอบคู่ขนานสอบกับกลุ่มผู้สอบกลุ่มเดียวกันแล้ว คะแนนที่สังเกตได้ต้องให้ค่าสัมประสิทธิสหสัมพันธ์กันสูงเป็น 1.00 แบบทดสอบนั้นย่อมมีความเที่ยงสมบูรณ์ ซึ่งตามทฤษฎีต้องเป็นเช่นนั้น แต่แนวปฏิบัติเราไม่สามารถสร้างแบบทดสอบ 2 ฉบับให้มีลักษณะเป็นแบบทดสอบคู่ขนานกันได้ ค่าสัมประสิทธิ์ของความเที่ยงนี้เป็นค่าที่แสดงความคงที่ของคะแนนจากการวัด ซึ่งมีความหมายเป็นคะแนนจริง แต่มิใช่คะแนนที่วัดได้นั้นวัดได้ตรงกับความเป็นจริง หรือตรงกับวัตถุประสงค์ของการวัด อธิบายความแปรปรวนของคะแนนได้ดังภาพประกอบนี้

ให้ X_1 และ X_2 เป็นความแปรปรวนของคะแนนที่วัดได้จากการสอบครั้งที่ 1 และ 2 ตามลำดับ E เป็นความแปรปรวนของคะแนนความคลาดเคลื่อนสุ่ม และ T เป็นความแปรปรวนของคะแนนจริง(กรณี

ตัวอย่างเป็นการตรวจสอบความเที่ยงด้วยการสอบซ้ำ เนื่องจากถ้าใช้แบบทดสอบคู่ขนานประกอบการอธิบายแล้วความแปรปรวนจะเท่ากับ 1 เสมอ)



ค่า $r_{X_1X_2}$ นี้เป็นพื้นฐานของมโนทัศน์เกี่ยวกับความเที่ยงของแบบทดสอบที่ได้จากการสอบซ้ำหรือเรียกว่า สัมประสิทธิ์ของความคงที่ (Coefficient of stability) แต่ถ้าเปลี่ยนจากวิธีการสอบซ้ำมาใช้วิธีการสอบโดยใช้แบบทดสอบคู่ขนาน (Parallel form) เพื่อขจัดปัญหาของการจำข้อสอบได้(Carry over effect)ใช้สัญลักษณ์ของสัมประสิทธิ์สหสัมพันธ์เป็น $r_{XX'}$ เรียกว่า ค่าสัมประสิทธิ์ของความเท่าเทียมกัน (Coefficient of equivalent) สำหรับลักษณะของแบบทดสอบคู่ขนานกันนั้นมีคุณสมบัติคือ 1) มีค่าเฉลี่ยและ

ความแปรปรวนเท่ากัน 2) มีค่าสหสัมพันธ์ภายในระหว่างข้อสอบ (Item Intercorrelation) เท่ากัน 3) มีเนื้อหาที่วัดเหมือนกัน 4) มีค่าสัมประสิทธิสหสัมพันธ์กับแบบทดสอบอื่นเท่ากัน แต่ถ้าเห็นว่ามีปัญหาเกี่ยวกับวิธีการตรวจสอบความเที่ยงของแบบทดสอบทั้ง 2 วิธีดังกล่าวมาข้างต้น เราจะใช้การตรวจสอบโดยวิธีการสอบเพียงครั้งเดียวเรียกว่า วิธีการหาค่าสัมประสิทธิ์ของความคงที่ภายในแบบทดสอบ (Coefficient of Internal Consistency) ซึ่งมีมโนทัศน์เดียวกันกับการใช้แบบทดสอบคู่ขนาน กล่าวคือ ใช้วิธีการแบ่งครึ่งข้อสอบ (Split-Half) โดยสมมติว่าแบบทดสอบที่แบ่งครึ่งนั้นแต่ละครั้งเป็นแบบทดสอบหนึ่งฉบับ ซึ่งถือเสมือนเป็นแบบทดสอบคู่ขนาน แล้วนำคะแนนที่ได้จากการวัด(ครั้งเดียว) นั้นมาหาค่าสัมประสิทธิ์สหสัมพันธ์ตามวิธีการที่ปรากฏในตำราการวัดผลทั่วไป หรือถ้าไม่แบ่งครึ่งข้อคำถามในแบบทดสอบก็ใช้การจินตนาการหรือสมมติว่าถ้าเราแบ่งข้อคำถามของแบบทดสอบออกเป็น 3 ฉบับ 4 ฉบับ (3 ส่วน 4 ส่วน ซึ่งมีจำนวนข้อคำถามส่วนละเท่า ๆ กัน) จนถึงสมมติไปถึงว่าแบ่งออกเป็นส่วนละ 1 ข้อคำถามให้แต่ละส่วนที่แบ่งออกนั้นเปรียบเสมือนว่าเป็นแบบทดสอบคู่ขนานเทียม (Allen and Yen. 1979 : 83-84) แล้วนำมาหาค่าสัมประสิทธิ์สหสัมพันธ์ภายในโดยใช้วิธีการของ Kuder-Richardson สูตร KR20,KR21 หรือ α(20), α(21) หรือของ Cronbach สูตร Coefficient α เป็นต้น ก็จะได้ค่าสัมประสิทธิ์สหสัมพันธ์ของแบบทดสอบที่แสดงให้เห็นว่าแบบทดสอบนั้นมี หรือไม่มีความเที่ยงได้

ความตรง (Validity)

ในการพิจารณาความตรงของแบบทดสอบมีคำถามพื้นฐานที่ผู้สร้างแบบทดสอบต้องตอบให้ได้ว่า แบบทดสอบนั้นวัดความรู้หรือคุณลักษณะต่างๆ ได้เหมาะสมหรือตรงกับจุดมุ่งหมายของการวัดได้หรือไม่ และโดยทั่วไปเราต้องพิจารณาตัดสินใจเกี่ยวกับความตรงโดยอาศัยความสัมพันธ์ระหว่างคะแนนการวัดจากแบบทดสอบกับตัวแปรอื่น ๆ ซึ่งจะใช้เป็นข้ออ้างได้ว่าแบบทดสอบมีความตรง อย่างไรก็ตามมโนทัศน์ของการวิเคราะห์นั้นไม่เหมือนกับการวิเคราะห์ความเที่ยงของแบบทดสอบ ซึ่งหมายถึง ในการวิเคราะห์ความตรงนี้ไม่มีวิธีการทางสถิติใดที่จะกำหนด หรือเป็นเหตุให้แบบทดสอบมีความตรงได้ (Friedenberg. 1995 : 221) กล่าวคือการวิเคราะห์ทางสถิติ เช่นการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ของแบบทดสอบสองฉบับ เป็นต้น ไม่ได้เป็นเหตุให้แบบทดสอบมีความตรง เพียงแต่เป็นวิธีการที่หาข้อมูล หลักฐานประจักษ์พยาน (Evidence) เพื่อสนับสนุน ยืนยันว่าแบบทดสอบมีความตรงเท่านั้น กระบวนการดังกล่าวเรียกว่า มาตรการเพื่อทราบความตรง (Validation measure) ซึ่งไม่เหมือนกับการวิเคราะห์ความเที่ยง เพราะความเที่ยงนิยามว่าเป็นความคงที่ของผลการวัด เช่น การที่จะทำให้แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนมีความตรงตามเนื้อหา ผู้สร้างแบบทดสอบจะต้องสร้างข้อสอบให้สามารถวัดได้ตรงตามตารางวิเคราะห์ข้อสอบ(Items specification table) หรือแบบทดสอบวัดบุคลิกภาพ ผู้สร้างแบบทดสอบต้อง อธิบายภาวะสันนิษฐาน

(Construct explication) เพื่อนิยามเชิงปฏิบัติการ(Operational definition) แล้วจึงสร้างข้อสอบให้ตรงกับนิยามนั้น เป็นต้น ดังนั้นการตรวจสอบความตรงโดยใช้วิธีการทางสถิตินั้นเป็นการพิจารณาที่ผล มิใช่เป็นการพิจารณาที่เหตุ

การวิเคราะห์ความตรงของแบบทดสอบ แบ่งออกเป็น 3 ประเภท แต่ละประเภทมีคำถามแตกต่างกัน ซึ่งได้แก่ 1) ความตรงตามเกณฑ์ (Criterion Validity) ซึ่งหมายถึง ความสามารถของแบบทดสอบที่จะใช้ทำนายการปฏิบัติ พฤติกรรม ความเชื่อ บุคลิกภาพ ฯลฯ เมื่อเทียบกับเกณฑ์อื่น(แบบทดสอบ หรือการวัดอื่น) ในช่วงเวลาเดียวกัน หรือต่างช่วงเวลา (บัจจุบัน-อนาคต) ถ้าแบบทดสอบใดมีคุณสมบัติความตรงตามเกณฑ์ในช่วงเวลาเดียวกัน หรือต่างช่วงเวลาแล้ว เมื่อนำแบบทดสอบไปใช้ในภายหลัง แบบทดสอบจะสามารถทำนายการปฏิบัติ พฤติกรรม ความเชื่อ บุคลิกภาพ ฯลฯ ได้ตามเกณฑ์นั้น 2) ความตรงตามเนื้อหา (Content Validity) หมายถึง การวัดที่ตรงกับเนื้อหาที่ปรากฏชัดเป็นรูปธรรม เช่นเนื้อหาในบทเรียน และ 3) ความตรงตามภาวะสันนิษฐาน หรือโครงสร้างทางทฤษฎี (Construct Validity) หมายถึง การวัดที่ตรงกับทฤษฎีเกี่ยวกับคุณลักษณะต่างๆ ของผู้สอบ ซึ่งเป็นภาวะสันนิษฐานที่สมติขึ้นตามทฤษฎี และมีความเป็นนามธรรม (Theoretical abstraction)

การวิเคราะห์ทางสถิติด้านความตรงของแบบทดสอบมีประเด็นบ่งชี้ที่แตกต่างกับความเที่ยง ความเที่ยงของแบบทดสอบนั้นแสดงหรือบ่งชี้ว่าเป็นความสามารถของแบบทดสอบที่จะให้ผลการวัดหรือคะแนนมีความคงที่ในตัวแปรเดียวกัน เพราะนิยามความเที่ยงเป็นนิยามโดยการใช้แบบทดสอบคู่ขนาน(ความแปรปรวนของตัวแปรตัวเดียวกัน) ส่วนความตรงนั้นเป็นการแสดงหรือบ่งชี้ว่า ความคงที่ของคะแนนที่วัดได้นั้นเป็นความคงที่ของคะแนนที่วัดได้จากคุณลักษณะใด (Characteristics) เพราะนิยามความตรงเป็นนิยามของการวัดลักษณะร่วมของต่างตัวแปร หรือเป็นความคงที่ของคะแนนต่างตัวแปร (ความแปรปรวนร่วม) นั่นคือมโนทัศน์ที่สำคัญและแสดงให้เห็นความสัมพันธ์ในเชิงสถิติระหว่างทั้งสองกระบวนการ

อย่างไรก็ตามคุณสมบัติทั้งสองประการนี้ไม่ได้เป็นสาเหตุให้แบบทดสอบมีคุณภาพ แต่กระบวนการสร้างแบบทดสอบต่างหากที่เป็นสาเหตุที่ทำให้แบบทดสอบมีคุณภาพ ดังนั้น การหาประจักษ์พยานยืนยันคุณภาพทั้งสองประการนี้เราจึงเรียกว่า การตรวจสอบคุณภาพของแบบทดสอบ อุปมาว่า มีดที่มีคุณภาพนั้นตรวจสอบได้โดยการสังเกตดูว่ามีดตัดไม้ได้ขาดได้ในเวลาอันรวดเร็วโดยไม่ชำรุดเสียหายหรือไม่ แต่สิ่งที่จะทำให้มีดมีคุณภาพนั้นอยู่ที่กระบวนการผลิตมีดต่างหาก แบบทดสอบก็เช่นเดียวกัน การที่แบบทดสอบจะมีคุณภาพนั้นหลักอยู่ที่ว่า ผู้สร้าง หรือผู้ออกข้อสอบ ออกข้อสอบดี ถูกต้องตามหลักวิชาหรือไม่ การพิจารณาดังนี้เป็นการพิจารณาเหตุที่ทำให้เกิดคุณภาพ มิใช่การพิจารณาที่ผลการใช้แบบทดสอบที่แสดงว่าแบบทดสอบมีคุณภาพ กรณีดังกล่าว ชวาล แพรัตกุล (2518: 123-138) ได้กล่าวถึงคุณลักษณะของแบบทดสอบที่ดี 10 ประการ (บัญญัติ 10 ประการ) ซึ่งประกอบด้วยคุณลักษณะด้านความตรง และความ

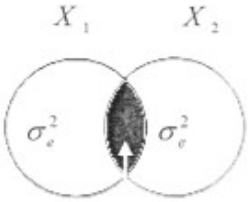
เที่ยง อยู่ด้วย และให้ทัศนะว่า คุณลักษณะต่างๆ เหล่านั้นทำให้มีขึ้นได้โดยการเขียนข้อสอบ หรือสร้างแบบทดสอบให้ดี มีคุณภาพ

"...การสอบวัดที่ดีๆคงจะไม่ได้มาง่ายๆดังที่ชาวบ้านนึก เอาแต่เพียงเขียนโจทย์คำถามก็พอ นึกหรือว่า ถ้าใครพูดภาษาคำถามเป็น หรือสามารถเขียนประโยคคำถามได้ ก็จะสามารถแต่งข้อสอบวัดได้ตรงจุด ตรงตามความปรารถนา และตรงตามความมุ่งหมายของวิชานั้น?..." (ชวาล แพรัตกุล.2518: 22)

"...บัญญัติ 10 ประการของแบบทดสอบที่ดีนี้ ถ้าระลึกไว้เสมอๆขณะเขียนข้อสอบก็จะเป็นลู่ทางให้การวัดผลของเรามีความหมาย และบริสุทธิผุดผาดยิ่งขึ้นด้วย..." (ชวาล แพรัตกุล.2518:138)

ทฤษฎีเบื้องหลัง และความสัมพันธ์ของการวิเคราะห์ความเที่ยง และความตรง

การวิเคราะห์ความเที่ยงเป็นการแสดงให้เห็นว่าคะแนนจากการวัดนั้นแต่ละครั้งมีความสัมพันธ์กันเท่าไร หรือมีความแปรปรวน (σ_T^2) เท่าไร สิ่งนี้จะเป็นสิ่งสะท้อนให้เห็นคะแนนจริง หรือความคงที่ของคะแนน เขียนเป็นภาพอธิบายได้ดังนี้



Reliability (σ_T^2 = True Score)

เมื่อ X_1 และ X_2 เป็นการวัด 2 ครั้ง โดยใช้กลุ่มตัวอย่างและแบบทดสอบเดียวกัน

σ_e^2 เป็นความแปรปรวนของความคลาดเคลื่อน

σ_T^2 เป็นความแปรปรวนของคะแนนจริง

แต่ความหมายของคะแนนจริง ในการวิเคราะห์ความตรงนั้น หมายถึง คะแนนจริงที่เป็นคุณลักษณะที่เกี่ยวข้องสัมพันธ์กันกับวัตถุประสงค์ของการทดสอบหรือลักษณะที่ต้องการวัด(เป็นทั้งคะแนนที่คงที่ และคะแนนที่วัดได้ตรงกับวัตถุประสงค์ของการวัดด้วย) จากสมการการวิเคราะห์ความเที่ยง $\sigma_X^2 = \sigma_T^2 + \sigma_E^2$ นั้น เราแบ่ง σ_T^2 ออกเป็น σ_R^2 และ σ_I^2 ดังนั้นจึงเขียนสมการใหม่ได้ดังนี้

$$\sigma_X^2 = (\sigma_R^2 + \sigma_I^2) + \sigma_E^2 \qquad \text{(Friedenberg. 1995 : 224)}$$

เมื่อ σ_X^2 เป็นความแปรปรวนทั้งหมด

σ_T^2 เป็นความแปรปรวนรวมของคะแนนจริงหรือความเที่ยง

σ_R^2 เป็นความแปรปรวนของคะแนนที่วัดตรงหรือเกี่ยวข้องสัมพันธ์กับวัตถุประสงค์ของการวัด

σ_I^2 เป็นความแปรปรวนของคะแนนที่วัดไม่ตรงหรือไม่เกี่ยวข้องสัมพันธ์กับวัตถุประสงค์ของการวัด

σ_E^2 เป็นความแปรปรวนของความคลาดเคลื่อน

การวิเคราะห์ความเที่ยงเราทดสอบว่า $\sigma_R^2 + \sigma_I^2 > \sigma_E^2$ ส่วนการวิเคราะห์ความตรงเราทดสอบว่า $\sigma_R^2 > \sigma_I^2$ เท่าไร ในทางทฤษฎีเราจึงนิยามความตรงว่า

$$\text{Val} = \frac{\sigma_R^2}{\sigma_X^2}$$

และนิยามความเที่ยงว่า

$$\text{Rel} = \frac{\sigma_T^2}{\sigma_X^2}$$

โดยสรุปแล้ว เราแบ่งส่วนของความแปรปรวนของการทดสอบความตรงได้ 3 ส่วนคือ (Friedenberg. 1995 : 224)

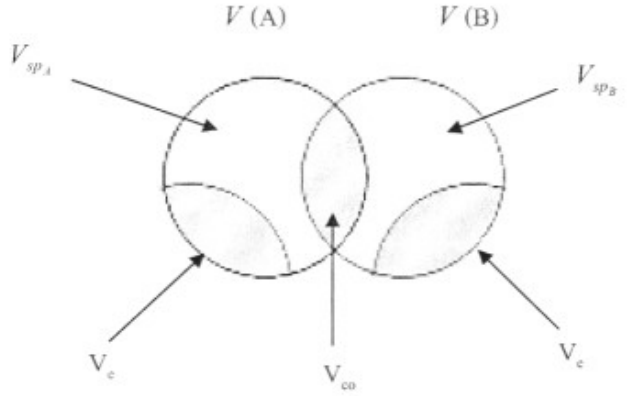
1. สัดส่วนของความแปรปรวนของความคลาดเคลื่อนสุ่ม ซึ่งถูกกำหนดโดยความเที่ยงของแบบทดสอบคือ $\text{Rel} = \frac{\sigma_T^2}{\sigma_X^2}$ ซึ่งแสดงให้เห็นขนาดของ $\frac{\sigma_E^2}{\sigma_X^2}$

2. สัดส่วนอันเนื่องมาจากการวัดที่เกี่ยวข้องกับวัตถุประสงค์ของการวัดถูกกำหนดโดยความสัมพันธ์ระหว่างคะแนนการสอบกับตัวแปรอื่นหรือแบบทดสอบอื่น

3. สัดส่วนอันเนื่องมาจากการวัดในองค์ประกอบอื่นหรือตัวแปรที่ไม่เกี่ยวข้องกับการวัดที่อธิบายไม่ได้ว่าเป็นตัวแปรใด

ตามเหตุผลแล้วถ้าแบบทดสอบมีความสมบูรณ์ จะให้ค่าการวัดที่เป็นความแปรปรวนของคะแนนจริงที่เกี่ยวข้องกับวัตถุประสงค์ของการวัดเพียงอย่างเดียว($\sigma_X^2 = \sigma_R^2$) แต่ตามข้อตกลงหรือสมมติฐานทางการวัดเชิงจิตวิทยาเชื่อว่าแบบทดสอบทั้งหลายไม่มีความสมบูรณ์ของการเป็นเครื่องมือวัดคะแนนจากการวัดซึ่งสะท้อนให้เห็นถึงคุณลักษณะต่าง ๆ (หลายคุณลักษณะ) ที่แปรผัน (Vary) อยู่ในคะแนนจริงซึ่งส่วนหนึ่งไม่ใช่คุณลักษณะที่เกี่ยวข้องกับวัตถุประสงค์ของการวัด

Kerlinger (1992. 428-431) ได้อธิบายเป็นแผนภาพเกี่ยวกับมโนทัศน์ของความตรงไว้ ดังนี้



เมื่อ V (A) และ V (B) เป็นความแปรปรวนของคะแนนจากแบบทดสอบ 2 ฉบับ ตามลำดับ

V_{co} ความแปรปรวนของคะแนนองค์ประกอบร่วมหรือความแปรปรวนของคุณลักษณะที่เกี่ยวข้องกับวัตถุประสงค์ของการวัด

V_{sp_A} และ V_{sp_B} เป็นความแปรปรวนขององค์ประกอบอื่น (Specific Variance) ซึ่งเป็นความแปรปรวนอย่างมีระบบขององค์ประกอบที่ไม่เกี่ยวข้องกับวัตถุประสงค์ของการวัด

V_e เป็นความแปรปรวนของความคลาดเคลื่อนสุ่ม

V_t เป็นความแปรปรวนรวมทั้งหมด

เขียนเป็นสมการ

$$V_t = V_{co} + V_{sp} + V_e\ (\sigma_R^2 + \sigma_I^2 + \sigma_E^2) \quad \text{.....1}$$

$$\frac{V_t}{V_t} = \frac{V_{co}}{V_t} + \frac{V_{sp}}{V_t} + \frac{V_e}{V_t} \quad \text{.....2}$$

$$\frac{V_{co}}{V_t} = \frac{V_t}{V_t} - \frac{V_{sp}}{V_t} - \frac{V_e}{V_t} \quad \text{.....3}$$

แต่นิยามของ Reliability คือ

$$r_{tt} = 1 - \frac{V_e}{V_t} \quad \text{ดังนั้น} \quad \text{.....4}$$

$$r_{tt} = \frac{V_t}{V_t} - \frac{V_e}{V_t} \quad \text{.....5}$$

ทำให้ $\frac{V_{co}}{V_t} = \frac{V_t}{V_t} - \frac{V_e}{V_t} - \frac{V_{sp}}{V_t}$ 6

และเมื่อ $r_{tt} = \frac{V_t}{V_t} - \frac{V_e}{V_t} = \frac{V_\infty}{V_t}$ (V_∞ = ความแปรปรวนของคะแนนจริง หรือ σ_T^2)

แทนค่า 6 $\frac{V_{co}}{V_t} = \frac{V_\infty}{V_t} - \frac{V_{sp}}{V_t}$ 7

ในการวิเคราะห์องค์ประกอบ (Factor Analysis) เรานิยามว่า

$V_{co} = V_A + V_B$ 8

เมื่อ V_{co} เป็นความแปรปรวนองค์ประกอบร่วมหรือความตรง

V_A เป็นความแปรปรวนขององค์ประกอบ A

V_B เป็นความแปรปรวนขององค์ประกอบ B

(Kerlinger.1992: 430)

ดังนั้น $\frac{V_{co}}{V_t} = \frac{V_A}{V_t} + \frac{V_B}{V_t}$ 9

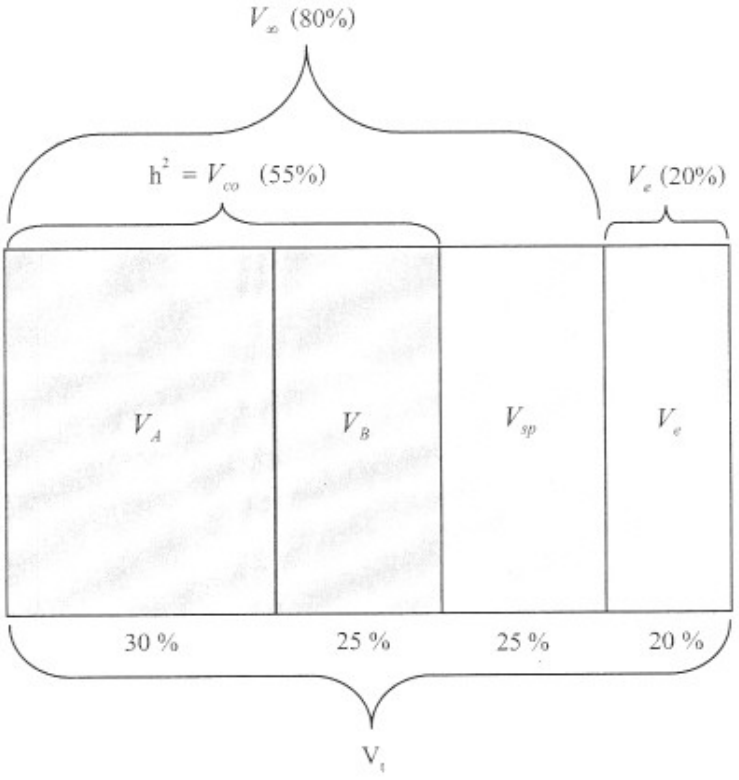
และสมการ $Val = \frac{\sigma_R^2}{\sigma_X^2} = \frac{V_{co}}{V_t}$ 10

ดังนั้น $Val = \frac{V_A}{V_t} + \frac{V_B}{V_t}$ 11

เมื่อพิจารณาสมการ 1 และ 2 แล้วแทนค่าสมการ 2 ด้วยสมการ 9 จะได้

$$\frac{V_t}{V_t} = \overbrace{\underbrace{\frac{V_A}{V_t} + \frac{V_B}{V_t} + \frac{V_{sp}}{V_t}}_{r_{tt}}}^{h^2} + \frac{V_e}{V_t}$$

เขียนและสมมติค่าสัดส่วนอธิบายได้ดังภาพ



จากภาพแสดงให้เห็นว่า V_∞ คือ ค่าความเที่ยง และ h^2 (ค่าการร่วม) หรือ V_{co} ก็คือค่าความตรงนั่นเอง

สรุป

การศึกษามโนทัศน์ของความเที่ยงและความตรงของแบบทดสอบจากเบื้องหลังทางทฤษฎี และความสัมพันธ์ระหว่างการวิเคราะห์ของมโนทัศน์ทั้งสองนี้ เป็นพื้นฐานของการศึกษาเกี่ยวกับคุณภาพของแบบทดสอบ เพราะจะทำให้ผู้พัฒนาและผู้ใช้แบบทดสอบสามารถทำความเข้าใจ และแปลความหมายของคุณสมบัติเครื่องมือได้ กล่าวโดยสรุปแล้วมโนทัศน์ทั้งสองมีความหมายใกล้เคียงกันอาจทำให้เข้าใจได้ว่าเป็นเรื่องเดียวกัน อย่างไรก็ตามอาจกล่าวว่ามีนัยแตกต่างกันเล็กน้อยทำให้มโนทัศน์ทั้งสองไม่ใช่สิ่งเดียวกัน เบื้องหลังทางทฤษฎีเป็นสิ่งที่บอก หรืออธิบายให้เห็นนัยของความแตกต่างกันระหว่างสองมโนทัศน์นี้ คำกล่าวว่า"แบบทดสอบที่มีความตรงย่อมมีความเที่ยง" เป็นตัวอย่างที่สามารถศึกษา ทำความเข้าใจได้จากเบื้องหลังทางทฤษฎี และความสัมพันธ์ระหว่างมโนทัศน์ทั้งสองนี้

บรรณานุกรม

ชวาล แพรัตกุล. (2518). เทคนิคการวัดผล. พิมพ์ครั้งที่ 6 กรุงเทพฯ: วัฒนาพานิช.

พระธรรมปิฎก (ป.อ.ปยุตฺโต).(2543). พจนานุกรมพุทธศาสตร์ฉบับประมวลศัพท์. พิมพ์ครั้งที่ 9. กรุงเทพฯ: โรงพิมพ์มหาจุฬาลงกรณราชวิทยาลัย.

Allen, Mary J. and Wendy M. Yen. (1979). Introduction to Measurement Theory. California : Brooks/Cole Publishing Company.

Friedenberg. Lisa. (1995). Psychological Testing Design, Analysis, and Use. Massachusetts : Allyn and Bacon.

Kerlinger. Fred N.(1992). Foundation of Behavioral Research. 3rd ed. U.S.A. : Holt, Rinehart and Winston, Inc,. Copyright Renewed.

Kline. P. (1986). A Handbook of Test Construction: Introduction to Psychometric Design. London: Methuen&Co.Ltd.