

นานาปัจจัยที่มีผลต่อความเที่ยง และความตรงของแบบทดสอบ

● สุทัศน์ สุกมลสันต์

คำนำ

ในการทดสอบนั้นไม่ว่าจะเป็นการทดสอบแบบอิงกลุ่ม (norm-referenced test) แบบอิงเกณฑ์ (criterion-referenced test) หรือแบบอิงบริเขต (domain-referenced test) และไม่ว่าจะเป็นการทดสอบตามแนวทฤษฎีการวัดแบบประเพณีนิยม (Classical Test Theory) แนวทฤษฎีการตอบสนองของข้อทดสอบ (Item Response Theory) หรือแม้แต่การทดสอบใหม่ล่าสุดที่ตามแนวทฤษฎีการอ้างอิงสรุป (Generalizability Theory) ก็ตาม ต่างก็ถือว่าความเที่ยง (reliability) และความตรง (Validity) ของแบบทดสอบมีความสำคัญมากที่สุดในการทดสอบและประเมินผล ดังนั้นนักทดสอบจึงมีความเพียรพยายามมากที่จะทำให้ข้อทดสอบและแบบทดสอบมีค่าความเที่ยงและความตรงสูงมากที่สุดเท่าที่จะสามารถทำได้ เนื่องจากมีปัจจัย (factor) นานาประการที่มีผลต่อค่าทั้งสองทั้งทางตรงและทางอ้อม จึงควรที่ครูอาจารย์และนักทดสอบทั้งหลายจะได้ตระหนักไว้ เพื่อหาทางแก้ไขและทำให้ค่าดังกล่าวของแบบทดสอบสูงยิ่งขึ้น

แต่อย่างไรก็ตาม เนื่องจากการทดสอบ มีหลายแบบและอาศัยแนวคิดหลายทฤษฎีดังกล่าวแล้วข้างต้น ซึ่งแต่ละแบบหรือแต่ละทฤษฎีมีแนวคิดเกี่ยวกับความเที่ยงและความตรงแตกต่างกัน ในที่นี้ผู้เขียนจะขอกล่าวถึง ความเที่ยงและความตรงของการทดสอบแบบอิงกลุ่มตามแนวทฤษฎีประเพณีนิยมเท่านั้น เนื่องจากยังเป็นแนวคิดที่ท่านผู้อ่านทั่วไปคุ้นเคยและใช้กันมากที่สุดในขณะนี้ และเพื่อให้ทุกท่านได้เข้าใจตรงกัน ผู้เขียนจะขอกล่าว

ถึงความหมายของความเที่ยงและความตรงนี้ ก่อน รวมทั้งความสัมพันธ์ระหว่างค่าทั้งสองนี้ก่อนที่จะได้กล่าวถึงนานาปัจจัยที่มีผลต่อค่าทั้งสองนี้ต่อไป

ความหมายของความเที่ยงและความตรง

ก. ความเที่ยง

ความเที่ยงของแบบสอบถามกลุ่ม มีความหมายสำคัญ 3 ประการ (Ebel, 1972 : 409 ; Stanley, 1971:374) คือ

1. ความเที่ยง หมายถึง ความคงที่ของผลการสอบของแบบสอบถามฉบับใดฉบับหนึ่งที่ต้องการวัดสิ่งใดสิ่งหนึ่ง

2. ความเที่ยง หมายถึง ร้อยละของความแปรปรวนของสิ่งที่วัดที่แบบสอบถามนั้นสามารถวัดได้

3. ความเที่ยง หมายถึง ค่าความสัมพันธ์ระหว่างผลการสอบแบบสอบถามคู่ขนานที่วัดสิ่งเดียวกัน คู่ใดคู่หนึ่ง

ข. ความตรง

โดยทั่วไปนั้น ความตรงหมายถึง ความถูกต้องของสิ่งที่วัดที่แบบสอบถามนั้นต้องการวัด (Ebel, 1972 : 405) แต่ความหมายเฉพาะของความตรงนั้นแตกต่างกันตามวิธีการคำนวณและจุดมุ่งหมายของการวัด กล่าวคือ (Ebel, 1972: 436 ; Cronbach, 1949 : 103)

1. ความตรงเชิงพยากรณ์ (Predictive Validity) ได้แก่ ความถูกต้องในการพยากรณ์พฤติกรรม หรือสิ่งที่ต้องการวัดที่จะเกิดขึ้นในอนาคตของแบบสอบ ซึ่งได้จากการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบสอบนั้นกับคะแนนการสอบของแบบสอบที่ใช้เป็นเกณฑ์

2. ความตรงร่วมสมัย (Concurrent Validity) ได้แก่ ความถูกต้องในการวัดของแบบสอบถามฉบับหนึ่ง ที่มีความถูกต้องเช่นเดียวกับผลการวัดของแบบสอบถามอีกฉบับหนึ่ง ที่มุ่งวัดสิ่งเดียวกันในปัจจุบัน ค่าความตรงนี้คำนวณได้จากค่าสัมประสิทธิ์สหสัมพันธ์ของการสอบของแบบสอบทั้ง 2 ฉบับ

3. ความตรงเชิงภาวะสันนิษฐาน (Construct Validity) ได้แก่ ความถูกต้องตามภาวะสันนิษฐานของโครงสร้างของสิ่งที่วัด ซึ่งมักได้จากการทดสอบโครงสร้างของสิ่งที่วัดโดยวิธีการทางสถิติวิธีใดวิธีหนึ่ง

4. ความตรงเชิงประจักษ์ (Face Validity) ได้แก่ ความถูกต้องของแบบสอบถามรูปแบบหรือลักษณะของแบบสอบที่ควรจะเป็นเท่าที่สังเกตได้โดยตาเปล่า

5. ความตรงเชิงเนื้อหา (Content Validity) ได้แก่ ความถูกต้อง และเพียงพอของสิ่งที่มุ่งวัด

ความสัมพันธ์ระหว่างความเที่ยงและความตรงของแบบสอบ

แบบทดสอบที่มีความตรงจำเป็นอย่างยิ่งที่จะต้องมีความเที่ยงด้วย ดังนั้น ค่าความเที่ยงจึงมีความจำเป็นต่อค่าความตรง แต่แบบทดสอบที่มีความเที่ยงสูงอาจมีค่าความตรงต่ำ เพราะแบบทดสอบนั้นอาจไม่ได้วัดสิ่งที่มุ่งทำการสอบก็ได้ (Ebel, 1972 : 444) ตามทฤษฎีการทดสอบแบบประเพณีนิยม แบบสอบที่มีความยาวเพิ่มขึ้น ด้วยข้อสอบที่วัดสิ่งเดียวกัน และมีคุณสมบัติเหมือนกับข้อสอบเดิมที่มีอยู่ จะทำให้ค่าความเที่ยงเพิ่มขึ้น และทำให้ค่าความตรงเพิ่มขึ้นด้วย แต่ค่าสูงสุดของความตรงมีได้ไม่เกินรากที่สองของค่าความเที่ยง (r_{tt}) ของแบบสอบนั้น (Magnuson, 1966 : 150)

ปัจจัยที่มีผลต่อความเที่ยงและความตรง

ปัจจัย (factors) ที่มีผลกระทบต่อค่าความเที่ยงและความตรงของแบบทดสอบนั้นมีมากมาย ผู้สร้างและผู้ใช้แบบทดสอบจึงควรพึงระวังถึงผลกระทบเหล่านี้ด้วยเสมอ เพื่อหลีกเลี่ยงหรือจำกัดขอบเขตของปัจจัยเหล่านี้ให้ได้มากที่สุด เพื่อให้ผลการสอบมีความเที่ยงและความตรงสูงที่สุด ปัจจัยต่าง ๆ ดังกล่าวนี้น่าที่สามารถรวบรวมได้มีดังต่อไปนี้

ก. ปัจจัยที่มีผลต่อความเที่ยง

ปัจจัยที่มีผลต่อความเที่ยงของแบบสอบมีหลายอย่าง เช่น ความยาวของแบบสอบ เวลาในการทำแบบสอบ ความเป็นเอกพันธ์ของกลุ่มตัวอย่าง ความยากของแบบสอบ ความเป็นปรนัย การกระจายของคะแนนสอบ และวิธีในการคำนวณหาค่าความเที่ยง เป็นต้น (Gronlund, 1976 : 117-122 ; Mehrens and Lehman, 1975 : 100-103) ปัจจัยดังกล่าวนี้มีอิทธิพลต่อความเที่ยง ดังนี้

1. ความยาวของแบบสอบ

เมื่อแบบสอบมีความยาวเพิ่มขึ้น ก็จะมีค่าความเที่ยงเพิ่มขึ้น ซึ่งอาจแสดงจากสูตรต่อไปนี้

$$r_{tt} = \frac{kr}{1 + (k-1)r}$$

r_{tt} = ค่าความเที่ยงที่คาดหวัง

เมื่อแบบสอบยาวขึ้น k เท่า

k = อัตราส่วนของข้อสอบใหม่ที่เพิ่มขึ้น

r = ค่าความเที่ยงเดิม

สมมติว่า แบบสอบชุดหนึ่งมีค่าความเที่ยง .60 ถ้าเพิ่มข้อสอบที่วัดสิ่งเดียวกัน และมีลักษณะคล้ายคลึงกันอีก 3 เท่า แบบสอบนั้นจะมีค่าความเที่ยงใหม่ ดังนี้

$$r_{tt} = \frac{3 \times .60}{1 + (2 \times .60)} = .818$$

2. เวลาในการทำแบบสอบ

ถ้าเวลาในการทำแบบสอบไม่เหมาะสม จะมีผลต่อค่าความเที่ยงของแบบสอบ เช่น ถ้าเวลาน้อยเกินไป ผู้สอบทำไม่ทัน ก็อาจเดาคำตอบ ซึ่งจะทำให้การกระจายของคะแนนแตกต่างจากแบบสอบที่ให้เวลาเพียงพอ แต่ถ้าหากว่า แบบสอบให้เวลาผู้สอบมากเกินไป จะทำให้คนที่เรียนเก่งและเรียนอ่อนทำแบบสอบได้คะแนนเท่า ๆ กัน คะแนนจะมีการกระจายน้อย และมีผลต่อความเที่ยงของแบบสอบเช่นกัน ในปี 1932 Cook (อ้างจาก Hopkins, 1964 : 271) พบว่าเวลาในการสอบมีผลต่อความเที่ยงอย่างมีนัยสำคัญ กล่าวคือแบบสอบการสะกดคำฉบับหนึ่ง ถ้าให้เวลาพอดีจะมีค่าความเที่ยง .91-.92 แต่ถ้าให้เวลามากเกินไป ความเที่ยงจะลดลงเป็น .71-.85 เท่านั้น

3. ความเป็นเอกพันธ์ของกลุ่มตัวอย่าง

ถ้าปัจจัยอย่างอื่นที่มีผลต่อแบบสอบ 2 ฉบับเท่าเทียมกัน แบบสอบที่สอบกับกลุ่มตัวอย่างวิวิธพันธ์มากเท่าใด ก็ยังมีค่าความเที่ยงมากขึ้นเท่านั้น ทั้งนี้จะสังเกตได้จากสูตรทั่วไปของค่าความเที่ยงดังนี้

$$r_{tt} = 1 - \frac{\sigma_c^2}{\sigma_t^2}$$

ในเมื่อ σ_c^2 = ความแปรปรวนของความคลาดเคลื่อน

σ_t^2 = ความแปรปรวนของคะแนนสอบทั้งหมด

เนื่องจาก σ_c^2 เป็นความแปรปรวนของคะแนนที่ปรากฏ (observed score) ของกลุ่มตัวอย่างเกี่ยวกับคะแนนจริง (true score) ว่ามีความคลาดเคลื่อนมากน้อยเพียงใด ดังนั้น σ_c^2 จึงเป็นค่าที่คงที่ แม้ว่ากลุ่มตัวอย่างจะเปลี่ยนไป แต่ค่า σ_c^2 จะมีมากเพิ่มขึ้นเมื่อกลุ่มตัวอย่างมีความแตกต่างกันมากขึ้น จึงทำให้ค่าความเที่ยงมีมากเพิ่มขึ้น ดังนั้น ถ้าแบบสอบนั้นใช้สอบกับกลุ่มเอกพันธ์ เช่น ใช้กับกลุ่มตัวอย่างที่คัดสรรแล้ว เช่น กลุ่มตัวอย่างที่สอบผ่านเข้ามหาวิทยาลัยแล้วในขณะใดขณะหนึ่งแล้ว ค่าความเที่ยงของแบบสอบจะลดลง (Cronbach, 1949 : 133)

4. ความสัมพันธ์ซึ่งกันและกันของข้อสอบ

แบบสอบใดที่ข้อสอบมีความสัมพันธ์ซึ่งกันและกันสูง แบบสอบนั้นจะมีค่าความเที่ยงสูง (Guilford, 1954 : 360) ในปี 1971 Bowers (Bowers, 1971 : 427) ได้เสนอสูตรในการคำนวณค่าความเที่ยงโดยอาศัยค่าสหสัมพันธ์ซึ่งกันและกันของข้อสอบดังนี้

$$r_{tt} = \frac{n r_{ij}}{1 + (n - 1) r_{ij}}$$

ในเมื่อ r_{ij} = ค่าเฉลี่ยของค่า
สัมประสิทธิ์สหสัมพันธ์
ซึ่งกันและกันของข้อสอบ
 n = จำนวนข้อสอบ

นอกจากนี้แบบสอบที่ข้อสอบมีความสัมพันธ์กับคะแนนรวมยิ่งมาก แบบสอบนั้นก็ยิ่งมีค่าความเที่ยงมาก ในปี 1969 Gaylord (Gaylord, 1969: 303) ได้เสนอสูตรในการคำนวณหาค่าความเที่ยง โดยอาศัยค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบ กับคะแนนรวมไว้ดังต่อไปนี้

$$r_{rr} = \frac{nr_{it}^2 - 1}{(n-1)r_{it}^2}$$

ในเมื่อ n = จำนวนข้อสอบ
 r_{it} = ค่าเฉลี่ยของค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบกับคะแนนรวม เช่น ค่า r_{bis} เป็นต้น

นอกจากนี้ Gaylord ได้พิสูจน์ว่า สูตรที่ Guilford เสนอไว้เพื่อการคำนวณหาค่าความเที่ยงจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างข้อสอบกับคะแนนรวมนั้นไม่ถูกต้องคือสูตรต่อไปนี้

$$r_{rr} = \frac{nr_{it}^2}{1 + (n-1)r_{it}^2}$$

5. ความยากง่ายของข้อสอบ

ข้อสอบที่ยากเกินไปหรือง่ายเกินไป ไม่มีอำนาจจำแนก แบบสอบที่มีข้อสอบที่ง่ายเกินไปหรือยากเกินไปจำนวนมาก จะมีค่าความเที่ยงต่ำ Gronlund (Gronlund, 1976: 121) แนะนำว่า แบบสอบเพื่อการคัดเลือกที่ดีควรมีค่าความยาก โดยเฉลี่ยเท่ากับค่าจุดกึ่งกลางระหว่างค่าโอกาสการเดา กับจำนวนข้อสอบทั้งหมด เช่น ถ้าข้อสอบแบบเลือกตอบมี 4 ตัวเลือก ยาว 50 ข้อ ความยากง่ายที่เหมาะสมควรเป็น $\frac{(25-50)}{2} + 25 = 38$ (หรือ $p = .38$) เป็นต้น ทั้งนี้เพื่อให้ได้แบบสอบที่มีค่าความเที่ยงสูง Cureton (Cureton, 1966: 15) เสนอแนะว่า แบบสอบที่มีข้อสอบง่าย ๆ อยู่ไม่ควรหาค่าความเที่ยงโดยวิธี KR_{21} เพราะค่าที่คำนวณได้จะต่ำกว่าความเป็นจริง ทั้งนี้เพราะว่า ค่าความแปรปรวน (σ^2) จะคงที่ แต่หาค่าความแปรปรวนรายข้อเพิ่มขึ้น ค่าจึงต่ำกว่าความเป็นจริง

6. ความเป็นปรนัยในการให้คะแนน

แบบสอบที่ให้คะแนนไม่เป็นปรนัย ค่าความเที่ยงจะไม่แน่นอน ดังนั้นควรทำเครื่องมือในการวัดให้การคิดคะแนนมีลักษณะเป็นปรนัยให้ได้มากที่สุดเท่าที่จะทำได้

7. วิธีการคิดค่าความเที่ยง

การคิดค่าความเที่ยงมีหลายวิธี แต่ละวิธีให้ผลต่างกัน และแต่ละวิธีก็เหมาะสมสำหรับจุดมุ่งหมายเฉพาะอย่างในการตีความของค่า

ความเที่ยง ค่าความเที่ยงที่คิดแบบแบ่งครึ่ง (Split-half method) มักให้ค่าสูงที่สุด ค่าความเที่ยงที่คิดแบบการสอบซ้ำ (Test-retest method) มักให้ค่าปานกลาง และค่าความเที่ยงที่คิดแบบ Kuder-Richardson มักให้ค่าต่ำ เป็นต้น (Gronlund, 1976 : 122)

8. การเดา

การเดามีผลต่อค่าความเที่ยงมาก เพราะการเดาเป็นแหล่งทำให้เกิดความคลาดเคลื่อนในการวัดมาก Magnusson (Magnusson, 1966 : 101) กล่าวว่า แหล่งความคลาดเคลื่อนในการวัดที่สำคัญ มี 3 แหล่ง คือ การบริหาร การสอบ การคิดคะแนน และการเดา ถ้ามีการเดามากแบบสอบจะมีค่าความเที่ยงลดลง Mattson (Mattson, 1965 : 729) ได้เสนอสูตรในการทำนายค่าความเที่ยงของแบบสอบที่มีการเดาถูกดังนี้

$$\frac{r_{rr}p_c(1-p_c)}{1-r_{rr}} = n(1-\pi)^2s_{pk}^2$$

- ในเมื่อ p_c = ค่าเฉลี่ยของกลุ่มผู้สอบที่ทำแบบสอบได้ถูกต้อง
- r_{rr} = ค่าความเที่ยงเดิม
- n = จำนวนข้อสอบ
- π = โอกาสที่ผู้สอบทั้งกลุ่มเดาได้ถูกต้อง
- s_{pk}^2 = ค่าความแปรปรวนคะแนนที่ทำถูกโดยไม่ได้เดา

ดังนั้น ถ้าแบบสอบชุดหนึ่งมี 100 ข้อ ผู้สอบโดยเฉลี่ยแล้วทำถูกร้อยละ 50 ค่าความแปรปรวนของคะแนนที่ผู้สอบทำถูกโดยไม่ได้เดา = .01 แบบสอบชุดนี้จะมีค่าความเที่ยง = .80 แต่ถ้าแบบสอบชุดเดียวกันนี้มีโอกาสเดาได้ถูกต้องร้อยละ 50 ค่าเฉลี่ยของกลุ่มผู้สอบจะทำข้อสอบได้ถูกร้อยละ 75 และความแปรปรวนของคะแนนที่ทำถูก = .10 แล้ว แบบสอบชุดเดิมนี้อาจมีค่าความเที่ยงเพียง .571 เท่านั้น

9. คำสั่งในการทำแบบสอบ

คำสั่งในการทำแบบสอบเป็นตัวแปรที่สำคัญมากในการทำให้แบบสอบมีค่าความเที่ยงเปลี่ยนไปได้ โดยเฉพาะอย่างยิ่งคำสั่งที่ให้ผู้สอบเดาข้อสอบที่ทำไม่ได้ จากการศึกษารายชื่อของ Sax และ Collet (Sax and Collet, 1968 : 134) ใช้กลุ่มตัวอย่าง 240 คน แบ่งเป็น 3 กลุ่มเท่ากันแล้ว ให้กลุ่มตัวอย่างทำแบบสอบ Nelson Test of Mental Ability ซึ่งยาว 50 ข้อ คำสั่งมี 3 อย่างคือ 1) บอกรให้ผู้สอบเดาคำตอบข้อสอบที่ทำไม่ได้ และไม่หักคะแนนการเดา 2) ห้ามผู้สอบที่ทำไม่ได้เดาคำตอบข้อสอบ ถ้าเดาผิดจะหักข้อละ $\frac{1}{4}$ คะแนน (แบบสอบมี 4 ตัวเลือก) และ 3) ห้ามผู้สอบที่ทำไม่ได้เดา แต่ถ้าบอกว่าตัวเลือกใดบ้างที่ผิดจะได้ข้อละ $\frac{1}{4}$ คะแนน ผลจากการวิจัยพบว่าค่าความเที่ยงของแบบสอบที่มีคำสั่งอย่างนี้

2 และ 3 เพิ่มขึ้นจากเดิมอย่างมีนัยสำคัญ นอกจากนี้ จากการศึกษาของ Traub และ Hambleton (Traub and Hambleton, 1972 : 737) พบว่า คำสั่งในการทำแบบสอบ 2 อย่าง คือ 1) ห้ามไม่ให้ผู้สอบเดาคำตอบที่ไม่รู้และให้รางวัลบางอย่างสำหรับข้อที่ทำไม่ได้แล้วเว้นไว้ และ 2) ให้เดาคำตอบได้ แต่ถ้าผิดจะหักคะแนน ผลปรากฏว่า แบบสอบศัพท์ยาว 50 ข้อ ที่ใช้คำสั่งที่ 1 มีค่าความเที่ยงเพิ่มขึ้นจากแบบสอบที่ใช้คำสั่งที่ 2 อย่างมีนัยสำคัญ ต่อมาเขาได้เพิ่มคำสั่งที่ 3 ขึ้นอีก คือ ให้ผู้สอบทำข้อสอบทุกข้อ หากทำไม่ได้ให้เดาคำตอบ จากการศึกษาในกลุ่มตัวอย่าง 1,091 คน โดยใช้แบบสอบคณิตศาสตร์ และแบบสอบศัพท์ยาว 20 ข้อ เป็นเครื่องมือ ผลปรากฏว่า แบบสอบที่ใช้คำสั่งที่ 1 มีค่าความเที่ยงสูงกว่าแบบสอบที่ใช้คำสั่งที่ 2 และ 3 อย่างมีนัยสำคัญ แบบสอบที่ใช้คำสั่งที่ 3 มีค่าความเที่ยงต่ำที่สุด

10. การคิดค่าน้ำหนักคะแนนของแบบสอบ

การให้ค่าน้ำหนักของข้อสอบรายข้อ และรายตัวเล็อก เป็นวิธีหนึ่งก็ตามทฤษฎีแล้วเชื่อว่า จะทำให้แบบสอบมีคะแนนกระจายเพิ่มมากขึ้นกว่าการให้คะแนนวิธีประเพณี ดังนั้น จะทำให้แบบสอบมีความเที่ยงมากกว่าการให้คะแนนแบบ 0,1 (Wang and Stanley, 1970 : 663)

11. ปัจจัยอื่น ๆ ที่มีผลต่อความเที่ยงปรាកฏู่ขนาน (Parallel Extrinsic Reliability)

Hopkins (Hopkins, 1964 : 271-281) ได้ทำการศึกษาปัจจัยต่าง ๆ ที่มีผลต่อความเที่ยงของแบบสอบ ซึ่งยังไม่มีวัดออกมาได้แน่นอน แต่มีความสำคัญต่อการสอบ เขาเรียกความเที่ยงนี้ว่า “ความเที่ยงปรากฏู่ขนาน” ปัจจัยต่าง ๆ นี้ได้แก่

1. ความเร็วในการทำแบบสอบ
2. การเดา
3. รูปแบบ (format) ของแบบสอบ
4. เวลาที่กำหนดให้ทำแบบสอบ
5. กิจนิสัยในการทำแบบสอบ (response style/response set)
6. คำสั่งในการทำแบบสอบ
7. การทำข้อสอบไม่ครบทุกข้อ
8. ระยะเวลาในการสอบซ้ำ
9. เวลาหรือโอกาสในการทำแบบสอบ
10. ความเหน็ดเหนื่อยและความเบื่อหน่ายในการทำแบบสอบ

Hopkins กล่าวว่า แบบสอบอาจมีค่าความเที่ยงวิธีประเพณีนิยม (Conventional Reliability) เท่ากัน แต่มีความเที่ยงปรากฏู่ขนานแตกต่างกันก็ได้

ข. ปัจจัยที่มีผลต่อความตรง

ปัจจัยที่มีผลต่อความตรงของแบบสอบมีมาก ปัจจัยบางอย่างนอกจากมีผลต่อค่าความเที่ยงแล้ว ยังมีผลต่อความตรงของแบบสอบ

ด้วย เช่น การเดา และคำสั่งในการทำแบบสอบ เป็นต้น ปัจจัยที่สำคัญต่อความตรงมีดังนี้ (Gronlund, 1976 : 98-102)

1. ปัจจัยที่มีผลต่อความตรง

ก. การเดา

จากการศึกษาของ Lord (Lord, 1964 : 745) เมื่อใช้กลุ่มตัวอย่าง 2,000 คน ทำแบบสอบศัพท์ยาว 25 50 และ 60 ข้อ โดยมีคำสั่งให้ผู้ตอบเดาคำตอบได้ถ้าไม่รู้คำตอบ ผลปรากฏว่า การบังคับให้เดาอย่างไม่มีระบบทำให้แบบสอบมีค่าความตรงเพิ่มขึ้น ถ้าการเดามีมากและผลของการเดาสัมพันธ์กับเกณฑ์และการบังคับให้เดาอย่างมีระบบ ทำให้แบบสอบมีค่าความตรงเพิ่มขึ้นอย่างมีนัยสำคัญด้วยเช่นกัน ดังนั้น Lord จึงแนะนำว่า การใช้คำสั่งให้ผู้สอบเดาคำตอบข้อที่ไม่รู้นั้นควรระวัง เพราะอาจทำให้ค่าความตรงเพิ่มขึ้นได้ แต่จากการศึกษาของ Sax และ Collet (Sax and Collet, 1968 ; 1127) พบว่า แบบสอบที่ห้ามผู้สอบเดาคำตอบ ถ้าเดาผิดจะหักคะแนนและแบบสอบที่ห้ามผู้สอบเดาคำตอบ แต่ถ้าบอกว่าตัวเลือกใดผิดได้ถูกต้อง จะให้คะแนนเพิ่ม ต่างก็มีค่าความตรงไม่แตกต่างกันอย่างมีนัยสำคัญ จากแบบสอบที่อนุญาตให้ผู้สอบเดาคำตอบได้

ข. กิจนิสัยในการทำแบบสอบ

ผู้สอบมักมีนิสัยในการทำแบบสอบต่างกัน เช่น บางคนทำข้อสอบเลือกตอบไม่ได้ จะเลือกข้อ ค เป็นคำตอบที่ถูก บางคนจะเลือกข้อ ก เป็นข้อถูกเสมอ เป็นต้น ผลของการทำแบบสอบเช่นนี้ทำให้ไม่อาจทราบได้ว่า แบบสอบสามารถวัดสิ่งที่ต้องการจะวัดได้มากน้อยเพียงใด

ค. ความไม่พร้อมทางร่างกายและจิตใจ

ความวิตกกังวล ความเจ็บป่วย ความเบื่อหน่าย และสภาพของร่างกายและจิตใจไม่ปกติต่าง ๆ ในระหว่างการสอบ ทำให้การสอบได้ไม่ดีเท่ากับการสอบในสภาวะปกติ

2. ปัจจัยที่เกิดจากการบริหารการสอบ

ก. การให้คะแนน

การให้คะแนนที่ไม่เป็นปรนัย ทำให้ผลการสอบได้รับการตัดสินไม่แน่นอน จึงทำให้แบบสอบมีความตรงไม่แน่นอน

ข. การบริหารการสอบ

สภาพต่าง ๆ ในการทำแบบสอบมีผลต่อความตรงของการสอบ เช่น ห้องสอบร้อนเกินไป แสงสว่างไม่เพียงพอ มีเสียงดังมาก ผู้บริหารการสอบให้คำแนะนำในการสอบได้ไม่ดี เช่น เสียงไม่ดังพอ หรือทำความเข้าใจให้กับผู้สอบได้ไม่ชัดเจนเพียงพอ เป็นต้น

ก. ระยะเวลาในการสอบ

ระยะเวลาในการสอบ นอกจากจะมีผลต่อความเที่ยงของแบบสอบแล้ว ยังมีผลต่อความตรงด้วย Boxter (อ้างจาก Kandall, 1964 : 789) พบว่า การให้เวลาในการทำแบบสอบนานเกินไป ทำให้แบบสอบมีความตรงลดลงอย่างมีนัยสำคัญ เขากล่าวว่า เวลาในการสอบนานเกินไป หรือน้อยเกินไป ทำให้ผู้สอบเดาคำตอบข้อที่ทำไม่ได้ ซึ่งมีผลต่อค่าความเที่ยงและความตรงด้วย แต่จากการศึกษาของ Kandall (Kandall, 1964 : 789) โดยใช้กลุ่มตัวอย่าง 739 คน ซึ่งเป็นทหารในกองทัพบก พบว่าแบบสอบที่ให้เวลาสอบมากพอดีจะมีค่าความตรงสูงที่สุด แต่ถ้าให้เวลามากเกินไป ค่าความเที่ยงจะคงที่ และถ้าให้เวลาน้อยเกินไป ค่าความเที่ยงจะลดลงอย่างมีนัยสำคัญ

3. ปัจจัยที่เกิดจากแบบสอบ

ก. คำสั่งในการทำแบบสอบไม่ชัดเจน

คำสั่งที่ไม่ชัดเจนทำให้ผู้สอบเข้าใจผิดได้ง่าย ซึ่งอาจเกิดจากไม่มีตัวอย่างประกอบใช้ภาษาไม่รัดกุม และพิมพ์ไม่ชัดเจน เป็นต้น

ข. คำศัพท์และโครงสร้างของภาษายากเกินไป

คำศัพท์และโครงสร้างของภาษาที่ใช้ยากเกินไป อาจทำให้ผู้สอบไม่เข้าใจส่วนที่ไม่เกี่ยวข้อง

กับสิ่งที่ต้องการสอบ แต่เสียเวลามาก จนทำแบบสอบได้ไม่เสร็จทันเวลา

ค. ข้อสอบมีความยากไม่เหมาะสม

ในกรณีแบบสอบอิงกลุ่ม ข้อสอบที่ยากเกินไปหรือง่ายเกินไป ทำให้แบบสอบมีค่าความตรงต่ำ และในกรณีของการสอบอิงเกณฑ์ ข้อสอบที่ง่ายมากเกินไปอาจไม่เหมาะสมกับจุดมุ่งหมายในการสอบ ดังนั้น จะทำให้แบบสอบมีความตรงต่ำด้วยเช่นเดียวกัน

ง. ข้อสอบไม่ดี

ข้อสอบแบบเลือกตอบ ถ้าสร้างไม่ดีตัวเลือกอาจแนะนำคำตอบได้โดยผู้ออกข้อสอบไม่ได้ตั้งใจ ทำให้ข้อสอบวัดสิ่งที่ต้องการวัดอย่างแท้จริงไม่ได้

จ. ข้อสอบกำกวม

ข้อสอบที่ใช้ภาษากำกวม อาจทำให้ผู้สอบที่เก่งทำไม่ได้ แต่ผู้สอบที่ไม่เก่งทำได้ และอาจทำให้ผู้สอบเข้าใจความหมายของภาษาผิดพลาดได้

ฉ. ชนิดของแบบสอบที่ใช้ไม่เหมาะสม

แบบสอบที่ไม่เหมาะสม อาจวัดสิ่งที่ต้องการวัดได้ไม่ดีเพียงพอ ทำให้ไม่อาจประเมินสิ่งที่ต้องการวัดได้อย่างถูกต้อง เช่น การสอบความสามารถในการเขียนก็ควรใช้แบบสอบที่ให้ผู้สอบเขียนมากกว่าจะให้เลือกคำตอบจากแบบสอบแบบเลือกตอบเท่านั้น เป็นต้น

ข. แบบสอบสั้นเกินไป

แบบสอบที่สั้นเกินไปอาจทำให้วัดสิ่งที่ต้องการวัดได้ไม่มาก และไม่ครอบคลุมเนื้อหาที่ต้องวัด ทำให้การประเมินผลการสอบผิดพลาดได้ เพราะสิ่งที่วัดไม่เป็นตัวแทนของเนื้อหาที่จะวัด

นอกจากนี้เป็นที่ทราบกันแล้วว่า ความตรงของแบบสอบขึ้นอยู่กับความเที่ยงและความเที่ยงของแบบสอบขึ้นอยู่กับความยาวของแบบสอบ ซึ่งเป็นปัจจัยที่สำคัญอย่างหนึ่ง ดังนั้น ความตรงของแบบสอบก็ขึ้นอยู่กับความเที่ยงด้วย Guilford ได้เสนอสูตรในการคำนวณค่าความตรงของแบบสอบที่มีความยาวเพิ่มขึ้น n เท่าดังนี้ (Guilford, 1954 : 407)

$$r_{nx,y} = \frac{r_{xy}}{\sqrt{\frac{1-r_{xx}}{n} + r_{xx}}}$$

ในเมื่อ $r_{nx,y}$ = ค่าความตรงของแบบสอบที่มีความยาวเพิ่มขึ้น

r_{xy} = ค่าสัมประสิทธิ์สหสัมพันธ์ของแบบสอบกับเกณฑ์

n = จำนวนเท่าของข้อสอบที่เพิ่มขึ้น

r_{xx} = ค่าความเที่ยงของแบบสอบ

ข. กระสวนของคำตอบของแบบสอบ

ถ้าผู้สอบสามารถรู้ได้ว่า กระสวนของคำตอบของแบบสอบเป็นอย่างไร เช่น คำตอบที่ถูกต้อง คือ ข้อ ค ทุกข้อ จะทำให้ผู้สอบเลือกคำตอบตามกระสวนของคำตอบที่ตนคิดว่าถูกต้องแล้ว ทำให้การวัดไม่อาจวัดสิ่งที่ต้องการจะวัดได้อย่างแท้จริง ดังนั้น แบบสอบจะมีความตรงต่ำ

4. ปัจจัยเกิดจากการสอนของครู

ปัจจัยที่เกิดจากครูที่มีผลต่อความตรงของแบบสอบก็คือ เนื้อหาที่สอนกับสิ่งที่สอบไม่สัมพันธ์กัน เช่น ถ้าต้องการสอบความรู้ความสามารถในการคำนวณ แต่คำตอบและวิธีการได้เฉลยให้ผู้สอบรู้แล้ว จึงวัดได้แต่ความจำเท่านั้น หรือสอบสิ่งที่ไม่ได้สอนเลย เป็นต้น

5. ปัจจัยที่เกิดจากกลุ่มตัวอย่างและเกณฑ์ในการอ้างอิง

ก. กลุ่มตัวอย่าง

ค่าความตรงและค่าความเที่ยงเป็นค่าของแบบสอบที่เกิดจากการสอบแต่ละครั้ง จะเปลี่ยนแปลงไปตามลักษณะของผู้สอบ เช่น ถ้าผู้สอบมีจำนวนมาก ความรู้ความสามารถแตกต่างกัน ค่าความตรงหรือความเที่ยงก็จะมีค่าอย่างหนึ่ง แต่ถ้ากลุ่มตัวอย่างมีขนาดเล็ก ความสามารถไม่แตกต่างกันมาก ก็จะทำให้แบบสอบมีค่าความตรงและความเที่ยงเปลี่ยนแปลง

ไป นอกจากนี้ เพศ อายุ พื้นฐานทางการศึกษา และพื้นฐานทางเศรษฐกิจและสังคมก็มีส่วนทำให้แบบสอบมีค่าความตรงและความเที่ยงแตกต่างกันได้

ข. เกณฑ์ในการอ้างอิง

ค่าความตรงของแบบสอบที่ต้องอาศัยเกณฑ์ในการอ้างอิง จะมีค่าแตกต่างกัน แล้วแต่ความสัมพันธ์ระหว่างเกณฑ์กับแบบสอบนั้นว่า จะมีมากน้อยเพียงใด ในการเลือกเกณฑ์นั้น ควรพิจารณาว่า ตามความเป็นจริงแล้ว เกณฑ์กับแบบสอบนั้นสามารถวัดสิ่งที่ต้องการวัดได้มากน้อยเพียงใด เช่น แบบสอบวัดความสามารถในการคิดเชิงวิพากษ์วิจารณ์ (critical thinking) สัมพันธ์กับผลการเรียนวิชาทางสังคมศาสตร์มากกว่าวิชาที่อาศัยแต่เพียงความรู้ความจำเท่านั้น เป็นต้น

สรุป

ดังนั้นจากรายละเอียดต่าง ๆ ดังได้กล่าวมาแล้วตั้งแต่ต้น จะเห็นได้อย่างชัดเจนว่าความเที่ยงและความตรงเป็นคุณลักษณะที่สำคัญยิ่งของแบบสอบที่ใช้วัดผลและประเมินผลแบบต่าง ๆ และตามแนวทฤษฎีต่าง ๆ และจะเห็นได้ว่ามีปัจจัยมากมายหลายอย่างที่มีผลโดยตรงและโดยอ้อมต่อค่าทั้งสองนี้ อันจะทำให้แบบสอบมีค่าความเที่ยงและความตรงต่ำลง

หรือกล่าวอีกนัยหนึ่งก็คือทำให้แบบสอบ การวัดและการประเมินผลไม่สามารถที่จะทดสอบสิ่งที่มุ่งทดสอบได้อย่างถูกต้อง และเชื่อถือได้ซึ่งจะมีผลโดยตรงกับคุณภาพของการตัดสินผลการสอบ และคุณภาพของการศึกษาในส่วนที่ท่านรับผิดชอบด้วยอย่างหลีกเลี่ยงไม่ได้ ด้วยเหตุนี้ท่านที่เกี่ยวข้องกับการทดสอบ การวัดและการประเมินผลการเรียนการสอนจึงควรอย่างยิ่งที่จะได้ตระหนักกับนานาปัจจัยที่มีผลต่อความเที่ยงและความตรงของแบบสอบดังได้กล่าวมาแล้ว และการหาทางทั้งป้องกัน (บางปัจจัย) และปรับปรุง (บางปัจจัย) เพื่อให้ผลการทดสอบมีค่าดังกล่าวสูงที่สุดเท่าที่จะสามารถทำได้ เพื่อที่ท่านจะได้ไม่เป็นผู้สร้างบาปให้แก่ตัวของตนเองโดยไม่รู้ตัวอันเนื่องมาจากการตัดสินผลการสอบผิดพลาด อันเนื่องมาจากคุณภาพของแบบสอบต่ำโดยที่ท่านไม่ได้ตระหนักไว้แต่แรกหรือไม่รู้!

เอกสารอ้างอิง

- Bowers, J. "A Note on Gaylord's Estimating Test Reliability from Item-test Correlation." *Educational and Psychological Measurement* 31 (Summer 1971) : 427-429.
- Cronbach, L.J. and Others. *The Dependability of Behavioral Measurements*. New York : John Wiley & Sons, Inc., 1972.
- Cureton, E.E. "Kuder-Richardson Reliability of Classroom Test." *Educational and Psychological Measurement* 26 (Spring 1966) : 13-15.
- "Reliability and Validity : Basic Assumptions and Experimental Design." *Educational and Psychological Measurement* 15 (Summer 1965) : 327-346.
- Ebel, R.L. *Essentials of Educational Measurement*. New Jersey : Prentice-Hall, Inc., 1972.
- Gaylord, R.H. "Estimating Test Reliability from the Item-test Correlation." *Educational and Psychological Measurement* 29 (Summer 1969) : 303-304
- Gronlund, N.E. *Measurement and Evaluation in Teaching*. New York : MacMillan, 1976.
- Guilford, J.P. *Psychometric Methods*. New York : McGraw-Hill Book Co., 1954.
- Hopkins, K.D. "Extrinsic Reliability : Estimating and Attenuating Variance from Response Styles, Chance and Other Relevant Sources." *Educational and Psychological Measurement* 24 (Summer 1964) : 271-281.
- Kandall, L.M. "The Effects of Varying Time Limits on Test Validity." *Educational and Psychological Measurement* 24 (Winter 1964) : 789-799.
- Lord, F.M. "The Effects of Random Guessing on Test Validating." *Educational and Psychological Measurement* 24 (Winter 1964) : 745-747.
- Magnusson, D. *Test Theory*. Stockholm : Stockholm University Press, 1966.
- Mattson, D. "The Effects of Guessing on the Standard Error of Measurement and the Relationship of Test Scores." *Educational and Psychological Measurement* 25 (Autumn 1965) : 727-732.
- Mehrens, W.A. and Lehmann, I.J. *Measurement and Evaluation in Education and Psychology*. New York : Holt, Rinehart and Winston, 1978.
- Patnaik, D. and Traub, R.E. "Differential Weighting by Judged Degree of Correctness." *Journal of Educational Measurement* 10 (Winter 1973) : 281-286.
- Sax, G. and Collet, L.V. "The Effects of Differing Instructions and Guessing Formula on Reliability and Validity." *Educational and Psychological Measurement* 28 (Winter 1968) : 1127-1136.
- Stanley, J.C. "Reliability." In *Educational Measurement*, pp. 356-441. Edited by R.L. Thorndike. Washington. D.C. : American Council on Education, 1971.
- Wang, M.W. and Stanley, J.C. "Differential Weighting : A Review of Methods and Empirical Studies." *Review of Educational Research* 40 (1970) : 663-705.