บทที่ 6
ความเชื่อมั่นและความเที่ยงตรง

ความสำคัญของความเชื่อมั่น

ความเชื่อมั่นนับว่าเป็นคุณลักษณะที่สำคัญของเครื่องมือในการวัดทุกชนิด กล่าวคือ ถ้าเครื่องมือในการวัดขาดความเชื่อมั่นแล้ว ผลที่ได้จากการวัดก็จะไม่มีความหมาย

แบบทดสอบที่มีความเชื่อมั่นสูง หมายถึงแบบทดสอบที่นำไปสอบกับนักเรียนกี่ครั้ง ๆ ผู้เข้าสอบก็จะได้คะแนนคงที่ แน่นอน ไม่เปลี่ยนแปลง หรือเปลี่ยนแปลงไปจากเดิมแต่เพียงเล็กน้อย

ในการวัดทางการศึกษานั้น มักจะพบปัญหาในการวัดมากมาย และปัญหาที่สำคัญอย่างหนึ่งก็คือ ความเชื่อมั่นในการวัดทางการศึกษาต่ำกว่าความเชื่อมั่นในการวัดทางกายภาพ

ดร.อนันต์ ศรีโสภา ได้สรุปสาเหตุที่ทำให้ค่าความเชื่อมั่นในการวัดทางการศึกษาต่ำไว้ดังนี้

1. การวัดทางกายภาพนั้นสามารถวัดได้โดยตรง ไม่เหมือนกับการวัดทางการศึกษา ซึ่งไม่สามารถวัดได้โดยตรง จึงทำให้ผลที่ได้จากการวัดไม่แน่นอน

2. เครื่องมือที่ใช้วัดทางกายภาพมีความถูกต้องชัดเจนกว่าเครื่องมือวัดทางการศึกษา

3. คุณลักษณะของสิ่งที่ถูกวัดทางกายภาพมีความคงที่แน่นอน ส่วนคุณลักษณะของสิ่งที่ถูกวัดทางการศึกษาเปลี่ยนแปลงอยู่ตลอดเวลา

ทฤษฎีเกี่ยวกับคะแนนจริง

คะแนนที่ได้จากการสอบ หรือคะแนนที่สังเกตได้ (observed score) จะประกอบด้วยคะแนน 2 ส่วน คือ คะแนนจริง (true score) และส่วนที่เป็นคะแนนอันเกิดจากความคลาดเคลื่อน (error score) คะแนนจริง คือส่วนของคะแนนที่วัดได้โดยไม่ถูกอิทธิพลของความคลาดเคลื่อน คะแนนส่วนนี้ถ้าวัดในสิ่งเดียวกันจะมีค่าคงที่เสมอไม่ว่าจะวัดกี่ครั้ง ๆ ก็ตาม คะแนนความคลาดเคลื่อนนั้นอาจเนื่องมาจากตัวข้อสอบ การควบคุมการสอบ การให้คะแนน และการเดาเป็นต้น นั่นคือนักเรียนอาจได้คะแนนโดยที่ไม่มีความรู้ในเรื่องนั้นเลยก็ได้ หรือนักเรียนอาจไม่ได้คะแนนทั้ง ๆ ที่มีความรู้ ความคลาดเคลื่อนในการวัดมี 2 ชนิด คือ ความคลาดเคลื่อนอย่างมีระบบ (systematic error) ความคลาดเคลื่อนชนิดนี้จะไม่มีผลกระทบกระเทือนต่อการวัด ความคลาดเคลื่อนชนิดนี้จะเป็นไปในทิศทางเดียว อาจเป็นลบหรือเป็นทางบวกทางใดทางหนึ่งเท่านั้น เช่น แถบวัดที่ใช้มานานอาจยืด เมื่อนำแถบวัดนี้ไปวัดสิ่งใดก็ตาม จะทำให้ความยาวของสิ่งนั้นสั้นกว่าความยาวจริง ๆ เช่น การที่แถบวัดบอกความยาวเกินไป 5 ซม. ทุก ๆ หนึ่งเมตร ก็จะไม่ทำให้ความเชื่อมั่นเปลี่ยนแปลงไป

แต่อย่างใด ส่วนความคลาดเคลื่อนอย่างสุ่ม (random error) นั้นเป็นไปได้หลายทิศทาง เช่นบางครั้งอาจเพิ่มขึ้น บางครั้งอาจลดลง ไม่แน่นอน ตัวอย่างเช่น คะแนนที่ได้จากการสอบซึ่งบางครั้งอาจได้คะแนนมากหรือน้อยกว่าคะแนนจริง

อย่างไรก็ตาม คะแนนความคลาดเคลื่อนนี้ยังไม่สามารถหาวิธีการตรวจสอบได้ ดังนั้นจึงอาจสรุปได้แต่เพียงว่า คะแนนที่ได้จากการสอบหรือการสังเกตจะไม่เท่ากับคะแนนจริง และสามารถเขียนสมการของคะแนนในการสอบได้ดังนี้

$$x = t + e \qquad \text{.....(6-1)}$$

เมื่อ x แทน คะแนนผลการสอบ (observed score)

t แทน คะแนนจริง (true score)

e แทน คะแนนความคลาดเคลื่อน (error score)

ความคลาดเคลื่อนอย่างสุ่มนี้ เป็นไปได้ทั้งทางบวกและทางลบ ดังนั้นในการศึกษาเกี่ยวกับความคลาดเคลื่อน จำเป็นต้องมีข้อตกลงเบื้องต้น ดังนี้

1. ค่าเฉลี่ยของคะแนนความคลาดเคลื่อนเท่ากับ 0

$$M_e = 0$$

2. ความสัมพันธ์ระหว่างคะแนนจริงกับคะแนนความคลาดเคลื่อนเท่ากับ 0

$$r_{te} = 0$$

3. ความสัมพันธ์ของคะแนนความคลาดเคลื่อนของแบบทดสอบ 2 ฉบับ ที่ขนานกัน เท่ากับ 0

$$r_{e_1e_2} = 0$$

จากข้อสมมุติดังกล่าว อาจเขียนสมการเกี่ยวกับความแปรปรวนได้ดังนี้

$$S_x^2 = S_t^2 + S_e^2 + 2r_{te}\,S_tS_e$$

แต่เนื่องจาก $r_{te} = 0$ จะได้ว่า

$$S_x^2 = S_t^2 + S_e^2 \qquad \text{.....(6-2)}$$

ความหมายของความเชื่อมั่น

ความเชื่อมั่น หมายถึงการวัดที่ให้ผลคงที่แน่นอนสม่ำเสมอ ไม่เปลี่ยนไปเปลี่ยนมา การวัดครั้งแรกเป็นอย่างไร เมื่อวัดซ้ำอีกกี่ครั้งก็ตาม ผลการวัดยังคงเหมือนเดิม โดยปกติจะนิยามความ

เชื่อมั่นว่าเป็นอัตราส่วนระหว่างความแปรปรวนของคะแนนจริง และความแปรปรวนของคะแนนที่ได้จากการสอบ เขียนเป็น สมการ ได้ดังนี้

$$r_{tt} = \frac{S_t^2}{S_x^2} \qquad \text{.....(6 - 3)}$$

เนื่องจาก $S^2_t = S^2_x - S^2_e$

ดังนั้น $r_{tt} = 1 - \frac{S^2_e}{S^2_x}$ .....(6 - 4)

จากสมการ (6 - 4) สามารถสรุปได้ดังนี้

1. r_{tt} จะมีค่าเป็นศูนย์ เมื่อ $S_e^2 = S_x^2$
2. r_{tt} จะมีค่าเป็น 1.00 เมื่อ $S_e^2 = 0$
3. ถ้าคะแนนของแบบทดสอบมีความคงเส้นคงวา เป็นที่เชื่อถือได้แล้ว S_e^2 จะมีค่าน้อยกว่า S_x^2 เสมอ
4. ความเชื่อมั่นของแบบทดสอบ (r_{tt}) ไม่เคยมีค่าติดลบ ทั้งนี้เพราะ S_e^2 ไม่เคยมีค่ามากกว่า S_x^2

การคำนวณหาค่าความเชื่อมั่น

วิธีการหาค่าความเชื่อมั่นของเครื่องมือในการวัด อาจทำได้หลายวิธี ดังนี้

1. **วิธีสอบซ้ำ** (test - retest method) เป็นการหาความเชื่อมั่นของเครื่องมือวัด โดยการนำเอาเครื่องมือนั้นไปสอบวัดกับนักเรียนกลุ่มใดกลุ่มหนึ่ง 2 ครั้ง ในเวลาที่ต่างกัน โดยเว้นระยะเวลาในการสอบทั้ง 2 ครั้งให้ห่างกันพอสมควร จุดมุ่งหมายก็เพื่อจะตรวจสอบดูว่าคะแนนหรือผลการวัดทั้ง 2 ครั้งนั้น สอดคล้องหรือสัมพันธ์กันเพียงใด แล้วนำคะแนนที่ได้จากการสอบทั้ง 2 ครั้ง มาคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ โดยใช้วิธี Pearson Product Moment ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้นี้ก็คือ ค่าความเชื่อมั่นของเครื่องมือวัดฉบับนั้นนั่นเอง ซึ่งสามารถคำนวณได้จากสูตร

$$r_{xy} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[N\Sigma X^2 - (\Sigma X)^2][N\Sigma Y^2 - (\Sigma Y)^2]}} \qquad \text{.....(6-5)}$$

ตัวอย่าง การหาค่าความเชื่อมั่นของแบบทดสอบฉบับหนึ่ง ซึ่งใช้ทดสอบกับนักเรียนกลุ่มเดียวกัน 2 ครั้ง ปรากฏผลของคะแนนดังนี้

คนที่	สอบครั้งที่ 1 X	สอบครั้งที่ 2 Y	X^2	Y^2	XY
1	1	2	1	4	2
2	2	3	4	9	6
3	2	2	4	4	4
4	3	5	9	25	15
5	4	5	16	25	20
6	5	7	25	49	35
7	6	5	36	25	30
8	8	6	64	36	48
9	9	7	81	49	63
10	10	8	100	64	80
N = 10	$\Sigma X = 50$	$\Sigma Y = 50$	$\Sigma X^2 = 340$	$\Sigma Y^2 = 290$	$\Sigma XY = 303$

แทนค่าจากสูตร

$$r_{xy} = \frac{10\,(303) - (50)\,(50)}{\sqrt{[10(340) - (50)^2]\,[10(290) - (50)^2]}}$$

$$= \frac{530}{\sqrt{(900)\,(400)}}$$

$$= \frac{530}{600} = 0.883$$

ดังนั้น แบบทดสอบฉบับนี้มีค่าความเชื่อมั่นเท่ากับ 0.88

ปัญหาที่สำคัญของการหาความเชื่อมั่นโดยวิธีนี้ คือ

1. ช่วงเวลาในการสอบทั้ง 2 ครั้ง ถ้าระยะเวลาระหว่างการสอบนานก็จะทำให้ค่า r ลดลง ค่า r ที่ลดลงนี้อาจเนื่องจากเกิดการเปลี่ยนแปลงในตัวผู้สอบ เช่น ผู้สอบอาจไปศึกษาค้นคว้าเพิ่ม

เติม ซึ่งจะทำให้คะแนนของนักเรียนแต่ละคนเปลี่ยนไป และอาจเปลี่ยนในระดับที่ต่างกัน มีผลทำให้ค่าความเชื่อมั่นของข้อสอบต่ำ และหากเว้นระยะเวลาระหว่างการสอบครั้งแรกกับครั้งหลังสั้นเกินไป จะทำให้เกิด Test wise คือนักเรียนหลายคนในชั้นจะยังจดจำคำตอบบางข้อได้ ทำให้การสอบทั้ง 2 ครั้ง ไม่เป็นอิสระต่อกัน นั่นคือครั้งแรกนักเรียนตอบว่าอย่างไร ครั้งหลังก็จะตอบอย่างนั้น ทำให้ค่าสหสัมพันธ์ที่หาได้ผิดไปจากความจริง และค่าความเชื่อมั่นก็ผิดไปจากความจริงด้วย

2. ความเปลี่ยนแปลงในตัวนักเรียนด้านต่าง ๆ เช่น อารมณ์ ความรู้สึก สภาพทางร่างกาย ในการสอบทั้ง 2 ครั้ง แตกต่างกัน เช่น อาจเกิดความวิตกกังวล เกิดการเจ็บป่วย ซึ่งสิ่งต่าง ๆ เหล่านี้นับว่ามีอิทธิพลต่อผลการสอบของนักเรียน

3. สภาพการสอบทั้ง 2 ครั้ง เช่น อากาศ แสง เสียง หรือเรื่องที่อาจเกิดขึ้นโดยนึกไม่ถึงขณะสอบ ซึ่งจะส่งผลต่อการทำข้อสอบของนักเรียน และเป็นสิ่งที่ควบคุมได้ยาก

4. การขาดหายของกลุ่มตัวอย่างในการสอบทั้ง 2 ครั้ง

จะเห็นได้ว่าการหาความเชื่อมั่นโดยวิธีนี้ จำเป็นต้องพิจารณาอย่างรอบคอบ โดยเฉพาะระยะเวลาในการสอบทั้ง 2 ครั้งนั้น ควรเว้นระยะเวลาในการสอบให้ห่างกันพอสมควร

2. วิธีใช้แบบทดสอบคู่ขนาน (Paralell test method) วิธีนี้โดยหลักการแล้ว เป็นวิธีเดียวกับการสอบซ้ำ แต่พยายามป้องกันการจำข้อสอบได้ จึงได้ใช้แบบทดสอบ 2 ฉบับ ที่คู่ขนานกันไปทดสอบแก่นักเรียนกลุ่มหนึ่ง โดยที่ข้อสอบคู่ขนานทั้ง 2 ฉบับนี้ต้องเป็นอิสระต่อกัน คือไม่ใช่ทำถูกฉบับหนึ่ง แล้วต้องไปทำถูกอีกฉบับหนึ่งด้วย ตามทฤษฎีแล้ว แบบทดสอบ 2 ฉบับ จะคู่ขนานกันก็ต่อเมื่อมีจำนวนข้อเท่ากัน ถามเนื้อหาเดียวกัน มีความยากง่ายเท่ากัน นั่นคือคะแนนเฉลี่ยและความแปรปรวนของแบบทดสอบทั้ง 2 ฉบับต้องเท่ากัน เมื่อนำแบบทดสอบคู่ขนานทั้ง 2 ฉบับไปสอบนักเรียนแล้ว ก็นำเอาคะแนนที่ได้จากการสอบทั้ง 2 ครั้ง มาหาค่าสัมประสิทธิ์สหสัมพันธ์แบบ Pearson Product Moment เช่นเดียวกับการหาความเชื่อมั่นโดยวิธีสอบซ้ำ

ปัญหาที่สำคัญของการหาความเชื่อมั่นโดยวิธีนี้ก็คือ เป็นการยากมากที่จะสร้างแบบทดสอบทั้ง 2 ฉบับ ให้คู่ขนานกัน เมื่อไม่สามารถสร้างแบบทดสอบทั้ง 2 ฉบับให้ขนานกันได้จริง ๆ ก็ย่อมทำให้เกิดความคลาดเคลื่อนในการหาความเชื่อมั่นได้

3. วิธีหาความคงที่ภายใน (Internal consistency method) เนื่องจากการหาความเชื่อมั่น 2 วิธีที่กล่าวมานั้น มีปัญหาบางประการ เช่น ต้องใช้การสอบถึง 2 ครั้ง ทำให้เกิดความไม่สะดวก ดังนั้นเพื่อขจัดปัญหานี้ จึงหลีกเลี่ยงด้วยการสอบเพียงครั้งเดียว แล้วนำผลที่ได้จากการสอบมาคำนวณหาค่าความเชื่อมั่น แต่ทั้งนี้ต้องเป็นไปตามเงื่อนไขของแต่ละวิธี วิธีการคำนวณหาค่าความ

คงที่ภายใน แบ่งออกเป็นดังนี้

3.1 วิธีแบ่งครึ่งข้อสอบ (Split - half method) การหาความเชื่อมั่นโดยวิธีนี้ คล้ายกับการหาความเชื่อมั่นโดยใช้แบบทดสอบคู่ขนาน คือแทนที่จะนำแบบทดสอบไปสอบกับนักเรียน 2 ครั้ง ก็เปลี่ยนเป็นนำแบบทดสอบนั้นไปสอบกับนักเรียนเพียงครั้งเดียว แล้วแบ่งข้อสอบออกเป็น 2 ชุดย่อย โดยพยายามทำให้แบบทดสอบทั้ง 2 ชุดนั้น มีลักษณะคล้ายคลึงกันมากที่สุดเท่าที่จะทำได้ ทั้งนี้อาจแบ่งข้อสอบออกเป็นข้อคู่ - ข้อคี่ หรือแบ่งเป็น 2 ตอน คือ ตอนแรก - ตอนหลัง เมื่อได้คะแนน 2 ชุดแล้ว ก็นำมาหาค่าสัมประสิทธิ์สหสัมพันธ์แบบ Pearson Product Moment ค่าสัมประสิทธิ์สหสัมพันธ์ที่ได้นี้จะเป็นค่าความเชื่อมั่นของแบบทดสอบเพียงครึ่งฉบับ ฉะนั้นถ้าต้องการหาค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับ ก็ต้องนำไปคำนวณขยายโดยใช้สูตรของสเปียร์-แมน - บราวน์ (Spearman - Brown) ดังนี้

$$r_{tt} = \frac{nr_{xy}}{1 + (n-1)\, r_{xy}} \qquad \text{.....(6-6)}$$

เมื่อ r_{tt} แทน ความเชื่อมั่นของแบบทดสอบที่มีความยาว n เท่า ของแบบทดสอบที่ให้ค่า r_{xy}

ถ้าแบ่งข้อสอบฉบับหนึ่งออกเป็น 2 ชุดย่อย และต้องการหาความเชื่อมั่นของข้อสอบเต็มฉบับ ก็จะต้องขยายความเชื่อมั่นเป็น 2 เท่าของฉบับเดิม สูตรขยายความเชื่อมั่นจากครึ่งฉบับเป็นเต็มฉบับ จะเป็นดังนี้

$$r_{tt} = \frac{2r_{xy}}{1 + (2-1)\, r_{xy}}$$

$$= \frac{2r_{xy}}{1 + r_{xy}} \qquad \text{.....(6-7)}$$

ตัวอย่าง แสดงการคำนวณหาความเชื่อมั่นโดยวิธีแบ่งครึ่ง

นักเรียน	ข้อคู่ a	ข้อคี่ b	a^2	b^2	ab
1	1	0	1	0	0
2	3	3	9	9	9
3	2	0	4	0	0
4	4	3	16	9	12
5	2	2	4	4	4
6	1	1	1	1	1
7	3	2	9	4	6
8	2	1	4	1	2
9	4	2	16	4	8
10	2	2	4	4	4
11	3	4	9	16	12
12	3	2	9	4	6
13	3	3	9	9	9
14	2	2	4	4	4
15	2	1	4	1	2
16	3	1	9	1	3
17	1	2	1	4	2
18	4	4	16	16	16
19	2	3	4	9	6
20	3	2	9	4	6
Σ	50	40	142	104	112

$$r_{ab} = \frac{112 - (50 \times 40)/20}{\sqrt{(142 - 50^2/20)\,(104 - 40^2/20)}} = 0.594$$

$$r_{tt} = \frac{2 \times 0.594}{1 + 0.594} = \frac{1.188}{1.594} = 0.745$$

ปัญหาที่สำคัญของการหาความเชื่อมั่นโดยวิธีนี้คือ ข้อสอบแต่ละข้อจะต้องมีความยากง่ายใกล้เคียงกัน จึงจะสามารถแบ่งเป็นข้อคู่ - ข้อคี่ หรือ ตอนแรก - ตอนหลัง ได้ การแบ่งข้อสอบเป็น ตอนแรก - ตอนหลัง จะมีปัญหาเรื่องเวลา เช่น นักเรียนทำข้อสอบไม่ทัน จะทำให้ผลที่ได้จากการหาความเชื่อมั่นนั้นไม่แน่นอน การหาความเชื่อมั่นโดยวิธีแบ่งครึ่งนี้ ไม่เหมาะกับข้อสอบที่เป็น Speed test เพราะถ้าให้เวลาน้อย นักเรียนทุกคนก็จะไม่ได้ทำข้อหลัง ๆ เหมือนกันหมด ทำให้ค่าความเชื่อมั่นที่ได้ผิดไปจากความจริง ในการหาค่าความเชื่อมั่นของวิธีแบ่งครึ่งอาจใช้สูตรของ Rulon ดังนี้

$$r_{tt} = 1 - S_d^2/S_x^2 \qquad \text{.....(6-8)}$$

เมื่อ r_{tt} แทน ค่าความเชื่อมั่นทั้งฉบับ

S_d^2 แทน ความแปรปรวนของคะแนนความแตกต่างของนักเรียนแต่ละคน

S_x^2 แทน ความแปรปรวนของข้อสอบทั้งฉบับ

สูตรของ Rulon นี้ Guttman ได้แปลงให้อยู่ในรูปที่ง่ายต่อการคำนวณ ดังนี้

$$r_{tt} = 2[1 - (S_a^2 + S_b^2)/S_x^2] \qquad \text{.....(6-9)}$$

เมื่อ a, b แทนข้อสอบครึ่งแรกและครึ่งหลัง ตามลำดับ

จากสมการ (6-8) และ (6-9) จะได้ว่า

$$S_d^2 = S_a^2 + S_b^2 - 2 r_{ab} S_a S_b$$

และ $S_x^2 = S_a^2 + S_b^2 + 2 r_{ab} S_a S_b$

ดังนั้น สมการ (6-8) สามารถเขียนใหม่ได้ดังนี้

$$r_{tt} = 4 r_{ab} S_a S_b / S_x^2 \qquad \text{.....(6-10)}$$

ตัวอย่าง แสดงการคำนวณหาความเชื่อมั่น โดยใช้สูตรของ Rulon

นักเรียน	ข้อคู่ a	ข้อคี่ b	d_j	d_j^2	x_j	x_j^2
1	1	0	1	1	1	1
2	3	3	0	0	6	36
3	2	0	2	4	2	4
4	4	3	1	1	7	49
5	2	2	0	0	4	16
6	1	1	0	0	2	4
7	3	2	1	1	5	25
8	2	1	1	1	3	9
9	4	2	2	4	6	36
10	2	2	0	0	4	16
11	3	4	−1	1	7	49
12	3	2	1	1	5	25
13	3	3	0	0	6	36
14	2	2	0	0	4	16
15	2	1	1	1	3	9
16	3	1	2	4	4	16
17	1	2	−1	1	3	9
18	4	4	0	0	8	64
19	2	3	−1	1	5	25
20	3	2	1	1	5	25
Σ	50	40	10	22	90	470

$$S_d^2 = \frac{22}{20} - \left(\frac{10}{20}\right)^2 = 1.10 - 0.25 = 0.85$$

$$S_x^2 = \frac{470}{20} - \left(\frac{90}{20}\right)^2 = 23.50 - 20.25 = 3.25$$

$$r_{tt} = 1 - \frac{0.85}{3.25} = 1 - 0.262 = 0.738$$

3.2 วิธีคูเดอร์ ริชาร์ดสัน (Kuder-Richardson method) เป็นการดูความคงที่ภายในของแบบทดสอบ นั่นคือจะใช้ข้อสอบฉบับเดียวทดสอบกับนักเรียนเพียงครั้งเดียว แล้วนำผลที่ได้มาคำนวณหาค่าความคงที่ภายใน

สมการความแปรปรวนของข้อสอบ ซึ่งประกอบด้วยข้อสอบ n ข้อ

$$S_x^2 = \Sigma S_i^2 + 2\Sigma r_{ik} S_i S_k,\ i < k$$

จากสมการนี้ เทอมสุดท้ายจะมีความแปรปรวนร่วม $n(n-1)$ เทอม ถ้าข้อสอบแต่ละข้อมีโอกาสที่จะตอบถูกเท่ากัน และสหสัมพันธ์ระหว่างข้อสอบแต่ละข้อเท่ากัน สามารถแทน S_iS_k ด้วย S_i^2 และผลบวกของเทอมที่เป็นความแปรปรวนร่วม สามารถเขียนแทนด้วย $n(n-1)\ \bar{r}_{ik}\ \overline{S_i^2}$

$$S_x^2 = \Sigma S_i^2 + n(n-1)\bar{r}_{ik}\overline{S_i^2} \qquad \text{.....(6-11)}$$

เนื่องจาก $n\overline{S_i^2} = \Sigma S_i^2$

ดังนั้น $$\bar{r}_{ik} = \frac{S_x^2 - \Sigma S_i^2}{(n-1)\Sigma S_i^2} \qquad \text{.....(6-12)}$$

สมการ 6-12 เป็นค่าความเชื่อมั่นของข้อสอบข้อเดียว ดังนั้นถ้าต้องการหาค่าความเชื่อมั่นของข้อสอบที่ประกอบด้วย n ข้อ ก็ต้องขยายโดยใช้สูตรของสเปียร์แมน-บราวน์ ดังนี้

$$r_n = \frac{n\bar{r}_{ik}}{1 + (n-1)\bar{r}_{ik}}$$

$$= n\frac{S_x^2 - \Sigma S_i^2}{(n-1)\Sigma S_i^2} \cdot \frac{1}{1 + (n-1)[(S_x^2 - \Sigma S_i^2)/(n-1)\Sigma S_i^2]}$$

$$= \frac{n}{n-1}\cdot\frac{S_x^2 - \Sigma S_i^2}{\Sigma S_i^2}\cdot\frac{1}{1 + (S_x^2/\Sigma S_i^2) - 1}$$

$$= \frac{n}{n-1}\cdot\frac{S_x^2 - \Sigma S_i^2}{\Sigma S_i^2}\cdot\frac{\Sigma S_i^2}{S_x^2}$$

$$= \frac{n}{n-1}\cdot\frac{S_x^2 - \Sigma S_i^2}{S_x^2} \qquad \text{.....(6-13)}$$

สมการ 6-13 เป็นการหาความเชื่อมั่นของข้อสอบที่มี n ข้อ เรียกการหาความเชื่อมั่นตามวิธีการนี้ว่า Coefficient Alpha (α)

สมการ 6-13 เป็นสูตรทั่วไปในการคำนวณหาค่าความเชื่อมั่นของแบบทดสอบ นั่นคือ ไม่ว่าการให้คะแนนจะเป็นแบบใดก็ตาม สามารถคำนวณหาความเชื่อมั่นของแบบทดสอบฉบับนั้นด้วย Coefficient Alpha ได้เสมอ

ถ้าเป็นข้อสอบแบบ 0, 1 คือตอบผิดให้ 0 คะแนน ตอบถูกให้ 1 คะแนน จะได้ว่า

$$S_i^2 = p_i q_i$$

ดังนั้น
$$r_{tt} = \frac{n}{n-1} \frac{S_x^2 - \Sigma pq}{S_x^2} \qquad \text{.....(6-14)}$$

หรือ
$$r_{tt} = \frac{n}{n-1} \left(1 - \frac{\Sigma pq}{S_x^2}\right) \qquad \text{.....(6-15)}$$

สมการ 6-14 และ 6-15 นี้ ก็คือสูตรการหาความเชื่อมั่นตามวิธีของคูเดอร์-ริชาร์ดสัน สูตรที่ 20 นั่นเอง สูตรนี้มีข้อจำกัดว่าการให้คะแนนจะต้องเป็นแบบ 0,1

ถ้าให้ข้อสอบแต่ละข้อมีระดับความยากเท่ากัน หรือค่า p คงที่ จะได้ว่า

$$r_{tt} = \frac{n}{n-1} \left(1 - \frac{n\bar{p}\bar{q}}{S_x^2}\right) \qquad \text{.....(6-16)}$$

$$= \frac{n}{n-1} \left(1 - \frac{\frac{n\Sigma p}{n} \cdot \frac{\Sigma(1-p)}{n}}{S_x^2}\right)$$

$$= \frac{n}{n-1} \left(1 - \frac{\Sigma p \cdot \frac{\Sigma 1 - \Sigma p}{n}}{S_x^2}\right)$$

$$= \frac{n}{n-1} \left(1 - \frac{\bar{X}\frac{(n-\bar{X})}{n}}{S_x^2}\right)$$

$$= \frac{n}{n-1} \left(1 - \frac{\bar{X}(n-\bar{X})}{nS_x^2}\right) \qquad \text{.....(6-17)}$$

สมการ 6-16 และ 6-17 เป็นการหาความเชื่อมั่นตามวิธีการของคูเดอร์-ริชาร์ดสัน สูตรที่ 21 สูตรนี้นับว่ามีความสะดวกในการคำนวณมาก เพราะเพียงแต่หาค่าเฉลี่ยและความแปรปรวนของแบบทดสอบก็สามารถคำนวณค่าความเชื่อมั่นออกมาได้ โดยนำค่าที่คำนวณได้ไปแทนในสมการ 6-17 สูตรนี้มีข้อจำกัดเพิ่มเติมจากสูตรที่ 20 ก็คือ ข้อสอบแต่ละข้อต้อง

มีความยากง่ายเท่ากัน (มีค่า p เท่ากันทุกข้อ) จะเห็นได้ว่าในทางปฏิบัติแล้วสูตรนี้ไม่มีโอกาสได้ใช้เลยเนื่องจากเป็นการยากมากที่จะสร้างข้อสอบทุกข้อให้มีความยากง่ายเท่ากัน

สูตรที่ 21 นี้เป็น under estimate ของสูตรที่ 20 ดังนั้นถ้าคำถามแต่ละข้อมีความยากง่ายต่างกันแล้ว ค่าความเชื่อมั่นที่คำนวณโดยใช้สูตรที่ 21 นี้ จะให้ค่าต่ำกว่าที่คำนวณโดยใช้สูตรที่ 20 เสมอ แต่ถ้าข้อสอบทุกข้อมีความยากง่ายเท่ากัน ไม่ว่าจะคำนวณโดยใช้สูตรที่ 20 หรือสูตรที่ 21 ก็จะให้ค่าความเชื่อมั่นออกมาเท่ากันเสมอ

ตัวอย่าง แสดงการหาค่าความเชื่อมั่นโดยใช้วิธีของคูเดอร์-ริชาร์ดสัน

นักเรียน	ข้อที่						คะแนน
คนที่	1	2	3	4	5	6	รวม
1	1	1	1	0	1	1	5
2	0	1	1	0	0	1	3
3	1	0	1	1	0	0	3
4	1	1	0	1	1	1	5
5	1	1	1	1	0	0	4
6	0	1	0	0	1	0	2
7	1	1	0	1	1	0	4
8	1	1	0	1	0	1	4
9	1	1	1	1	1	1	6
10	1	1	1	1	1	1	6
p	.8	.9	.6	.7	.6	.6	—
q	.2	.1	.4	.3	.4	.4	—
pq	.16	.09	.24	.21	.24	.24	—

$$\Sigma pq = 1.18$$

$$\Sigma X = 42 \qquad N = 10 \qquad \overline{X} = 4.2 \qquad n = 6$$

$$\Sigma X^2 = 192$$

$$S_x^2 = \frac{192}{10} - (4.2)^2 = 19.20 - 17.64 = 1.56$$

$$\text{KR 20} : r_{tt} = \frac{n}{n-1}\left[1 - \frac{\Sigma pq}{S_x^2}\right]$$

$$= \frac{6}{5}\left(1 - \frac{1.18}{1.56}\right)$$

$$= 1.2\left(\frac{0.38}{1.56}\right)$$

$$= 0.29$$

$$\text{KR 21} : r_{tt} = \frac{n}{n-1}\left[1 - \frac{\overline{X}(n-\overline{X})}{nS_x^2}\right]$$

$S_x^2 = 1.56 \qquad \overline{X} = 4.2 \qquad n = 6$

$$r_{tt} = \frac{6}{5}\left[1 - \frac{4.2\,(6-4.2)}{6\,(1.56)}\right]$$

$$= \frac{6}{5}\left[1 - \frac{(0.7)\,(1.8)}{1.56}\right]$$

$$= \frac{6}{5}\,\frac{(1.56 - 1.26)}{1.56}$$

$$= \frac{6}{5}\left(\frac{0.30}{1.56}\right)$$

$$= 0.23$$

การหาความเชื่อมั่นของแบบทดสอบอิงเกณฑ์ (Criterion Referenced Test)

การวัดผลแบบอิงเกณฑ์นั้นมีจุดประสงค์เพื่อตัดสินการเรียนรู้ของผู้เรียนว่าผ่านเกณฑ์ที่กำหนดไว้หรือไม่ ในการตัดสินนั้นมีสิ่งที่ต้องคำนึงถึงอยู่ 2 ประการคือ

1. A false - negative decision เป็นการตัดสินว่าผู้เรียนไม่ผ่านเกณฑ์ทั้ง ๆ ที่ผู้เรียนมีความสามารถ

2. A false - positive decision เป็นการตัดสินว่าผู้เรียนผ่านเกณฑ์ทั้ง ๆ ที่จริง ๆ แล้วผู้เรียนมีความสามารถไม่ถึงเกณฑ์

จะเห็นได้ว่าการหาความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์มุ่งไปในส่วนที่เกี่ยวข้องกับความถูกต้องในการตัดสินใจ (Consistency of decision making) เกี่ยวกับสภาวะการเรียนรู้ของผู้เรียน ส่วนการวัดแบบอิงกลุ่มนั้นสนใจเกี่ยวกับความถูกต้องของคะแนน กล่าวคือ คะแนนผลการสอบควรจำแนกความแตกต่างระหว่างบุคคลในกลุ่มผู้สอบ ดังนั้นการหาความเชื่อมั่นของแบบทดสอบโดยวิธีอิงกลุ่มจึงไม่เหมาะที่จะนำมาใช้หาความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์ (CRT) ทั้งนี้เพราะคะแนนที่ได้จากแบบทดสอบแบบอิงเกณฑ์อาจมีความแปรปรวนเพียงเล็กน้อย ดังนั้นหากนำคะแนนจากแบบทดสอบแบบอิงเกณฑ์มาหาความเชื่อมั่นโดยวิธีอิงกลุ่มแล้วก็จะทำให้ความเชื่อมั่นของแบบทดสอบต่ำ

นักวัดผลได้พยายามคิดหาวิธีการหาความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์ไว้หลายวิธีด้วยกัน แต่ละวิธียังมีข้อจำกัดและข้อโต้แย้งมาก และยังหาข้อสรุปไม่ได้ว่าวิธีใดเหมาะสมที่สุด ในที่นี้จึงขอยกตัวอย่างวิธีการหาความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์เพียงบางวิธีเท่านั้น

1. Livingston (1972) ได้เสนอแนะวิธีหาความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์โดยทำการปรับปรุงวิธีการหาความเชื่อมั่นของแบบทดสอบแบบอิงกลุ่มดังนี้

$$K^2(X, T_x) = \frac{\rho^2(X, T_x)\,\sigma_x^2 + (\mu_x - C_x)^2}{\sigma_x^2 + (\mu_x - C_x)^2} \qquad \text{.....(6-18)}$$

เมื่อ $K^2(X, T_x)$ แทนความเชื่อมั่นของแบบทดสอบอิงเกณฑ์

$\rho^2(X, T_x)$ แทนความเชื่อมั่นของแบบทดสอบอิงกลุ่ม

σ_x^2 แทน ความแปรปรวนของคะแนน

μ_x แทน ค่าเฉลี่ยของคะแนนจากแบบทดสอบ

C_x แทน เกณฑ์มาตรฐานที่กำหนดไว้สำหรับตัดสินว่าใครมีความรู้มากพอที่จะให้ผ่านหรือให้สอบได้

จากสูตร (6-18) สามารถสรุปได้ดังนี้

1. ถ้า $\rho^2 (X, T_X)$ เพิ่ม ค่า $K^2 (X, T_X)$ จะเพิ่มขึ้นด้วย

2. ถ้า $\rho^2 (X, T_X) = 1.00$ แล้ว $K^2 (X, T_X)$ จะมีค่าเท่ากับ 1.00 ด้วย

3. ถ้า $\mu_x = C_X$ แล้วจะได้ว่า $K^2 (X, T_X) = \rho^2 (X, T_X)$

จะเห็นได้ว่าค่าความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์ที่คำนวณโดยวิธีของ Livingston จะมีค่ามากกว่าหรือเท่ากับค่าความเชื่อมั่นของข้อสอบฉบับเดียวกันที่คำนวณโดยใช้สูตรของการวัดแบบอิงกลุ่ม

2. Harris (1972) ได้เสนอวิธีการหาความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์โดยใช้วิธีการวิเคราะห์ความแปรปรวนแบบทางเดียว (One Way ANOVA) ซึ่งมีสูตรในการคำนวณดังนี้

$$\mu_c^2 = \frac{SS_B}{SS_B + SS_W} \qquad \text{.....(6-19)}$$

เมื่อ μ_c^2 แทนความเชื่อมั่นของข้อสอบแบบอิงเกณฑ์

SS_B แทน ผลบวกกำลังสองระหว่างกลุ่ม (Sum Square Between)

SS_W แทน ผลบวกกำลังสองภายในกลุ่ม (Sum Square Within)

3. Hambleton และคณะ (1974) ได้เสนอแนะวิธีการหาความเชื่อมั่นโดยใช้การสอบ 2 ครั้ง แล้วดูสัดส่วนของนักเรียนที่ถูกกำหนดอย่างคงเส้นคงวาในแต่ละสภาพ (ผ่าน-ไม่ผ่าน) ของการเรียนรู้ทั้ง 2 ครั้ง โดยใช้ค่าสัมประสิทธิ์แคปป้า (Kappa - K) ของ Cohen ดังนี้

$$K = (P_O - P_e)/(1 - P_e) \qquad \text{.....(6-20)}$$

เมื่อ K แทน ความเชื่อมั่นของแบบทดสอบแบบอิงเกณฑ์

P_O แทน สัดส่วนของความสอดคล้องที่สังเกตได้จากการสอบซ้ำ

P_e แทน สัดส่วนของความสอดคล้องที่คาดหวัง

สอบครั้งที่ 2 / สอบครั้งที่ 1	Master	Non-Master	รวม
Master	.66	.14	.80
Non-master	.04	.16	.20
รวม	.70	.30	

P_o = .66 + .16

= .82

P_e = (.80) (.70) + (.20) (.30)

= .62

K = (.82 - .62) / (1 - .62)

= .53

4. Carver ได้เสนอแนะวิธีหาความเชื่อมั่นของแบบทดสอบ โดยใช้แบบทดสอบอิงเกณฑ์ 2 ฉบับ ที่คู่ขนานกัน ทดสอบกับนักเรียนกลุ่มเดียวกัน แล้วแบ่งนักเรียนดังตารางข้างล่าง

		แบบทดสอบ B: ตก	แบบทดสอบ B: ได้
แบบทดสอบ A	ได้	b	a
	ตก	c	d

Reliability = (a + c)/N

N = a + b + c + d

ความคลาดเคลื่อนมาตรฐานของการวัด

ถ้าเครื่องมือในการวัดมีความเชื่อมั่นสูง (r_{tt} = 1.00) ความคลาดเคลื่อนในการวัดก็จะไม่มี แต่ในทางการศึกษาจะไม่มีเครื่องมือใดที่มีความเชื่อมั่นถึง 1.00 และเมื่อใดก็ตามถ้าความเชื่อมั่นของเครื่องมือในการวัดต่ำแล้ว ค่าความคลาดเคลื่อนก็จะสูง จากนิยามของความเชื่อมั่นจะเห็นว่าความเชื่อมั่นจะสูงขึ้น เมื่อความแปรปรวนของคะแนนความคลาดเคลื่อนลดลง โดยที่ความแปรปรวนของคะแนนที่ได้จากการสอบคงที่ หรือเมื่อความแปรปรวนของคะแนนความคลาดเคลื่อนคงที่ และความแปรปรวนของคะแนนที่ได้จากการสอบเพิ่มขึ้น ความเชื่อมั่นก็จะสูงขึ้นเช่นกัน

เนื่องจาก $r_{tt} = 1 - \frac{S_e^2}{S_x^2}$

$$S_e^2 = S_x^2 (1 - r_{tt})$$

หรือ $$S_e = S_x \sqrt{1 - r_{tt}} \qquad \text{.....(6-21)}$$

เมื่อ S_e แทน ความคลาดเคลื่อนมาตรฐานในการวัด

S_x แทน ความเบี่ยงเบนมาตรฐานในการวัด

r_{tt} แทน ความเชื่อมั่น

ตัวอย่าง แบบทดสอบฉบับหนึ่งมีความเชื่อมั่นเท่ากับ 0.84 ความเบี่ยงเบนมาตรฐานในการวัดเท่ากับ 10 จงหาความคลาดเคลื่อนมาตรฐานของการวัด

$$S_e = 10\sqrt{1 - 0.84}$$
$$= 10\sqrt{0.16}$$
$$= 10\ (0.4)$$
$$= 4.0$$

สมมุติว่าคะแนนที่ได้จากการสอบเป็น 40 คะแนน โอกาสที่จะได้คะแนนจริงอยู่ระหว่าง 40 ± 4 หรือ 36 - 44 คะแนน ประมาณ 68 เปอร์เซ็นต์ และโอกาสที่จะได้คะแนนจริงอยู่ระหว่าง 40 ± 1.96 (4.0) หรือระหว่าง 32 - 48 คะแนนประมาณ 95 เปอร์เซ็นต์

องค์ประกอบที่มีอิทธิพลต่อค่าความเชื่อมั่น

เป็นที่ทราบกันดีอยู่แล้วว่า ความเชื่อมั่นของข้อสอบไม่คงที่ อาจเพิ่มมากขึ้นหรือลดน้อยลงก็ได้ วิธีการต่าง ๆ ในการคำนวณหาความเชื่อมั่นก็เป็นสาเหตุหนึ่งที่ทำให้ค่าความเชื่อมั่นเปลี่ยนไป

นอกจากนี้แล้วยังมีองค์ประกอบอื่น ๆ อีก ซึ่งมีอิทธิพลต่อค่าความเชื่อมั่นของแบบทดสอบ ซึ่งพอสรุปได้ดังนี้

1. **ความยาวของแบบทดสอบ** (Test length) ความยาวของแบบทดสอบมีอิทธิพลต่อความเชื่อมั่นของแบบทดสอบ แบบทดสอบที่ยาวจะมีความเชื่อมั่นสูงกว่าแบบทดสอบที่สั้น ทั้งนี้พิจารณาได้จากสูตรของ Spearman - Brown ดังนี้

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

เมื่อ r_{tt} แทน ความเชื่อมั่นของแบบทดสอบใหม่ที่มีความยาวเป็น n เท่า ของข้อสอบเดิม

r_{xy} แทน ความเชื่อมั่นของข้อสอบเดิม

ตัวอย่าง ถ้าความเชื่อมั่นของแบบทดสอบฉบับหนึ่งซึ่งมีอยู่ 20 ข้อ มีค่าเท่ากับ .60 และถ้าต้องการเพิ่มข้อสอบเป็น 60 ข้อ แบบทดสอบใหม่จะมีค่าความเชื่อมั่นเป็นเท่าไร

$$r_{tt} = \frac{3\ (.60)}{1 + (3 - 1)\ (.60)}$$

$$= 0.818$$

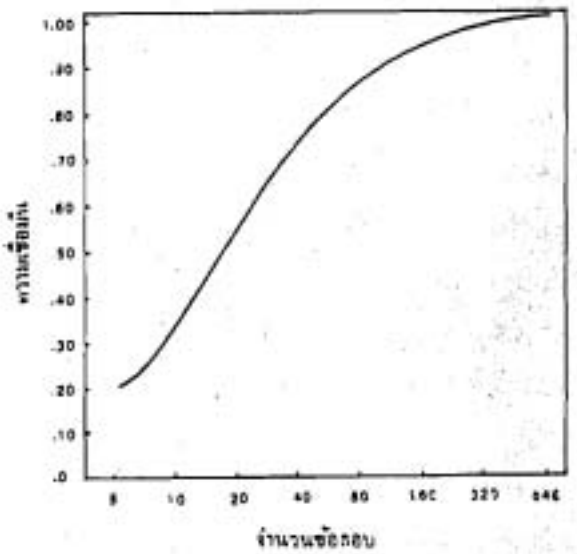
จากตัวอย่างนี้ จะเห็นว่าข้อสอบยิ่งยาว ความเชื่อมั่นของแบบทดสอบจะมีค่าสูงขึ้นและในทำนองเดียวกัน ถ้าลดจำนวนข้อในแบบทดสอบก็จะทำให้ค่าความเชื่อมั่นของแบบทดสอบนั้นลดลงด้วย

การเพิ่มค่าความเชื่อมั่นของแบบทดสอบโดยใช้สูตรนี้ มีข้อตกลง (Assumption) ว่าข้อสอบที่เพิ่มขึ้นนั้นจะต้องมีคุณสมบัติทางสถิติเหมือนกับข้อสอบเดิม คือข้อสอบที่เพิ่มเข้าไปจะต้องมีค่าความยากง่ายเหมือนเดิม หรือกล่าวง่าย ๆ ว่า ข้อสอบที่เพิ่มเข้าไปนั้นจะต้องขนานกับข้อสอบเดิม แต่ในทางปฏิบัติเมื่อเพิ่มข้อสอบแล้วนำมาหาความเชื่อมั่นใหม่ก็จะไม่ได้เท่ากับความเชื่อมั่นที่เพิ่มขึ้นตามสูตรของ Spearman - Brown ทั้งนี้เพราะข้อสอบที่เพิ่มเข้าไปอาจไม่คู่ขนานกับข้อสอบเดิมในแง่ของความยาก อำนาจจำแนก และความแปรปรวนของข้อสอบแต่ละข้อ นอกจากนี้สภาพการสอบก็เปลี่ยนไป คือเมื่อเพิ่มจำนวนข้อสอบมากขึ้นนักเรียนจะต้องใช้เวลาทำนานขึ้น จะทำให้นักเรียนเกิดความเหนื่อยล้า และเบื่อหน่ายในการทำข้อสอบ ซึ่งนับว่าเป็นสาเหตุสำคัญที่ทำให้สูตรของ Spearman - Brown มีความคลาดเคลื่อน

อย่างไรก็ตาม ค่าความเชื่อมั่นจากสูตรของ Spearman - Brown ก็มิได้แปรผันโดยตรงกับความยาวที่เปลี่ยนไป ดังตัวอย่างแสดงความสัมพันธ์ระหว่างความเชื่อมั่นกับจำนวนข้อในแบบทดสอบ ดังนี้

จำนวนข้อ	ความเชื่อมั่น
5	0.20
10	0.33
20	0.50
40	0.67
80	0.80
160	0.89
320	0.94
640	0.97
∞	1.00

จากตารางจะเห็นว่า เมื่อข้อสอบมีความเชื่อมั่นสูงแล้ว การเพิ่มความยาวของข้อสอบจะทำให้ความเชื่อมั่นของข้อสอบเพิ่มขึ้นเพียงเล็กน้อยเท่านั้น เช่น เพิ่มข้อสอบ 80 ข้อ เป็น 160 ข้อ จะทำให้ความเชื่อมั่นเพิ่มขึ้นเพียง .09 เท่านั้น ดังเห็นได้จากกราฟแสดงความสัมพันธ์ระหว่างจำนวนข้อในแบบทดสอบกับความเชื่อมั่นของแบบทดสอบ ดังนี้



ภาพแสดงความสัมพันธ์ระหว่างจำนวนข้อในแบบทดสอบกับความเชื่อมั่นของแบบทดสอบ

อนึ่ง ในการเพิ่มจำนวนข้อสอบเพื่อให้แบบทดสอบมีความเชื่อมั่นสูงนั้นย่อมจะต้องมีขอบเขตจำกัด เช่น ช่วงเวลาที่กำหนดไว้สำหรับการสอบ และยังจะต้องคำนึงถึงประสิทธิภาพด้วย กล่าวคือถ้าใช้ข้อสอบจำนวนน้อยข้อแล้วให้ผลไม่แตกต่างจากข้อสอบที่มีจำนวนมากข้อ ก็ให้เลือกใช้ข้อสอบที่มีจำนวนน้อยข้อ เพราะเป็นการประหยัดทางเศรษฐกิจแต่ทั้งนี้จะต้องน้อยพอที่จะมีประสิทธิภาพในทางปฏิบัติ

2. **การกระจายของคะแนน** จากนิยามความเชื่อมั่นที่ว่า $r_{tt} = 1 - \frac{S_e^2}{S_x^2}$ จะเห็นได้ว่า เมื่อความแปรปรวนของคะแนนความคลาดเคลื่อนคงที่ และความแปรปรวนของคะแนนรวมจากแบบทดสอบมีค่าสูงจะทำให้ค่าความเชื่อมั่นสูง แบบทดสอบที่การกระจายของคะแนนมีมาก จะส่งผลให้ความแปรปรวนของคะแนนรวมจากแบบทดสอบมีค่าสูง และส่งผลให้ความเชื่อมั่นของแบบทดสอบมีค่าสูงด้วย ในเรื่องการกระจายของคะแนนนี้เป็นผลเนื่องมาจาก

2.1 **ความเป็นเอกพันธ์ของกลุ่มนักเรียน** (Group Homogeneity) ถ้านักเรียนที่ไปทดสอบมีความเป็นเอกพันธ์มากจะทำให้การกระจายของคะแนนมีน้อย ซึ่งจะส่งผลให้ความเชื่อมั่นของแบบทดสอบต่ำด้วย ในทำนองกลับกัน หากกลุ่มที่ไปทดสอบมีความแตกต่างกันมากหรือมีความเป็นวิวิธพันธ์ จะทำให้คะแนนกระจายมาก ซึ่งมีผลทำให้ความแปรปรวนของคะแนนสูง อันจะทำให้ความเชื่อมั่นสูงตามไปด้วย

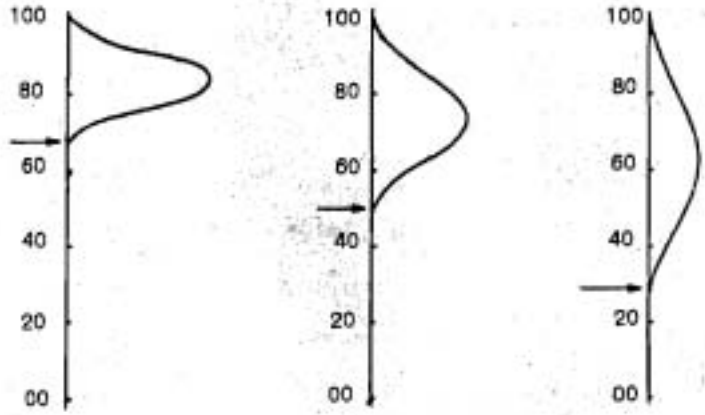
2.2 **ความยากง่ายของข้อสอบ** ความเชื่อมั่นของแบบทดสอบย่อมขึ้นอยู่กับความแปรปรวนของคะแนนผลการสอบ กล่าวคือถ้าความแปรปรวนระหว่างคะแนนผลการสอบมีน้อย ย่อมจะทำให้ค่าความเชื่อมั่นต่ำกว่าคะแนนผลการสอบที่มีความแปรปรวนมาก แบบทดสอบที่ยากเกินไปนักเรียนทำผิดเกือบทุกคน หรือแบบทดสอบที่ง่ายเกินไปนักเรียนทำถูกเกือบทุกคน จะจำแนกความสามารถของนักเรียนได้น้อย คะแนนจะไปรวมกันที่ใดที่หนึ่ง ทำให้ความแปรปรวนของคะแนนมีน้อย ซึ่งมีผลทำให้ความเชื่อมั่นของแบบทดสอบต่ำ ข้อสอบที่มีค่าความยากเท่ากับ .50 จะมีผลต่อความเชื่อมั่นมาก ทั้งนี้เพราะจะทำให้ความแปรปรวนในแต่ละข้อสูงสุด คือ pq = .25

ดังนั้น ในการสร้างข้อสอบเพื่อให้มีความเชื่อมั่นสูง ๆ จึงไม่ควรใช้ข้อสอบที่ยากหรือง่ายเกินไป ควรใช้ข้อสอบที่มีความยากง่ายพอเหมาะ คือข้อสอบที่มีค่า p ใกล้ ๆ .50

2.3 **ชนิดของแบบทดสอบ** แบบทดสอบต่างชนิดกันมีผลทำให้คะแนนกระจายต่างกันและมีผลทำให้ความเชื่อมั่นของแบบทดสอบแตกต่างกันด้วย ตัวอย่างเช่น ข้อสอบบรรยายคะแนนเต็ม 100 คะแนน เวลาให้คะแนนก็มักจะให้ระหว่าง 70 ถึง 100 ซึ่งทำให้พิสัยเท่ากับ 30 สำหรับข้อสอบแบบปรนัยนั้นพิสัยตามทฤษฎีคือผลต่างระหว่างคะแนนเต็มกับคะแนนที่ได้จากการเดาถูก คะแนนที่ได้จากการเดาถูกนั้น หมายถึงคะแนนเต็มคูณด้วยโอกาสที่จะเดาถูกแต่ละข้อ เป็นต้นว่าข้อสอบแบบถูกผิด โอกาสที่จะเดาถูกเป็น $\frac{1}{2}$ ถ้าเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก โอกาสที่จะ

เดาถูกเป็น $\frac{1}{4}$ ถ้าข้อสอบถูกผิด 100 ข้อ คะแนนเต็ม 100 คะแนน คะแนนที่ได้จากการเดาก็เป็น 50 ดังนั้นพิสัยของคะแนนก็เท่ากับ 100-50 = 50 ถ้าเป็นข้อสอบเลือกตอบ 4 ตัวเลือก 100 ข้อ 100 คะแนน คะแนนที่ได้จากการเดาก็จะเป็น 25 ดังนั้นพิสัยของคะแนนก็เท่ากับ 100-25 = 75

Robert L. Ebel ได้สร้างแบบของการกระจายของคะแนน ดังรูป



ภาพแสดงการเปรียบเทียบความเชื่อมั่นของแบบทดสอบทั้งสาม

Ebel ได้สรุปว่า ข้อสอบประเภทที่มีพิสัยของคะแนนมาก จะทำให้ค่าความเชื่อมั่นของข้อสอบสูงขึ้น

3. **ความเป็นปรนัยในการให้คะแนน** ข้อสอบที่มีความเป็นปรนัยในการให้คะแนนสูงจะมีความเชื่อมั่นสูง ส่วนข้อสอบที่ไม่มีความเป็นปรนัยในการให้คะแนน เช่น ข้อสอบประเภทความเรียง ซึ่งการตรวจให้คะแนนขึ้นอยู่กับอารมณ์ของผู้ตรวจ ไม่มีเกณฑ์การให้คะแนนที่แน่นอนจะมีผลทำให้ค่าความเชื่อมั่นของแบบทดสอบฉบับนั้นต่ำ

ไพศาล หวังพานิช ได้เสนอแนะวิธีการที่จะสร้างเครื่องมือในการวัดให้มีความเชื่อมั่นสูงไว้ดังนี้

1. มีความชัดเจนทั้งตัวคำถามและคำอธิบายหรือคำชี้แจงในการตอบ
2. เวลาที่ใช้ในการตอบต้องเหมาะสมพอดี ไม่มากหรือน้อยเกินไป
3. ต้องให้คะแนนอย่างมีหลักเกณฑ์แน่นอน
4. ใช้คำถามที่ถามตามหลักเกณฑ์หรือหลักวิชา เช่น ถ้าเป็นข้อสอบก็ต้องถามโดยสามารถหาคำตอบถูกได้โดยใช้หลักวิชา ไม่ใช่ถูกหรือผิดตามความรู้สึกนึกคิดของบุคคล
5. พยายามใช้ข้อคำถามมาก ๆ ข้อ

ความเที่ยงตรง (Validity)

ความเที่ยงตรงนับว่าเป็นคุณลักษณะที่สำคัญที่สุดของเครื่องมือในการวัดผล และเป็นยอดปรารถนาของทุกท่าน เพราะถ้าเครื่องมือในการวัดมีความเที่ยงตรงดีแล้วจะทำให้ผลที่ได้จากการวัดมีความหมายตามที่ต้องการจะวัด คงจะไม่มีใครปฏิเสธว่าตาชั่งใช้วัดน้ำหนัก ไม้เมตรใช้วัดความยาว และนาฬิกาใช้วัดเวลา แม้ว่านาฬิกาบางเรือนจะเดินช้าหรือเดินเร็วไปบ้างก็ตาม แต่ก็ยังเป็นเครื่องมือที่ใช้วัดเวลาอยู่นั่นเอง และก็คงจะไม่มีใครเอาเทอร์โมมิเตอร์ไปวัดน้ำหนักแน่ เครื่องมือในการวัดที่สำคัญของครูได้แก่แบบทดสอบ ทำอย่างไรจึงจะทำให้แบบทดสอบทดสอบที่ครูสร้างขึ้นมีความเที่ยงตรงสูง นับว่าเป็นเรื่องที่จะต้องศึกษากันต่อไป

ความเที่ยงตรงเป็นความสัมพันธ์ระหว่างค่าวัดที่ได้จากแบบทดสอบกับสิ่งที่ต้องการวัด สิ่งที่ต้องการวัดในที่นี่อาจเป็นเนื้อหาวิชาคณิตศาสตร์ ความสามารถในอนาคต เป็นต้น โดยทั่วไปเรียกสิ่งที่ต้องการวัดนี้ว่าตัวแปรเกณฑ์ ดังนั้นถ้าจะให้นิยามของความเที่ยงตรงตามนัยนี้ก็จะได้ว่า ความเที่ยงตรงเป็นความสัมพันธ์ระหว่างค่าวัดที่ได้จากแบบทดสอบกับค่าวัดของตัวแปรเกณฑ์

เมื่อความเที่ยงตรงเป็นความสัมพันธ์ระหว่างคะแนนที่ได้จากแบบทดสอบกับตัวแปรเกณฑ์ ดังนั้นความเที่ยงตรงจะมีมากหรือน้อยขึ้นอยู่กับว่าสัมพันธ์กันมากหรือน้อยโดยทั่วไปแบบทดสอบทุกฉบับจะมีความเที่ยงตรงทั้งนั้น เพียงแต่ว่าจะมากหรือน้อย ในการหาความเที่ยงตรงนั้นนิยมหาโดยใช้สัมประสิทธิ์สหสัมพันธ์

ดังได้กล่าวแล้วว่าตัวแปรเกณฑ์มีหลายชนิด ดังนั้นความเที่ยงตรงจึงมีหลายชนิดด้วย เช่น ถ้าใช้เนื้อหาเป็นเกณฑ์ ก็จะเรียกว่าความเที่ยงตรงเชิงเนื้อหา ถ้าใช้ความสามารถในอนาคตเป็นเกณฑ์ เรียกว่า ความเที่ยงตรงเชิงพยากรณ์ แต่ในที่นี้จะแบ่งชนิดของความเที่ยงตรง ซึ่งเสนอแนะโดย American Educational Research Association ออกเป็น 4 ชนิด คือ

1. ความเที่ยงตรงเชิงเนื้อหา (Content Validity)

2. ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity)

3. ความเที่ยงตรงเชิงสภาพ (Concurrent Validity)

4. ความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity)

1. **ความเที่ยงตรงเชิงเนื้อหา** หมายถึงความสอดคล้องของเนื้อหาที่ต้องการวัดกับเนื้อหาของแบบทดสอบ นั่นคือ ความเที่ยงตรงชนิดนี้มุ่งจะตรวจสอบว่า เครื่องมือนั้นสามารถวัดเนื้อหาที่ต้องการวัดได้จริงหรือไม่ เช่น สมมุติว่าต้องการวัดความสามารถทางคณิตศาสตร์ คำถามที่ใช้ในแบบทดสอบก็ควรที่จะมุ่งวัดเฉพาะเนื้อหาหรือเรื่องราวทางคณิตศาสตร์เท่านั้น มิใช่ไปวัดความสามารถทางภาษาไทย นอกจากนี้แบบทดสอบนั้นยังจะต้องสามารถวัดเนื้อหาต่าง ๆ ตามที่กำหนดไว้ในหลักสูตรได้อย่างครบถ้วนทุกเนื้อหาและวัดแต่ละเนื้อหาเหล่านั้นมากน้อยตามลำดับความสำคัญ แต่ในการสอบครั้งหนึ่ง ๆ นั้นไม่สามารถจะเอาเนื้อหาทั้งหมดมาถามได้ เนื่องจากมีข้อจำกัดหลายประการ จึงมีความจำเป็นที่จะต้องสุ่มเนื้อหาบางส่วนมาใช้ในการสอบ ดังนั้นจะต้องพยายามทำให้เนื้อหาของข้อสอบเป็นตัวแทนที่ดีของเนื้อหาที่เรียน

ในการที่จะสร้างข้อสอบให้มีความเที่ยงตรงเชิงเนื้อหา จำเป็นจะต้องสร้างข้อสอบทั้งฉบับให้ได้น้ำหนักหรือเป็นสัดส่วนตามตารางวิเคราะห์หลักสูตร ซึ่งได้กล่าวมาแล้วในบทต้น ๆ

ความเที่ยงตรงเชิงเนื้อหานี้ ไม่สามารถแสดงตัวเลขบอกระดับของความเที่ยงตรงออกมาได้ โดยทั่วไปจะใช้การพิจารณาความเที่ยงตรงเชิงเนื้อหาเป็น 2 ลักษณะ คือ

1. พิจารณาดูว่าคำถามแต่ละข้อวัดเนื้อหาที่ต้องการวัดได้แท้จริงเพียงไร

2. พิจารณาแบบทดสอบโดยตลอดทั้งฉบับ โดยดูว่าคำถามทั้งหมดสามารถวัดเนื้อหาได้ครบทุกเนื้อหาหรือไม่ และมีจำนวนข้อที่ถามแต่ละเนื้อหาเหมาะสมกับสัดส่วนในตารางวิเคราะห์หลักสูตรหรือยัง

ในการพิจารณาความเที่ยงตรงเชิงเนื้อหานั้น นิยมให้ผู้เชี่ยวชาญหรือผู้ชำนาญในสาขาวิชานั้น ๆ หลาย ๆ คน เป็นผู้ร่วมกันพิจารณา อย่างไรก็ตาม การที่บุคคลหนึ่งบอกว่าแบบทดสอบฉบับนี้มีความเที่ยงตรงสูง ก็ไม่จำเป็นที่คนอื่นจะต้องเห็นด้วยเสมอไป ทั้งนี้เพราะแต่ละคนมีเกณฑ์ในการพิจารณาต่างกัน

2. **ความเที่ยงตรงเชิงโครงสร้าง** ในทางจิตวิทยา คำว่าโครงสร้างหมายถึงลักษณะต่าง ๆ ของมนุษย์ เช่น ความเข้มแข็ง ความก้าวร้าว แต่ในทางการวัดผลโครงสร้างหมายถึงความสามารถทางสมอง เช่น ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ ความเที่ยงตรงเชิงโครงสร้างต่างจากความเที่ยงตรงเชิงเนื้อหา เพราะความเที่ยงตรงเชิงเนื้อหายึดเนื้อหาเป็นเกณฑ์ ส่วนความเที่ยง

ตรงเชิงโครงสร้างจะยึดคุณลักษณะต่าง ๆ หรือความสามารถทางสมองเป็นเกณฑ์ การสร้างเครื่องมือให้มีความเที่ยงตรงเชิงโครงสร้างจึงจำเป็นจะต้องรู้ว่าความสามารถทางสมองและคุณลักษณะที่ต้องการวัดนั้นคืออะไร ตัวอย่างเช่น ต้องการวัดความก้าวร้าว ความมีเหตุผล ก็จะต้องมาพิจารณาว่าโครงสร้างที่ต้องการวัดนั้นประกอบด้วยคุณลักษณะ (trait) ใดบ้างแล้วนำคะแนนที่ได้จากแบบทดสอบที่สร้างขึ้นตามคุณลักษณะนั้น ๆ ไปหาสหสัมพันธ์กับแบบทดสอบที่ถือว่าเป็นเกณฑ์หรือมาตรฐานของเรื่องนั้น หรือหากต้องการวัดความสามารถทางสมองด้านความเข้าใจ การนำไปใช้ฯลฯ ก็จะต้องออกข้อสอบให้สอดคล้องกับน้ำหนักในตารางวิเคราะห์หลักสูตรในด้านพฤติกรรม จึงจะนับได้ว่าแบบทดสอบนั้นมีความเที่ยงตรงเชิงโครงสร้าง

วิธีการพิจารณาหาความเที่ยงตรงชนิดนี้ก็ทำได้เช่นเดียวกับการหาความเที่ยงตรงเชิงเนื้อหา

3. **ความเที่ยงตรงเชิงสภาพ** หมายถึงความสอดคล้องของค่าวัดที่ได้จากแบบทดสอบกับสภาพความเป็นจริงของเด็กในปัจจุบัน นั่นคือจะเป็นการตรวจสอบความเที่ยงตรงของแบบทดสอบที่สร้างขึ้น โดยใช้สถานภาพอันแท้จริงของเด็กในปัจจุบันมาเป็นตัวเกณฑ์ในการเปรียบเทียบ ความเที่ยงตรงชนิดนี้ จะช่วยให้เห็นสภาพที่เป็นจริงของเด็กเช่นใครที่ทำข้อสอบเกี่ยวกับการบวกลบได้คะแนนสูง หากให้นักเรียนผู้นั้นไปประจำอยู่ร้านค้า ก็จะต้องสามารถทอนเงิน คิดเงินได้คล่องแคล่วจริง ๆ จึงจะนับได้ว่าแบบทดสอบที่สร้างขึ้นฉบับนั้นมีความเที่ยงตรงเชิงสภาพสูง เพราะมีความสอดคล้องกับความเป็นจริง จะเห็นได้ว่าความเที่ยงตรงเชิงสภาพนี้ต้องการที่จะชี้ให้เห็นว่า เมื่อครูได้สอนไปตามทฤษฎีที่วางไว้แล้ว นักเรียนปฏิบัติได้จริงหรือไม่ ดังนั้นในเชิงปฏิบัติ การหาความเที่ยงตรงเชิงสภาพก็จะนำเอาคะแนนที่ได้จากการสอบไปเปรียบเทียบหาความสอดคล้องกับความสามารถที่แท้จริง เช่น คะแนนภาคปฏิบัติ เป็นต้น ค่าสหสัมพันธ์ที่ได้นี้จะเป็นค่าที่บอกระดับความเที่ยงตรงเชิงสภาพ นอกจากนี้ในการหาความเที่ยงตรงเชิงสภาพยังอาจพิจารณาดูว่าแบบทดสอบที่สร้างขึ้นนั้นสามารถจำแนกนักเรียนที่เก่งกับอ่อน คือคนเก่งได้คะแนนสูง คนอ่อนได้คะแนนต่ำ ได้หรือไม่ นั่นคือถ้าแบบทดสอบใดสามารถให้คะแนนกระจายก็แสดงว่าแบบทดสอบนั้นมีความเที่ยงตรงเชิงสภาพสูง หากแบบทดสอบใดไม่สามารถจำแนกคะแนนระหว่างนักเรียนที่เก่งกับอ่อนได้ ก็แสดงว่าแบบทดสอบนั้นขาดความเที่ยงตรงเชิงสภาพ

4. **ความเที่ยงตรงเชิงพยากรณ์** หมายถึงความสอดคล้องระหว่างผลการสอบกับความสำเร็จในอนาคตของนักเรียน โดยพิจารณาว่าแบบทดสอบที่สร้างขึ้นนี้สามารถจะพยากรณ์ได้ตามวัตถุประสงค์จริงหรือไม่ ผลการวัดสามารถทำนายคุณลักษณะและความสามารถในอนาคตของบุคคลได้ตรงกับความเป็นจริงเพียงไร วิธีการหาความเที่ยงตรงชนิดนี้จะใช้วิธีการหาความสัมพันธ์ระหว่าง

คะแนนจากแบบทดสอบที่สร้างขึ้น กับผลสัมฤทธิ์ทางการเรียนในภายภาคหน้า ซึ่งถือว่าเป็นตัวเกณฑ์ ความเที่ยงตรงเชิงพยากรณ์นี้เป็นยอดปรารถนาของครูทุกคน เพราะสามารถใช้ผลจากการสอบนั้นสำหรับวัดความถนัดหรือใช้ในการแนะแนวได้ ตัวอย่างเช่น ในการสอบคัดเลือก นักเรียนเข้าเรียนโดยใช้แบบทดสอบความถนัดทางการเรียน ก็มุ่งหวังจะคัดเอาคนเก่งหรือคนฉลาดเข้ามาเรียน ถ้านักเรียนที่ได้รับการคัดเลือกเข้ามาสามารถเรียนได้สำเร็จทุกคน ก็แสดงว่าแบบทดสอบความถนัดฉบับนั้นมีความเที่ยงตรงเชิงพยากรณ์สูง แต่ในทางตรงกันข้าม หากปรากฏว่านักเรียนที่ได้รับการคัดเลือกเข้ามาส่วนใหญ่เรียนไม่สำเร็จ ก็ย่อมแสดงว่าแบบทดสอบความถนัดที่ใช้ฉบับนั้นขาดความเที่ยงตรงเชิงพยากรณ์

จะเห็นได้ว่าความเที่ยงตรงเชิงพยากรณ์ คล้ายกับความเที่ยงตรงเชิงสภาพในแง่ที่ว่าเป็นการหาความสัมพันธ์ระหว่างคะแนนผลการสอบกับเกณฑ์ภายนอก แต่ความเที่ยงตรงนั้น 2 ชนิดนี้แตกต่างกันตรงเวลา กล่าวคือ ความเที่ยงตรงเชิงพยากรณ์ดูพฤติกรรมในอนาคตเป็นเกณฑ์ ส่วนความเที่ยงตรงเชิงสภาพนั้นดูพฤติกรรมในปัจจุบันเป็นเกณฑ์

วิธีการหาความเที่ยงตรง

ในการหาความเที่ยงตรงเชิงเนื้อหา และความเที่ยงตรงเชิงโครงสร้างนั้นนิยมใช้วิธีการพิจารณาเป็นหลัก โดยทำการตรวจสอบว่าข้อสอบฉบับนั้นสามารถวัดเนื้อหาและพฤติกรรมต่าง ๆ ตามตารางวิเคราะห์หลักสูตรหรือไม่ ซึ่งอาจกล่าวได้ว่าเป็นการหาความเที่ยงตรงตามหลักสูตร (Curriculum validity) ส่วนการหาความเที่ยงตรงเชิงสภาพ และความเที่ยงตรงเชิงพยากรณ์ นิยมหาโดยการคำนวณค่าสหสัมพันธ์ระหว่างคะแนนผลการสอบกับเกณฑ์ที่ปรากฏในปัจจุบัน หรืออนาคต ซึ่งเป็นการตรวจสอบโดยอาศัยความสัมพันธ์กับเกณฑ์ชนิดอื่น จึงอาจเรียกว่าความเที่ยงตรงตามเกณฑ์สัมพันธ์ (Criterion related validity)

นอกจากนี้ยังอาจหาความเที่ยงตรงโดยวิธีการที่เรียกว่า Known group technique ซึ่งมีวิธีการดังนี้คือ เมื่อสร้างเครื่องมือวัดคุณลักษณะชนิดหนึ่งขึ้นมาแล้วให้นำเครื่องมือนั้นไปสอบวัดบุคคลกลุ่มหนึ่งที่ทราบอย่างแน่ชัดว่ามีคุณลักษณะที่จะวัดนั้นจริง ๆ ทั้งนี้เพื่อนำผลไปเปรียบเทียบกับผลการวัดของบุคคลทั่วไปอีกกลุ่มหนึ่ง ถ้าบุคคลกลุ่มที่มีลักษณะนั้นตอบได้คะแนนสูงกว่ากลุ่มทั่วไป ก็แสดงว่าเครื่องมือนั้นวัดคุณลักษณะนั้นได้จริง การทดสอบความแตกต่างของรายเฉลี่ยระหว่างกลุ่มที่มีคุณลักษณะนั้นกับกลุ่มทั่วไป สามารถทดสอบโดยใช้สูตร

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{N_1} + \frac{S_2^2}{N_2}}} \qquad \text{.....(6-22)}$$

เมื่อ t แทน ค่าสถิติที่ใช้ตัดสินระดับของความแตกต่าง

$\bar{X}_1$ แทน ค่าเฉลี่ยคะแนนของกลุ่มที่รู้ลักษณะ

$\bar{X}_2$ แทน ค่าเฉลี่ยคะแนนของกลุ่มทั่วไป

S_1^2, S_2^2 แทน ค่าความแปรปรวนของคะแนนในแต่ละกลุ่ม ตามลำดับ

N_1, N_2 แทน จำนวนคนในแต่ละกลุ่ม

การหาความเที่ยงตรง โดยใช้ Known group technique นิยมใช้หาความเที่ยงตรง 2 ประเภท ต่อไปนี้คือ

1. ความเที่ยงตรงเชิงโครงสร้าง เกณฑ์ที่ใช้เป็นกลุ่มบุคคลที่มีพฤติกรรมและคุณลักษณะต่าง ๆ ที่จะวัดนั้นจริง เช่น เครื่องมือวัดความก้าวร้าว ก็ต้องนำไปทดลองสอบกับคนที่ทราบแน่ชัดแล้วว่าเป็นคนที่มีความก้าวร้าวสูง

2. ความเที่ยงตรงเชิงสภาพ เกณฑ์ที่ใช้เป็นกลุ่มที่มีสภาพต่าง ๆ ในปัจจุบันเป็นจริง เช่น เครื่องมือวัดความสนใจในอาชีพตำรวจ ก็ต้องนำไปทดลองสอบกับนักเรียนตำรวจ

ความสัมพันธ์ระหว่างความเที่ยงตรงกับความเชื่อมั่น

ความเที่ยงตรงของแบบทดสอบจะมีค่าไม่เกินรากที่สองของค่าความเชื่อมั่น นั่นคือ ความเที่ยงตรงสูงสุด $= \sqrt{\text{ความเชื่อมั่น}}$

หรือ $r_{xy} = \sqrt{r_{tt}}$ .....(6-23)

เมื่อ r_{xy} แทน ความเที่ยงตรงสูงสุดของแบบทดสอบที่สร้างขึ้น เมื่อเทียบกับเกณฑ์

r_{tt} แทน ความเชื่อมั่นของแบบทดสอบที่สร้างขึ้น

ตัวอย่าง แบบทดสอบฉบับหนึ่งมีค่าความเชื่อมั่นเท่ากับ 0.64 แบบทดสอบฉบับนี้จะมีความเที่ยงตรงสูงสุดไม่เกินเท่าไร

$$r_{xy} = \sqrt{0.64}$$
$$= 0.80$$

การแก้ค่าความเที่ยงตรง

ความเชื่อมั่นของคะแนนเกณฑ์ก็คล้ายกับความเชื่อมั่นของคะแนนจากแบบทดสอบที่สร้างขึ้น คือจะมีความคลาดเคลื่อนทางการวัดเกิดขึ้นเสมอ เช่น นำข้อสอบที่เป็นเกณฑ์ไปสอบนักเรียน 2 ครั้ง ก็ย่อมจะได้คะแนนแตกต่างกัน เช่น ครั้งแรกนักเรียนอาจทำถูก 30 ข้อ ครั้งที่สองอาจทำถูกเพียง 20 ข้อ เป็นต้น สาเหตุที่นักเรียนได้คะแนนแตกต่างกันในการสอบทั้งสองครั้ง ก็เนื่องมาจากความเคลื่อนในการวัดของเกณฑ์นั่นเอง

โดยทั่วไปหากคะแนนที่ได้จากเกณฑ์ขาดความเชื่อมั่น ก็จะทำให้คะแนนของแบบทดสอบที่สร้างขึ้นขาดความเที่ยงตรงไปด้วย ในทางสถิติมีวิธีการแก้คะแนนอันเนื่องมาจากความคลาดเคลื่อนของเกณฑ์ให้มีความเที่ยงตรงสูงขึ้น โดยใช้สูตร ดังนี้

$$r_{cc} = \frac{r_{xy}}{\sqrt{r_{yy}}} \qquad \text{.....(6-24)}$$

เมื่อ r_{cc} แทน ความเที่ยงตรงที่แก้แล้ว อันเนื่องมาจากความคลาดเคลื่อนของเกณฑ์

r_{xy} แทน สัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบทั้งสอง

r_{yy} แทน ความเชื่อมั่นของเกณฑ์

ตัวอย่าง เกณฑ์มีความเชื่อมั่น 0.49 และมีสหสัมพันธ์กับคะแนนจากแบบทดสอบที่สร้างขึ้นเป็น 0.35 จงหาค่าความเที่ยงตรงที่แก้แล้ว อันเนื่องมาจากความคลาดเคลื่อนของเกณฑ์

$$r_{cc} = \frac{0.35}{\sqrt{0.49}}$$

$$= \frac{0.35}{0.70}$$

$$= 0.50$$

ในกรณีที่คะแนนจากแบบทดสอบที่สร้างขึ้น และคะแนนจากเกณฑ์มีความคลาดเคลื่อนในการวัดสูง ก็จะทำให้คะแนนที่ได้จากแบบทดสอบที่สร้างขึ้น และคะแนนจากเกณฑ์มีค่าความเชื่อมั่นต่ำ ซึ่งจะส่งผลต่อค่าความเที่ยงตรงด้วย วิธีการที่จะกำจัดความคลาดเคลื่อนอันเกิดจากแบบทดสอบที่สร้างขึ้นและเกณฑ์ เพื่อให้ได้ค่าความเที่ยงตรงสูงสุดในทางสถิติ อาจทำได้โดยใช้สูตร

$$r_{cxy} = \frac{r_{xy}}{\sqrt{r_{xx}\, r_{yy}}} \qquad \text{.....(6-25)}$$

เมื่อ r_{cxy} แทน ความเที่ยงตรงที่แก้แล้ว

r_{xx} แทน ความเชื่อมั่นของแบบทดสอบที่สร้างขึ้น

r_{yy} แทน ความเชื่อมั่นของเกณฑ์

r_{xy} แทน สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบทดสอบที่สร้างขึ้นกับเกณฑ์

ตัวอย่าง แบบทดสอบที่สร้างขึ้นมีความเชื่อมั่นเท่ากับ 0.64 และแบบทดสอบที่ใช้เป็นเกณฑ์มีความเชื่อมั่นเท่ากับ 0.49 สัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบทั้งสองเป็น 0.35 จงหาค่าความเที่ยงตรงที่แก้แล้ว

$$r_{cxy} = \frac{0.35}{\sqrt{(0.64)\ (0.49)}}$$

$$= \frac{0.35}{(0.80)\ (0.70)}$$

$$= \frac{35}{56}$$

$$= 0.62$$

จะเห็นว่าค่าความเที่ยงตรง 0.62 นี้ ได้จากการกำจัดความคลาดเคลื่อนอันเกิดจากแบบทดสอบที่สร้างขึ้น และเกณฑ์โดยใช้สถิติเข้ามาช่วย และเป็นผลทำให้แบบทดสอบที่สร้างขึ้น และแบบทดสอบที่เป็นเกณฑ์ มีความเชื่อมั่นอย่างสมบูรณ์ หรือไม่มีความคลาดเคลื่อนเลย ซึ่งมีผลทำให้ค่าความเที่ยงตรงของแบบทดสอบที่สร้างขึ้นมีค่าสูงกว่าค่าความเที่ยงตรงที่ได้จากการแก้ตัวแปรเกณฑ์แต่เพียงอย่างเดียว ดังจะเห็นได้จากตัวอย่างข้างต้น ซึ่งมีค่าความเที่ยงตรงเท่ากับ 0.50 และ 0.62 ตามลำดับ

องค์ประกอบที่มีอิทธิพลต่อความเที่ยงตรง

สาเหตุที่ทำให้แบบทดสอบขาดความเที่ยงตรงมีอยู่หลายอย่าง บางอย่างเห็นได้ชัดเจนและสามารถหลีกเลี่ยงได้ เช่น การออกข้อสอบวิทยาศาสตร์เป็นภาษาอังกฤษ หรือการออกข้อสอบภาษาอังกฤษโดยต้องใช้ความรู้ทางภูมิศาสตร์มาตอบ กรณีเช่นนี้จะทำให้ข้อสอบขาดความเที่ยงตรง ส่วนสาเหตุบางอย่างที่ทำให้ข้อสอบขาดความเที่ยงตรง อาจเห็นได้ไม่ชัดเจนนัก เช่น คำสั่งของข้อสอบไม่ชัดเจน เข้าใจยาก ต้องนักเรียนที่มีความสามารถอย่างดีในการอ่านจึงจะเข้าใจได้ กรณีเช่นนี้จะเป็นว่าแบบทดสอบฉบับนั้นวัดความเข้าใจในการอ่านคำสั่งชี้แจง แทนที่จะวัดเนื้อหาวิชาจริง ๆ ซึ่งคงจะไม่ตรงกับวัตถุประสงค์ที่จะวัดความรู้ในเนื้อหาวิชาที่ต้องการวัด เช่น ความสามารถทางคณิตศาสตร์ เป็นต้น สาเหตุดังกล่าวทำให้ข้อสอบมีความเที่ยงตรงต่ำ

องค์ประกอบที่มีอิทธิพลต่อความเที่ยงตรงของข้อสอบ พอสรุปได้ดังนี้

1. องค์ประกอบในตัวข้อสอบ เป็นความบกพร่องที่เกิดจากตัวข้อสอบเอง เช่น คำสั่งไม่ชัดเจน ศัพท์ที่ใช้ยากเกินระดับเด็ก ข้อสอบง่ายหรือยากเกินไป ข้อสอบบางข้อแนะคำตอบ ข้อสอบไม่วัดพฤติกรรมที่ต้องการวัด ฯลฯ

2. องค์ประกอบจากกระบวนการสอน บางครั้งการสอนของครูก็มีผลต่อความเที่ยงตรงของข้อสอบ เช่น ครูต้องการออกข้อสอบเพื่อวัดเหตุผลทางคณิตศาสตร์ แต่หากครูเคยนำข้อสอบข้อนั้นมาให้นักเรียนทำในห้องก่อนการสอบแล้ว ข้อสอบข้อนั้นจะเป็นเพียงข้อสอบที่วัดความจำ

ซึ่งผิดกับวัตถุประสงค์เดิมของครูที่ต้องการให้ข้อสอบ ข้อนั้นวัดพฤติกรรมทางด้านเหตุผลทางคณิตศาสตร์ สิ่งนี้ย่อมทำให้ข้อความขาดความเที่ยงตรงได้

3. องค์ประกอบจากตัวผู้สอบ ผู้สอบอาจเกิดการกลัว ประหม่า หรือตกใจในขณะสอบ สิ่งเหล่านี้มีผลต่อความเที่ยงตรงของข้อสอบเช่นกัน

ไพศาล หวังพานิช ได้เสนอแนะวิธีการที่จะสร้างเครื่องมือให้มีความเที่ยงตรงสูง ไว้ดังนี้

1. ต้องรู้จักพฤติกรรมหรือคุณลักษณะที่ต้องการจะวัด และสร้างคำถามวัดได้ตรงตามคุณสมบัติเหล่านั้น

2. ต้องสร้างเครื่องมือให้มีคุณภาพด้านยุติธรรม (fair) คือต้องวัดได้ครบทุกเรื่อง ทุกเนื้อหา และทุกพฤติกรรม หรือทุกคุณลักษณะ อย่างได้สัดส่วนความสำคัญ

3. ต้องใช้คำถามที่ชัดเจน (definite) ไม่คลุมเครือ หรือใช้ภาษาวกวน

4. ต้องถามในสิ่งที่เกี่ยวกับชีวิตประจำวัน เป็นเรื่องใกล้ตัว และเด็กรู้จักสิ่งต่าง ๆ เหล่านั้น

5. ต้องใช้คำถามที่มีอำนาจจำแนกสูง