การวิเคราะห์ข้อสอบ 9

วัตถุประสงค์

เมื่อผู้อ่านได้ศึกษาเนื้อหาบทที่ 9 แล้ว ควรจะมีความสามารถดังนี้

1. บอกจุดมุ่งหมายของการวิเคราะห์ข้อสอบได้
2. อธิบายความหมายและวิธีการในการหาอำนาจจำแนกและความยากง่ายได้
3. สามารถวิเคราะห์ข้อสอบตามทฤษฎีการวัดผลแบบอิงเกณฑ์ได้

เนื้อหา

ในการประกอบอาชีพที่ใช้เทคนิคทุกสาขา จำเป็นต้องมีเครื่องมือในการปฏิบัติงานทั้งสิ้น เช่น อาชีพแพทย์ จำเป็นต้องมีเครื่องมือผ่าตัด อาชีพช่างไม้จำเป็นต้องมี กบ สิ่ว ฆ้อน อาชีพครูก็จำเป็นต้องมีเครื่องมือในการปฏิบัติงานเช่นเดียวกับอาชีพอื่น ๆ และเครื่องมือที่สำคัญในการปฏิบัติงานของครูก็คือแบบทดสอบ การที่จะทราบว่าเครื่องมือที่ใช้มีคุณภาพดีเพียงใด จำเป็นต้องมีการตรวจสอบโดยนำเกณฑ์ต่าง ๆ มาเป็นหลักในการพิจารณา ทั้งนี้เพื่อจะสรุปว่าเครื่องมือนั้นดี เหมาะสมที่จะนำมาใช้งานหรือไม่ หากไม่ดีจะทำการปรับปรุงอย่างไร วิธีการในการพิจารณาตรวจสอบคุณภาพของข้อสอบเรียกว่าการวิเคราะห์ข้อสอบ

การวิเคราะห์ข้อสอบ คือเทคนิคสำหรับตรวจสอบคุณภาพของข้อคำถามเป็นรายข้อ ทั้งนี้โดยมีความมุ่งหมายที่จะทราบคุณลักษณะที่สำคัญของแบบทดสอบ 2 ประการ คือ ระดับความยากและอำนาจจำแนก

อำนาจจำแนก (Discriminating power) หมายถึงความสามารถของข้อคำถามที่จะแบ่งนักเรียนออกเป็น 2 กลุ่ม คือ กลุ่มเก่งกับกลุ่มอ่อน อำนาจจำแนกนี้เปรียบได้กับความไวของตาชั่งที่มีต่อน้ำหนักวัตถุที่วางบนเครื่องชั่ง ข้อสอบที่ง่ายเกินไป (ทุกคนตอบถูกหมด) หรือข้อสอบที่ยากเกินไป (ทุกคนตอบผิด) จะไม่มีอำนาจการจำแนก คือไม่สามารถจำแนกเด็กเก่งและอ่อนออกจากกันได้ ดังนั้นข้อสอบที่มีลักษณะเช่นนี้จึงเป็นข้อสอบที่ไม่พึงปรารถนา

ระดับความยาก (Difficulty) หมายถึงจำนวนเปอร์เซ็นต์ หรือสัดส่วนที่นักเรียนทั้งหมดตอบข้อนั้นถูก ข้อสอบที่ดีจะต้องมีความยากง่ายพอเหมาะ

จะหาอำนาจจำแนกของข้อสอบได้อย่างไร?

ตารางที่ 9.1 แสดงคะแนนรวมของนักเรียน และคะแนนในข้อที่ 5 ของนักเรียนจำนวน 10 คน

นักเรียนคนที่	คะแนนรวม	คะแนนข้อที่ 5
1	6	0
2	4	1
3	1	0
4	10	1
5	9	1
6	1	0
7	2	0
8	8	1
9	7	1
10	8	1

จากตาราง 9-1 จะเห็นว่าข้อสอบข้อที่ 5 นักเรียนที่ได้คะแนนสูงมีแนวโน้มที่จะตอบถูกมากกว่านักเรียนที่ได้คะแนนรวมต่ำ จึงสามารถหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรวมกับคะแนนในข้อที่ 5 ได้ และค่าสัมประสิทธิ์สหสัมพันธ์นี้ก็คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ 5 นั่นเอง

9.1 วิธีหาสหสัมพันธ์

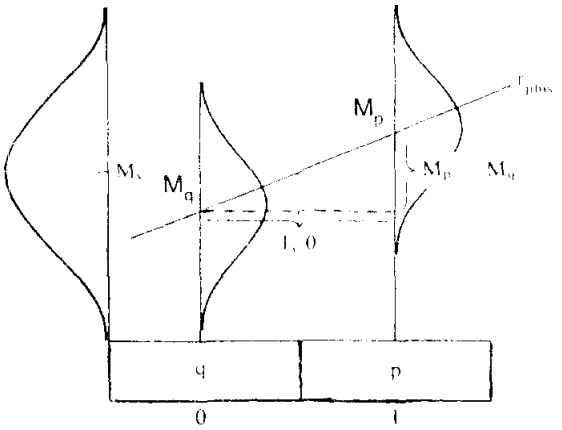
ก. Point - biserial correlation

ในบทที่ 4 ได้กล่าวถึงการหาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรทั้งสองที่มีลักษณะต่อเนื่อง (Continuous variables) โดยใช้สูตร

$$r_{xy} = \frac{N\Sigma XY - \Sigma X \Sigma Y}{\sqrt{[\, N\Sigma X^2 + (\Sigma X)^2 \,][\, N\Sigma Y^2 - (\Sigma Y)^2 \,]}}$$

และหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปรที่เป็น Dichotomus ทั้งคู่ โดยใช้ phi - coefficient

หากตัวแปรที่ต้องการศึกษามีลักษณะไม่เหมือนกับที่กล่าวมาข้างต้น เช่น ถ้านำแบบทดสอบไปทดสอบนักเรียนชาย 50 คน หญิง 50 คน เพื่อจะดูว่าเพศมีความสัมพันธ์กับคะแนนผลการสอบหรือไม่ ดังแสดงในภาพ 9-1



รูป 9-1 แสดงสหสัมพันธ์แบบ Point - biserial

จากภาพ 9-1 แกน X จะเป็นตัวแปร dichotomus โดยให้ 1 แทนชาย และ 0 แทนหญิง แทน Y แทนคะแนนผลการสอบซึ่งมีการแจกแจงเป็นโค้งปกติ ในภาพจะแบ่งการแจกแจงออกเป็น 2 ส่วน คือส่วนหนึ่งเป็นการแจกแจงคะแนนของนักเรียนชาย และอีกส่วนหนึ่งเป็นการแจกแจงคะแนนของนักเรียนหญิง

ถ้าสหสัมพันธ์ระหว่างคะแนนผลการสอบกับเพศเป็น 0 แล้ว ค่าเฉลี่ยของการแจกแจงในกลุ่มนักเรียนชายและหญิงจะต้องเท่ากัน และต้องเท่ากับค่าเฉลี่ยของการแจกแจงของแบบทดสอบ (M_y)

จากตัวอย่างในภาพ 9 - 1 แสดงให้เห็นว่าเพศกับคะแนนผลการสอบมีความสัมพันธ์กันในทางบวก โดยที่ค่าสหสัมพันธ์นี้บ่งให้เห็นว่าค่าเฉลี่ยของคะแนนในกลุ่มนักเรียนชายสูงกว่าค่าเฉลี่ยของคะแนนในกลุ่มนักเรียนหญิง ($M_p > M_q$)

ในการทำนายค่า y จาก x จะได้ว่า

$$b_{yx} = r_{xy} \frac{S_y}{S_x}$$

หรืออาจเขียน

$$r_{xy} = b_{yx} \frac{S_x}{S_y} \qquad \text{.....(9-1)}$$

เมื่อ b_{yx} แทนความลาด (slope) ของเส้นถดถอย ซึ่งเท่ากับค่า tangent ของมุมระหว่างแกน X กับเส้นถดถอยนั่นเอง

ค่าเฉลี่ยในตัวแปรเพศหญิงเป็น 0 และค่าเฉลี่ยในตัวแปรเพศชายเป็น 1 ดังนั้นความแตกต่างของค่าเฉลี่ยในตัวแปรทั้งสอง (ตัวแปรในแกน X) จึงเป็น 1 จึงได้ว่า

$$b_{yx} = \frac{M_p \quad M_q}{1}$$

สมการ 9-1 สามารถเขียนใหม่ได้ดังนี้

$$r_{xy} = \frac{(M_p - M_q) \quad S_x}{S_y}$$

เนื่องจากคะแนนในแกน X ประกอบด้วย 0 และ 1 เท่านั้น ดังนั้นจะได้ว่า $S_x = \sqrt{pq}$ ดังนั้น สามารถเขียนสมการใหม่ได้ดังนี้

$$r_{pbis} = \frac{M_p - M_q}{S_y} \cdot \sqrt{pq} \qquad \ldots\ldots(9\text{-}2)$$

เมื่อ M_p แทนค่าเฉลี่ยของคะแนนสำหรับตัวแปรที่เป็น 1

M_q แทนค่าเฉลี่ยของคะแนนสำหรับตัวแปรที่เป็น 0

S_y แทนความเบี่ยงเบนมาตรฐานของคะแนนรวมทั้งหมด

p แทนสัดส่วนจำนวนคนในตัวแปรที่เป็น 1 กับจำนวนคนทั้งหมด

q แทนสัดส่วนของจำนวนคนในตัวแปรที่เป็น 0 กับจำนวนคนทั้งหมด

ความแตกต่างระหว่าง $M_p - M_y$ และ $M_p - M_q$ จะมีค่าความลาดเท่ากัน ทั้งนี้เพราะเป็นเส้นตรงเส้นเดียวกัน จะเห็นได้ว่าค่า M_q ต้องคำนวณมาจากข้อสอบทุก ๆ ข้อ ส่วน M_y เป็นค่าคงที่ซึ่งคำนวณเพียงครั้งเดียว ดังนั้นเพื่อให้ง่ายขึ้น จึงใช้ $M_p - M_y$ แทน $M_p - M_q$ ทั้งนี้เพราะความแตกต่างดังกล่าวเท่ากัน ดังที่กล่าวมาแล้ว ซึ่งจะได้สมการใหม่ดังนี้

จากสมการ (9-2) ที่ว่า

$$r_{pbis} = \frac{M_p - M_q}{S_y}\sqrt{pq}$$

ถ้าให้ P แทน จำนวนคนในกลุ่มที่เป็น 1 (ชาย)

Q แทน จำนวนคนในกลุ่มที่เป็น 0 (หญิง)

N แทน จำนวนคนทั้งหมด (N = P + Q)

$$p = \frac{P}{N} \Rightarrow N_p = P$$

$$q = \frac{Q}{N} \Rightarrow N_q = Q$$

$$X_y = X_p + X_q$$

$$\Sigma X_y = \Sigma X_p + \Sigma X_q \qquad .(9\text{-}3)$$

$$\frac{\Sigma X_y}{N} = M_y$$

$$\Sigma X_y = NM_y$$

ในทำนองเดียวกัน

$$\Sigma X_p = PM_p$$

$$\Sigma X_q = QM_q$$

แทนค่าในสมการ (9-3)

$$NM_y = PM_p + QM_q$$

$$M_q = \frac{NM_y - PM_p}{Q} \qquad .(9\text{-}4)$$

นำค่า M_q ในสมการ (9-4) ไปแทนในสมการ (9-2)

$$r_{pbis} = \frac{M_p - \frac{NM_y - PM_p}{Q}}{S_y}\sqrt{pq}$$

$$= \frac{\frac{QM_p - NM_y + PM_p}{Q}}{S_y}\sqrt{pq}$$

$$= \frac{\frac{M_p(Q + P) - NM_y}{Q}}{S_y}\sqrt{pq}$$

$$= \frac{\frac{M_p N - N M_y}{Q}}{S_y} \sqrt{pq}$$

$$= \frac{\frac{N (M_p - M_y)}{N_q}}{S_y} \sqrt{pq}$$

$$= \frac{\frac{M_p - M_y}{q}}{S_y} \sqrt{pq}$$

$$= \frac{M_p - M_y}{S_y} \sqrt{\frac{pq}{q^2}}$$

$$r_{pbis} = \frac{M_p - M_y}{S_y} \sqrt{\frac{p}{q}} \quad \ldots (9\text{-}5)$$

ตาราง 9-2 แสดงคะแนนจากแบบสอบถามวัดความสนใจเกี่ยวกับเครื่องยนต์ของนักเรียน ชาย 50 คน หญิง 50 คน

คะแนน	หญิง (X_q)	ชาย (X_p)	รวม (X_t)
20		1	1
19			
18		1	1
17		3	3
16		4	4
15		6	6
14		6	6
13		7	7
12	4	4	8
11	4	6	10
10	8	5	13
9	10	3	13
8	9	2	11
7	3	2	5
6	4		4
5	5		5
4			
3	3		3
2			
1			
N	50	50	100
Σ	41 3	639	
M	8.26	12.70	
	S_y = 3.49		

จากตาราง 9-2 จะได้ว่า $M_p = 12.78$, $M_q = 8.26$, $S_y = 3.49$, และ $p = q = 0.50$ ดังนั้นสามารถคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ point biserial ระหว่างเพศกับความสนใจได้ดังนี้

$$r_{pbis} = \frac{12.78 - 8.26}{3.49} \times \sqrt{0.50 \times 0.50}$$

$$= 0.648$$

หากนำไปแทนในสมการ 9-5 ก็จะได้ค่าเช่นเดียวกัน

จะใช้สหสัมพันธ์ แบบ point biserial ก็ต่อเมื่อตัวแปรหนึ่งเป็น dichotomus และอีกตัวแปรหนึ่งเป็นตัวแปรต่อเนื่อง (continuous)

ค่าสัมประสิทธิ์สหสัมพันธ์จะขึ้นอยู่กับความสัมพันธ์ระหว่างสัดส่วนของ p และ q ในตัวแปร dichotomus กล่าวคือ ถ้า $p = q$ ค่าสูงสุดของสัมประสิทธิ์สหสัมพันธ์จะน้อยกว่า 1.0 และจะมีค่าลดลงเรื่อย ๆ ถ้าความแตกต่างระหว่างสัดส่วนของ p และ q เพิ่มขึ้น

จากสมการ 9-2 จะเห็นว่าสัมประสิทธิ์สหสมพันธ์ point biserial จะสูงสุดเมื่อ $p = q$ ในกรณีที่ตัวอื่นคงที่

จากสมการ 9-5 ถ้า $p = q$, $M_y = 0$ และ $S_y = 1.0$ จะได้ว่า $M_p - M_y = M_p = 1.0$ เมื่อ $r_{pbis} = 1.0$

ถ้า $M_y = 0$ และ $S_y = 1.0$ แสดงว่าคะแนนทุกตัวในแกน Y สามารถแสดงได้โดย z - score นั่นคือ ถ้า $p = q$ จะได้ว่าค่าสูงสุดของ $r_{pbis} = M_p$

ตัวอย่าง ถ้าค่าเฉลี่ยของคะแนน Z - score (M_p) เป็น 0.7978

จะได้ว่า $r_{pbis} = 0.7978$

สรุปเกี่ยวกับการหาสัมประสิทธิ์สหสัมพันธ์โดยใช้ point biserial correlation

การหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่นักเรียนทำได้ในข้อสอบแต่ละข้อกับคะแนนรวม โดยใช้ point biserial correlation จะทำได้ต่อเมื่อ

1. คะแนนรวมของข้อสอบฉบับนั้นมีการกระจายเป็นโค้งปกติ (normal curve)
2. การให้คะแนนข้อสอบแต่ละข้อเป็นแบบ 0, 1 คือ ตอบผิดให้ 0 ตอบถูกให้ 1

ค่าสัมประสิทธิ์ที่หาได้นี้ก็คือ ค่าอำนาจจำแนกของข้อสอบนั่นเอง ซึ่งจะมีค่าสูงสุดเมื่อ $p = q = .50$

ข. Biserial correlation

ในการหาสัมประสิทธิ์สหสัมพันธ์ แบบ point biserial นั้น มีข้อตกลงเบื้องต้นว่า คะแนนของนักเรียนแต่ละคนมีการแจกแจงเป็นโค้งปกติ ซึ่งข้อตกลงพื้นฐานอันนี้ พบกันบ่อยมากในทางทฤษฎี การทดสอบที่เกี่ยวข้องกับคะแนนของตัวแปรทางด้านจิตวิทยา

ให้ถ้า 1 แทนคะแนนที่ตอบถูก และ 0 แทนคะแนนที่ตอบผิด การแจกแจงของตัวแปร dichotomus จะเป็นผลจากการแบ่งโค้งการแจกแจงปกติออกเป็น 2 ส่วน คือส่วนที่ได้คะแนนเหนือระดับผ่าน กับส่วนที่ได้คะแนนต่ำกว่าระดับผ่าน

นั่นคือถ้าตัวแปรหนึ่งเป็นตัวแปรแบบต่อเนื่อง และอีกตัวแปรหนึ่งเป็นตัวแปร dichotomus ที่ทำมาจากตัวแปรต่อเนื่อง จะสามารถหาค่าสัมประสิทธิ์สหสัมพันธ์ของตัวแปรทั้งสอง โดยใช้ biserial

สำหรับการแจกแจงปกติ ซึ่งแบ่งส่วนโค้งออกเป็น 2 ส่วน คือ p และ q ค่าความแตกต่างระหว่างค่าเฉลี่ยของคะแนนในส่วนที่เป็นสัดส่วนของ p และค่าเฉลี่ยของการแจกแจงของคะแนนรวมจะเท่ากับ $\frac{y}{p}$ และค่าความแตกต่างระหว่างค่าเฉลี่ยของคะแนนในส่วนที่เป็นสัดส่วนของ q และค่าเฉลี่ยของคะแนนรวมจะเท่ากับ y/q โดยที่ y เป็น ordinate ที่แบ่งจุดระหว่างการแจกแจงออกเป็น 2 ส่วน ซึ่งมีพื้นที่เท่ากับ 1.0

$$\frac{y}{p} + \frac{y}{q} = \frac{yq + yp}{pq} = \frac{y\,(p + q)}{pq}$$

แต่ p + q = พื้นที่ทั้งหมด = 1.0 ดังนั้นสมการ 9 - 1 สามารถเขียนใหม่ได้ดังนี้

$$r_{xy} = b_{yx}\,\frac{S_x}{S_y} = \frac{(M_p - M_q)\,pqS_x}{yS_y} \qquad \text{.(9-6)}$$

เนื่องจากความเบี่ยงเบนมาตรฐานบนแกน $X(S_x)$ เป็น 1.0 ดังนั้น

$$r_{bis} = \frac{M_p - M_q}{S_x} \cdot \frac{pq}{y} \qquad \text{.....(9-7)}$$

เมื่อ y แทน ordinate ที่แบ่งโค้งปกติออกเป็น 2 ส่วน คือ p และ q

$\therefore M_p - M_y = M_p - M_q$ ดังนั้น เพื่อความสะดวกจึงเขียนสมการ 9-7 ในรูปใหม่ ดังนี้

$$r_{bis} \qquad \frac{M_p - M_x}{S_x} \cdot \frac{p}{y} \qquad \text{.(9-8)}$$

จากตาราง 9-2 ถ้าให้คอลัมน์ที่ 2 แทนนักเรียนที่ไม่ผ่านการฝึก และคอลัมน์ที่ 3 แทนนักเรียนที่ผ่านการฝึก

ขั้นแรกจะต้องยอมรับว่าความสามารถในการประสบความสำเร็จในการฝึก มีการแจกแจงเป็นโค้งปกติ ซึ่งสามารถคำนวณหาค่าสหพันธ์ระหว่างความสำเร็จในการฝึกกับคะแนนจากแบบสอบถามความสนใจ โดยใช้ biserial จากสมการ 9-7 หรือ 9-8 ได้

แทนค่าในสมการ 9-7 จะได้ดังนี้

$$r_{bis} = \frac{12.78 \quad 8.26}{3.49} \times 0.6267$$

$$= 0.812$$

สหสัมพันธ์แบบ biserial นี้ เป็นกรณีพิเศษของสหสัมพันธ์ product moment และค่าสหสัมพันธ์แบบ biserial จะไม่ขึ้นอยู่กับขนาดของ p และ q ดังนั้นข้อสอบที่ค่อนข้างยาก เช่น ข้อสอบที่ใช้ในการคัดเลือก หรือข้อสอบที่ค่อนข้างง่าย เช่นข้อสอบ mastery test จึงควรหาอำนาจจำแนกด้วยวิธีนี้ ส่วนค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์สหสัมพันธ์แบบ biserial จะสูงกว่าสัมประสิทธิ์สหสัมพันธ์แบบ product moment ซึ่งสามารถประมาณค่าความคลาดเคลื่อนมาตรฐานของ biserial ได้ดังนี้

$$S_{r_{bis}} = \frac{1}{y} \sqrt{pq/N}$$

ซึ่งจะเห็นได้ว่าความคลาดเคลื่อนมาตรฐานขึ้นอยู่กับความสัมพันธ์ระหว่าง p และ q กับจำนวนคน

ค่าความคลาดเคลื่อนมาตรฐานจะน้อยที่สุด เมื่อ p = q = 0.50

จากสมการ 9-2 และ 9-7 สามารถเขียนความสัมพันธ์ระหว่าง r_{pbis} กับ r_{bis} ได้ดังนี้

$$\frac{r_{bis}}{r_{pbis}} = \frac{\sqrt{pq}}{y} \qquad \text{.....(9-9)}$$

หรือ $r_{pbis} = r_{bis} \cdot \frac{y}{\sqrt{pq}}$ ..... (9-10)

จากตาราง Q-2 ถ้า p = 0.50 จะได้ว่า

$r_{pbis} = 0.648$ และ $r_{bis} = 0.812$

ดังนั้น $r_{pbis} = 0.8\ r_{bis}$ .....(9-11)

จะเห็นได้ว่าค่าของ r_{pbis} ต่ำกว่าค่าของ r_{bis}

9.2 การหาค่าอำนาจจำแนกและความยากง่ายโดยใช้สูตร

ค่า biserial correlation และ point biserial correlation ที่กล่าวมานี้ ก็คือค่าอำนาจจำแนกของข้อสอบ แต่ยังมีวิธีในการหาค่าอำนาจจำแนกของข้อสอบอีกหลายวิธี เช่น การหาค่าอำนาจจำแนกอย่างง่ายโดยใช้สูตร

$$r = \frac{R_U - R_L}{T/2} \qquad \text{.....(9-12)}$$

เมื่อ r แทน ค่าอำนาจจำแนก

R_U แทน จำนวนนักเรียนที่ตอบถูกในกลุ่มสูง

R_L แทน จำนวนนักเรียนที่ตอบถูกในกลุ่มต่ำ

T แทน จำนวนนักเรียนทั้งหมดที่นำมาวิเคราะห์

ตัวอย่าง มีข้อสอบข้อหนึ่ง นำไปทดสอบนักเรียน 30 คน เมื่อนำผลการสอบมาวิเคราะห์แล้วปรากฏผลดังนี้

1. คนเก่งตอบถูกทุกคน คนอ่อนตอบผิดทุกคน ข้อสอบข้อนี้จะมีค่าอำนาจจำแนกสูงสุด หรือมีค่า $r = 1.00$

$$r = \frac{15 - 0}{15} = 1.00$$

2. คนเก่งและคนอ่อนตอบถูกทุกคน ข้อสอบข้อนี้จะไม่มีค่าอำนาจจำแนกเลย หรือมีค่า $r = 0$

$$r = \frac{15 - 15}{15} = 0$$

3. คนเก่งและคนอ่อนตอบผิดทุกคน ข้อสอบนี้ก็จะไม่มีค่าอำนาจจำแนกเช่นกัน หรือมีค่า $r = 0$

$$r = \frac{0 - 0}{15} = 0$$

4. คนเก่งทุกคนตอบผิด ส่วนคนอ่อนทุกคนตอบถูก ข้อสอบข้อนี้มีอำนาจจำแนกสูงทางลบ ซึ่งเป็นข้อสอบที่ครูไม่ต้องการ หรือมีค่า $r = -1.00$

$$r = \frac{0 - 15}{15} = -1.00$$

5. คนเก่งตอบถูกบ้าง คนอ่อนตอบถูกบ้าง โดยที่คนเก่งตอบถูกมากกว่าคนอ่อน เช่น คนเก่งตอบถูก 10 คน คนอ่อนตอบถูก 3 คน

$$r = \frac{10 - 3}{15} = \frac{7}{15} = .47$$

ข้อสอบข้อนี้ถือว่ามีอำนาจจำแนกปานกลาง

6. คนเก่งตอบถูกบ้าง คนอ่อนตอบถูกบ้าง โดยที่คนอ่อนตอบถูกมากกว่าคนเก่ง เช่น คนเก่งตอบถูก 3 คน คนอ่อนตอบถูก 10 คน

$$r = \frac{3 - 10}{15} = \frac{-7}{15} = -.47$$

ข้อสอบข้อนี้มีอำนาจจำแนกปานกลางทางลบ ซึ่งเป็นข้อสอบที่ครูไม่ต้องการ

โดยทั่ว ๆ ไป ครูมักจะต้องการข้อสอบที่มีค่าอำนาจจำแนกสูง ๆ ยิ่งสูงเท่าไรยิ่งดี ข้อสอบที่ถือว่ามีอำนาจจำแนกสูงคือข้อสอบที่มีค่า r ตั้งแต่ .20 ขึ้นไป

ไพศาล หวังพานิช ได้เสนอแนะวิธีการสร้างคำถามให้มีอำนาจจำแนกสูงดังนี้

1. คำถามต้องมีความชัดเจน หรือมีความเป็นปรนัยสูง

2. คำถามนั้นต้องถามคุณลักษณะที่ใช้ความคิด หรือต้องใช้ดุลพินิจในการตอบ ไม่ถามผิวเผินเพียงข้อเท็จจริงธรรมดา

3. เป็นข้อคำถามที่เป็นไปตามหลักวิชา หรือมีความสมเหตุสมผล

4. กรณีที่เป็นข้อสอบแบบเลือกตอบ ตัวเลือกที่ใช้ต้องมีความเป็นเอกพันธุ์

ความยากง่ายของข้อสอบ (Item Difficulty) ค่าความยากของข้อสอบ คือสัดส่วนที่คนทำข้อสอบข้อนั้นถูก กับจำนวนคนทั้งหมดที่เข้าสอบ โดยทั่วไปเขียนแทนค่าความยากง่ายด้วย p และ p นี้สามารถคำนวณได้จากสูตร

$$p = \frac{R}{T} \qquad \text{.....(9-13)}$$

เมื่อ p แทน ความยากง่ายของข้อสอบ

R แทน จำนวนนักเรียนที่ตอบถูก

T แทน จำนวนนักเรียนทั้งหมดที่นำมาวิเคราะห์

ตัวอย่าง ข้อสอบข้อหนึ่งมีคนทำถูก 25 คน ในจำนวนคน 100 คน

$$p = \frac{25}{100} = .25$$

ถ้า p น้อย แสดงว่าข้อสอบนั้นยาก เช่น p = 0 แสดงว่าข้อสอบนั้นไม่มีคนทำถูกเลย ถ้า p = 1 แสดงว่าข้อสอบง่าย นักเรียนทุกคนทำข้อสอบนั้นถูกหมด ถ้า p = .50 แสดงว่านักเรียนในกลุ่มนั้นทำถูกครึ่งหนึ่งทำผิดครึ่งหนึ่ง

ในการวัดผลทั่ว ๆ ไป นิยมเลือกใช้ข้อสอบที่มีความยากง่ายปานกลาง กล่าวคือไม่ยากเกินระดับความสามารถของนักเรียน จนนักเรียนส่วนใหญ่ทำไม่ได้ ขณะเดียวกันก็จะต้องไม่ง่ายเกินไปจนนักเรียนส่วนใหญ่ตอบถูกหมด ข้อสอบที่ง่ายหรือยากเกินไปจะมีประโยชน์ในการสอบวัดน้อยมาก ข้อสอบที่มีความยากง่ายปานกลาง คือข้อสอบที่มีค่า p อยู่ระหว่าง .20 - .80

9.3 การวิเคราะห์ข้อสอบแบบอิงเกณฑ์

เตือนใจ เกตุษา ได้กล่าวว่าเหตุผลสำคัญประการหนึ่งของการสร้างข้อสอบแบบอิงเกณฑ์ คือ ความต้องการที่จะกำหนดเป้าหมายของการสอนให้กระจ่างชัดที่สุด จนกระทั่งผู้เรียนส่วนใหญ่สัมฤทธิ์ผลในการเรียนสอดคล้องกับพฤติกรรมที่กำหนดไว้ หากการสอนมีประสิทธิภาพแล้วผู้เรียนทุกคนควรสัมฤทธิ์ผลตามเกณฑ์ที่กำหนดไว้ ซึ่งจะเป็นสาเหตุให้ทุกคนสอบได้คะแนนเต็ม จึงทำให้ไม่มีความแปรปรวนของคะแนน ดังนั้นการหาอำนาจจำแนกของข้อสอบโดยวิธีสหสัมพันธ์ (correlation) จึงไม่อาจทำได้

ในการประเมินผลแบบอิงเกณฑ์นั้นโดยทฤษฎีแล้วมีหลักอยู่ว่าก่อนที่เด็กจะเริ่มเรียนบทใดบทหนึ่ง เด็กควรจะไม่มีความรู้ในเรื่องนั้น (กล่าวคือถ้าเด็กรู้แล้วครูก็ไม่จำเป็นต้องสอนอีกเพราะจะทำให้เสียเวลาทั้งครูและเด็ก) และหลังจากที่ครูสอนจบแล้วนักเรียนควรจะต้องมีความรู้ถึงเกณฑ์ที่กำหนดไว้ (หากการสอนมีประสิทธิภาพ) หรือรู้เรื่องที่ครูสอนเป็นอย่างดี จะเห็นได้ว่าตามทฤษฎีแล้วการประเมินผลแบบอิงเกณฑ์มุ่งที่จะให้นักเรียนมีความรู้เท่าเทียมกัน ซึ่งถ้าเป็นไปตามนี้แล้วคะแนนที่ได้จากการสอบหลังจากที่เรียนจบบทเรียนแล้วจะมีความแปรปรวนน้อย หรือไม่มีความแปรปรวนเลย ดังนั้นการวิเคราะห์ข้อสอบโดยวิธีการตามที่เคยวิเคราะห์กันในการวัดผลแบบอิงกลุ่มคงจะใช้ไม่ได้ นักวัดผลทั้งหลายต่างก็พยายามเสนอแนะวิธีการวิเคราะห์ข้อสอบแบบอิงเกณฑ์ไว้ต่าง ๆ กัน ซึ่งยังไม่สามารถหาข้อสรุปได้ว่าวิธีการของผู้ใดดีที่สุด ในที่นี้ผู้เขียนขอเสนอแนะวิธีการวิเคราะห์ข้อสอบอิงเกณฑ์ง่าย ๆ ซึ่งสามารถนำไปใช้ได้ดังนี้

1. A priori approaches **หรือ** logical review

วิธีการนี้จะให้ผู้เชี่ยวชาญที่มีความชำนาญในเนื้อหาวิชานั้น ๆ หลาย ๆ คนช่วยกัน

พิจารณา ข้อสอบแต่ละข้อนั้นวัดได้ตรงตามจุดประสงค์ที่ต้องการวัดหรือไม่ โดยให้ผู้เชี่ยวชาญเหล่านั้นช่วยกันพิจารณาให้คะแนนโดยถือเกณฑ์ดังนี้

ถ้ารู้สึกแน่ใจว่าข้อสอบนั้นวัดจุดประสงค์ที่ตั้งไว้จริงก็ให้ 1 คะแนน

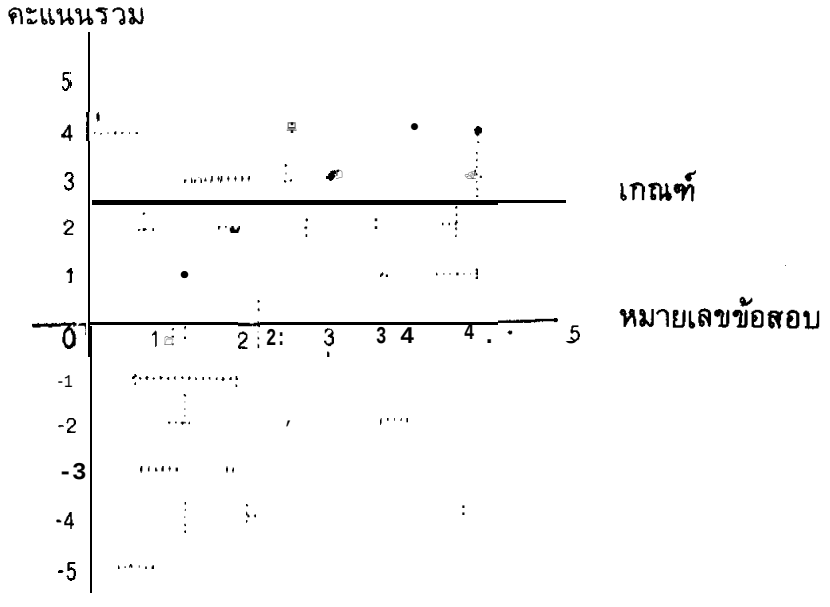
ถ้ารู้สึกไม่แน่ใจว่าข้อสอบนั้นจะวัดจุดประสงค์ที่ตั้งไว้ก็ให้ 0 คะแนน

ถ้ารู้สึกแน่ใจว่าข้อสอบนั้นไม่ได้วัดจุดประสงค์ที่ตั้งก็ให้ -1 คะแนน

เสร็จแล้วนำคะแนนจากผู้เชี่ยวชาญทั้งหมดมารวมคะแนนสำหรับข้อสอบแต่ละข้อ ตัวเลขคะแนนที่ได้นี้เรียกว่า ดัชนีความสอดคล้องกันระหว่างข้อสอบแต่ละข้อกับจุดประสงค์ (Index of item objective congruence)

จากนั้นก็นำคะแนนรวมของข้อสอบแต่ละข้อไปเทียบกับเกณฑ์ โดยเกณฑ์นี้อาจกำหนดไว้ว่า ต้องเกินครึ่งหนึ่งของคะแนนเต็มในแต่ละข้อ จึงจะถือว่าข้อสอบข้อนั้นวัดจุดประสงค์นั้นจริง ในทางตรงกันข้ามหากผลรวมของคะแนนไม่ถึงครึ่งหนึ่งของคะแนนเต็ม ถือว่าข้อสอบข้อนั้นไม่ได้วัดจุดประสงค์นั้นจริง ดังตัวอย่าง

มีข้อสอบที่ต้องการใช้วัดจุดประสงค์หนึ่ง 4 ข้อ และให้ผู้ชำนาญ 5 คน เป็นผู้พิจารณา ถ้าคะแนนรวมของแต่ละข้อเกินกว่า 2.5 ถือว่าข้อสอบข้อนั้นวัดจุดประสงค์นั้นจริง ซึ่งผลรวมของคะแนนแต่ละข้ออาจเสนอเป็นรูปกราฟได้ดังนี้



รูป 9-2 กราฟแสดงผลรวมของคะแนนแต่ละข้อ

จากกราฟนี้แสดงว่าข้อที่ 3 และที่ 4 เท่านั้นที่ผู้เชี่ยวชาญในเนื้อหามีความเห็นตรงกันว่าสามารถวัดจุดประสงค์ที่ต้องการ ส่วนข้ออื่น ๆ ถือว่าไม่สามารถวัดจุดประสงค์นั้นจริง

2. A Posteriori Approaches **หรือ** Empirical review

เป็นวิธีการพิจารณาผลการตอบข้อสอบของกลุ่มตัวอย่าง ซึ่งมีหลายวิธีด้วยกันดังนี้

2.1 Item - response chart

วิธีการนี้คือการนำข้อสอบไปสอบกับนักเรียน 2 ครั้ง โดยนำไปสอบครั้งแรกก่อนที่จะมีการเรียนการสอนและเมื่อสอนจบบทเรียนแล้วก็นำข้อสอบฉบับเดิมไปสอบกับนักเรียนกลุ่มเดิมอีก แล้วนำผลการสอบของเด็กแต่ละคนมาบันทึกลงในตาราง ดังตัวอย่าง

ตาราง 9-3 แสดงผลการสอบของนักเรียน 6 คน ในข้อสอบ 5 ข้อ ซึ่งผลการสอบทั้งก่อนสอนและหลังสอน เป็นดังนี้

ข้อสอบข้อที่	1		2		3		4		5	
	ก่อน	หลัง	ก่อน	หลัง	ก่อน	หลัง	ก่อน	หลัง	ก่อน	หลัง
สมบัติ	–	+	+	+	–	–	+	–	–	+
สาธิต	–	+	+	+	–	–	+	–	+	+
มงคล	–	+	+	+	–	–	+	–	–	+
กิตติ	–	+	+	+	–	–	+	–	–	+
กานดา	–	+	+	+	–	–	+	–	+	+
ปราณี	–	+	+	+	–	–	+	–	–	–

\+ แทน ตอบถูก

– แทน ตอบผิด

ข้อมูลจากตาราง 9-3 สามารถแปลความหมายได้ดังนี้

ข้อ 1 ก่อนสอนไม่มีใครทำถูกเลย แต่หลังจากสอนแล้วนักเรียนทำข้อสอบถูกหมด ข้อสอบข้อนี้จัดว่าเป็นข้อสอบที่ดีมาก เป็นข้อสอบในอุดมคติ ซึ่งคงจะหาได้ไม่ง่ายนัก

ข้อ 2 ก่อนสอนนักเรียนทำถูกหมด และหลังจากสอนนักเรียนก็ยังคงทำถูกเหมือนเดิม ข้อสอบข้อนี้จัดว่าเป็นข้อสอบที่ง่าย ซึ่งที่จริงแล้ว ถ้าก่อนสอนนักเรียนทำถูกหมดแล้วครูก็ไม่จำเป็นต้องสอน เพราะทำให้เสียเวลาทั้งครูและเด็ก

ข้อ 3 ทั้งก่อนสอนและหลังสอนไม่มีใครทำข้อสอบข้อนี้ถูกเลย แสดงว่าข้อสอบยากมาก หรือครูตั้งจุดประสงค์ในการสอนสูงเกินไป ทำให้นักเรียนไปไม่ถึงจุดประสงค์ที่ตั้งไว้ ในลักษณะเช่นนี้ครูจำเป็นต้องแก้ไขจุดประสงค์ในการสอนเสียใหม่

ข้อ 4 ข้อสอบข้อนี้ค่อนข้างแปลก คือก่อนสอนนักเรียนทุกคนทำถูกหมด แต่หลังจากสอนไปแล้วนักเรียนทุกคนกลับตอบผิดหมด แสดงว่าข้อสอบอาจกำกวม หรืออาจเป็นเพราะการสอนไม่ถูกต้อง ข้อสอบข้อนี้จึงไม่ควรใช้

ข้อ 5 ข้อสอบข้อนี้ก่อนสอนมีนักเรียนบางคนทำถูกบ้าง ทำผิดบ้าง และหลังการสอนแล้วมีนักเรียนที่ทำข้อสอบผิดกลับทำถูก ส่วนคนที่ทำถูกก่อนสอนก็คงยังทำถูกเหมือนเดิม ข้อสอบข้อนี้จัดว่าเป็นข้อสอบที่ดีข้อหนึ่ง และมักจะพบบ่อย ๆ ในทางปฏิบัติ

2.2 Sensitivity Index

Cox ได้เสนอแนะวิธีวัดความไวของข้อสอบไว้ดังนี้

$$S = P_{post} - P_{pre} \qquad (9\text{-}14)$$

เมื่อ S แทน ความไวของข้อสอบ

P_{post} แทน สัดส่วนที่ตอบถูกหลังสอน

P_{pre} แทน สัดส่วนที่ตอบถูกก่อนสอน

ตัวอย่าง จงหาค่า sensitivity Index จากตารางต่อไปนี้

ตาราง 9-4 แสดงค่าความยากง่ายของข้อสอบจำนวน 7 ข้อ

ข้อสอบข้อที่	P_{post}	P_{pre}	$S = P_{post} - P_{pre}$
1	1.00	.00	1.00
2	.00	1.00	-1.00
3	1.00	.50	.50
4	.00	.00	.00
5	.60	.05	.55
6	.90	.10	.80
7	1.00	1.00	.00

ข้อ 1 เป็นข้อสอบในอุดมคติ คือก่อนสอนไม่มีนักเรียนตอบถูกเลย P_{pre} = .00 แต่หลังจากสอนไปแล้วนักเรียนทุกคนตอบถูกหมด P_{post} = 1.00 จัดว่าเป็นข้อสอบที่ดี เพราะมีความไวในการวัดผลมากสมควรเก็บไว้ใช้ต่อไป

ข้อ 2 เป็นข้อสอบที่ไม่ดี สมควรตัดทิ้ง ทั้งนี้เนื่องจากก่อนสอนนักเรียนทุกคนตอบถูกหมด แต่หลังจากสอนแล้วนักเรียนกลับตอบผิดหมด ที่เป็นเช่นนี้อาจเป็นเพราะคำถามกำกวมหรือการสอนของครูไม่ถูกต้อง

ข้อ 3 จัดว่าเป็นข้อสอบที่พอใช้ได้ แต่อาจต้องปรับปรุงจุดประสงค์ที่จะวัดให้ซับซ้อนขึ้น

ข้อ 4 และ 7 เป็นข้อสอบที่ไม่ดีทั้งคู่ เพราะ S มีค่าเท่ากับ .00

ข้อ 5 เป็นข้อสอบที่ไม่ดี ควรตั้งทิ้ง ทั้งนี้เพราะหลังจากสอนแล้ว มีนักเรียนเพียง 60% เท่านั้นที่ตอบข้อสอบข้อนี้ถูก

ข้อ 6 จัดว่าเป็นข้อสอบที่ดี มีค่า S = .80 ก่อนสอนมีนักเรียนตอบถูกเพียง 10% แต่หลังจากสอนมีนักเรียนตอบถูก 90%

เตือนใจ เกตุษา ได้เสนอแนะวิธีพิจารณาค่า Sensitivity index ไว้ดังนี้

1. ในการพิจารณาค่าความยากง่ายของข้อสอบก่อนสอน มีหลักดังนี้

P_{pre} .41 ขึ้นไป	เป็นข้อสอบที่ง่ายเกินไป ควรตัดทิ้ง
P_{pre} .21-.40	เป็นข้อสอบที่พอใช้ได้ แต่ต้องปรับปรุงจุดประสงค์เชิงพฤติกรรมให้ซับซ้อนขึ้น
P_{pre} ต่ำกว่า .21	เป็นข้อสอบที่ดี

2. ในการพิจารณาค่าความยากง่ายของข้อสอบหลังสอน มีหลักดังนี้

P_{post} .80 ขึ้นไป	เป็นข้อสอบที่ดี
P_{post} .70-.79	เป็นข้อสอบที่พอใช้ได้ แต่ครูอาจต้องปรับปรุงตัวคำถามหรือตัวเลือกบ้างเล็กน้อย
P_{post} ต่ำกว่า .69	เป็นข้อสอบที่ไม่ดี ควรตัดทิ้ง

3. ในการพิจารณาค่าความไวของข้อสอบ มีหลักดังนี้

S เป็น 1.00	เป็นข้อสอบที่ดี เป็นไปตามทฤษฎี
S .80-.99	เป็นข้อสอบที่ดี หาได้ในเชิงปฏิบัติ
S .30-.79	เป็นข้อสอบที่พอใช้ได้

S .00-.29 เป็นข้อสอบที่ไม่ดี ควรตัดทิ้ง

S -1.00 ถึง .00 เป็นข้อสอบที่ไม่ดีอย่างยิ่ง ต้องตัดทิ้ง

2.3 Phi correlation

ในการหาค่าที่เป็นตัวแสดงคุณภาพของข้อสอบอีกวิธีหนึ่ง คือ การหาค่าสัมประสิทธิ์สหสัมพันธ์ไฟระหว่างการสอบก่อนสอน และการสอบหลังสอน กับการทำข้อสอบข้อนั้นได้ (ผ่าน) และไม่ได้ (ไม่ผ่าน) ซึ่งผลการตอบข้อสอบแต่ละข้อสามารถสรุปเป็นตารางได้ดังนี้

ผลการสอบ / ชนิดการสอบ	ผ่าน	ไม่ผ่าน	รวม
ก่อนสอน	A	B	A + B
หลังสอน	C	D	C + D
รวม	A + C	B + D	

เมื่อ A คือ จำนวนผู้ผ่านจุดประสงค์ก่อนสอน

B คือ จำนวนผู้ไม่ผ่านจุดประสงค์ก่อนสอน

C คือ จำนวนผู้ผ่านจุดประสงค์หลังสอน

D คือ จำนวนผู้ไม่ผ่านจุดประสงค์หลังสอน

จากนั้นก็นำค่าตัวเลขจากตารางไปคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบไฟ โดยใช้สูตร ดังนี้

$$\phi = \frac{BC - AD}{\sqrt{(A + B)(C + D)(A + C)(B + D)}} \qquad \text{.....(9-15)}$$

ตัวอย่าง นำแบบทดสอบไปสอบกับนักเรียน 5 คน ปรากฏว่าก่อนสอนมีผู้ตอบข้อสอบข้อ 10 ผ่าน 2 คน และไม่ผ่าน 3 คน และหลังจากสอนแล้ว มีผู้ตอบข้อสอบข้อ 10 ผ่าน 4 คน ไม่ผ่าน 1 คน ดังแสดงในตาราง

ตัวอย่างผลการสอบข้อที่ 10 ของนักเรียน 5 คน ก่อนและหลังการสอน

ชื่อผู้สอบ	ข้อที่ 10	
	ก่อนสอน	หลังสอน
นายสมศักดิ์	0	1
นายวิโรจน์	1	0
นายสมชัย	0	1
น.ส.วรรณา	0	1
น.ส.นภา	1	1
จำนวนผู้สอบผ่าน	2	4
จำนวนผู้สอบไม่ผ่าน	3	1

หรือสรุปเป็นตารางได้ ดังนี้

	ผ่าน	ไม่ผ่าน	รวม
ก่อนสอน	2	3	5
หลังสอน	4	1	5
รวม	6	4	

จากตาราง

A = 2 A + B = 5

B = 3 C + D = 5

c = 4 A + C = 6

D = **I** **B+D** = **4**

$$\text{ดังนั้น } \varnothing = \frac{BC - AD}{\sqrt{(A + B)(C + D)(A + C)(B + D)}}$$

$$= \frac{(3)(4) \quad (2)(1)}{\sqrt{(5)\ (5)\ (6)\ (4)}}$$

$$= .40$$

นั่นคือข้อสอบข้อ 10 นี้วัดการเปลี่ยนแปลงของผู้เรียนได้พอสมควร

$\varnothing$ มีค่าอยู่ระหว่าง -1.00 ถึง 1.00 เช่นเดียวกับค่า S (ดัชนีความไว)

สรุปเนื้อหาบทที่ 9

1. การวิเคราะห์ข้อสอบเป็นเทคนิคในการตรวจสอบคุณภาพของข้อสอบเป็นรายข้อ โดยมีจุดมุ่งหมายที่จะทราบคุณลักษณะที่สำคัญของแบบทดสอบ 2 ประการคือ ค่าความยากง่าย และค่าอำนาจจำแนก
2. ในการหาอำนาจจำแนกของข้อสอบนั้นอาจหาได้โดยหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนในข้อนั้นกับคะแนนรวม
3. การวิเคราะห์ข้อสอบแบบอิงเกณฑ์นั้นจะมีวิธีการวิเคราะห์แตกต่างไปจากวิธีวิเคราะห์แบบอิงกลุ่ม ทั้งนี้เพราะข้อสอบแบบอิงเกณฑ์นั้นคะแนนจะมีความแปรปรวนน้อย

คำถามท้ายบทที่ 9

1. การวิเคราะห์ข้อสอบคืออะไร การวิเคราะห์ข้อสอบมีประโยชน์อย่างไร
2. ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากข้อสอบข้อใดข้อหนึ่งกับคะแนนรวมจากแบบทดสอบทั้งฉบับ คือค่าอะไรของข้อสอบ
3. ในการวิเคราะห์ข้อสอบแบบอิงเกณฑ์จะใช้วิธีวิเคราะห์เช่นเดียวกับการวิเคราะห์ข้อสอบแบบอิงกลุ่มได้หรือไม่ เพราะเหตุใด