บทที่ 2

วิธีวัดผลทางการศึกษา

เมื่อครูให้คะแนนแก่คำตอบนักเรียน มักจะมีปัญหาเกี่ยวกับการพิจารณาความหมายของคะแนนที่ให้ และถ้านักบริหารจะนำคะแนนเหล่านั้นไปใช้เพื่อตัดสินใจในปัญหาบางอย่าง ก็จำเป็นจะต้องทราบว่า คะแนน (Scores) ที่ให้เหล่านั้นอยู่ในมาตราการวัด (Scale) ระดับใด และใช้วิธีการทางสถิติอะไรมาอธิบายเกี่ยวกับข้อเท็จจริงของข้อมูลเหล่านั้น อาทิเช่น การหาแนวโน้ม หรือ ตำแหน่งที่การวัดการกระจาย หรือ การวัดความสัมพันธ์ของข้อมูลเหล่านั้น ดังนั้นการที่จะตีความหมายของคะแนนต่าง ๆ ได้ถูกต้องตามข้อเท็จจริง จึงจำเป็นต้องทราบถึงระดับคะแนนเหล่านั้นเสียก่อนว่าวัดมาจากมาตราอะไร และนำวิธีการทางสถิติอะไรมาใช้

นอกจากนั้น จะกล่าวถึงการวิเคราะห์แบบทดสอบ แบบทดสอบนั้นจะรู้ว่าดีหรือไม่ จะต้องนำมาวิเคราะห์เสียก่อน การวิเคราะห์แบบทดสอบมีเทคนิคหลายวิธี โดยจะกล่าวถึง การวิวิเคราะห์เป็น 3 ประเภท คือ

(1) การวิเคราะห์ความเชื่อมั่นของแบบทดสอบ (Reliability)

(2) การวิเคราะห์ความเที่ยงตรงของแบบทดสอบ (Validity)

(3) การวิเคราะห์ข้อสอบเป็นรายข้อ (Item Analysis)

2.1 การวัดแนวโน้มเข้าสู่ส่วนกลาง (Measures of Central Tendency)

บางครั้งการนำคะแนนผลการทดสอบมาหาการแจกแจงความถี่ แล้วก็ยังให้ข้อเท็จจริงเกี่ยวกับคะแนนชุดนั้นไม่เพียงพอ จึงจำเป็นต้องใช้วิธีการทางสถิติอื่น ๆ มาเพิ่มอีก เพื่อที่จะช่วยให้ทราบข้อเท็จจริงของคะแนนชุดนั้นได้มากขึ้น สถิติดังกล่าวได้แก่ การวัดแนวโน้มเข้าสู่ส่วนกลาง หรือการวัดตำแหน่งที่ (Measure of Location) ของข้อมูลชุดนั้น ซึ่งได้แก่ การหาค่าเฉลี่ย (Mean) การหาค่ามัธยฐาน (Median) และการหาค่าฐานนิยม (Mode)

2.1.1 การหาค่าเฉลี่ย (Mean : $\overline{X}$)

ค่าเฉลี่ยเป็นค่าที่ได้จากผลบวกของคะแนนทั้งหมดในการแจกแจง หารด้วยจำนวนคะแนนหรือจำนวนนักเรียนที่เข้าสอบทั้งหมด ซึ่งเขียนเป็นสูตรดังนี้

$$\bar{X} = \frac{\Sigma X}{N} \qquad (3)$$

เมื่อ $\bar{X}$ = ค่าเฉลี่ย

X = คะแนนของนักเรียนแต่ละคน

N = จำนวนนักเรียนที่เข้าสอบ

Σ = การรวมกันของคะแนนทั้งหมด

ตัวอย่างที่ 2.1 หาค่าเฉลี่ยของคะแนน 5, 8, 10, 12, 15

$$\bar{X} = \frac{\Sigma X}{N}$$

$$= \frac{5+8+10+12+15}{5}$$

$$= 10$$

2.1.2 ค่ามัธยฐาน (Median : Mdn)

ค่ามัธยฐาน คือ จุดที่แบ่งครึ่งของคะแนนชุดหนึ่งซึ่งเรียงจากมากไปหาน้อย หรือจุดที่มีคะแนนสูงกว่า 50 เปอร์เซนต์ และมีคะแนนต่ำกว่า 50 เปอร์เซนต์

ตัวอย่างที่ 2.2 ค่ามัธยฐานของข้อมูลชุดหนึ่งที่ประกอบด้วย 2,4,5,9,11 และ 13

ค่ามัธยฐานจะเท่ากับ $\frac{5+9}{2} = \frac{14}{2} = 7$ เป็นต้น

2.1.3 ค่าฐานนิยม (Mode : Mo)

ฐานนิยม คือ ค่าของคะแนนหรือข้อมูลที่มีความถี่มากที่สุด

ตัวอย่างที่ 2.3 ข้อมูลชุดหนึ่งมีการแจกแจงดังนี้ 14,16,16,17,18,19,19, 19,21,22 ค่าฐานนิยมของข้อมูลชุดนี้เท่ากับ 19 เพราะว่า 19 มีความถี่มากที่สุด บางครั้งเราอาจพบการแจกแจงของข้อมูล ที่มีฐานนิยมสองฐานหรือเรียกว่า Bimodal

ตัวอย่างที่ 2.4 ถ้ามีข้อมูลที่มีการแจกแจงดังนี้ 14,16,16,16,18,19,19,19,21,22, ค่าฐานนิยมทั้งสองของข้อมูลชุดนี้ คือ 16 และ 19

ใช้ฐานนิยมในกรณีที่การแจกแจงของข้อมูลไม่เป็นแบบปกติและไม่มีความคงที่ (Unstable)

การวิเคราะห์ในทางสถิติมักจะใช้ค่าเฉลี่ยในการวัดแนวโน้มเข้าสู่ส่วนกลางของข้อมูล เพราะว่า ค่าเฉลี่ยนี้ ได้จากการรวบรวมของคะแนนทุกคะแนน หารด้วยจำนวนคะแนนทั้งหมด บางครั้งคะแนนการทดสอบชุดหนึ่งมีนักเรียนบางคนหรือนักเรียนจำนวนน้อยสอบได้คะแนนสูง หรือบางทีก็มีนักเรียนบางคนสอบได้คะแนนต่ำมากปนอยู่ด้วย ในการวัดที่จะจัดแนวโน้มเข้าสู่ส่วนกลางของคะแนนชุดนี้ นิยมใช้ค่ามัธยฐานแทนคะแนนชุดนี้ เพราะจะเป็นตัวแทนของคะแนนชุดนี้ได้ดีกว่าการใช้ค่าเฉลี่ย ส่วนค่าฐานนิยมนั้นใช้ในกรณีที่การแจกแจงเบี้ยวไปทางบวกหรือเบี้ยวไปทางลบ และต้องการจะทราบเฉพาะตำแหน่งข้อมูลที่ปรากฏบ่อยที่สุดเท่านั้น อย่างไรก็ดี ถ้าจะให้ทราบแนวโน้มเข้าสู่ส่วนกลางที่แท้จริงของข้อมูล ควรจะนำค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยมมาพิจารณารวมกัน อนึ่ง ถ้าการแจกแจงของคะแนนเป็นแบบปกติ (Normal Distribution) เส้นโค้งทั้งสองข้างจะมีลักษณะสมมาตร (Symmetry) ในกรณีนี้ จะได้ค่าเฉลี่ย ค่ามัธยฐาน และค่าฐานนิยม อยู่ในตำแหน่งเดียวกันหมด หรือมีค่าเท่ากัน

2.2 <u>การวัดการกระจาย</u> (Measures of Variability)

การที่เราทราบคะแนนของนักเรียนแต่ละคน ให้ความหมายแก่เราน้อยมาก ถ้าทราบว่าคะแนนนั้นอยู่สูงกว่าหรืออยู่ต่ำกว่าค่าเฉลี่ยเท่าใด ก็ให้ความหมายมากขึ้น แต่ถ้าหากทราบการกระจายของคะแนนชุดนั้นด้วย ก็จะทำให้ทราบความหมายของคะแนนนั้นมากยิ่งขึ้น ตัวอย่างเช่น การที่คะแนนสองชุดมีค่าเฉลี่ยเท่ากัน แต่อาจมีการกระจายแตกต่างกันก็ได้ เช่น

คะแนนชุดที่ 1 ประกอบด้วย 3,5,8,4 และ 10 ซึ่งมีค่าเฉลี่ยเท่ากับ 6

คะแนนชุดที่ 2 ประกอบด้วย 1,2,4,11 และ 12 ซึ่งมีค่าเฉลี่ยเท่ากับ 6 เหมือนกัน แต่คะแนนชุดที่ 2 มีการกระจายมากกว่า การกระจายของคะแนนชุดที่ 1

โดยทั่วไปเราวัดการกระจายของคะแนนด้วย พิสัย (Range) ความแปรปรวน (Variance) และความเบี่ยงเบนมาตรฐาน (Standard Deviation)

2.2.1 <u>พิสัย</u> (Range)

มีค่าเท่ากับผลต่างระหว่างคะแนนสูงสุดกับคะแนนต่ำสุดของข้อมูลชุดนั้น เช่น พิสัยของคะแนนชุดที่ 1 ข้างต้น จะมีค่าเท่ากับ 10-3 = 7

ส่วนพิสัยของคะแนนชุดที่ 2 จะมีค่าเท่ากัย 12-1 = 11 เป็นต้น

2.2.2 ความแปรปรวน (Variance : V)

เป็นส่วนที่เบี่ยงเบนเฉลี่ยจาก Mean ยกกำลังสอง ซึ่งคำนวณได้จากสูตร

$$V = \frac{\Sigma(X-\bar{X})^2}{N-1}$$

เมื่อ V = ค่าความแปรปรวนของคะแนน

X = คะแนนผลการทดสอบ

$\bar{X}$ = ค่าเฉลี่ย

N = จำนวนนักเรียนที่เข้าสอบ

จากสมการ (1) เราอาจทำให้อยู่ในเทอมที่ง่ายต่อการคำนวณได้ดังนี้

$$V = \frac{(\Sigma_{i=1}^{N} X_i^2) - N\bar{X}^2}{N-1} \quad (3)$$

2.2.3 ความเบี่ยงเบนมาตรฐาน (Standard Deviation : SD)

หาได้จากการถอดรากที่สอง ของค่าความแปรปรวน ดังนี้

$$SD = \sqrt{V} \quad (3)$$

แต่โดยปกติในทางสังคมศาสตร์มักจะใช้ตัวอย่างที่มีขนาดใหญ่ดังนั้นจะลบจำนวน N ด้วย 1 หรือไม่ ไม่สู้จะมีความสำคัญเท่าใดนัก นักวิชาการทางสถิติได้ให้ข้อสรุปไว้ว่าถ้าจำนวน N มากกว่าหรือเท่ากับ 30 ไม่ต้องใช้สูตรที่หารด้วย N-1 ถ้าตัวอย่างจำนวนน้อย ๆ เช่น N=10 จึงจะต้องใช้สูตรที่หารด้วย N-1

2.3 ความคลาดเคลื่อนมาตรฐาน (Standard Error : SE)

มีการแจกแจงที่เป็นปกติอยู่ประเภทหนึ่งซึ่งได้จากความคลาดเคลื่อนของการวัด

(มิใช่ตัวคะแนนจริงที่ได้จากการวัด) คะแนนจริงที่ได้จากการวัดหรือคะแนนที่ถือว่าเป็นตัวแทนของกลุ่มก็คือค่าเฉลี่ย ค่าอื่น ๆ ที่ต่างจากค่าเฉลี่ย เราถือว่าเป็นตัวแทนของกลุ่มก็คือค่าเฉลี่ย เราถือว่าเป็นค่าความแตกต่างที่เกิดขึ้นจากการวัด และข้อมูลทุกชุดย่อมจะไม่มีค่าเดียวกันเหมือนกันทุกคน ดังนั้นจึงต้องมีคะแนนที่ผิดไปจากค่าเฉลี่ยซึ่งเรียกว่าความแตกต่างหรือค่าที่เบี่ยงเบนไปจากค่าเฉลี่ย คือคะแนนนั้น ๆ อาจจะมากกว่าหรือน้อยกว่าค่าเฉลี่ย ซึ่งค่าที่มากกว่าหรือน้อยกว่านี้ ถ้านำมารวมกันแล้วนำมาเฉลี่ยอีกครั้งหนึ่ง ก็จะได้ค่าเฉลี่ยของความแตกต่างที่เกิดขึ้น ค่าเฉลี่ยของความแตกต่างที่ได้นี้เรียกว่า ค่าความเบี่ยงเบนมาตรฐาน ค่าความเบี่ยงเบนนี้ถ้านำมาหารด้วยกรณ์ที่ 2 ของจำนวนคนทั้งหมด จะเรียกว่า ความคลาดเคลื่อนมาตรฐาน (Standard Error) แทนด้วยสูตร

$$SE = \frac{SD}{\sqrt{N}} \qquad (3)$$

ความคลาดเคลื่อนมาตรฐานนี้มีการแจกแจงเป็นรูปโค้งปกติ ถ้าความคลาดเคลื่อนมีน้อยเท่าใดก็แสดงว่า การวัดจะมีความถูกต้องมากเท่านั้น

2.4 <u>ค่ามากที่สุดและน้อยที่สุด</u> (Maximum & Minimum)

2.4.1 <u>ค่ามากที่สุด</u> (Maximum)

เป็นค่าของข้อมูลที่น้อยที่สุด สำหรับข้อมูลที่กำหนดให้

2.4.2 <u>ค่าน้อยที่สุด</u> (Minimum)

เป็นค่าของข้อมูลน้อยที่สุด สำหรับข้อมูลที่กำหนดให้

<u>ตัวอย่างที่ 2.5</u> ข้อมูลชุดหนึ่งมี 5,10,8,6,7

ค่ามากที่สุด คือ 10

ค่าน้อยที่สุด คือ 5

2.5 <u>ความเบ้</u> (Skewness)

ข้อมูลชุดใดที่มีการแจกแจงสมมาตร เส้นโค้งที่ได้จากการแจกแจงของข้อมูลชุดนั้นจะมีลักษณะเป็นรูประฆังที่สมมาตรกันที่มัชฌิมเลขคณิต เส้นโค้งทางด้านขวาของมัชฌิมเลขคณิตและทางด้านซ้ายของมัชฌิมเลขคณิต จะมีลักษณะเหมือนกันทุกประการ มัชฌิมเลขคณิต มัธยฐานและฐานนิยม จะมีค่าเท่ากันหรือทับกันสนิท

แต่ถ้าข้อมูลที่มีการแจกแจงไม่สมมาตร มีลักษณะเบ้ไปข้างใดข้างหนึ่ง มัชฌิมเลขคณิต มัธยฐาน และ ฐานนิยม ก็จะมีค่าต่างกัน

การแจกแจงความเบ้มี 2 แบบ คือ เบ้ขวา และ เบ้ซ้าย การแจกแจงความเบ้นี้ หากมีการแจกแจงปกติ ค่าเฉลี่ยและการกระจายที่นำไปใช้ย่อมไม่มีปัญหา จะมีปัญหาก็คือการแจกแจงความเบ้ ดังนั้นก่อนจะนำค่าเฉลี่ยและค่าการกระจายไปใช้ ควรแน่ใจเสียก่อนว่า การแจกแจงข้อมูลเดิมมีลักษณะปกติ หรือไม่ หากมี การแจกแจงความเบ้เพียงเล็กน้อยก็พอจะนำไปใช้ได้

สูตรที่ใช้ในการวัดความเบ้คือ

$$\text{Skewness} = \frac{\left\{\left[\sum_{i=1}^{N} X_i^3 - 3\bar{X}\left(\sum_{i=1}^{N} X_i^2\right) + 3\bar{X}^2\left(\sum_{i=1}^{N} X_i\right)\right/N\right\} - \bar{X}^3}{\left\{\left[\left(\sum_{i=1}^{N} X_i^2\right) - N\bar{X}^2\right] \Big/ (N-1)\right\}^{3/2}} \qquad (3)$$

ถ้าค่าที่คำนวณได้	คือ	Sk	=	0	ถือว่าเป็นการแจกแจงปกติ
	ถ้า	Sk	>	0	ถือว่าเบ้ขวา
	ถ้า	Sk	<	0	ถือว่าเบ้ซ้าย

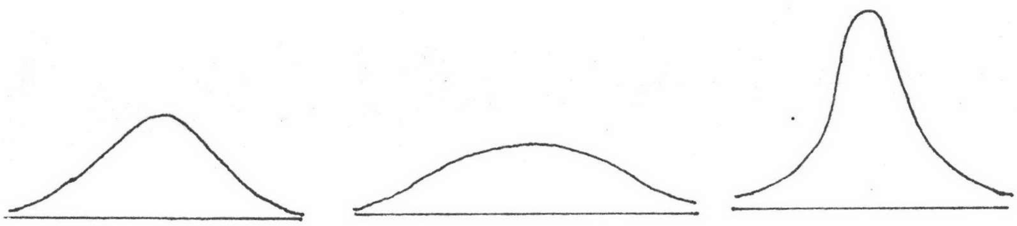
2.6 ความโด่ง (Kurtosis)

การวัดความโด่ง คือการวัดเส้นโค้งว่าจะมีความโด่งมากน้อยเพียงไร เส้นโค้งที่เราเรียกว่าเส้นโค้งปกติ นอกจากเป็นรูประฆังที่มีลักษณะสมมาตร ไม่เบ้แล้วยังต้องเป็นเส้นโค้งที่มีความโด่งตามสัดส่วนอีกด้วย ดังนั้นเส้นโค้งใดที่โด่งผิดปกติจากเส้นโค้งปกติ ก็นับเป็นเส้นโค้งที่ไม่ปกติทั้งสิ้น แม้แต่จะเป็นรูประฆังที่สมมาตรก็ตาม

เส้นโค้งที่มีความโด่งเป็นปกติ ค่าที่คำนวณได้เป็นศูนย์ เรียกว่าเส้นโค้งชนิด Meso Kurtic

เส้นโค้งที่แบนราบกว่าปกติ ค่าที่คำนวณได้เป็นลบ เรียกว่า เส้นโค้งชนิด Platy Kurtic

เส้นโค้งที่โด่งกว่าปกติ ค่าที่คำนวณได้เป็นบวก เรียกว่า เส้นโค้งชนิด Lepto Kurtic



Meso Kurtic Platy Kurtic Lepto Kurtic

มีสูตรที่ใช้คำนวณดังนี้

(3)

$$\text{Kurtosis}= \frac{\left\{\left[\Sigma_{i=1}^{N} X_{i}^{4}-4\bar{X}(\Sigma_{i=1}^{N}X_{i}^{3})+\bar{X}^{2}(\Sigma_{i=1}^{N}X_{i}^{2})-4\bar{X}^{3}(\Sigma_{i=1}^{N}X_{i})\right]/N\right\}+\bar{X}^{4}-3}{\left\{\left[(\Sigma_{i=1}^{N}X_{i}^{2})-N\bar{X}^{2}\right]/(N-1)\right\}^{3/2}}$$

2.7 <u>การวัดความสัมพันธ์</u> (Measure of Relationship)

ถ้ามีคะแนนการทดสอบสองวิชาจากการทดสอบนักเรียนในชั้นหนึ่ง เราอยากทราบว่าคะแนนการทดสอบของวิชานี้มีความสัมพันธ์กันเพียงใด ก็หาสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ระหว่างคะแนนสองชุดนี้ ตัวอย่างเช่น เราอาจหาความสัมพันธ์ระหว่างคะแนนรวมของวิชาฟิสิกส์กับคะแนนรวมของวิชาอื่น ๆ เพื่อที่จะทราบว่าคนที่เก่งวิชาฟิสิกส์จะเก่งวิชาอื่นด้วยหรือไม่ เป็นต้น สูตรที่ใช้คำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์ ส่วนมากใช้สูตรของ Pearson Product Moment Correlation(r) ซึ่งมีสมการดังนี้

(3)

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

เมื่อ

r_{XY} = ค่าสหสัมพันธ์ระหว่างตัวแปร X และตัวแปร Y

X = คะแนนการทดสอบของวิชาหนึ่ง

Y = คะแนนการทดสอบของอีกวิชาหนึ่ง (ของนักเรียนคนเดียวกัน)

XY = ผลคูณของคะแนน X และ Y ของแต่ละคน

X^2, Y^2 = กำลังสองของคะแนน X และคะแนน Y

Σ = ผลบวกของคะแนน

N = จำนวนนักเรียนที่เข้าสอบทั้งหมด

ในการคำนวณมีขั้นตอนดังนี้

ขั้นที่ 1 ยกกำลังสองของคะแนน X และ Y

ขั้นที่ 2 คูณคะแนน X และ Y ของแต่ละคน

ขั้นที่ 3 รวมคะแนน X,Y,XY,X^2 และ Y^2

ขั้นที่ 4 ใช้สูตรดังกล่าวข้างต้น

การทดสอบนัยสำคัญของค่าสัมประสิทธิ์สหสัมพันธ์ ทดสอบโดยใช้ค่า t ดังนี้

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}}$$

โดยใช้ค่า t ที่ N-2 degree of freedom

df 60 ค่า $t \geqslant 2.00$ จะมีนัยสำคัญที่ .05

<u>ตัวอย่างที่ 2.6</u> การคำนวณสัมประสิทธิ์สหสัมพันธ์ (r_{XY}) ระหว่างคะแนนการทดสอบสองวิชา คือ X และ Y

	X	Y	X^2	Y^2	XY
	10	9	100	81	90
	9	10	81	100	90
	8	8	64	64	64
	7	6	49	36	42
	6	7	36	49	42
รวม	40	40	330	330	328

$$r_{XY} = \frac{(5x328) - (40x40)}{\sqrt{(5x330-40x40)\ (5x330-40x40)}}$$

$$= \frac{1640-1600}{1650-1600}$$

$$= \frac{40}{50}$$

$$r_{XY} = 0.80$$

$$t = \frac{0.80\ \sqrt{5-2}}{\sqrt{1-(.80)^2}}$$

$$= \frac{1.385}{.6}$$

$$= 2.308$$

ไม่มีนัยสำคัญ

2.7.1 <u>การตีความค่าสัมประสิทธิ์สหสัมพันธ์</u> (r)

ค่า r นั้นจะมีค่าอยู่ระหว่าง -1.00 กับ + 1.00 ความสัมพันธ์ระหว่างคะแนน 2 กลุ่มนั้น สามารถแยกออกได้ดังนี้

(1) <u>ความสัมพันธ์ทางบวก</u> คือ ค่าของ r เป็นบวก หมายความว่าความสัมพันธ์ที่เกิดขึ้นเป็นความสัมพันธ์ตามกัน เช่น คนหนึ่งสอบได้คะแนนสูงในวิชาแรก ก็จะสอบได้คะแนนสูงในวิชาหลังด้วย ตัวอย่างของความสัมพันธ์ตามกัน เช่น ความเร็วของรถยนต์กับระยะทางที่ได้ เมื่อถือว่าเวลาคงที่ จะเห็นว่า ถ้าความเร็วสูงก็จะได้ระยะทางมากด้วย

(2) <u>ความสัมพันธ์ทางลบ</u> คือ ค่าของ r เป็นลบ หมายความว่าความสัมพันธ์ที่เกิดขึ้นเป็นความสัมพันธ์กลับกัน เช่น นักเรียนคนที่สอบได้คะแนนสูง ในวิชาแรกจะสอบได้คะแนนต่ำในวิชาหลัง ตัวอย่างความสัมพันธ์กลับกัน เช่น ความเร็วของรถยนต์กับเวลาที่ใช้ในการเดินทาง เมื่อถือว่าระยะทางคงที่ จะเห็นได้ว่าถ้าความเร็วมากจะเสียเวลาน้อย แต่ถ้าความเร็วต่ำจะเสียเวลาเดินทางมาก

(3) <u>ไม่มีความสัมพันธ์กัน</u> คือ ค่าของ r เป็นศูนย์ ตัวอย่างของความสัมพันธ์ประเภทนี้ เช่น ความสามาถทางปัญญากับขนาดของรองเท้า คนที่ใช้รองเท้าใหญ่ไม่จำเป็นต้องมีปัญญาสูง หรือ คนสวมรองเท้าเล็กไม่จำเป็นต้องเป็นคนมีปัญญาต่ำ

2.8 <u>การแปลงคะแนน</u>

คะแนนจากการสอบทั่วไป อยู่ในลักษณะของคะแนนดิบ (Raw Score) ซึ่งวัดมาโดย

ใช้หน่วยหรือมาตราวัดต่างกันออกไป ไม่สามารถที่จะนำมาเปรียบเทียบกันได้ซึ่งมีวิธีแปลงคะแนนดิบต่าง ๆ เพื่อให้อยู่ในสภาพที่เหมาะในการเปรียบเทียบขึ้น เช่น

วิธีแปลงคะแนนดิบให้เป็นคะแนนมาตรฐานซี (Z-Score) หรือคะแนนมาตรฐานที (T-Score) เนื่องจากคะแนนมาตรฐาน (Standard Score) โดยทั่วไปจะมีการแจกแจงเป็นโค้งปกติ และมีหน่วยของคะแนนเท่ากัน จึงสามารถที่จะรวมคะแนนต่าง ๆ เข้าด้วยกันได้แล้วเปรียบเทียบกันได้ว่า ใครมีคะแนนดีกว่ากัน เช่น คะแนนมาตรฐานซี นั้น จะมีค่าเฉลี่ยคะแนน (Mean) เท่ากับ 0 และมีค่าเบี่ยงเบนมาตรฐาน (SD) เท่ากับ 1 จึงใช้คะแนนมาตรฐานซีของการวัดแบบต่าง ๆ มาเปรียบเทียบกันได้ สูตรที่ใช้ในการแปลงให้เป็นคะแนน ซี คือ

$$Z = \frac{X-\bar{X}}{SD} \quad (5)$$

เมื่อ Z = คะแนนมาตรฐาน ซี

X คะแนนดิบของแต่ละคน

$\bar{X}$ คะแนนเฉลี่ยของกลุ่ม

SD = คะแนนเบี่ยงเบนมาตรฐาน ซึ่งคำนวณได้จาก

$$SD = \sqrt{\frac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}}$$

N = จำนวนคนที่สอบ

เนื่องจากคะแนนมาตรฐานซี ที่คำนวณได้มีทั้งค่า + (บวก) และ - (ลบ) ซึ่งขึ้นอยู่กับคะแนนดิบ คือ คะแนนดิบที่สูงกว่าค่าเฉลี่ย ($\bar{X}$) คะแนนซีที่แปลงได้จะมีค่าเป็น + ตรงกันข้าม คะแนนดิบที่ต่ำกว่าค่าเฉลี่ยก็จะได้ค่าซีเป็น - ดังนั้น ในทางปฏิบัติจึงนิยมแปลงคะแนนซี ให้เป็นคะแนนมาตรฐานที (T-Score) ซึ่งนิยมให้มี Mean = 50 และ SD = 10 ดังสูตรต่อไปนี้

$$T = 10Z + 50 \quad (5)$$

หรือ $T = 50 - 10Z$ (ถ้าคะแนนน้อยหมายถึง มีความสามารถสูง เช่น เวลาในการวิ่ง เป็นต้น)

2.9 คะแนนเปอร์เซนไทล์ และตำแหน่งเปอร์เซนไทล์

(Percentile & Percentile Rank)

เพื่อที่จะให้คะแนนมีความหมายขึ้น จึงมักบอกผลการสอบเป็นอันดับที่ เช่น สอบได้ที่ 8 สอบได้ที่ 15 เป็นต้น ตัวเลขนี้ไม่สู้จะมีความหมายเท่าไรนัก เพราะไม่ทราบว่าได้ที่ 8 จากนักเรียนกี่คน ถ้าเป็นที่ 8 จากนักเรียน 200 คนก็คงจะเก่ง แต่ถ้าเป็นที่ 8 จากนักเรียน 8 คน คงจะไม่เก่งแน่ วิธีการหนึ่งที่จะช่วยให้คะแนนมีความหมายขึ้นก็คือ คิดคะแนนเป็นเปอร์เซนไทล์ และตำแหน่งเปอร์เซนไทล์

2.9.1 คะแนนเปอร์เซนไทล์ (Percentile : P_X)

คะแนนเปอร์เซนไทล์ เป็นคะแนนที่แสดงให้ทราบว่า (Percentile:P_X) มีนักเรียนกี่เปอร์เซนต์สอบได้ต่ำกว่าคะแนนนี้ เช่น แสงสอบได้ 63 คะแนน คิดเป็นเปอร์เซนไทล์ที่ 70 หมายความว่า มีนักเรียน 70 % ของนักเรียนที่สอบเข้า สอบได้คะแนนต่ำกว่า 63 คะแนน

เปอร์เซนไทล์ ที่ X ใช้ P_x เป็นสัญญลักษณ์ เช่น เปอร์เซนไทล์ที่ 25 สามารถเขียนสัญญลักษณ์ได้เป็น P_{25} มีสูตรดังนี้

$$P_x = \frac{cfx}{N} \times 100$$

ลำดับขั้นตอนในการคำนวณ

1. เรียงลำดับคะแนนจากสูงไปต่ำ
2. หาความถี่สะสมของแต่ละคะแนน (f)
3. หาความถี่สะสมจากคะแนนต่ำไปหาคะแนนสูง (ct)
4. รวมจำนวนผู้เข้าสอบทั้งหมด (N)
5. คำนวณหาค่าเปอร์เซนไทล์ จากสูตรดังกล่าวข้างต้น

2.9.2 ตำแหน่งเปอร์เซนไทล์ (Percentile Ranks : PR)

การรายงานคะแนนผลการทดสอบในลักษณะ ตำแหน่งเปอร์เซนไทล์ ก็เป็นที่นิยมใช้กันมากอีกชนิดหนึ่ง คือ เป็นการระบุว่ามีผู้ที่เข้าสอบด้วยกันกี่เปอร์เซนต์ของผู้เข้าสอบ

ทั้งหมด ได้คะแนนต่ำกว่าคะแนนกึ่งกลาง (Mid Point) ของแต่ละคะแนนหรือของแต่ละช่วงคะแนน การคำนวณหาตตำแหน่งเปอร์เซนไทล์ เป็นการนำเอาจำนวนความถี่ของผู้ที่ได้คะแนนต่ำกว่าคะแนนนั้นทั้งหมดมารวมกัน และบวกด้วยครึ่งหนึ่งของจำนวนความถี่ของผู้ที่ได้คะแนนนั้นหารด้วยจำนวนความถี่ทั้งหมด หรือจำนวนผู้เข้าสอบทั้งหมด ตำแหน่งเปอร์เซนไทล์เป็นการรายงานผลการสอบที่มีความหมายในตัว (Self-Interpreting) และใช้ในลักษณะเปรียบเทียบระหว่างผลการสอบของนักเรียนแต่ละคน โดยใช้สัญญลักษณ์ว่า PR (3)

$$PR = \frac{m}{N} \times 100$$

เมื่อ

$$m = feb + \frac{1}{2} f$$

<u>ลำดับขั้นตอนในการคำนวณ</u>

1. เรียงลำดับคะแนนจากสูงไปต่ำ
2. หาความถี่สะสมของแต่ละคะแนน (f) จากคะแนนน้อยไปหาคะแนนมาก
3. หาความถี่สะสมจากคะแนนน้อยไปหาคะแนนมาก (feb)
4. หาความถี่สะสมที่แท้จริงจากคะแนนน้อยไปหาคะแนนมาก (m) โดยการเอาความถี่สะสมที่ต่ำกว่าชั้นที่ต้องการ 1 ชั้น (feb) รวมกับครึ่งหนึ่งของความถี่ในชั้นที่ต้องการ

   การที่ใช้ความถี่สะสมต่ำกว่า (b = below) ก็เพื่อจะรวมคนทั้งหมดก่อนที่จะถึงชั้นที่ต้องการ และรวมกับครึ่งหนึ่งของความถี่ของชั้นที่ต้องการก็เพราะว่า ครึ่งหนึ่งจะทำให้มีค่าผิดพลาดน้อยที่สุดในการคำนวณจำนวนคนในชั้นนั้น ๆ ที่ต้องการ
5. รวมจำนวนผู้เข้าสอบทั้งหมด (N)
6. คำนวณค่าตำแหน่งเปอร์เซนไทล์จากสูตรดังกล่าวข้างต้น

ตารางที่ 2.1 แสดงการคำนวณหาตำแหน่งเปอร์เซนไทล์ของคะแนนผลการสอบ

คะแนน	ความถี่ f	ความถี่สะสม m	2(ความถี่สะสม)	เปอร์เซนต์ความถี่	ตำแหน่งเปอร์เซนต์ไทล์
44		8.0	16	100.00	100
43	/	7.5	15	93.75	94
42	///	5.5	11	68.75	69
41	/	3.5	7	43.75	44
40	//	2.0	4	25.00	25
39		1.0	2	12.50	13
38	/	0.5	1	6.25	6
37		0	0	0.00	0

บางครั้งจำนวนความถี่ ในแต่ละคะแนนเป็นเลข เมื่อหารด้วยสอง ทำให้ได้เลขจุดทศนิยม จึงนิยมเอาสองคูณตลอด แล้วแปลงความถี่สะสมที่ได้นั้นเป็นเปอร์เซนต์ของความถี่ โดยเอาความถี่ทั้งหมดไปหารความถี่สะสมทีละตัวแล้วคูณด้วย 100 แล้วเปลี่ยนให้เป็นตำแหน่งเปอร์เซนไทล์ โดยปัดเศษที่เป็นจุดทศนิยมมาก

จากตารางที่ 2.1 คะแนน 42 คะแนน มีตำแหน่งเปอร์เซนไทล์เท่ากับ 94 หมายความว่านักเรียน 94 % ของนักเรียนที่เข้าสอบ สอบได้คะแนนต่ำกว่า 42 คะแนน

มักจะมีความสับสนเสมอระหว่างคะแนนเปอร์เซนไทล์ (P_x)กับ ตำแหน่งเปอร์เซนไทล์ (PR) ในการหาค่าของเปอร์เซนไทล์ และตำแหน่งเปอร์เซนไทล์เราเทียบจำนวนนักเรียนเป็น 100 คะแนนเปอร์เซนไทล์ คือ คะแนนที่มีนักเรียนจำนวนกี่เปอร์เซนต์สอบได้ต่ำกว่า ส่วนตำแหน่งเปอร์เซนไทล์ ของคะแนนคือร้อยละของนักเรียนที่สอบได้ต่ำกว่าคะแนนนั้น ดังนั้น เปอร์เซนไทล์ คือ คะแนน ส่วนตำแหน่งเปอร์เซนไทล์คือร้อยละ

2.10 การวิเคราะห์ค่าความเชื่อมั่นของแบบทดสอบ

การคำนวณหาค่าความเชื่อมั่นของแบบทดสอบทำได้หลายวิธี ซึ่งแต่ละวิธีมีค่าใกล้เคียงกัน

การที่จะเลือกใช้วิธีใดนั้นขึ้นอยู่กับความมุ่งหมายและลักษณะของคะแนนที่ได้จากการทดสอบนั้น วิธีหาค่าความเชื่อมั่นมี 4 วิธี คือ

(1) วิธีสอบซ้ำ (Measures of Stability)

(2) วิธีใช้ข้อสอบคล้ายกัน (Measures of Equivalence)

(3) วิธีใช้ข้อสอบคล้ายกันและสอบซ้ำ (Measures of Equivalence and Stability)

(4) วิธีวัดความคงที่ภายใน (Measures of Internal Consistency)

- วิธีแบ่งครึ่งจำนวนข้อสอบ (Split-Half)
- วิธีของคูเดอร์-ริชาร์ดสัน (Kuder-Richardson Estimates)
- วิธีของฮอยท์ (Hoyt's Analysis of Variance Procedure)
- วิธีของครอนบาค แอลฟา (Cronbach Alpha)

สำหรับวิธีที่ 1,2 และ 3 นั้นไม่ค่อยเป็นที่นิยมกันในทางปฏิบัติ เพราะใช้เวลาในการทดสอบนาน ดังนั้นจะกล่าวถึงวิธีคำนวณหาค่าความเชื่อมั่นเฉพาะ แบบที่ 4 เท่านั้น

(4) วิธีวัดความคงที่ภายใน (Internal Consistency) โดยทั่วไปข้อสอบในแบบทดสอบชุดเดียวกัน ควรมีความสัมพันธ์อันแสดงถึงการวัดในเรื่องเดียวกัน ในทางปฏิบัตินิยมใช้กันมาก เนื่องจากให้ความสะดวกและเหมาะสมกับสถานการณ์ทั่วไป ซึ่งมักจะมีการสอบครั้งเดียว การหาสัมประสิทธิ์ความเชื่อมั่น โดยวิธีนี้มี

2.10.1 วิธีแบ่งครึ่ง (Split-Half Method)

วิธีนี้เป็นการวัดความเชื่อมั่นที่เกี่ยวกับ Internal Consistency คือคือแทนที่เราจะนำแบบทดสอบที่แยกนั้นไปทดสอบคนละเวลา เราก็นำมาทดสอบในเวลาเดียวกัน โดยอาจแบ่งข้อสอบออกเป็นตอน ก. และตอน ข. หรือ โดยวิธีแบ่งข้อสอบเป็นข้อคี่ และข้อคู่ (Odd-Even Item) ซึ่งแบบทดสอบเดียวกันแต่แยกออกเป็น 2 ส่วน แล้วคำนวณหาค่าสัมประสิทธิ์ความเชื่อมั่นโดยวิธีของเพียร์สัน (Pearson) ซึ่งจะได้ค่า r_{12} จากนั้นก็คำนวณสัมประสิทธิ์ความเชื่อมั่นของแบบสอบชุดนั้นโดยวิธีของเสปียร์แบน-บราวน์ (Speaernan-Brown Formula) ดังสูตรต่อไปนี้

$$r_{tt} = \frac{2\ r_{12}}{1+ r_{12}}^{(6)}$$

เมื่อ r_{12} คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบสอบส่วนที่ 1 และส่วนที่ 2

r_{tt} คือ ค่าสัมประสทธิ์ความเชื่อมั่นของแบบทดสอบทั้งฉบับ

ตัวอย่างที่ 2.7 ถ้าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบทดสอบย่อยสองฉบับ เท่ากับ .80 หรือ r_{12} เท่ากับ .80 ค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับจะเท่ากับ

$$r_{tt} = \frac{2(.80)}{1+.80}$$

$$= \frac{1.60}{1.80}$$

$$= .89$$

2.10.2 วิธีของคูเดอร์-ริชาร์ดสัน สูตร 20 หรือ สูตร 21 (Kuder-Richardson Formular 20 or Formular 21)

เป็นวิธีการหาค่าความเชื่อมั่นของแบบทดสอบอีกวิธีหนึ่ง ซึ่งวิธีนี้ข้อสอบจะต้องให้คะแนนเมื่อถูกเท่ากับ 1 และผิดเท่ากับ 0 นอกจากนั้นต้องทราบอัตราส่วนของผู้ตอบถูกและผิด คือ ค่า p และ q แล้วคำนวณตามสูตรต่อไปนี้

(6)

$$\text{K-R}_{20} \; : \; r_{tt} = \frac{n}{n-1}\left[1 - \frac{\Sigma pq}{\sigma_X^2}\right] \quad \ldots (1)$$

(6)

$$\text{K-R}_{21} \; : \; r_{tt} = \frac{n}{n-1}\left[1 - \frac{\bar{X}\,(n-\bar{X})}{n\sigma_X^2}\right] \quad \ldots (2)$$

เมื่อ r_{tt} = ค่าความเชื่อมั่นแบบทดสอบ

n = จำนวนข้อสอบในแบบทดสอบ

p = สัดส่วนของผู้ที่ตอบถูก

q = สัดส่วนของผู้ตอบผิด ($q = 1-p$)

Σpq = ความแปรปรวนของข้อสอบแต่ละข้อ (ในกรณีที่ให้คะแนนแบบศูนย์-หนึ่ง)

σ_X^2 = ความแปรปรวนของคะแนนรวมทั้งหมด

$\bar{X}$ = ค่าเฉลี่ยของคะแนนรวมทั้งหมด

การหาความเชื่อมั่นโดยใช้สูตร K-R_{21} ปกติจะได้ค่าต่ำกว่า K-R_{20} เพราะคำนวณจากค่าเฉลี่ย ครูตามโรงเรียนต่าง ๆ มักนิยมใช้สูตร K-R_{21} เพราะว่าใช้การคำนวณน้อยกว่าและทำได้รวดเร็วกว่า เพียงแต่แทนค่าจำนวนข้อในแบบทดสอบ (n) ค่าเฉลี่ย ($\bar{X}$) และความแปรปรวน (Σ_X^2) ลงในสมการ (2) ก็สามารถคำนวณค่าความเชื่อมั่นออกมาได้

<u>ตัวอย่างที่ 2.8</u> แสดงการหาค่าความเชื่อมั่นโดยใช้สูตร K-R_{20} และสูตร K-R_{21} ของแบบทดสอบหนึ่ง ซึ่งมีจำนวน 6 ข้อ ทดสอบกับนักเรียน 10 คน ให้คะแนนตามวิธี ศูนย์-หนึ่ง (Zero One Method)

ตารางที่ 2.2 แสดงการหาค่าความเชื่อมั่นของแบบทดสอบจำนวน 6 ข้อ ทดสอบกับนักเรียน 10 คน

นักเรียนคนที่	ข้อที่ 1	2	3	4	5	6	คะแนนรวม
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	.29	.24	-

$\Sigma pq = 1.18$

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

$\Sigma X^2 = 192$

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

$$\text{K-R}_{20} : r_{tt} = \frac{n}{n-1}\left[1 - \frac{\Sigma pq}{\sigma_X^2}\right]$$

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

$$= 1.2 \ \left(\frac{.38}{1.56}\right)$$

$$= .29$$

$$K\text{-}R_{21} \ : \ r_{tt} = \frac{n}{n-1}\left[1 - \frac{\bar{X}\,(n-\bar{X})}{N\sigma^2_X}\right]$$

$$N = 10,\ \sigma^2_X = 1.56$$

$$\bar{X} = 4.2,\ n = 6$$

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

$$= \frac{6}{5}\left[1 - \frac{(.7)(1.8)}{1.56}\right]$$

$$= \frac{6}{5}\left[\frac{1.56-1.26}{1.56}\right]$$

$$= \frac{6}{5}\left[\frac{.30}{.156}\right]$$

$$r_{tt} = .23$$

จะเห็นว่าการคำนวณค่าความเชื่อมั่น โดยใช้สูตร $K\text{-}R_{21}$ จะได้ค่าต่ำกว่าการคำนวณโดยใช้สูตร $K\text{-}R_{20}$ ทั้งเนื่องจากสูตร $K\text{-}R_{21}$ ใช้ค่าเฉลี่ย (Mean) ของทุกข้อแทน Σpq ของแต่ละข้อ

2.10.3 วิธีของครอนบาค แอลฟา (Cronbach's Alpha)

ในกรณีที่ระบบการให้คะแนนไม่เป็น ศูนย์ หรือ หนึ่ง. ครอนบาค จึงได้สร้างสูตร Coefficient Alpha ขึ้นเพื่อหาค่าความเชื่อมั่นของแบบทดสอบในปี ค.ศ. 1951 สูตรนี้มีลักษณะเหมือน $K\text{-}R_{20}$ แต่ต่างกันที่แทน Σpq ด้วย σ^2 ซึ่ง σ^2_i เป็นค่าความแปรปรวนของข้อสอบแต่ละข้อ และเครื่องหมาย Σ แทนการรวมกันของ Item Variance ทั้งหมด ในแบบทดสอบฉบับนั้น สูตรนี้มีประโยชน์ ในการหาค่าความเชื่อมั่นของแบบทดสอบ ที่มีข้อสอบเป็นแบบเรียงความ (Essay Questions) ซึ่งคะแนนในแต่ละข้อมีพิสัยไม่เท่ากัน Coefficient Alpha มีสูตรดังนี้ (6)

เมื่อ $$\alpha = \frac{n}{n-1}\left(1 - \frac{\Sigma\sigma^2_i}{\sigma^2_x}\right)$$

α = ความเชื่อมั่นของแบบทดสอบ

σ_i^2 = ความแปรปรวนของข้อสอบแต่ละข้อ (Single Item Variance)

σ_x^2 = ความแปรปรวนของข้อสอบทั้งหมด(Total Test Variance)

n = จำนวนข้อในแบบทดสอบ

ตัวอย่างที่ 2.9 ข้อสอบเรียงความฉบับหนึ่งมีข้อสอบจำนวน 5 ข้อ ซึ่งมีคะแนนเต็ม 10,5,8,5, และ 2 ตามลำดับ ทดสอบกับนักเรียน 10 คน ได้คะแนนผลการสอบออกมาดังตารางข้างล่างนี้ จงหาความเชื่อมั่นของแบบทดสอบฉบับนี้

นักเรียน	ข้อที่					คะแนน
เลขที่	1	2	3	4	5	รวม
1	9	4	7	4	1	25
2	8	3	6	2	0	19
3	4	5	4	1	0	14
4	5	5	3	4	1	18
5	0	1	2	3	0	6
6	3	3	1	2	1	10
7	6	4	0	4	2	16
8	8	2	5	3	2	20
9	10	1	6	1	1	19
10	7	2	7	2	0	18
σ_i^2	8.40	2.00	5.69	1.24	0.56	

$$\Sigma\sigma_i^2 = 17.89$$

$$\sigma_x^2 = 26.05$$

$$\alpha = \frac{5}{4}\left(- \frac{17.89}{26.05} \right)$$

$$= 0.392$$

2.10.4 การหาความเชื่อมั่นของฮอยท์ (Hoyt's Analysis of Variance)

ฮอยท์ ได้พบวิธีหาค่าความเชื่อมั่นโดยการวิเคราะห์ค่าความแปรปรวน (Analysis of Variance) ในปี ค.ศ. 1941 วิธีนี้ให้ความเชื่อมั่นเหมือนกับสูตร $K\text{-}R_{20}$

ลำดับขั้นในการคำนวณค่าความเชื่อมั่น

1. นำคะแนนที่ตรวจได้บรรจุลงในตารางสำหรับคะแนนที่ได้จะเป็นคะแนนแบบ 0 กับ 1

ตารางที่ 2.3 แสดงการบรรจุคะแนนลงในตาราง

นักเรียน	ข้อที่				คะแนน
เลขที่	1	2	3	K	รวม
1	X_{11}	X_{12}	X_{13} ...	X_{1K}	T_1
2	X_{21}	X_{22}			T_2
3	X_{31}				T_3
.	.				.
.	.				.
.	.				.
N	X_{N1} ...			X_{NK}	T_N
ผลรวม	R_1	R_2	R_3 ...	R_K	$\Sigma R = T$

2. คำนวณค่าต่าง ๆ จากสูตรในตาราง

ตารางที่ 2.4 แสดงสูตรในการหาค่าความเชื่อมั่นวิธีของฮอยท์

Source of Variance	Degree of Freedom	Sum of Squares	Variance
Among Students	N-1	$\frac{\Sigma T^2}{n} - \frac{(\Sigma T)^2}{nN}$	SS_{St} / (N-1)
Among Item	n-1	$\frac{\Sigma R^2}{N} - \frac{\Sigma T)^2}{nN}$	SS_{it} / (N-1)
Remainder	(n-1)(N-1)	$SS_{tot} - SS_{St} - SS_{it}$	SS_{rem} / (n-1)(N-1)
Total	nN-1	$\frac{(\Sigma T)\ (nN - \Sigma T)}{nN}$	

ตัวอย่างที่ 2.10 จากข้อมูลในตารางที่ 2.2

	ข้อที่						รวม
	1	2	3	4	5	6	
R	8	9	6	7	6	6	42
R^2	64	81	36	49	36	36	302
W	2	1	4	3	4	4	

$\Sigma T = 42 \quad N = 10 \quad R^2 = 302 \quad W = 18$

$\Sigma T^2 = 192 \quad n = 6 \quad R = 42$

$$SS_{st} = \frac{\Sigma T^2}{n} - \frac{(\Sigma T)^2}{nN} = \frac{192}{6} - \frac{(42)^2}{60} = 32-29.4 = 2.6$$

$$SS_{it} = \frac{\Sigma R^2}{N} - \frac{(\Sigma T)^2}{nN} = \frac{302}{10} - 29.4 = 30.2-29.4 = .8$$

$$SS_{tot} = \frac{(\Sigma T)\ (nN-\Sigma T)}{nN} = \frac{(42)(18)}{60} = \frac{756}{60} = 12.6$$

$$SS_{rem} = 12.6-2.6-.8 = 9.2$$

ANOVA

Source of Variation	Degree of Freedom	Sum of Squares	Variance
Among Students	9	216	.289
Among Item	5	.8	.160
Remainder	45	9.2	.204
Total	59	12.6	

$$r_{tt} = 1 - \frac{S^2_{rem}}{S^2_{among\ st}}$$

$$= 1 - \frac{.204}{.289}$$

$$= 0.294$$

2.10.5 <u>การตีความค่าความเชื่อมั่น</u> (r)

เครื่องมือใดที่มีความเชื่อมั่นสูง จะให้ความมั่นใจได้ว่ามีความคลาดเคลื่อนมาตรฐานในการวัดต่ำ ย่อมเป็นประโยชน์ ในการประมาณ (Estimate)คะแนนที่แท้จริง (True Score)ของผู้ตอบได้ใกล้เคียง โดยทฤษฎีแล้วคะแนนจากการวัดทั่วไปจะประกอบด้วยคะแนนส่วนซึ่งเป็นคะแนนคลาดเคลื่อน (Error Score) อยู่ ฉะนั้น ถ้าค่าความคลาดเคลื่อนไม่มีหรือมีน้อยย่อมหมายความว่าคะแนนที่สังเกตได้ (Observed Score)หรือ คะแนนจากการสอบจะเท่าหรือใกล้คะแนนใกล้คะแนนจริง (True Score)ในโอกาสที่ไม่มีความคลาดเคลื่อนของการวัด ค่าสัมประสิทธิ์ของความเชื่อมั่นจะเท่ากับ 1 โดยปกติแล้วแบบทดสอบที่สร้างขึ้นควรมีค่าความเชื่อมั่นตั้งแต่ .60 ขึ้นไป และถ้าให้มาตรฐานแล้วอาจได้ถึง .90 ขึ้นไป อย่างไรก็ตาม ไม่มีมาตรฐานใดจะบอกค่าความเชื่อมั่นนั้นว่าควรจะมีค่าเท่าใดอย่างแน่นอน ถ้าแบบทดสอบใดมีค่าความเชื่อมั่น เช่น ต่ำกว่า .05 ลงไป ก็จะต้องมีการปรับปรุงข้อสอบให้ดี ตลอดจนอาจจะมีการเพิ่มความยาวของแบบทดสอบ เป็นต้น

การตรวจสอบประสิทธิภาพของเครื่องมือในการวัดผลการศึกษา ดังที่ได้กล่าวมาแล้ว คือ ค่าความเชื่อมั่น ค่าความเที่ยงตรงแล้ว ยังต้องมีการตรวจสอบคุณลักษณะอื่น ๆ ต่อไป เช่น

การวิเคราะห์ข้อสอบ เพื่อตรวจสอบระดับความยากง่าย (Level of Difficulty) ของข้อสอบว่าเหมาะสมกับวัตถุประสงค์ของการวัดพฤติกรรมที่ต้องการ โดยเฉพาะมีความเหมาะสมกับกลุ่มทดสอบ (Validation Group) หรือไม่ และข้อสอบต่าง ๆ มีอำนาจจำแนก (Power of Discrimination) ในการแยกผู้มีความสามารถได้ดีเพียงไร

2.11 การวิเคราะห์ค่าความเที่ยงตรงของแบบทดสอบ (Test Validity Analysis)

เป็นการวิเคราะห์แบบทดสอบทั้งฉบับ ความเที่ยงตรง (Validity) ของเครื่องมือวัดผลทางการศึกษานั้น ขึ้นอยู่กับวัตถุประสงค์ของการวัด คือ ต้องการวัดสิ่งใดบ้าง ซึ่งในทางปฏิบัติก็คือ วัตถุประสงค์ของกระบวนการเรียนการสอนนั่นเอง เครื่องมือวัดผลใดถ้าสามารถวัดในสิ่งที่ต้องการวัดตามเป้าหมายที่กำหนด ก็แสดงว่าเครื่องมือนั้นมีความเที่ยงตรงในการวัดความเที่ยงตรงของเครื่องมืออาจจำแนกตามลักษณะของการวัด ดังนี้คือ

(1) ความเที่ยงตรงตามเนื้อหา (Content Validity)

(2) ความเที่ยงตรงตามทฤษฎี (Construct Validity)

(3) ความเที่ยงตรงตามเกณฑ์สัมพันธ์ (Criterion Related Validity)

3.1 ความเที่ยงตรงตามสถานการณ์ (Concurrent Validity)

3.2 ความเที่ยงตรงตามพยากรณ์ (Predictive Validity)

(1) ความเที่ยงตรงตามเนื้อหา หมายถึง เครื่องมือการวัดที่มีความสามารถวัดความรู้จากตัวแทนของเนื้อหา ที่สุ่มมา อันเป็นพฤติกรรมที่ต้องการตามจุดมุ่งหมายการสอน ให้ผู้เรียนเกิดการเรียนรู้ ผู้สร้างเครื่องมือวัดผลความเที่ยงตรงตามเนื้อหา จำเป็นต้องมีการตรวจสอบเนื้อหา ที่สอนอย่างรอบคอบ เป็นต้นว่าจากหลักสูตรประมวลการสอน และแหล่งอื่น ๆ ที่แสดงขอบข่ายของความรู้ เช่น ตำราต่าง ๆ นอกจากนั้นผู้สร้างจะต้องมีการวางแผนการสอนควบคู่กับการเตรียมตารางวิเคราะห์เนื้อวิชา เช่น ทำตารางวิเคราะห์หลักสูตร (Table of Specification) เครื่องมือที่วัดความเที่ยงตรงตามเนื้อหาทั่วไปได้แก่ แบบทดสอบวัดผลสัมฤทธิ์ (Achievement Tests) ซึ่งประกอบด้วยข้อสอบต่าง ๆ ที่มีประสิทธิภาพ

(2) ความเที่ยงตรงตามทฤษฎี หมายถึง เครื่องมือนั้น สามารถวัดทฤษฎีหรือคุณลักษณะของพฤติกรรมใด ที่ได้อธิบายไว้หรือตามสมมุติฐานที่ตั้งไว้ โดยทั่วไปถ้าแบบทดสอบใดมีความเที่ยงตรงตามทฤษฎีดังกล่าว แสดงว่าคะแนนจากแบบทดสอบต้องมีความสัมพันธ์กับทฤษฎี

หรือคุณลักษณะที่กำหนด ตัวอย่าง เช่น แบบทดสอบที่สร้างขึ้นวัดพฤติกรรม เกี่ยวกับความวิตกกังวล คนที่ทำคะแนนได้สูง ก็ควรมีความวิตกกังวล คนที่ทำคะแนนได้สูง ก็คงมีความวิตกกังวลสูงด้วย แสดงให้เห็นว่ามีความเที่ยงตรงตามทฤษฎี นั้น ซึ่งจะชี้ให้เห็นถึงความสัมพันธ์ระหว่างทฤษฎี (Theory) และผลการปฏิบัติของแบบทดสอบ (Actual Test Performance)

สำหรับความเที่ยงตรงทั้ง 2 แบบที่ได้กล่าวมาแล้วนั้น จะวัดออกมาเป็นตัวเลขสถิติไม่ได้ แต่จะเที่ยงตรงแค่ไหนขึ้นอยู่กับว่าตรงกับตารางวิเคราะห์หลักสูตร มากน้อยเพียงไร ถ้าออกได้ตามที่วิเคราะห์มาก็เรียกว่ามีความเที่ยงตรงตามเนื้อเรื่องและทฤษฎี

(3) <u>ความเที่ยงตรงตามเกณฑ์สัมพันธ์</u> (Criterian Related Validity) หมายถึงเครื่องมือการวัดที่สามารถทำนายพฤติกรรมของบุคคลในสถานการณ์จำเพาะตามต้องการ เครื่องมือการวัดความเที่ยงตรงประเภทนี้โดยทั่วไปได้แก่แบบทดสอบวัดความถนัด (Aptitude Test) แบบทดสอบวัดความสนใจ (Interest)ต่าง ๆ เช่น ความสนใจในอาชีพเป็นต้น เครื่องมือการวัดที่จะมีความเที่ยงตรงดังกล่าวมักจะตรวจสอบจากคะแนนที่ได้จากแบบทดสอบ (คือวัดเป็นตัวเลขสถิติได้) ว่ามีความสัมพันธ์กับเกณฑ์ที่ตั้งไว้เพียงไร และความเที่ยงตรงตามเกณฑ์สัมพันธ์ สามารถจำแนกตามเกณฑ์ที่ใช้วัดดังนี้

3.1 <u>ความเที่ยงตรงตามพยากรณ์</u> (Predictive Validity) คือ ความเที่ยงตรงซึ่งแสดงให้เห็นว่าคะแนนจากการทำแบบทดสอบสามารถที่จะทำนายผล ที่เกิดขึ้นในอนาคต ตัวอย่าง เช่น ถ้าคะแนนการสอบเข้ามหาวิทยาลัยมีความเที่ยงตรง ตามพยากรณ์ความสามารถของผู้เรียน ก็หมายความว่า ผู้สอบเข้าได้จะสามารถเรียนสำเร็จในมหาวิทยาลัยได้ หรือ ถ้าคะแนนจากแบบทดสอบความถนัด ซึ่งมีความเที่ยงตรงตามพยากรณ์ ย่อมแสดงให้เห็นว่า ผู้ได้คะแนนสูง จะต้องมีความสามารถเรียนในสาขาที่ตนถนัดได้สำเร็จ ฯลฯ จากตัวอย่างที่กล่าวมาจะเห็นว่าเกณฑ์ที่จะใช้ในการตรวจสอบความเที่ยงตรงของเครื่องมือเป็นเกณฑ์ในอนาคต โดยทั่วไปต้องรอช่วงระยะเวลาหนึ่งจึงจะทำการวิเคราะห์ได้

3.2 <u>ความเที่ยงตรงตามสถานการณ์</u> (Concurrent Validity) คือ ความเที่ยงตรงที่แสดงพฤติกรรมที่วัดได้ว่าตรงสภาพที่เป็นอยู่ในปัจจุบัน ซึ่งเป็นเกณฑ์ที่มีอยู่และสัมพันธ์กับสิ่งที่ต้องการวัด การตรวจสอบความเที่ยงตรงประเภทนี้ สามารถตรวจสอบได้โดยไม่ต้องรอเกณฑ์ในอนาคต ตัวอย่าง เช่น คะแนนจากการสอบภาคทฤษฎีได้ดี ก็น่าจะสัมพันธ์กับคะแนนภาคปฏิบัติ ซึ่งเป็นเรื่องเดียวกัน ถ้าแบบทดสอบดังกล่าวมีความเที่ยงตรงตามสถานการณ์

การประมาณค่าความเที่ยงตรงตามเกณฑ์สัมพันธ์ ตรวจสอบได้จากการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบทดสอบกับเกณฑ์ โดยใช้สูตรของเพียร์สัน (Pearson Product Moment Correlation Coefficient) ดังนี้คือ

$$r_{XY} = \frac{N\,\Sigma XY - \Sigma X\,\Sigma Y}{\sqrt{\left[N\Sigma X^2 - (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}} \quad (6)$$

เมื่อ r_Y = ค่าสัมประสิทธิ์สหสัมพันธ์ (ค่าความเที่ยงตรง)

X = คะแนนจากแบบทดสอบ

Y = คะแนนจากเกณฑ์

Σ = ผลรวม

2.12 การวิเคราะห์ข้อสอบเป็นรายข้อ (Item Analysis)

ความมุ่งหมายที่สำคัญของการวิเคราะห์ข้อสอบ เพื่อทราบคุณลักษณะของข้อสอบ (Item) 2 ประการคือ

- ระดับความยากง่าย (Level of Difficulty)
- อำนวจการจำแนกของข้อสอบ (Item Discriminating Power)

การวิเคราะห์ข้อสอบโดยทั่วไป จึงหมายถึงการวิเคราะห์เป็นรายข้อ โดยใช้เทคนิคทางสถิติ ข้อสอบที่ทำการวิเคราะห์จะเป็นข้อสอบประเภทเลือกตอบ เช่น ข้อสอบ 2 ตัวเลือก แบบ ถูก-ผิด (True-False Item) หรือ หลายตัวเลือกแบบคำตอบถูกหรือดีที่สุดเพียงคำตอบเดียว (Multiple-Choice Item) สำหรับการวิเคราะห์ข้อสอบแบบหลายตัวเลือก นอกจากจะตรวจสอบระดับความยากง่ายและอำนาจจำแนกของคำตอบแล้ว ยังต้องตรวจสอบประสิทธิภาพของตัวลวง (Distractors)ได้อีกด้วย

2.12.1 การคำนวณค่าความยากของข้อสอบ (Level of Item Difficulty)

ค่าความยากของข้อสอบ หมายถึง สัดส่วนของจำนวนผู้ตอบข้อสอบข้อนั้นต่อจำนวนผู้สอบทั้งหมด หรือ หมายถึงเปอร์เซนต์ (%) ของผู้ตอบข้อสอบข้อนั้นถูก[2]

ตัวอย่างที่ 2.11 มีผู้เข้าสอบ 100 คน ตอบข้อสอบข้อที่ 1 ถูก 60 คน

ดังนั้นข้อสอบข้อที่ 1 จึงมีค่าระดับความยากเท่ากับ $\frac{60}{100}$ = .60 สรุปได้ว่า วิธีหาระดับความยากง่าย (P) คำนวณได้ดังสูตรต่อไปนี้

$$P = \frac{R}{T}$$

เมื่อ P = ระดับความยากของข้อสอบ

R = จำนวนนักเรียนที่สอบตก

T = จำนวนนักเรียนที่นำมาวิเคราะห์

วิธีการคำนวณดังกล่าวได้จากแนวความคิดที่ว่า คะแนนเฉลี่ยของแบบทดสอบเป็นตัวชี้ถึงความยากง่ายของแบบทดสอบนั้น ในกรณีที่ข้อสอบนั้นตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 ถ้าให้ R เท่ากับจำนวนผู้ตอบถูก และให้ W เป็นจำนวนผู้ตอบผิด

ดังนั้น คะแนนเฉลี่ยสำหรับความยากง่ายของข้อสอบ $= \frac{R.1+W.0}{R+W} = \frac{R}{N}$

เมื่อ R = จำนวนผู้ตอบถูก

W = ผู้ตอบผิด

R + W = จำนวนผู้ตอบทั้งหมด (N)

จะเห็นได้ว่าแนวความคิดดังกล่าว สัดส่วนของผู้ตอบข้อสอบถูก หรือคะแนนเฉลี่ยของข้อสอบ คือ ดัชนีความยากง่ายของข้อสอบ (Index of Item Difficulty) และเพื่อไม่ให้เป็นค่าทศนิยม จึงนิยมให้เป็นค่าเปอร์เซนต์โดยคำนวณตามสูตร

$$P = \frac{R}{N} \times 100$$

2.12.2 ค่าอำนาจจำแนกของข้อสอบ (Power of Item Discrimination)

หมายถึง ความสามารถของข้อสอบที่จะแยกหรือจำแนกผู้เข้าสอบตามลำดับ ความสามารถ เช่น จำแนกคนเก่งและคนไม่เก่ง หรือ คนที่มีความถนัด และไม่มีความถนัด ฯลฯ

การหาอำนาจจำแนกนั้น ต้องมีเกณฑ์ภายนอก (External Criterion) หรือเกณฑ์อิสระ (Independent Criterion) เป็นตัวพยากรณ์บ่งชี้ถึงความแตกต่างที่เห็นชัด

แต่การทดสอบผลสัมฤทธิ์ทางการศึกษาโดยทั่วไปแล้วยากที่จะหาเกณฑ์เมื่อตัดสินให้เห็นชัดได้ จึงมักใช้เกณฑ์ภายใน (Internal Criterian) เป็นต้นว่าคะแนนรวมของแบบทดสอบ โดยศึกษาหาความสัมพันธ์ระหว่างข้อสอบแต่ละข้อกับแบบทดสอบทั้งฉบับ ถ้าแบบทดสอบใดมีค่าสหสัมพันธ์ระหว่างข้อสอบแต่ละข้อกับแบบทดสอบทั้งฉบับสูง แสดงว่ามีค่าอำนาจดี แนวความคิดที่ใช้เกณฑ์ดังกล่าวก็เนื่องจากว่าแบบทดสอบที่มีความเที่ยงตรงในการวัด จะต้องวัดในสิ่งที่ต้องการวัด ดังนั้นข้อสอบใดที่มีความสัมพันธ์ทางบวกกับคะแนนรวมของแบบทดสอบ แสดงว่าข้อสอบนั้นวัดความรู้ในเรื่องที่ต้องการวัด จึงน่าจะมีอำนาจจำแนกผู้มีความรู้จากผู้ไม่มีความรู้ ด้วยเหตุผลดังกล่าว เทคนิคการหาอำนาจจำแนกทางสถิติ จึงใช้วิธีที่นิยม 3 วิธี คือ

(1) วิธีไบซีเรียล (Biserial Correlation)

(2) วิธีพอยท์-ไบซีเรียล (Point-Biserial Correalation)

(3) วิธีแบ่ง กลุ่มสูง-กลุ่มต่ำ 27 เปอร์เซนต์ (Upper-Lower 27 Percent Method)

(1) สูตรสำหรับหาค่าอำนวจจำแนกโดยวิธีไบซีเรียล (Biserial Correlation Cocfficient : r_{bis})

$$r_{bis} = \frac{Mr-Mw}{St} \times \frac{P(1-p)}{y} \quad (5)$$

เมื่อ

r_{bis} = ค่าสัมประสิทธิ์สหสัมพันธ์

M_r = ค่าเฉลี่ยของคะแนนรวมสำหรับนักเรียนที่ตอบถูก

M_w = ค่าเฉลี่ยของคะแนนรวมสำหรับนักเรียนที่ตอบผิด

S_t = ความเบี่ยงเบนมาตรฐานของคะแนนรวมทั้งหมด

P = สัดส่วนของนักเรียนที่ตอบถูกหรือค่าความยากของข้อสอบ

Y = ค่า Ordinate ของโค้งปกติตรงจุดที่แบ่ง และ 1 - P

ตัวอย่างที่ 2.12 แสดงการแจกแจงคะแนนรวมของนักเรียนที่เลือกตอบถูกและผิดในข้อ 15

	จำนวนนักเรียนที่เลือก		
คะแนนรวม	ข้อผิด	ข้อถูก	รวม
15		3	3
14	1	3	4
13	1	8	9
12	4	16	20
11	11	31	42
10	14	38	54
9	34	24	58
8	57	23	80
7	49	12	61
6	30	5	35
5	13	1	14
4	10		10
3	5		5
2	2		2
1	3		3
รวม	236	164	400

$M_w = 7.49$ $\quad p = .41$ $\quad M_t = 8.48$

$M_t = 9.90$ $\quad y = .389$ $\quad S_t = 2.36$

$$r_{bis} = \frac{9.90-7.49}{2.36} \times \frac{(.41)(.59)}{.389}$$

$$= .64$$

ค่า r_{bis} เท่ากับ .64 แสดงว่าข้อที่ 15 นี้ มีค่าอำนาจการจำแนกค่อนข้างสูง จัดว่าเป็นข้อสอบที่พอใช้ได้

(2) สูตรสำหรับหาค่าอำนาจจำแนกโดยวิธีพอยท์-ไบซีเรียล

(Point-Biserial Correlation Coefficient : r_{pbis})

$$r_{pbis} = \frac{M_r - M_w}{S_t} \times \sqrt{P(1-p)} \quad (6)$$

เมื่อ r_{pbis} = ค่าสัมประสิทธิ์สหสัมพันธ์

M_r = ค่าเฉลี่ยของคะแนนรวมสำหรับนักเรียนที่ตอบถูก

M_w = ค่าเฉลี่ยของคะแนนรวมสำหรับนักเรียนที่ตอบผิด

S_t = ความเบี่ยงเบนมาตรฐานของคะแนนรวมทั้งหมด

p = สัดส่วนของนักเรียนที่ตอบถูกหรือค่าความยากของข้อสอบ

(3) วิธีแบ่งกลุ่มสูง-กลุ่มต่ำ 27 เปอร์เซนต์

(Upper-Lower 27 Percent Method)

โดยการนำข้อสอบของนักเรียนที่ได้นักเรียนสูงสุดร้อยละ 27 และของนักเรียนที่ได้คะแนนต่ำสุดร้อยละ 27 หรือ ร้อยละ 54 ของนักเรียนทั้งหมดมาทำการวิเคราะห์ การวิเคราะห์ โดยระบบ 27-27 นี้ต้องมีข้อตกลงว่า ผลการสอบของนักเรียนกระจายเป็นโค้งปกติ มีผู้เสนอแนะว่าการใช้วิธีนี้ควรจะมีจำนวนนักเรียนอย่างน้อย 100 คน จึงจะทำให้ค่าต่าง ๆ ที่วิเคราะห์ออกมามีความถูกต้อง นั่นหมายความว่า จำนวนนักเรียนที่สอบจะต้องไม่น้อยกว่า 370คน หากนักเรียนน้อยกว่านี้ การใช้เทคนิค 27 เปอร์เซนต์ หรือระบบ 27-27 จะได้ค่าสถิติอย่างประมาณที่ ไม่ใกล้เคียงกับค่าที่แท้จริง

วิธีวิเคราะห์โดยวิธีของจอห์นสัน

จอห์นสัน (Johnson) ได้เสนอวิธีวิเคราะห์อย่างง่าย ซึ่งสะดวกในการวิเคราะห์ข้อสอบที่ใช้การทดสอบในชั้นเรียนทั่วไป โดยเทคนิคการแบ่งกลุ่มสูง-กลุ่มต่ำ ร้อยละ 27 แล้วคำนวณดังสูตรง่าย ๆ ต่อไปนี้

$$r = \frac{R_H - R_L}{N_H \text{ หรือ } N_L} \qquad (7)$$

$$p = \frac{100}{N_H+N_L}(R_H + R_L) \qquad (7)$$

เมื่อ R_H , R_L คือ จำนวนคนในกลุ่มสูง กลุ่มต่ำ ตอบถูกตามลำดับ

N_H , N_L คือ จำนวนคนทั้งหมดในกลุ่มสูง กลุ่มต่ำตามลำดับ

r คือ ค่าอำนาจจำแนก

P คือ ค่าความยากง่าย

ค่าอำนาจจำแนก และค่าความยากที่คำนวณจากสูตรดังกล่าว ไม่คำนึงถึงว่านักเรียนจะเลือกตอบ หรือ เว้นว่างทำไม่ทัน แต่ในกรณีที่เป็นข้อสอบแบบเร่งรีบ (Speed Test)จะต้องคำนึงถึงข้อที่นักเรียนตอบไม่ทันหรือเว้นว่างไว้ โดยใช้สูตร

$$r = \frac{R_H}{N_H-B_H} - \frac{R_L}{N_L-B_L} \qquad (8)$$

$$p = \frac{100}{(N_H+N_L) - (B_H+B_L)} (R_H+R_L)$$

เมื่อ B_H = จำนวนข้อที่ทำไม่ทันหรือเว้นว่างไว้ในกลุ่มสูง

B_L = จำนวนข้อที่ทำไม่ทันหรือเว้นว่างไว้ในกลุ่มต่ำ

2.12.3 <u>การตีความค่าความยากง่าย</u> (R)

ระดับความยากของข้อสอบจะมีค่าอยู่ระหว่าง 0 ถึง 1 หรือ 0 ถึง 100 % ข้อสอบที่มีค่า P มาก หมายความว่า ข้อสอบข้อนั้นมีคนเลือกตอบถูกเป็นจำนวนมาก แสดงว่าข้อสอบง่าย

ข้อสอบที่มีค่า P น้อย หมายความว่า ข้อสอบนั้นมีคนเลือกตอบถูกเป็นจำนวนน้อย แสดงว่าข้อสอบยาก

การแปลความหมายของระดับความยากของข้อสอบ (8)

ค่าอำนาจจำแนก	ความหมาย
.81 - 1.00	ง่ายมาก
.61 - .80	ง่าย
.51 - .60	ค่อนข้างง่าย
.50	ยากง่ายพอเหมาะ
.40 - .49	ค่อนข้างยาก
.20 - .39	ยาก
.00 - .19	ยากมาก

ถ้าพิจารณาตามประเภทของข้อสอบแล้ว นักวัดผลบางท่านได้เสนอแนะว่า ข้อสอบประเภทถูกผิด (True-False Item) ค่า P ควรอยู่ระหว่าง .60 - .95 ส่วนข้อสอบปรประเภทหลายตัวเลือก (Multiple Choice Item) เช่น 3 ตัวเลือก ควรมีค่า P ระหว่าง .45 - .95 และถ้า 4 ตัวเลือก ควรมีค่า P ระหว่าง .35 - .85 เป็นต้น ส่วนในกรณีที่ใช้แบบทดสอบเพื่อคัดเลือก ก็จำเป็นต้องมีข้อสอบที่ยาก หรือ ค่อนข้างยากเป็นส่วนใหญ่ เพื่อคัดเลือกผู้ที่มีความรู้สูงเท่านั้น

2.12.4 <u>การตีความค่าอำนาจจำแนก</u> (r)

ค่าอำนสจจำแนกสำหรับข้อสอบที่ถูก จะมีค่าได้ระหว่าง -1.00 ถึง 1.00 แต่ค่าที่ถือว่ามีอำนาจจำแนกดี จะต้องเป็นค่าบวก (+) แสดงว่ากลุ่มสูงหรือกลุ่มเก่งตอบได้มากกว่ากลุ่มต่ำ ตรงกันข้าม ถ้ากลุ่มสูงตอบน้อยกว่ากลุ่มต่ำ ค่าอำนาจจำแนกจะเป็นลบ (-) ซึ่งถือว่าข้อสอบนั้นจำแนกผิดทิศทาง แต่ในกรณีที่เป็นตัวลวงก็น่าจะลวงคนในกลุ่มต่ำมากกว่าคนในกลุ่มสูง ค่า r ที่ได้จึงควรเป็นลบ (-) จึงถือว่าเป็นตัวลวงที่ดี สำหรับค่าอำนาจจำแนกของข้อสอบ

ที่ยอมรับกันว่าพอใช้ได้ คือ ตั้งแต่ +.20 ขึ้นไป

ดังนั้นในการเลือกข้อสอบไว้ใช้คัดเลือกข้อที่มีค่า r เป็นบวกสูง ๆ Ebel ได้ให้หลักเกณฑ์ของข้อสอบที่ครูสร้างขึ้นใช้เองในห้องเรียนว่า ข้อสอบที่มีค่าอำนาจจำแนกระดับใดควรจะปรับปรุงแก้ไขดังนี้

ค่าอำนาจจำแนก	การประเมินผล
.40 ขึ้นไป	จัดเป็นข้อสอบที่ดีมาก
.30 - .39	จัดเป็นข้อสอบที่ดีแต่ควรจะนำไปปรับปรุง
.20 - .29	จัดเป็นข้อสอบที่พอใช้ได้ควรนำไปปรับปรุงใหม่อีก
ต่ำกว่า .19	จัดเป็นข้อสอบที่ไม่ดี ไม่ควรใช้