วิธีการวิเคราะห์ข้อสอบ

Item Analysis Method

สิรินธร สินจินดาวงศ์*

บทคัดย่อ

วิธีการวิเคราะห์ข้อสอบมี 2 ประเภทคือ การวิเคราะห์หาคุณภาพรายข้อ และการวิเคราะห์หาคุณภาพทั้งฉบับ สำหรับการวิเคราะห์หาคุณภาพรายข้อ สามารถวิเคราะห์ได้ 2 ลักษณะ คือ การหาค่าความยาก เป็นการหาตัวเลขที่ใช้บอกความยากง่ายของข้อสอบ ซึ่งคำนวณจากสัดส่วนของผู้สอบที่ตอบถูกต้อง และการหาค่าอำนาจจำแนกเป็นการแยกแยะผู้สอบออกเป็นกลุ่มเก่งกับกลุ่มอ่อน ส่วนการวิเคราะห์หาคุณภาพทั้งฉบับสามารถวิเคราะห์ได้ 2 ด้าน คือ หาค่าความเชื่อมั่น เป็นการหาความคงที่ในการวัดผล ที่เกิดจากการวัดผลในสถานการณ์และเงื่อนไขเดียวกัน และหาค่าความเที่ยงตรง เป็นความสามารถในการตรวจสอบผลการวัดได้ว่าความถูกต้อง แม่นยำ ดังนั้น ข้อสอบที่ใช้ในการวัดผลการเรียนรู้ จะต้องมีความถูกต้องและน่าเชื่อถือ ซึ่งเครื่องมือวัดที่ดี จึงจะตัดสินคุณค่าของสิ่งที่จะวัดได้ การหาประสิทธิภาพของเครื่องมือ จึงใช้วิธีทางสถิติเข้ามาช่วยในการวิเคราะห์ให้มีความเชื่อถือมากยิ่งขึ้น

Abstract

There are two kinds of item analysis, one is analyzing each item and the other one is analyzing the whole test. Each item analysis is used for finding difficulty in the numeric of the test, and discrimination indices were used to find out the high and low group. The whole test analysis is used for finding reliability coefficient which resulted in a consistency in measurement given the same situation and condition. On the other hand, the validity of the test was used to evaluate the result of the measurement accuracy. Therefore, it is necessary to analyze and have an achievement test. So the test that was analyzed would be an accurate and effective measurement.

* อาจารย์ประจำสำนักประกันคุณภาพ มหาวิทยาลัยศรีปทุม

บทนำ

การจัดการเรียนการสอน จำเป็นต้องมีการวัดผลเพื่อทราบการพัฒนาของผู้เรียน ว่าเกิดการเรียนรู้หรือไม่ ซึ่งการวัดผลที่นิยมใช้ ได้แก่ การวัดผลแบบอิงเกณฑ์ (Criterion - Referenced test) และการวัดผลแบบอิงกลุ่ม (Norm-Referenced test) การวัดผลแบบอิงเกณฑ์มีวัตถุประสงค์เพื่อต้องการทราบว่าผู้เรียนเกิดความรอบรู้หรือเรียนผ่านเกณฑ์ที่กำหนดไว้หรือไม่ ส่วนการวัดผลแบบอิงกลุ่ม เป็นการวัดผลว่าผู้เรียนอยู่ในระดับใดเมื่อเทียบกับกลุ่ม การวัดผลจะได้ข้อมูลที่ถูกต้องชัดเจนเพียงใดขึ้นอยู่กับเครื่องมือที่ใช้วัด สำหรับการวัดความรู้นิยมใช้ข้อสอบ (Item) เป็นเครื่องมือ ซึ่งข้อสอบที่ใช้วัดผล จำเป็นต้องมีคุณภาพ จึงจะสามารถวัดผลออกมาได้อย่างถูกต้อง จึงจำเป็นต้องผ่านการวิเคราะห์ ซึ่งในการวิเคราะห์ข้อสอบแบ่งออกเป็นการวิเคราะห์ข้อสอบก่อนนำข้อสอบไปใช้และการวิเคราะห์ข้อสอบหลังจากการนำไปใช้ โดยมีจุดมุ่งหมายเพื่อหาประสิทธิภาพและค้นหาข้อบกพร่องของข้อสอบ เพื่อเป็นข้อมูลในการปรับปรุงข้อสอบและการเรียนการสอน การหาคุณภาพของข้อสอบคือ การหาคุณภาพรายข้อและคุณภาพทั้งฉบับ เพื่อสร้างแบบทดสอบให้มีมาตรฐานและจะนำไปใช้ประโยชน์ต่อผู้สอน ผู้เรียนเองและต่อการพัฒนาหลักสูตรการเรียนการสอนต่อไป

การวิเคราะห์ข้อสอบรายข้อ

การวิเคราะห์ข้อสอบรายข้อ เพื่อให้ทราบคุณภาพของข้อสอบ แต่ละข้อว่าจะสามารถวัดผลได้ถูกต้องเพียงใด ได้แก่ การวิเคราะห์หาค่าความยาก (Difficulty) และค่าอำนาจจำแนก (Discrimination) ของข้อสอบ

ประโยชน์ของความยาก (Difficulty) ของข้อสอบ

ข้อสอบที่ดีจะต้องมีความยากพอเหมาะ คือ ประกอบด้วยข้อที่ยาก ง่าย ค่อนข้างยาก และค่อนข้างง่าย ประกอบอยู่ด้วยกัน ซึ่งการพิจารณาความยากของข้อสอบขึ้นอยู่กับวัตถุประสงค์ของการสอบ ถ้าเป็นข้อสอบเพื่อคัดเลือกบุคคล ควรจะมีความยากค่อนข้างสูง ส่วนข้อสอบเพื่อพิจารณาว่าบุคคลเกิดการเรียนรู้หรือไม่ ควรเป็นข้อสอบระดับปานกลาง ข้อสอบที่ดีควรมีความยากตั้งแต่ 0.2 - 0.8

โดยทั่วไปนิยมคำนวณหาค่าความยากจากสูตรดัชนีค่าความยากหรือดัชนีค่าความง่าย (Difficulty index or Easiness index) Hopkins and Antes, 1985.

วิธีที่ 1 ดัชนีค่าความยาก

$$P_D = \frac{N_w}{N_t}$$

P_D แทน ดัชนีค่าความยากของข้อสอบ
N_w แทน จำนวนผู้ที่ตอบข้อนั้นผิด
N_t แทน จำนวนผู้ที่ทำข้อสอบข้อนั้นทั้งหมด

วิธีที่ 2 ดัชนีค่าความง่าย

$$P_E = \frac{N_r}{N_t}$$

P_E แทน ดัชนีค่าความง่ายของข้อสอบ
N_r แทน จำนวนผู้ที่ตอบข้อนั้นถูก
N_t แทน จำนวนผู้ที่ทำข้อสอบข้อนั้นทั้งหมด

โดยการแปลความหมายของดัชนีความยาก (P_D) และดัชนีความง่าย (P_E) ได้ดังนี้

ความยาก (P_D)	=	.00 - .19	ง่ายมาก
	=	.20 - .39	ค่อนข้างง่าย
	=	.40 - .60	ยากปานกลาง
	=	.61 - .80	ค่อนข้างยาก
	=	.81 - 1.00	ยากมาก
ความง่าย (P_E)	=	.00 - .19	ยากมาก
	=	.20 - .39	ค่อนข้างยาก

= .40 - .60 ยากปานกลาง
= .61 - .80 ค่อนข้างง่าย
= .81 - 1.00 ง่ายมาก

ถ้าวิเคราะห์ข้อสอบแล้วข้อใด มีค่าความยากในระดับยากมากหรือง่ายมาก ควรตัดทิ้งหรือปรับปรุงข้อคำถามและตัวเลือกใหม่

ตัวอย่าง

ข้อที่	ผู้ที่ตอบถูก (คน)	ผู้ที่ตอบผิด (คน)	จำนวนผู้เข้าสอบทั้งหมด
1	25	25	50
2	45	5	50
3	35	15	50
4	20	30	50
5	5	45	50

ตัวอย่าง การคำนวณหาค่าดัชนีความยาก จากผู้เข้าสอบ 50 คน

ข้อ 1 $P_D = \frac{25}{50} = 0.50$ ยากปานกลาง

ข้อ 2 $P_D = \frac{5}{50} = 0.10$ ง่ายมาก

ข้อ 3 $P_D = \frac{15}{50} = 0.30$ ค่อนข้างง่าย

ข้อ 4 $P_D = \frac{30}{50} = 0.60$ ยากปานกลาง

ข้อ 5 $P_D = \frac{45}{50} = 0.90$ ยากมาก

ดังนั้น จะเห็นได้ว่า ข้อสอบข้อ 2 และข้อ 5 ควรตัดทิ้ง เนื่องจากง่ายมากและยากมาก จึงไม่ควรใช้หรือถ้าจะนำมาใช้ต้องปรับปรุงแล้วนำไปหาคุณภาพอีกครั้งหนึ่งก่อนนำไปใช้

วิธีการหาค่าอำนาจจำแนก (Discrimination)

การหาค่าอำนาจจำแนกเป็นการพิจารณาแยกแยะผู้สอบ ออกเป็นกลุ่มเก่งและกลุ่มอ่อน พิจารณาจากคะแนนสอบว่ามีการกระจายมากน้อยเพียงไร ค่าอำนาจจำแนก (r) มีค่าอยู่ระหว่าง -1.00 ถึง 1.00 โดยค่าอำนาจจำแนกรายข้อ ต้องมีค่ามากกว่า หรือเท่ากับ 0.20 ($r \geq 0.20$) หรือ .20 ถึง 1.00 จึงจะถือว่าใช้ได้

ถ้าอำนาจจำแนก r เป็นบวก + แสดงว่า การตอบข้อคำถามข้อนั้นคนเก่งตอบถูก คนอ่อนตอบผิด

แต่ถ้าอำนาจจำแนก r เป็นลบ - แสดงว่า การตอบข้อคำถามข้อนั้นคนอ่อนตอบถูก คนเก่งตอบผิด

แต่ถ้าอำนาจจำแนก r เป็นศูนย์ หรือใกล้ศูนย์ แสดงว่าไม่มีอำนาจจำแนก คนเก่งและคนอ่อนตอบถูกใกล้เคียงกัน

การแปลความหมายของของค่าอำนาจจำแนก (r) ได้ดังนี้

- r มีค่าระหว่าง .20 ถึง 1.00 แสดงว่า จำแนกได้
- r มีค่าระหว่าง -.19 ถึง +.19 แสดงว่า จำแนกไม่ได้
- r มีค่าระหว่าง -0.20 ถึง -1.00 แสดงว่า จำแนกกลับ(ใช้ไม่ได้)

จะมีวิธีใดบ้างที่จะหาค่าอำนาจจำแนกของข้อสอบ

วิธีที่ 1 จะใช้กับข้อสอบที่ตรวจให้คะแนนชนิดถูกได้ 1 คะแนน ผิดได้ 0 คะแนนเท่านั้น โดยแบ่งกลุ่มสูงและกลุ่มต่ำ (Kelley, 1939) เป็น 27% หรือ 33% หรือ 50% โดยถ้าผู้สอบมีจำนวนมากกว่า 100 คนขึ้นไป มักใช้เทคนิค 27% แต่ถ้าต่ำกว่า 100 คน ควรใช้เทคนิค 50% หรือ 33% ของจำนวนทั้งหมดในการแบ่งกลุ่มสูง หรือกลุ่มต่ำ (Cureton,1957) ซึ่งมีสูตรคำนวณดังนี้

$$D = \frac{U}{n_U} - \frac{L}{n_L}$$

D แทน ดัชนีค่าอำนาจจำแนกของข้อสอบ
U แทน จำนวนผู้สอบที่ตอบถูกในกลุ่มคะแนนสูง
L แทน จำนวนผู้สอบที่ตอบถูกในกลุ่มคะแนนต่ำ
n_U แทน จำนวนผู้สอบทั้งหมดในกลุ่มคะแนนสูง
n_L แทน จำนวนผู้สอบทั้งหมดในกลุ่มคะแนนต่ำ

วิธีที่ 2 ค่าสหสัมพันธ์แบบพอยต์ไบซีเรียล (Point biserial Correlation : $r_{p.bis}$) วิธีนี้มีข้อตกลงเบื้องต้นว่าข้อสอบแต่ละข้อจะต้องให้คะแนนโดยทำถูกได้ 1 และทำผิดได้ 0 เท่านั้น

จากสูตร $r_{p.bis} = \frac{\mu_p - \mu}{\sigma}\sqrt{\frac{p}{q}}$

เมื่อ $r_{p.bis}$ แทน ดัชนีค่าอำนาจจำแนกแบบพอยต์ไบซีเรียล
μ_p แทน คะแนนเฉลี่ยของกลุ่มที่ทำข้อนั้นถูก
μ แทน คะแนนเฉลี่ยของข้อสอบ
σ แทน คะแนนเบี่ยงเบนมาตรฐานของข้อสอบ
P แทน สัดส่วนของผู้สอบที่ทำข้อนั้นถูก
q แทน สัดส่วนของผู้ที่ทำข้อนั้นผิดเท่ากับ (1 - P)
n_p แทน จำนวนผู้สอบที่ตอบถูก
n_q แทน จำนวนผู้สอบที่ตอบผิด

วิธีที่ 3 ค่าสหสัมพันธ์แบบไบซีเรียล (Biserial Correlation : r_{bis}) มีข้อจำกัดว่า ข้อมูลจะมีการแจกแจงแบบโค้งปกติเท่านั้น และกลุ่มผู้สอบจำนวนมากกว่า 100 คนขึ้นไป

จากสูตร $r_{bis} = \frac{\mu_p - \mu}{\sigma} \times \frac{p}{y}$

วิธีที่ 4 ใช้การเปิดตารางจุง-เตห์-ฟาน (Chung-Teh-Fan) โดยต้องแบ่งกลุ่มผู้สอบออก 2 กลุ่ม คือ กลุ่มเก่งและกลุ่มอ่อน ใช้เทคนิค 27% คำนวณหาค่า P_H กลุ่มเก่ง และ P_L กลุ่มอ่อน แล้วไปเปิดตารางหาค่าอำนาจจำแนก (r) ค่าความยาก (P) และค่าความยากมาตรฐาน (Δ) เงื่อนไข ข้อสอบนั้นต้องตรวจให้คะแนน ถูก 1 ผิด 0 โดยต้องมีจำนวนผู้สอบมากกว่า 100 คน

ตัวอย่าง การหาค่าความยาก อำนาจจำแนก และความยากมาตรฐานโดยวิธีของ จุง-เตห์-ฟาน จากผู้สอบ 100 คน

กลุ่มเก่ง (P_H)					กลุ่มอ่อน (P_L)				
ผู้สอบคนที่	ข้อที่ 1				ผู้สอบคนที่	ข้อที่ 1			
	ก	ⓑ	ค	ง		ก	ⓑ	ค	ง
1	1	0	0	0	1	0	1	0	0
2	0	1	0	0	2	0	0	1	0
3	0	0	1	0	3	1	0	0	0
...					...				
รวม	**3**	**20**	**1**	**2**	**รวม**	**8**	**10**	**4**	**5**
$P_H = R_H / n_H$	.11	.74	.04	.07	$P_L = R_L / n_L$	.30	.37	.15	.19

หมายเหตุ : จำนวนผู้สอบในกลุ่มสูงและกลุ่มต่ำ = 27 คน

ในข้อสอบข้อที่ 1 นี้ คำตอบที่ถูก คือ ข้อ ข. โดยค่า $P_H = .74$ และ $P_L = .37$ นำไปเปิดตาราง จุง-เตห์-ฟาน ที่ P_H.74 และ P_L.37

ได้ค่าความยาก $P = .56$ ค่าอำนาจจำแนก $r = .38$ ค่าความยากมาตรฐาน $\Delta = 12.4$ และเมื่อนำทั้งตัวถูกและตัวลวงมาวิเคราะห์ได้ดังตารางต่อไปนี้ คือ

ข้อที่	ตัวเลือก	P_L	P_H	P	R	Δ
1	ก	.30	.11	.20	.28	16.4
	ⓑ	.37	.74	.56	.38	12.4
	ค	.15	.04	.09	.28	18.4
	ง	.19	.07	.13	.24	17.6

❋ สำหรับตัวลวงที่ดี ควรมีคนเลือกตอบมากกว่า 5% และกลุ่มเก่งเลือกตอบน้อยกว่ากลุ่มอ่อน

โดยค่า Δ (Delta) ค่าความยากมาตรฐาน พิจารณาได้ดังนี้

- Δ น้อยกว่า 13.00 ค่อนข้างง่าย
- Δ เท่ากับ 13.00 ปานกลาง
- Δ มากกว่า 13.00 ค่อนข้างยาก

ประสิทธิภาพของการทำนายค่าอำนาจจำแนก (E)

เพื่อจะบอกถึงระดับความสามารถในการจำแนกคนเก่งหรือคนอ่อน เมื่อทำข้อสอบได้ถูกหรือผิดนั้น คำนวณจาก $E=100\left\{1-\sqrt{1-r^2}\right\}$

การวิเคราะห์ข้อสอบอัตนัย

ข้อสอบอัตนัย ไม่ได้ให้คะแนน ถูก 1 ผิด 0 จึงไม่สามารถวิเคราะห์ได้ตามวิธีดังที่กล่าวมาแล้วข้างต้น ดังนั้นจึงวิเคราะห์โดยใช้สูตร D.R Whitney and D.L Sabers, 1970 ต้องแบ่งกลุ่มเก่ง กลุ่มอ่อน โดยใช้เทคนิค 25% ของกลุ่มผู้ได้คะแนนสูงและคะแนนต่ำ ดังนี้

1. ดัชนีค่าความง่าย (P_E) จากสูตร

$$P_E=\frac{S_U+S_L-(2N\times X_{\min})}{2N(X_{\max}-X_{\min})}$$

2. ดัชนีค่าอำนาจจำแนก คำนวณจากสูตร

$$D=\frac{S_U-S_L}{N(X_{\max}-X_{\min})}$$

เมื่อ P_E แทน ดัชนีค่าความง่าย
D แทน ดัชนีค่าอำนาจจำแนก
S_U แทน ผลรวมของคะแนนกลุ่มเก่ง
S_L แทน ผลรวมของคะแนนกลุ่มอ่อน
N แทน จำนวนผู้เข้าสอบของกลุ่มเก่งหรือกลุ่มอ่อน
$X_{\max}$ แทน คะแนนสูงสุด
$X_{\min}$ แทน คะแนนต่ำสุด

สรุปได้ว่า การหาคุณภาพของข้อสอบรายข้อ โดยพิจารณาค่าความยาก อยู่ระหว่าง 0.20 ถึง 0.80 และค่าอำนาจจำแนกมากกว่า 0.20 ขึ้นไป สำหรับข้อสอบปรนัยควรพิจารณาทั้งตัวถูกและตัวลวง

การวิเคราะห์คุณภาพแบบทดสอบทั้งฉบับ

การวิเคราะห์คุณภาพของแบบทดสอบทั้งฉบับ ได้แก่ การหาค่าความเชื่อมั่น (Reliability) และการหาค่าความเที่ยงตรง (Validity)

ความเชื่อมั่น (Reliability)

ความเชื่อมั่น (Reliability) คือ ความคงที่ในผลการวัด (Consistency) ไม่ว่าจะวัดกี่ครั้งผลจะต้องเท่ากัน ภายใต้สถานการณ์และเงื่อนไขเดียวกัน (Same condition)

การวิเคราะห์หาค่าความเชื่อมั่นของแบบทดสอบกระทำได้ 2 กรณี ดังนี้ คือ

1. การหาความเชื่อมั่นโดยความสอดคล้องจากการสอบซ้ำ ได้แก่ วิธีการสอบซ้ำ (Test -retest) แบบทดสอบคู่ขนาน (Parallel form)

2. การหาค่าความเชื่อมั่นจากความสอดคล้องภายใน ได้แก่ วิธีแบ่งครึ่งฉบับ (Split-half method) วิธี Kuder Richardson Procudure (KR-20, KR-21) วิธี Cornbach Alpha Procudure

1. การหาความเชื่อมั่นโดยความสอดคล้องจากการสอบซ้ำ

1.1) การสอบซ้ำ (Test -retest Method) เป็นการนำแบบทดสอบชุดเดียวกันไปสอบกับผู้สอบกลุ่มเดียวกัน 2 ครั้ง แล้วคำนวณหาความสัมพันธ์ของคะแนนจากสูตรสหสัมพันธ์ของเพียร์สัน (Pearson Product Moment Coefficient Correlation)

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

เมื่อ r_{XY} แทน ค่าความเชื่อมั่นของแบบทดสอบ โดยการสอบซ้ำ
$\Sigma X, \Sigma Y$ แทน ผลรวมจากคะแนนจาก

การสอบครั้งที่ 1 และครั้งที่ 2 ของกลุ่ม ที่ทดสอบกลุ่มเดียวกัน

$\Sigma X^2, \Sigma Y^2$ แทน ผลรวมกำลังสองของคะแนนจากการสอบครั้งที่ 1 และครั้งที่ 2

ΣXY แทน ผลรวมของผลคูณระหว่างคะแนนสอบครั้งที่ 1 และครั้งที่ 2 ของแต่ละคน

N แทน จำนวนคนที่สอบแบบทดสอบฉบับนั้น

ตัวอย่าง ในตารางต่อไปนี้ เป็นผลการสอบจากวิธีการสอบซ้ำ โดยการแทนค่าลงใน สูตรของเพียร์สัน ได้ดังนี้

ผู้สอบ	คะแนนสอบครั้งที่ 1 (X)	คะแนนสอบครั้งที่ 2 (Y)
1	16	27
2	21	33
3	18	26
4	25	31
5	23	37
6	17	29
7	19	25
8	24	32
9	16	28
10	24	34

เมื่อ $N = 10$ $\Sigma XY = 6,219$

$\Sigma X = 203$ $\Sigma X^2 = 4,233$

$\Sigma Y = 302$ $\Sigma Y^2 = 9,254$

ดังนั้น แทนค่าในสูตรจะได้

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

$$r_{XY} = \frac{(10\times 6,219) - (203\times 302)}{\sqrt{\left[(10\times 4,233) - (233)^2\right]\left[(10\times 9,254) - (302)^2\right]}}$$

$$= \frac{884}{1223.79}$$

$$= 0.72$$

จากตัวอย่าง ค่าความเชื่อมั่นของแบบทดสอบชุดนี้ คำนวณโดยวิธีการสอบซ้ำเท่ากับ 0.72

แบบทดสอบที่มีความเชื่อมั่น ใกล้เคียง 1 จะมีค่าความเชื่อมั่นสูง

ข้อเสีย ของการคำนวณโดยวิธีการสอบซ้ำ ได้แก่

1. การสอบโดยใช้แบบทดสอบฉบับเดียวกันหลายๆ ครั้ง ผู้สอบจะเกิดการเบื่อหน่าย จำข้อสอบได้
2. เสียเวลาในการสอบมาก (ต้องเว้นระยะการสอบครั้งที่ 1 และครั้งที่ 2 ให้เหมาะสม) ไม่ควรเกิน 3 สัปดาห์
3. ผู้สอบเกิดการเรียนรู้มากขึ้น จากการสอบครั้งแรก จึงเกิดความคลาดเคลื่อน

1.2) วิธีใช้แบบทดสอบคู่ขนาน (Parallel Tests or Equivalence Tests) เป็นการสอบโดยการใช้แบบทดสอบ 2 ชุด ที่มีคุณลักษณะใกล้เคียงกัน ทั้งด้านเนื้อหา ความยาก และค่าอำนาจจำแนก การหาค่าความเชื่อมั่นโดยการนำแบบทดสอบทั้ง 2 ชุด ไปสอบกับกลุ่มที่ศึกษากลุ่มเดียวกัน และนำมาหาค่าความเชื่อมั่น โดยใช้สูตรเดียวกับการสอบซ้ำ โดยที่ X แทน แบบทดสอบชุดที่ 1 และ Y แทน แบบทดสอบชุดที่ 2

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

โดยที่ r_{XY} แทน ความเชื่อมั่นของแบบทดสอบคู่ขนาน

ข้อจำกัดของแบบทดสอบคู่ขนาน คือ สร้างยากที่ให้มีลักษณะใกล้เคียงกันทั้งเนื้อหา ความยากและอำนาจจำแนก

2. การหาค่าความเชื่อมั่นจากความสอดคล้องภายใน

2.1) วิธีแบ่งครึ่งฉบับ (Split-half) เป็นการหาค่าความเชื่อมั่นภายใน คือ การใช้แบบทดสอบชุดเดียวสอบครั้งเดียว แต่นำคะแนนของแต่ละคนมาแบ่งครึ่งข้อคู่-ข้อคี่ และนำมาคำนวณ หาค่าความเชื่อมั่นครึ่งฉบับ

การหาความเชื่อมั่นครึ่งฉบับ คำนวณได้จากสูตรของเพียร์สันเช่นกัน คือ

$$r_{\frac{1}{2}} = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 - (\Sigma X)^2\right]\left[N\Sigma Y^2 - (\Sigma Y)^2\right]}}$$

โดยที่ $r_{\frac{1}{2}}$ แทน ค่าความเชื่อมั่นของแบบทดสอบครึ่งฉบับ

เมื่อคำนวณหาค่าความเชื่อมั่นของแบบทดสอบครึ่งฉบับแล้ว จึงมาปรับขยายเป็นความเชื่อมั่นทั้งฉบับจากสูตรของ Spearman -Brown (Mary J.Allen and Wendy M. Yen, 1979)

$$r_{tt} = \frac{2r_{\frac{1}{2}}}{1 + r_{\frac{1}{2}}}$$

โดยที่ r_{tt} แทน ค่าความเชื่อมั่นทั้งฉบับ

2.2) วิธีของคูเดอร์ริชาร์ดสัน (Kuder-Richardson Procedure) เป็นการหาค่าความเชื่อมั่นของแบบทดสอบ โดยหาความคงที่ภายใน (Internal consistency) ได้แก่ สูตร (Kuder and Richardson , 1939) $KR-20$ และ $KR-21$ จำเป็นต้องมีการตรวจให้คะแนน แบบถูกได้ 1 ผิดได้ 0 เท่านั้น

สูตร $KR-20$ คือ

$$r_{tt} = \frac{n}{n-1}\left(1 - \frac{\Sigma pq}{S^2}\right)$$

โดยที่ r_{tt} แทน ค่าความเชื่อมั่นของแบบทดสอบ

n แทน จำนวนข้อของแบบทดสอบ

P แทน สัดส่วนของผู้ตอบถูกจากผู้สอบทั้งหมด

q แทน สัดส่วนของผู้ตอบผิดจากผู้สอบทั้งหมดเท่ากับ $(1 - p)$

S^2 แทน ความแปรปรวนของคะแนนรวม

สูตร $KR-21$ คือ

$$r_{tt} = \frac{n}{n-1}\left\{1 - \frac{\bar{x}(n - \bar{X})}{ns^2}\right\}$$

เป็นสูตรที่แปลงมาจาก $KR-20$ ในการแปลงค่า pq เป็นค่าเฉลี่ยให้ง่ายต่อการคิดคำนวณ

โดยที่ r_{tt} แทน ค่าความเชื่อมั่นของแบบทดสอบ

n แทน จำนวนข้อของแบบทดสอบ

$\bar{X}$ แทน ค่าเฉลี่ยของคะแนนสอบ

S^2 แทน ความแปรปรวนของคะแนนรวม

ตัวอย่าง ข้อมูลจ

ผู้สอบ	
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
ผู้ตอบถูก	
P	
q	
pq	
$\Sigma pq = 1.39$	

คำนวณห

ค่าความเ

มีค่าเท่ากับ 0.7

คำนวณ

r_{tt}

ตัวอย่าง ข้อมูลจากการสอบของผู้สอบ 10 คน จำนวนแบบทดสอบ 8 ข้อ

ผู้สอบ	ข้อคำถาม								คะแนนรวม
	1	2	3	4	5	6	7	8	
1	1	0	0	0	0	0	0	0	1
2	1	1	0	0	0	0	0	0	2
3	1	1	1	0	0	0	0	0	3
4	1	1	1	1	0	0	0	0	4
5	1	1	1	1	1	0	0	0	5
6	1	1	1	1	1	1	0	0	6
7	1	1	1	0	1	1	0	1	6
8	1	1	0	1	1	1	1	0	6
9	1	1	1	1	1	0	1	1	7
10	1	1	1	1	1	1	0	0	6
ผู้ตอบถูก	**10**	**9**	**7**	**6**	**6**	**4**	**3**	**2**	**47**
P	**1.0**	**0.9**	**0.7**	**0.6**	**0.6**	**0.4**	**0.3**	**0.2**	
q	**0**	**0.1**	**0.3**	**0.4**	**0.4**	**0.6**	**0.7**	**0.8**	
pq	**0**	**.09**	**.21**	**.24**	**.24**	**.24**	**.21**	**.16**	
$\Sigma pq = 1.39$	$\overline{X} = 4.70$				$S_t^2 = 4.01$				

คำนวณหาค่าความเชื่อมั่น จากสูตร $KR-20$

$$r_{tt} = \frac{n}{n-1}\left(1-\frac{\Sigma pq}{S^2}\right)$$

$$r_{tt} = \frac{8}{8-1}\left(1-\frac{1.39}{4.01}\right)$$

$$= 0.75$$

ค่าความเชื่อมั่นจากการคำนวณโดยสูตร $KR-20$ มีค่าเท่ากับ 0.75

คำนวณหาค่าความเชื่อมั่น จากสูตร $KR-21$

$$r_{tt} = \frac{n}{n-1}\left\{1-\frac{\overline{X}(n-\overline{X})}{ns^2}\right\}$$

$$r_{tt} = \frac{8}{8-1}\left\{1-\frac{4.70(8-4.70)}{8\times 4.01}\right\}$$

$$= 0.59$$

ค่าความเชื่อมั่นจากการคำนวณโดยสูตร $KR-21$ มีค่าเท่ากับ 0.59

จะเห็นได้ว่าการคำนวณโดยใช้สูตร $KR-21$ จะได้ค่าความเชื่อมั่นต่ำกว่าสูตร $KR-20$ เพราะว่าคำนวณจากค่าเฉลี่ยไม่ได้คำนวณจากสัดส่วนผู้ตอบถูกรายข้อ แต่สามารถคำนวณได้สะดวกกว่า

2.3) วิธีของครอนบัค (Cronbach Alpha Procedure) ในรูปสัมประสิทธิ์สหสัมพันธ์แอลฟา (α-Coefficient) พัฒนามาจากสูตร KR-20 ซึ่ง**ไม่จำเป็น**ต้องตรวจให้คะแนน แบบถูก 1 ผิด 0 ก็ได้ (Cronbach, 1951)

$$\alpha = \frac{n}{n-1}\left\{1-\frac{\Sigma S_i^2}{S_t^2}\right\}$$

โดยที่ α แทน ค่าความเชื่อมั่น
n แทน จำนวนข้อของแบบทดสอบ
S_i^2 แทน ความแปรปรวนรายข้อของแบบทดสอบ
S_t^2 แทน ความแปรปรวนทั้งฉบับ

สรุป การหาความเชื่อมั่นโดยวิธีหาความสอดคล้องภายใน (Internal Consistency) จะใช้เมื่อแบบทดสอบนั้นมีการสอบครั้งเดียวและวัดในสิ่งเดียวกันเท่านั้น การที่จะให้แบบทดสอบมีความเชื่อมั่นสูง ต้องสร้างข้อสอบให้มีจำนวนข้อมาก เนื่องจากจำนวนข้อมาก ทำให้มีความแปรปรวนสูง ความเชื่อมั่นก็จะสูงด้วยและถ้าจำนวนข้อมากก็จะทำให้สามารถวัดได้ครอบคลุมเนื้อหาและพฤติกรรมที่ต้องการวัดนั้นเอง

การวิเคราะห์หาค่าความเที่ยงตรงของแบบทดสอบ (Validity)

ความเที่ยงตรง (Validity) หมายถึง ความสามารถในการวัดสิ่งต่างๆ ที่ต้องการวัดได้ถูกต้อง แม่นยำหรือหมายถึง ความสามารถในการให้ความหมายในสิ่งที่วัดได้อย่างไม่ผิดพลาด (Gulliksen, 1950) ความเที่ยงตรงสามารถแบ่งออกเป็น 3 ประเภท ใหญ่ๆ ดังนี้ คือ

1. ความเที่ยงตรงตามเนื้อหา (Content Validity)
2. ความเที่ยงตรงตามเกณฑ์สัมพันธ์ (Criterion-Related Validity)
3. ความเที่ยงตรงตามโครงสร้าง (Construct Validity)

1.) ความเที่ยงตรงตามเนื้อหา (Content Validity)

ความเที่ยงตรงตามเนื้อหา หมายถึง เครื่องมือที่สามารถวัดได้ตามเนื้อหาที่ต้องการจะวัด พิจารณาวิเคราะห์อย่างมีเหตุผล

1.1 ความเที่ยงตรงเชิงเหตุผล (Logical Validity) ก่อนที่จะมีการวิเคราะห์หาความเที่ยงตรงเชิงเหตุผลนั้น จะต้องสร้างตารางวิเคราะห์รายละเอียด และเขียนจุดมุ่งหมายเชิงพฤติกรรมให้ชัดเจน โดยให้ผู้เชี่ยวชาญเป็นผู้พิจารณาให้ความสอดคล้องของเนื้อหากับจุดมุ่งหมายเชิงพฤติกรรม และความสอดคล้องระหว่างจุดมุ่งหมายเชิงพฤติกรรมกับแบบทดสอบ โดยให้ผู้เชี่ยวชาญอย่างน้อย 5 ท่าน พิจารณาความสอดคล้องเป็น +1 ไม่สอดคล้อง เป็น -1 และ ไม่แน่ใจ เป็น 0 และนำค่าคะแนนการพิจารณามาคำนวณหาค่าดัชนีความสอดคล้อง (Index of Consistency : IOC) ใช้สูตรของโรวิเนลรี และแฮมเบลตัน (Rovinelli and Hambleton, 1977)

$$IOC = \frac{\Sigma R}{N}$$

โดยที่ IOC แทน ดัชนีความสอดคล้องมีค่าอยู่ระหว่าง -1 ถึง +1
ΣR แทน ผลรวมของการพิจารณาของผู้เชี่ยวชาญ
N แทน จำนวนผู้เชี่ยวชาญ

1.2 ความเที่ยงตรงเชิงพินิจ (Face Validity) เป็นการพิจารณาว่าแบบทดสอบสามารถวัดได้ตรงตามนิยามที่กำหนดไว้หรือไม่ เหมาะสมสำหรับแบบทดสอบวัดความรู้สึก วัดบุคลิกภาพ วัดค่านิยมฯลฯ ซึ่งแบบทดสอบวัดผลสัมฤทธิ์เหมาะกับการหาความเที่ยงตรงเชิงเนื้อหาด้านเหตุผลมากกว่า

2) ความเที่ยงตรงตามเกณฑ์สัมพันธ์ (Criterion-Related Validity)

ความเที่ยงตรงตามเกณฑ์สัมพันธ์ หมายถึง ความสัมพันธ์ระหว่างคะแนนที่ผู้สอบทำได้จากแบบทดสอบกับเกณฑ์ภายนอกที่เป็นอิสระ

2.1 ความเที่ยงตรงเชิงสภาพ (Concurrent Validity) เป็นการนำเอาผลการทดสอบจากแบบทดสอบที่สร้างขึ้นไปหาความสัมพันธ์กับเกณฑ์หรือสภาพใน

ปัจจุบัน ได้แก่ การหาความสัมพันธ์ระหว่างแบบทดสอบที่มีความเที่ยงตรงเชิงสภาพ ได้แก่ แบบทดสอบวัดความสามารถทางสมอง (Mental ability) กับแบบทดสอบวัดเชาวน์ปัญญา (Intelligence or I.Q.) คะแนนจากแบบทดสอบปลายภาคของแต่ละวิชากับคะแนน G.P.A. ถ้ามีความสัมพันธ์กันสูง แสดงว่าแบบทดสอบนั้นมีความเที่ยงตรงเชิงสภาพ

2.2 ความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) หมายถึง ความสัมพันธ์ระหว่างคะแนนจากการสอบกับผลที่คาดหวังว่าจะเกิดขึ้นในอนาคต ได้แก่ ความสัมพันธ์จากคะแนนสอบเข้ามหาวิทยาลัยกับความสำเร็จในการศึกษาหรือความสัมพันธ์ระหว่างคนที่สอบเข้าในสาขาวิชากับอาชีพในอนาคต แบบทดสอบที่ใช้ เช่น แบบทดสอบวัดความถนัด (AptitudeTest) ความถนัดทางช่าง วัดแววความเป็นครู ที่เป็นแบบทดสอบมาตรฐานมีความเที่ยงตรงเชิงพยากรณ์

3) ความเที่ยงตรงตามโครงสร้าง (Construct Validity)

ความเที่ยงตรงตามโครงสร้าง (Construct Validity) หมายถึง คุณภาพของเครื่องมือที่สามารถวัดได้ตรงตามลักษณะหรือตามทฤษฎีต่างๆ ของโครงสร้างนั้น (Cronbach & Meehl, 1955)

วิธีการคำนวณหาค่าความเที่ยงตรงตามโครงสร้างสามารถหาได้จากวิธีต่างๆดังนี้

3.1 คำนวณจากค่าความสัมพันธ์ของแบบทดสอบ ที่ต้องการหาความเที่ยงตรงกับแบบทดสอบมาตรฐานที่วัดลักษณะเดียวกัน โดยใช้สูตรสัมประสิทธิ์ สหสัมพันธ์ เช่น แบบทดสอบวัดความสามารถทางสมองกับแบบทดสอบมาตรฐานของ Standford-Binet

3.2 ลักษณะหลากและวิธีหลาย (The Multitrait-Multimethod Matrix) ซึ่งแคมพ์เบลและฟิสต์ (Campbell and Fiske, 1959) ได้กล่าวถึงการวัดความเที่ยงตรงแบบลักษณะหลากวิธีหลายว่าเป็นการวัดแบบทดสอบที่ประกอบด้วยลักษณะที่วัดสองลักษณะหรือมากกว่า และวิธีวัดสองวิธีหรือมากกว่า แบ่งออกเป็น 2 ลักษณะดังนี้

3.2.1 ความเที่ยงตรงเชิงเหมือน (Convergent Validity) เกิดจากความสัมพันธ์ระหว่างผลการวัดลักษณะเดียวกันหรือวิธีวัดเดียวกัน (Reliability of test-retest) และลักษณะเดียวกันแต่ต่างวิธีจะมีความสัมพันธ์กันสูง

3.2.2 ความเที่ยงตรงเชิงจำแนก (Discriminant Validity) เป็นความเที่ยงตรงที่เกิดจากความสัมพันธ์ระหว่างผลการวัดที่ต่างลักษณะกัน ใช้วิธีเดียวกันหรือต่างกัน จะมีความสัมพันธ์กันต่ำ

3.3 วิธีคำนวณจากการวิเคราะห์องค์ประกอบ (Factor Analysis) เป็นการคำนวณหาค่าสหสัมพันธ์ภายใน (Intercorrelation)ของแบบทดสอบแต่ละข้อ หรือแบบทดสอบย่อย วัดว่าแบบทดสอบแต่ละข้อหรือแต่ละฉบับวัดองค์ประกอบเดียวกันหรือไม่ ถ้าคำนวณแล้วมี 1 องค์ประกอบแสดงว่า มีความเที่ยงตรงเชิงโครงสร้าง

3.4 วิธีคำนวณจากกลุ่มรู้ชัด (Known-Group Technique) เป็นการเปรียบเทียบผลของการสอบระหว่างกลุ่มที่รู้ว่ามีลักษณะที่ต้องการวัดกับกลุ่มที่รู้ว่าไม่มีลักษณะที่ต้องการวัด แล้วคำนวณหาค่าเฉลี่ยของทั้ง 2 กลุ่ม โดยใช้ สถิติ t-test แล้วหาค่านัยสำคัญทางสถิติ ถ้าแตกต่างอย่างมีนัยสำคัญทางสถิติ แสดงว่าสามารถวัดได้ตามโครงสร้าง

ความเที่ยงตรงตามโครงสร้างจะใช้คำนวณหาความเที่ยงตรงของแบบทดสอบความถนัดทางการเรียน แบบวัดบุคลิกภาพ แบบวัดจริยธรรม ฯลฯ ที่ไม่มีเนื้อหาที่ชัดเจน

สรุป ความเที่ยงตรงแบ่ง ออกเป็น 3 ประเภทใหญ่ ได้แก่

1. ความเที่ยงตรงตามเนื้อหา ได้แก่ แบบทดสอบวัดผลสัมฤทธิ์ที่สอดคล้องกับจุดมุ่งหมายเชิงพฤติกรรม แบบทดสอบวัดบุคลิกภาพ วัดค่านิยมที่สร้างตามนิยามที่กำหนด เป็นต้น

2. ความเที่ยงตรงตามเกณฑ์สัมพันธ์ ได้แก่ แบบทดสอบวัดเชาว์ปัญญา ที่สัมพันธ์กับคะแนน GPA และแบบทดสอบวัดความถนัด วัดแววความเป็นครูที่สัมพันธ์กับอนาคต เป็นต้น

3. ความเที่ยงตรงตามโครงสร้าง ได้แก่ แบบทดสอบวัดเชาว์ปัญญา วัดความถนัดทางการเรียน วัดจริยธรรมที่สร้างได้ตรงตามลักษณะและทฤษฎีของโครงสร้างนั้น เป็นต้น

ล้วน สายยศ และ
Campbell, D.T.
Multi
Cornbach , L, J
Cronbach, Lee
Cronbach,L.J.,
Bullet
Cureton, E.E. (
Ebel, Robert L
New
Gronlund, N.E.
Gulliksen, H. (1
Gulliksen, Har
Meas
Hopkins, C.D. &
F.E. P
Kelley. T.L. (19
Journ
Kuder. G.F., & P
Mary J. A. & V
Wads
Richardson, M.
Psych

บรรณานุกรม

ล้วน สายยศ และอังคณา สายยศ. (2539). **เทคนิคการวัดผลการเรียนรู้.** กรุงเทพฯ : ชมรมเด็ก.

Campbell, D.T., & Fiske, D.W. (1959). **Convergent and Discriminant Validation by the Multitrait Multimethod Matrix.** Psychological Bulletin.

Cornbach , L, J. (1951). **Coefficient Alpha and the Internal Structure of Tests.** Psychometrika.

Cronbach, Lee J. (1984). **Essentials of Psychological Testing.** 4th ed. New York : Harper & Row.

Cronbach,L.J., & Meehl, P.E. (1955). **Construct Validity in Psychological Tests.** Psychological Bulletin.

Cureton, E.E. (1957). **The Upper and Lower twenty-seven per cent rule.** Psychometrika.

Ebel, Robert L. and Frisbie, David A. (1991). **Essentials of Educational Measurement.** 5th ed. New Jersey : Prentice Hall.

Gronlund, N.E. (1982). **Constructing Achievement Tests.** Englewood Cliffs: Prentice-Hall INC.

Gulliksen, H. (1950). **Theory of Mental Tests.** New York : Wiley.

Gulliksen, Harold. (1986). **Perspective on Educational Measurement.** Applied Psychological Measurement V10(2) June.

Hopkins, C.D. and Antes, R.L. (1985). **Classroom Measurement and Evaluation.** 2nd ed. Itasca, III : F.E. Peacock.

Kelley. T.L. (1939). **The Selection of Upper and Lower groups for the Validation of Test items.** Journal of Educational Psychology.

Kuder. G.F., & Richardson, M.W. (1937). **The Theory of the Estimation of Test Reliability**. Psychometrika.

Mary J. A. & Wendy M.Y. (1979). **Introduction to Measurement Theory**. California : A Division of Wadsworth, Inc.

Richardson, M.W. (1936). **The Relation between the Difficulty and the Differential Validity of a Test**. Psychometrika.