

วารสาร "มนุษยศาสตร์ปริทรรศน์"
ปีที่ ๒ ฉบับที่ ๑ ประจำปีการศึกษา ๒๕๖๓

การประเมินผล ข้อสอบภาษาอังกฤษ แบบปรนัย

บุษกร วิชชุตา
ศิริกุล มาร์แชลล์

บทนำ

ข้อสอบที่ดีจะต้องมีคุณสมบัติ ๓ ประการคือ มีความเที่ยงตรง (Validity) มีความเชื่อถือได้ (reliability) และมีความเหมาะสมใช้งานได้ (practicality) ผู้ออกข้อสอบไม่มีโอกาสทราบล่วงหน้าว่าข้อสอบแต่ละชุดที่ตนออกไปนั้นมีคุณสมบัติดังกล่าวมากน้อยประการใด จนกว่าจะได้นำไปทดลองใช้และนำผลสอบมาวิเคราะห์ ซึ่งการวิเคราะห์ข้อสอบนั้นไม่เป็นสิ่งที่ยากแต่ต้องใช้เวลาค่อนข้างมาก.

ทำไมต้องวิเคราะห์ข้อสอบ

เพื่อที่จะดูว่าข้อสอบนั้นได้วัดสิ่งที่ต้องการวัด (Validity) เพียงใด มีความเชื่อถือ (reliability) ได้เพียงไหน และมีอำนาจในการจำแนกผู้สอบ (power of discrimination) สูงพอหรือไม่ นั่นคือข้อสอบนั้นได้ครอบคลุมเนื้อหาที่ต้องการวัดหรือไม่ และถ้าใช้ข้อสอบนั้นในชั้นเวลาอื่นที่คล้ายคลึงกันหรือเหมือนกันผลจะออกมาเหมือนกันหรือไม่ และข้อสอบนั้นสามารถจำแนกผู้ที่มีความสามารถแตกต่างกันออกจากกันได้หรือไม่

โดยทั่วไปการวัดความเที่ยงตรงของข้อสอบ (Validity) มีอยู่ ๔ ประเภท (Gronbach, 1961, : 106) คือ

๑. Content Validity - ข้อสอบนั้นได้ครอบคลุมเนื้อหาที่สำคัญทุกหัวข้อหรือไม่ ข้อสอบควรจะเป็นตัวแทนของหลักสูตรที่ใช้ในการเรียนการสอนนั้น
๒. Predictive Validity - ผลของข้อสอบนั้นสามารถที่จะพยากรณ์พฤติกรรม (performance) ในอนาคต
๓. Concurrent Validity - ผลของข้อสอบนั้นสามารถจะประเมินความสามารถของผู้สอบในปัจจุบันหรือไม่
๔. Construct Validity - ข้อสอบนั้นตรงกับทฤษฎีแนวความเชื่ออันใดอันหนึ่ง การวัดความเชื่อถือของข้อสอบ (Reliability) ข้อสอบใด ถ้าใช้สอบกับผู้สอบกลุ่มเดียวกัน ๒ ครั้งแล้วให้ผลสอบเหมือนเดิมถือว่าข้อสอบนั้นมีความเชื่อถือสูงที่สุดคือมี perfect reliability (Ingram, 1977 : 16) แต่สภาวะเช่นนี้เกิดได้ยาก ค่าของความเชื่อถือสามารถหาได้หลายวิธี เช่น จาก Test-Retest, Parallel Form, Split-Half และ Kuder-Richardson²¹ เป็นต้น แต่วิธีนี้ทั้งข้อดีและข้อเสียแตกต่างกันออกไป.

การวิเคราะห์ข้อสอบแบบปรนัย (Objective Tests)

ข้อสอบแบบปรนัยคือข้อสอบชนิดใด ๆ ที่ไม่ว่าผู้ใดตรวจสอบผลคะแนนจะได้เท่ากัน เช่น ข้อสอบประเภท multiple-choice การวิเคราะห์ข้อสอบแบบนี้ทำได้เป็น ๒ ตอนคือ

๑. การวิเคราะห์ข้อสอบแบบรายข้อ (Item Analysis) เพื่อ
๑. หาความยากง่ายของข้อสอบแต่ละข้อ (difficulty index)
๒. หาค่าอำนาจจำแนกข้อสอบแต่ละข้อ (discrimination index)
๓. หาค่าอำนาจจำแนกของตัวลวง (distractors)

Validity) มี
อกข้อสอบไม่มี
น้อยประการใด
นั้นไม่เป็นสิ่งที่

ข. การวิเคราะห์ข้อสอบทั้งฉบับ (Test Analysis) เพื่อหาค่าความเชื่อถือ (reliability) ของข้อสอบทั้งฉบับ

การคำนวณหาค่าความยากง่ายของข้อสอบ (difficulty index) ทำได้ดังนี้

๑. เรียงผลคะแนน (total score) ของผู้สอบทั้งหมดจากมากไปหาน้อยหรือน้อยไปหามาก

๒. แบ่งคะแนนออกเป็น ๓ กลุ่ม คือกลุ่มคะแนนในช่วงสูงสุด กลุ่มคะแนนในช่วงปานกลาง และกลุ่มคะแนนในช่วงต่ำสุด โดยหลักการทางสถิติต้องใช้ ๒๗% ของกลุ่มคะแนนช่วงสูงกับกลุ่มคะแนนช่วงต่ำ แต่ตัวเลขนี้ไม่จำเป็นเสมอไป คืออาจจะแบ่งคะแนนออกเป็น ๓ กลุ่ม เท่า ๆ กัน กลุ่มละ $\frac{100}{3}\%$ หรือเอา ๒๕% ของกลุ่มคะแนนสูงกับกลุ่มคะแนนต่ำ หรือถ้ามีจำนวนผู้สอบน้อยก็อาจแบ่งออกเป็น ๒ กลุ่ม คือกลุ่มคะแนนสูงและกลุ่มคะแนนต่ำก็ได้ แต่สิ่งที่สำคัญคือต้องให้จำนวนผู้สอบในกลุ่มคะแนนสูงและกลุ่มคะแนนต่ำเท่ากัน

๓. นับคะแนนของคำตอบข้อที่ถูกที่ผู้สอบทั้ง ๒ กลุ่ม (กลุ่มคะแนนสูงกับกลุ่มคะแนนต่ำ) ตอบได้แต่ละข้อแล้วใช้สูตรต่อไปคำนวณหาค่าดัชนีของความยากง่ายของข้อสอบ

$$\text{difficulty index} = \frac{U + L}{N_u + N_l} \times 100 = \text{---} \%$$

เมื่อ U - ผลรวมของคะแนนข้อที่ถูกของกลุ่มคะแนนช่วงสูง

L - ผลรวมของคะแนนข้อที่ถูกของกลุ่มคะแนนช่วงต่ำ

N_u - จำนวนผู้สอบของกลุ่มคะแนนช่วงสูง

N_l - จำนวนผู้สอบของกลุ่มคะแนนช่วงต่ำ

การคำนวณหาค่าอำนาจจำแนก (discrimination index) ของข้อสอบแต่ละข้อ ใช้วิธีการเตรียมเช่นเดียวกับการหาค่าดัชนีความยากง่ายแต่ใช้สูตรดังต่อไปนี้

$$\text{discrimination index} = \frac{U - L}{N_u} = 0. \text{---} \text{ ทศนิยม}$$

การคำนวณหาค่าอำนาจจำแนกของตัวลวง (distractors) ก็ใช้สูตรข้างบนนี้เช่นกันต่อไปนี้เป็นตัวอย่างในการวิเคราะห์ข้อสอบตามวิธีการที่ได้กล่าวมาข้างต้นนี้เพื่อให้เห็นและเข้าใจในวิธีการอย่างชัดเจนยิ่งขึ้น

ตัวอย่าง ในการสอบครึ่งหนึ่งผู้สอบทั้งหมด ๑๐ คน และข้อสอบมีทั้งหมด ๕ ข้อ แต่ละข้อ มีตัวเลือก ๕ ตัว คือ a, b, c, d, e

วิธีการ ๑. เรียงผลคะแนนรวมจากน้อยไปหามาก

๒. แบ่งผู้สอบออกเป็น ๒ กลุ่มตามลำดับของคะแนน คือ ๕๐% ของกลุ่มที่ได้คะแนนในช่วงสูง และ ๕๐% ของกลุ่มที่ได้คะแนนในช่วงต่ำ เพราะในที่นี้ผู้สอบมีเพียง ๑๑ คนเท่านั้น

๓. นับคะแนนของคำตอบข้อที่ถูกของผู้สอบในแต่ละกลุ่มแล้วนำมาคำนวณหาค่าเฉลี่ย ค่าความสุทธที่ได้กล่าวไว้แล้ว

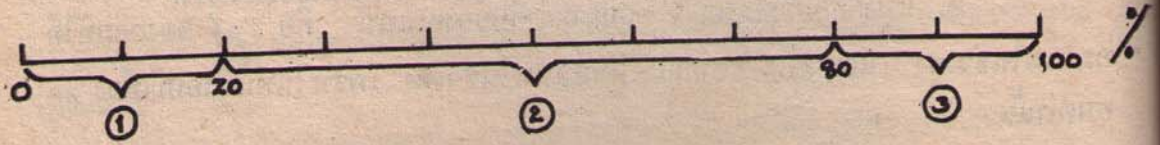
ตัวอย่างและวิธีการได้แสดงไว้ในตารางต่อไปนี้

		*					*					*					*					*					คะแนนรวม					
ข้อสอบ	ผู้สอบ	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e	a	b	c	d	e		a	b	c	d	e
				(1)					(2)					(3)					(4)					(5)								
		/					/					/					/					/					1					
		/					/					/					/					/					1					
		/					/					/					/					/					2					
		/					/					/					/					/					2					
		/					/					/					/					/					3					
		0	4	1	0	0	3	0	1	0	1	1	1	2	0	1	1	1	2	1	0	1	3	0	0	1						
		(1)					(2)					(3)					(4)					(5)										
		/					/					/					/					/					3					
		/					/					/					/					/					3					
		/					/					/					/					/					3					
		/					/					/					/					/					4					
		/					/					/					/					/					5					
		0	1	0	4	0	2	2	0	1	0	4	1	0	0	0	1	0	4	0	0	0	4	0	0	0	4	0	1	0		
		0	4	1	0	0	3	0	1	0	1	1	1	2	0	1	1	1	2	1	0	1	3	0	0	1						
		0	.8	.2	.8	0	.2	.4	.2	.2	.2	.6	.0	.4	0	.2	0	.2	.4	.2	0	.2	.2	0	.2	.2						
		40%					50%					50%					60%					70%										

*คือตัวเลือกที่ถูกของแต่ละข้อ

การแปลผลคะแนน

๑. คำนีความยากง่ายของข้อสอบ จาก scale ข้างล่างนี้จะเห็นว่า



ข้อสอบที่มีค่านีความยากง่าย

- | | |
|-------------------|-----------------------------|
| ระหว่าง (๑) ๑-๒๐% | ถือเป็นข้อสอบที่ยาก |
| (๒) ๒๑-๘๐% | ถือเป็นข้อสอบที่อยู่ในเกณฑ์ |
| (๓) ๘๑-๑๐๐% | ถือเป็นข้อสอบที่ง่าย |

ข้อสอบชุดหนึ่ง ๆ ควรจะมีข้อสอบทั้งยากและง่ายคละกัน ไปคือควรมีค่าต่ำกว่า ๕๐% และสูงกว่า ๕๐% วิธีจ้กัก็ควรจะเรียงข้อสอบจากง่ายไปหายาก ทั้งนี้เพื่อเป็นการให้กำลังใจแก่ผู้สอบ

๒. คำนีอำนาจในการจำแนกของข้อสอบ

ตามหลักสถิติ ข้อสอบที่ถูกถ้ามีค่าอำนาจจำแนกตั้งแต่ ๑.๓ ขึ้นไป ถือว่าใช้ได้ คือสามารถจำแนกความสามารถของผู้สอบได้ ถ้าน้อยกว่าต้องแก้ไข ถ้ามีค่าเป็นศูนย์แสดงว่าไม่มีอำนาจในการจำแนกเลย และถ้ามีค่าเป็นลบถือว่าเป็นข้อสอบที่ใช้ไม่ได้ ควรตัดทิ้งไป ค่าอำนาจจำแนกยิ่งสูงยิ่งสามารถจำแนกความสามารถได้มาก

ส่วนข้อสอบที่เป็นตัวลวง (distractors) ถ้ามีอำนาจจำแนกมีค่าเป็นลบกลับถือว่าเป็นตัวลวงที่ดี เพราะแสดงว่าผู้สอบกลุ่มอ่อนเลือกตัวลวงตัวนั้นมากกว่ากลุ่มเก่ง

ผลของคะแนนในตัวอย่างที่แสดงไว้ในตารางสามารถนำมาแปลผลได้ดังนี้ ข้อสอบข้อที่ ๑ มีค่าของค่านีความยากง่ายเท่ากับ ๔๐% มีค่าของค่านีอำนาจจำแนกเท่ากับ ๑.๘ จะเห็นว่าเป็นข้อสอบที่อยู่ในเกณฑ์ไม่ค่อยยากมากและสามารถจำแนกผู้สอบได้ดีเพราะมีค่าอำนาจจำแนกสูงถือเป็นข้อสอบที่ใช้ได้ แต่เมื่อคูนอำนาจในการจำแนกของตัวลวงทั้ง ๔ ตัวของ

ข้อ ๒ ข้อนี้ไม่มีอำนาจจะลงผู้สอบได้ ถือเป็นข้อสอบที่ควรแก้ไขเพื่อให้ยุติเคียงกับคำตอบที่ถูกต้อง ส่วนข้อ (b) และ (c) มีค่าอำนาจจำแนกเป็นลบ คือ -๑.๖ กับ -๑.๒ แสดงว่าเป็นตัวลงที่ใช้ได้ เพราะสามารถลงผู้สอบกลุ่มอ่อนได้มากกว่าผู้สอบกลุ่มเก่ง ข้อสอบข้อที่ ๒ มีค่าของดัชนีความยากง่ายเท่ากับ ๕๐% มีค่าของดัชนีอำนาจจำแนกเท่ากับ -๑.๒ จะเห็นว่าเป็นข้อสอบที่อยู่ในเกณฑ์ปานกลางไม่ยากและไม่ง่ายเกินไป เมื่อดูแต่ค่าของดัชนีความยากง่าย แต่เมื่อดูที่อำนาจจำแนกแล้วจะเห็นว่าต่ำกว่าเกณฑ์ คือ ๑.๓ ที่วางไว้ คนกลุ่มเก่งทำได้น้อยกว่ากลุ่มอ่อน แสดงว่าคนกลุ่มอ่อนทำได้ อาจมาจากการเดาถูก ข้อสอบข้อนี้จึงควรระคัตทั้งหรือแก้ไขใหม่ให้มีอำนาจจำแนกสูงกว่านี้และมีค่าเป็นบวก แต่ลองหันมาดูอำนาจจำแนกของตัวลงทั้ง ๔ ตัวดู จะเห็นว่าข้อ (c) และ (e) ซึ่งมีค่าเป็นลบคือ -๑.๒ ถือเป็นตัวลงที่ใช้ได้ ส่วนข้อ (b) และ (d) ซึ่งมีค่าเป็นบวก คือ ๑.๔ กับ ๑.๒ นั้นควรแก้ไขใหม่เพราะมีประสิทธิภาพในการลงสูงเกินไป คือสามารถลงคนกลุ่มเก่งได้มากกว่าคนกลุ่มอ่อน แสดงว่าอาจเป็นตัวลงที่ให้คำตอบคลุมเครือ คือดูแล้วอาจมีโอกาสเป็นตัวคำตอบที่ถูกต้อง ถ้าเป็นเช่นนั้นควรแก้ไขใหม่ให้ยุติเป็นคำตอบที่ผิดมากขึ้นกว่าเดิม

ข้อสอบข้อที่ ๓ มีค่าของดัชนีความยากง่ายเท่ากับ ๕๐% มีค่าของดัชนีอำนาจจำแนกเท่ากับ ๑.๒ จะเห็นว่าเป็นข้อสอบที่อยู่ในเกณฑ์ความยากง่ายปานกลาง และสามารถจำแนกผู้สอบได้ดี คนกลุ่มเก่งสามารถทำได้มากกว่าคนกลุ่มอ่อน นับเป็นข้อสอบที่ใช้ได้ควรเก็บไว้ ส่วนข้อสอบที่เป็นตัวลงอีก ๔ ข้อนั้นจะเห็นว่า ข้อ (c) และ (e) ซึ่งมีค่าอำนาจจำแนกเป็นลบ ถือเป็นตัวลงที่ดี แต่ข้อ (b) และ (d) มีค่าอำนาจจำแนกเป็นศูนย์ควรแก้ไขใหม่ เพราะไม่มีประสิทธิภาพในการลงผู้สอบได้เลยไม่ว่าจะเป็นกลุ่มเก่งหรือกลุ่มอ่อน

ข้อสอบข้อที่ ๔ มีค่าของดัชนีความยากง่ายเท่ากับ ๖๐% และค่าของดัชนีอำนาจจำแนกเท่ากับ ๑.๔ แสดงว่าถึงแม้ว่าข้อสอบนั้นจะไม่สู้ยากมากนักแต่ก็สามารถจำแนกผู้สอบได้ดี เพราะคนกลุ่มเก่งสามารถตอบได้มากกว่าคนกลุ่มอ่อน ถือเป็นข้อสอบที่ดีควรเก็บไว้ ส่วนข้อสอบที่เป็นตัวลงอีก ๔ ข้อนั้น ข้อ (a) และ (e) มีค่าอำนาจจำแนกเป็นศูนย์ทั้งคู่ถือว่าเป็นตัวลงที่ไม่ดี ควรแก้ไขใหม่ ส่วนข้อ (b) และ (d) มีค่าอำนาจจำแนกเป็นลบทั้งคู่ถือว่าเป็นตัวลงที่ใช้ได้

$\frac{1}{100} \%$

มีค่าต่ำกว่า
เพื่อเป็นการให้

ถือว่าใช้ได้
เป็นศูนย์แสดงว่า
ควรตัดทิ้งไป

เป็นลบกลับคือ

เก่ง
ก็ดังนี้
อำนาจจำแนกเท่ากับ
ได้ดีเพราะมีค่า
ทั้ง ๔ ตัวของ

ข้อสอบข้อที่ ๕ มีค่าของดัชนีความยากง่ายเท่ากับ ๗๐% และค่าดัชนีอำนาจจำแนกเท่ากับ ๑.๒ แสดงว่าเป็นข้อสอบที่ค่อนข้างง่าย แต่ค่าอำนาจจำแนกไม่สูงจะสูงนักอาจจะแก้ไขเล็กน้อยให้มีค่าอำนาจจำแนกสูงขึ้น ส่วนข้อสอบที่เป็นตัวลวงอีก ๔ ข้อนี้จะเห็นว่า ข้อ (c) มีค่าอำนาจจำแนกเป็นศูนย์ซึ่งเป็นตัวลวงที่ไม่ดี เพราะไม่มีประสิทธิภาพในการลวงเลย ข้อ (d) มีค่าเท่ากับ ๑.๒ ซึ่งสามารถลวงคนกลุ่มเก่งได้เพียงคนเดียวและไม่สามารถลวงคนกลุ่มอ่อนเลย แสดงว่าเป็นตัวลวงที่ไม่ดีต้องแก้ไขใหม่ให้เหมาะสมกว่านี้

ข. การวิเคราะห์ข้อสอบทั้งฉบับ (Test Analysis) ทั้งนี้เพื่อความเชื่อถือของข้อสอบซึ่งจะต้องทำการวิเคราะห์หาสิ่งดังต่อไปนี้

๑. หาค่าของ mean ของข้อสอบทั้งฉบับซึ่งใช้สูตร

$$\text{mean (M)} = \frac{\sum x}{N}$$

x = ผลรวมของคะแนนทั้งหมด

N = จำนวนผู้สอบทั้งหมด

จากตัวอย่างข้อสอบที่แสดงไว้ข้างต้น ๕ ข้อ ข้อสอบนี้จะหาค่า mean ได้ดังนี้

$$\frac{1 + 1 + 2 + 2 + 3 + 3 + 3 + 3 + 4 + 5}{10} = \frac{27}{10} = 2.7$$

ค่าของ mean ถ้าอยู่ระหว่าง ๕๐%-๗๐% ถือเป็นข้อสอบที่ยาก

สำหรับตัวอย่างนี้ มีข้อสอบ ๕ ข้อ และมี mean = 2.7

ถ้าตอบถูก ๕ ข้อ คิดเป็น ๑๐๐%

ถ้าตอบถูก ๒.๗ ข้อ คิดเป็น $\frac{๑๐๐ \times ๒.๗}{๕} = ๕๔\%$

แสดงว่าข้อสอบชุดนี้ค่อนข้างยาก

๒. หาค่า Standard Deviation ใช้สูตร

$$SD = \sqrt{\frac{\sum d^2}{N}}$$

d = ค่าเบี่ยงเบนจาก mean = M-x

Σ = ผลรวมของ - - - -

N = จำนวนผู้สอบทั้งหมด

จากตัวอย่าง

คะแนน X	d	d ²
1	1.7	2.89
1	1.7	2.89
2	.7	.49
2	.7	.49
3	-.3	.49
3	-.3	.09
3	-.3	.09
3	-.3	.09
4	-1.3	16.9
5	-2.3	5.29
		$\Sigma d^2 = 14.5$

$$SD = \sqrt{\frac{14.5}{10}} = \sqrt{1.45} = 1.204$$

ค่า SD เป็นตัวอธิบายการกระจายของคะแนนว่าเป็นอย่างไร สำหรับตัวอย่างนี้ N=10, $\Sigma = 27$, SD = 1.2 แสดงว่าประมาณ ๒ ใน ๓ ของผู้สอบกลุ่มนี้จะได้คะแนนอยู่ระหว่าง ๑.๓ = (๒.๗ + ๑.๒) และ ๑.๕ = (๒.๗ - ๑.๒)

๓. ค่า Reliability ซึ่งมีวิธีการให้หาค่านี้ได้หลายวิธีในที่นี้จะลองใช้วิธีการของ Kuder (1977 : 145) ซึ่งเป็นวิธีที่ง่ายและรวบรัดไม่ต้องเสียเวลามาก เพราะเป็นการง่ายที่จะเปรียบเทียบค่าได้เลย

วิธีหาค่า Reliability

๑. Mean score เป็นตัวตัดสินว่าข้อสอบชุดนั้นง่ายหรือยาก คือถ้า mean มีค่าระหว่าง ๕๑-๗๐% ถือว่าเป็นข้อสอบที่ง่าย ถ้า mean มีค่าระหว่าง ๕๑-๗๐% ถือว่าเป็นข้อสอบที่ยาก ถ้าข้อสอบง่ายใช้ตาราง A ถ้าข้อสอบยากใช้ตาราง B

๒. Standard Deviation แล้วหารด้วยจำนวนข้อทั้งหมดของข้อสอบแล้วเทียบเป็นเปอร์เซ็นต์ของจำนวนข้อของข้อสอบใช้เทียบ ถ้าได้ค่า SD ใกล้กับ ๑๑ เปอร์เซ็นต์ของจำนวนข้อของข้อสอบใช้เทียบ

กับบรรทัดแรกของตารางและถ้าค่า SD ใกล้เคียงกับ ๑๕ เปอร์เซนต์ใช้เทียบกับบรรทัดที่ ๒ และ บรรทัดที่ ๓ ถ้าค่า SD ใกล้เคียงกับ ๒๑ เปอร์เซนต์
 ๓. เลือกแถวที่จะอ่านค่าความเชื่อถือจากแถวที่มีจำนวนข้อใกล้เคียงกับจำนวนข้อสอบที่ ต้องการหาค่า r (reliability co-efficients)

ตารางจาก Harris (1977 : 145)

Table 7. Estimating the Approximate Reliability of a Test*

Number of items (n)	A. Easy Test (Average 70 to 90 per cent Correct)								
	20	30	40	50	60	70	80	90	100
If SD is .10n	.21	.48	.62	.69	.75	.78	.81	.83	.85
If SD is .15n	.68	.80	.84	.88	.90	.91	.92	.93	.94
If SD is .20n	.84	.90	.92	.94	.95	.96	.96	.97	.97

Number of items (n)	B. Difficult Test (Average 50 to 70 per cent Correct)								
	20	30	40	50	60	70	80	90	100
If SD is .10n	—	.21	.41	.53	.61	.66	.71	.74	.77
If SD is .15n	.49	.67	.75	.80	.84	.86	.88	.89	.90
If SD is .20n	.74	.83	.87	.90	.92	.93	.94	.94	.95

* Taken from Paul B. Diederich, *Short-cut Statistics for Teacher-made Tests*, Evaluation and Advisory Service Series, No. 5, 2d ed., Princeton, N.J.: Educational Testing Service, 1964, p. 31. By permission of Educational Testing Service.

ตัวอย่างของการหาค่า Reliability co-efficients (r)

ข้อสอบชุดหนึ่งมี ๘๐ ข้อ ใช้กับนักเรียน ๑๐๐ คน มีค่า mean = 62 และค่า SD = 9.21
 วิธีการ mean = 62 จากข้อสอบ ๘๐ ข้อ แสดงว่า

ถ้าตอบถูก ๘๐ ข้อ คิดเป็น ๑๐๐%

$$\text{,, } 62 \text{ ข้อ คิดเป็น } \frac{100 \times 62}{80} = 77.5\%$$

แสดงว่าข้อสอบนี้อยู่ในเกณฑ์ง่ายเพราะมีค่า mean อยู่ระหว่าง ๗๐-๘๐% ฉะนั้น

จะต้องใช้ตาราง A เป็นตัวเทียบ

ค่า SD = 9.21 จากข้อสอบ ๘๐ ข้อ คิดเป็นเปอร์เซนต์ได้เท่ากับ $\frac{100 \times 9.21}{80}$

= 11.5% หรือ $\approx .11r$

แสดงว่าต้องใช้บรรทัดที่หนึ่งในการหาค่า r

จากตาราง A จะได้ค่า r = 0.81

ถ้าครูมีเวลาอยากจะใช้วิธีคำนวณจากสูตรของ Kuder-Richardson₂₁ ก็ได้ ซึ่งมีสูตรว่า

$$KR_{21} = 1 - \frac{M(K-M)}{K(SD)^2} \quad (\text{Dierich : 1964})$$

หมายเหตุ สูตรนี้เป็นสูตรที่ดัดแปลงให้ง่ายขึ้นสำหรับครูซึ่งมีเวลาน้อย

เมื่อ M = mean ของข้อสอบ

K = จำนวนข้อของข้อสอบ

SD = ความเบี่ยงเบนมาตรฐาน

เพราะฉะนั้นจากข้อมูลข้างต้นจะหาค่าได้เป็น

$$= 1 - \frac{62(80-62)}{(80)(9.21)^2}$$

$$= 1 - 0.165$$

$$= 0.835$$

โดยทั่วไปข้อสอบที่ถือว่าเป็นข้อสอบมาตรฐานจะมีค่า R มากกว่า ๐.๕๐ ส่วนข้อสอบในห้องเรียนนั้นถ้ามีค่ามากกว่า ๐.๗๐ ก็ถือว่าใช้ได้

๔. หาค่า Standard Error of Measurement (SE_{mean}) ซึ่งมีสูตรดังนี้

$$SE_{\text{mean}} = DS \sqrt{1-r}$$

ในที่นี้จะใช้ข้อมูลอันเดียวกับที่ใช้หาค่าความเชื่อถือ ซึ่งจะแทนค่าสูตรได้ดังนี้

$$SE_{\text{mean}} = 9.21 \sqrt{1-0.81}$$

$$= 9.21 \sqrt{0.19}$$

$$= 9.21 \times 0.435$$

$$= 4.00635$$

ค่า SE_{mean} จะเป็นตัวชี้ให้เห็นว่าคะแนนของผู้สอบจะสามารถเบี่ยงเบนไปจากคะแนนจริงได้เป็นเท่าไรนั่นคือ จากข้อมูลนี้ถ้าผู้สอบที่ได้คะแนนจริง ๗๐ คะแนน เราสามารถทำนายได้ว่าคะแนนของเขาจะไม่มากไปกว่า ๗๘ หรือน้อยกว่า ๖๒ นั่นคือ จะอยู่ระหว่าง $\pm 2 SE_{\text{mean}}$ ของคะแนนจริงของเขา

100
.85
.94
.97
100
.77
.90
.95

SD = 9.21

๗๐% ฉะนั้น

$$\frac{100 \times 9.21}{80}$$

สรุป

การประเมินผลข้อสอบหลังจากได้นำไปทดสอบแล้ว ถือว่าเป็นสิ่งจำเป็นอย่างยิ่งที่ครูควรจะทำเพื่อจะได้ศึกษาและวิเคราะห์ว่าข้อสอบนั้นมีความเชื่อถือ เทียงตรง มากน้อยแค่ไหน และสามารถจะวัดผลตรงตามวัตถุประสงค์ที่ตั้งไว้หรือไม่ การวิเคราะห์ด้วยการทำ Item Analysis ของข้อสอบรายข้อและการวิเคราะห์ข้อสอบทั้งฉบับคงได้แสดงไว้ในที่นี้จะเป็นประโยชน์ต่อการปรับปรุงแก้ไขเพื่อให้ได้มาซึ่งข้อสอบที่ดี วิธีการเช่นนี้ซึ่งต้องอาศัยสถิติเป็นเครื่องมือจะเป็นการช่วยยืนยันข้อเท็จจริง และยังช่วยในการตัดสินใจของเราไปเป็นอย่างดี มีหลักการและเหตุผล ที่สำคัญให้ทุนเวลาและแรงงานของเราในการออกข้อสอบใหม่หมดทั้งชุดในแต่ละครั้ง เพราะเราสามารถจะเลือกแก้ไขในส่วนที่ไม่ดีเท่านั้น ถ้าข้อไหนดีก็เก็บไว้เข้า data bank ของข้อสอบเพื่อใช้ในโอกาสต่อไปได้

บทความฉบับนี้เขียนขึ้นเพื่อต้องการให้ครูที่มีเวลาน้อยมีโอกาสทำการวิเคราะห์ข้อสอบของตัวเองที่ออกไปได้ในเวลาอันรวดเร็ว ฉะนั้นสูตรและวิธีการบางอย่างในที่นี้จึงได้รับการดัดแปลงเพื่อความเหมาะสมและประหยัดเวลาแต่ผลที่ได้ก็เป็นที่น่าเชื่อถือได้เช่นเดียวกัน ถ้าครูท่านใดต้องการทำการวิเคราะห์ให้ละเอียดยิ่งขึ้นกว่านี้ก็สามารที่จะศึกษาวิธีการต่างๆ ได้จากตำราทางสถิติเพื่อการศึกษาโดยทั่วๆ ไปที่เชื่อถือได้

บรรณานุกรม

- Allen, J.P.B and A. Davies, eds, 1977 Testing and Experimental Method. London : Oxford University Press.
- Crocker, A.C. 1974. Statistics for the Teacher. London : NFER Publishing Company Ltd.
- Cronbach, L.J. 1961. Essentials for Psychological Testing. New York : Harper and Row.
- Diedrich, P.B. 1964. Short-Cut Statistics for Teacher-Made Tests, Evaluation and Adivisory Service Series, No. 5 Princeton, N.J. : Educational Testing Service.
- Harris, D.P. 1977. Testing English as A Second Language. India : Tata McGraw -Hill
- Ingram, E. 1977. "Basic Concepts in Testing" *The Edinburgh Course in Applied Linguistics : Testing and Experimental Methods*. Vol IV. eds by Allen, J.B.P. and A. Davies. Oxford University Press.
- Valette, Rebecca M. 1977. Modern Language Testing. 2nd Ed. New York : Harcourt Brace Jovanovick.