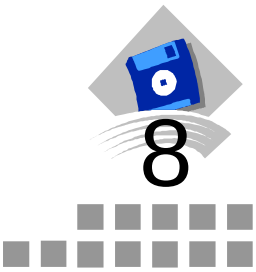
การหาคุณภาพของแบบทดสอบ
Quality Testing of Tests



■ บทนำ

ข้อสอบหรือแบบทดสอบที่ใช้ในบทเรียนคอมพิวเตอร์ ไม่ว่าจะเป็นแบบทดสอบก่อนบทเรียน แบบทดสอบหลังบทเรียน หรือแบบฝึกหัดระหว่างบทเรียนก็ตาม หลังจากที่ผ่านกระบวนการออกแบบทดสอบตามวัตถุประสงค์เชิงพฤติกรรมของบทเรียนแล้ว ก่อนที่จะนำไปใช้งานจะต้องผ่านการทดสอบหาคุณภาพก่อน เนื่องจากแบบทดสอบเป็นเครื่องมือที่ใช้ประเมินผลการเรียนรู้ของผู้เรียนว่าบรรลุตามวัตถุประสงค์หรือไม่ อีกทั้งยังใช้ในการประเมินประสิทธิภาพของบทเรียนหากแบบทดสอบไม่มีคุณภาพ ย่อมส่งผลให้คุณภาพของบทเรียนคอมพิวเตอร์ไม่มีคุณภาพตามไปด้วย ในทางตรงกันข้ามหากแบบทดสอบที่ใช้ในบทเรียนคอมพิวเตอร์มีคุณภาพดี ย่อมส่งผลให้บทเรียนมีคุณภาพดีตามไปด้วยเช่นกัน โดยเฉพาะอย่างยิ่งการพัฒนาบทเรียนคอมพิวเตอร์ในการวิจัยเพื่อเปรียบเทียบกับวิธีการเรียนการสอนรูปแบบต่าง ๆ เช่น การเปรียบเทียบผลสัมฤทธิ์ทางการเรียนระหว่างการใช้บทเรียนคอมพิวเตอร์กับวิธีการสอนแบบปกติ ย่อมต้องใช้แบบทดสอบที่มีคุณภาพผ่านตามเกณฑ์มาตรฐาน เพื่อถ่ายทอดผลการทดสอบที่เป็นจริง เนื่องจากส่วนนี้เป็นปัจจัยสำคัญที่ทำให้ผลการวิจัยสอดคล้องกับวัตถุประสงค์ที่ตั้งไว้หรือไม่ เพียงใด

เป็นที่ทราบกันดีอยู่แล้วว่า ไม่ว่าจะเป็นการหาประสิทธิภาพของบทเรียนคอมพิวเตอร์ตามเกณฑ์ที่นิยมกำหนดกันเพื่อใช้เป็นเกณฑ์มาตรฐานทั่วไป เช่น 90/90 การหาผลสัมฤทธิ์ทางการเรียนของผู้เรียนหรือการตรวจวัดความคงทนทางการเรียน ล้วนเป็นผลมาจากคุณภาพของแบบทดสอบทั้งสิ้น ดังนั้น การหาคุณภาพของแบบทดสอบจึงเป็นกระบวนการที่ต้องดำเนินการอย่างรอบคอบและอาจจะต้องทำการทดสอบซ้ำ ๆ หลายครั้ง จนกว่าจะได้แบบทดสอบที่เป็นมาตรฐานและมีจำนวนข้อเพียงพอกับความต้องการ คุณภาพของแบบทดสอบ เกี่ยวข้องกับปัจจัยที่สำคัญจำนวน 5 องค์ประกอบ ดังนี้ (วัญญา. 2531 : 118 - 124)

1. ความเที่ยงตรง (Validity)
2. ความเชื่อมั่น (Reliability)
3. ความยากง่าย (Difficulty)
4. อำนาจจำแนก (Discrimination)
5. ความเป็นปรนัย (Objectivity)

■ ความเที่ยงตรง (Validity)

ความเที่ยงตรง (Validity) หมายถึง ความถูกต้องของแบบทดสอบในสิ่งที่ต้องการจะวัดหรือความถูกต้องแม่นยำที่แบบทดสอบวัดได้ตามวัตถุประสงค์ที่วางไว้ หรืออาจกล่าวได้อีกนัยหนึ่งว่าเป็นความสามารถของแบบทดสอบ ที่จะสะท้อนความหมายที่แท้จริงของแนวคิดที่ต้องการศึกษาออกมาได้อย่างสมบูรณ์และถูกต้อง แบบทดสอบจะไม่ได้มีความเที่ยงตรงโดยตัวเองแต่จะมีความเที่ยงตรงในจุดมุ่งหมายเฉพาะกับกลุ่มที่ต้องการวัดเท่านั้น ความเที่ยงตรงจำแนกออกได้เป็น 4 ประเภท ดังนี้

1. ความเที่ยงตรงตามเนื้อหา (Content Validity) หมายถึง ระดับความสามารถของแบบทดสอบที่วัดในเนื้อหาที่ต้องการจะวัด เช่น หากต้องการวัดเรื่องความสนใจ ข้อคำถามในแบบทดสอบหรือข้อสอบก็ต้องเป็นเรื่องของความสนใจ โดยการพิจารณาว่าเนื้อหาของแบบทดสอบสะท้อนแนวความคิดที่ตามที่ต้องการหรือไม่ ความเที่ยงตรงตามเนื้อหาจึงมีความสำคัญยิ่งในการวัดผลสัมฤทธิ์ทางการเรียน การวัดผลการเรียนการสอนที่ใช้แบบทดสอบไม่ตรงหรือไม่ครอบคลุมเนื้อหาที่เรียน จึงเป็นการวัดผลที่ขาดความเที่ยงตรงตามเนื้อหา

การทดสอบความเที่ยงตรงตามเนื้อหา ทำได้โดยพิจารณาจากกระบวนการสร้างแบบทดสอบหรือข้อสอบว่าวัดได้จริงตามที่ต้องการจะวัดหรือไม่ หรือโดยการตรวจสอบคำตอบกับข้อเท็จจริงที่ปรากฏ เช่น การสังเกตจากพฤติกรรมที่เกิดขึ้นว่าสอดคล้องกับพฤติกรรมที่ตอบในแบบทดสอบหรือไม่ กระบวนการทดสอบดังกล่าวนี้ต้องอาศัยผู้เชี่ยวชาญทางด้านเนื้อหา เพื่อตรวจสอบกระบวนการสร้างแบบทดสอบ เพื่อตัดสินใจว่าข้อคำถามในแบบทดสอบว่าสามารถใช้เป็นตัวแทนของเนื้อหาที่จะถามได้หรือไม่ โดยการเปรียบเทียบสิ่งที่ปรากฏในแบบทดสอบกับสิ่งที่ควรจะถามว่ามีความสอดคล้องกันมากเพียงใด การทดสอบความเที่ยงตรงตามเนื้อหาโดยผู้เชี่ยวชาญในลักษณะนี้ เรียกว่า การหาค่าความสอดคล้องระหว่างวัตถุประสงค์กับแบบทดสอบ หรือเรียกว่า การหาค่า IOC (Index of Item-objective Congruence)

สูตรสำหรับหาค่า IOC

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

เมื่อ

IOC = ความสอดคล้องระหว่างวัตถุประสงค์กับแบบทดสอบ

$\sum R$ = ผลรวมของคะแนนการพิจารณาจากผู้เชี่ยวชาญ

N = จำนวนผู้เชี่ยวชาญ

การพิจารณาความสอดคล้องระหว่างวัตถุประสงค์กับแบบทดสอบ จะเป็นการพิจารณาแบบทดสอบรายข้อจากความคิดเห็นของผู้เชี่ยวชาญ โดยใช้แบบสอบถามที่แนบไปพร้อมกับ

แบบทดสอบที่ต้องการให้ผู้เชี่ยวชาญประเมินความสอดคล้อง โดยมีเกณฑ์การให้คะแนนเพื่อหาค่า IOC ของผู้เชี่ยวชาญกำหนดเป็น 3 ระดับ ดังนี้

+1 หมายถึง แน่ใจว่า แบบทดสอบวัดตรงตามวัตถุประสงค์หรือตรงตามเนื้อหา

0 หมายถึง ไม่แน่ใจว่า แบบทดสอบวัดตรงตามวัตถุประสงค์หรือตรงตามเนื้อหา

-1 หมายถึง แน่ใจว่า แบบทดสอบไม่ได้วัดตรงตามวัตถุประสงค์หรือตรงตามเนื้อหา

แบบทดสอบหรือข้อสอบที่ถือว่ามีความเที่ยงตรงตามเนื้อหาในระดับดี สามารถนำไปวัดผลได้ จะต้องมีค่า IOC เกินกว่า 0.5 เป็นต้นไป

ตัวอย่างแบบทดสอบเพื่อหาค่า IOC และการแปลผล

วัตถุประสงค์/เนื้อหา	แบบทดสอบ	ระดับของการพิจารณา		
		+1	0	-1
1.	1.1 1.2 1.3			
2.	2.1 2.2			

วัตถุประสงค์	แบบทดสอบ	คะแนนความคิดเห็นของผู้เชี่ยวชาญ					รวม	ค่า IOC	สรุปผล
		คนที่ 1	คนที่ 2	คนที่ 3	คนที่ 4	คนที่ 5			
1	1.1	+1	+1	+1	0	+1	4	.80	ใช้ได้
	1.2	0	+1	+1	0	+1	3	.60	ใช้ได้
	1.3	+1	+1	-1	+1	+1	3	.60	ใช้ได้
2	2.1	0	-1	-1	-1	+1	-2	-.40	ใช้ไม่ได้
	2.2	+1	-1	0	+1	-1	0	0	ใช้ไม่ได้

2. ความเที่ยงตรงตามโครงสร้าง (Construct Validity) หมายถึง ความสามารถของแบบทดสอบที่วัดได้ตามลักษณะคุณสมบัติ ทฤษฎี และประเด็นต่าง ๆ ของโครงสร้างนั้น โครงสร้างเป็นคุณลักษณะที่อธิบายพฤติกรรมต่าง ๆ โดยแท้จริงแล้วโครงสร้างคือสิ่งที่ประดิษฐ์ขึ้นมาเพื่ออธิบายพฤติกรรม เช่น โครงสร้างของคอมพิวเตอร์ประกอบไปด้วยหน่วยรับข้อมูล หน่วยแสดงผลข้อมูล และหน่วยประมวลผลกลาง เป็นต้น ถ้าแบบทดสอบมีความเที่ยงตรงตามโครงสร้าง

จะต้องประกอบด้วยแนวคำถามที่สามารถวัดประเด็นต่าง ๆ ครบทั้ง 3 ส่วนประกอบการทดสอบความเที่ยงตรงตามโครงสร้าง ซึ่งทำได้ 2 วิธี ดังนี้

2.1 การหาค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) ทำได้โดยการหาค่าสัมประสิทธ์ความสัมพันธ์ของคะแนนของแบบทดสอบ 2 ชุด ที่วัดในเรื่องเดียวกัน เช่น แบบทดสอบมาตรฐานกับแบบทดสอบที่สร้างขึ้นเพื่อต้องการหาความเที่ยงตรงตามโครงสร้าง โดยใช้สูตรการหาค่าสัมประสิทธิ์ความสัมพันธ์ ถ้าค่าสัมประสิทธิ์ความสัมพันธ์มีค่าสูงและมีทิศทางเดียวกัน แสดงว่าแบบทดสอบที่สร้างขึ้นมีความเที่ยงตรงตามโครงสร้างสูง สามารถนำไปใช้งานได้

2.2 การเปรียบเทียบกับกลุ่มที่มีลักษณะที่ต้องการวัดอย่างเด่นชัด หรือเรียกว่าวิธีนี้ว่า Known Group Technique โดยการนำแบบทดสอบที่สร้างขึ้นไปทดสอบกับกลุ่มตัวอย่าง 2 กลุ่ม ได้แก่ กลุ่มที่มีลักษณะตามที่กำหนดขึ้นอย่างเด่นชัด กับกลุ่มที่ไม่มีลักษณะดังกล่าว หลังจากนั้นจึงนำค่าที่ได้มาเปรียบเทียบโดยใช้ t-test แบบ Independent ถ้าพบว่าผลการเปรียบเทียบมีความแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 หรือ .05 แสดงว่าแบบทดสอบที่สร้างขึ้นมีความเที่ยงตรงตามโครงสร้างสูง

การทดสอบหาค่าความเที่ยงตรงตามโครงสร้าง จึงเป็นเรื่องที่ซับซ้อนและมีขั้นตอนมากกว่าการทดสอบหาค่าความเที่ยงตรงตามเนื้อหา สำหรับแบบทดสอบแนวอิงเกณฑ์ที่ใช้ในการเรียนการสอน มีวิธีการทดสอบความเที่ยงตรงตามโครงสร้างอยู่หลายวิธี ในที่นี้ขอนำเสนอ 2 วิธี ดังนี้

1. วิธีของคาร์เวอร์ (Carver Method)

วิธีการทดสอบความเที่ยงตรงตามโครงสร้างของคาร์เวอร์ ทำได้โดยการนำแบบทดสอบที่สร้างขึ้นไปทดสอบกับกลุ่มผู้เรียนที่เรียนแล้วกับกลุ่มผู้เรียนที่ยังไม่เคยเรียน แล้วนำอัตราส่วนระหว่างผลรวมของจำนวนผู้เรียนที่ยังไม่เคยเรียนที่สอบไม่ผ่าน กับ จำนวนผู้เรียนที่เรียนแล้วหรือสอบผ่านต่อจำนวนผู้เรียนทั้งหมดมาคำนวณตามสูตร

สูตรการหาค่าความเที่ยงตรงตามโครงสร้างของคาร์เวอร์

$$Construct - Validity = \frac{a + c}{N}$$

เมื่อ

a = จำนวนผู้เรียนที่เรียนแล้วและสอบผ่าน

b = จำนวนผู้เรียนที่ยังไม่เคยเรียนและสอบไม่ผ่าน

N = จำนวนผู้เรียนทั้งหมด

ตัวอย่าง

แบบทดสอบวิชาโครงสร้างข้อมูลคะแนนเต็ม 10 คะแนน นำไปทดสอบกับผู้เรียน 2 กลุ่ม ได้แก่ กลุ่มผู้เรียนที่เรียนแล้ว กับกลุ่มผู้เรียนที่ไม่เคยเรียน ให้หาความเที่ยงตรงตามโครงสร้างของแบบทดสอบฉบับนี้ ถ้ากำหนดว่าเกณฑ์ตัดสินผ่านเท่ากับ 5 คะแนน

ผู้ที่ไม่เคยเรียน	5	4	2	3	3	4	4	6	5	2	1	4
ผู้ที่เรียนแล้ว	5	7	6	8	4	7	6	5	3	7	6	8

จำนวนผู้เรียนที่เรียนแล้วและสอบผ่าน (a) = 10 คน

จำนวนผู้เรียนที่ยังไม่เคยเรียนและสอบไม่ผ่าน (b) = 9 คน

จำนวนผู้เรียนทั้งหมด (N) = 24 คน

$$Construct - Validity = \frac{a+c}{N}$$

$$= \frac{10+9}{24} = \frac{19}{24} = .79$$

แสดงว่าแบบทดสอบวิชาโครงสร้างข้อมูลมีความเที่ยงตรงตามโครงสร้าง .79

2. วิธีการหาค่าสหสัมพันธ์แบบฟี (Phi-Correlation)

วิธีการทดสอบความเที่ยงตรงตามโครงสร้างโดยการหาค่าสหสัมพันธ์แบบฟี โดยการหาความสัมพันธ์ของผู้เรียน 2 กลุ่ม ได้แก่ กลุ่มผู้เรียนที่ยังไม่ได้รับการสอนหรือไม่ได้สอบก่อนเรียนกับกลุ่มผู้เรียนที่เรียนแล้วหรือผ่านการสอบหลังเรียนแล้ว โดยกำหนดเกณฑ์การผ่านไว้ก่อน หลังจากนั้นจึงนำไปแทนค่าในสูตรการหาค่าสหสัมพันธ์แบบฟี

สูตรการหาค่าสหสัมพันธ์แบบฟี

$$f = \frac{ac - bd}{\sqrt{(a+b)(c+d)(a+d)(b+c)}}$$

เมื่อ

f = ความเที่ยงตรงตามโครงสร้าง

a = จำนวนผู้เรียนที่สอบก่อนเรียนและสอบไม่ผ่าน

b = จำนวนผู้เรียนที่สอบหลังเรียนและสอบไม่ผ่าน

c = จำนวนผู้เรียนที่สอบหลังเรียนและสอบผ่าน

d = จำนวนผู้เรียนที่สอบก่อนเรียนและสอบผ่าน

ตัวอย่าง

แบบทดสอบชุดหนึ่ง เมื่อนำไปทดสอบกับผู้เรียน ได้ผลปรากฏดังตาราง ให้หาค่าความเที่ยงตรงตามโครงสร้าง โดยวิธีการหาค่าสหสัมพันธ์แบบฟี

	ก่อนเรียน	หลังเรียน
สอบไม่ผ่าน	16	5
สอบผ่าน	3	14

จากสูตร $$f = \frac{ac - bd}{\sqrt{(a+b)(c+d)(a+d)(b+c)}}$$

$$f = \frac{(16)(14) - (5)(3)}{\sqrt{(16+5)(14+3)(16+3)(5+14)}}$$

$$f = \frac{224 - 15}{\sqrt{(21)(17)(19)(19)}} = \frac{209}{\sqrt{128877}}$$

$$f = \frac{209}{358.99} = .53$$

แสดงว่าแบบทดสอบชุดนั้นมีความเที่ยงตรงตามโครงสร้าง .53

3. ความเที่ยงตรงตามสภาพ (Concurrent Validity) หมายถึง แบบทดสอบที่สามารถวัดได้ตามสภาพความเป็นจริงของกลุ่มตัวอย่าง เช่น ถ้าผู้เรียนคนหนึ่งที่ในเวลาเรียนเป็นผู้เรียนที่เก่งที่สุดในชั้นเรียน เมื่อทำข้อสอบปรากฏว่าผู้เรียนผู้นั้นทำคะแนนได้สูงสุด แสดงว่าแบบทดสอบนั้นมีความเที่ยงตรงตามสภาพดี แต่ถ้าหากว่าผลการสอบออกมาตรงกันข้าม ผู้ที่ได้คะแนนสูงกลับไปเป็นผู้ที่เรียนอ่อนขณะที่เรียนในชั้นเรียน แสดงว่าแบบทดสอบนั้นมีความเที่ยงตรงตามสภาพไม่ดี

การทดสอบความเที่ยงตรงตามสภาพ ทำได้โดยนำคะแนนของแบบทดสอบที่สร้างขึ้นใหม่ไปหาค่าสหสัมพันธ์กับคะแนนของแบบทดสอบเดิมที่มีความเที่ยงตรง ความสัมพันธ์ระหว่างคะแนนของแบบทดสอบทั้งสอง ก็คือสหสัมพันธ์ของความเที่ยงตรง (Validity Coefficient) ซึ่งจะเป็นเครื่องชี้บ่งความเที่ยงตรงตามสภาพ ถ้าสหสัมพันธ์มีค่าสูงก็หมายความว่าแบบทดสอบที่สร้างขึ้นใหม่นั้นมีความเที่ยงตรงตามสภาพอยู่ในเกณฑ์ดี

4. ความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) หมายถึง การหาความสัมพันธ์ระหว่างคะแนนผลการสอบกับเกณฑ์ของความสำเร็จที่จะเกิดขึ้นในอนาคต โดยใช้คะแนนผลการสอบในการพยากรณ์ในอนาคต ถ้าหากแบบทดสอบมีความเที่ยงตรงเชิงพยากรณ์สูงและบุคคลผู้ใดทำคะแนนได้ดี จะสามารถพยากรณ์ได้ว่าบุคคลผู้นั้นย่อมมีความสำเร็จในสาขาวิชาที่เกี่ยวข้องในประเด็นของแบบทดสอบ ความเที่ยงตรงเชิงพยากรณ์เป็นสิ่งที่มีความสำคัญมากในปัจจุบันโดยเฉพาะการสอบคัดเลือกบุคคลเข้าทำงานหรือศึกษาต่อ หากข้อสอบคัดเลือกมีความเที่ยงตรงเชิงพยากรณ์อยู่ในเกณฑ์ดี ผู้ที่ได้คะแนนสูงและสอบการคัดเลือกผ่าน อาจจะพยากรณ์ได้ว่าบุคคลผู้นั้นจะพบกับความสำเร็จในการทำงานหรือการศึกษาต่อในอนาคต

การทดสอบความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบ ทำได้โดยการสร้างความสัมพันธ์ระหว่างคะแนนที่ได้จากแบบทดสอบกับเกณฑ์ที่ใช้ในการวัดความสำเร็จ แบบทดสอบที่ใช้เพื่อพยากรณ์ความสำเร็จเรียกว่า ตัวพยากรณ์ (Predictor) และพฤติกรรมที่ถูกพยากรณ์เรียกว่า เกณฑ์ (Criterion) ซึ่งจะต้องวัดอย่างเที่ยงตรงของพฤติกรรมที่จะถูกพยากรณ์ ในการหาความเที่ยงตรงเชิงพยากรณ์นั้นจะต้องนิยามตัวพยากรณ์และเกณฑ์เสียก่อน หลังจากนั้นจึงทดสอบตัวแปรที่ใช้เป็นตัวพยากรณ์ จากนั้นจึงรอจนกว่าพฤติกรรมที่จะถูกพยากรณ์เกิดขึ้นแล้วจึงวัดเกณฑ์จากกลุ่มเดิม หลังจากนั้นจึงหาความสัมพันธ์ของคะแนนทั้งสองชุดโดยใช้สูตรสหสัมพันธ์ของเพียร์สัน (วัญญา. 2531)

สูตรที่ใช้ในการหาค่าสหสัมพันธ์ของคะแนนทั้งสองชุด

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

เมื่อ

r_{XY} = สัมประสิทธิ์สหสัมพันธ์

N = จำนวนคู่ของคะแนน

X = คะแนนของคะแนนชุดที่หนึ่ง

Y = คะแนนของคะแนนชุดที่สอง

■ ความเชื่อมั่น (Reliability)

ความเชื่อมั่น (Reliability) หมายถึง ความคงที่ ความมั่นคง หรือความสม่ำเสมอของผลการวัด เช่น ถ้านำแบบทดสอบไปวัดสิ่งเดียวกันสองครั้งแล้วได้ผลไม่แตกต่างกัน ถือว่ามีความคงที่ของผลคะแนนที่ได้สูง อีกกรณีหนึ่งก็คือถ้าให้ทำแบบทดสอบฉบับเดียวกันสองครั้งในเวลาต่างกันและได้คะแนนเกือบเท่ากันทั้งสองครั้ง ก็จะหมายความว่าแบบทดสอบนั้นมีความเชื่อมั่นสูง ค่าของความเชื่อมั่นแสดงเป็นตัวเลขที่มีค่าไม่เกิน 1.00 หรือ 100% ซึ่งเรียกว่า สัมประสิทธิ์ (Coefficient) ถ้าแบบทดสอบมีค่าสัมประสิทธิ์สูง ก็แสดงว่ามีความเชื่อมั่นสูง

การหาค่าความเชื่อมั่น สามารถทำได้หลายวิธี ดังต่อไปนี้

1. การทดสอบซ้ำ (Test-Retest Reliability)
2. การทดสอบแบบใช้ข้อสอบเหมือนกัน (Equivalent-Forms Reliability)
3. การทดสอบแบบแบ่งครึ่ง (Split-Half Reliability)
4. การทดสอบโดยวิธีหาความคงที่ภายในโดยใช้ KR-20 และ KR-21
5. การทดสอบโดยวิธีหาสัมประสิทธิ์แอลฟา (α-Coefficient)

รายละเอียดแต่ละวิธี มีดังนี้

1. การทดสอบซ้ำ (Test-Retest Reliability) เป็นการทดสอบหาความเชื่อมั่นของแบบทดสอบโดยการทำแบบทดสอบฉบับเดียวกัน 2 ครั้งในเวลาต่างกัน หลังจากนั้นจึงนำค่าที่ได้จากการทดสอบทั้ง 2 ครั้งไปหาค่าสหสัมพันธ์ เพื่อหาความสอดคล้องของผลการทดสอบ โดยใช้สูตรของเพียร์สัน (Pearson Product-Moment Correlation) ค่าสัมประสิทธิ์ที่คำนวณได้เรียกว่าสัมประสิทธิ์ของความคงที่ (Coefficient of Stability) ถ้าได้ค่าสัมประสิทธิ์สูงก็หมายความว่าแบบทดสอบฉบับนี้มีความเชื่อมั่นสูง ปัญหาของการทดสอบเพื่อหาค่าความเชื่อมั่นวิธีนี้ก็คือ ระยะห่างของเวลาการทดสอบทั้ง 2 ครั้ง ถ้าระยะเวลาใกล้กันมากเกินไป การทดสอบครั้งแรกก็ย่อมส่งผลถึงการทดสอบครั้งหลัง เนื่องจากผู้สอบยังจำข้อสอบได้อยู่ ซึ่งมีผลต่อค่าสหสัมพันธ์ที่ได้ แต่ถ้าระยะเวลาห่างกันมาก ก็จะให้ผลที่ตรงกันข้าม

สูตรที่ใช้ในการหาค่าสหสัมพันธ์ของเพียร์สัน

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

เมื่อ

r_{XY} = สัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบ

N = จำนวนกลุ่มตัวอย่าง

X = คะแนนจากการทดสอบครั้งแรก

Y = คะแนนจากการทดสอบครั้งที่สอง

2. การทดสอบแบบใช้ข้อสอบเหมือนกัน (Equivalent-Forms Reliability) การหาความเชื่อมั่นวิธีนี้ทำได้โดยใช้แบบทดสอบ 2 ฉบับที่เหมือนกัน ทำในระยะเวลาที่ห่างกันเพียงเล็กน้อย แบบทดสอบที่เหมือนกันในที่นี้หมายความว่าทั้งสองฉบับวัดในสิ่งเดียวกัน จำนวนข้อเท่ากัน มีโครงสร้างเหมือนกัน มีความยากง่ายในระดับเดียวกัน มีวิธีการทดสอบ การตรวจให้คะแนนและการแปลความหมายของคะแนนเหมือนกัน จากนั้นจึงนำคะแนนจากผลการทดสอบทั้ง 2 ฉบับไปหาค่าสหสัมพันธ์ ค่าสัมประสิทธิ์ที่คำนวณได้เรียกว่าสัมประสิทธิ์ของความเหมือนกัน (Coefficient of Equivalence) โดยใช้สูตรในการคำนวณเช่นเดียวกันกับการหาค่าความเชื่อมั่นแบบทดสอบซ้ำ ปัญหาของการหาค่าความเชื่อมั่นวิธีนี้ก็คือ เป็นไปได้ยากที่จะจัดหาแบบทดสอบที่มีคุณลักษณะเหมือนกันทั้ง 2 ฉบับ ตามเงื่อนไขที่กำหนดข้างต้น

3. การทดสอบแบบแบ่งครึ่ง (Split-Half Reliability) การหาความเชื่อมั่นวิธีนี้เป็นการหาค่าความเชื่อมั่นแบบคงที่ภายใน (Internal Consistency Reliability) หาได้โดยการทดสอบเพียงครั้งเดียวโดยใช้แบบทดสอบเพียงฉบับเดียว จากนั้นจึงแบ่งออกเป็น 2 ส่วน ได้แก่ ข้อคู่กับข้อคี่

แล้วจึงนำไปหาค่าสหสัมพันธ์ระหว่างคะแนนข้อคู่กับข้อคี่ ค่าสหสัมพันธ์ที่ได้เป็นสัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบครึ่งฉบับ จากนั้นจึงไปหาค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับโดยใช้สูตรของสเปียร์แมน บราวน์ (Spearman-Brown)

สูตรการหาค่าความเชื่อมั่นทั้งฉบับของสเปียร์แมน บราวน์

$$r_t = \frac{2r_{1/2}}{1+r_{1/2}}$$

เมื่อ

r_t = สัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบทั้งฉบับ

$r_{1/2}$ = สัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบครึ่งฉบับ

ตัวอย่าง

ผู้ตอบ	คะแนนข้อคู่	คะแนนข้อคี่	XY	X^2	Y^2
1	6	7	42	36	49
2	9	10	90	81	100
3	7	7	49	49	49
4	8	9	72	64	81
5	10	9	90	100	81
6	8	7	56	64	49
7	7	8	56	49	64
8	6	7	42	36	49
9	7	8	56	49	64
10	7	6	42	49	36
	$\sum X = 75$	$\sum Y = 78$	$\sum XY = 595$	$\sum X^2 = 577$	$\sum Y^2 = 622$

จากตาราง เป็นผลจากการนำแบบทดสอบจำนวน 20 ข้อ นำไปทดสอบกับกลุ่มตัวอย่างจำนวน 10 คน จงหาความเชื่อมั่นแบบแบ่งครึ่งของแบบทดสอบฉบับนี้

สูตรที่ใช้ในการหาค่าสหสัมพันธ์ของเพียร์สัน

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

$$r_{XY} = \frac{10(595) - (75)(78)}{\sqrt{\left\{10(577) - 75^2\right\}\left\{10(622) - 78^2\right\}}} = \frac{100}{\sqrt{(145)(136)}} = .71$$

แสดงว่าแบบทดสอบฉบับนี้มีค่าสัมประสิทธิ์ความเชื่อมั่นครึ่งฉบับเท่ากับ .71
การหาค่าสัมประสิทธิ์ความเชื่อมั่นทั้งฉบับ ใช้สูตรดังนี้

แทนค่าในสูตร $r_t = \frac{2r_{1/2}}{1+r_{1/2}}$ ดังนั้น $r_t = \frac{2(.71)}{1+.71} = .83$

สรุปได้ว่า แบบทดสอบฉบับนี้มีสัมประสิทธิ์ความเชื่อมั่นทั้งฉบับเท่ากับ .83 หรือ 83%

4. การทดสอบโดยวิธีหาความคงที่ภายในโดยใช้ KR-20 และ KR-21 สำหรับการหาค่าความเชื่อมั่นแบบคูเดอร์-ริชาร์ดสัน (Kuder-Richardson) โดยใช้สูตร KR-20 และ KR-21 นั้น มิได้หาโดยการหาค่าสัมประสิทธิ์ความสัมพันธ์ แต่เป็นการทดสอบว่าแบบทดสอบแต่ละข้อมีความสัมพันธ์กับข้ออื่น ๆ ในฉบับเดียวกันหรือไม่ และมีความสัมพันธ์กับแบบทดสอบทั้งฉบับอย่างไร โดยใช้สูตร KR-20 หรือ KR-21 ก็ได้ ซึ่งค่าที่ได้จากการใช้สูตร KR-21 จะมีค่าต่ำกว่าสูตร KR-20 เล็กน้อย แต่ก็อยู่ในเกณฑ์ยอมรับได้เช่นเดียวกัน ปัญหาของการทดสอบโดยวิธีการหาความคงที่ภายในก็คือ จะต้องแปลงผลคำตอบก่อนนำไปแทนค่าในสูตร โดยกำหนดให้ข้อที่ตอบถูกมีค่าเท่ากับ 1 และตอบผิดมีค่าเท่ากับ 0 จึงมีข้อจำกัดในการใช้งานที่ใช้ได้เฉพาะแบบทดสอบชนิดเลือกตอบ (Multiple Choice) หรือแบบทดสอบอื่น ๆ ที่ให้คะแนนเป็น 0 และ 1 เท่านั้น สำหรับสูตร KR-20 และ KR-21 มีดังนี้

สูตรการหาค่าความคงที่ภายใน KR-20

$$r_t = \frac{n}{n-1}\left[1 - \frac{\sum pq}{s_t^2}\right]$$

เมื่อ

r_t = สัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบทั้งฉบับ
n = จำนวนข้อของแบบทดสอบ (ไม่ควรน้อยกว่า 20 ข้อ)
p = อัตราส่วนของผู้ที่ตอบแบบทดสอบข้อนี้ถูก (หาได้จากจำนวนผู้ที่ตอบถูก หารด้วยจำนวนทั้งหมด)
q = อัตราส่วนของผู้ที่ตอบข้อนี้ผิด (เท่ากับ 1 – p)
s_t^2 = ความแปรปรวนของคะแนนที่สอบได้ทั้งฉบับ

สูตรการหาค่าความคงที่ภายใน KR-21

$$r_t = \frac{n}{n-1}\left[1 - \frac{\overline{X}(n-\overline{X})}{ns_t^2}\right]$$

เมื่อ

r_t = สัมประสิทธิ์ของความเชื่อมั่นของแบบทดสอบทั้งฉบับ
n = จำนวนข้อในแบบทดสอบฉบับนั้น

$\overline{X}$ = ค่าเฉลี่ยของคะแนน

s_t^2 = ความแปรปรวนของคะแนนที่สอบได้ทั้งฉบับ

ตัวอย่าง

แบบทดสอบแบบเลือกตอบจำนวน 5 ข้อ เมื่อนำไปใช้ทดสอบกับกลุ่มตัวอย่างจำนวน 10 คน ปรากฏผลดังตารางต่อไปนี้ จงหาความเชื่อมั่นโดยใช้สูตรของคูเดอร์-ริชาร์ดสัน

ผู้เรียน	ข้อที่ 1	ข้อที่ 2	ข้อที่ 3	ข้อที่ 4	ข้อที่ 5	คะแนนรวม	X^2
1	1	1	1	1	1	5	25
2	1	1	1	0	1	4	16
3	1	1	1	1	1	5	25
4	1	0	1	0	1	3	9
5	1	0	0	1	1	3	9
6	1	1	0	1	0	3	9
7	1	0	0	0	1	2	4
8	1	0	0	1	0	2	4
9	1	0	1	0	0	2	4
10	1	0	0	0	0	1	1
$p_{ถูก}$	1.0	.4	.5	.5	.6	$\sum X = 30$	$\sum X^2 = 106$
$q_{ผิด}$	0	.6	.5	.5	.4		
pq	0	.24	.25	.25	.24	$\sum pq = .98$	

จากสูตรการหาความคงที่ภายใน KR-20

$$r_t = \frac{n}{n-1}\left[1 - \frac{\overline{X}(n-\overline{X})}{ns_t^2}\right]$$

สูตรหาค่าความแปรปรวน

$$s_t^2 = \frac{N\sum X^2 - \left(\sum X\right)^2}{N^2}$$

$$s_t^2 = \frac{10(106)^2 - (30)^2}{10^2} = 1.6$$

แทนค่าในสูตร $$r_t = \frac{5}{5-1}\left[1 - \frac{.98}{1.6}\right] = .48$$

ความเชื่อมั่นของแบบทดสอบ โดยใช้สูตร KR-20 มีค่าเท่ากับ .48 หรือ 48%

เมื่อใช้สูตร KR-21

$$r_t = \frac{n}{n-1}\left[1 - \frac{\overline{X}(n-\overline{X})}{ns_t^2}\right]$$

$$r_t = \frac{5}{5-1}\left[1 - \frac{3(5-3)}{5\times 1.6}\right] = .31$$

ความเชื่อมั่นของแบบทดสอบ โดยใช้สูตร KR-21 มีค่าเท่ากับ .31 หรือ 31%

แสดงว่าความเชื่อมั่นของแบบทดสอบยังมีคุณภาพไม่ดี เนื่องจากมีค่าสัมประสิทธิ์อยู่ที่ .48 เมื่อใช้สูตร KR-20 และมีค่าเท่ากับ .31 เมื่อใช้สูตร KR-21 โดยที่คุณภาพของแบบทดสอบที่อยู่ในเกณฑ์ที่ยอมรับได้ สามารถนำไปใช้เก็บรวบรวมข้อมูลในการวิจัยได้นั้น ควรมีค่าความเชื่อมั่นไม่ต่ำกว่า .60 หรือ 60% ขึ้นไป

สำหรับสูตร KR-21 จะให้ค่าที่คำนวณได้ต่ำกว่าที่ควรจะเป็น (Underestimate) จึงไม่เหมาะสำหรับแบบทดสอบหรือข้อสอบที่ง่ายมาก ๆ หรือยากมาก ๆ นอกจากนี้ยังต้องเป็นข้อสอบที่มีความยากแต่ละข้อใกล้เคียงกัน (บุญเรียง. 2543 : 58) จึงจะใช้สูตร KR-21 คำนวณได้เหมาะสม

5. การทดสอบโดยวิธีหาสัมประสิทธิ์แอลฟา (α-Coefficient) การหาค่าความเชื่อมั่นโดยวิธีนี้พัฒนามาจากสูตร KR-20 เนื่องจากวิธีการของคูเดอร์-ริชาร์ดสัน จะต้องแปลงคำตอบถูกให้เป็น 0 และคำตอบผิดให้เป็น 1 ก่อนวิเคราะห์ข้อมูลและแทนค่าในสูตร จึงเป็นข้อจำกัดอย่างหนึ่งในการนำไปใช้ซึ่งอาจส่งผลให้เกิดการผิดพลาดในการแปลงคำตอบได้ ถ้าหากแบบทดสอบมีเป็นจำนวนมาก วิธีนี้จึงพัฒนาขึ้นเพื่อให้ใช้ได้กับแบบทดสอบที่ไม่ได้ตรวจให้คะแนน เป็น 0 กับ 1 เช่น ข้อสอบแบบอัตนัยหรือข้อสอบแบบเติมคำ เป็นต้น เนื่องจากสูตรสัมประสิทธิ์แอลฟาจะใช้กับคะแนนที่ทำได้จริงหรือใช้กับแบบทดสอบที่ให้คะแนนแต่ละข้อเป็น 3, 2, 1 หรือ 5, 4, 3, 2, 1 ก็ได้ ดังนั้น การทดสอบโดยวิธีหาสัมประสิทธิ์แอลฟา จึงใช้ได้ทั้งแบบทดสอบแบบเลือกตอบ (Multiple Choice) และแบบทดสอบทั่ว ๆ ไป โดยใช้สูตรการหาค่าความเชื่อมั่นของครอนบัค (Cronbach)

สูตรการหาค่าความเชื่อมั่นของครอนบัค

$$a = \frac{n}{n-1}\left[1 - \frac{\sum S_i^2}{S_t^2}\right]$$

เมื่อ

a = สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบ

N = จำนวนข้อในแบบทดสอบ

S_i^2 = ความแปรปรวนของแบบทดสอบเป็นรายข้อ

S_t^2 = ความแปรปรวนของแบบทดสอบทั้งฉบับ

การทดสอบโดยวิธีสัมประสิทธิ์แอลฟา เพื่อหาค่าความเชื่อมั่นของแบบทดสอบที่ใช้ในการวิจัยจะให้ผลลัพธ์เป็นค่าสัมประสิทธิ์แอลฟา (Alpha Coefficient) ที่บ่งชี้ถึงลักษณะของแบบทดสอบว่าดีหรือไม่ดีซึ่งหมายถึงความผันแปรว่ามีมากหรือไม่ วิธีนี้จึงเป็นวิธีการหาค่าความเชื่อมั่นที่ให้รายละเอียดทางสถิติมากกว่าวิธีการอื่น ๆ ทำให้การหาค่าความเชื่อมั่นโดยใช้สูตรของครอนบัคได้รับความนิยมในการวิจัยค่อนข้างสูง โดยที่คะแนนของแบบทดสอบจะต้องเป็นคะแนนแบบมาตราเรียงลำดับหรืออันตรภาค

ตัวอย่าง

แบบทดสอบแบบอัตนัยจำนวน 10 ข้อ เมื่อนำไปทดลองใช้กับกลุ่มตัวอย่างจำนวน 12 คน ปรากฏผลดังตารางข้างล่าง จงหาความเชื่อมั่นของแบบทดสอบโดยวิธีการหาสัมประสิทธิ์แอลฟา

กลุ่มตัวอย่าง	ข้อ 1	ข้อ 2	ข้อ 3	ข้อ 4	ข้อ 5	ข้อ 6	ข้อ 7	ข้อ 8	ข้อ 9	ข้อ 10	X	X^2
1	7	6	8	9	5	6	4	6	7	8	66	4356
2	7	5	7	6	6	4	7	6	5	3	56	3136
3	4	8	6	3	2	7	4	2	8	6	50	2500
4	3	6	3	7	4	9	6	3	6	5	52	2704
5	6	5	4	5	6	3	5	2	7	4	47	2209
6	8	4	2	4	5	5	4	1	5	6	44	1936
7	2	3	5	3	3	4	3	6	4	3	36	1296
8	3	2	4	5	1	7	7	5	3	5	42	1764
9	2	4	6	4	2	6	6	4	2	4	40	1600
10	4	6	3	3	1	3	3	3	4	3	33	1089
11	1	1	2	2	2	4	2	4	3	1	22	484
12	5	2	1	2	5	2	2	2	1	2	24	576
$\sum X$	52	52	51	53	42	60	53	44	55	50	$\sum X = 512$	23650
$\sum X^2$	282	272	269	283	186	346	269	196	303	250		$\sum X^2$
S_i^2	4.72	3.89	4.35	4.08	3.25	3.83	2.91	2.86	4.24	3.47		

จากตาราง หาค่า

$$\sum S_i^2 = 4.72 + 3.89 + 4.35 + 4.08 + 3.25 + 3.83 + 2.91 + 2.86 + 4.24 + 3.47 = 37.64$$

สูตรหาค่าความแปรปรวน

$$s_t^2 = \frac{N\sum X^2 - \left(\sum X\right)^2}{N^2}$$

$$s_t^2 = \frac{N\sum X^2 - \left(\sum X\right)^2}{N^2}$$

$$s_t^2 = \frac{12 \times 12360 - (512)^2}{12^2} = 150.39$$

สูตรการหาค่าความเชื่อมั่นของครอนบัค

$$a = \frac{n}{n-1}\left[1 - \frac{\sum S_i^2}{S_t^2}\right]$$

$$a = \frac{10}{10-1}\left[1 - \frac{37.64}{150.39}\right] = .83$$

สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบมีค่าเท่ากับ .83 หรือ 83%

สำหรับการหาค่าความเชื่อมั่นของแบบทดสอบแนวอิงเกณฑ์ โดยเฉพาะแบบทดสอบที่ใช้ในการเรียนการสอน มีวิธีการทดสอบความเชื่อมั่นอยู่หลายวิธี ในที่นี้ขอนำเสนอ 2 วิธี ดังนี้

1. วิธีของคาร์เวอร์ (Carver Method)

วิธีการทดสอบความเชื่อมั่นของคาร์เวอร์ เป็นวิธีการหาความเชื่อมั่นแบบสอดคล้องในการตัดสินใจ (Decision Consistency Reliability) โดยการทดสอบกับผู้เรียนกลุ่มเดียวกันจำนวน 2 ครั้ง หรือใช้แบบทดสอบแบบคู่ขนานจำนวน 2 ฉบับแล้วทดสอบเพียงครั้งเดียว หลังจากนั้นจึงนำค่าที่ได้ทั้ง 2 ครั้ง (หรือ 2 ฉบับ) มาคำนวณในสูตรเพื่อหาค่าความเชื่อมั่นของแบบทดสอบ

สูตรการหาค่าความเชื่อมั่นของคาร์เวอร์

$$r = \frac{a+c}{N}$$

เมื่อ

r = ความเชื่อมั่นของแบบทดสอบ

a = จำนวนผู้เรียนที่สอบผ่านทั้งฉบับที่ 1 และฉบับที่ 2

b = จำนวนผู้เรียนที่สอบไม่ผ่านทั้งฉบับที่ 1 และฉบับที่ 2

N = จำนวนผู้เรียนทั้งหมด

ตัวอย่าง

แบบทดสอบวิชาโครงสร้างข้อมูล จำนวน 2 ฉบับ ๆ ละ 10 ข้อ คะแนนเต็ม 10 คะแนน นำไปทดสอบกับผู้เรียนกลุ่มเดียวกันจำนวน 12 คน 2 ครั้ง ปรากฏว่าได้ผลตามตาราง ให้หาความเชื่อมั่นของแบบทดสอบฉบับนี้ ถ้ากำหนดเกณฑ์ตัดสินผ่านเท่ากับ 5 คะแนน

ฉบับที่ 1	5	9	7	3	3	4	8	6	5	2	1	6
ฉบับที่ 2	5	7	6	2	4	7	6	5	8	4	6	8

จำนวนผู้เรียนที่สอบผ่านทั้งฉบับที่ 1 และฉบับที่ 2 (a) = 7
จำนวนผู้เรียนที่สอบไม่ผ่านทั้งฉบับที่ 1 และฉบับที่ 2 (c) = 3
จำนวนผู้เรียนทั้งหมด (N) = 12 คน

$$r = \frac{a+c}{N}$$
$$= \frac{7+3}{12} = \frac{10}{12} = .83$$

แสดงว่าแบบทดสอบวิชาโครงสร้างข้อมูลมีความเชื่อมั่น .83

2. วิธีของแฮมเบิลตันและโนวิก (Hambleton and Novick Method)

วิธีการทดสอบความเชื่อมั่นของแฮมเบิลตันและโนวิก เป็นวิธีการหาความเชื่อมั่นโดยการทดสอบกับผู้เรียนกลุ่มเดียวกันจำนวน 2 ครั้ง หรือใช้แบบทดสอบแบบคู่ขนานจำนวน 2 ฉบับ ทำการทดสอบเพียงครั้งเดียว หลังจากนั้นจึงนำค่าที่ได้ทั้ง 2 ครั้ง (หรือ 2 ฉบับ) มาคำนวณในสูตรเพื่อหาค่าความเชื่อมั่นของแบบทดสอบ

สูตรการหาค่าความเชื่อมั่นของแฮมเบิลตันและโนวิก

$$P_0 = P_{11} + P_{22}$$

เมื่อ

P_0 = ความเชื่อมั่นของแบบทดสอบ
P_{11} = สัดส่วนของผู้เรียนที่สอบผ่านทั้งฉบับที่ 1 และฉบับที่ 2
P_{22} = สัดส่วนของผู้เรียนที่สอบไม่ผ่านทั้งฉบับที่ 1 และฉบับที่ 2

ตัวอย่าง

แบบทดสอบวิชาระบบปฏิบัติการ นำไปทดสอบกับผู้เรียนกลุ่มเดียวกันจำนวน 50 คน 2 ครั้ง ปรากฏว่ามีผู้เรียนที่สอบผ่านทั้งสองครั้งจำนวน 32 คน และผู้เรียนที่สอบไม่ผ่านทั้งสองครั้งจำนวน 6 คน ให้หาความเชื่อมั่นของแบบทดสอบฉบับนี้

สัดส่วนของผู้เรียนที่สอบผ่านทั้งฉบับที่ 1 และฉบับที่ 2 (P_{11}) = 32/50 = .64
สัดส่วนของผู้เรียนที่สอบไม่ผ่านทั้งฉบับที่ 1 และฉบับที่ 2 (P_{22}) = 6/50 = .12

$$P_0 = P_{11} + P_{22}$$
$$P_0 = .64 + .12 = .76$$

แสดงว่าแบบทดสอบวิชาระบบปฏิบัติการมีความเชื่อมั่น .76

นอกจากวิธีต่าง ๆ ที่ได้นำเสนอมานี้ ยังมีวิธีอื่น ๆ ที่ใช้ในการหาค่าความเชื่อมั่นของแบบทดสอบหรือข้อสอบ เช่น วิธีของโลเวตต์ (Lowett Method) เป็นต้น

■ ความยากง่าย (Difficulty)

ความยากง่าย (Difficulty) มีความหมายตรงตัว หมายถึง ระดับความยากง่ายของแบบทดสอบหรือข้อสอบ โดยปกติแบบทดสอบที่ควรหาค่าความยากง่ายนั้นจะเป็นแบบทดสอบที่วัดทางด้านสติปัญญา (Cognitive Domain) ของผู้เรียน เช่น แบบทดสอบผลสัมฤทธิ์ทางการเรียน แบบทดสอบวัดความถนัด เป็นต้น แบบทดสอบประเภทนี้จะต้องมีคุณภาพทางด้านความยากง่าย (P) พอเหมาะ กล่าวคือ ผู้เรียนจะต้องทำได้ถูกต้อง 50% และทำผิด 50% หรือคิดเป็นสัดส่วนเท่ากับ .50 (P = .50) แต่การที่จะออกแบบทดสอบให้มีค่าความยากง่ายพอดี คือ P = .50 นั้นเป็นเรื่องยากมาก จะต้องนำไปทดสอบซ้ำหลายครั้งและทำการปรับปรุงจนกว่าข้อคำถามในแบบทดสอบมีค่าระดับความยากง่ายใกล้เคียงกับ P = .50

ในทางปฏิบัติ ข้อคำถามที่ถือว่ามีความยากง่ายใช้ได้มีค่าอยู่ระหว่าง .20 - .80 ถ้า P มีค่าต่ำกว่า .20 ถือว่าข้อคำถามนั้นยากเกินไป แต่ถ้าค่า P สูงกว่า .80 แสดงว่าง่ายเกินไป ค่าความยากง่ายจึงเป็นองค์ประกอบที่สำคัญด้านคุณภาพของแบบทดสอบที่ใช้วัดทางด้านสติปัญญา โดยเฉพาะอย่างยิ่ง การพัฒนาบทเรียนคอมพิวเตอร์ที่มักจะหาประสิทธิภาพของตัวบทเรียนด้วยคะแนนของผู้เรียนที่ทำได้จากแบบทดสอบก่อนและหลังบทเรียน แม้ว่าจะตั้งเกณฑ์ไว้สูงมาก เช่น 95/95 หากแบบทดสอบที่ใช้ตัดสินเกณฑ์มีค่าความยากง่ายอยู่สูงเกินไป (P เกินกว่า .80) การที่จะเข้าถึงเกณฑ์ที่กำหนดก็ไม่ใช่เรื่องยากอีกต่อไป ซึ่งเป็นเรื่องที่ไม่ถูกต้อง ดังนั้น แบบทดสอบที่ใช้ในการเรียนการสอน จึงต้องผ่านการหาค่าความยากง่ายมาก่อนและคัดเลือกข้อคำถามที่มีค่าความยากง่ายพอเหมาะเพื่อนำไปใช้งาน

สูตรคำนวณหาค่าความยากง่าย

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

เมื่อ

P = ค่าความยากง่ายของแบบทดสอบ
R = จำนวนผู้เรียนที่ตอบข้อคำถามข้อนั้นถูกต้อง
N = จำนวนผู้เรียนทั้งหมด

ตัวอย่าง

แบบทดสอบวิชาการโปรแกรมภาษาซี จำนวน 5 ข้อ เฉพาะข้อที่ 1 มีผู้เรียนทำถูกต้อง 14 คน จากผู้เรียนทั้งหมด 30 คน จงคำนวณหาค่าความยากง่ายของแบบทดสอบข้อที่ 1

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

$$P = \frac{14}{30} = .46$$

แสดงว่า แบบทดสอบข้อที่ 1 นี้ มีค่าความยากง่าย .46 หรือเท่ากับ 46 % สามารถแปลความได้ว่าเป็นข้อสอบที่มีความยากง่ายอยู่ในเกณฑ์เหมาะสม (มีค่าระหว่าง .20 - .80) สามารถนำไปใช้งานได้

ความยากง่าย จัดว่าเป็นเกณฑ์การหาคุณภาพของแบบทดสอบหรือข้อสอบที่มีความหมายตรงตัวและหาได้ง่าย แต่มีประโยชน์ต่อการนำไปใช้ โดยพิจารณาจากสัดส่วนของผู้ตอบถูกและตอบผิด หากแบบทดสอบข้อใดมีผู้ตอบผิดมากกว่าตอบถูกก็แสดงว่ายาก หากตอบถูกมากกว่าตอบผิดก็แสดงว่าง่าย

■ อำนาจจำแนก (Discrimination)

อำนาจจำแนก (Discrimination) หมายถึง ความสามารถของแบบทดสอบในการจำแนกกลุ่มตัวอย่างซึ่งอาจหมายถึงผู้เรียนหรือผู้ตอบแบบทดสอบ ออกเป็นกลุ่มต่าง ๆ เช่น กลุ่มเก่งและกลุ่มอ่อน กลุ่มที่เห็นด้วยและกลุ่มที่ไม่เห็นด้วย เป็นต้น ค่าอำนาจจำแนกแทนด้วยสัญลักษณ์ D ซึ่งมีค่าอยู่ระหว่าง +1.00 ถึง –1.00 ถ้าคำถามข้อใดมีค่า D เป็นบวกสูง แสดงว่าข้อคำถามนั้นสามารถจำแนกกลุ่มเก่งออกจากกลุ่มอ่อนได้ดี ซึ่งมีการแจกแจงระดับของค่าอำนาจจำแนกสำหรับแบบทดสอบที่ใช้วัดผลสัมฤทธิ์ทางการเรียนไว้ดังนี้

D > .40	หมายถึง	มีอำนาจจำแนกดีมาก
D .30 - .39	หมายถึง	มีอำนาจจำแนกดี
D .20 - .29	หมายถึง	มีอำนาจจำแนกพอใช้ได้ แต่ควรนำไปปรับปรุงใหม่
D < .19	หมายถึง	มีอำนาจจำแนกไม่ดี ต้องตัดทิ้งไป

การหาค่าอำนาจจำแนกมีหลายวิธี ดังนี้

1. การใช้วิธีการตรวจให้คะแนน
2. การใช้สูตรสัดส่วน
3. การใช้ค่าสหสัมพันธ์แบบพอยท์-ไบซีเรียล (Point-Biserial Correlation)
4. การใช้ตารางสำเร็จของจุงเตฟาน (Chung Teh Fan)

รายละเอียดแต่ละวิธี มีดังนี้

1. การใช้วิธีการตรวจให้คะแนน

การใช้วิธีการตรวจให้คะแนน เริ่มจากนำแบบทดสอบที่ต้องการหาค่าอำนาจจำแนกไปทดสอบกับกลุ่มตัวอย่างแล้วตรวจให้คะแนน จากนั้นจึงเรียงผลคะแนนที่ได้จากคะแนนสูงไปหาต่ำ แล้วทำการคัดเลือกกลุ่มที่ได้คะแนนสูงออกมา 1/3 ของจำนวนกลุ่มตัวอย่างทั้งหมด เรียกว่ากลุ่มเก่งหรือกลุ่มสูง หลังจากนั้นจึงคัดเลือกกลุ่มที่ได้คะแนนต่ำออกมา 1/3 ของจำนวนกลุ่มตัวอย่างทั้งหมด เรียกว่ากลุ่มอ่อนหรือกลุ่มต่ำ แล้วนำมาแทนค่าในสูตร

สูตรการหาค่าอำนาจจำแนกแบบตรวจให้คะแนน

$$D = \frac{R_U - R_L}{N/2} \quad \text{หรือ} \quad D = \frac{R_U - R_L}{R_U}$$

เมื่อ

D = ค่าอำนาจจำแนก

R_U = จำนวนกลุ่มตัวอย่างที่ตอบถูกในกลุ่มเก่ง

R_L = จำนวนกลุ่มตัวอย่างที่ตอบถูกในกลุ่มอ่อน

N = จำนวนกลุ่มตัวอย่างทั้งหมด

ตัวอย่าง

ผลการทดสอบวิชาภาษาอังกฤษ มีผู้เข้าทดสอบจำนวน 40 คน เมื่อนำมาวิเคราะห์หาค่าความยากง่าย ปรากฏว่ากลุ่มเก่งทำได้ทั้ง 13 คน และกลุ่มอ่อนทำได้ 6 คน จงหาค่าอำนาจจำแนก

จำนวนผู้เรียนที่นำมาวิเคราะห์หาค่าอำนาจจำแนก = 1/3 x40 = 13.33 คน

แทนค่าในสูตร $D = \frac{R_U - R_L}{N/2}$

$$D = \frac{13-6}{26/2} = .53$$

แสดงว่า ข้อคำถามข้อนี้มีค่าอำนาจจำแนก .53 ซึ่งจัดว่าเป็นข้อคำถามที่ดี

การหาค่าอำนาจจำแนกด้วยวิธีนี้ สามารถแบ่งกลุ่มสูงและกลุ่มต่ำได้หลายวิธี เช่น 25%, 27%, 33% หรือแบ่งเป็น 1/2 (50%) เพื่อให้ง่ายต่อการคำนวณ สำหรับวิธีที่นิยมใช้มากที่สุดก็คือ 27% อย่างไรก็ตามจำนวนประชากร 50% ทั้งกลุ่มสูงและกลุ่มต่ำ จะต้องมีจำนวนไม่น้อยกว่า 1/3 ของจำนวนทั้งหมด

2. การใช้สูตรสัดส่วน

การหาค่าอำนาจจำแนกโดยการใช้สูตรสัดส่วนมีวิธีการคล้ายคลึงกับวิธีแรก โดยนำผลคะแนนที่ผู้เรียนทำได้มาเรียงลำดับจากคะแนนสูงไปต่ำ หลังจากนั้นจึงนำมาแทนค่าในสูตรสัดส่วน ซึ่งเป็นวิธีขั้นพื้นฐาน

สูตรที่ใช้ในการหาค่าอำนาจจำแนกแบบสูตรสัดส่วน

$$D = P_H - P_L$$

เมื่อ

P_H = สัดส่วนของกลุ่มเก่ง

P_L = สัดส่วนของกลุ่มอ่อน

ตัวอย่าง

ข้อคำถามข้อหนึ่ง กลุ่มเก่งทำถูก 12 คน กลุ่มอ่อนทำถูก 5 คน จะมีค่าอำนาจจำแนกเท่าไร ถ้าแต่ละกลุ่มมีจำนวนทั้งหมด 12 คน

จากสูตรที่ใช้ในการหาค่าอำนาจจำแนกแบบสูตรสัดส่วน

$$D = P_H - P_L$$

เมื่อ $P_H = 12/12 = 1$

$P_H = 5/12 = .42$

แทนค่าในสูตร $D = 1 - .42 = .58$

แสดงว่า ข้อคำถามข้อนี้มีค่าอำนาจจำแนก .58

3. การใช้ค่าสหสัมพันธ์แบบพอยท์-ไบซีเรียล (Point-Biserial Correlation)

การหาค่าอำนาจจำแนกโดยวิธีการใช้ค่าสหสัมพันธ์แบบพอยท์-ไบซีเรียล มีข้อตกลงเบื้องต้นว่า ถ้าผู้เรียนทำถูกให้คะแนน 1 และทำผิดได้คะแนน 0 หลังจากนั้นจึงนำมาแทนค่าในสูตรสหสัมพันธ์แบบพอยท์-ไบซีเรียล

สูตรที่ใช้ในการหาค่าอำนาจจำแนกแบบพอยท์-ไบซีเรียล

$$r_{p.bis} = \frac{\overline{X}_p - \overline{X}_f}{S_t} \cdot \sqrt{pq}$$

เมื่อ

$r_{p.bis}$ = ค่าอำนาจจำแนกแบบพอยท์-ไบซีเรียล

$\overline{X}_p$ = คะแนนเฉลี่ยของกลุ่มตัวอย่างที่ทำแบบทดสอบข้อนั้นได้

$\overline{X}_f$ = คะแนนเฉลี่ยของกลุ่มตัวอย่างที่ทำแบบทดสอบข้อนั้นไม่ได้

S_t = ค่าเบี่ยงเบนมาตรฐานของแบบทดสอบฉบับนั้น

p = สัดส่วนของกลุ่มตัวอย่างที่ทำแบบทดสอบข้อนั้นได้

q = สัดส่วนของกลุ่มตัวอย่างที่ทำแบบทดสอบข้อนั้นไม่ได้ (1-p)

ตัวอย่าง

ผลการวิเคราะห์แบบทดสอบข้อที่ 10 ของวิชาการโปรแกรมภาษาซี ซึ่งมีผู้เรียนเข้าทดสอบจำนวน 15 คน ได้ผลตามตาราง จงหาค่าอำนาจจำแนกของแบบทดสอบข้อนี้

ผู้เรียน	คะแนน X	ข้อที่ 10	XY	X^2
1	25	1	25	625
2	23	1	23	529
3	18	0	0	324
4	24	0	0	576
5	23	1	23	529
6	20	0	0	400
7	19	0	0	361
8	22	1	22	484
9	21	1	21	441
10	23	1	23	529
11	21	0	0	441
12	20	0	0	400
13	21	1	21	441
14	21	1	21	441
15	22	1	22	484
			$\sum XY = 201$	$\sum X^2 = 7005$

จาก $$\overline{X}_p = \frac{\sum XY}{n_p} = \frac{201}{9} = 22.33$$

$$\overline{X}_f = \frac{\sum X - \sum XY}{n_f} = \frac{323 - 201}{6} = 20.33$$

เมื่อ $$p = \frac{9}{15} = .60 \quad \text{และ} \quad q = 1 - .60 = .40$$

เมื่อ $$S_t = \sqrt{\frac{N\sum X^2 - \left(\sum X\right)^2}{N(N-1)}}$$

ดังนั้น $$S_t = \sqrt{\frac{15\times 7005 - (323)^2}{15(14)}} = 1.82$$

จากสูตร $$r_{p.bis} = \frac{22.33 - 20.33}{1.82}.\sqrt{.60\times .40} = .54$$

แสดงว่าแบบทดสอบข้อที่ 10 มีค่าอำนาจจำแนกเป็น .54

4. การใช้ตารางสำเร็จของจุงเตฟาน (Chung Teh Fan)

จุงเตฟาน (Chung Teh Fan) ได้คิดค้นตารางสำเร็จ เพื่อแก้ปัญหาการคำนวณที่ซับซ้อนของวิธีการหาค่าอำนาจจำแนกโดยวิธีคำนวณ ตารางสำเร็จรูปของจุงเตฟานจะใช้วิธีแบ่งกลุ่มเก่งและกลุ่มอ่อนโดยใช้วิธี 27% โดยถือว่าการกระจายของคะแนนอยู่ในลักษณะของเส้นโค้งปกติ วิธีการนี้จึงเหมาะสำหรับการวิเคราะห์แบบทดสอบที่มีผู้เข้าสอบเป็นจำนวนมาก โดยมีข้อกำหนดเบื้องต้นว่าถ้าผู้ทำถูกได้ 1 และทำผิดได้ 0 เช่นเดียวกับแบบพอยท์-ไบซีเรียล จากนั้นจึงนำคะแนนมาเรียงลำดับจากสูงไปยังต่ำ แล้วคัดเลือก 27% ของกลุ่มที่ได้คะแนนสูงเป็นกลุ่มเก่ง และคัดเลือก 27% ของกลุ่มที่ได้คะแนนต่ำเป็นกลุ่มอ่อน ต่อจากนั้นก็นำมาแจกแจงแต่ละข้อคำถามของแบบทดสอบนั้นว่ากลุ่มเก่งทำถูกกี่คนและกลุ่มอ่อนทำถูกกี่คน

เมื่อแจกแจงกลุ่มเก่งและกลุ่มอ่อนว่าทำแบบทดสอบถูกกี่คนแล้ว จึงเปลี่ยนเป็นสัดส่วนของ P_H และ P_L แล้วนำสองค่านี้ไปเปิดตารางสำเร็จของจุงเตฟาน ในตารางจะบอกค่าอำนาจจำแนกเป็นจุดทศนิยม พร้อมทั้งบอกค่าความยากง่ายของแบบทดสอบอีกด้วย

ตัวอย่างแบบทดสอบข้อหนึ่ง มีกลุ่มเก่งทำถูก 40 คน จากจำนวนกลุ่มผู้เรียนเก่งทั้งหมด 50 คน ดังนั้น สัดส่วนของกลุ่มเก่ง (P_H) เท่ากับ 40/50 = .80 และกลุ่มอ่อนทำถูก 10 คน จากจำนวนกลุ่มผู้เรียนอ่อนทั้งหมด 50 คน สัดส่วนของกลุ่มอ่อน (P_L) เท่ากับ 10/50 = .20 หลังจากนั้นจึงนำค่า P_H = .80 และ P_L = .20 ไปเปิดตารางสำเร็จของจุงเตฟาน จะพบว่าได้ค่าอำนาจจำแนก (r) เท่ากับ .50 และได้ค่าความยากง่าย (p) เท่ากับ .50 และได้ค่าความยากง่ายมาตรฐาน (Δ) = 12.9

ค่าอำนาจจำแนกเป็นองค์ประกอบที่สำคัญ ที่ส่งผลต่อคุณภาพของแบบทดสอบ โดยเฉพาะแบบทดสอบที่ใช้ในการเรียนการสอน ถ้าแบบทดสอบมีค่าอำนาจจำแนกต่ำ ๆ เช่น .20 แสดงว่าจะมีประสิทธิภาพในการพยากรณ์ว่าเป็นผู้เรียนเก่งหรือผู้เรียนอ่อนได้ถูกต้องเพียง 2% เท่านั้น ซึ่งสามารถแปลความได้ว่ามีเพียง 2 คนเท่านั้นที่ทำแบบทดสอบข้อนั้นถูกแล้วเป็นผู้เรียนเก่ง ดังนั้นถ้าค่าอำนาจจำแนกต่ำแล้ว ประสิทธิภาพในการพยากรณ์ก็จะมีค่าต่ำตามไปด้วย ในทางปฏิบัติจึงต้องออกแบบทดสอบให้มีคุณภาพ โดยพยายามให้มีค่าอำนาจจำแนกสูง ๆ เนื่องจากยิ่งค่าสูงก็

สามารถจำแนกผู้เรียนได้ดี แต่ถ้าค่าอำนาจจำแนกติดลบจะแสดงว่าแบบทดสอบข้อนั้นผู้เรียนอ่อนตอบถูกมากกว่าผู้เรียนเก่ง ซึ่งเป็นข้อคำถามที่ใช้ไม่ได้ ต้องตัดทิ้งไป

■ ความเป็นปรนัย (Objectivity)

ความเป็นปรนัย (Objectivity) หมายถึง ความชัดเจนของแบบทดสอบที่ทุกคนอ่านแล้วตีความตรงกัน รวมทั้งการตรวจให้คะแนนมีเกณฑ์ที่แน่นอนไม่ว่าผู้ใดจะเป็นผู้ตรวจก็ตาม ลักษณะของแบบทดสอบที่มีความเป็นปรนัย จึงเกี่ยวข้องกับองค์ประกอบ 3 ประการ ได้แก่

1. ความแจ่มชัดในความหมายของแบบทดสอบ
2. ความแจ่มชัดในวิธีตรวจหรือมาตรฐานการให้คะแนน
3. ความแจ่มชัดในการแปลความหมายของคะแนน

แม้ว่าความเป็นปรนัยของแบบทดสอบ จะไม่มีเครื่องมือหรือวิธีการที่แน่นอนในการบ่งชี้คุณภาพ แต่การหาคุณภาพด้านนี้ของเครื่องมือจะหลีกเลี่ยงไม่ได้ เนื่องจากเป็นการทำให้เกิดคุณภาพทางด้านความเชื่อมั่นสูงและสร้างความเที่ยงตรงของการวัด นับตั้งแต่คำสั่งและเงื่อนไขในการทำแบบทดสอบ รวมถึงข้อคำถามต่าง ๆ ต้องมีความชัดเจนว่าต้องการสิ่งใด คำตอบที่ต้องการเป็นอะไร ไม่ว่าผู้ใดอ่านก็ตามจะเข้าใจตรงกันว่าถามอะไร และการตรวจให้คะแนนต้องมีเกณฑ์ในการให้คะแนนที่แน่นอน รวมทั้งการแปลความหมายของคะแนนที่ได้ต้องมีความชัดเจน โดยเฉพาะอย่างยิ่งแบบทดสอบที่ใช้ในบทเรียนคอมพิวเตอร์ ซึ่งผู้เรียนเป็นผู้ควบคุมกิจกรรมการเรียนด้วยตนเองทั้งหมด ความชัดเจนของแบบทดสอบที่ใช้ จึงต้องผ่านการหาคุณภาพมาก่อน โดยผ่านการทดลองใช้เพื่อหาความเหมาะสมกับผู้เรียนที่เป็นกลุ่มเป้าหมายโดยตรง หรือผ่านการประเมินโดยผู้เชี่ยวชาญมาก่อน

■ การหาคุณภาพของแบบทดสอบสำหรับบทเรียนคอมพิวเตอร์

แบบทดสอบหรือข้อสอบสำหรับบทเรียนคอมพิวเตอร์ จำแนกออกตามลักษณะของการใช้งานได้หลายประเภท ดังต่อไปนี้

1. แบบทดสอบที่ใช้สำหรับการประเมินผลระหว่างดำเนินการ (Formative Evaluation) ซึ่งประกอบด้วย

1.1 แบบทดสอบก่อนบทเรียน (Pretest) เป็นแบบทดสอบที่ใช้ประเมินผลผู้เรียนก่อนที่จะเข้าสู่กระบวนการเรียนรู้ โดยมีวัตถุประสงค์เพื่อตรวจสอบความรู้พื้นฐานของผู้เรียนว่ามีเพียงพอหรือไม่ และเพื่อเป็นการจัดระดับความสามารถของผู้เรียน โดยผู้เรียนที่มีระดับความสามารถสูงกว่าผู้เรียนคนอื่น ๆ บทเรียนอาจจะแนะนำหรือดำเนินการให้ข้ามไปศึกษาในบทเรียนถัดไปก็ได้

1.2 แบบฝึกหัดระหว่างบทเรียน (Exercise) เป็นแบบทดสอบที่ใช้วัดความก้าวหน้าทางการเรียนเป็นระยะ ๆ ในระหว่างกระบวนการเรียนรู้ โดยทั่วไปจะเป็นแบบฝึกหัดหลังบทเรียนแต่ละบทหรือแต่ละโมดูล

1.3 แบบประเมินผลอื่น ๆ เช่น ใบงาน ใบการบ้าน หรือแบบประเมินอื่น ๆ ที่ใช้วัดและประเมินผลระหว่างกระบวนการเรียนรู้ ซึ่งอาจจะมีการบันทึกผลคะแนนของผู้เรียนหรือไม่ก็ตาม

2. แบบทดสอบที่ใช้สำหรับการประเมินผลสรุป (Summative Evaluation) ซึ่งหมายถึงแบบทดสอบหลังบทเรียน (Posttest) เป็นแบบทดสอบที่ใช้ประเมินผลผู้เรียนหลังสิ้นสุดกระบวนการเรียนรู้ โดยมีวัตถุประสงค์เพื่อตรวจสอบความรู้ของผู้เรียนว่าผ่านเกณฑ์ที่กำหนดไว้หรือไม่ และเพื่อนำผลคะแนนไปตัดสินผลการสอบได้-ตก รวมทั้งการนำไปใช้ประโยชน์อย่างอื่น ๆ เช่น นำไปพิจารณาปรับปรุงแก้ไขบทเรียน นำไปใช้หาคุณภาพของบทเรียน หาความคงทนทางการเรียน และนำไปเปรียบเทียบกับการเรียนรู้แบบอื่น ๆ เป็นต้น

หลังจากที่ออกแบบทดสอบเพื่อใช้ในบทเรียนคอมพิวเตอร์ โดยพิจารณาจากพฤติกรรมที่คาดหวังในแต่ละข้อของวัตถุประสงค์เชิงพฤติกรรมของบทเรียน ซึ่งปกติมักจะออกแบบทดสอบจำนวน 2 - 4 ข้อต่อหนึ่งวัตถุประสงค์ ขั้นต่อไปจะเป็นการรวบรวมแบบทดสอบทั้งหมด รวมทั้งแบบฝึกหัดที่จะใช้ระหว่างบทเรียน เพื่อนำไปทดลองหาคุณภาพกับผู้เรียนกลุ่มเดียวกันกับกลุ่มเป้าหมายที่จะเป็นผู้ใช้บทเรียน เพียงแต่ว่ากลุ่มที่จะใช้ทดลองหาคุณภาพของแบบทดสอบจะต้องผ่านการเรียนหัวเรื่องดังกล่าวมาก่อน เช่น ต้องการพัฒนาบทเรียนคอมพิวเตอร์ในหัวเรื่องโครงสร้างข้อมูล ซึ่งเป็นรายวิชาในหลักสูตรประกาศนียบัตรวิชาชีพชั้นสูง ปีที่ 1 ภาคเรียนที่ 2 (ป.วส. 1) กลุ่มผู้เรียนที่จะใช้ทดลองหาคุณภาพของแบบทดสอบ จะต้องเป็นผู้เรียนที่เคยศึกษาหัวเรื่องนี้มาก่อน ได้แก่ นักศึกษา ป.วส. ชั้นปีที่ 2 ที่กำลังศึกษาอยู่ในปัจจุบัน

ขั้นตอนการหาคุณภาพของแบบทดสอบสำหรับบทเรียนคอมพิวเตอร์ มีดังนี้

1. วางแผนการดำเนินงาน

การวางแผนการดำเนินงาน เป็นการเตรียมการทดลองเพื่อหาคุณภาพของแบบทดสอบของบทเรียนคอมพิวเตอร์ ได้แก่ การศึกษาหลักสูตรรายวิชา การรวบรวมแบบทดสอบที่ออกไว้ใช้ในบทเรียน จัดเตรียมและพิมพ์แบบทดสอบ เตรียมการด้านสถานที่ และกำหนดการต่าง ๆ รวมทั้งการศึกษาถึงวิธีการหาคุณภาพของแบบทดสอบ และเทคนิคต่าง ๆ ที่ใช้ในการวิเคราะห์คุณภาพโดยใช้สถิติ

2. กำหนดกลุ่มตัวอย่าง

กลุ่มตัวอย่างในการทดลองหาคุณภาพของแบบทดสอบ ก่อนที่จะนำไปใช้ในบทเรียนคอมพิวเตอร์ จำแนกออกเป็น 2 กลุ่ม ได้แก่ กลุ่มทดลองแบบทดสอบขั้นต้น และกลุ่มทดลองแบบทดสอบขั้นใช้งานจริง ซึ่งทั้งสองกลุ่มต้องเป็นผู้เรียนที่เคยศึกษาในหัวเรื่องดังกล่าวมาก่อน

กลุ่มทดลองแบบทดสอบขั้นต้น ควรมีจำนวน 3 – 9 คน คัดเลือกโดยวิธีเฉพาะเจาะจงให้ได้ผู้เรียนที่มีความสามารถทางการเรียนระดับเก่ง ปานกลาง และอ่อน จำนวนเท่า ๆ กัน โดยใช้ผลการเรียนในรายวิชาพื้นฐานของหัวเรื่องที่พัฒนาเป็นบทเรียนคอมพิวเตอร์ หรือใช้ผลคะแนน GPA

ของผู้เรียนเป็นเกณฑ์ สำหรับกลุ่มทดลองแบบทดสอบขั้นใช้งานจริง ส่วนใหญ่จะคัดเลือกมาจากผู้เรียนทั้งชั้น โดยมีจำนวนไม่น้อยกว่า 35 คน

ในขั้นตอนนี้ ยังต้องพิจารณาถึงกลุ่มผู้เชี่ยวชาญที่จะเป็นผู้ให้คำปรึกษาเกี่ยวกับคุณภาพของแบบทดสอบอีกด้วย เนื่องจากองค์ประกอบด้านคุณภาพของแบบทดสอบ สามารถให้ผู้เชี่ยวชาญตรวจสอบและให้คำแนะนำได้ เช่น ค่าความเที่ยงตรง และความเป็นปรนัยของแบบทดสอบ

3. ทดลองใช้ขั้นต้น

เป็นขั้นตอนการทดลองใช้แบบทดสอบกับกลุ่มทดลองกลุ่มย่อยจำนวน 3 – 9 คนที่คัดเลือกไว้แล้ว โดยให้กลุ่มทดลองดำเนินการทำแบบทดสอบโดยลำพังตั้งแต่ต้นจนจบเหมือนกับการใช้งานจริง ผู้ควบคุมการดำเนินการมีหน้าที่เพียงแต่อธิบายวิธีการทำเท่านั้น หลังจากนั้นจึงนำผลที่ได้ไปวิเคราะห์ข้อมูลว่าผลที่ได้เป็นอย่างไร ขณะเดียวกันอาจต้องมีการสัมภาษณ์กลุ่มทดลองถึงความเข้าใจในข้อคำถาม ความยากง่าย และบันทึกเวลาที่ใช้ทำแบบทดสอบ เพื่อนำข้อมูลเหล่านี้ไปปรับปรุงแบบทดสอบต่อไป

4. จัดเตรียมและพิมพ์แบบทดสอบ

เมื่อทำการปรับปรุงแก้ไขแบบทดสอบตามข้อมูลที่ได้รับจากการทดลองใช้งานขั้นต้นแล้ว ในขั้นนี้จะเป็นการจัดเตรียมและพิมพ์แบบทดสอบให้เพียงพอกับความต้องการ เพื่อที่จะได้นำไปทดลองใช้งานจริงกับกลุ่มตัวอย่าง เพื่อวิเคราะห์หาคุณภาพของแบบทดสอบต่อไป ถ้าเป็นแบบทดสอบที่นำเสนอโดยใช้คอมพิวเตอร์ ในขั้นนี้จะหมายถึงการเตรียมพร้อมทางด้านซอฟท์แวร์ให้สอดคล้องกับฮาร์ดแวร์

5. ทดลองใช้จริง

การทดลองใช้จริงเป็นขั้นตอนที่คล้ายกับขั้นตอนใช้ขั้นต้นที่ผ่านมา เพียงแต่เป็นการกระทำกับกลุ่มตัวอย่างจำนวนไม่ต่ำกว่า 35 คน เหมือนกับการสอบวัดผลซึ่งเป็นกิจกรรมปกติของผู้เรียน ในขั้นนี้ผู้ดำเนินการจะต้องจดบันทึกข้อคำถามที่ผู้เรียนมีปัญหา เพื่อนำไปปรับปรุงแก้ไขต่อไป หลังจากนั้นจึงนำแบบทดสอบที่ผ่านการทำเสร็จสิ้นแล้วไปวิเคราะห์หาคุณภาพต่อไป

6. วิเคราะห์หาคุณภาพ

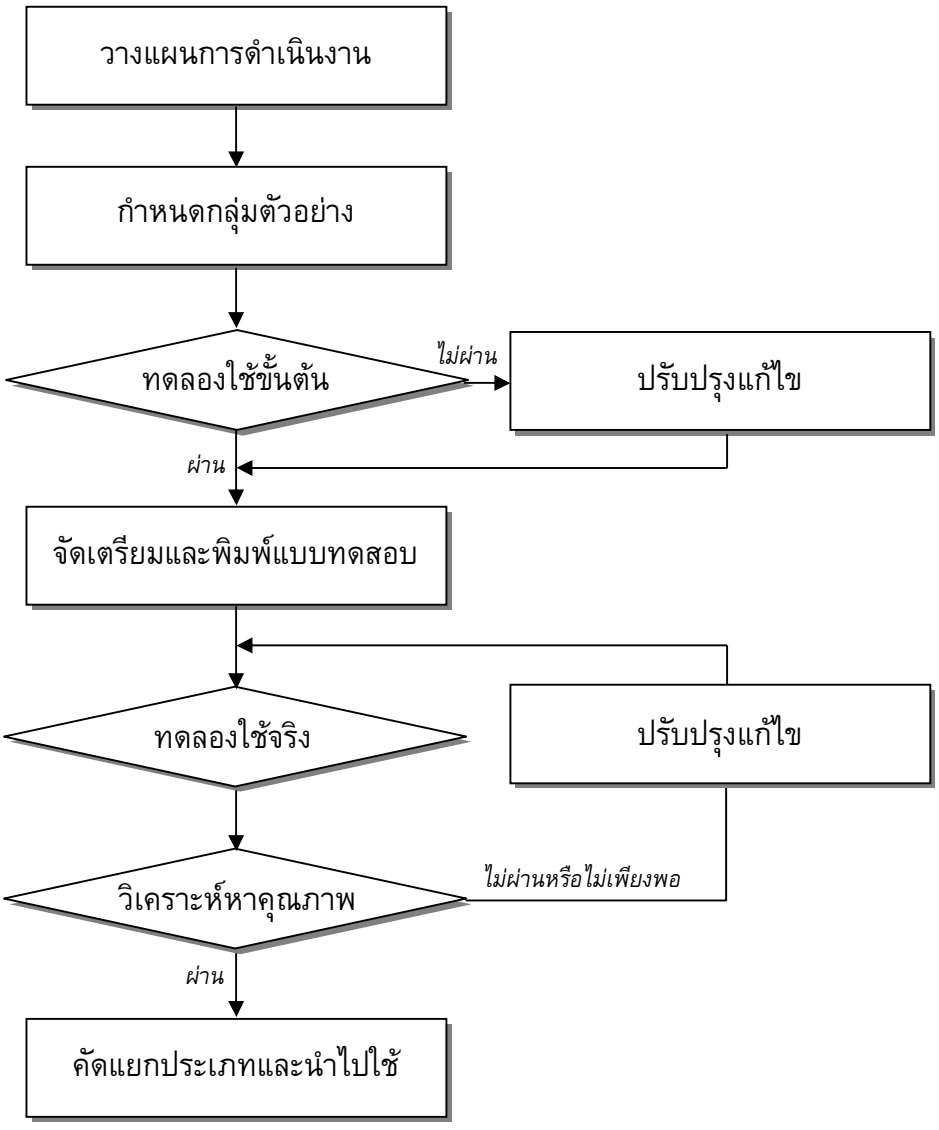
การวิเคราะห์หาคุณภาพของแบบทดสอบ ได้แก่ ค่าความเชื่อมั่น ค่าความยากง่าย และค่าอำนาจจำแนก ซึ่งต้องใช้การคำนวณทางสถิติหรือเปิดตารางสำเร็จรูป ส่วนค่าความเที่ยงตรง และความเป็นปรนัย จะได้จากการสอบถามผู้เชี่ยวชาญ โดยการหาค่า IOC

ผลลัพธ์ที่ได้ในขั้นตอนนี้ก็คือ แบบทดสอบที่ผ่านเกณฑ์มาตรฐาน ซึ่งผู้ออกแบบบทเรียนจะต้องวิเคราะห์แบบทดสอบที่ผ่านเกณฑ์ เพื่อนำไปพิจารณากับวัตถุประสงค์เชิงพฤติกรรมของบทเรียนที่ตั้งไว้ ถ้าแบบทดสอบมีจำนวนไม่เพียงพอกับการใช้งานหรือไม่ผ่านเกณฑ์ จะต้องปรับปรุงแก้ไขแบบทดสอบ แล้วนำไปทดลองใหม่กับกลุ่มตัวอย่างกลุ่มใหม่ที่มีจำนวนใกล้เคียงกันและดำเนินการซ้ำ ๆ จนกว่าจะได้แบบทดสอบที่มีคุณภาพตรงตามเกณฑ์และมีจำนวนข้อเพียงพอสำหรับแบบทดสอบก่อนบทเรียน แบบทดสอบหลังบทเรียน และแบบฝึกหัดระหว่างบทเรียนของ

บทเรียนคอมพิวเตอร์ จะเห็นได้ว่าขั้นตอนนี้จะใช้เวลามาก โดยทั่วไปแล้วในการทดสอบครั้งแรก ถ้าได้แบบทดสอบเกินกว่า 50% ถือได้ว่าแบบทดสอบที่สร้างขึ้นมานั้นมีคุณภาพในขั้นดี

7. คัดแยกประเภทและนำไปใช้

ขั้นสุดท้ายเป็นการคัดแยกประเภทของแบบทดสอบที่ผ่านเกณฑ์ด้านคุณภาพ เพื่อนำไปใช้ในบทเรียน การคัดแยกประเภทจะพิจารณาที่วัตถุประสงค์เชิงพฤติกรรมของบทเรียนเป็นเกณฑ์ โดยมีข้อตกลงเบื้องต้นว่า วัตถุประสงค์เชิงพฤติกรรมแต่ละข้อควรมีแบบทดสอบอย่างน้อย 2 ข้อ หากได้น้อยกว่านี้จะต้องกลับไปทดลองซ้ำในขั้นที่ 5 ใหม่ โดยเฉพาะบทเรียนคอมพิวเตอร์ที่ใช้แบบทดสอบที่มีการสุ่ม ควรมีแบบทดสอบอย่างน้อย 3 ข้อ ต่อวัตถุประสงค์เชิงพฤติกรรมแต่ละข้อ การสุ่มแบบทดสอบจากธนาคารข้อสอบจึงจะได้ผลดี



ภาพที่ 8-1 ขั้นตอนการหาคุณภาพของแบบทดสอบของบทเรียนคอมพิวเตอร์

■ บทสรุป

คุณภาพของแบบทดสอบที่ใช้ในบทเรียนคอมพิวเตอร์มีความสำคัญยิ่ง ซึ่งนอกจากจะเป็นสิ่งที่ใช้ประเมินผู้เรียนว่าบรรลุผลตามวัตถุประสงค์หรือไม่แล้ว ยังใช้เป็นเป็นตัวบ่งชี้ประสิทธิภาพของบทเรียนอีกด้วย การที่ผู้เรียนทำคะแนนได้สูงจากแบบทดสอบที่มีคุณภาพ ย่อมแสดงถึงความสามารถที่แท้จริงของผู้เรียน แต่ถ้าแบบทดสอบที่ใช้ไม่มีคุณภาพ ผลคะแนนที่ผู้เรียนทำได้จะไม่สามารถบ่งชี้ถึงความสามารถที่แท้จริงได้เลย ดังนั้น หลังจากที่ออกแบบทดสอบแล้ว ก่อนที่จะนำไปใช้งานจะต้องผ่านการหาคุณภาพของแบบทดสอบก่อนทุกครั้ง

องค์ประกอบที่เป็นตัวบ่งชี้คุณภาพของแบบทดสอบ ได้แก่ ความเที่ยงตรง ความเชื่อมั่น อำนาจจำแนก ความยากง่าย และความเป็นปรนัย แบบทดสอบที่ถือว่ามีความเที่ยงตรงในระดับดีสามารถนำไปวัดผลได้ จะต้องมีค่า IOC เกินกว่า .50 ขึ้นไป ส่วนค่าความเชื่อมั่นของแบบทดสอบในเกณฑ์ที่ยอมรับได้ ควรมีค่าเกินกว่า .60 เป็นต้นไป สำหรับค่าอำนาจจำแนก ควรมีค่าสูงเกิน .40 ขึ้นไป และค่าความยากง่ายที่เหมาะสมมีค่าเท่ากับ .50 แต่การออกแบบทดสอบให้มีค่าความยากง่ายเท่ากับ .50 เป็นเรื่องยาก ดังนั้น ความยากง่ายจึงควรมีค่าใกล้เคียง .50 ในทางปฏิบัติจึงกำหนดไว้ที่ .2 – .80

ขั้นตอนการหาคุณภาพของแบบทดสอบสำหรับบทเรียนคอมพิวเตอร์ ได้แก่ วางแผนการดำเนินงาน กำหนดกลุ่มตัวอย่าง ทดลองใช้ขั้นต้น จัดเตรียมและพิมพ์แบบทดสอบ ทดลองใช้จริง วิเคราะห์หาคุณภาพ และคัดแยกประเภทและนำไปใช้

■ แบบฝึกหัดท้ายบท

จงตอบคำถามต่อไปนี้

1. ค่า IOC มีประโยชน์อย่างไรต่อเครื่องมือวัดสำหรับการวิจัย
2. ความเที่ยงตรงเชิงพยากรณ์ มีความหมายอย่างไร
3. แบบทดสอบที่มีค่า P = .50 หมายความว่าอย่างไร
4. ค่าอำนาจจำแนกติดลบ มีความหมายอย่างไร
5. ความเป็นปรนัยของแบบทดสอบ หมายถึงอะไร
6. ความเชื่อมั่น มีความสำคัญอย่างไรต่อคุณภาพของแบบทดสอบ
7. สัมประสิทธิ์สหสัมพันธ์ หมายถึงอะไร
8. ความเป็นปรนัยของแบบทดสอบ หมายถึงอะไร
9. ให้อธิบายขั้นตอนการหาคุณภาพของแบบทดสอบ
10. ให้บอกวิธีการหาความเป็นปรนัยของแบบทดสอบที่ใช้ในบทเรียนคอมพิวเตอร์
11. ถ้าดำเนินการทดลองแบบทดสอบกับกลุ่มตัวอย่างที่ใช้งานจริงจำนวน 3 ครั้งแล้ว แต่ก็ยังได้แบบทดสอบไม่ครบตามจำนวนที่ต้องการ ควรจะดำเนินการอย่างไรต่อไป
12. องค์ประกอบใดที่มีผลต่อคุณภาพของแบบทดสอบมากที่สุดตามความคิดเห็นของท่าน

13. ผู้เชี่ยวชาญสามารถให้ข้อมูลด้านใดเกี่ยวกับคุณภาพของแบบทดสอบ
14. จากข้อมูลในตารางต่อไปนี้ จงวิเคราะห์หาคุณภาพของแบบทดสอบดังนี้
    14.1 ความเชื่อมั่นทั้งฉบับ
    14.2 ความยากง่ายรายข้อ
    14.3 ค่าอำนาจจำแนกรายข้อ
    14.4 สรุปผลการวิเคราะห์ด้วยว่า แบบทดสอบแต่ละข้อมีคุณภาพเป็นอย่างไร

คนที่	ข้อ 1	ข้อ 2	ข้อ 3	ข้อ 4	ข้อ 5	ข้อ 6	ข้อ 7	ข้อ 8	ข้อ 9	ข้อ 10	ข้อ 11	ข้อ 12	ข้อ 13	ข้อ 14
1	8	4	5	6	3	3	4	7	4	8	7	8	5	4
2	7	5	7	4	6	4	7	3	5	3	6	8	6	4
3	8	8	5	3	2	5	6	2	8	6	5	4	8	8
4	3	4	3	7	3	9	6	3	4	5	4	4	9	8
5	6	5	5	7	6	5	5	5	7	4	3	5	2	7
6	4	6	2	4	5	5	7	1	5	6	1	5	2	7
7	2	3	5	5	3	9	3	6	6	3	9	3	3	6
8	3	3	3	5	7	7	7	7	3	5	7	3	6	5
9	4	4	6	4	2	9	4	4	6	4	5	6	6	2
10	4	6	3	6	7	3	3	5	4	3	4	6	3	6
11	5	6	2	2	2	7	5	5	5	1	5	4	3	3
12	6	2	6	7	8	7	2	5	1	2	6	7	4	3
13	5	5	3	7	7	9	1	9	4	9	2	4	5	7
14	5	4	4	6	3	8	2	6	7	8	4	9	4	4
15	7	4	3	8	5	6	5	5	6	6	3	6	1	3
16	8	7	7	3	5	5	1	4	6	5	4	3	2	7

15. จากข้อมูลในตารางต่อไปนี้ จงวิเคราะห์หาคุณภาพของแบบทดสอบดังนี้
    15.1 ความเชื่อมั่นทั้งฉบับ
    15.2 ความยากง่ายรายข้อ
    15.3 ค่าอำนาจจำแนกรายข้อ
    15.4 สรุปผลการวิเคราะห์ด้วยว่า แบบทดสอบแต่ละข้อมีคุณภาพเป็นอย่างไร

คนที่	ข้อ 1	ข้อ 2	ข้อ 3	ข้อ 4	ข้อ 5	ข้อ 6	ข้อ 7	ข้อ 8	ข้อ 9	ข้อ 10	ข้อ 11	ข้อ 12
1	0	1	1	1	1	1	0	1	1	1	1	1
2	1	1	1	1	0	1	0	1	1	1	1	1
3	1	1	1	1	1	1	1	0	1	1	0	1
4	1	0	1	0	1	0	1	0	1	1	1	0
5	1	0	1	0	1	1	1	1	1	0	1	0
6	0	0	0	0	1	1	1	1	1	0	1	0
7	1	1	1	1	1	1	0	1	0	1	0	1
8	1	1	1	1	0	1	0	1	0	0	1	1
9	0	0	0	1	1	1	1	0	1	1	1	1
10	0	0	1	0	1	1	1	0	1	0	1	1
11	1	1	1	0	1	1	0	0	0	1	0	0
12	0	1	0	1	0	1	0	1	1	1	1	1