

บทที่ 4

การหาคุณภาพของเครื่องมือ

การหาคุณภาพของเครื่องมือที่ใช้ในการวิจัยทางการเรียนการสอนวิทยาศาสตร์นั้นนอกจากจะขึ้นอยู่กับประเภทของเครื่องมือแล้ว ยังขึ้นอยู่กับขั้นตอนของการสร้างเครื่องมือ โดยมีการหาคุณภาพของเครื่องมือในขั้นตอนการสร้างหรือก่อนนำไปทดลองใช้ และหลังการทดลองใช้

1. การหาคุณภาพของเครื่องมือในขั้นตอนการสร้าง

การหาคุณภาพของเครื่องมือในขั้นตอนการสร้าง เริ่มตั้งแต่ขั้นตอนตรวจสอบความสอดคล้องของข้อความในเครื่องมือที่สร้างว่ามีความสอดคล้องกับการนิยามตัวแปรที่ต้องการวัดหรือไม่ โดยปกติมักจะให้ผู้ทรงคุณวุฒิ 3-5 คน ตรวจสอบความถูกต้องของความสอดคล้องระหว่างข้อความที่แสดงการวัดตัวกับนิยามของตัวแปรนั้น

การหาคุณภาพของเครื่องมือในขั้นตอนนี้ เรียกว่า การตรวจสอบความตรงเชิงเนื้อหา (Content Validity) ด้วยเทคนิคเผชิญหน้า (Face Validity) ตัวอย่าง

นิยามตัวแปร ความคิดเห็นที่มีต่อพฤติกรรมการสอนของครู หมายถึง ความรู้สึกนึกคิดของนักเรียนต่อการแสดงออกของครูในระหว่างการสอน ซึ่งนักเรียนสามารถสังเกตเห็นและรับรู้ความรู้สึกถึงอารมณ์และความคิดของครู ด้วยปริมาณความถี่ของพฤติกรรมที่แสดงออก

ข้อความ	ระดับของการปฏิบัติ				สำหรับผู้ทรงคุณวุฒิ		
	มาก	ปานกลาง	น้อย	ไม่เคย	-1	0	1
1.ครูผู้สอนปฏิบัติต่อนักเรียนทุกคนอย่างเป็นธรรม							
2.							

- | | | |
|----|---------|----------------------------|
| 1 | หมายถึง | ข้อความสอดคล้องกับนิยาม |
| 0 | หมายถึง | ไม่แน่ใจ |
| -1 | หมายถึง | ข้อความไม่สอดคล้องกับนิยาม |

จากนั้นจึงนำแต่ละข้อมาหาค่า IOC (Item Objective Congruence) ดังนี้

$$\text{IOC ของข้อ } i = \frac{\sum R_i}{N}$$

$\sum R_i$ คือ คะแนนรวมของผู้ทรงคุณวุฒิ

N คือ จำนวนผู้ทรงคุณวุฒิ

โดยคัดข้อที่มีค่า IOC ตั้งแต่ 0.5 ขึ้นไป

2. การหาคุณภาพของเครื่องมือหลังการทดลองใช้

ภายหลังที่ข้อความในเครื่องมือวัดตัวแปรได้ผ่านการตรวจสอบจากผู้ทรงคุณวุฒิแล้ว จึงนำแบบวัดหรือเครื่องมือวัดตัวแปรไปทดลองใช้กับกลุ่มที่มีลักษณะกลุ่มตัวอย่างที่มีจำนวนมากเกินพอ (ประมาณ 100 คน)

การวัดตัวแปรทางด้านการสอนวิทยาศาสตร์ มีทั้งตัวแปรที่วัดได้ในระดับนามบัญญัติ (Nominal Scale) ระดับลำดับ (Ordinal Scale) ระดับช่วง (Interval Scale) และระดับอัตราส่วน (Ratio Scale) ดังนั้นการหาคุณภาพของเครื่องมือที่ใช้วัดตัวแปรในแต่ละระดับจึงไม่เหมือนกัน

2.1 การหาคุณภาพของเครื่องมือที่วัดระดับนามบัญญัติ (Nominal Scale)

ตัวแปรที่อยู่ในระดับของการวัดชั้นนามบัญญัติ ตัวเลขที่อยู่ในเครื่องมือหรือแบบสอบถามเป็นตัวเลขที่ระบุคุณลักษณะของตัวแปร เช่น เพศของนักเรียนชั้นม.1 ระดับชั้นของผู้เรียน อาชีพของบิดามารดา ขนาดของโรงเรียน สังกัดของโรงเรียน ฯลฯ ตัวแปรประเภทนี้เป็นการวัดข้อเท็จจริงมากกว่าความคิดเห็น หรือเชาว์ปัญญา การหาคุณภาพของเครื่องมือประเภทนี้ จะใช้เทคนิคของการวัดซ้ำ โดยทิ้งช่วงเวลาของการวัดทั้ง 2 ครั้งห่างกัน ประมาณ 2 สัปดาห์หรือ 1 เดือน แล้วตรวจสอบความคงเส้นคงวาในการตอบคำถามแต่ละข้อ ว่ามีการตอบคำถามทั้ง 2 ครั้งตรงกันหรือไม่ หากไม่ตรงกันก็แปลว่าข้อคำถามของผู้วิจัยเขียนไว้มั่วชัดเจน ทำให้การตีความหมายไม่ตรงกันระหว่างการตอบครั้งที่ 1 กับคำตอบครั้งที่ 2

$$\text{ค่าความเชื่อมั่นของการตอบคำถามข้อ } i = \frac{\text{จำนวนคนที่ตอบข้อ } i \text{ ครั้งเหมือนกัน}}{\text{จำนวนคนทั้งหมด}}$$

ตัวอย่างคำถามในแบบสอบถามถามว่ารายได้ของผู้ปกครอง

- ☐ ต่ำกว่า 6,000 ☐ ระหว่าง 6,000 – 10,000
☐ ระหว่าง 10,000 – 20,000 ☐ มากกว่า 20,000

ด้วยความคลุมเครือของข้อคำถาม ผู้ตอบอาจตีความครั้งแรกเป็นรายได้ของพ่อ และเมื่อให้คำตอบครั้งที่ 2 ก็อาจตีความใหม่เป็นรายได้ของผู้ปกครองรวมกัน จึงทำให้การตอบครั้งที่ 1 และครั้งที่ 2 ไม่ตรงกัน ดังนั้นถ้าถามผู้เรียน 100 คน ตอบตรงกันทั้ง 2 ครั้งเป็น 60 คน

$$\text{ค่าความเชื่อมั่นของการตอบคำถาม} = \frac{60}{100} = .6$$

ซึ่งควรจะเป็น 1.00 หรือ 0.99 จึงจะเชื่อถือได้

2.2 การหาคุณภาพของเครื่องมือที่วัดระดับนามลำดับ (Ordinal Scale)

การที่จะตรวจสอบว่าผู้ตอบมีการ Rank ข้อมูลอย่างเชื่อมั่น ก็ตรงให้ผู้ตอบทำการ Rank หรือจัดอันดับ 2 ครั้ง โดยการวัดแต่ละครั้งให้ห่างกัน 2 สัปดาห์หรือ 1 เดือน

ตัวอย่าง ให้นักเรียนจัดลำดับความชอบในการอ่านหนังสือประเภทต่างๆ ดังนี้ โดยลำดับที่ 1 คือชอบมากที่สุด

ประเภทหนังสือ	ลำดับความชอบ
การ์ตูน	
หนังสือเรียน	
สารคดี	
นิยาย	
ข่าววิชาการ	
ข่าวกีฬา	
ข่าวบันเทิง	

การหาคุณภาพของเครื่องมือก็ควรนำการลำดับความชอบทั้ง 2 ครั้ง มาหาค่าสหสัมพันธ์แบบ Rank หรือ Spearman Rank Correlation มีสูตรดังนี้

$$r_{rank} = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

เมื่อ d คือความต่างของลำดับของ 2 ครั้ง

ค่าฐานนิยมของลำดับครั้งที่ 1		ค่าฐานนิยมของลำดับครั้งที่ 2	
การ์ตูน	1	การ์ตูน	1
หนังสือเรียน	6	หนังสือเรียน	5
สารคดี	7	สารคดี	5
นิยาย	1	นิยาย	2
ข่าววิชาการ	2	ข่าววิชาการ	2
ข่าวกีฬา	3	ข่าวกีฬา	3
ข่าวบันเทิง	4	ข่าวบันเทิง	4

ฐานนิยมของลำดับ = ลำดับที่ที่นักเรียนให้ไว้มากที่สุด

สมมติว่ามีกลุ่มตัวอย่าง 100 คน

	ลำดับที่ ครั้งที่ 1	ลำดับที่ ครั้งที่ 2	d	d ²
การ์ตูน	1.5	1	.5	.25
นิยาย	1.5	2.5	1	1
ข่าววิชาการ	3	2.5	.5	.25
ข่าวกีฬา	4	4	0	0
ข่าวบันเทิง	5	5	0	0
หนังสือเรียน	6	5.5	1.5	2.25
สารคดี	7	5.5	2.5	6.25

แทนค่า

$$\sum d^2 = 10$$

$$r_{rank} = 1 - \frac{6(10)}{7(49 - 1)} = 1 - \frac{60}{330} = .821$$

สรุปได้ว่าการจัดลำดับของความชอบในการอ่านหนังสือประเภทต่างๆ ของนักเรียนทั้ง 2 ครั้ง มีความสัมพันธ์กัน .821 แสดงว่า คำตอบมีความเชื่อมั่น .821 ซึ่งถือว่าอยู่ในระดับดี

2.3 การหาคุณภาพของเครื่องมือวัดระดับช่วงและระดับอัตราส่วน (Interval and Ratio Scale)

เครื่องมือวัดตัวแปรที่อยู่ในรูปของมาตราส่วนประมาณค่า (Rating Scale) ตามแบบของ Likert ที่ประมาณคะแนนเป็น 5 ช่วงนั้น โดยปกติแล้วจะเป็นการวัดตัวแปรแค่ในระดับลำดับเท่านั้น เพราะผู้ตอบว่าเห็นด้วยอย่างยิ่งซึ่งได้ 5 คะแนน ก็ไม่ได้หมายความว่ามีความต่างจากการเห็นด้วย 1 คะแนน เท่ากันกับค่าความต่างไม่แน่ใจกับไม่เห็นด้วยและเท่ากันกับระหว่างความแตกต่างระหว่างไม่เห็นด้วยกับไม่เห็นด้วยอย่างยิ่ง การหาคุณภาพของเครื่องมือประเภทมาตราส่วนประมาณค่าจึงควรใช้เทคนิคของการตอบซ้ำ (Test-Retest) ดังที่กล่าวมาแล้ว

แต่เนื่องจาก Likert (1932 : อ้างถึงใน Edward,1957) ได้ ทำการพิสูจน์และพบว่า มาตราประมาณค่า (Rating Scale) ของเขาที่กำหนดเป็นคะแนนน้ำหนัก เมื่อคิดเป็นคะแนนน้ำหนักรวมกับคิดเป็นสัดส่วนของน้ำหนักแล้วมีความสัมพันธ์กันถึง .99 ดังนั้น มาตราประมาณค่าที่เป็นน้ำหนักจึงสามารถนำแต่ละข้อมารวมกันได้ เสมือนว่าค่าน้ำหนักเหล่านี้เป็นมาตราช่วง (Interval Scale) Bird (1940,อ้างถึงใน Edward,1957) ได้เรียกวิธีการสร้างมาตราของ Likert ว่าเป็น “Method of summated ratinge” (วิธีการบวกค่าน้ำหนัก) จึงเป็นที่มาของการนำเอาน้ำหนักจาก Rating Scale มาเป็นมาตราอันตรภาค

2.3.1 การวิเคราะห์คุณภาพรายข้อ

การวิเคราะห์รายข้อ : ในการวิเคราะห์รายข้อนั้นสามารถวิเคราะห์ได้ทั้งในแง่ของคุณภาพซึ่งอยู่ในเทอมของเนื้อหา (Content) และรูปแบบ (Form) และวิเคราะห์ในแง่ของปริมาณ ซึ่งจะอยู่ในเทอมของคุณสมบัติทางสถิติ การวิเคราะห์ทางด้านคุณภาพนั้นรวมถึงการพิจารณาความตรงในเนื้อหา (Content Validity) และการประเมินรายข้อในเทอมของวิธีการเขียนข้อกระทงที่มีประสิทธิภาพ ไม่ว่าจะเป็นความตรงหรือค่าความเที่ยงของแบบสอบ ย่อมขึ้นอยู่กับคุณสมบัติรายข้อ การปรับปรุงค่าความเที่ยงหรือความตรงที่ดีก็ต้องไปปรับที่รายข้อ

การวิเคราะห์ข้อกระทงต้องทำให้แบบสอบสั้นลงเท่าที่เป็นไปได้แต่ในขณะเดียวกัน ความเที่ยงและความตรงต้องเพิ่มขึ้น ตามที่เราทราบอยู่แล้วว่าแบบสอบยิ่งมีความยาว ค่าความเที่ยงก็จะยิ่งเพิ่มมากขึ้น แต่ถ้าเรามีการวิเคราะห์รายข้อของแบบสอบ แล้วทำให้แบบสอบสั้นลงโดยการตัดข้อกระทงไม่ดีทิ้งไป แบบสอบที่ไม่สั้นลงนี้อาจจะมีค่าความตรงและความเที่ยงสูงกว่าฉบับเดิมที่ยาวกว่าก็ได้ (Anastasi 1982 : 192)

ก) การวิเคราะห์รายข้อสำหรับแบบสอบในห้องเรียน (Item Analysis For Classroom Test)

ในการวิเคราะห์รายข้อนั้น จำเป็นต้องอาศัยความสัมฤทธิ์ผลของการตอบคำถามของนักเรียนในชั้น คำตอบจากนักเรียนเพียงคนเดียวไม่เพียงพอที่จะทำการตัดสินใจเชิงวินิจฉัยได้ คำตอบจากนักเรียนเป็นกลุ่มเท่านั้นที่จะทำให้ข้อกระทงมีความหมายและมีความเที่ยง (Hopkin & Stanley 1981 : 269)

Coffman (1969) กล่าวว่า เป็นธรรมดาที่เราอาจจะพบว่ามีบางกระทงที่ปรากฏเห็นชัดว่า น่าพอใจ แม้ผู้เชี่ยวชาญเองก็ลงความเห็นว่าเป็นเช่นนั้น แต่ก็ยังพบจากการวิเคราะห์รายข้อว่ามีความกำกวมและทำให้นักเรียนเข้าใจผิดได้ จุดมุ่งหมายโดยทั่วไปของการวิเคราะห์รายข้อนั้นคือ การหาค่าความยากและค่าอำนาจจำแนกของแต่ละข้อ

การวิเคราะห์รายข้อสำหรับแบบสอบที่ใช้ในห้องเรียนนั้นได้มีการวิเคราะห์หาค่าความยากรายข้อ (Item Difficulty) โดยใช้ทฤษฎีตั้งเดิมและหาค่าอำนาจจำแนกรายข้อ (Item Discrimination) ซึ่งได้แบ่งวิธีการออกเป็นการใช้กลุ่มสุดโต่ง (Use of extreme groups) การวิเคราะห์กับกลุ่มเล็ก (Simple Analysis with small groups) และการหาดัชนีอำนาจจำแนกด้วยค่าสัมประสิทธิ์สหสัมพันธ์

1.1 ค่าความยากรายข้อ (Item Difficulty)

ในการหาค่าความยากของข้อกระทงส่วนมากจะกำหนดในเทอมของเปอร์เซ็นต์ (หรือสัดส่วน) ของคนที่ตอบข้อกระทงนั้นถูก ข้อที่ง่ายก็จะมีเปอร์เซ็นต์การตอบถูกสูงกว่าข้อที่ยาก ในกระบวนการสร้างแบบสอบ หลักที่ใช้ในการคัดเลือกข้อกระทงอย่างหนึ่งก็คือ จะต้องเลือกเอาข้อกระทงที่ไม่ยากเกินไปและไม่ง่ายเกินไป (ค่าความยากเข้าใกล้ 1 หรือ 0) เพราะข้อเหล่านี้จะไม่ให้สาระสนเทศเกี่ยวกับความแตกต่างของผู้เข้าสอบเลย

เนื่องจากว่าข้อเหล่านี้ไม่มีผลกระทบต่อการเล่นแปลงของคะแนนสอบและไม่มีผลต่อค่าความเที่ยงและความตรงของแบบสอบด้วย ซึ่งตรงข้ามกับข้อที่มีค่าความยากยิ่งเข้าใกล้ .5 ข้อนั้นจะสามารถจำแนกความแตกต่างได้มากขึ้น เช่น คน 100 คน มี 50 คนที่ตอบข้อสอบข้อนี้ได้ และอีก 50 คนที่ตอบไม่ได้ ($P = .5$) ข้อนี้สามารถจะบอกเราถึงความแตกต่างระหว่างคนตอบได้กับคนที่ตอบไม่ได้ ดังนั้นเราจะมี 50×50 คู่ของการเปรียบเทียบหรือหน่วยของสารสนเทศที่บอกความแตกต่าง (Bit of Differential Information) อีกข้อหนึ่งซึ่งมีคนตอบได้ 70% จะให้สารสนเทศที่บอกความแตกต่างเป็น 70×30 หรือ 2,100 หน่วย (Bit) และอีกข้อหนึ่งมีคนตอบได้ 90% ก็ให้สารสนเทศเพียง 90×10 หรือ 900 หน่วย และข้อที่คนตอบได้ 100% ก็ให้สารสนเทศ 100×0 หรือ 0 หน่วย จะเห็นว่าค่าที่บอกความแตกต่างได้สูงสุดคือข้อที่มีค่าความยาก 50% หรือ .5 อย่างไรก็ตามถ้าเราพิจารณาถึงความจริงที่ว่า ถ้าข้อกระทงของแบบสอบมีค่าสหสัมพันธ์ระหว่างกัน (Intercorrelation) สูง แบบสอบนั้นก็จะเป็นเอกพันธ์ (Homogeneous) สูงด้วย ในกรณีของกลุ่มชุดใดก็จะพบว่าถ้าข้อกระทงเหล่านั้นมีค่าสหสัมพันธ์ระหว่างกันอย่างสมบูรณ์ (Perfect Intercorrelation) แล้วข้อกระทงทั้งหมดมีค่าความยาก .5 จะได้ 50 คนเดิมจาก 100 คนที่ตอบข้อกระทงแต่ละข้อได้ ดังนั้นครึ่งหนึ่งของผู้เข้าสอบจะได้คะแนนเต็ม ในขณะที่อีกครึ่งหนึ่งได้คะแนน 0 ทั้งนี้ก็เพราะว่า การหาค่าสหสัมพันธ์ระหว่างข้อเป็นวิธีที่ดีที่สุดในการเลือกข้อกระทงที่มีค่าระดับความยากกระจาย นอกเสียจากค่าเฉลี่ยของความยากคือ .5 ยิ่งไปกว่านั้นค่าสหสัมพันธ์ระหว่างข้อสูงขึ้น การกระจายของค่าความยากก็ควรกว้างขึ้น

ตัวเลือกของข้อกระทงที่มีค่าความยากพอเหมาะ ยังเกี่ยวข้องกับโอกาสการเดาในข้อสอบเลือกตอบอีกด้วย ซึ่งต้องยอมรับความจริงที่ว่า สัดส่วนของผู้เข้าสอบจำนวนหนึ่งจะเลือกตัวเลือก ตัวเลือกที่ถูกต้องจากการเดา ซึ่งทำให้เราต้องกำหนดสัดส่วนของการตอบถูกให้สูงกว่าที่ควรจะเป็น Lord (1952) กำหนดว่าสำหรับข้อกระทงที่มี 5 ตัวเลือก สัดส่วนของการตอบถูกโดยเฉลี่ยแล้วควรประมาณ .69 (Anastasi 1982 : 193-194)

ค่าความยากของข้อคือ ผลบวกของจำนวนคนตอบข้อนั้นได้จากคนทั้ง 3 กลุ่ม ($U + M + L$)

จากผลการวิเคราะห์ในตารางข้อ 2 และข้อ 7 ควรจะตัดทิ้งเพราะว่าข้อ 2 ง่ายเกินไป มีคนตอบได้ถึง 56 คน จาก 60 คน และข้อ 7 ยากเกินไป มีคนตอบได้เพียง 5 คน

ส่วนข้อ 4 และข้อ 5 มีค่าความยากเป็นที่น่าพอใจแล้ว แต่ค่าจำแนกกลับติดลบและเป็น 0 ตามลำดับ เรายังต้องพิจารณาถึงข้อที่มีค่าจำแนกพอใช้ได้ในกลุ่มเล็กอาจจะเป็นค่าที่ใช้ไม่ได้ในกลุ่มใหญ่ก็ได้ ดังนั้นเราจึงควรที่จะทำการวิเคราะห์เพิ่มเติมโดยพิจารณาที่การตอบของกลุ่มสูงและกลุ่มต่ำบนตัวลงแต่ละตัวด้วยเพื่อประกอบการพิจารณา ให้ดูที่ตารางที่ 3 (Anastasi 1982 : 206) ดังนี้

ตาราง 3

Response Analysis of Individual Items

Item	Group	ผลตอบของแต่ละตัวเลือก				
		1	2	3	4	5
2	Upper	0	0	0	20	0
	Lower	2	0	1	16	1
4	Upper	0	10	9	0	1
	Lower	2	16	2	0	0
5	Upper	2	3	2	11	2
	Lower	1	3	3	11	2
7	Upper	5	3	5	4	3
	Lower	0	5	8	3	4
.						
.						
.						

*ตัวเอียงคือตัวเลือกที่ถูก

ข้อ 2 มีคนในกลุ่มต่ำตอบผิดเพียง 4 คน และคนในกลุ่มสูงไม่มีใครตอบผิดเลยซึ่งครูผู้สอนจะต้องมาทบทวนดูว่า (1) ข้อสอบข้อนี้ยากเกินไปใช่หรือไม่ (2) หรือมีอะไรบางอย่างที่ทำให้ไปชี้แนะคำตอบที่ถูก (3) หรือว่ามันเป็นข้อที่ดี ซึ่งแสดงให้เห็นว่าการสอน

ประสิทธิภาพ และเด็กมีความจำที่ดี เมื่อเป็นเช่นนี้ถ้าเป็นกรณีที่ 1 ก็ควรตัดทิ้ง ถ้าเป็นกรณีที่ 2 ก็ควรปรับปรุงและถ้าเป็นกรณีที่ 3 ก็ให้คงไว้

ส่วนข้อที่ 4 ตัวเลือกที่ 3 มีบางสิ่งบางอย่างผิดปกติที่ทำให้คนเก่งหันไปเลือกวิธีการที่ดีที่สุดก็คือให้นักเรียนอธิบายว่าทำไมจึงเลือกตัวเลือกนี้ ข้อ 5 ดูเหมือนว่าความผิดอยู่ที่ตัวคำถามหรือตัวเลือกที่ถูก เพราะนักเรียนที่ตอบผิดกระจายเป็นแบบเดียวกันบน 4 ตัวลง สำหรับข้อ 7 เป็นข้อที่มีความยากผิดปกติ ซึ่งมีคนกลุ่มสูงตอบถูกเพียง 5 คน และคนกลุ่มต่ำตอบไม่ได้เลย ตัวเลือกที่ 3 มีสิ่งดึงดูดใจให้คนกลุ่มต่ำตอบไม่ได้เลย ตัวเลือกที่ 3 มีสิ่งดึงดูดใจให้คนกลุ่มต่ำตอบในขณะที่ตัวเลือกที่ 1 ซึ่งเป็นตัวที่ถูกคนกลุ่มต่ำกลับตอบไม่ได้เลย ถ้าหากว่าคนกลุ่มต่ำให้เหตุผลว่าตัวเลือกที่ 1 นั้นน่าสงสัยมากจนดูเหมือนว่ามันผิด ด้วยการพิจารณาลักษณะตัวลงทั้ง 2 นี้ ข้อสอบข้อ 7 นี้ก็อาจจะดีซึ่งสามารถชี้ให้เห็นว่า มีเพียงไม่กี่คนในห้องที่เรียนรู้ได้ถูกต้อง

1.2 ค่าจำแนกรายข้อ (Item Discrimination)

ค่าจำแนกรายข้อจะบอกถึงระดับของความแตกต่างในการตอบข้อกระทงข้อหนึ่งๆ ได้ถูกต้องในบรรดาผู้เข้าสอบ ตัวเลขที่บ่งชี้ถึงระดับของความแตกต่างนี้เรียกว่าค่าดัชนีจำแนก (The Index of Discrimination) ค่าดัชนีจำแนกคิดเป็นเปอร์เซ็นต์ ซึ่งได้มาจากผลต่างของเปอร์เซ็นต์ของกลุ่มสูงที่ตอบข้อนั้นถูกกับเปอร์เซ็นต์ของคนกลุ่มต่ำที่ตอบข้อนั้นถูก ดัชนีจำแนก (D) นี้จะมีค่าอยู่ระหว่าง +100 ถึง -100 คุณสมบัติที่น่าสนใจของ D มีหลายข้อ ซึ่ง Ebel และ Findley (1956) ได้แสดงให้เห็น (Anastasi 1982 : 207-208) ดังนี้

1. ค่า D เป็นสัดส่วนโดยตรงของความแตกต่างระหว่างจำนวนของการจำแนกตัวเลือกที่ถูก กับจำแนกตัวลงของข้อๆหนึ่ง โดยที่ค่าจำแนกตัวถูก (Correct) ขึ้นอยู่กับจำนวนคนตอบถูกในกลุ่มสูงต่อด้วยจำนวนคนตอบผิดในกลุ่มต่ำ

2. ค่าเฉลี่ยของดัชนี D มีความสัมพันธ์กับสัมประสิทธิ์ความเที่ยงของแบบสอบ โดยที่ค่าเฉลี่ยของ D สูงขึ้น ค่าความเที่ยงก็จะสูงขึ้นด้วย

3. ค่าดัชนี D นี้เกี่ยวข้องกับค่าดัชนีอื่นๆอีกหลายตัว ค่า D ไม่อิสระจากค่าความยาก แต่มีความลำเอียงเกิดขึ้นในระหว่างระดับของความยาก ให้พิจารณาจากตาราง 1 ดังนี้

Relation of Maximum Value of D to Difficulty

Percentage Passing Item	Maximum Value of D
100	0
90	20
70	60
50	100
30	60
10	20
0	0

จากตารางแสดงให้เห็นว่า ค่าสูงสุดที่เป็นไปได้ของ D จะให้ค่าเปอร์เซ็นต์ของการตอบข้อนั้นถูกแตกต่างกันหลายค่า เช่น ถ้าค่า D สูงสุดที่เป็นไปได้เป็น 0 จะมีเปอร์เซ็นต์ของการตอบข้อนั้นถูกเป็น 100% หรือ 0% ก็ได้ และถ้าค่า D สูงสุดที่เป็นไปได้เป็น 20 จะมีเปอร์เซ็นต์ของการตอบข้อนั้นถูกเป็น 10% หรือ 90% ก็ได้

วิธีการคิดค่าอำนาจจำแนกหรือค่าจำแนกรายข้อนี้ วิธีการคิดหลายวิธี ทั้งนี้ขึ้นอยู่กับลักษณะของกลุ่มที่ใช้ข้อสอบ และประเภทของข้อสอบ ให้หัวข้อนี้ได้เสนอวิธีการวิเคราะห์ค่าอำนาจจำแนกรายข้อ โดยการใช้กับกลุ่มสุดโต่ง การวิเคราะห์กับกลุ่มเล็กและการหาดัชนีจำแนกด้วยค่าสัมประสิทธิ์สหสัมพันธ์

1.2.1 การใช้กลุ่มสุดโต่ง (Use of extreme groups)

ในทางปฏิบัติการวิเคราะห์รายข้อก็คือการเปรียบเทียบสัดส่วนของกลุ่มคนที่ผ่านมาข้อหนึ่งๆ ซึ่งพิจารณาจากเกรดวิชา การให้น้ำหนักของงาน หรือคะแนนรวมของแบบสอบ กลุ่มเกณฑ์สูงและกลุ่มเกณฑ์ต่ำ เหล่านี้จะถูกเลือกขึ้นมาจากกลุ่มปลายสุดโต่งของการกระจายจะเห็นได้ชัดว่ากลุ่มที่ยังมีความสุดโต่งกันมากๆ เช่น กลุ่มสูง 10% กลุ่มต่ำ 10% จะไปลดความเที่ยง (Reliability) ของผลลัพธ์ เพราะเนื่องจากการใช้กลุ่มคนจำนวนน้อยในการกระจายของโค้งปกติ จุดที่เหมาะสมที่สุดซึ่งทำให้เงื่อนไขทั้ง 2 สมดุลย์ คือ กลุ่มสูง 27% กลุ่มต่ำ 27% (Kelley 1939) เมื่อการกระจายแบบราบกว่าโค้งปกติ เปอร์เซ็นต์ที่เหมาะสมค่อนข้างจะมากกว่า 27 คือ เข้าใกล้ 33 (Cureton 1957)

ถ้าเป็นกลุ่มเล็กๆเช่น ในห้องเรียนธรรมดา การคลาดเคลื่อนแบบสุ่มของค่าสถิติรายข้อจะใหญ่มาก ซึ่งเราก็จะได้ค่าที่หยابๆ ดังนั้นภายใต้เงื่อนไขเหล่านี้ เราไม่จำเป็นต้องสนใจมากเกี่ยวกับเปอร์เซ็นต์ที่แน่นอนของกลุ่มทั้งสอง ตัวเลขสะดวกใดๆจะเป็น 25% หรือ 33% ก็พอใจแล้ว

ถ้ากลุ่มตัวอย่างมีขนาดใหญ่และมีการกระจายแบบโค้งปกติ ต้องการจะพัฒนาแบบสอบถามมาตรฐานเรามักจะคุ้นเคยกับการใช้กลุ่มสูง 27% และกลุ่มต่ำ 27% ของการกระจาย (Criterion Dist") ตารางทั้งหลายที่สร้างขึ้นเพื่อสะดวกในการคำนวณค่าดัชนีจำแนกรายข้อนั้นมีข้อตกลงว่าจะต้องเป็นไปตาม "กฎ 27%"

ไม่ว่าจะใช้เทคนิค 27% หรือ 33% หรือ 25% การหาค่าจำแนกรายข้อก็คือ การเปรียบเทียบสัดส่วนของการทำข้อสอบข้อนั้นถูกของกลุ่มสุดโต่งทั้งสอง ค่าอำนาจจำแนกรายข้อที่ $I =$ สัดส่วนของคนกลุ่มสูงทำข้อ i ถูก - สัดส่วนของคนกลุ่มต่ำทำข้อที่ i ถูก

โดยที่ค่าของผลต่างที่อยู่ระหว่าง -1 ถึง 1 ค่าติดลบคือค่าที่แสดงว่าข้อนั้นคนกลุ่มต่ำ ทำได้ดีกว่าคนกลุ่มสูง การจำแนกรายข้อนั้นจึงไม่ดี ถ้าผลต่างยิ่งมากและมีค่าเป็นบวกก็แสดงว่าข้อนั้นมีความจำแนกได้มาก กล่าวคือโอกาสที่คนกลุ่มสูงตอบข้อนั้นได้ถูกมากกว่าโอกาสที่คนกลุ่มต่ำตอบข้อนั้นได้ถูก โดยปกติถือค่าอำนาจจำแนก .2 ขึ้นไป ก็น่าจะพอใจแล้ว

1.2.2 การวิเคราะห์กับกลุ่มเล็ก (Simple Analysis with Small Groups)

บ่อยครั้งที่เราพบว่าการวิเคราะห์รายข้อของเราต้องกระทำกับกลุ่มเล็กๆ เช่น กลุ่มนักเรียนในห้องเรียน ซึ่งก่อนอื่นเราก็ต้องพิจารณาถึงวิธีการเฉพาะที่เหมาะสมกับสถานการณ์เช่นนี้ สมมุติว่าในชั้นหนึ่งมีนักเรียน 60 คน เราเลือกคนกลุ่มสูงสุดมา 20 คน (33%) และอีก 20 คน (33%) เป็นกลุ่มต่ำสุด และที่เหลือเป็นกลุ่มกลางๆให้คนในกลุ่มทั้ง 3 ใช้สัญลักษณ์เป็น U L และ M ตามลำดับ

ตัวอย่างของการวิเคราะห์รายข้อ ให้ดูในตาราง 2 (Anastasi 1982 : 205) เพื่อประกอบอธิบายดังนี้

ตาราง 2

Simple Item Analysis Procedure : Number of Persons Giving Correct Response in Each Criterion Group

Item	U (20)	M (20)	L (20)	Difficulty (U+M+L)	Discrimination (U-L)
1	15	9	7	31	8
2	20	20	16	56	4
3	19	18	9	46	10
4	10	11	16	37	-6
5	11	13	11	35	0
6	16	14	9	39	7
7	5	0	0	5	5
.					
.					
75					

*Item chosen for discussion

ค่าดัชนีชี้แจงของข้ออย่างหายๆ ก็คือผลต่างของจำนวนคนกลุ่มสูงตอบถูกกับคนกลุ่มต่ำถูก (U - L)

1.2.3 การหาดัชนีชี้แจงด้วยค่าสัมประสิทธิ์สหสัมพันธ์ ϕ (Phicoefficient)

ในบรรดาดัชนีชี้แจงรายข้อทั้งหลายที่รายงานด้วยค่าสหสัมพันธ์ระหว่างข้อ กระทบกับเกณฑ์ในรูปของสัมประสิทธิ์สหสัมพันธ์ตัว ϕ เป็นค่าหนึ่งที่คำนวณได้จากตาราง 2x2 ดังตาราง 4

ตาราง 4

	กลุ่มสูง	กลุ่มต่ำ
ตอบถูก	n_{11}	n_{12}
ตอบผิด	n_{21}	n_{22}

ค่า ϕ ขึ้นอยู่กับสัดส่วนของกลุ่มคนที่ตอบถูกและตอบผิดในข้อหนึ่งของกลุ่มเกณฑ์สูงและกลุ่มเกณฑ์ต่ำ เช่นเดียวกับค่าสหสัมพันธ์ทั่วไป ค่า ϕ มีค่าระหว่าง -1 ถึง 1 และเช่นเดียวกับค่า D คือ ดัชนี ϕ ลำเอียงต่อค่าความยากปานกลาง นั่นคือ ϕ มีค่าสูงสุดเมื่อกลุ่มเกณฑ์ทั้ง 2 เข้าใกล้ 50-50 สูตรในการหาค่าคือ (จากตาราง 4)

$$\phi = \frac{(n_{11}n_{12} - n_{21}n_{22})}{\sqrt{(n_{11}n_{12})(n_{21}n_{22})(n_{11}n_{21})(n_{12}n_{22})}}$$

การทดสอบนัยสำคัญของค่า ϕ สามารถคำนวณได้ เนื่องจากว่าค่าสหสัมพันธ์ ϕ นี้ สัมพันธ์กับทั้งค่าอัตราส่วน chi square กับอัตราส่วนโค้งปกติ ในการประยุกต์ใช้นั้น เราสามารถกำหนดค่าต่ำสุดของ ϕ ที่มีระดับนัยสำคัญทางสถิติที่ .05 หรือ .01 โดยการ ใช้สูตร (Anastasi 1982, 209)

$$\phi .05 = \frac{1.96}{\sqrt{N}}$$

$$\phi .01 = \frac{2.58}{\sqrt{N}}$$

N แทนจำนวนกรณีทั้งหมดทั้งในกลุ่มเกณฑ์สูงและต่ำรวมกัน ดังนั้นถ้าจำนวนกรณีของกลุ่มสูงมี 50 คน กลุ่มต่ำ 50 คน ค่า N จะเท่ากับ 100 ค่าของ ϕ ต่ำสุดที่มีนัยสำคัญทางสถิติระดับ .05 จะเท่ากับ $1.96/\sqrt{100} = .196$ ก็จะเป็นข้อที่มีค่าจำแนกใช้ได้

1.2.3 สัมประสิทธิ์สหสัมพันธ์ Biserial และ Point Biserial

การวิเคราะห์รายข้อที่มีจุดมุ่งหมายเพื่อการจำแนกรายข้อนั้น สัมประสิทธิ์สหสัมพันธ์ Biserial (r_{bis}) ต่างจากค่าสหสัมพันธ์ ϕ ตรงที่ว่า (1) ค่า r_{bis} มีข้อตกลงเบื้องต้นว่า คุณลักษณะ (trait) ซึ่งมีค่าของคำตอบเป็น 2 แบบ (Dichotomons) นั้น มีความต่อเนื่อง (Continuous) และมีการกระจายเป็นปกติ (Normal distribution) (2) ยอมให้มีการวัดความสัมพันธ์ระหว่างกระทงกับเกณฑ์ ซึ่งอิสระจากค่าความยากของกระทง

ในการคำนวณค่า r_{bis} โดยตรงจากข้อมูลนั้น เราจำเป็นต้องมีคะแนน เกณฑ์เฉลี่ยของคนที่ผ่านมาและไม่ผ่านข้อนั้น รวมทั้งสัดส่วนของกรณีที่ผ่านมาและไม่ผ่าน ข้อนั้นจากกลุ่ม

ตัวอย่างทั้งหมด และค่าส่วนเบี่ยงเบนมาตรฐานของคะแนนเกณฑ์สูตรสำหรับค่าสัมประสิทธิ์สหสัมพันธ์ Biserial เป็นดังนี้

$$r_{bis} = \frac{(\overline{X_i} - \overline{X})P_i}{s_x h} \quad (\text{กรณีข้อกระทงเทียบกับคะแนนรวมของแบบสอบนั้น})$$

หรือ

$$r_{bis} = \frac{(\overline{X_i} - \overline{Y})P_i}{s_y h} \quad (\text{กรณีข้อกระทงเทียบกับเกณฑ์มาตรฐานภายนอก})$$

เมื่อ $\overline{X_i}$ คือค่าเฉลี่ยของคะแนนรวมของคนที่ตอบข้อ i ถูก

$\overline{X}$ และ s_x เป็นค่าเฉลี่ยของคะแนนรวมและส่วนเบี่ยงเบนมาตรฐาน และคะแนนรวม ตามลำดับ และผู้เข้าสอบทั้งหมด

P_i คือ โอกาสของคนจะตอบข้อ i ถูก (ค่าความยากของข้อกระทง i)

Q_i คือ โอกาสของคนจะตอบข้อ i ผิด

h คือ ค่าความสูงของโค้งปกติ ณ จุดที่จำแนกตัวแปร

ส่วน $\overline{Y}$ และ s_y เป็นค่าเฉลี่ยของคะแนนรวมของเกณฑ์ และส่วนเบี่ยงเบน มาตรฐานของคะแนนรวมเกณฑ์ ตามลำดับ

สำหรับสัมประสิทธิ์สหสัมพันธ์ Point Biserial มีข้อตกลงเบื้องต้นต่างจาก สหสัมพันธ์แบบ Biserial ตรงที่ว่า ค่าของคะแนนแต่ละข้อมีค่าเป็น 0 กับ 1 นั้น มีค่าเป็น 2 ลักษณะแต่คือ มี 0 กับ 1 เท่านั้น ส่วนค่าระหว่างนี้ไม่มี สูตรการคำนวณจึงออกมาเป็น ดังนี้ ซึ่งต่างจากค่าสหสัมพันธ์ Biserial เล็กน้อย คือ

$$r_{pbis} = \frac{(\overline{X_i} - \overline{X})}{s_x} \sqrt{\frac{P_i}{Q_i}} \quad (\text{กรณีข้อกระทงเทียบกับคะแนนรวมของแบบสอบนั้น})$$

$$r_{pbis} = \frac{(\overline{X_i} - \overline{Y})}{s_y} \sqrt{\frac{P_i}{Q_i}} \quad (\text{กรณีข้อกระทงเทียบกับเกณฑ์มาตรฐานภายนอก})$$

ความหมายของสัญลักษณ์ต่าง ๆ ในสูตร เป็นเช่นเดียวกับในสูตรของสหสัมพันธ์ Biserial

ในการตกลงว่าข้อสอบแบบเลือกตอบที่ให้คะแนนเป็น 0 กับ 1 นั้น เป็นแบบต่อเนื่องตั้งแต่ 0 ถึง 0 หรือเป็นแบบไม่ต่อเนื่องมี 2 ค่า คือ 0 และ 1 เท่านั้น นั้นยังหาข้อยุติแบบแน่นอนไม่ได้ ดังนั้น ในหนังสือบางเล่ม (เช่น Allen & Yen) จะใช้สูตรสัมประสิทธิ์สหสัมพันธ์ Point Biserial ในการค่าหาอำนาจจำแนก แต่ในหนังสือบางเล่ม (เช่น Anastasi) จะใช้สูตรสัมประสิทธิ์สหสัมพันธ์ Biserial แทน อย่างไรก็ตามค่า r_{pbis} จะให้ค่าที่น้อยกว่าค่า r_{bis} ดังนั้นในการเลือกใช้ค่า r_{pbis} จึงเป็นเครื่องรับประกันได้ว่า ค่าอำนาจจำแนกที่ให้จะไม่ต่ำไปกว่าค่าที่หาได้นั้น

เนื่องจากยังไม่มีวิธีการคำนวณหาค่าระดับนัยสำคัญที่แน่นอน เพื่อทดสอบค่า r_{bis} ที่คำนวณได้ จึงได้มีการใช้ค่าความคลาดเคลื่อนมาตรฐานของมันมาประมาณดูว่า มันจะใหญ่กว่าค่า r_{bis} ที่คำนวณจากข้อมูลทั้งหมดหรือไม่ นั่นคือ ค่า r_{bis} ที่ประมาณได้จากในตาราง Fan ซึ่งจะแปรผัน (Fluctuates) จากตัวอย่างหนึ่งไปอีกตัวอย่างหนึ่งได้ง่ายกว่าการแปรผันของค่า r_{bis} ที่คำนวณได้จากสูตรด้วยสาระสนเทศนี้ เราสามารถใช้ค่าความคลาดเคลื่อนมาตรฐานของ r_{bis} ในการประมาณขนาดของค่าสหสัมพันธ์ที่มีนัยสำคัญทางสถิติได้ ซึ่งมันควรจะได้ มีการทำซ้ำแล้วซ้ำอีก ๆ ด้วยเครื่องคอมพิวเตอร์ ค่าสหสัมพันธ์ไบซีเรียลก็จะหาได้จากกลุ่มตัวอย่างทั้งหมด

ข) การวิเคราะห์ข้อสอบอิงเกณฑ์ (Item Analysis for Criterion Referenced test)

การทดสอบความรอบรู้ (Mastery testing) กลายเป็นที่นิยมใช้ในห้องเรียนมากขึ้นระหว่าง 10 ปีที่แล้วมา ทั้งนี้ด้วยเหตุผล 2 อย่าง คือ (1) มีการใช้โปรแกรมการศึกษาเป็นรายบุคคลเพิ่มมากขึ้น และ (2) การเคลื่อนไหวทางการศึกษาที่ต้องการยกระดับของตนเองและต้องการให้ครูและระบบโรงเรียน มีความรับผิดชอบมากขึ้นโดยใช้การสอนของครูเป็นตัววัด

การทดสอบความรอบรู้นับได้ว่าเป็นวิธีการประเมินผลงานที่ครอบคลุมและจุดตัดซึ่งทำหน้าที่แยกแยะความแตกต่างระหว่างกลุ่มคนที่รอบรู้ ในงานที่ได้กำหนดไว้แต่แรก กับกลุ่มคนที่ไม่รอบรู้ที่ทำได้อย่างประสบความสำเร็จ Glaser (1963) กล่าวว่า คะแนนจากแบบสอบอิงเกณฑ์ควรจะขยายความแตกต่างระหว่างกลุ่มและควรจะลดความ

แตกต่างกันภายในกลุ่ม สำหรับแบบสอบการรอบรู้นั้น การกระทำเช่นนี้เป็นการเลือกข้อ
 กระทบที่มีการจำแนกระหว่างคนที่รอบรู้ (mastery) กับคนที่ไม่รอบรู้ (non mastery) ข้อ
 กระทบที่ได้รับประกันว่าเป็นข้อที่ดีสำหรับแบบสอบความรอบรู้ก็คือ ข้อที่กลุ่มคนรอบรู้
 ตอบคำถามได้ถูกต้อง และกลุ่มคนไม่รอบรู้ตอบผิด และดัชนีที่เสนอ ณ ที่นี้ก็คืออัสยัมโน
 ทศน์พื้นฐานนี้ มีสถิติมากมายที่เสนอวิธีการเลือกข้อกระทบสำหรับแบบสอบความรอบรู้
 (mastery test) ซึ่งโชคไม่ดีเลยที่ค่าสถิติบางตัวให้มโนทัศน์ที่หละหลวม และบางสถิติก็มี
 วิธีการที่ซับซ้อนเกินกว่าที่จะมาใช้ในห้องเรียน

ค่าสถิติการทดสอบก่อน-หลัง (The Pretest-Posttest Statistic)

ในปี 1966 Cox และ Vargas ได้เสนอค่าสถิติการทดสอบก่อน-หลังซึ่งออกแบบไว้
 เพื่อคัดเลือกข้อกระทบของแบบสอบอิงเกณฑ์ ค่าสถิติการทดสอบก่อน-หลังซึ่งกำหนด
 สัญลักษณ์ D_{pp} เกิดจาก

สัดส่วนของนักเรียนที่ทำข้อนั้นถูกหลังการเรียน – สัดส่วนของนักเรียนที่ทำข้อถูก
 นั้นถูกต้องก่อนการเรียน

เพื่อให้ดูง่ายขึ้น อาจจะคำนวณค่าสถิติ D_{pp} จากตาราง ดังนี้

ตารางแสดงจำนวนนักเรียนที่ทำข้อ i ถูกต้องและทำผิดทั้งก่อนและหลังการเรียนการสอน

	Pre-test	Posttest
ตอบถูก	a	b
ตอบผิด	c	d
รวม	a+c	b+d

$$D_{pp} = \frac{b}{b+d} - \frac{a}{a+c}$$

วิธีนี้ ความสำคัญอยู่ที่การสอบ 2 ครั้ง Cox และ Vargas ได้แสดงให้เห็นว่าค่า
 D_{pp} และค่าอำนาจจำแนกที่ได้จากค่าสถิติอิงกลุ่มดั้งเดิม ต่างก็ให้ผลที่มีความแตกต่าง
 เพียงพอในการใช้คัดเลือกข้อกระทบ

ค่าสถิติการทดสอบก่อน-หลัง (Pretest-Posttest Statistic) นี้ เป็นวิธีเลือกข้อ
 กระทบของแบบสอบอิงเกณฑ์วิธีแรก เพราะเป็นวิธีที่อัสยัมโนทศน์ง่าย ๆ และง่ายต่อการ
 คำนวณ เหมาะที่จะใช้ในห้องเรียน

แต่ค่าสถิติการทดสอบก่อน-หลังนี้ ก็มีข้อเสียบางประการ ซึ่ง Berk (1980) และ Van der Linden (1981) ได้บอกข้อจำกัดของมันไว้ดังนี้

1. จำเป็นต้องมีการสอบถึง 2 ครั้ง
2. ปัญหาเกี่ยวกับการบริหารการสอบ 2 ครั้งในกระทงเดียวกัน
3. ปัญหาอิทธิพลของตัวแปรแทรกซ้อนที่มีผลต่อการเปลี่ยนแปลงของคะแนน
4. กลุ่มประชากรไม่อิสระจากกัน
5. ขาดความไว ในการจำแนกข้อกระทง ณ ที่จุดตัด

ค่าสถิติความสอดคล้อง (The Agreement Statistic)

ขณะที่สถิติคุณลักษณะแฝงได้ปรากฏชัดแล้วว่าเป็นดัชนีตัวแปรที่ดีที่สุด ในการเลือกข้อกระทงของแบบสอบอิงเกณฑ์ ในแง่มนทัศน์ของการทดสอบความรอบรู้ (mastery testing) แต่ก็พบว่าวิธีการนี้มีการคำนวณและมนทัศน์ที่ซับซ้อนไม่เหมาะที่จะนำมาใช้ในห้องเรียน ซึ่งในทางตรงกันข้าม ค่าสถิติการทดสอบก่อน-หลังสามารถคำนวณได้ง่าย แต่ไม่เหมาะสมด้วยเหตุผลที่กล่าวมาแล้ว ดังนั้น Harris & Subkoviak (1986) ได้เสนอสถิติอีกตัวหนึ่ง ใช้สัญลักษณ์ว่า $P(x_c)$ ใช้เป็นดัชนีในการเลือกข้อกระทงที่เข้ากับค่าสถิติ คุณลักษณะแฝงมากกว่า ค่าสถิติการทดสอบก่อน-หลัง แต่มีวิธีการคำนวณที่ง่าย

ค่า $P(x_c)$ คำนวณได้จากตาราง ซึ่งแสดงสถานภาพของการรอบรู้และไม่รอบรู้ (ตัดสินจากคะแนนรวม) และการตอบกระทง (ถูก-ผิด)

ตาราง 2x2 เพื่อใช้ในการคำนวณค่า $P(x_c)$

	รอบรู้	ไม่รอบรู้
ถูก	a_{11}	a_{12}
ผิด	a_{21}	a_{22}

โดยค่า $P(x_c)$ ถูกกำหนดอยู่ในสมการ (5) ดังนี้

$$P(x_c) = \frac{a_{11} + a_{22}}{N}$$

เมื่อ a_{11} คือ จำนวนคนรอบรู้ที่ตอบข้อกระทงนั้นถูก

a_{22} คือ จำนวนคนที่ไม่รอบรู้ที่ตอบข้อกระทงนั้นผิด

N คือ จำนวนผู้เข้าสอบทั้งหมด

ดังนั้น ค่า $P(x_c)$ สามารถตีความได้เหมือนกับค่าความน่าจะเป็นของความสอดคล้อง (agreement) ระหว่างผลลัพธ์ของข้อนี้กับผลลัพธ์ของข้อสอบทั้งฉบับ ข้อตกลงในอุดมการณ์ควรจะมีค่า $P(x_c)$ เป็น 1 ถ้าไม่มีความสัมพันธ์เกิดขึ้นระหว่างสถานภาพของการรอบรู้กับการตอบของข้อกระทง ชี้ดจำกัดล่างในทางปฏิบัติของค่า $P(x_c)$ คำนวณได้จาก

$$P(x_c) = \frac{(a_{11} + a_{12})(a_{11} + a_{21}) + (a_{21} + a_{22})(a_{12} + a_{22})}{N^2}$$

ค) การวิเคราะห์ข้อสอบวัดความเร็ว

เมื่อไรก็ตามที่การวัดนั้นจำเป็นต้องอาศัยความเร็วในการทำข้อสอบ การวิเคราะห์ข้อสอบโดยวิธีธรรมดาตามที่กล่าวมาย่อมเกิดปัญหา ดังนั้นรายชื่อที่คำนวณจากแบบสอบวัดความเร็วจะนำไปสู่การสรุปที่ผิด ค่าดัชนีรายชื่อที่ได้จากแบบสอบวัดความเร็วเกิดจากผลสะท้อนของตำแหน่งของข้อในแบบสอบมากกว่าเกิดจากความยาก และความมีอำนาจจำแนกด้วยตัวของมันเอง ข้อที่ปรากฏอยู่ที่ตำแหน่งท้ายๆ ของแบบสอบจะมีเปอร์เซ็นต์ของคนตอบถูกน้อย เพราะว่า จะมีเพียงไม่กี่คนที่ทำถึงข้อเหล่านี้ โดยไม่คำนึงว่าข้อนั้นอาจจะง่าย แต่ถ้ามันไปปรากฏอยู่ตรงท้ายๆ ของแบบสอบมันจะกลายเป็นข้อที่ยาก

ในลักษณะที่คล้ายกันนี้ ดัชนีจำแนกรายข้อมีแนวโน้มจะเกิดการประมาณค่าเกินความจริง (overestimated) ต่อข้อที่ไม่มีใครทำถึงเลย เพราะคนที่มีทักษะมากกว่าแนวโน้มจะทำให้ได้เร็วกว่า แล้วพวกเขาก็ค่อนข้างจะถึงข้อสุดท้ายๆ ได้มากกว่าเกณฑ์ขึ้นเพียงเล็กน้อย ถ้าข้อนั้นอยู่ที่ท้ายๆ ของแบบสอบวัดความเร็ว

เพื่อหลีกเลี่ยงค่าความยากบางข้อเหล่านี้ เราอาจจะจำกัดการวิเคราะห์ข้อเฉพาะที่มีคนทำทันเวลาเท่านั้น ซึ่งวิธีการนี้ก็ไม่ใช่วิธีแก้ปัญหานั้นที่น่าพอใจ แต่ก็พอใจได้ถ้ามีคนน้อยคนที่ไม่ถึงข้อท้ายๆ วิธีการเช่นนี้ทำให้จำนวนกรณีตัวอย่างลดลง และมีผลทำให้ข้อต่อๆ มาไม่ค่อยจะมีความเที่ยง ยิ่งไปกว่านั้นคนที่ถูกนำมาใช้วิเคราะห์ข้อท้ายๆ จะกลายเป็นคนที่ถูกคัดเลือกขึ้นมา ไม่ใช่เป็นตัวแทนของกลุ่มใหญ่ แต่เป็นกลุ่มที่มีทักษะสูงกว่า ซึ่งได้กล่าวมาแล้วว่า คนที่ทำเร็วก็มีแนวโน้มที่จะเป็นคนที่มีความสามารถสูง จึงมีผลทำให้การวิเคราะห์ในข้อท้ายๆ มีค่าความยากลดลงกว่าที่ควรจะเป็น เพราะเปอร์เซ็นต์ของคนทำได้ในกลุ่มทักษะสูงนี้มีมากกว่าในกลุ่มทั้งหมด ซึ่งเป็นที่น่าสังเกตว่า

วิธีแก้ไขที่แนะนำมานี้ มีความคลาดเคลื่อนตรงข้ามกับเวลาเราคำนวณ โดยใช้เปอร์เซ็นต์คนทำได้จากกลุ่มตัวอย่างทั้งหมด ซึ่งทำให้ค่าความยากเพิ่มขึ้นกว่าที่ควรจะเป็น

ผลกระทบจากวิธีการที่เสนอไว้ข้างต้นนี้ มีต่อดัชนีค่าจำแนกของข้อกระทงเห็นได้ชัดน้อยกว่า แต่ก็ไม่ใช่ค่าที่แท้จริง ยกตัวอย่างให้สังเกตว่า คนเข้าสอบที่ได้คะแนนต่ำบางคนทำเร็วโดยลวกๆ เช็ดหมดทุกข้อโดยมีบางข้อที่สับสนตั้งใจทำเท่านั้น อย่างนี้ก็จะทำให้ได้คะแนนน้อยกว่าปกติธรรมดาเมื่อเทียบกับกลุ่มผู้เข้าสอบกลุ่มคะแนนสูง ดังนั้น การวิเคราะห์ข้อท้ายๆ ก็จะติดเอาคนที่ตอบแถมมากบางคนเข้ามาซึ่งทำสอบแบบไม่ดีกว่าที่ควรจะเป็น และกลุ่มที่มีทักษะมากเป็นคนตอบเร็วจะมีจำนวนมาก ซึ่งคำตอบของพวกเขาจะถูกตัด เมื่อเป็นเช่นนั้น ค่าสหสัมพันธ์ระหว่างข้อกระทงกับเกณฑ์ ก็อาจมีโอกาสสูงกว่าที่ควรจะเป็นในกลุ่มตัวแทนกลุ่มใหญ่ แต่ถ้ามองในอีกแง่หนึ่งว่าผู้ตอบที่หายไปไม่ได้ถูกนำมาวิเคราะห์ในข้อท้ายๆ มีผลทำให้การวิเคราะห์ในข้อท้ายๆ มีพิสัยของความสามารถแคบลง ภายใต้งื่อนไขเช่นนี้ค่าดัชนีจำแนกของข้อกระทงท้ายๆ จะต่ำกว่าที่มันจะเป็นเมื่อคำนวณจากกลุ่มตัวอย่างทั้งหมด

ผลกระทบที่มีต่อความยากและค่าจำแนกข้อกระทงของแบบสอบวัดความเร็วตามที่กล่าวมาข้างต้นทั้งหมดนั้น ได้มีการพิสูจน์ให้เห็นประจักษ์โดย Wesman (1949) และ Mollenkopf (1950) (Anastasi 1982, 216) ต่อมาได้มีศึกษาเปรียบเทียบกลุ่มนักเรียนระดับมัธยมกลุ่มต่างๆ ที่ได้รับการทดสอบด้วยแบบสอบทางภาษาคู่ขนาน 2 ฟอรัม และแบบสอบทางด้านคณิตศาสตร์คู่ขนานอีก 2 ฟอรัม แต่ละฟอรัมจะประกอบด้วยข้อกระทงชุดเดียวกัน แต่ฟอรัมหนึ่งจะเรียงข้อกระทงจากข้อท้ายสุดของอีกฟอรัมหนึ่งมาเป็นข้อแรก และเรียงตามลำดับสวนทางกับอีกฟอรัมหนึ่ง และในแต่ละฟอรัมยังได้รับการบริหารการสอบ 2 แบบ คือ แบบที่จำกัดเวลา (speed conditions) และแบบจำกัดเวลาอย่างเสรีภาพมาก (power condition) ผลการวิจัยพบว่าตำแหน่งของข้อกระทงมีผลต่อดัชนีความยากและการจำแนก และเมื่อให้ลำดับตำแหน่งของข้อเดียวกันก็พบว่า กลุ่มที่พยายามทำจะมีเปอร์เซ็นต์ในการตอบถูกสูงกว่าและให้ค่าสหสัมพันธ์ระหว่างข้อกับเกณฑ์สูงกว่าด้วย

ง) การวิเคราะห์ข้อสอบอัตนัย (Analysis of Essay Test)

1. ความเที่ยงรายข้อของแบบสอบอัตนัย

การศึกษาความเที่ยงของข้อสอบอัตนัยส่วนมากจะต้องพิจารณาถึงความสอดคล้องระหว่างผู้ตรวจหลายคนหรือความสอดคล้องภายในตัวผู้ตรวจคนเดียวกับการให้คะแนนไม่ใช่ตัววัดความเที่ยงของการสอบเอง (Hopkins & Stanley 1981 : 210) สังเกตได้จากการเอาบทความจาก ETS มาศึกษา พบว่า ความเที่ยงของผู้อ่าน ในหัวข้อเดียวกันของข้อสอบอัตนัยใช้เวลาประมาณ 20-40 นาที จากผู้อ่านคนเดียวกัน ค่าความเที่ยงจากผู้อ่านคนเดียวกันเป็น .40 แต่ความเที่ยงของแบบสอบเองมีเพียง .25 ผู้อ่านไม่มีความเที่ยงในการจำกัด ค่าความเที่ยงของแบบสอบ แต่มันเป็นไปได้ตามทฤษฎีที่จะมีผู้อ่านที่สมบูรณ์แบบ มีความเที่ยงในการอ่านโดยไม่มีผลกระทบต่อความเที่ยงของแบบสอบ มีการศึกษาน้อยมากที่รายงานถึงสหสัมพันธ์ระหว่างแบบสอบอัตนัย 2 รูปแบบที่ออกแบบเพื่อวัตถุประสงค์เฉพาะอันหนึ่ง และให้นักเรียนกลุ่มเดียวกันทำและตรวจโดยผู้อ่านข้อสอบที่เชี่ยวชาญ Coffman (1971) ได้ทำการศึกษาด้วยวิธีนี้โดยใช้แบบสอบใน 16 วิชา ให้นักเรียนเกรด 8 จำนวน 952 คน ใน 11 รัฐ ทำแบบสอบแต่ละฉบับมีการสอบ 2 ชุด ที่ให้คะแนนเป็นอิสระกันจากครูที่เชี่ยวชาญ 2 คน การศึกษาได้มีการจับคู่ทั้งหมดที่เป็นไปได้เพื่อเปรียบเทียบความเที่ยงของแบบสอบที่สอดคล้องกับการให้คะแนนของแบบสอบ (ความเที่ยงของผู้อ่าน) การให้คะแนนที่อิสระจากกัน 2 ครั้ง ด้วยแบบสอบเดียวกัน มีความสัมพันธ์กัน .62 แต่การสอบต่างกัน 2 ชุด มีการให้คะแนนโดยครูคนเดียวกัน มีความสัมพันธ์กันเพียง .43 ในการศึกษาแบบเดียวกันนี้ของการสอบ New York Regent's Examinations ผลการเปรียบเทียบ คือ ค่าเฉลี่ยของความสอดคล้อง (average agreement) ของการให้คะแนนอิสระจากกัน 2 ครั้งในแบบสอบเดียวกันเป็น .72 แต่ค่าเฉลี่ยของความสอดคล้องของแบบสอบ 2 ชุด ที่ให้คะแนนโดยครูคนเดียวกันเป็น .42 ซึ่งก็แน่นอนถ้าข้อสอบ 2 ชุด ได้รับการตรวจด้วยครูที่ต่างกัน ความสอดคล้องที่คาดหวังก็ควรจะต่ำ

จากการศึกษาอีกครั้งหนึ่งของ Coffman (1971) การให้คะแนนที่อิสระจากกัน 2 ชุด โดยผู้เชี่ยวชาญในการตรวจข้อสอบอัตนัย มีค่าสหสัมพันธ์กันเป็น .94 สำหรับแบบฟอร์ม A และมีค่าสหสัมพันธ์กันเป็น .84 สำหรับแบบฟอร์ม B แต่ค่าสหสัมพันธ์ระหว่างแบบฟอร์ม A และแบบฟอร์ม B มีค่าเพียง .60 Coffman (1971) พบว่า

สหสัมพันธ์ระหว่างคำถามอัตนัย (45 นาที) ต่างกัน 2 ข้อ กับผลการตรวจจากผู้อ่านต่างกัน 2 คน ในการสอบเลื่อนชั้นวิชาประวัติศาสตร์อเมริกัน มีค่าเพียง .28 ถึง .48 ซึ่งขึ้นอยู่กับความถนัด 2 ข้อรวมอยู่ด้วย ในขณะที่ค่าสหสัมพันธ์ระหว่างผู้ตรวจ 2 คน ในการตรวจคำถามข้อเดียวกันเป็น .54 ถึง .67 Michael, Cooper, Shaffer & Wallis (1980) ได้ชี้ให้เห็นจุดสำคัญที่ได้จากการศึกษาเหล่านี้ว่า “ความสอดคล้องในการให้คะแนนแบบสอบอัตนัย มีค่าสูงกว่าค่าความเที่ยงที่แท้จริงของตัวแบบสอบเอง” การมีผู้ตรวจที่มีความเที่ยงก็สำคัญ แต่คะแนนที่ได้รับมีความสามารถในการสรุปพาดพิง (generalizable) ไปยังจักรวาลของเนื้อหา (content universe) มากน้อยนั้น มีความสำคัญมากกว่า ซึ่งการสอบเป็นเพียงผิวกัดตัวอย่างของเนื้อหาเท่านั้น

ค่าสัมประสิทธิ์ความเที่ยงของข้อสอบอัตนัยข้อหนึ่ง (Hopkins & Stanley 1981 : 211) คือ “ค่าสหสัมพันธ์ระหว่างแบบฟอร์มที่ต่างกัน 2 ฟอร์มในแบบสอบอัตนัย ซึ่งการให้คะแนนเป็นอิสระจากกัน” นั่นคือสัมประสิทธิ์ความเที่ยงของข้อสอบอัตนัยข้อหนึ่งเป็นค่าสหสัมพันธ์ระหว่างการให้คะแนนของครูคนที่ 1 ในข้อสอบอัตนัยฟอร์ม A กับการให้คะแนนของครูคนที่ 2 ในข้อสอบอัตนัยฟอร์ม B ของนักเรียนคนเดียวกัน ถ้าครู 2 คนให้คะแนนในข้อสอบอัตนัย ฟอร์ม A และครูอีก 2 คน ให้คะแนนในข้อสอบอัตนัยฟอร์ม B ค่าความเที่ยงก็ควรจะดีขึ้น สำหรับคะแนนรวมของนักเรียนแต่ละคนก็ควรจะอยู่ในฟอร์ม A ซึ่งได้รับการตรวจจากครู 2 คน ในฟอร์ม และคะแนนรวมในฟอร์ม B ก็ควรมาจากครูอีก 2 คน ที่ตรวจฟอร์ม B คะแนนรวมของนักเรียนในฟอร์ม A ก็ควรจะเป็นความสัมพันธ์กันกับคะแนนรวมในฟอร์ม B หรือพูดอีกนัยหนึ่งว่าความเที่ยงของแบบสอบจะดีขึ้น เมื่อผู้ตรวจมีความเที่ยงเพิ่มขึ้น

อย่างไรก็ดี จะสังเกตว่าคะแนนรวมของนักเรียนในแบบสอบทั้ง 2 ฟอร์มรวมกัน จะให้ความเที่ยงสูงกว่าที่ได้จากการเอาคะแนนรวมเฉพาะของฟอร์ม A หรือฟอร์ม B เท่านั้น ซึ่งสามารถพิสูจน์ให้เห็นได้โดยสูตรหาความเที่ยงของ Spearman – Brown เนื่องจากในที่นี้ต้องการที่จะรายงานการวิเคราะห์รายข้อของแบบสอบอัตนัย จึงไม่ขอพูดถึงการหาความเที่ยงของแบบสอบอัตนัยทั้งฉบับให้ละเอียด ณ ที่นี้

การวิเคราะห์หาความเที่ยงรายข้อของแบบสอบอัตนัย เป็นการวิเคราะห์รายข้อแบบหนึ่ง ซึ่งต้องขึ้นอยู่กับความเที่ยงของผู้ตรวจด้วย

2. ค่าความยากและอำนาจจำแนกรายข้อของแบบสอบอัตนัย

ดังที่กล่าวมาแล้วว่า ความเที่ยงรายข้อของแบบสอบอัตนัยขึ้นอยู่กับความเที่ยงของผู้ตรวจ การหาค่าความยากและค่าอำนาจจำแนกรายข้อของแบบสอบอัตนัยก็พบปัญหาแบบเดียวกัน คือ ผู้ตรวจข้อสอบจะต้องมีความเที่ยงในการตรวจด้วย มิเช่นนั้นแล้วค่าความยากและค่าอำนาจจำแนกรายข้อที่หามาได้ก็ไม่มี ความหมาย ดังนั้นก่อนที่จะมีการหาค่าความยากและค่าอำนาจจำแนกรายข้อ ก็ควรที่จะมีการตรวจสอบความเที่ยงของผู้ตรวจข้อสอบเสียก่อน วิธีการหาความเที่ยงของผู้ตรวจที่ได้กล่าวมาแล้วในตอน 4.1 เป็นวิธีหนึ่งที่สามารถตรวจสอบความเที่ยงของผู้ตรวจได้ดี ในที่นี้จะขอเสนอวิธีการตรวจสอบความเที่ยงของผู้ตรวจอีกวิธีหนึ่งโดยใช้แบบสอบคู่ขนานที่เป็นแบบเลือกตอบ (Multiple choice) ซึ่งแบบสอบคู่ขนานแบบเลือกตอบนี้ อาจจะประกอบด้วยจำนวนข้อสอบชุดหนึ่งประมาณ 2-5 ข้อ ที่มีเนื้อหาครอบคลุม และวัดในจุดประสงค์เดียวกันกับข้อสอบอัตนัยข้อหนึ่งนั้น การตรวจสอบความเป็นคู่ขนานของแบบสอบทั้ง 2 รูปแบบนี้คงต้องเป็นไปตามทฤษฎีของแบบสอบคู่ขนาน อย่างไรก็ตาม มีบางเนื้อหาและบางจุดประสงค์เท่านั้นที่สามารถสร้างแบบสอบคู่ขนานแบบเลือกตอบมา เทียบเคียงกับข้อสอบอัตนัยข้อหนึ่งๆได้ เนื่องจากแบบสอบแบบเลือกตอบมีความเที่ยงในการให้คะแนน จึงนำข้อสอบแบบเลือกตอบนี้มาเป็นเกณฑ์ในการเปรียบเทียบการให้คะแนนข้อสอบอัตนัย นักเรียนที่ทำแบบสอบคู่ขนานแบบเลือกตอบได้คะแนนสูงก็ควรที่จะทำข้อสอบแบบอัตนัยข้อนั้นได้คะแนนสูงด้วย ค่าสหสัมพันธ์ของคะแนนที่ได้จากแบบสอบ 2 รูปแบบของนักเรียนคนเดียวกันเป็นตัวชี้ความเที่ยงของผู้ตรวจข้อสอบอัตนัย

เมื่อเราสามารถตรวจสอบได้ว่าผู้ตรวจข้อสอบอัตนัยมีความเที่ยงมากเพียงไรแล้ว การหาค่าความยากของข้อสอบอัตนัยก็ไม่ยาก เนื่องจากข้อสอบอัตนัยมีคะแนนเต็มไม่ใช่ 1 แต่เป็นเลขใดๆตามที่ผู้ออกข้อสอบจะเป็นผู้กำหนด สมมติให้มีคะแนนเต็มเป็น x_r ผู้ออกข้อสอบยังต้องกำหนดจุดตัดของคะแนนที่เป็นตัวบอกว่านักเรียนมีความรอบรู้ (Mastered) ในข้อนี้ สมมติให้มีคะแนนจุดตัดเป็น x_c ดังนั้น สัดส่วนหรือเปอร์เซ็นต์ของนักเรียนที่ผ่านจุดตัดหรือรอบรู้ในข้อสอบอัตนัยข้อ i ก็คือค่าความยากของข้อสอบอัตนัยข้อ i นั้นเอง

สำหรับการหาค่าอำนาจจำแนกรายข้อก็กระทำในทำนองเดียวกัน เราสามารถหาสัดส่วนของคนที่ตอบผ่านเกณฑ์ในข้อที่ i ของกลุ่มคะแนนสูงได้ และของคนในกลุ่มต่ำได้ ความแตกต่างของสัดส่วนทั้ง 2 กลุ่ม จะเป็นตัวบ่งชี้ค่าอำนาจของข้อนั้น ค่าอำนาจ

จำแนกรายชื่อของข้อสอบอัตนัยนี้ก็มีค่าตั้งแต่ -1 ถึง 1 เช่นเดียวกับค่าอำนาจจำแนกรายชื่อของข้อสอบแบบปรนัยหรือแบบเลือกตอบ

การหาค่าอำนาจจำแนกรายชื่อด้วยวิธีการหาค่าสหสัมพันธ์นั้นก็สามารถทำได้ โดยหาค่าสหสัมพันธ์ระหว่างคะแนนที่ได้จากข้อสอบอัตนัยกับคะแนนที่ได้จากแบบสอบที่เป็นเกณฑ์ ซึ่งแบบสอบที่เป็นเกณฑ์ก็คือแบบสอบที่คู่ขนานกับข้อสอบอัตนัย หรือแบบสอบอัตนัยทั้งฉบับ แบบสอบคู่ขนานแบบเลือกตอบที่ได้กล่าวไว้ในข้างต้นนั้น สามารถนำมาเป็นแบบสอบที่เป็นเกณฑ์เพื่อหาค่าสหสัมพันธ์ระหว่างข้อสอบอัตนัยข้อนั้นๆ กับเกณฑ์ ซึ่งค่าสหสัมพันธ์ที่สูงก็แสดงว่าข้อสอบอัตนัยข้อนั้นมีค่าอำนาจสูงและเช่นเดียวกับค่าสหสัมพันธ์ทั่วไป ค่าอำนาจนี้มีค่าตั้งแต่ -1 ถึง 1

Whitney และ Sabers (1970) ได้เสนอวิธีหาค่าความยาก และค่าอำนาจจำแนกของแบบสอบแบบอัตนัย

1. จำแนกผู้เข้าสอบออกเป็นกลุ่มสูง 25% กลุ่มต่ำ 25%
2. คำนวณผลบวกของคะแนนในข้อ i ของคนกลุ่มสูง ($\sum H$) และคนกลุ่มต่ำ ($\sum L$)
3. ใช้สูตรดังนี้

$$\text{ค่าความยากรายข้อ (P)} = \frac{\sum H + \sum L - (2nx_{\min})}{2n(x_{\max} - x_{\min})}$$

$$\text{ค่าอำนาจจำแนกรายข้อ (D)} = \frac{\sum H - \sum L}{n(x_{\max} - x_{\min})}$$

เมื่อ $\sum H$ = ผลบวกของคะแนนของคนกลุ่มสูง 25%
 $\sum L$ = ผลบวกของคะแนนของคนกลุ่มต่ำ 25%
 N = 25% ของผู้เข้าสอบ
 $x_{\max}$ = คะแนนสูงสุดข้อข้อ
 $x_{\min}$ = คะแนนต่ำสุดของข้อ

P		D	
0.80 – 1.00	ง่ายเกินไป	.40 – 1.00	ดีมาก
.21 – .79	ปานกลาง	.20 - .39	พอใช้ได้แต่ต้องปรับปรุงบ้าง
0 -.20	ยากเกินไป	.10 -.19	ส่วนใหญ่ยังต้องแก้ไข
		0 - .09	แย่มาก ควรตัดทั้งหมดหรือแก้ไขใหม่

สรุปเกณฑ์การพิจารณา

1. ข้อสอบปรนัย

เนื่องจาก ค่าความยากมีค่าอยู่ระหว่าง 0 ถึง 1 ดังนั้นการแปลผลจึงกำหนดดังนี้

ค่าความยากระหว่าง .33 - .66 คือ ยากปานกลาง

ค่าความยากสูงกว่า .66 คือ ง่าย

ค่าความยากต่ำกว่า .33 คือ ยาก

เนื่องจากอำนาจจำแนกมีค่าตั้งแต่ค่าติดลบจนถึง ค่า 1 ดังนั้นการแปลผลจึงกำหนดดังนี้

อำนาจจำแนกระหว่าง .33 - .66 คือ จำแนกได้พอสมควร

อำนาจจำแนกสูงกว่า .66 คือ จำแนกสูง

อำนาจจำแนกที่มีค่าติดลบจนถึง .32 คือ จำแนกต่ำ

2. ข้อสอบที่ให้คะแนนต่างจาก 0.1

เนื่องจากค่าความยากมีตั้งแต่ 0 ถึง 100% ดังนั้นการแปลผลจึงกำหนดดังนี้

(อุทุมพร จามรมาน 2546, หน้า 64)

ค่าความยากระหว่าง 33% - 66% ของคะแนนเต็ม คือ ยากปานกลาง

ค่าความยากสูงกว่า 66% ของคะแนนเต็ม คือ ง่าย

ค่าความยากต่ำกว่า 33% ของคะแนนเต็ม คือ ยาก

เนื่องจากค่าอำนาจจำแนกมีค่าตั้งแต่ค่าติดลบจนถึง 100% ดังนั้นการแปลผลจึงกำหนดดังนี้ (อุทุมพร จามรมาน 2546, หน้า 64)

ค่าอำนาจจำแนกระหว่าง 33% - 66% ของคะแนนเต็ม คือ จำแนกได้พอสมควร

ค่าอำนาจจำแนกสูงกว่า 66% ของคะแนนเต็ม คือ จำแนกสูง

ค่าอำนาจจำแนกติดลบจนถึง 32% ของคะแนนเต็ม คือ จำแนกต่ำ

อย่างไรก็ตามการกำหนดเกณฑ์แปรผลข้อสอบ อาจเปลี่ยนแปลงค่าได้เช่น อาจกำหนดความยากปานกลาง ระหว่าง .25 - .75 (หรือ 25% - 75%) ก็ได้ ซึ่งควรใช้พิจารณาเองด้วย

2.3.2 การวิเคราะห์คุณภาพทั้งฉบับ

ค่าความเที่ยง (Reliability) ของแบบทดสอบ

ชื่อ	สูตร	หมายเหตุ
KR – 20	$\frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n p_i q_i}{s_x^2} \right)$	ใช้เมื่อข้อมูลเป็น 0,1 และข้อสอบมีความยากง่ายใกล้เคียงกัน
Alpha (a)	$\frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n s_i^2}{s_x^2} \right)$	ใช้เมื่อข้อมูลมีค่าหรือคะแนนต่อเนื่อง
Test-retest	$r_{xy} = \frac{s_{xy}}{s_x s_y}$	เมื่อสิ่งที่วัดเป็นระดับของการปฏิบัติหรือข้อเท็จจริง

เมื่อ x คือ คะแนนที่สอบครั้งแรก

Y คือ คะแนนที่สอบครั้งที่ 2

N คือ จำนวนข้อสอบ

p_i คือ สัดส่วนผู้ตอบข้อสอบข้อ i ถูก

q_i คือ สัดส่วนผู้ตอบข้อสอบข้อ i ผิด

s_i^2 คือ ความแปรปรวนของข้อสอบข้อ i

s_x^2 คือ ความแปรปรวนของข้อสอบทั้งหมด

$\bar{x}$ คือ คะแนนเฉลี่ยของกลุ่มที่ทำข้อสอบได้ถูก

ตัวอย่างการคำนวณ KR – 20

ข้อสอบ	สัดส่วน		
	q ตอบถูก	q ตอบผิด	pq
1	.8	.2	.16
2	.6	.4	.24
3	.8	.2	.16
4	.6	.4	.24
5	.4	.6	.24
รวม			1.04

ตัวอย่างข้อสอบ 5 ข้อ มีผู้เข้าสอบ 30 คน เมื่อตรวจข้อสอบพบว่าข้อ 1 มีผู้ตอบถูก 24 คน หรือมีผู้ตอบถูก .8 และตอบผิด .2 ข้อ 5 มีผู้ตอบถูก .4 ตอบผิด .6 การประมาณค่าความเที่ยงโดยใช้สูตร KR 20 ทำได้ดังนี้ เมื่อ $n = 5$ และสมมุติค่าความแปรปรวน (s_x^2) มีค่าเท่ากับ 1.68

ตัวอย่าง การคำนวณหาความเที่ยงของเครื่องมือที่วัดพฤติกรรมหรือระดับของการปฏิบัติ โดยใช้เทคนิคการวัดซ้ำ (test-retest)

การวัดครั้งที่ 1 และครั้งที่ 2 ให้มีความห่างกันประมาณ 3-4 สัปดาห์

คนที่	คะแนนครั้งที่ 1 (x)	คะแนนครั้งที่ 2 (y)
1	7	8
2	4	3
3	8	7
4	6	5
5	3	4

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y}) / n}{S_x S_y} = .860$$