หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

บทนี้กล่าวถึงหลักการ แนวคิด ทฤษฎีและผลงานวิจัยที่เกี่ยวข้อง โดยแบ่งออกเป็น 3 ตอน คือ ตอนที่ 1 มโนทัศน์เกี่ยวกับดัชนีชี้คุณภาพแบบทดสอบตามแนวคิดของแบบแผนการตอบของกัตแมน ตอนที่ 2 มโนทัศน์เกี่ยวกับแบบแผนการตอบข้อกระทง ตามแนวคิดของซาโต้ และตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 มโนทัศน์เกี่ยวกับดัชนีชี้คุณภาพแบบทดสอบตามแนวคิดของแบบแผนการตอบของกัตแมน

ในตอนนี้ผู้วิจัยจะเสนอนิยาม พร้อมทั้งวิธีการวิเคราะห์ดัชนีต่าง ๆ ที่ได้มาจากแบบแผนการตอบตามแนวคิดของกัตแมน ได้แก่ ดัชนี Reproducibility และดัชนีอื่น ๆ

Reproducibility หมายถึง ระดับการจัดเรียงได้ของข้อกระทงในแบบสอบฉบับใด ๆ หลังจากดำเนินการสอบมาแล้วด้วยวิธีการเรียงที่ยึดคะแนนรวมของแต่ละบุคคลและระดับความยากของข้อกระทงเป็นเกณฑ์ ซึ่งคิดโดยกัตแมน (Guttman's reproducibility) โดยสรุปจะดำเนินการเป็นขั้นตอนดังนี้คือ

1. ดำเนินการสอบโดยใช้แบบสอบฉบับหนึ่งกับกลุ่มผู้สอบกลุ่มหนึ่ง
2. ตรวจคำตอบโดยระบบการให้คะแนน 1 คะแนน สำหรับคำตอบที่ถูก และ 0 คะแนน สำหรับคำตอบที่ผิด (dichotomous)
3. นำข้อกระทงมาจัดเรียงใหม่ โดยเรียงตามลำดับมากน้อยของคะแนนรวม และระดับความยากของข้อกระทง

ตัวอย่างการดำเนินการ ปรากฏดังตารางที่ 2 และ 3 ต่อไปนี้

ตารางที่ 2 แบบแผนการตอบของผู้สอบ 10 คน และข้อสอบ 6 ข้อกระทง ก่อนการจัดเรียงข้อกระทง

ผู้สอบ	ข้อกระทง 1	2	3	4	5	6	คะแนนรวม
A	0	0	0	1	0	0	1
B	1	0	1	1	1	1	5
C	0	0	1	0	0	0	1
D	1	1	1	0	1	1	5
E	1	0	1	1	1	0	4
F	0	0	1	0	1	0	2
G	0	1	1	0	0	0	2
H	1	0	1	0	1	0	3
I	1	0	1	0	1	1	4
J	1	0	0	1	1	0	3
ความยากของข้อกระทง	6	2	8	4	7	3	

ตารางที่ 3 แบบแผนการตอบของผู้สอบ 10 คน และข้อสอบ 6 ข้อกระทง หลังจากจัดเรียงข้อกระทงตามแนวคิดของกัตแมน

ผู้สอบ	ข้อกระทง 2	6	4	1	5	3	คะแนนรวม
D	1	1	0	1	1	1	5
E	0	1	1	1	1	1	5
B	0	0	1	1	1	1	4
I	0	1	0	1	1	1	4
H	0	0	0	1	1	1	3
J	0	0	1	1	1	0	3
G	1	0	0	0	0	1	2
F	0	0	0	0	1	1	2
C	0	0	0	0	0	1	1
A	0	0	1	0	0	0	1
ความยากของข้อกระทง	2	3	4	6	7	8	

จากตารางที่ 2 และ 3 จะเห็นว่า ในกรณีตารางที่ 2 นั้น ผู้สอนได้นำคะแนนจากการทำการทดสอบนักเรียน 10 คน ด้วยแบบสอบที่ประกอบด้วยข้อกระทง 6 ข้อ มาจัดเรียงตามลำดับข้อกระทงจากข้อที่ 1 ถึงข้อที่ 6 (เรียงตามปกติ) ส่วนในตารางที่ 3 เป็นการจัดเรียงผลการสอบในรูปแบบใหม่ คือ เรียงตามลำดับคะแนนรวม และตามลำดับความยากของข้อกระทง กล่าวคือ เรียงลำดับคะแนนจากผู้ได้คะแนนสูงสุด คือ D ซึ่งได้ 5 คะแนน จนคนสุดท้ายคือ A ซึ่งได้คะแนนเพียง 1 คะแนน ในขณะเดียวกันก็เรียงลำดับข้อกระทงจากข้อที่ยากที่สุดคือ ข้อ 2 (มีผู้ตอบถูกเพียง 2 คน) จนถึงข้อที่ง่ายที่สุดคือข้อ 3 ซึ่งมีผู้ตอบถูก 8 คน (การจัดเรียงในกรณีนี้ถือเป็นการจัดเรียงใหม่ตามแนวคิดของ Guttman)

สาระสรุปของการจัดเรียงลำดับใหม่

1. ถ้าการจัดเรียงสามารถทำได้อย่างสมบูรณ์ รูปแบบของตารางใหม่จะได้ในรูปแบบที่มีเลข 1 ในรูปสามเหลี่ยมด้านบน และเป็นเมตริกซ์ของเลข 0 ในรูป สามเหลี่ยมด้านล่าง (ในกรณีเรียงลำดับข้อสอบจากข้อที่ยากไปยังข้อที่ง่าย) รูปแบบการจัดเรียงที่สมบูรณ์จะได้เมตริกซ์ที่เรียกว่า Triangular Matrix

2. ในกรณีที่มีผู้ได้คะแนนรวมซ้ำกัน (ลำดับที่ซ้ำกัน) ผลการเรียงลำดับจะไม่อยู่ในรูปแบบที่สมบูรณ์ กล่าวคือ จะมีเลข 0 เข้าไปกระจายอยู่ในกลุ่มของเลข 1 หรือมีเลข 1 เข้าไปปนอยู่กับกลุ่มเลข 0

3. จำนวนเลข 0 ที่เข้าปะปนอยู่ในกลุ่มของเลข 1 หรือ เลข 1 ที่เข้าไปปะปนอยู่ในกลุ่มของเลข 0 เราถือว่าเป็นความคลาดเคลื่อน (Errors)

วิธีการคำนวณหาค่าดัชนีเพื่อบ่งชี้ระดับการจัดเรียงได้ (Reproducibility Index)

การคำนวณค่า Reproducibility Index จะทำการจัดระบบของตัวเลขโดยดำเนินการดังนี้

1. ลากเส้นเกณฑ์เพื่อแบ่งเขตระหว่างกลุ่มของเลข 1 กับกลุ่มของเลข 0
2. นับจำนวนเลข 1 หรือเลข 0 ซึ่งถือว่าเป็นความคลาดเคลื่อน
3. แทนค่าต่าง ๆ ในสูตรที่ (1) และ (2)

การดำเนินการตามขั้นตอนนี้สรุปได้ดังตารางที่ 4 ดังนี้

ตารางที่ 4 สรุปความคลาดเคลื่อนที่เกิดจากการจัดเรียงลำดับข้อกระทง

ผู้สอบ	ข้อกระทง 2	6	4	1	5	3	คะแนนรวม
D	1	1	(0)	1	1	1	5
E	0	1	1	1	1	1	5
B	0	0	1	1	1	1	4
I	0	(1)	0	1	1	1	4
H	0	0	0	1	1	1	3
J	0	0	(1)	1	1	(0)	3
G	(1)	0	0	0	(0)	1	2
F	0	0	0	0	1	1	2
C	0	0	0	0	0	1	1
A	0	0	(1)	0	0	0	1
ความยากของข้อกระทง	2	3	4	6	7	8	
ความคลาดเคลื่อน	1	1	3	0	1	1	

จากตาราง ปรากฏความคลาดเคลื่อนในแต่ละข้อเป็น 1,1,3,0,1 และ 1 ตามลำดับ รวมความคลาดเคลื่อน 7 ค่า

สูตรที่ใช้ในการคำนวณ

ค่า Reproducibility Index ของข้อกระทงรายข้อ

$$R_i = 1 - \frac{E}{N} \qquad \text{.....(1)}$$

ค่า Reproducibility Index ของแบบสอบทั้งฉบับ

$$R_t = 1 - \frac{\sum_{i=1}^{K} E_i}{NK} \qquad \text{.....(2)}$$

E คือ ความคลาดเคลื่อน ($\sum_{i=1}^{K} E_i$ คือ รวมความคลาดเคลื่อน)

N คือ จำนวนผู้สอบ

K คือ จำนวนข้อกระทง

R_i คือ ค่าดัชนีรายข้อ

R_t คือ ดัชนีทั้งฉบับ

จากกรณีตัวอย่าง ได้ค่าดัชนี Reproducibility ของแบบสอบทั้งฉบับเท่ากับ .88 หรือ 88 %

การประมาณค่า Reproducibility โดยวิธีการของแจคสัน (Jackson's plus percentage ratio (PPR))

แจคสันได้พัฒนาแนวความคิดในการประมาณค่า Reproducibility ที่แตกต่างไปจากวิธีการของ Guttman ทั้งนี้โดยมุ่งหวังที่จะปรับค่าดัชนี Reproducibility ให้มีค่าสูงสุดเป็น 1.00 และสามารถประมาณค่าต่ำสุดได้ด้วย

วิธีการของแจคสัน เริ่มต้นด้วยการจัดระบบตาราง ดังตารางที่ 5

ตารางที่ 5 การคำนวณค่า Reproducibility (R) ค่าต่ำสุดของ Reproducibility (MR) และ Plus Percentage Ratio (PPR)* โดยวิธีของแจคสัน

	Item											
	2		6		4		1		5		3	
Subject	+	−	+	−	+	−	+	−	+	−	+	−
D	1		1			(0)	1		1		1	
E		0	1		1		1		1		1	
B		0		0	1		1		1		1	
I		0	(1)			0	1		1		1	
H		0		0		0	1		1		1	
J		0		0	(1)		1		1			(0)
G	(1)			0		0		0		(0)	1	
F		0		0		0		0	1		1	
C		0		0		0		0		0	1	
A		0		0	(1)			0		0		0
# right (P)	2		3		4		6		7		8	
# wrong (Q)		8		7		6		4		3		2
Errors	1		1		3		0		1		1	
R_i	.90		.90		.70		1.00		.90		.90	
MR_i	.80		.70		.60		.60		.70		.50	
PP_i	.10		.20		.10		.40		.20		.10	
PPR_i	.50		.67		.25		1.00		.67		.50	

* Rights are listed under +. Wrongs are listed under −. Total errors = 7; R_t = 88%; MR_t = 70%; PP_t = 18%; PPR_t = .61.

ตามตารางที่ 5 ในแต่ละข้อกระทงจะแยกคะแนนออกเป็น 2 คอลัมน์คือ "เลข 0" และ "เลข 1 "

รวมจำนวนคนที่ตอบถูก (# right P) และจำนวนคนที่ตอบผิด (# wrong Q) ในข้อกระทง

ประมาณค่าต่าง ๆ ด้วยสูตรต่อไปนี้

1. คำนวณ Reproducibility ที่ต่ำสุดของแต่ละข้อกระทง (Minimum Reproducibility)

$$MR_i = \frac{\text{\# rights, or \# wrongs (whichever is larger)}}{N} \quad .....(3)$$

2. คำนวณค่าดัชนี Reproducibility ที่ต่ำที่สุดของแบบสอบ

$$MR_t = \frac{\sum_{i=1}^{k} \text{\#rights, or \# wrongs (whichever is larger)}}{KN} \quad ....(4)$$

3. คำนวณค่า Plus-Percentage ของแต่ละข้อกระทง

$$PP_i = R_i - MR_i \quad ....(5)$$

4. คำนวณค่า Plus-Percentage Ratio ของแต่ละข้อกระทง

$$PPR_i = \frac{R_i - MR_i}{1 - MR_i} \quad ....(6)$$

5. คำนวณค่า Plus-Percentage Ratio ของแบบสอบทั้งสองฉบับ

$$PPR_t = \frac{R_t - MR_t}{1 - MR_t} \quad ....(7)$$

สาระโดยสรุปตามวิธีการของแจคสัน

1. PPR_t เป็นดัชนีชี้คุณภาพการจัดเรียงได้ของข้อกระทง มีค่าระหว่าง 0 กับ 1

2. MR_t เป็นการประมาณค่า Reproducibility ที่ต่ำที่สุดของแบบสอบ

3. PPR_t ค่าที่ได้ต่ำกว่าค่า R_t ที่คำนวณตามวิธีการของ Guttman

4. วิธีการของแจคสันแก้ปัญหาเรื่องของการไม่เป็นลำดับในเรื่องค่าความยากของข้อกระทง (นักเรียนบางคนอาจทำผิดที่ข้อง่าย แต่กลับทำถูกในข้อที่ยาก)

5. จากตารางได้ค่า PPR_t เท่ากับ .61

ดัชนีบ่งชี้ความเป็นเอกพันธ์ของข้อกระทงของโลวิงเจอร์ (Loevinger's index of Homogeneity (H) Homogeneity of Test (H_t)

Loevinger (1948) ได้ทำการศึกษาเกี่ยวกับความเป็นเอกพันธ์ของข้อกระทงในแบบสอบ เขากล่าวว่า

"ความเป็นเอกพันธ์ของแบบสอบอย่างสมบูรณ์ สามารถพูดได้ในรูปของความน่าจะเป็น กล่าวคือ แบบสอบที่มีความเป็นเอกพันธ์อย่างสมบูรณ์นั้น ถ้าเราจัดเรียงข้อกระทงตามระดับความยากแล้ว ในกรณีที่ทราบว่าบุคคลผู้หนึ่งได้คะแนนหรือสอบผ่านในข้อใดข้อหนึ่ง ก็จะสรุปได้ว่า เขาจะต้องได้คะแนนหรือสอบผ่านในทุก ๆ ข้อที่ผ่านมา (ที่ง่ายกว่า) และในทางกลับกัน แบบสอบที่มีลักษณะวิวิธพันธ์อย่างสมบูรณ์ ความน่าจะเป็นที่บุคคลใด ๆ จะได้คะแนนในแต่ละข้อจะเป็นอิสระขาดจากกัน"

จากนิยามของความเป็นเอกพันธ์ของ Loevinger ดังกล่าวข้างต้น จะเห็นว่ามีลักษณะที่คล้ายคลึงกับนิยาม Reproducibility ของ Guttman คือ

1. Loevinger ใช้วิธีเรียงลำดับข้อกระทงตามระดับความยาก

2. คำนวณดัชนี S ตามสูตรที่ (8) ซึ่งใช้วิธีการเปรียบเทียบข้อกระทงเป็นคู่

3. คำนวณดัชนี Smax ตามสูตรที่ (9)

4. ใช้สูตรที่ (10) หรือ (11) เป็นสูตรในการคำนวณดัชนีความเป็นเอกพันธ์ของข้อกระทงในแบบสอบใด ๆ

$$S = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} P_{ij} - P_i P_j \quad \text{.....(8)}$$

$$S_{max} = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} P_j - P_iP_j \quad .....(9)$$

$$H_t = \frac{S}{S_{max}} \quad .....(10)$$

$$H_t = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} (1-P\hat{i}_j)-(1-P\hat{i}P_j)}{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} 1-(1-P_iP_j)} \quad ..(11)$$

P_{ij} คือ สัดส่วนจำนวนผู้สอบที่ผ่านทั้งข้อกระทงที่ i และข้อกระทงที่ j

P_iP_j คือ สัดส่วนในเชิงทฤษฎีของการผ่านข้อกระทงที่ i และข้อกระทงที่ j ซึ่งเป็นอิสระจากกัน

H_t คือ ดัชนีความเป็นเอกพันธ์ของแบบสอบ

สาระโดยสรุปเกี่ยวกับดัชนีความเป็นเอกพันธ์ของ Loevinger

1. เป็นดัชนีชี้ความเป็นเอกพันธ์ของข้อกระทงทั้งหมดในแบบสอบ

2. สูตรที่ (11) เป็นสูตรที่ทำการเปรียบเทียบดัชนี H_t กับดัชนี Minimum Reproducibility ที่คำนวณโดยสูตรที่ (4) ซึ่งเป็น Reproducibility ตามวิธีการของแจคสัน

3. ดัชนี H_t จะมีค่าน้อยกว่า PPR_t ของแจคสัน

4. Loevinger เสนอว่าในการใช้สูตรที่ (10) หรือ (11) ควรใช้ในกรณีที่มีผู้รับการทดสอบไม่น้อยกว่า 100 คน

วิธีการประมาณค่าดัชนี Reproducibility ของ Green

Green (1956) ได้ทำการศึกษาเกี่ยวกับรูปแบบการจัดเรียงผลการสอบของนักเรียน พร้อมทั้งเสนอดัชนีความคงที่ภายใน (Index of Consistency) ดังสูตรการคำนวณต่อไปนี้

$$Rep = 1 - \frac{1}{NK}\sum_{i=1}^{k-1} n_{\hat{i}}\,,\,_{i+1} - \frac{1}{NK}\sum_{i=2}^{k-2} n_{\overline{i-1}}\,,\,_{\hat{i},i+1,i+2} \quad \ldots\ldots(12)$$

$$Rep_{ind} = 1 - \frac{1}{N^2K}\sum_{i=1}^{k-1} n_{\hat{i}}n_{i+1} - \frac{1}{N^4K}\sum_{i=2}^{k-2} n_{\hat{i}}n_{i+1}\,n_{i+2}n_{\overline{i-1}} \quad \ldots\ldots(13)$$

$$1 = \frac{Rep - Rep_{ind}}{1.00 - Rep_{ind}} \quad \ldots\ldots(14)$$

Rep คือ ดัชนี Reproducibility ของแบบสอบ

Rep_{ind} คือ ดัชนี Reproducibility ของแบบสอบในกรณีข้อกระทงทั้งหลายเป็นอิสระขาดจากกัน

N คือ จำนวนผู้สอบ

K คือ จำนวนข้อกระทง

$\sum_{i=1}^{k-1} n_{\hat{i}}\,,\,_{i+1}$ คือ จำนวนผู้ที่ผ่านข้อกระทงที่ i+1 แต่ไม่ผ่านข้อกระทงที่ i

$\sum_{i=2}^{k-2} n_{i-1},_{\hat{i},i+1,i+2}$ คือ จำนวนผู้ที่ผ่านข้อกระทงที่ i+1 และ i+2 แต่ไม่ผ่านข้อกระทงที่ i และ i-1

ตัวอย่างการคำนวณปรากฏดังตารางที่ 6

ตารางที่ 6 การคำนวณค่า Reproducibility (Rep) Chance Reproducibility (Rep_{ind}) และดัชนีความคงที่ภายในโดยวิธีของกวีน

	ข้อกระทง						คะแนนรวม
ผู้สอบ	2	6	4	1	5	3	
D	1	1	0	1	1	1	5
E	0	1	1	1	1	1	5
B	0	0	1	1	1	1	4
I	0	1	0	1	1	1	4
H	0	0	0	1	1	1	3
J	0	0	1	1	1	0	3
G	1	0	0	0	0	1	2
F	0	0	0	0	1	1	2
C	0	0	0	0	0	1	1
A	0	0	1	0	0	0	1
การเรียงค่าความยาก	6	5	4	3	2	1	
n_i	2	3	4	6	7	8	
$n\hat{i}$	8	7	6	4	3	2	
$n\hat{i}_{,i-1}$		1	2	1	0	1	
$n\hat{i}_{-1,i.i+1,i+2}$			0	0	0		

$$Rep = 1 - \frac{1}{(10)(6)}(1+0+1+2+1) - \frac{1}{(10)(6)}(0+0+0) = .917$$

$$Rep_{ind} = \frac{1}{(10^2)(6)}(7.2+6.3+4.4+3.6+2.7) - \frac{1}{10^4(6)}(4.6.3.2 + 3.4.4.3 + 2.3.6.4)$$

$$= .860$$

$$I = \frac{Rep - Rep_{ind}}{1.00 - Rep_{ind}} = \frac{.916 - .860}{1.000 - .860} = .407$$

สาระโดยสรุปตามวิธีการของ Green

1. เป็นวิธีการรวมเอาข้อเด่นของ Loevinger และ Jackson เข้ามาวิเคราะห์รวมกัน

2. ดัชนีที่คำนวณได้ตามสูตรที่ 14 แสดงถึงความคงที่ภายในของแบบสอบ

3. จากสูตรที่ (14) ถ้าแบบสอบฉบับใดให้ค่า Reproducibility อย่างสมบูรณ์ จะทำให้ Rep_{ind} มีค่าเป็น "0" และจะให้ค่าดัชนีความคงที่ภายในเป็น 1.00

การประมาณค่าดัชนี Reproducibility ด้วยวิธีการใช้สัมประสิทธิ์ฟี PHI COEFFICIENT(ϕ_{it})

การประมาณค่า Reproducibility สามารถวิเคราะห์จากสูตรของสัมประสิทธิ์ ฟี ซึ่งให้ค่าสูงสุดเท่ากับ 1.00 และมีค่าต่ำสุด 0.00 แนวคิดเบื้องหลังของวิธีการนี้ก็คือ

"สมมุติว่ามีข้อกระทงหนึ่ง ซึ่งเป็นข้อกระทงที่บรรจุอยู่ในแบบสอบที่เป็น Perfectly Reproducible Test ถ้าปรากฏว่ามีผู้ตอบข้อกระทงข้อนี้ถูก 30 % และตอบผิด 70 % ก็จะหมายถึง กลุ่มผู้ได้คะแนนรวมต่ำ จะต้องตอบข้อกระทงข้อนั้นผิด" โดยนัยนี้ ข้อกระทงก็สามารถจัดเรียงใหม่ได้ตามลำดับความยากและลำดับคะแนนรวมของผู้สอบ

แนวคิดในการประมาณค่าสัมประสิทธิ์ แสดงได้ดังตารางที่ 7 และสูตรที่ (15)

ตารางที่ 7 คะแนนรวมและคะแนนข้อกระทงของสัมประสิทธิ์ฟี

		คะแนนรวม*		
		ต่ำ	สูง	รวม
คะแนนของข้อกระทง	ผ่านข้อกระทง i	A	B	A+B
	ไม่ผ่านข้อกระทง i	C	D	C+D
	รวม	A+C	B+D	N

* การแจกแจงของคะแนนรวมถูกแยกดังนั้นจำนวนผู้สอบในกลุ่มต่ำเท่ากับจำนวนผู้สอบที่ไม่ผ่านข้อกระทง i (C+D = A+C)

$$\phi_{it} = \frac{BC - AD}{(A + B)(C + D)} \qquad \text{..... (15)}$$

สาระโดยสรุปของสัมประสิทธิ์ ฟี

1. ผู้ตรวจสอบจะต้องกำหนดจุดตัดในการตัดสินการผ่าน - ไม่ผ่าน เพื่อจะได้ค่าสัมประสิทธิ์ ฟี ของแต่ละข้อกระทงที่สามารถจัดเรียงในรูปแบบของ "Reproducible"

2. วิธีการใช้สัมประสิทธิ์ ฟี จะไม่ถูกกระเทือนจากข้อกระทงที่มีความยากสุดขั้ว

3. เป็นเทคนิคที่ใช้ในการคัดเลือกข้อกระทงที่ไม่เหมาะสมออกจากแบบสอบ ซึ่งจะทำให้ได้แบบสอบที่มีความเป็นเอกพันธ์สูง การทดสอบนัยสำคัญของสัมประสิทธิ์ ฟี ทดสอบโดยไคสแควร์

สรุปแนวความคิดที่เกี่ยวข้องกับดัชนี Reproducibility

ผู้ศึกษาจะทำการสรุปในรูปของ Proposition ดังนี้

1. "Reproducibility เป็นเรื่องราวที่ว่าด้วยการจัดรูปแบบของผลการสอบของนักเรียนโดยคำนึงถึงระดับความยาก และคะแนนรวมของแต่ละบุคคล"

2. "Reproducibility Index เป็นดัชนีชี้บ่งระดับความสามารถในการจัดเรียงได้ของผลการสอบ"

3. "Guttman" เป็นผู้ริเริ่มแนวความคิดในเรื่อง Reproducibility และในเวลาต่อมาได้รับการพัฒนาโดยนักวัดผลอีกหลายท่าน เช่น Jackson ได้เสนอ Minimum Reproducibility of test และ Plus- Percentage Ratio และ Loevinger (1948) ได้เสนอดัชนีการตรวจสอบความเป็นเอกพันธ์ของข้อกระทง ซึ่งโดยนิยามแล้วจะมีความผูกพันกับดัชนี Reproducibility เป็นอย่างมาก และในเวลาต่อมา Green (1956) ได้พยายามรวบรวมข้อเด่นของ Loevinger และ Jackson เข้าด้วยกัน แล้วเสนอสูตรสำหรับการประมาณค่า Reproducibility (Rep) และ Index of Consistency รวมทั้งดัชนี ฟี ก็สามารถนำมาเพื่อประยุกต์ในการเลือกข้อกระทงเพื่อให้ได้ข้อกระทงที่มีความเป็นเอกพันธ์และสามารถจัดรูปแบบตามแนวความคิดในเรื่อง Reproducibility ได้"

4. "โดยสรุปแล้ว แนวความคิดในเรื่อง Reproducibility ,index of Homogeneity รวมทั้ง Index of Consistency ล้วนแล้วแต่เป็นเรื่องที่สอดคล้องกันและให้ข้อสรุปในลักษณะที่คล้ายคลึงกัน"

5. "แนวความคิดที่เกี่ยวข้องกับดัชนี Reproducibility ที่กล่าวมาแล้วข้างต้นมีลักษณะที่แตกต่างกันก็เพียงเฉพาะวิธีการในการประมาณ วิธีการในการประมาณ "ความคลาดเคลื่อน" หรือการนับจากจำนวนที่คลาดเคลื่อนในการจัดเรียงลำดับข้อกระทงเป็นสำคัญ"

6. "ถ้าแบบสอบใดสามารถจัดได้ในลักษณะที่เป็น "Perfectly Reproducible" ก็จะทำให้ค่าดัชนีความคงที่ภายใน และความเป็นเอกพันธ์ของแบบสอบเป็น 1.00 เช่นกัน"

REPRODUCIBILITY AND FACTOR ANALYSIS

หากพิจารณาเทคนิคการวิเคราะห์ตัวประกอบที่นำมาใช้ในการตรวจสอบคุณภาพของเครื่องมือ ก็นิยมนำมาใช้ใน 2 ลักษณะที่สำคัญ คือ "การตรวจสอบว่าแบบสอบฉบับใดบ้างที่มุ่งวัดคุณลักษณะภายในอย่างเดียวกัน หรือวัดองค์ประกอบร่วมกัน โดยนัยนี้เป็นการวิเคราะห์ความสอดคล้องกันระหว่างแบบสอบหลาย ๆ ฉบับ" อีกแนวทางหนึ่งก็คือ "การใช้เทคนิคการวิเคราะห์ตัวประกอบในการตรวจสอบว่าข้อกระทงทั้งหลายในแบบสอบฉบับใด ๆ ได้มุ่งวัดองค์ประกอบเดียวกันหรือไม่ หรือการตรวจสอบความเป็น Homogeneous ของแบบสอบใด ๆ" วิธีการที่เห็นได้ชัดเจนในกรณีหลังก็คือวิธีการของ Alpha Factor Analysis ของ Henry F.Kaiser (1965)

จากที่กล่าวมาจะเห็นได้ว่า เทคนิคการวิเคราะห์ตัวประกอบก็เป็นเทคนิคหนึ่งในการตรวจสอบความเป็นเอกพันธ์ของข้อกระทง ถ้าข้อกระทงในแบบสอบใดมีความเป็นเอกพันธ์สูงภายหลังจากที่ใช้เทคนิคการวิเคราะห์ตัวประกอบ ก็จะได้องค์ประกอบเดียว (one Factor) ซึ่งในการนี้ก็คือข้อกระทงในแบบสอบนั้นจะให้ค่าดัชนี Reproducibility สูงด้วย ซึ่งเป็นสภาพความสัมพันธ์ที่ชัดเจนระหว่าง Reproducibility กับ Factor Analysis (White and Saltz 1978: 252)

ตอนที่ 2 มโนทัศน์เกี่ยวกับแบบแผนการตอบข้อกระทงตามแนวความคิดของซาโต้

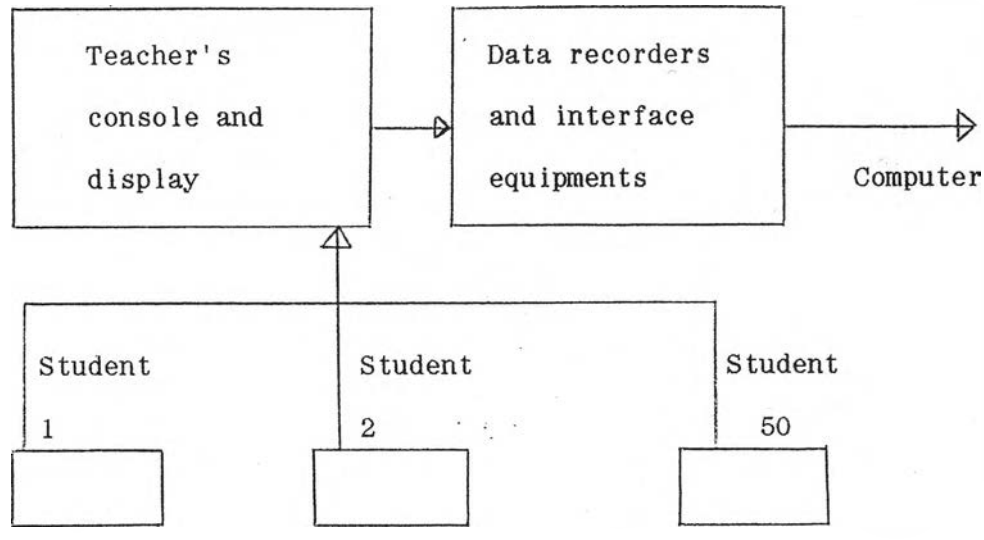
ตอนนี้เป็นเรื่องเกี่ยวกับความเป็นมาของแผนภูมิเอส - พี (S - P chart) ทฤษฎีเส้นโค้งเอส - พี (S - P Curve Theory) ดัชนีบ่งชี้ของซาโต้

1. **ความเป็นมาของแผนภูมิเอส - พี (S - P Chart)**

ในปี ค.ศ 1963 ทากาฮิโร ซาโต้ (Takahiro Sato 1980: 3) นักการศึกษาชาวญี่ปุ่นได้สร้างเครื่องกลชนิดหนึ่งขึ้นเพื่อใช้ในการเก็บรวบรวมข้อมูลในการตอบคำถามของนักเรียน เรียกว่า เครื่องวิเคราะห์การตอบคำถาม (The Response Analysis : RA)

ลักษณะของเครื่องกลจะประกอบไปด้วย 2 ส่วนคือ เครื่องรับและเครื่องส่ง เครื่องส่งจะอยู่ที่โต๊ะนักเรียน โดยจะประกอบไปด้วยแป้นกด (Keyboard) สำหรับนักเรียนกดเพื่อตอบคำถามของครูและมีเครื่องรับคำตอบของนักเรียนอยู่ที่โต๊ะครูดังภาพประกอบ (Sato 1980: 2)

ภาพประกอบ 1 แสดงระบบการทำงานของเครื่องกล ในการวิเคราะห์ผลการตอบข้อสอบ



การตอบข้อสอบของนักเรียน เครื่องวิเคราะห์ข้อสอบจะเก็บรวบรวมข้อมูลไว้ว่านักเรียนแต่ละคนเลือกตอบข้อสอบข้อใด ตัวเลือกแต่ละตัวเลือกมีจำนวนนักเรียนเลือกตอบเป็นจำนวนเท่าไร เวลาที่ใช้ในการตอบข้อสอบแต่ละข้อเป็นจำนวนเท่าไร เครื่องมือวิเคราะห์ผลการสอบจะเอาข้อมูลเกี่ยวกับเวลาที่ใช้ในการทำข้อสอบของนักเรียนแต่ละคนมาแจกแจงความถี่ว่าจะมีการแจกแจงของเวลาที่ใช้ในการทำข้อสอบเป็นรูปแบบใด เรียกว่า การแจกแจงแบบไวบูล (The Weibull distribution)(Tatsuoka and Tatsuoka 1980: 157) ต่อมาศาสตราจารย์ฮิโรอิชิ ฟูจิต้า แห่งมหาวิทยาลัยเคโอ (Keio University) ร่วมกับทากาฮิโร ซาโต้ ได้ร่วมกันพัฒนาการวิเคราะห์เวลาที่ใช้ในการตอบข้อสอบหรือการแจกแจงแบบไวบูลได้พัฒนาเป็นการวิเคราะห์คะแนนผลการสอบโดยได้คิดค้นการวิเคราะห์แบบแผนการตอบข้อสอบ โดยใช้ สถิติแบบนันพาราเมตริกแบบใหม่ (Non-Parametric) เรียกว่า แผนภูมิเอส-พี (S - P Chart) (Sato 1980: 3)

การวิเคราะห์ผลการสอบด้วยแผนภูมิเอส-พี ต่อมาได้ถูกเสนอให้เป็นทฤษฎีทางการวัดและซาโต้ได้ศึกษาเกี่ยวกับคุณลักษณะเบื้องต้นของแผนภูมิเอส-พี และได้ประยุกต์ใช้ทฤษฎีเส้นโค้งเอส-พี ในการปรับปรุงคุณภาพทางการศึกษาและในปัจจุบันทฤษฎีเส้นโค้งเอส-พี ได้รับความนิยมมาก ในการปรับปรุงคุณภาพการเรียนการสอนนักเรียนในประเทศญี่ปุ่น โดยเฉพาะโรงเรียนประถมศึกษา มัธยมศึกษาตอนต้น มัธยมศึกษาตอนปลาย เนื่องจากเป็นการวิเคราะห์ที่สะดวกและเส้นกราฟที่เสนอก็เข้าใจง่าย นอกจากนั้นยังทำให้สามารถใช้เป็นเครื่องมือวินิจฉัยข้อบกพร่องของนักเรียนเพื่อเป็นข้อมูลในการสอนซ่อมเสริมนักเรียน ปรับปรุง วิธีการสอนของครูและปรับปรุงหลักสูตรให้เหมาะสมและสอดคล้องกับความต้องการ

2. **ทฤษฎีเส้นโค้งเอส-พี (S - P Curve Theory)**

เป็นทฤษฎีที่แสดงความสัมพันธ์ของเส้นโค้งเอส (S - Curve) และเส้นโค้งพี (P - Curve) โดยการจัดเรียงข้อมูลหรือแบบแผนการตอบข้อสอบของนักเรียนเป็นรายบุคคลมากำหนดเป็นเมตริกซ์ (Matrix) ว่าข้อใดนักเรียนแต่ละคนตอบถูกตอบผิด และได้คะแนนรวมเท่าไร โดยกำหนดให้แนวนอน (Row) เกิดจากการเรียงลำดับข้อสอบจากข้อที่ง่ายไปหาข้อที่ยากโดยเรียงจากซ้ายไปขวา ซึ่งข้อสอบที่ง่าย หมายถึง ข้อสอบที่นักเรียนทำถูกจำนวนมากจากจำนวนผู้สอบทั้งหมด

ถ้าให้คะแนนการสอบเป็นระบบ 0,1 โดยตอบถูกให้ 1 คะแนน และตอบผิดให้ 0 คะแนน นำผลที่ได้จากการสอบไปแสดงในภาพประกอบที่ 2 สมมุติให้นักเรียน 18 คน ข้อสอบมี 5 ข้อ นำผลจากการสอบมาจัดเป็นเมตริกซ์ (Matrix) ดังนี้ (Harnisch and Linn 1982: 7)

ภาพประกอบ 2 แผนภูมิเอส-พี จำนวนนักเรียน 18 คน ข้อสอบจำนวน 5 ข้อ ดัชนีชี้เตือนของซาโต้ และดัชนีชี้เตือนที่ปรับแก้ของซาโต้

ผู้สอบ i	ข้อสอบ j 1	2	3	4	5	คะแนนรวมของผู้สอบ $n_{i.}$	ดัชนีของซาโต้ C_i	ดัชนีที่ปรับแก้ของซาโต้ C_i^*
1	1	1	1	1	0	4	.00	.00
2	1	1	1	0	1	4	.65	.33
3	1	1	1	0	0	3	.00	.00
4	1	1	0	1	0	3	.16	.08
5	1	1	0	0	1	3	.65	.31
6	1	0	1	0	1	3	1.13	.54
7	1	1	0	0	0	2	.00	.00
8	1	1	0	0	0	2	.00	.00
9	1	0	1	0	0	2	.44	.23
10	1	0	0	1	0	2	.59	.31
11	0	1	1	0	0	2	.74	.39
12	0	1	0	1	0	2	.88	.47
13	1	0	0	0	0	1	.00	.00
14	1	0	0	0	0	1	.00	.00
15	0	1	0	0	0	1	.45	.22
16	0	0	1	0	0	1	1.14	.56
17	0	0	0	1	0	1	1.36	.67
18	0	0	0	1	0	1	1.36	.67
จำนวนผู้ตอบถูก $n_{.j}$	12	10	7	6	3			
ในแต่ละข้อ								
Sato's Caution								
Index C_j	.13	.28	.42	.95	.21			
Modified Caution								
Index C^*_j	.14	.21	.50	.13				

เส้นโค้งเอส (S - Curve : Student Curve) เป็นเส้นที่แสดงถึงการตอบข้อสอบของนักเรียน ตามคะแนนรวมที่นักเรียนแต่ละคนได้ เช่น ภาพประกอบ 2 นักเรียนคนที่ 1 และ 2 ได้คะแนนรวม 4 คะแนน การลากเส้นโค้งเอส จะลากตามจำนวนคะแนนรวมของข้อสอบที่เรียงตามลำดับความยากแล้ว จากข้อสอบข้อที่มีความยากเป็นอันดับ 4 ลากลงในแนวดิ่ง 2 แถวตามจำนวนนักเรียนที่ได้คะแนนรวม 4 คะแนน ซึ่งมีจำนวน 2 คน นักเรียนคนที่ 3,4,5 และ 6 ได้คะแนนรวม 3 คะแนน เส้นโค้งเอสจึงถูกลากจากข้อสอบที่มีความยากเป็นอันดับ 3 หรือคอลัมภ์ที่ 3 และแถวนอนที่ 3 ลงมาในแนวดิ่งตามจำนวนนักเรียนที่ได้คะแนนเท่ากัน ลากเช่นนี้ต่อเนื่องกันไป จากภาพประกอบ เส้นโค้งเอสจะลากด้วยเส้นทึบ (Solid line) และถ้าเป็นไปตามหลักที่ว่า คนที่มีความ สามารถเท่ากันจะมีแบบแผนการตอบข้อสอบเหมือนกันหรือใกล้เคียงกันข้อต่อข้อ และถ้าการตอบข้อสอบของนักเรียนเป็นไปตามทฤษฎีของกัตแมน คือ ทางซ้ายของเส้นโค้งเอส นักเรียนจะตอบถูกทุกข้อ จะได้คะแนน 1 ในแต่ละข้อ ส่วนทางขวาของเส้นโค้งเอส จะได้คะแนน 0 คะแนนทุกข้อเช่นกัน เส้นโค้งเอสลักษณะนี้เราเรียกว่าเส้นโค้งเอสอุดมคติ (Ideal) ดังภาพประกอบ 3.1 (Tatsuoka and Linn 1983: 84-85)

เส้นโค้งพี (P-Curve : Problem Curve) เป็นเส้นที่แสดงถึงจำนวนนักเรียนที่ตอบข้อสอบแต่ละข้อได้ถูกต้องเช่นภาพประกอบ 2 ข้อสอบข้อที่ 1 จำนวนนักเรียนที่ตอบข้อสอบถูกต้องมีจำนวน 12 คน ข้อที่ 2 ตอบถูกจำนวน 10 คน ดังนั้นเส้นโค้งพี จะลากจากนักเรียนคนที่ 12 (คะแนนรวมเป็นอันดับที่ 12) ในข้อที่ 1 และลากจากนักเรียนคนที่ 10 ในข้อที่ 2 ลากต่อเนื่องกันไป จากภาพประกอบ 2 เส้นโค้งที่จะเป็นเส้นประ (Dotted line) และถ้าเป็นไปตามหลักที่ว่า ข้อสอบข้อที่ง่าย นักเรียนควรตอบถูกเป็นจำนวนมาก และข้อสอบข้อที่ยากนักเรียนจะตอบถูกเป็นจำนวนน้อยคน ดังนั้นทางซ้ายของเส้นโค้งพี นักเรียนจะตอบข้อสอบถูกทุกข้อได้คะแนน 1 คะแนนทุกข้อ ทางขวาของเส้นโค้งพี นักเรียนจะตอบข้อสอบผิดทุกข้อได้คะแนน 0 คะแนนทุกข้อเช่นกัน เส้นโค้งพีลักษณะนี้เรียกว่า เส้นโค้งพีอุดมคติ (Ideal) ดังภาพประกอบ 3.2 (Tatsuoka and Linn 1983: 84-85)

ภาพประกอบ 3 แสดงเส้นโค้งเอสอุดมคติ และเส้นโค้งพีอุดมคติ

Subject i	Item j										$y_{i.}$
	1	2	3	4	5	6	7	8	9	10	
1	1	1	1	1	1	1	1	1	1	1	10
2	1	1	1	1	1	1	1	1	1	1	10
3	1	1	1	1	1	1	1	1	1	1	10
4	1	1	1	1	1	1	1	1	1	1	10
5	1	1	1	1	1	1	1	1	1	0	9
6	1	1	1	1	1	1	1	0	0	0	7
7	1	1	1	1	1	1	0	0	0	0	6
8	1	1	1	1	1	1	0	0	0	0	6
9	1	1	1	1	0	0	0	0	0	0	4
10	1	1	1	0	0	0	0	0	0	0	3
11	1	1	0	0	0	0	0	0	0	0	2
12	1	0	0	0	0	0	0	0	0	0	1
13	1	0	0	0	0	0	0	0	0	0	1
14	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0
$y_{.j}$	13	11	10	9	8	6	6	5	5	4	79

3.1 แสดงเส้นโค้งเอสอุดมคติ (Perfect S-Curve)

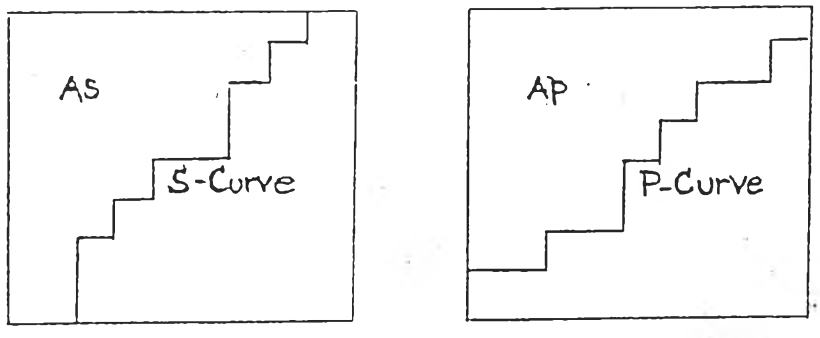
Subject i	Item j										$y_{i.}$
	1	2	3	4	5	6	7	8	9	10	
1	1	1	1	1	1	1	1	1	1	1	10
2	1	1	1	1	1	1	1	1	1	0	9
3	1	1	1	1	1	1	1	1	0	0	8
4	1	1	1	1	1	1	0	0	0	0	6
5	1	1	1	1	1	1	0	0	0	0	6
6	1	1	1	1	1	1	0	0	0	0	6
7	1	1	1	1	1	0	0	0	0	0	5
8	1	1	1	1	1	0	0	0	0	0	5
9	1	1	1	1	1	0	0	0	0	0	5
10	1	1	1	1	1	0	0	0	0	0	5
11	1	1	1	1	0	0	0	0	0	0	4
12	1	1	1	1	0	0	0	0	0	0	4
13	1	1	1	0	0	0	0	0	0	0	3
14	1	1	0	0	0	0	0	0	0	0	2
15	1	0	0	0	0	0	0	0	0	0	1
$y_{.j}$	15	14	13	12	10	6	3	3	2	1	79

3.2 แสดงเส้นโค้งพีอุดมคติ (Perfect P-Curve)

i17306292

2.1 คุณสมบัติของเส้นโค้ง เอส - พี (Properties of S-P Curve)(Sato 1980: 5)

2.1.1 ขนาดพื้นที่ของเส้นโค้งเอส (AS-Curve) ซึ่งได้แก่พื้นที่ทางซ้ายมือของเส้นโค้งเอส จะมีขนาดพื้นที่เท่ากับพื้นที่ของเส้นโค้งพี (AP-Curve) ซึ่งได้แก่พื้นที่ทางซ้ายมือของเส้นโค้งพีทั้งหมด ดังภาพ

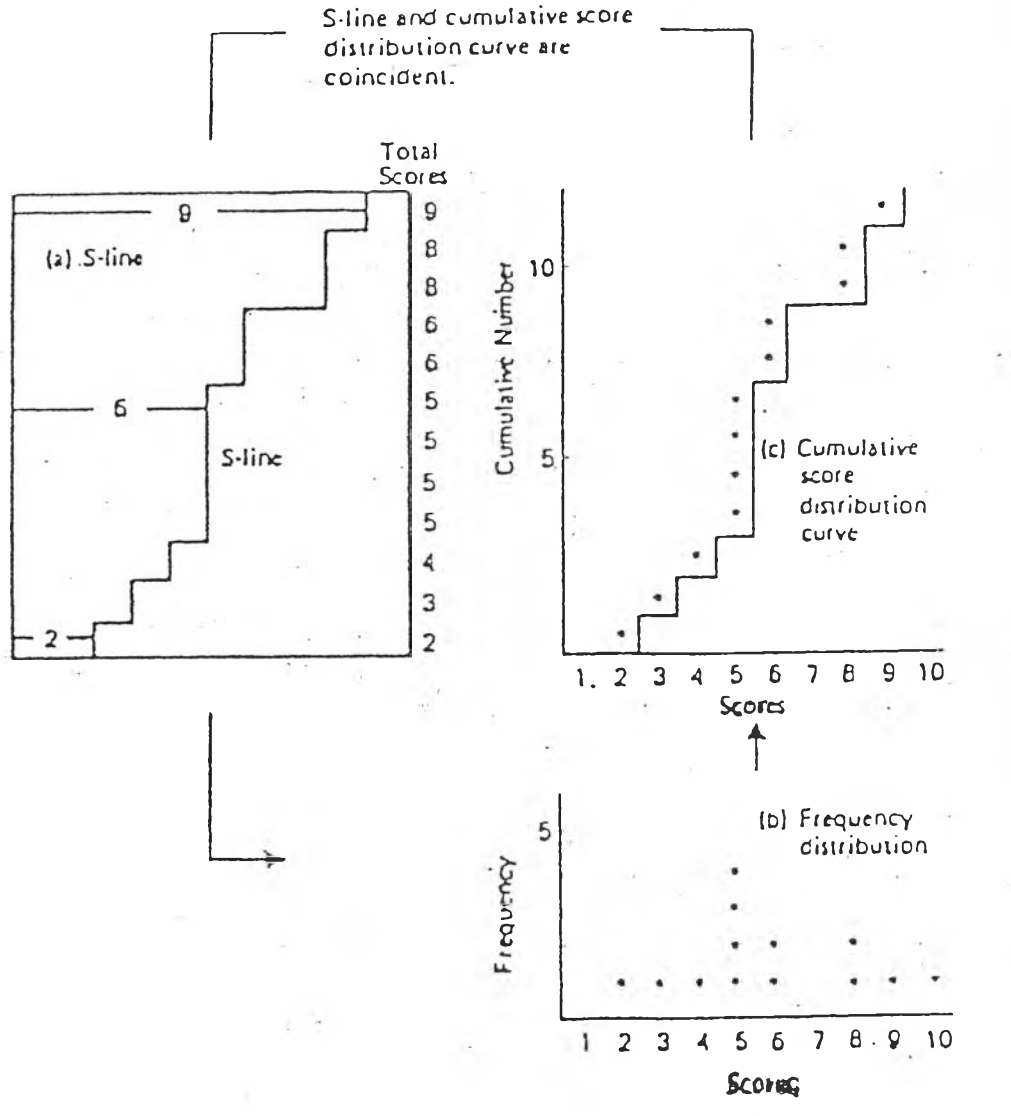


AS = AP

2.1.2 เส้นโค้งเอส เป็นเส้นที่แสดงถึงการแจกแจงความถี่สะสมของคะแนนการสอบส่วนเส้นโค้งพี (P-Curve) เป็นเส้นที่แสดงถึงจำนวนข้อสอบ (ภาพประกอบ 4)

2.1.3 ถ้าแบบทดสอบ และกลุ่มนักเรียนที่สอบ ในแผนภูมิเอส-พี มีลักษณะเป็นเอกพันธ์ (Homogeneous) ดังนั้นเส้นโค้งเอสและเส้นโค้งพีจะทับกันสนิทหรือเป็นเส้นโค้งเดียวกัน

ภาพประกอบ 4 การเปรียบเทียบเส้นโค้งเอสกับเส้นโค้งการแจกแจงความถี่สะสมของคะแนนการสอบ (Cummulative score distribution curve)



จากภาพประกอบ 4 แสดงให้เห็นว่าเส้นโค้งเอส (S-Curve) กับเส้นโค้งการแจกแจงความถี่สะสมของคะแนนการสอบของนักเรียนจะเป็นเส้นโค้งเดียวกันดังภาพประกอบ เส้นทั้งสองจะทับกันสนิท (Sato and Kurata 1977: 64)

2.2 การสร้างแผนภูมิเอส-พี (S-P Chart)

การสร้างแผนภูมิเอส-พี มีลำดับขั้นในการสร้าง สรุปได้ดังนี้

1. นำแบบทดสอบไปสอบนักเรียน และให้คะแนนระบบ 1,0 ให้คะแนนเป็นรายข้อ ตอบถูกได้ 1 คะแนน ถ้าตอบผิดได้ 0 คะแนน และรวมคะแนนเป็นรายบุคคล

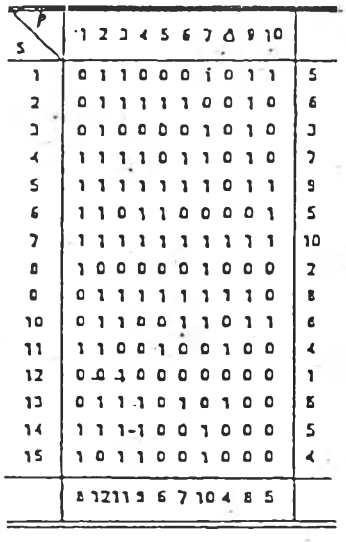
2. นำคะแนนการสอบของนักเรียน มาเรียงเป็นตารางเมตริกซ์ (Matrix) ให้แนวนอน (Row) แทนนักเรียนแต่ละคน (S_i) ซึ่งจะมีรายละเอียดการตอบข้อสอบเป็นรายข้อและคะแนนรวมแต่ละคน ส่วนในแนวดิ่ง (Column) แทนข้อสอบแต่ละข้อ (P_j) เรียงจากข้อ 1 ถึงข้อสุดท้ายจากซ้ายไปขวา ดูภาพประกอบ 5.1

3. จัดเรียงลำดับคะแนนรวมของนักเรียนโดยเรียงจากคนที่ได้คะแนนมากไปหาน้อยในตารางเมตริกซ์ หรือเรียงจากข้างบนลงข้างล่าง ดูภาพประกอบ 5.2

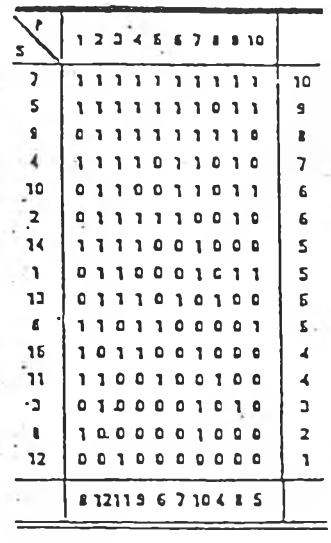
4. จัดเรียงข้อสอบในเมตริกซ์จากง่ายไปยาก หรือจากซ้ายไปขวาในตารางข้อสอบง่าย หมายถึง ข้อสอบข้อที่มีนักเรียนตอบถูกจำนวนมาก ดูภาพประกอบ 5.3

5. การลากเส้นโค้งเอส (S-Curve) และเส้นโค้งพี (P-Curve) โดยลากต่อเนื่องกันไปเป็นบันได เส้นโค้งเอส (S-Curve) จะลากตามคะแนนรวมของนักเรียนแต่ละคนด้วยเส้นทึบ (Solid line) เส้นโค้งพี (P-Curve) จะลากตามจำนวนนักเรียนที่ตอบถูกในข้อสอบแต่ละข้อนั่นคือการลากเส้นโค้งพี การลากจะต้องพิจารณาลากตามค่าจำนวนความถี่ของการตอบถูกข้อสอบแต่ละข้อ ดูภาพประกอบ 3.2 หรือ 5.4

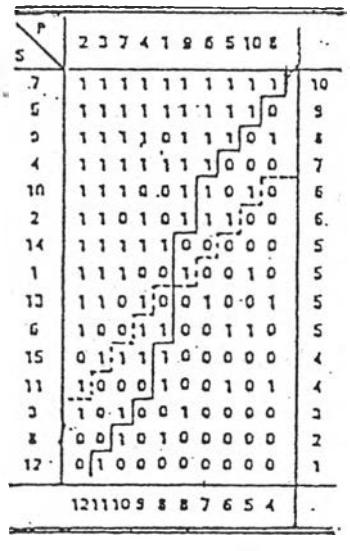
ภาพประกอบ 5 แสดงลำดับขั้นการสร้างแผนภูมิเอส-พี (S-P table) (Sato 1983 : 3)



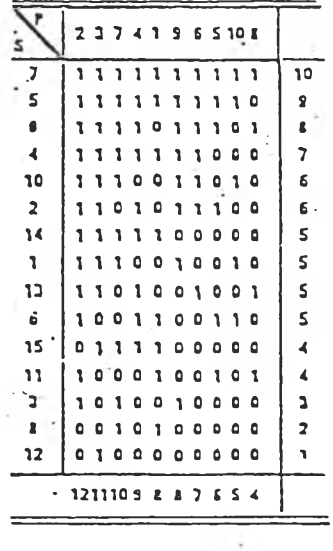
5.1 การเรียงข้อสอบตามลำดับข้อและเรียงนักเรียนตามเลขที่



5.2 เรียงคะแนนรวมจากคนที่ได้คะแนนมาก ไปหาคนที่ได้คะแนนน้อย



5.4 แผนภูมิเอส-พี

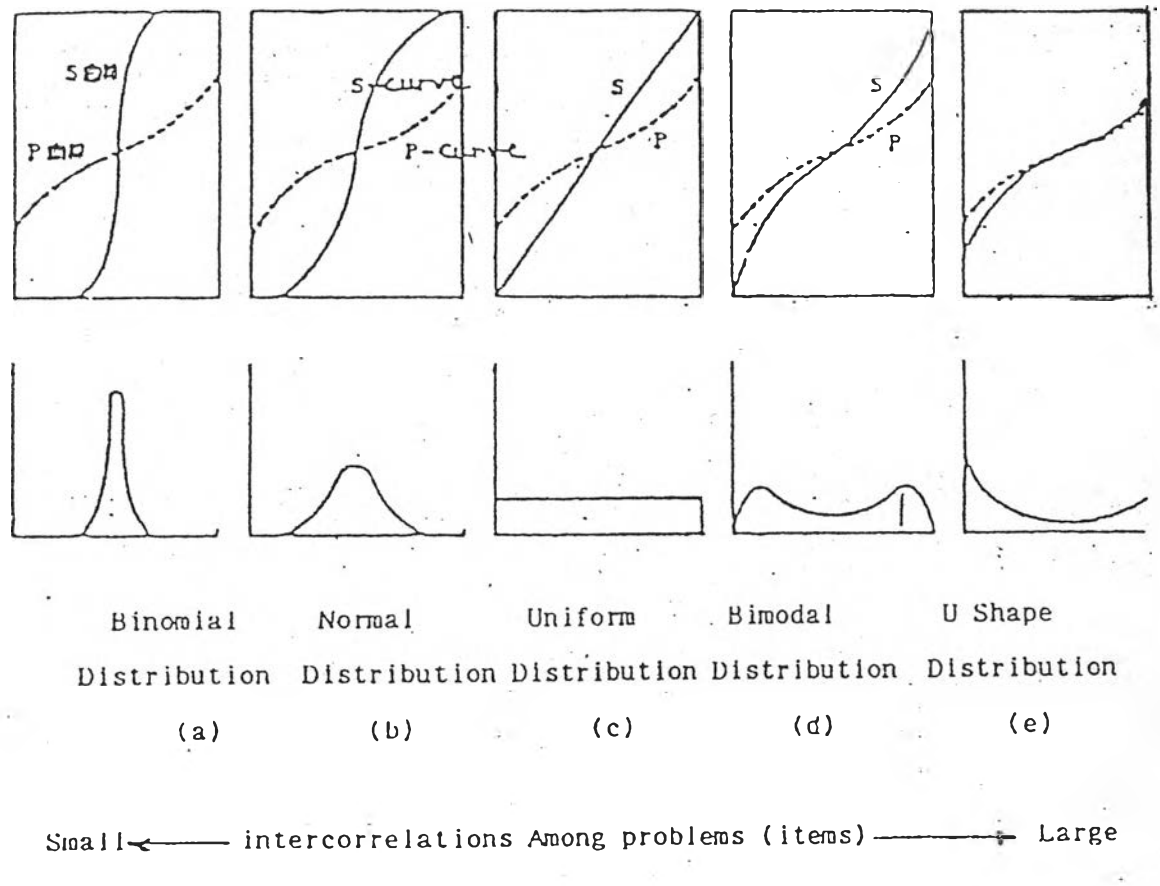


5.3 การเรียงข้อสอบจากข้อง่ายไปหายาก

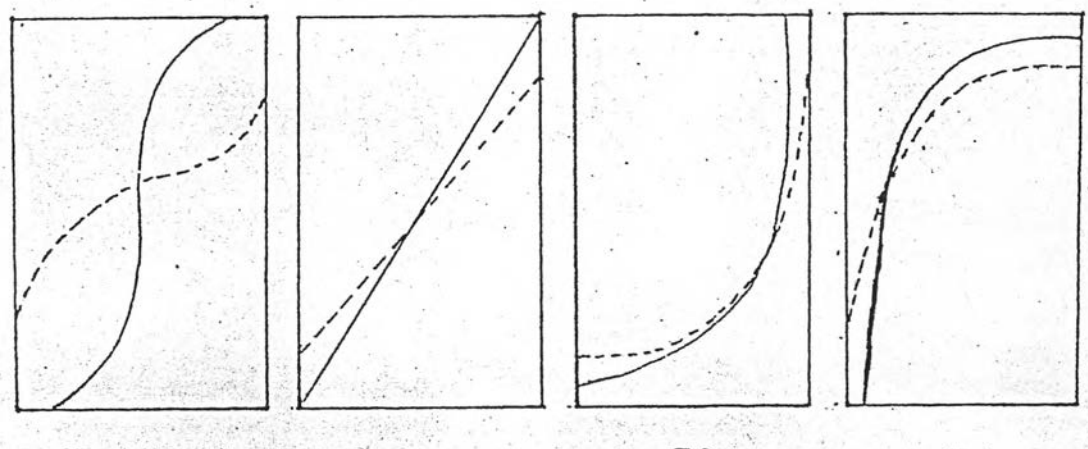
2.3 ความสัมพันธ์ระหว่างลักษณะของเส้นโค้งเอส (S - Curve) และความสัมพันธ์ภายในระหว่างข้อสอบ

เนื่องจากรูปร่างลักษณะของเส้นโค้งเอส (S - Curve) จะขึ้นอยู่กับลักษณะของกลุ่มข้อสอบ (Set of Problems) ทั้งนี้การจะดูลักษณะของกลุ่มผู้สอบว่ามีการแจกแจงเป็นอย่างไร เราก็ต้องพิจารณาความสัมพันธ์ระหว่างรูปร่างลักษณะของเส้นโค้งและระบบของความสัมพันธ์ภายในระหว่างข้อสอบหรือคำถาม ความสัมพันธ์นี้จะมีหลายลักษณะ ดังรูปภาพประกอบที่ 6 (Sato 1980: 9)

ภาพประกอบ 6 แสดงรูปร่างลักษณะของเส้นโค้งเอส (S-Curve) และเส้นโค้งพี (P-Curve) และการแจกแจงความถี่ของนักเรียนจากเส้นโค้งเอส



ภาพประกอบ 7 รูปแบบความสัมพันธ์ของเส้นโค้งเอส และเส้นโค้งพี



จากภาพประกอบ จะแสดงถึงความสัมพันธ์ของเส้นโค้งเอส และเส้นโค้งพีในลักษณะที่แตกต่างกัน ซึ่งแต่ละรูปแบบจะมีความหมายที่แตกต่างกันดังนี้

7.1 ลักษณะเช่นนี้ การแจกแจงของคะแนนการสอบจะมีลักษณะการแจกแจงเป็นโค้งปกติ (Normal Curve) โดยพิจารณาจากเส้นโค้งเอส (S - Curve) และลักษณะของข้อสอบจะมีระดับความยากประมาณ 0.50 ซึ่งเป็นระดับความยากของข้อสอบที่ดีมาก

7.2 ความสัมพันธ์ของเส้นโค้งเอส และเส้นโค้งพี เกือบจะมีค่าเท่ากับ 1.00 เกือบจะเป็นเส้นโค้งเดียวกัน (Highly Homogeneous) นั่นคือ แบบทดสอบกับนักเรียนมีความเหมาะสมกันการแจกแจงความถี่ของคะแนนการสอบจะเป็นรูปสี่เหลี่ยมผืนผ้า (Uniform distributions)

7.3 ความสัมพันธ์ของเส้นโค้งเอส เส้นโค้งพีในลักษณะเช่นนี้ จะแสดงให้เห็นว่าข้อสอบค่อนข้างง่าย การแจกแจงความถี่จะมีลักษณะไปทางลบ (High Negative Skewness) แบบทดสอบลักษณะเช่นนี้ ได้แก่ แบบทดสอบ Post-test แบบทดสอบวัดถุประสงค์หรืออิงเกณฑ์

7.4 ความสัมพันธ์ลักษณะนี้แสดงให้เห็นว่า ข้อสอบ แบบทดสอบฉบับนั้นค่อนข้างยาก การแจกแจงความถี่ของคะแนนนักเรียนจะเป็นลักษณะเบ้ทางบวก ที่มีความเบ้สูงมาก (High Positive Skewness) แบบทดสอบเช่นนี้ ได้แก่ แบบทดสอบ Pre-test หรือแบบทดสอบที่ผู้สอบยังไม่เคยได้เรียนมาก่อน เป็นต้น

2.4 **แบบแผนการตอบข้อสอบที่บกพร่อง**

แผนภูมิเอส - พี (S - P Chart) นอกจากจะบ่งบอกถึงลักษณะของแบบสอบและลักษณะของผู้สอบแล้ว แผนภูมิดังกล่าวยังมีประสิทธิภาพมากในการบ่งบอกถึงความบกพร่องในการตอบข้อสอบของนักเรียนด้วย โดยการพิจารณาตามแนวทางดังนี้

จากการให้คะแนนการสอบของนักเรียน 1 คะแนนถ้าตอบถูก และ 0 คะแนนถ้าตอบผิด ดังนั้นในแผนภูมิเอส - พี จะบ่งบอกถึงรูปแบบการตอบของนักเรียนเป็นรายบุคคลในแนวนอน (Row) ของตารางเมตริกซ์ ถ้าหากการตอบข้อสอบของนักเรียนไม่มีความบกพร่องจะไม่มี 0 ทางด้านซ้ายของเส้นโค้งเอส ซึ่งแสดงว่านักเรียนคนนั้นทำข้อสอบข้อง่ายไม่ผิดเลย เช่น ภาพประกอบ 5.4 นักเรียนคนที่ 4 และ 14 ซึ่งไม่มี 0 ในด้านซ้ายของเส้นโค้งเอส (S -Curve) ซึ่งถือว่าเป็นรูปแบบการตอบข้อสอบที่มีความเหมาะสมที่สุดเป็นรูปแบบในอุดมคติ (Ideal) แต่ถ้าหากนักเรียนคนใดทำข้อสอบที่ง่ายผิด แต่ทำข้อสอบที่ยากถูก ดังภาพประกอบ 5.4 นักเรียนคนที่ 6 มีแบบแผนการตอบข้อสอบที่บกพร่อง

ในทำนองเดียวกัน ถ้าหากพิจารณาข้อสอบในแนวคิ่ง (Column) จะต้องพิจารณาเส้นโค้งพี (P -Curve) เส้นโค้งพีจะเป็นตัวบ่งบอกให้ทราบเกี่ยวกับข้อสอบว่ามีความเหมาะสมหรือไม่ ถ้าเป็นข้อสอบที่เป็นไปตามหลักการหรือข้อสอบในอุดมคติ จะไม่มี 0 เหนือเส้นโค้งพีของข้อสอบแต่ละข้อ นั่นคือ เหนือเส้นโค้งพีจะมีเลข 1 ทุกข้อ แต่ถ้าเหนือเส้นโค้งพีมีเลข 0 หรือใต้เส้นโค้งพี มีเลข 1 แสดงว่าเกิดความบกพร่องในข้อสอบข้อนั้น เช่น คนที่มีความสามารถสูงตอบผิดแต่คนมีความสามารถต่ำกว่าตอบข้อนั้นถูก อาจจะเป็นข้อสอบที่ไม่ดี ข้อสอบอาจจะมีลักษณะแตกต่างจากข้ออื่น ๆ หรือวัดองค์ประกอบอื่น เช่นข้อสอบข้อที่ 7 ในภาพประกอบ 5.4

ดังนั้นจะเห็นได้ว่าแผนภูมิเอส - พี (S - P Chart) เป็นวิธีการวิเคราะห์แบบแผนการตอบข้อสอบของนักเรียน เพื่อพิจารณาถึงความบกพร่องในแบบแผนการตอบของแต่ละคน และความไม่เหมาะสมของกลุ่มตัวอย่างของข้อสอบ

แบบแผนการตอบข้อสอบที่บกพร่องของนักเรียนจะเป็นลักษณะการตอบข้อสอบของนักเรียนแต่ละคน คือ ตอบข้อสอบข้อที่ง่ายผิด แต่ตอบข้อสอบข้อที่ยากบางข้อถูก ซึ่งมีสาเหตุหลายประการที่ทำให้นักเรียนบางคนตอบข้อสอบที่ยากถูก อาทิ การเดา หรือตอบข้อสอบที่ง่ายแต่ตอบผิด อาทิ ความสะเพร่า ความวิตกกังวล หรือ ความผิดปกติบางอย่างในขณะสอบอาจจะมีสาเหตุมาจากนักเรียนหรือสิ่งแวดล้อม

ภาพประกอบ 8 แสดงการวินิจฉัยความบกพร่องในด้านสมรรถภาพหรือด้านเนื้อหาของนักเรียน

Categorized S-P Chart

Student Number	Test Score (Raw)	(%)	Modified Caution Ind/Sgn		Problem Number j 136	9570	842
					Category 111	3333	222
10018	10	100	.00	A	111	1111	111
10009	9	90	.83	B	111	0111	111
10019	9	90	.17	A	110	1111	111
10020	9	90	.00	A	111	1111	110
10010	8	80	.50	B	110	0111	111
10011	8	80	.40	B	101	1111	011
10003	7	70	.00	A	111	1110	100
10012	7	70	.62	B	000	1111	111
10017	7	70	.08	A	111	1111	000
10001	6	60	.00	A	110	1100	110
10004	6	60	.53	B	111	0000	111
10002	5	50	.06	C	111	1100	000
10008	5	50	.63	D	101	0010	101
10005	4	40	.27	C	100	1001	100
10013	3	30	.62	D	001	1001	000
10014	3	30	.08	C	110	1000	000
10015	3	30	.69	D	000	1000	011
10006	2	20	.10	C	100	0100	000
10016	2	20	.30	D	010	1000	000
10007	1	10	.33	D	000	0100	000
Problem Total					111 520	1111 4300	11 109
Percent Correct by Category					6 2	5 9	5 0

*A, B, C, D เป็นสัญลักษณ์ที่แสดงการวินิจฉัยความบกพร่องในแบบแผนการตอบของนักเรียน รายละเอียดอยู่หน้า 54

2.5 การวินิจฉัยกลุ่มทักษะของนักเรียนจากแผนภูมิเอส - พี

การวินิจฉัยนักเรียนแต่ละคนว่ามีความสามารถสูงในทักษะ หรือสมรรถภาพใดสามารถที่จะประยุกต์ใช้ทฤษฎีเส้นโค้งเอส - พี (S - P Curve Theory) จากการพิจารณาแบบแผนการตอบของนักเรียนแต่ละคนมาวินิจฉัยข้อบกพร่องได้ ซึ่งเป็นการประยุกต์ใช้ทฤษฎีที่เหมาะสมมากสำหรับนักเรียนแต่ละชั้น สำหรับครูประจำชั้นเพื่อจะได้ทราบว่านักเรียนเด่น และบกพร่องในสมรรถภาพใดหรือเนื้อหาใด ซึ่งวิธีการสร้างแผนภูมิเอส - พี เหมือนกันกับการสร้างแผนภูมิเอส - พี ที่กล่าวมาข้างต้น แตกต่างกันตรงที่วินิจฉัยความบกพร่องเป็นสมรรถภาพหรือเนื้อหาวิธีการสร้างดังนี้

1. จัดประเภทของข้อเป็นการวัดสมรรถภาพแต่ละสมรรถภาพหรือทักษะประกอบด้วยข้อใดบ้าง หรือแต่ละเนื้อหาประกอบด้วยข้อสอบข้อใดบ้าง

2. จัดเรียงนักเรียนจากคนที่ได้คะแนนรวมมากไปหาน้อยเรียงลงมาจากข้างบนสู่ล่าง

3. จัดเรียงสมรรถภาพต่าง ๆ ตามที่จัดไว้ (ข้อ 1) การเรียงลำดับให้คิดร้อยละเฉลี่ยของความยากของข้อสอบแต่ละสมรรถภาพ

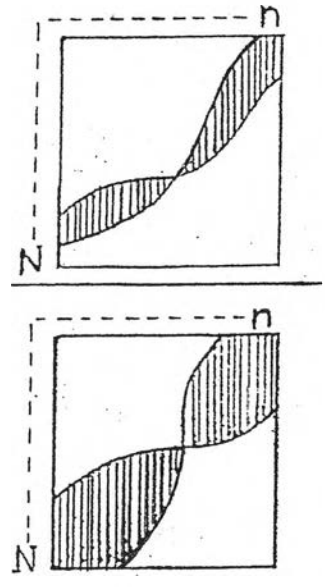
การพิจารณาว่านักเรียนแต่ละคนมีข้อบกพร่องสมรรถภาพใด แฮนิส (Harnisch 1987: 12) ได้ใช้เกณฑ์การวินิจฉัยว่านักเรียนบกพร่องหรือไม่ในสมรรถภาพนั้นหรือในเนื้อหานั้นโดยใช้เกณฑ์ 70 % ของเนื้อหาหรือสมรรถภาพนั้น ๆ คือ ถ้าตอบข้อสอบในแต่ละสมรรถภาพได้น้อยกว่า 70 เปอร์เซ็นต์ของเนื้อหานั้นหรือสมรรถภาพนั้น แสดงว่า นักเรียนมีข้อบกพร่องในสมรรถภาพนั้น ๆ (Harnisch 1983: 204) ดังภาพประกอบ 8

2.6 สัมประสิทธิ์ของความสอดคล้อง (Disparity Coefficient : D*)

จากความสัมพันธ์ของเส้นโค้งเอส (S -Curve) เส้นโค้งพี (P -Curve) จะทำให้ทราบว่าแบบทดสอบกับนักเรียนมีความเหมาะสมกันมากน้อยแค่ไหน ซึ่งตามหลักทฤษฎีเส้นโค้งเอสและเส้นโค้งพีต้องทับกัน ซึ่งแสดงว่านักเรียนและข้อสอบมีลักษณะเป็นเอกพันธ์ (Homogeneous) แต่ถ้าเส้นโค้งเอสและเส้นโค้งพีห่างกันมากเท่าไร ก็จะแสดงถึงความผิดปกติหรือมีความสอดคล้องกันน้อย ระหว่างแบบทดสอบชุดนั้นกับนักเรียนผู้สอบแบบทดสอบชุดนั้น นั่นคือ แบบทดสอบชุดนั้นไม่เหมาะที่จะนำมาวัดกับเด็กกลุ่มนั้น เพราะแบบแผนการตอบข้อสอบของนักเรียนแตกต่างกันมาก ซาโต้เรียกความผิดปกติดังกล่าวนี้ว่าสัมประสิทธิ์ความสอดคล้อง (Disparity Coefficient : D*)

การคำนวณหาค่าสัมประสิทธิ์ความสอดคล้องหาได้จากสูตร (Sato 1980: 15)

$$D^* = \frac{A(N,n,\bar{p})}{A_n(N,n,\bar{p})}$$

นั่นคือ $D^* =$ 

เมื่อ $A(N,n,\bar{P})$ คือ พื้นที่ระหว่างเส้นโค้งเอสกับเส้นโค้งพี S-P Chart เมื่อมีนักเรียน N คน ข้อสอบ n ข้อ และค่าเฉลี่ยของการตอบข้อสอบถูกของนักเรียนกลุ่มนั้น

และ $A_n(N,n,\bar{P})$ คือ พื้นที่ระหว่างเส้นโค้งเอสกับเส้นโค้งพีซึ่งมีลักษณะการแจกแจงแบบไบโนเมียล (Binomial Distribution) ด้วยค่าพารามิเตอร์ N,n และ $\bar{P}$

เพื่อสะดวกในการคำนวณหาค่าสัมประสิทธิ์ความสอดคล้อง (Disparity Coefficient) ชาโต้ได้เสนอสูตรที่ง่ายต่อการคำนวณ ดังนี้ (Sato 1980: 16)

$$D^* = \frac{C}{4Nn\bar{P}(1-\bar{P})Dn(M)}$$

เมื่อ C คือ จำนวนคะแนน (นับทั้ง 1,0) ในระหว่างเส้นโค้งเอสและเส้นโค้งพี

N คือ จำนวนนักเรียน

n คือ จำนวนข้อสอบ

$\bar{P}$ คือ ค่าเฉลี่ยของอัตราส่วนการตอบถูกในแต่ละข้อ คือ

เมื่อ $$\bar{P} = \frac{1}{n} \sum_{j=1}^{n} (Y_j/N)$$

หรือ $$\bar{P} = \frac{1}{Nn} \sum_{i=1}^{N} \sum_{j=1}^{n} X_{ij}$$

X_{ij} คือ คะแนนของนักเรียนคนที่ i ที่ตอบข้อสอบที่ j ถ้าตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน

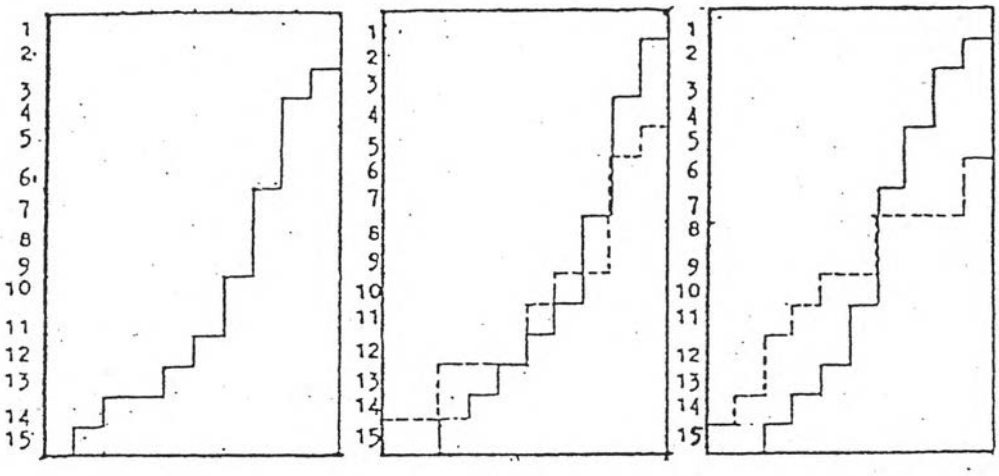
Y_j คือ จำนวนนักเรียนที่ตอบข้อสอบถูกในแต่ละข้อ

M คือ ค่าที่หาได้จากสูตร $M = \sqrt{Nn} + .5$ จะต้องทำให้ค่า Nn ออกให้เป็นเลขจำนวนเต็ม

Dn(M) คือ ค่าความเกี่ยวข้องของ M เป็นค่าที่หาค่าได้จากการเปิดตาราง

ค่าสัมประสิทธิ์ความสอดคล้องของข้อสอบกับกลุ่มนักเรียนผู้สอบมีค่าอยู่ระหว่าง 0 ถึง 1 ค่าสัมประสิทธิ์ความสอดคล้องจะมีค่ามากหรือน้อยขึ้นอยู่กับความเป็นทวิพันธ์ (Heterogeneous) ของเส้นโค้งเอส และเส้นโค้งพี ถ้าเป็นทวิพันธ์ค่าสัมประสิทธิ์ความสอดคล้อง (D*) จะมีค่ามาก แต่ถ้าค่าความสอดคล้องมีค่าเป็น 0 แสดงว่าเส้นโค้งเอสและเส้นโค้งพีจะมีลักษณะเป็นเอกพันธ์ (Homogeneous) เส้นทั้งสองจะทับกันหรือเป็นเส้นเดียวกัน เหมือนกับโมเดลของกัตแมนที่สมบูรณ์ ยิ่งค่า D* มากก็ยิ่งแสดงถึงความสอดคล้องกันน้อยระหว่างข้อสอบกับผู้สอบ ซาโต้กล่าวว่าค่าสัมประสิทธิ์ความสอดคล้องของข้อสอบกับผู้สอบจะมีค่าประมาณ 0.4 แต่ถ้าค่าเกิน 0.6 แสดงว่าข้อสอบกับผู้สอบกลุ่มนั้น ไม่เหมาะสมกัน (Sato 1980: 16) ดังนั้นการจะตัดสินว่าข้อสอบกับผู้สอบมีความสอดคล้องกันหรือไม่จะต้องใช้ค่า D* = .6 เป็นเกณฑ์ตัดสิน

ภาพประกอบ 9 ความสัมพันธ์ระหว่างเส้นโค้งเอส เส้นโค้งพี และส้มประสิทธิ์ของความสอดคล้องระหว่างแบบทดสอบกับกลุ่มนักเรียน



ก. $D^* = 0$ ข. $D^* = 0.36$ ค. $D^* = 0.63$

ความสัมพันธ์ของเส้นโค้งเอส เส้นโค้งพี จะแสดงให้เห็นถึงความเหมาะสมหรือความสอดคล้องระหว่างนักเรียนกับแบบทดสอบได้อย่างชัดเจน ดังนี้

ก. เส้นโค้งเอส และเส้นโค้งพีทับกัน ซึ่งจะทำให้ค่าส้มประสิทธิ์ความสอดคล้องมีค่าเป็นศูนย์ ($D^* = 0$) หมายความว่า แบบทดสอบฉบับนั้นมีความเหมาะสมที่จะวัดนักเรียนกลุ่มนั้นได้ดีที่สุด เป็นความสัมพันธ์กันแบบอุดมคติซึ่งเป็นไปได้ยากมากในการปฏิบัติจริง

ข. ความสัมพันธ์ของเส้นโค้งเอส เส้นโค้งพี มีค่าส้มประสิทธิ์ความสอดคล้องเท่ากับ 0.36 หมายความว่า แบบทดสอบฉบับนั้นกับนักเรียนที่สอบมีความเหมาะสมอยู่ในเกณฑ์ปกติ เพราะน้อยกว่า 0.6 หรือเกณฑ์แบบทดสอบฉบับนั้นนำไปใช้วัดนักเรียนกลุ่มนั้นได้ การแปลผลประเมินผล และตัดสินนักเรียนได้อย่างมั่นใจ

ค. ความสัมพันธ์ของเส้นโค้งเอส - พี มีค่าสัมประสิทธิ์ความสอดคล้องเกิน 0.6 ซึ่งหมายความว่า แบบทดสอบฉบับนั้นไม่เหมาะสมที่จะนำมาวัดนักเรียนกลุ่มนั้น การแปลผล หรือการตัดสินนักเรียนจากผลการสอบแบบทดสอบฉบับนั้นมีความคลาดเคลื่อนมาก เป็นการไม่ยุติธรรมแก่นักเรียน ซึ่งอาจจะเป็นผลมาจากความลำเอียงของข้อสอบก็เป็นได้

3. **ดัชนีชี้เตือนของซาโต้ (Sato's Caution Index)**

ดัชนีชี้เตือน (Caution Index) เป็นดัชนีที่จะบอกให้ทราบถึง ความบกพร่องในแบบแผนการตอบของนักเรียนเป็นรายบุคคลหรือกลุ่ม ทากายิโร ซาโต้ (Sato 1975) นักการศึกษาชาวญี่ปุ่น ได้เสนอดัชนีตัวหนึ่งซึ่งใช้ในการวิเคราะห์ผลการสอบ โดยมีแนวความคิดว่าคนที่มีความสามารถเท่ากัน จะต้องทำข้อสอบได้คะแนนรวมเท่ากัน และควรมีแบบแผนการตอบถูก หรือผิดในข้อที่เหมือนกันหรือใกล้เคียงกัน นักเรียนที่มีความสามารถต่ำหรือคะแนนรวมต่ำ ควรตอบข้อสอบผิดในข้อที่ยากจากแนวคิดดังกล่าว เขาได้จัดเรียงข้อมูลในลักษณะเมตริกซ์ของกัตแมน (Guttman 1941) และใช้ทฤษฎีเส้นโค้งเอส - พี (S-P Curve Theory) ในการพิจารณาความบกพร่องในแบบแผนการตอบของนักเรียนเป็นรายบุคคล และใช้ค่าสถิติตัวหนึ่ง ช่วยในการตัดสินความบกพร่องในแบบแผนการตอบข้อสอบเรียกว่า ดัชนีชี้เตือนของซาโต้ (Sato's Caution Index) จะเป็นค่าที่บ่งบอกถึงความบกพร่องของนักเรียนรายบุคคล ซึ่งคะแนนรวมจากการวัดผลสัมฤทธิ์บอกไม่ได้ (Harnisch and Linn 1982: 9)

ดัชนีของซาโต้มี 2 ชนิด คือ ดัชนีชี้เตือนนักเรียนของซาโต้ ในการศึกษาครั้งนี้ใช้สัญลักษณ์ คือ $C(S_i)$ และดัชนีชี้เตือนข้อสอบของซาโต้ ใช้สัญลักษณ์คือ $C(P_j)$

3.1 ดัชนีชี้เตือนนักเรียนของซาโต้ ใช้สัญลักษณ์ คือ $C(S_i)$ คำนวณได้จากค่าผลต่างของ 1 กับอัตราส่วนความแปรปรวนของเวคเตอร์ของการตอบข้อสอบของนักเรียน 2 ประประเภท คือ ความแปรปรวนร่วมของเวคเตอร์ของคะแนนที่ได้จากการตอบข้อสอบ (Observed Score) แต่ละข้อ กับจำนวนนักเรียนที่ตอบข้อสอบนั้นถูกเป็นตัวเศษ และความแปรปรวนร่วมของคะแนนของนักเรียนเมื่อข้อสอบเป็นไปตามเมตริกซ์ของกัตแมนที่สมบูรณ์ (Guttman 1941) กับจำนวนนักเรียนที่ตอบข้อสอบข้อนั้นถูกเป็นตัวส่วน เขียนเป็นสูตรได้ดังนี้ (Sato 1980: 18)

$$C(S_i) = 1 - \frac{COV(X_{ij}, Y_j)}{COV(U_{ij}, Y_j)}$$

เมื่อ X_{ij} คือ คะแนน (0,1) ของผู้สอบข้อที่ j คนที่ i

Y_j คือ ความถี่หรือจำนวนผู้สอบที่ตอบข้อสอบข้อที่ j ถูก

U_{ij} คือ คะแนนในอุดมคติเมื่อข้อสอบเป็นไปตามเมตริกซ์ของกัตแมน (U_{ij} จะเท่ากับ 1 ถ้า $j \leqslant X_{i.}$ และ $U_{ij} = 0$ เมื่อ $j > X_{i.}$)

$X_{i.}$ คือ คะแนนรวมของผู้สอบคนที่ i

j คือ ข้อสอบข้อที่ j ที่เรียงตามความยากของข้อกระทง

$C(S_i)$ คือ ดัชนีบ่งชี้ผู้สอบของชาโด้ของผู้สอบคนที่ i

$$COV(X_{ij}, Y_j) = \frac{1}{n} \sum_{j=1}^{n} (X_{ij} - \frac{X_{i.}}{n})(Y_j - \mu')$$

$$= \frac{1}{n} \left(\sum_{j=1}^{n} X_{ij}Y_j - X_{i.}\mu' \right)$$

$$COV(U_{ij},Y_j) = \frac{1}{n}\left[\sum_{j=1}^{x_{i.}} (1-\frac{X_{i.}}{n})(Y_j - \mu') + \sum_{j=X_{i.}+1}^{n} (0-\frac{X_{i.}}{n})(Y_j - \mu')\right]$$

$$= \frac{1}{n}\left(\sum_{j=1}^{X_{i.}} Y_j - X_{i.}\,\mu'\right)$$

เมื่อ μ' คือ คะแนนเฉลี่ยของผู้สอบที่ตอบข้อสอบถูก มีค่า

$$\mu' = \frac{1}{n}\sum_{j=1}^{n} Y_j$$

ดังนั้น

$$C(S_i) = 1 - \frac{\frac{1}{n}\left(\sum_{j=1}^{n} X_{ij}\,Y_j - X_{i.}\,\mu'\right)}{\frac{1}{n}\left(\sum_{j=1}^{X_{i.}} Y_j - X_{i.}\,\mu'\right)}$$

ในการคำนวณหาค่า $C(S_i)$ เพื่อความสะดวกใช้สูตรคำนวณคือ

$$C(S_i) = \frac{\sum_{j=1}^{X_{i.}} (1 - X_{ij})\,Y_j - \sum_{j=X_{i.}+1}^{n} X_{ij}Y_j}{\sum_{j=1}^{X_{i.}} Y_j - X_{i.}\,\mu'}$$

สูตรการคำนวณหาค่าดัชนีของผู้สอบรายบุคคล ได้มีการปรับปรุงเพื่อสะดวกในการคำนวณมากขึ้น ปัจจุบันยิ่งสะดวกมากขึ้นเนื่องจากสามารถคำนวณด้วยคอมพิวเตอร์ (Harnisch and Linn 1981: 135)

$$C(S_i) = \frac{\sum_{j=1}^{n_{i.}} (1 - U_{ij})\, n_{.j} - \sum_{j=n_{i.}+1}^{J} U_{ij}\, n_{.j}}{\sum_{j=1}^{n_{i.}} n_{.j} - n_{i.} \left[\frac{\sum_{j=1}^{J} n_{.j}}{J} \right]}$$

เมื่อ i คือ ผู้สอบคนที่ 1,2,3,..... I

j คือ ข้อสอบข้อที่ 1,2,3,..... J

U_{ij} คือ { 1 คะแนนถ้าผู้สอบคนที่ i ตอบคำถามข้อที่ j ได้ถูกต้อง
0 คะแนนถ้าผู้สอบคนที่ i ตอบคำถามข้อที่ j ผิด

$n_{i.}$ คือ คะแนนรวมของผู้สอบคนที่ i

$n_{.j}$ คือ คะแนนรวมของการตอบข้อสอบข้อที่ j ได้ถูกต้องของผู้สอบกลุ่มนั้น

$C(S_i)$ คือ ดัชนีชี้เตือนของซาโต้ของผู้สอบคนที่ i

ดัชนีชี้เตือนของซาโต้ (Sato's Caution Index) จะมีค่าตั้งแต่ -0.5 ถึง +1.5 ถ้า $C(S_i) = 0$ แสดงว่าสัดส่วนความแปรปรวนร่วมของแบบแผนการตอบสนองของผู้สอบเป็นไปตามทฤษฎีที่คาดหวังไว้ ดังนั้นความแปรปรวนร่วมของสิ่งที่สังเกตจะเท่ากับทฤษฎี

ค่าของดัชนีชี้เตือนผู้สอบของซาโต้ ยิ่งใกล้ 1.50 มากเท่าไร จะแสดงถึงความบกพร่องมากขึ้นทวีคูณ ซาโต้ได้กำหนดเกณฑ์ตัดสินความบกพร่องของดัชนีไว้ว่า ถ้าผู้สอบทำข้อสอบมีค่าดัชนีชี้เตือนมากกว่า 0.5 แสดงว่า ผู้สอบคนนั้นมีความบกพร่องในแบบแผนการตอบข้อสอบ (Sato 1980: 20)

ในปี 1981 แฮนิส และลินน์ ได้ปรับแก้สูตรการคำนวณของซาโต้ โดยการตัดข้อสอบบางข้อที่ผู้สอบทุกคนตอบถูกและตอบผิด ตัดผู้สอบคนที่ตอบข้อสอบถูกทุกข้อ และคนที่ตอบผิดทุกข้อออก และปรับแก้สูตรดัชนีชี้เดือนของซาโต้ ดังนี้ (Harnisch and Linn 1981: 135)

$$C^*(S_i) = \frac{\sum_{j=1}^{n_i.} (1 - U_{ij})\, n_{.j} - \sum_{j=n_i.+1}^{J} U_{ij}\, n_{.j}}{\sum_{j=1}^{n_i.} n_{.j} - \sum_{j=j+1-n_i.}^{J} n_{.j}}$$

เมื่อ i คือ ผู้สอบคนที่ 1,2,3.....I

j คือ ข้อสอบข้อที่ 1,2,3.....J

U_{ij} คือ คะแนนที่ผู้สอบคนที่ i ตอบข้อสอบข้อที่ j ถ้าตอบถูกได้ 1 ถ้าตอบผิดได้ 0

n_i คือ คะแนนรวมของผู้สอบคนที่ i

n_j คือ คะแนนรวมของการตอบข้อสอบข้อที่ j ถูกของผู้สอบทุกคน

$C^*(S_i)$ คือ ดัชนีชี้เดือนที่ปรับแก้ของซาโต้ ของผู้สอบคนที่ i

การแปลความหมายของดัชนีชี้เดือนที่ปรับแก้ของซาโต้ (Modified Caution Index: $C^*(S_i)$ มีค่าอยู่ระหว่าง 0 ถึง 1 การแปลความหมายของดัชนีจะแปลเช่นเดียวกับดัชนีชี้เดือนของซาโต้ (Caution Index) คือ ถ้าค่าของดัชนีมีค่าใกล้ 1.00 ความบกพร่องของผู้สอบผู้นั้นในการตอบข้อสอบก็จะยิ่งมากขึ้น แฮนิสได้กำหนดจุดตัด สำหรับเป็นเกณฑ์ตัดสินความบกพร่องในแบบแผนการตอบข้อสอบของดัชนีชี้เดือนที่ปรับแก้ของซาโต้ (Modified Caution Index) โดยกำหนดให้ 0.3 เป็นจุดตัด ถ้าดัชนีชี้เดือนที่ปรับแก้ของซาโต้ $C^*(S_i)$ มีค่ามากกว่า 0.3 แสดงว่าแบบแผนการตอบของผู้สอบผู้นั้นบกพร่อง (Harnisch 1987: 11)

การแปลความหมายความบกพร่องของดัชนีชี้เตือนของซาโต้

จากการที่ดัชนีชี้เตือนของซาโต้ และดัชนีชี้เตือนที่ปรับแก้ของซาโต้ ซึ่งทั้งสองดัชนีชี้เตือนจะมีเกณฑ์สำหรับบ่งชี้ถึงความบกพร่องในแบบแผนการตอบของนักเรียนว่ามีความบกพร่อง แต่ไม่ทราบว่าสาเหตุของความบกพร่องคืออะไร ดังนั้นซาโต้จึงได้เสนอตารางแสดงความสัมพันธ์ระหว่างผลการสอบของนักเรียน หรือคะแนนรวมกับดัชนีชี้เตือน $C(S_i)$ มากกว่าหรือเท่ากับ 0.5 ใช้เป็นเกณฑ์ในการตัดสินและวินิจฉัยความบกพร่องของนักเรียน ตารางความสัมพันธ์จะเป็นตาราง 2X2 ดังภาพประกอบ 10 (Harnisch and Staff nd : 16, Harnisch 1983: 202)

ภาพประกอบ 10 แสดงตารางการวินิจฉัยความบกพร่องของนักเรียน จากความสัมพันธ์ของดัชนีชี้เตือนของซาโต้ ดัชนีชี้เตือนที่ปรับแก้ของซาโต้ กับคะแนนผลการสอบ

	ไม่ บกพร่อง		Student Performance	Student Cell	Student Correct	Modified Caution Index
HIGH	A	B	100 %	A	>50 %	<=0.30 <=0.50 ของซาโต้
			50 %	B	>50 %	> 0.30 > 0.50 ของซาโต้
LOW	C	D	0 %	C	<=50 %	<=0.30 <=0.50 ของซาโต้
	0.30					
	0.50			D	<=50 %	> 0.30 > 0.50 ของซาโต้

Modified Caution Index

การวินิจฉัยความบกพร่องในแบบแผนการตอบของนักเรียนจากตาราง 2X2 จากภาพประกอบ 10 สามารถแปลผลได้ดังนี้

เครื่องหมาย A หมายถึง ช่องที่นักเรียนตอบข้อสอบถูกต้องมากกว่าร้อยละ 50 และค่าดัชนีบ่งชี้เตือน $C(S_i)$ น้อยกว่าหรือเท่ากับ .50

B หมายถึง ช่องที่นักเรียนตอบข้อสอบถูกต้องมากกว่าร้อยละ 50 และค่าดัชนี $C(S_i)$ มากกว่า .50

C หมายถึง ช่องที่นักเรียนตอบข้อสอบถูกน้อยกว่าหรือเท่ากับร้อยละ 50 และค่าดัชนีชี้เตือน $C(S_i)$ น้อยกว่าหรือเท่ากับ .50

D หมายถึง ช่องที่นักเรียนตอบข้อสอบถูกต้องน้อยกว่าหรือเท่ากับร้อยละ 50 และค่าดัชนีชี้เตือน $C(S_i)$ มากกว่า .50

การแปลความหมายความบกพร่อง (Sato 1975; citing Harnisch 1983: 198)

A หมายความว่า นักเรียนเป็นผู้มีความรู้จริง ทำข้อสอบด้วยความรู้จริงไม่มีความบกพร่องในแบบแผนการตอบข้อสอบ

B หมายความว่า นักเรียนเป็นผู้มีความรู้จริง แต่มีความบกพร่องในแบบแผนการตอบข้อสอบ เพราะตอบข้อสอบด้วยความสะเพร่า

C หมายความว่า นักเรียนยังไม่มีความรู้ในเรื่องที่เรียนเพียงพอ ต้องเรียนเนื้อหาเรื่องนั้นใหม่อีกครั้งหนึ่ง

D หมายความว่า นักเรียนยังไม่มีความรู้ในเรื่องที่เรียนเพียงพอและมีความบกพร่องในแบบแผนการตอบข้อสอบต้องเรียนเนื้อหานั้นใหม่อีกครั้งหนึ่ง

ในปี ค.ศ. 1981 แฮนิส และลินน์ (Harnisch and Linn 1981: 135) ได้ปรับปรุงสูตรดัชนีชี้เตือนของซาโต้ และได้ปรับปรุงเกณฑ์ความบกพร่องในแบบแผนการตอบข้อสอบของนักเรียน .5 เป็น .3 ดังนั้นการวินิจฉัยความบกพร่องในแบบแผนการตอบของนักเรียนจึงเปลี่ยนไปด้วย ความหมายของสัญลักษณ์ A , B , C และ D แปลเช่นเดียวกับของซาโต้

3.2 ดัชนีชี้เตือนข้อสอบของซาโต้ สัญลักษณ์ที่ใช้คือ $C(P_j)$ คำนวณได้จากสูตรดังนี้ (Sato 1980: 20)

$$C(P_j) = 1 - \frac{COV(X_{ij} , X_{i.})}{COV(V_{ij} , X_{i.})}$$

เมื่อ X_{ij} คือ คะแนนผู้สอบคนที่ i ข้อสอบข้อที่ j

V_{ij} คือ คะแนนข้อสอบข้อที่ j ของผู้สอบคนที่ i

เมื่อการตอบของผู้สอบเป็นไปตามเมตริกซ์ของกัตแมน

เช่น $V_{ij} = 1$ ถ้า $i =< Y_j$ และ

$V_{ij} = 0$ ถ้า $i > Y_j$

$X_{i.}$ คือ คะแนนรวมของผู้สอบคนที่ i

$C(P_j)$ คือ ดัชนีชี้เตือนข้อสอบของซาโต้ ข้อที่ j

i คือ ผู้สอบคนที่ i

ในการคำนวณหาค่าความบกพร่องของข้อสอบใช้สูตร (Sato 1980: 22)

$$C(P_j) = 1 - \frac{\frac{1}{N}\left(\sum_{i=1}^{N} X_{ij} X_{i.} - Y_j \mu\right)}{\frac{1}{N}\left(\sum_{i=1}^{Y_j} X_{i.} - Y_j \mu\right)}$$

เพื่อความสะดวกและง่ายในการคำนวณมากขึ้น ใช้สูตร

$$C(P_j) = \frac{\sum_{i=1}^{Y_j} X_{i.} - \sum_{i=1}^{N} X_{ij}X_{i.}}{\sum_{i=1}^{Y_j} X_{i.} - Y_j \mu}$$

$$= \frac{\sum_{i=1}^{Y_j} (1 - X_{ij}) X_{i.} - \sum_{i=Y_j+1}^{N} X_{ij}X_{i.}}{\sum_{i=1}^{Y_j} X_{i.} - Y_j \mu}$$

เมื่อ μ คือ ค่าคะแนนรวมเฉลี่ยของคะแนนการสอบ ซึ่งมีค่า

$$= \frac{1}{N} \left(\sum_{i=1}^{N} X_{i.} \right)$$

$X_{i.}$ คือ ค่าคะแนนรวมของผู้สอบคนที่ i

N คือ จำนวนผู้สอบกลุ่มที่สอบ

ในการคำนวณหาค่าดัชนีชี้เตือนข้อสอบของชาโต้เป็นรายข้อ เพื่อสะดวกในการคำนวณด้วยเครื่องคอมพิวเตอร์ สามารถคำนวณได้จากสูตรดังนี้

$$C(P_j) = \frac{\sum_{i=1}^{n.j} (1 - V_{ij})n_{i.} - \sum_{i=n.j+1}^{I} V_{ij}\, n_{i.}}{\sum_{i=1}^{n.j} n_{i.} - n.j \left(\sum_{i=1}^{I} \frac{n_{i.}}{I} \right)}$$

เมื่อ i คือ ผู้สอบคนที่ 1,2,3.....I

j คือ ข้อสอบข้อที่ 1,2,3.....J

V_{ij} คือ 1 ถ้าผู้สอบคนที่ i ตอบข้อสอบข้อที่ j ได้ถูกต้อง

0 ถ้าผู้สอบคนที่ i ตอบข้อสอบข้อที่ j ผิด

$n_{i.}$ คือ คะแนนรวมของผู้สอบคนที่ i

n_j คือ จำนวนผู้สอบที่ตอบข้อสอบข้อที่ j ถูกต้อง

และค่าดัชนีชี้เตือนข้อสอบที่ปรับแก้ของชาโต้ ซึ่งแฮนิส และลินน์ ได้ปรับปรุงจากสูตรดัชนีของชาโต้ (Modified Caution Index) สามารถคำนวณได้ดังนี้

$$C^*(P_j) = \frac{\sum_{i=1}^{n.j} (1 - V_{ij})\, n_{i.} - \sum_{i=n.j+1}^{I} V_{ij}\, n_{i.}}{\sum_{i=1}^{n.j} n_{i.} - \sum_{i=I+1-n.j}^{I} n_{i.}}$$

เมื่อ i คือ ผู้สอบคนที่ 1,2,3,...I

j คือ ข้อสอบข้อที่ 1,2,3,...J

$n_{i.}$ คือ คะแนนรวมของผู้สอบคนที่ i

$n_{.j}$ คือ จำนวนผู้สอบที่ตอบข้อสอบข้อที่ j ถูกต้อง

V_{ij} คือ คะแนนการตอบข้อสอบข้อที่ j ของผู้สอบคนที่ i
ถ้าตอบถูกได้ 1 ถ้าตอบผิดได้ 0

$C^*(P_j)$ คือ ดัชนีชี้เตือนข้อสอบที่ปรับแก้ของซาโต้

ค่าดัชนีชี้เตือนข้อสอบที่ปรับแก้ของซาโต้ หรือ $C^*(P_j)$ จะมีค่าอยู่ระหว่าง 0 กับ 1 เช่นเดียวกับดัชนีชี้เตือนนักเรียนที่ปรับแก้ของซาโต้หรือ $C^*(S_i)$ เกณฑ์ใช้ในการตัดสินว่าข้อสอบข้อใดมีความบกพร่องคือ 0.3 สำหรับดัชนีชี้เตือนที่ปรับแก้ของซาโต้ (Modified Caution Index : C*) และ 0.5 สำหรับดัชนีชี้เตือนของซาโต้ (Sato' Caution Index : C_i)

การพิจารณาความบกพร่องของข้อสอบ แฮนิสได้เสนอวิธีการพิจารณาความบกพร่องของข้อสอบเป็นรายข้อ โดยการแสดงความสัมพันธ์ของดัชนีชี้เตือนข้อสอบที่ปรับแก้ของซาโต้หรือ ($C^*(P_j)$) กับระดับความยากง่ายของข้อสอบ โดยเสนอเป็นตาราง 2X2 เช่นเดียวกับการพิจารณาการแปลผลความบกพร่องของนักเรียน ดังภาพประกอบ 10 (Harnisch 1983: 202 ,Harnisch and Staff nd: 16)

ภาพประกอบ 11 แสดงการวินิจฉัยความบกพร่องของข้อสอบจากความสัมพันธ์ของดัชนีชี้เตือนข้อสอบที่ปรับแก้ของซาโต้ หรือ C* (P_3) กับความยากง่ายของข้อสอบ

	ไม่บกพร่อง	บกพร่อง
HARD	W	X
EASY	Y	Z

0.3

Modified Caution Index

Item difficulty	Item Cell	Percent Correct	Modified Caution Index
0 %			
50 %			
	W	<=50 %	<=0.30
100 %	X	<=50 %	>0.30
	Y	> 50 %	<=0.30
	Z	> 50 %	> 0.30

ในการพิจารณาความบกพร่องจากตาราง 2 X 2 ในภาพประกอบ 11 แต่ละสัญลักษณ์ในตารางดังนี้ (Harnisch 1983: 200)

สัญลักษณ์ W หมายถึง นักเรียนตอบข้อสอบนั้นถูกต้อง มีจำนวนน้อยกว่าหรือเท่ากับ 50 เปอร์เซ็นต์ และค่าดัชนีประจำข้อน้อยกว่าหรือเท่ากับ .30

X หมายถึง นักเรียนตอบข้อสอบข้อนั้นถูกต้อง มีจำนวนน้อยกว่าหรือเท่ากับ 50 เปอร์เซ็นต์ และค่าดัชนีประจำข้อมากกว่า .30

Y หมายถึง นักเรียนตอบข้อสอบข้อนั้นถูกต้อง มีจำนวนมากกว่า 50 เปอร์เซ็นต์ (ง่าย) และค่าดัชนีประจำข้อสอบน้อยกว่าหรือเท่ากับ .30

Z หมายถึง นักเรียนตอบข้อสอบนั้นถูกต้อง มีจำนวนมากกว่า 50 เปอร์เซ็นต์ (ง่าย) และค่าดัชนีชี้ประจำข้อสอบมีค่ามากกว่า .30

การแปลความหมายของสัญลักษณ์ซึ่ง แฮนิส (Harnisch 1983: 203) ได้แปลความหมายสัญลักษณ์ต่าง ๆ ดังนี้

สัญลักษณ์	W	หมายถึง	เป็นข้อสอบที่ค่อนข้างยาก มีค่าอำนาจจำแนกสูง สามารถแยกนักเรียนได้
	X	หมายถึง	ข้อสอบที่มีการใช้คำไม่เป็นปรนัย หรือคำลวงเด่นกว่าตัวเลือก หรือเป็นข้อสอบที่วัดเนื้อหาแตกต่างจากเนื้อหาที่จะวัด
	Y	หมายถึง	ข้อสอบที่มีค่าอำนาจจำแนกนักเรียนต่ำและข้อความบางคำจะแนะคำตอบ
	Z	หมายถึง	ข้อสอบที่ต้องปรับปรุงแก้ไข หรือบางข้ออาจจะต้องตัดทิ้งเพราะง่ายเกินไป นักเรียนทุกคนตอบถูก

ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

การศึกษาเปรียบเทียบดัชนีชี้เดือนของชาโต้กับดัชนีอื่น ๆ

แฮนิส และลินน์ (Harnisch and Linn 1981: 141) ได้ศึกษาแบบแผนการตอบข้อสอบที่บกพร่องของนักเรียน ปี 1981 โดยใช้ข้อมูลจากการประเมินคุณภาพการศึกษา ในปี 1978 ได้ศึกษาเปรียบเทียบดัชนีบ่งชี้ถึงความบกพร่องในการตอบข้อสอบ โดยการวิเคราะห์แบบแผนการตอบข้อสอบของนักเรียนเป็นรายบุคคลไป ดัชนีที่ศึกษาใช้หลักการเดียวกันทั้ง 8 ดัชนี คือ ดัชนีของชาโต้ (C_i) ดัชนีที่ปรับแก้ของชาโต้ (C^*_i) ดัชนีความสัมพันธ์ระหว่างคะแนนแต่ละข้อกับคะแนนรวม (r_i) ดัชนีความสัมพันธ์ของคะแนนแต่ละคนกับคะแนนรวม (r'_i) ดัชนีการเห็นด้วยของเคนและเบรนัน (A) ดัชนีความไม่เห็นด้วยของเคนและเบรนัน (D) ดัชนีความสามารถที่สัมพันธ์กับการตอบผิด (θ) ดัชนียูของเวนเดอร์ไฟเออ (U_i) และดัชนีความสามารถของบุคคลของทาทซูโอกะ (NCI) เครื่องมือที่ใช้เป็นแบบสำรวจความก้าวหน้าทางการศึกษาวิชาคณิตศาสตร์ 40 ข้อ การอ่าน 28 ข้อ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนเกรด 4,8 และ 11 ซึ่งสุ่มจากโรงเรียน 110 โรงเรียน สุ่มนักเรียนแต่ละระดับชั้นในโรงเรียนต่าง ๆ มา 18 - 22 คน ได้นักเรียนกลุ่มตัวอย่างแต่ละระดับชั้น 2,100 คน 3 ระดับชั้น คือ ระดับประถมศึกษาหรือเกรด 4 มัธยมศึกษาตอนต้นหรือเกรด 8 และมัธยมศึกษาตอนปลายหรือ

เกรด 11 รวม 6,300 คน เครื่องมือที่ใช้นั้นนอกจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์และการอ่านแล้ว ยังประกอบด้วยแบบสำรวจข้อมูลพื้นฐานเกี่ยวกับตัวนักเรียนและตัวแปรอื่น ๆ ด้านอารมณ์ของนักเรียน ในการศึกษาความสัมพันธ์ของดัชนีทั้ง 8 ดัชนี ได้วิเคราะห์โดยใช้กลุ่มตัวอย่างนักเรียนระดับประถมศึกษาหรือเกรด 4 จำนวน 2,094 คน ผลการศึกษาพบว่าดัชนีที่ถูกคัดเลือกว่ามีประสิทธิภาพมากคือ ดัชนีที่ปรับแก้ของซาโต้ หรือ C^*_i

ทาทซูโอกะ และทาทซูโอกะ (Tatsuoka and Tatsuoka 1982: 143) ได้ศึกษาดัชนี Extended Caution Indices (ECI) ที่เป็นมาตราฐาน และการเปรียบเทียบอัตราการตรวจจับ (Detection Rates) โดยใช้ดัชนี ECI หลายตัว ซึ่งได้มาจากวิธีการที่ต่างกัน 2 วิธี คือ วิธีที่ทำให้เป็นมาตราฐาน กับวิธีที่ได้มาจากพื้นฐานทฤษฎี IRT ซึ่งค่าเฉลี่ย และความแปรปรวนของ ECI บางตัวที่เบี่ยงเบนและคุณสมบัติทางสถิติของดัชนีเหล่านี้นำมาเปรียบเทียบกันพบว่า ECI มาตราฐานนั้นเหมาะกับการกระจายแบบปกติเป็นอย่างดี และเปรียบเทียบอัตราการตรวจจับบนพื้นฐานทฤษฎี IRT ไม่พบความแตกต่างกัน

ทาทซูโอกะ และทาทซูโอกะ (Tatsuoka and Tatsuoka 1983: 226-227) ได้อภิปรายถึงจุดอ่อนของดัชนีความคล้อยตามกลุ่ม (Norm Coformity Index) และดัชนีชี้เตือน (Caution Index) ว่าการคำนวณดัชนี 2 ตัวนี้ ขึ้นอยู่กับการเรียงลำดับข้อตามความยากใน Norm Group เมื่อ Norm Group เปลี่ยนไปลำดับข้อตามความยากจะเปลี่ยนตามซึ่งมีผลกระทบต่อค่าดัชนีทั้งสองนี้ สำหรับรูปแบบการตอบสนองเดียวกันของผู้สอบคนนั้น ดัชนีดังกล่าวจึงไม่เหมาะที่จะใช้วินิจฉัยความรุนแรงของความผิดพลาดของผู้สอบ ทาทซูโอกะ และทาทซูโอกะ จึงได้เสนอดัชนีที่เป็นอิสระจากกลุ่มคือ Individual Consistency Index (ICI) ค่าดัชนีนี้จะขึ้นอยู่กับความยากของงานที่กำหนดโดยความสามารถของผู้สอบเอง แต่ ICI ก็มีข้อจำกัดอยู่ที่ต้องใช้แบบสอบคู่ขนานตั้งแต่ 2 ชุดขึ้นไป จึงจะสามารถคำนวณหาค่าดัชนีนี้ได้ซึ่งเป็นไปได้ยากในทางปฏิบัติ

รัดเนอร์ (Rudner 1983: 207-219) ได้ประเมินความตรงของดัชนีที่ยึดหลักการพิจารณาแบบแผนการตอบข้อสอบของนักเรียนที่พิจารณาการตอบถูกและผิดของนักเรียนและคะแนนรวม กับดัชนีที่ยึดหลักทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) เครื่องมือที่ใช้วัดคือ แบบทดสอบ 2 ฉบับ คือ แบบทดสอบความถนัดด้านภาษา 80 ข้อ ได้ปรับคะแนนการสอบโดยใช้หลักการของลอร์ด (Lord 1968) และแบบทดสอบวัดผลสัมฤทธิ์

วิชาชีววิทยาทั่วไป จำนวน 45 ข้อ ได้ปรับคะแนนนักเรียนในห้องเรียนโดยครู ตามหลักของบิจา ไวลส์ และคิงส์บุรี (Bijar Weiss and Kingsbary 1977) ดัชนีที่ใช้ในการศึกษาเปรียบเทียบครั้งนี้มี 9 ดัชนี ดัชนีที่ใช้หลักการของ เบิอร์นบอม (Birnbaum 1968) 3 วิธี คือ ดัชนีการใช้น้ำหนักคะแนนเบิอร์นบอม (W_3) ดัชนีการไม่ใช้น้ำหนักของคะแนน (U_3) และแบบแผนการตอบ Likelihood โดยใช้โมเดลของเบิอร์นบอม ดัชนีที่ใช้หลักการของ ราชส์โมเดล 2 วิธี (Rasch 1960) ดัชนีการให้น้ำหนักคะแนนราชส์โมเดล ที่เหมาะสมกับความแปรปรวน (MS) คือ W_i และดัชนีการไม่ให้น้ำหนักคะแนนของราชส์โมเดล (U_i) ดัชนีที่ใช้วิธีการสัมพันธ์ 2 วิธีคือ สหสัมพันธ์แบบพอยส์ไบซีเรียล (r_i) และดัชนีสหสัมพันธ์แบบไบซีเรียลของบุคคล (r_b) และดัชนีความต่อเนื่องของทาทซูโอกะ 2 วิธีคือ ดัชนีของทาทซูโอกะและทาทซูโอกะ (NCI) ดัชนีปรับแก้ของซาโต้ (C^*_i) ของแฮนิส และลินน์ กลุ่มตัวอย่างที่ใช้ ในการศึกษาแบบทดสอบฉบับที่ 1 จำนวน 2,800 คน แบบทดสอบฉบับที่ 2 จำนวน 2,600 คน ผลการศึกษาพบว่า ดัชนีความสัมพันธ์แบบไบซีเรียล (r_i) พอยส์ไบซีเรียล (r_b) ดัชนีทาทซูโอกะ คือ (NCI) ดัชนีที่ปรับแก้ของซาโต้ (C^*_i) และดัชนีการให้น้ำหนักคะแนนของราชส์ (W_i) มีความเหมาะสมกับค่าสถิติจากค่าสหสัมพันธ์ของดัชนีแต่ละดัชนีในแบบทดสอบ 2 ฉบับ พบว่า ดัชนีบางตัวมีความสอดคล้องและสัมพันธ์กับแบบทดสอบบางฉบับ แต่ไม่สอดคล้องหรือเหมาะสมกับแบบทดสอบบางฉบับ มีเพียงดัชนีทาทซูโอกะ และทาทซูโอกะ (NCI) ดัชนีที่ปรับแก้ของซาโต้ (C^*_i) เป็นค่าดัชนีที่มีความคงที่ไม่แปรเปลี่ยนไปตามแบบทดสอบ เป็นดัชนีที่มีความสอดคล้องและสัมพันธ์กับแบบทดสอบทั้ง 2 ฉบับ หรือทุกสถานการณ์ให้ค่าสถิติที่มีความคงที่กว่าดัชนีอื่น ๆ ซึ่งขัดแย้งกับการศึกษาของรัดเนอร์ (Rudner 1982 ceting Harnisch 1983: 194) ซึ่งได้ศึกษาเปรียบเทียบดัชนีชี้เตือนนักเรียนที่ปรับแก้ของซาโต้หรือ ($C^*(S_i)$) โดยหลักการพิจารณาความบกพร่องในแบบแผนการตอบข้อสอบกับดัชนีที่ใช้หลักการของทฤษฎีการตอบสนองของข้อสอบ (Item Response Thoery) มีประสิทธิภาพไม่แตกต่างกันกับดัชนีที่ใช้หลักการพิจารณาแบบแผนการตอบสนองของนักเรียนของข้อถูกข้อผิด แต่ดัชนีที่ใช้หลักการทฤษฎีการตอบสนองข้อสอบ มีข้อจำกัดที่ต้องใช้กลุ่มตัวอย่างเป็นจำนวนมาก ไม่เหมาะสมที่จะนำมาใช้ในห้องเรียนหนึ่ง ๆ สอดคล้องกับการศึกษาของทาทซูโอกะและลินน์ (Tatsuoka and Linn 1983: 81-96) ได้ศึกษาความสัมพันธ์ของดัชนีที่จะวินิจฉัยแบบแผนการตอบข้อสอบที่บกพร่องของนักเรียน 2 ทฤษฎี คือ ทฤษฎีเส้นโค้งเอส - พี (S-P Curve Theory) กับทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) โดยได้แสดงให้

เห็นถึงความสัมพันธ์ของข้อสอบ หรือมีลักษณะเช่นเดียวกันอย่างแน่นอนระหว่างเส้นโค้งของข้อสอบ (Item Response Curve : IRC) กับเส้นโค้งของนักเรียน (The Person Response Curve : PRC) และเส้นโค้งของแบบทดสอบ (Test Response Curve : TRC) กับเส้นโค้งของกลุ่มนักเรียนที่สอบ (Group Response Curve : GRC) ซึ่งแสดงให้เห็นถึงความสัมพันธ์ของทั้ง 2 ทฤษฎี แต่ข้อจำกัดอาจจะต่างกันเพียงเล็กน้อย

เบิอร์นบอม (Birnbaum 1985: 523-534) ได้ศึกษาเปรียบเทียบดัชนีบ่งชี้ความบกพร่องในแบบแผนการตอบข้อสอบนักเรียนที่พิจารณาข้อถูกข้อผิด คะแนนรวมซึ่งใช้หลักการของทฤษฎีการตอบสนองข้อสอบในการศึกษาได้แบ่งกลุ่มออกเป็น 3 กลุ่ม คือ กลุ่มนักเรียนที่ตั้งใจทำข้อสอบ (Coperative examines) กลุ่มที่ไม่ให้ความร่วมมือ (Uncoperative examines) และกลุ่มที่สุ่มทั่ว ๆ ไป เกณฑ์การตัดสินว่านักเรียนตั้งใจสอบหรือไม่ โดยการวิเคราะห์จากการเขียนชื่อบนกระดาษคำตอบข้อสอบของนักเรียน ถ้าเขียนชื่อแสดงว่า นักเรียนตั้งใจตอบ หรือให้ความร่วมมือ ถ้าไม่เขียนแสดงว่าไม่ตั้งใจสอบ เครื่องมือที่ใช้วัดคือ แบบทดสอบวัดทักษะในการอ่านวิชาภาษาอังกฤษและภาษาที่ 2 ของมหาวิทยาลัยเทลอะวิฟ เป็นข้อสอบคู่ขนาน ฉบับละ 20 ข้อ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนระดับเกรด 10 ใน 77 โรงเรียน รวมจำนวนกลุ่มตัวอย่าง 1,864 คน ดัชนีที่ใช้เปรียบเทียบทั้งหมด 9 ดัชนี คือ ดัชนีของทาทซูโอกะ ซึ่งยังไม่ได้ปรับความเป็นมาตราฐาน ได้แก่ ดัชนีบ่งชี้ที่ขยายออกไป (The Extended Caution Index : ECI) คือ ECI_1 , ECI_2 , ECI_4 และดัชนีที่ปรับขยายของทาทซูโอกะที่ได้มาตราฐาน คือ ECI_z , ECI_{2z} , ECI_{4z} ดัชนีของเลวิฟ และรูบิน (L_o) ดัชนีของเลวินและคราสโกว์ (L_z) และดัชนีที่เสนอโดยไรท์ (U_z) ผลการเปรียบเทียบความสัมพันธ์พบว่าดัชนีของเลวินและคราสโกว์ (L_z) มีความสัมพันธ์กับดัชนีอื่น ๆ สูงมากที่สุดและเป็นดัชนีที่มีความสัมพันธ์กับคะแนนรวมน้อยที่สุด และผลจากการเปรียบเทียบพบว่าแบบแผนการตอบข้อสอบแตกต่างกันหรือไม่ของกลุ่มตัวอย่างทั้ง 3 กลุ่ม กับดัชนีทั้ง 9 ดัชนี พบว่าดัชนีของไรท์ (U_z) เพียงดัชนีเดียวที่พบว่ากลุ่มที่ตั้งใจสอบ ไม่ตั้งใจสอบและกลุ่มที่สุ่มโดยทั่ว ๆ ไป ทั้ง 3 กลุ่ม มีแบบแผนการตอบไม่แตกต่างกัน

บูซี (Buxie 1986: 372) ได้ศึกษาผลของความยาก ความยาวของแบบสอบ และค่าการเดาที่มีผลต่อลักษณะการแจกแจงของดัชนีปรับแก้ของซาโต้ (MCI) และดัชนี Individual Consistency Index (ICI) ซึ่งแหล่งข้อมูลของข้อกระทงมาจาก 2 แหล่งคือ จากข้อมูลเชิงประจักษ์กับข้อมูลที่ได้จำลองขึ้นมา การศึกษาครั้งนี้ศึกษากับข้อกระทงที่มีค่าความยาก 3 ระดับ และความยาวของแบบสอบ 4 ระดับ ผลการศึกษาพบว่าค่าเฉลี่ยเลขคณิตของ MCI และ ICI มีค่าต่ำในการตอบแบบสอบที่จำนวนข้อกระทงน้อยที่สุด ดัชนี MCI บ่งชี้ผู้ตอบแบบทดสอบที่ผิดปกติได้น้อยเมื่อแบบสอบมีความยากมากและมีจำนวนข้อกระทงในระดับมากและปานกลาง ค่าของ ICI มีความคงที่สูงสุดเมื่อแบบสอบมีความยากปานกลางและมีความคงที่ต่ำสุดเมื่อแบบสอบมีความยากมาก ดัชนี MCI สามารถบ่งชี้ถึงหลักฐานของการเดาได้สูงสุดเมื่อแบบสอบมีความยากมาก แต่ ICI ไม่สามารถบ่งชี้ถึงหลักฐานของการเดา ค่าเฉลี่ยเลขคณิตและความแปรปรวนของ MCI มีค่าต่ำเมื่อแบบสอบมีความยากมากในความยาวของแบบสอบทุกระดับ ค่าของ ICI บ่งชี้การตอบที่ผิดปกติได้น้อยในแบบสอบที่มีความยากปานกลาง

ทอมสิก (Tomsic 1986: 2555) ได้ศึกษาความคงที่ของ Extended Caution Indices (ECI) ที่เป็นมาตราฐานสำหรับการสอบในโรงเรียนรัฐบาล ซึ่งการศึกษานี้สนใจ ECI ที่เป็นมาตราฐานและไม่มาตราฐานของ ECI2 และ ECI4 โดยใช้นักเรียนเกรด 3 และ 7 ในปี 1984 และ 1985 จำนวน 500 คน และ 700 คน ใช้แบบสอบคำศัพท์และแบบสอบความเข้าใจในทักษะพื้นฐาน พบว่าค่าเฉลี่ยของ ECI ที่เป็นมาตราฐานและไม่มาตราฐานมีค่าโดยประมาณใกล้เคียง 0 ในทุกแบบสอบย่อยของแต่ละระดับชั้น ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง Caution Indices ของสองระดับชั้นมีค่าอยู่ระหว่าง .05 ถึง .30 มีค่าสัมประสิทธิ์สหสัมพันธ์สูงในชั้นเด็กโต แต่ไม่มีความแตกต่างกัน ค่าสหสัมพันธ์ของคะแนนสอบทั้งชุดกับ ECI จะเป็นลบและมีค่าต่ำมาก

ทอมสิก (Tomsic 1987: 29) ได้ทำการศึกษาผลของข้อกระทงที่ไม่เหมาะสมต่อการแจกแจงของดัชนีชี้เตือนขยาย (Extended Caution Indices) ซึ่งการวิจัยนี้เพื่อตรวจสอบว่าการแจกแจงของ ECI ใกล้เคียงปกติหรือไม่ ถ้าข้อกระทงที่ไม่เหมาะสมถูกตัดออกจากการวิเคราะห์ โดยศึกษากับนักเรียนเกรด 3 จำนวน 700 คน และ เกรด 7 จำนวน 737 คน กับแบบสอบวัดความเข้าใจในทักษะเบื้องต้นพบว่า การคัดข้อกระทงที่ไม่เหมาะสมออกไม่ได้ช่วยให้การแจกแจงของ ECI เป็นปกติ

ในปี 1982 แฮนิส (Harnisch 1982: 194) ได้ศึกษาเปรียบเทียบดัชนีที่บ่งบอกถึงข้อบกพร่องในแบบแผนการตอบข้อสอบของนักเรียน ซึ่งเป็นดัชนีที่พัฒนาขึ้นโดยทาทซูโอกะ คือ ดัชนีความสามารถของบุคคล (NCI) กับดัชนีความสามารถของแต่ละบุคคล (ICI) กลุ่มตัวอย่างที่ใช้ได้แก่ นักเรียนระดับเกรด 3 , 6 และ เกรด 7 กับ 8 จำนวน 357,396 และ 643 ตามลำดับ เครื่องมือที่ใช้วัดเป็นแบบทดสอบวัดทักษะพื้นฐานวิชาคณิตศาสตร์ ผลการศึกษาพบว่า ดัชนีทั้งสองมีความสัมพันธ์ทางบวกสูงมาก และทั้งสามระดับชั้นคือ 3,6 และ 7 กับ 8 คะแนนรวมมีความสัมพันธ์กันสูงมากเช่นกัน

การศึกษาเกี่ยวกับการใช้ดัชนีชี้เตือนของซาโต้และทฤษฎีเส้นโค้งเอส - พี

แฮนิส (Harnisch 1983: 191-206) ได้ศึกษาแบบแผนการตอบข้อสอบของนักเรียน โดยใช้ทฤษฎีเส้นโค้งเอส - พี (S - P Curve Theory) กับดัชนีของซาโต้เป็นเกณฑ์ตัดสินข้อบกพร่องในการตอบแบบทดสอบของนักเรียนรายบุคคล วิเคราะห์ความบกพร่องของข้อสอบด้วยดัชนีของซาโต้ (C^*_i) เช่นกัน พิจารณาความเหมาะสมของนักเรียนกับแบบทดสอบ (D^*) ตลอดจนการแปลผลความบกพร่องโดยเสนอในรูปเส้นกราฟ (Profile) ของนักเรียนและข้อสอบโดยการศึกษาครั้งนี้ ได้ยกตัวอย่างเพื่อให้สามารถนำมาใช้ได้ ตัวอย่างที่ใช้บรรยายมีจำนวนนักเรียน 24 คน ระดับชั้นเกรด 4 และจำนวนข้อสอบ 44 ข้อ ได้เสนอในรูปของเส้นภาพโดยใช้ทฤษฎีเส้นโค้งเอส - พี ซึ่งประกอบด้วยเลขที่นักเรียน คะแนนรวม ดัชนีของซาโต้ สัญลักษณ์ของการแปลผลข้อบกพร่อง ของซาโต้ เป็นรายบุคคลและได้แสดงให้เห็นถึงวิธีการแปลผลข้อสอบและข้อบกพร่องของนักเรียน ตลอดจนความสอดคล้องของข้อสอบกับกลุ่มนักเรียนโดยใช้ในการคำนวณจากค่าของสถิติมาช่วยในการตัดสินใจเรียก ค่าสัมประสิทธิ์ความสอดคล้องของนักเรียนและข้อสอบ (Disparity coefficient) การแปลผลความบกพร่องของนักเรียนรายบุคคล จะแปลผลด้วยสัญลักษณ์ A , B , C และ D ส่วนการแปลผลของความบกพร่องของข้อสอบ จะแปลโดยใช้สัญลักษณ์ W , X , Y และ Z และในปี 1987 แฮนิส (Harnisch 1987: 840) ได้พัฒนารูปแบบของเส้นภาพ (Profile) ซึ่งประยุกต์ใช้ในการศึกษา จากการพิจารณาความบกพร่องในการตอบข้อสอบของนักเรียน โดยจะเขียนทั้งข้อที่นักเรียนตอบผิดในข้อสอบแต่ละข้อลงในเส้นภาพ (Profile) ด้วยเป็นรายบุคคล เพื่อสะดวกในการวินิจฉัยนักเรียน กลุ่มตัวอย่างที่ใช้เป็นนักเรียนฝึกหัดครู

จาก 303 ห้อง จำนวน 5,954 คน เครื่องมือที่ใช้เป็นแบบสำรวจทักษะพื้นฐาน ในการเสนอรูปกราฟ จะใช้สัญลักษณ์ในรูปแบบแผนการตอบข้อสอบของนักเรียน 3 สัญลักษณ์ด้วยกันคือ เครื่องหมายบวก (+) หมายถึง ข้อสอบข้อนั้นนักเรียนตอบถูก M , L เป็นสัญลักษณ์ที่แสดงถึงทักษะในสมรรถภาพด้านนั้น ๆ มีปานกลาง (middle) และต่ำ เพื่อสะดวกและง่าย ในการแปลผล หรือวินิจฉัยนักเรียนเป็นรายบุคคล นอกจากนั้น ยังแยกทักษะเป็นกลุ่มทักษะ เพื่อบ่งบอกถึงความบกพร่องหรือความเด่น ในแต่ละสมรรถภาพของนักเรียนเป็นรายบุคคล ในการคำนวณหาค่าของดัชนีต่าง ๆ ตลอดจนการแปลผล และค่าสถิติต่าง ๆ แฮนิส และคณะ (Harnisch and Staff nd : 1-18) ได้พัฒนาการคำนวณค่าดัชนีค่าสถิติต่าง ๆ การเขียนแผนภูมิ เอส - พี ด้วยเครื่องคอมพิวเตอร์ ซึ่งเป็นโปรแกรมคอมพิวเตอร์ที่มีความเหมาะสมสำหรับสถานศึกษาต่าง ๆ แบ่งออกเป็น 2 โปรแกรม คือ โปรแกรมสำเร็จรูปสำหรับวิเคราะห์ผลการสอบ (Test Analysis Package : TAP) ซึ่งโปรแกรมนี้พัฒนาขึ้นโดย แฮนิส ฮอวิทซ์ และแวง (Harnisch Horwitz and Wang 1984) สำหรับวิเคราะห์ข้อมูลการสอบของนักเรียน และโปรแกรมสำเร็จรูปการเขียนแผนภูมิเอส - พี (Student - Package : SPP) เป็นโปรแกรมที่ แฮนิส และโรมี ได้พัฒนาขึ้นในปี 1985 (Harnisch and Romy 1985) ใช้กับเครื่องคอมพิวเตอร์ IBM - PC SPP เป็นโปรแกรมที่มีประโยชน์ในการศึกษามาก สะดวกรวดเร็วในการสร้างแผนภูมิเอส - พี และสร้างเส้นภาพ (Profile) ประจำชั้นได้ดีมาก ตลอดจนวิเคราะห์ความบกพร่องของนักเรียนเป็นรายสมรรถภาพ รายบุคคลได้เช่นกัน นอกจากนั้นโปรแกรม SPP ยังบ่งบอกความบกพร่องของนักเรียนโดยใช้สัญลักษณ์ A , B , C และ D และความบกพร่องของข้อสอบโดยใช้สัญลักษณ์ W , X , Y และ Z

สุนันท์ ศลโกสุม (2530: 122-124) ได้ศึกษาเปรียบเทียบการวิเคราะห์ผลการสอบโดยทฤษฎีดั้งเดิม ทฤษฎีการตอบสนองข้อสอบ ดัชนีของซาโต้ และดัชนีของทาทซูโอกะ ใช้กลุ่มตัวอย่างที่ศึกษานักเรียนมัธยมศึกษาปีที่ 6 วิชาภาษาไทย ในโครงการประเมินคุณภาพการศึกษา ปีการศึกษา 2528 กรมวิชาการ กระทรวงศึกษาธิการ นักเรียนเขตการศึกษา 9,10 และ 11 จำนวน 17 โรง จำนวนนักเรียน 5,774 คน ในการแปลผลดัชนีของซาโต้ในการวิจัยจะแปลผลควบคู่กับค่าคะแนนเฉลี่ยของกลุ่มเพื่อให้เกิดความชัดเจนขึ้น โดยการแปลผลวินิจฉัยจะแปลผลใน 3 ลักษณะคือ ลักษณะที่หนึ่ง ค่า $\bar{X}$ ต่ำ และค่า C* น้อยกว่า .30 แสดงว่า นักเรียนมีความสามารถต่ำจริง เพราะการตอบข้อสอบจะตอบได้เฉพาะ

ข้อสอบที่ง่าย ๆ เท่านั้น ต้องปรับปรุงการเรียนการสอนเพื่อให้ทั้งกลุ่มมีความสามารถเพิ่มขึ้น ถ้าค่า $\bar{x}$ สูง แสดงว่า นักเรียนมีความสามารถสูงจริง การปรับปรุงการเรียนการสอนต้องปรับปรุงเป็นรายบุคคล เพราะสภาพการเรียนการสอนไม่บกพร่อง ลักษณะที่สอง ค่า C* มากกว่า .30 แสดงว่า มีความผิดปกติในการเรียนการสอน ต้องพิจารณาค่า $\bar{x}$ สูงหรือต่ำ ถ้าค่า $\bar{x}$ ต่ำ แสดงว่า อาจจะมีความผิดปกติในการเรียนการสอน หรือข้อสอบไม่เหมาะสมกับกลุ่มตัวอย่าง ถ้าค่า $\bar{x}$ สูง และ C* มากกว่า .30 ต้องพิจารณาว่า บกพร่องในเนื้อหาใด ลักษณะที่ 3 โรงเรียนใดมีความผิดปกติ ในการทำข้อสอบมากกว่ากัน โดยพิจารณาค่าดัชนีบ่งชี้ก่อน แล้วพิจารณาค่า $\bar{x}$ ถ้าค่า $\bar{x}$ สูง แต่ค่าดัชนีผิดปกติ แสดงว่า ต้องพิจารณาว่าบกพร่องด้านใด การวิเคราะห์หาค่า Weighted Standardized mean residaul ประกอบด้วย สุนันท์ ศลโกสุม (2530: 126) ได้สรุปว่าการวิเคราะห์ผลการสอบด้วยดัชนีของทาทซูโอกะ ได้ผลไม่สอดคล้องกับผลการวิเคราะห์ด้วยดัชนีของ ซาโต้ แสดงว่าเกิดความแตกต่างกันในเรื่องการเรียงลำดับข้อสอบตามความยากกับการแทนค่าความยากลงในตารางเมตริกซ์

มิลเลอร์ (Miller 1986: 147-156) ได้ศึกษาแบบแผนการตอบข้อสอบของนักเรียนโดยใช้ดัชนีของซาโต้ ในการหาความสัมพันธ์ของแบบแผนการตอบข้อสอบ กับเวลาที่กำหนดให้ ในการจัดกิจกรรมการเรียนการสอนของเนื้อหาที่สอน กลุ่มตัวอย่างที่ใช้เป็นนักเรียนระดับชั้น เกรด 2 และ 5 จำนวน 123 คน สุ่มมาจาก 21 ห้องเรียน เครื่องมือที่ใช้วัดคือ แบบทดสอบวัดผลสัมฤทธิ์วิชาคณิตศาสตร์ จำนวน 101 ข้อ เนื้อหาทั้งหมด 11 เนื้อหา กำหนดในการสอบ 17 สัปดาห์ การหาค่าความสัมพันธ์ใช้ตาราง 2×2 Contigency โดยใช้ค่าดัชนีของซาโต้ซึ่งใช้เกณฑ์ .50 กับค่าเบี่ยงเบนมาตราฐาน ซึ่งใช้เกณฑ์ตัดสินคือ 2.0 ค่าสหสัมพันธ์ของการกำหนดเวลาในการกำหนดกิจกรรมการเรียนการสอนกับแบบแผนการตอบข้อสอบมีค่า r = .44 หมายความว่า รูปแบบแผนในการตอบข้อสอบของนักเรียนมีความสัมพันธ์กับการกำหนดเวลาในการเรียนการสอน ในแต่ละห้องเรียนอย่างมีนัยสำคัญทางสถิติ

เจเกอร์ และบูลช์ (Jager and Busch 1986: 193-194) ได้ศึกษาสำรวจความบกพร่องของนักเรียน โดยใช้ดัชนีของซาโต้เป็นตัวบ่งบอกถึงความสามารถของนักเรียน ข้อมูลที่ใช้ในการศึกษาครั้งนี้ ได้มาจากโครงการประเมินสมรรถภาพความสามารถมาตราฐาน เครื่องมือที่ใช้เป็นแบบทดสอบความสามารถ (Competency test) ระดับมัธยมศึกษาตอนปลาย วิชาการอ่านจำนวน 120 ข้อ กลุ่มตัวอย่างจำนวน 83 คน และวิชาคณิตศาสตร์ 120 ข้อ กลุ่มตัวอย่าง 84 คน จัดเรียงผลเป็นรูปเมตริกซ์ข้อต่อข้อ ด้วยทฤษฎีเส้นโค้งเอส-พี และค่าดัชนีของซาโต้ซึ่งผลการศึกษาพบว่า ดัชนีของซาโต้มีประสิทธิภาพในการบ่งบอกความบกพร่องของการสอบได้ดีกว่าการพิจารณาความสามารถจากคะแนนรวมแต่เพียงอย่างเดียว ดังนั้นกล่าวได้ว่า ดัชนีของซาโต้มีประโยชน์ในการคัดเลือก ตัดสินความสามารถของนักเรียนได้ดีกว่า และชี้ให้เห็นว่าการคัดเลือกนักเรียนควรจะใช้ดัชนีของซาโต้เพราะเป็นดัชนีที่สามารถแยกแยะความบกพร่องได้ชัดเจนยิ่งขึ้น

จากงานวิจัยที่ผู้วิจัยนำมาเสนอนี้แบ่งได้เป็น 2 กลุ่มคือ กลุ่มที่หนึ่งเป็นการศึกษาเปรียบเทียบดัชนีชี้เตือนของซาโต้กับดัชนีอื่น ๆ โดยการหาค่าสหสัมพันธ์ระหว่างดัชนีต่าง ๆ เพื่อดูว่าดัชนีคู่ใดมีความสัมพันธ์กันมาก ดัชนีคู่ใดสัมพันธ์กันน้อย แต่การศึกษาเหล่านี้ไม่ได้แสดงว่าดัชนีตัวใดดีพอที่จะนำไปใช้ได้ และการศึกษาเกี่ยวกับการนำดัชนีชี้เตือนของซาโต้ไปใช้ในการตรวจสอบความบกพร่องของผู้ตอบแบบทดสอบ ดังงานวิจัยของแฮนิส (Harnisch 1983: 191-206) และงานวิจัยของ สุนันท์ ศลโกสุม (2530: 122-124) ซึ่งส่วนใหญ่ใช้ข้อมูลจริง โดยสรุปแล้วงานวิจัยทั้งหมดเป็นการนำดัชนีที่มีอยู่แล้วมาเปรียบเทียบกันกับการนำดัชนีเหล่านั้นมาใช้เท่านั้น ไม่มีงานวิจัยชิ้นใดเลยที่แสดงการพัฒนาดัชนีต่าง ๆ เพื่อแก้ไขข้อบกพร่องของดัชนีที่มีอยู่แล้ว และไม่มีงานวิจัยชิ้นใดที่สามารถบอกได้ว่าดัชนีตัวใดดี โดยมีเกณฑ์ในการหาคุณภาพที่ชัดเจน