บทที่ 2

วรรณคดีและงานวิจัยที่เกี่ยวข้อง

ผู้วิจัยนำเสนอแนวคิดและทฤษฎีที่เกี่ยวข้อง เป็น 5 ตอน คือ

1. แนวคิดเกี่ยวกับสเกลของกัตแมน (Guttman Perfect Scale)
2. ดัชนีบ่งชี้ความเหมาะสมของบุคคล (Person-fit indices)
3. ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)
4. วิธีการตรวจสอบคุณภาพของแบบสอบ
5. รายงานการวิจัยที่เกี่ยวข้อง

1. <u>แนวคิดเกี่ยวกับสเกลของกัตแมน</u> (Guttman Perfect Scale)

กัตแมน เสนอความคิดเกี่ยวกับ แบบแผนการตอบข้อสอบ (Item Response Pattern) โดยจัดเรียงลำดับข้อสอบจากข้อที่ง่ายที่สุด ไปจนถึงข้อที่ยากที่สุด ตามแนวแถว และนำคะแนนรายข้อของผู้สอบแต่ละคนเรียงลำดับจากคะแนนรายข้อของผู้สอบที่มีคะแนนสูงสุดถึงคะแนนต่ำสุดตามแนวแถวจากบนลงล่าง เป็นเมตริกซ์ของคะแนนรายข้อของผู้สอบ N คน ที่ตอบข้อสอบ K ข้อ ลักษณะเมตริกซ์ของกัตแมนนี้จะมีความสอดคล้องสมบูรณ์เป็นไปตามข้อตกลงเบื้องต้น เมื่อคะแนนรายข้อที่อยู่มุมบนด้านซ้ายมือของเมตริกซ์มีค่าเป็น 1 หมดและคะแนนรายข้อที่อยู่มุมล่างด้านขวามือมีค่าเป็น 0 หมด และเรียกเมตริกซ์ที่มีลักษณะเช่นนี้ว่า กัตแมนสมบูรณ์ (Perfect Guttman)

ข้อตกลงเบื้องต้นของกัตแมน (Allen and Yen, 1979)

1.1 ข้อสอบที่ผู้สอบที่มีความสามารถระดับหนึ่งตอบได้ถูกตอบน่าจะเป็นข้อสอบที่ผู้สอบทุกคนที่มีความสามารถสูงกว่าตอบได้ถูกต้อง ในทางตรงกันข้ามข้อสอบข้อที่ผู้ที่มีความสามารถระดับหนึ่งตอบผิดย่อมทำให้ผู้สอบทุกคนที่มีความสามารถต่ำกว่าผู้สอบคนนั้นตอบผิดด้วย

1.2 ผู้สอบที่ตอบข้อสอบที่มีความยากระดับหนึ่งได้ถูกต้อง น่าจะตอบข้อสอบทุกข้อที่ง่ายกว่าข้อดังกล่าวได้ถูกต้องด้วย ในทางตรงกันข้ามเมื่อผู้สอบตอบข้อสอบที่มีความยากระดับหนึ่งผิด เขาย่อมจะตอบข้อสอบทุกข้อที่ยากกว่าข้อดังกล่าวผิดด้วย

ตารางที่ 1 สเกลของกัตแมน (GUTTMAN PERFECT SCALE)

ข้อสอบ / ผู้สอบ	1	2	3	4	5	6	7	8	9	10	รวม
1	1	1	1	1	1	1	1	1	1	0	9
2	1	1	1	1	1	1	1	1	0	0	8
3	1	1	1	1	1	1	1	0	0	0	7
4	1	1	1	1	1	1	0	0	0	0	6
5	1	1	1	1	1	0	0	0	0	0	5
6	1	1	1	1	0	0	0	0	0	0	4
7	1	1	1	0	0	0	0	0	0	0	3
8	1	1	0	0	0	0	0	0	0	0	2
9	1	0	0	0	0	0	0	0	0	0	1
10	0	0	0	0	0	0	0	0	0	0	0

จากตารางผู้สอบทุกคนมีแบบแผนการสอบที่เป็นปกติ ตามรูปแบบของกัตแมน

2. **ดัชนีบ่งชี้ความเหมาะสมของบุคคล** (Person-fit index)

ดัชนีบ่งชี้ความเหมาะสมของบุคคล (Person-fit index) หมายถึง ค่าสถิติที่บ่งบอกว่าบุคคลนั้นมีแบบแผนการตอบที่สอดคล้องกับโมเดลการตอบข้อสอบ ตามแนวความคิดของกัตแมนที่เรียกว่า กัตแมนสมบูรณ์ (Perfect Guttman) มากน้อยเพียงใด

2.1 วัตถุประสงค์ในการวิเคราะห์บุคคลที่เหมาะสม

นักวัดผลการศึกษาวิเคราะห์บุคคลที่เหมาะสม (person-fit) เพื่อจุดมุ่งหมายที่แตกต่างกันซึ่งสามารถกล่าวโดยสรุปได้ดังนี้

2.1.1 เพื่อบ่งชี้ความผิดพลาดในการแปลผลของคะแนนจากแบบสอบชุดนั้นๆ เนื่องจากความสามารถ (ability) ของผู้สอบ ที่ตอบข้อสอบแตกต่างไปจากที่คาดหวังไว้ในโมเดล ได้

จากการประมาณค่าที่คลาดเคลื่อน (Drasgow & Levine,1987; Drasgow,Levine& Williams,1985; Levine & Rubin,1979 cited in Drasgow,1996)

2.1.2 เพื่อบ่งชี้บุคคลที่มีแบบแผนการตอบที่ผิดปกติเนื่องจากมีมโนทัศน์ที่คลาดเคลื่อน (misconseption) วิธีการวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนนั้น ทาทซูโอกะใช้วิธีการ "rule space" วิเคราะห์แบบการคิดที่นักเรียนใช้ในการแก้ปัญหา จากแบบแผนการตอบข้อสอบ (Benbaum,Kelly & Tatsuoka,1992; Tatsuoka & Tatsuoka,1992; cited in Merjer,1996; ศิริเดช สุชีวะ, 2537)

2.1.3 เพื่อบ่งชี้แบบแผนการตอบที่ผิดปกติของผู้สอบซึ่งอาจเกิดจากพฤติกรรมการตอบข้อสอบของผู้สอบ การบริหารการสอบ หรือความบกพร่องของข้อสอบ (Hanisch & Linn,1981; Rudner,1983; Birenbaum,1986; Meijer,1996) ซึ่งในส่วนที่เกี่ยวกับพฤติกรรมการตอบข้อสอบนั้น เมเจอร์ (Meijer,1996) ได้กล่าวถึง รูปแบบของพฤติกรรมการตอบข้อสอบที่ทำให้เกิดแบบแผนการตอบที่ผิดปกติไว้หลายรูปแบบ โดยเมเจอร์กำหนดว่า ในแบบสอบแบบเลือกตอบชุดหนึ่งซึ่งมีข้อสอบ 12 ข้อ แบ่งเป็นข้อสอบซึ่งมีความยาก 3 ระดับ คือ ง่าย ยาก และปานกลาง อย่างละ 4 ข้อ เรียงลำดับข้อสอบที่ง่ายที่สุดจนถึง ข้อที่มีค่าความยากสูงสุด ตรวจให้คะแนนแบบ 2 ค่า โดยข้อที่ตอบถูกให้ 1 คะแนน ข้อที่ผิดให้ 0 คะแนน เมื่อพิจารณารูปแบบการตอบของผู้สอบ จะมีพฤติกรรมการตอบที่ผิดปกติในรูปแบบต่าง ๆ กัน ดังต่อไปนี้

1.) ความสับสนเนื่องจากตัวแบบสอบ (Sleeping Behavior) ผู้สอบไม่สามารถตอบข้อสอบในข้อแรกๆได้ถูกต้อง เนื่องจากสับสนในรูปแบบของแบบสอบ และไม่ได้ทบทวนคำตอบ แต่ตอบได้ถูกต้องในข้อหลัง ๆ ซึ่งเป็นข้อที่ยาก เนื่องจากปรับตัวได้

2.) การเดาคำตอบ (Guessing Behavior) ผู้สอบซึ่งมีความสามารถต่ำ ไม่สามารถทำข้อสอบซึ่งมีค่าความยากสูงกว่าความสามารถของตนได้ เลยตัดสินใจเดาสุ่มคำตอบในข้อเหล่านั้น ซึ่งโอกาสที่จะเดาได้ถูกต้องขึ้นอยู่กับจำนวนตัวเลือกของข้อสอบชุดนั้น

3.) การทุจริตในการสอบ (Cheating Behavior) ผู้สอบซึ่งมีความสามารถต่ำทำข้อสอบซึ่งง่าย และความยากปานกลาง ด้วยความสามารถของตนเอง แต่ลอกคำตอบในข้อที่ยากมากจากผู้สอบที่มีความสามารถสูงกว่าตน ดังนั้นผู้สอบกลุ่มที่จะตอบข้อสอบซึ่งง่ายและยากได้ถูกต้องแต่จะตอบข้อสอบซึ่งมีความยากปานกลางผิด

4.) การทำข้อสอบไม่ทันตามกำหนด (Plodding Behavior) ผู้สอบซึ่งทำข้อสอบช้าและตอบข้อสอบไปที่ละข้อโดยไม่ข้ามไปทำข้ออื่นๆ ทำให้ไม่สามารถตอบข้อสอบได้ทันตามกำหนดเวลาจึงมีข้อสอบบางข้อซึ่งผู้สอบไม่ได้ตอบ ทั้งๆ ที่เป็นข้อที่มีโอกาสจะตอบได้ถูกต้อง

5.) ความสะเพร่า (Alignment Errors) ผู้ตอบตัดสินใจเว้นคำตอบข้อซึ่งไม่สามารถตอบได้ในขณะนั้น เมื่อทำข้อสอบข้อต่อไปกลับทำผิดข้อเนื่องจากลืมข้อที่เว้นคำตอบไว้ทำให้คำตอบในข้อหลัง ๆ ผิด

6.) การเปลี่ยนคำตอบ (Extremely Creative Examinees) ผู้ตอบทำข้อสอบครบทุกข้อแล้วทบทวนคำตอบใหม่ ตัดสินใจแก้คำตอบในข้อง่าย ๆ ทำให้ตอบข้อเหล่านั้นผิด

7.) แบบสอบวัดหลายคุณลักษณะ (Deficiency of Subabilities) แบบสอบชุดเดียววัดเนื้อหา 2 เนื้อหา ผู้ตอบมีความรู้ในแบบสอบตอนหลัง ซึ่งเป็นข้อที่ยากทำให้ตอบข้อสอบถูกแต่จะตอบในข้อที่ง่าย ๆ ไม่ได้ เนื่องจากไม่มีความรู้ในเรื่องนั้น

พฤติกรรมต่าง ๆ เหล่านี้ เมื่อปรากฏในแบบแผนการตอบของผู้สอบคนใด แบบแผนการตอบของบุคคลนั้นจะเป็นแบบแผนการตอบซึ่งผิดปกติ วิธีการในการตรวจค้นผู้ตอบซึ่งมีแบบแผนการตอบผิดปกตินั้นนักวัดผลใช้วิธีการตรวจสอบความเหมาะสมของบุคคลบโมเดลการตอบข้อสอบ โดยใช้ค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลจะบอกได้แต่เพียงว่าเป็นแบบแผนการตอบที่ผิดปกติ แต่ไม่สามารถระบุไว้ว่าเป็นแบบแผนซึ่งเกิดจากพฤติกรรมแบบใดเนื่องจากแบบแผนการตอบซึ่งมีสาเหตุมาจากพฤติกรรมการตอบข้อสอบที่แตกต่างกัน จะปรากฏผลในแบบแผนการตอบที่คล้ายคลึงกันได้ เมื่อผู้สอบ 7 คน ตอบข้อสอบ 12 ข้อ ซึ่งมีค่าความยากเรียงจากง่ายไปหายาก จะปรากฏพฤติกรรม การตอบข้อสอบที่แตกต่างกัน ดังตาราง

ตารางที่ 2 แบบแผนการตอบข้อสอบที่ผิดปกติ

ผู้สอบ	ข้อสอบข้อที่												พฤติกรรมการตอบข้อสอบ
	1	2	3	4	5	6	7	8	9	10	11	12	
1	0	0	1	1	1	1	1	1	1	1	0	1	ความสับสนในรูปแบบข้อสอบ
2	1	1	1	1	0	0	1	,0	0	0	0	1	การเดาคำตอบ
3	1	1	0	1	0	1	0	0	0	1	1	1	การทุจริตในการสอบ
4	1	1	1	1	1	1	0	0	0	0	0	0	การทำข้อสอบไม่ทัน
5	1	1	1	1	1	0	1	1	0	0	0	0	ความสะเพร่า
6	0	0	0	0	1	1	1	1	0	1	1	1	การเปลี่ยนคำตอบ
7	0	0	1	0	1	1	1	0	1	1	1	0	แบบสอบวัดหลายคุณลักษณะ

จากตารางผู้สอบคนที่ 1 และผู้สอบคนที่ 6 จะมีพฤติกรรมการตอบข้อสอบที่คล้ายคลึงกัน และในทำนองเดียวกันลักษณะการตอบข้อสอบของผู้สอบคนที่ 2 กับคนที่ 3 ก็ไม่สามารถระบุได้ว่าเกิดแบบแผนการตอบข้อสอบที่ผิดปกติเนื่องจากสาเหตุใด

2.2 ประเภทของดัชนีบ่งชี้ความเหมาะสมสำหรับบุคคล

ดัชนีที่ใช้วิเคราะห์แบบแผนการตอบข้อสอบรายบุคคลนั้น มีหลายดัชนี ซึ่ง สุนันท์ ศลโกสุม (2530) ได้เสนอเป็น 2 กลุ่ม ตามเทคนิคของการวิเคราะห์ ดังนี้

กลุ่มแรก วิเคราะห์โดยการหาความสัมพันธ์ของแบบแผนการตอบถูกและผิดเป็นรายบุคคลกับคะแนนรวม เป็นทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ดัชนีพวกนี้ได้แก่ ดัชนีชี้นำของซาโต้ (Sato's caution index) ดัชนียู (U - index) ของเวนเดอร์ (vender, 1977) ดัชนีความสามารถของบุคคลภายในกลุ่ม (Norm conformity index) ของ ทาทซูโอกะและทาทซูโอกะ (Tatsuoka and Tatsuoka,1980) ดัชนีที่แสดงความสอดคล้องและไม่สอดคล้อง (Agreement and disagreement indices) ของเคนและแบรนแนน (Kane and Brennan,1980) และดัชนีความสอดคล้องของแบบแผนการตอบข้อสอบรายบุคคลของ พนิชา สังข์เพ็ชร(2536) แบบแผนของการตอบข้อสอบที่ถูกต้องและตามลักษณะของข้อสอบที่มีคุณภาพดี ผู้สอบควรจะตอบข้อสอบที่ง่ายๆได้ และตอบข้อสอบที่ยากไม่ได้ และคนที่มีคะแนนรวมเท่ากันควรตอบคำถามได้เหมือนๆกัน นั่นคือ ต้องมีแบบแผนการตอบถูกและผิดเหมือนกันในกลุ่มที่มีคะแนนเท่ากัน แต่ถ้าในกลุ่มผู้ตอบที่ได้คะแนนเท่ากัน มีการตอบถูกและผิดไม่เป็นรูปแบบเดียวกัน แสดงว่ามีความผิดปกติเกิดขึ้น อาจจะเนื่องจากข้อสอบ หรือตัวของผู้สอบ

กลุ่มที่สอง ใช้หลักการของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ในปี ค.ศ.1952 ลอร์ด (Lord and Novick,1968) เสนอทฤษฎีโค้งคุณลักษณะข้อสอบ (Item Characteristic Curve:ICC) ทฤษฎีการตอบสนองข้อสอบมีความเชื่อว่าพฤติกรรมการตอบสนองข้อสอบ ถูกกำหนดด้วยคุณลักษณะ (Trait) ภายในตัวบุคคลซึ่งไม่สามารถสังเกตได้ แต่สามารถอธิบายความสัมพันธ์ระหว่างคุณลักษณะนั้นกับพฤติกรรมการตอบสนองข้อสอบ โดยอาศัยแบบจำลองทางคณิตศาสตร์ ซึ่งมีพารามิเตอร์ที่ต้องประมาณค่า 4 ตัว คือ ความสามารถที่แท้จริงของผู้สอบ (θ) ค่าความยากของข้อสอบ (b) ค่าอำนาจจำแนกของข้อสอบ (a) และค่า

โอกาสการเดาถูก ในกรณีที่เป็น โมเดลแบบ 3 พารามิเตอร์ ดัชนีที่บ่งชี้ความเหมาะสมของบุคคลในกลุ่มนี้ มี ดัชนีของเลอวินและรูบิน (Levine and Rubin, 1979) ดัชนีที่พัฒนาโดย ดราสโกว์ (Drasgow,1978) การทดสอบไค-สแควร์ของบุคคล (The Chi-Square Test of Person-fit) ประยุกต์ใช้ในราส์ชโมเดล ดัชนี W_1 พัฒนาโดย ไรท์ และ สโตน (Wright & Stone,1979) ไรท์ (Wright,1980) และ ราช์ส (Rasch,1980) และสถิตินอนพาราเมตริกซ์ซึ่งพัฒนาโดย โมแคนและเลวิส (Mokken and Lewis,1982) หรือที่เรียกสถิติ U_3 (U_3 Statistic) ดัชนี DFC_1 และ DFC_2 ของ สำราญ มีแจ้ง(2533) และดัชนี ECIZ4 ของทาทซูโอกะ (Tatsuoka,1984) ดัชนีกลุ่มนี้คำนวณโดยการประมาณค่าความน่าจะเป็นที่ผู้สอบแต่ละคนจะตอบข้อสอบแต่ละข้อได้ถูกต้อง เปรียบเทียบกับผลการตอบข้อสอบในข้อนั้นๆ เพื่อจะดูว่าผิดไปจากแบบแผนการตอบที่ปกติหรือไม่ ถ้าผิดปกติแสดงว่าผู้สอบคนนั้นไม่เหมาะสมกับ โมเดลการตอบข้อสอบ (person misfit) หรือผู้สอบคนนั้นมีแบบแผนการตอบข้อสอบที่ผิดปกติ (aberrant response pattern)

2.3 ลักษณะของดัชนีบ่งชี้ความเหมาะสมของบุคคลที่มีประสิทธิภาพ

พนิชา สังข์เพชร (2535) ได้กล่าวถึง คุณลักษณะที่พึงประสงค์ของดัชนีไว้ 5 ประการ ดังนี้ คือ

ประการที่ 1 ดัชนีต้องมีค่าคงที่เสมอ เมื่อผู้สอบมีแบบแผนการตอบข้อสอบเป็นกัตแมนสมบูรณ์

ประการที่ 2 ดัชนีต้องมีค่าเป็นอิสระจากการเปลี่ยนแปลงของแบบแผนการตอบข้อสอบของผู้สอบคนอื่นๆภายในกลุ่ม โดยเฉพาะอย่างยิ่งในกรณีที่การเปลี่ยนแปลงนั้นไม่มีผล ให้ลำดับที่ของข้อสอบเปลี่ยนไป

ประการที่ 3 ดัชนีต้องมีค่าที่ตรงกับความหมายที่แท้จริงของแบบแผนการตอบข้อสอบของผู้สอบในทุกกรณี

ประการที่ 4 ดัชนีต้องมีช่วงพิสัยที่แน่นอนไม่ขึ้นอยู่กับค่าความยากของข้อสอบ หรือจำนวนข้อสอบ

ประการที่ 5 ดัชนีต้องมีความไวต่อความคลาดเคลื่อน

2.4 วิธีประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลตามทฤษฎีการตอบสนองข้อสอบ

ดัชนีที่ใช้วิเคราะห์ความเหมาะสมของบุคคล (Person-fit) ตามทฤษฎีการตอบสนองข้อสอบ มีวิธีการประมาณค่าดัชนี 3 วิธี (Rost and Davier,1994) คือ

2.4.1 การประมาณค่าด้วยฟังก์ชันไลด์ลิฮูด (The Likelihood-based Approach)

2.4.2 การประมาณค่าจากคะแนนส่วนที่เหลือ (The Score Residual Approach)

2.4.3 การทดสอบไค-สแควร์ (The Chi-Square Test)

เนื่องจากการทดสอบไค-สแควร์ของบุคคล (The Chi-Square Test) วิเคราะห์บุคคลซึ่งมีแบบแผนการตอบไม่เหมาะสมกับโมเดลการตอบข้อสอบ (person misfit) โดยมีความไวต่อความคลาดเคลื่อนในการตอบข้อสอบ จึงมีจุดด้อยเนื่องจากมีความลำเอียงในการปฏิเสธสมมติฐานเกี่ยวกับความเหมาะสมของบุคคลและข้อสอบในโมเดล (Reise,1990) ผู้วิจัยจึงไม่นำมาศึกษาในการวิเคราะห์ผู้สอบที่เหมาะสมกับโมเดลการตอบข้อสอบ ใช้วิธีการตรวจสอบความเหมาะสมของบุคคลเพียง 2 วิธี คือ การประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลด้วยฟังก์ชันไลด์ลิฮูด (The Likelihood-based Approach) และ การประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลจากคะแนนส่วนที่เหลือ (The Score Residual Approach) ส่วนการทดสอบไค-สแควร์ จะใช้เป็นวิธีการตรวจสอบข้อสอบที่เหมาะสมกับโมเดลการตอบข้อสอบ (Item fit) หลังจากวิเคราะห์ผู้สอบที่เหมาะสม ด้วยวิธีประมาณค่าทั้งสองวิธี และตัดผู้สอบที่ไม่เหมาะสมออกจากการวิเคราะห์ข้อสอบแล้ว

ขั้นตอนการวิเคราะห์ของดัชนีแต่ละวิธีมี ดังนี้

1) การประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลด้วยฟังก์ชันไลด์ลิฮูด (The Likelihood-based Approach) พัฒนาโดย เลวิน และ รูบิน ((Leven & Rubin,1979) และ ดราสโกว์ (Drasgow,1985) คำนวณค่าดัชนีโดยการประมาณค่าความน่าจะเป็นที่ผู้สอบแต่ละคนจะตอบข้อสอบแต่ละข้อได้ถูกต้อง เปรียบเทียบกับผลการตอบข้อสอบในข้อนั้นๆ แปลงค่าสถิติความเหมาะสมของบุคคลเป็นค่ามาตรฐาน (Z- Value) ทดสอบความแตกต่างแบบสองทาง (two-tailed test) ถ้าดัชนีบ่งชี้ความเหมาะสมของบุคคลใดมีค่าแตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ ที่ระดับ 0.05 คือมี $|L_z|$ มากกว่า 1.96 บุคคลนั้นมีแบบแผนการตอบไม่เหมาะสมกับโมเดลการตอบข้อสอบ (person misfit) หรือผู้สอบคนนั้นมีแบบแผนการตอบข้อสอบที่ผิดปกติ (aberrant response pattern) ดัชนีซึ่งประมาณค่าด้วยวิธีนี้คือดัชนี L_z ซึ่ง ไรส์ (Reise,1990) ได้ศึกษาเปรียบเทียบประสิทธิภาพของดัชนี L_z และ χ^2 พบว่า ดัชนี L_z มีประสิทธิภาพในการวิเคราะห์ความเหมาะสมของบุคคลและข้อสอบสูงกว่า ดัชนีกลุ่มนี้มีข้อดีคือ คำนวณ

หอสมุดกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ได้ง่าย ไม่มีความลำเอียง มีการแจกแจงเป็นโค้งปกติ อัตราการตรวจค้นจะขึ้นอยู่กับความยาวของแบบสอบและความสามารถของผู้สอบ

วิธีการประมาณค่าดัชนี L_z

1. ประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบโดย โมเดลโลจิสติกแบบ 3 พารามิเตอร์

2. ประมาณค่าความน่าจะเป็นที่ผู้สอบคนที่ i จะตอบข้อสอบข้อที่ j ได้ถูกต้องโดยใช้ค่าพารามิเตอร์ที่ประมาณได้จากข้อ 1

$$P_i(\theta) = C_i + (1 - C_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

เมื่อ

$P_i(\theta)$ คือ โอกาสที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ

a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

b_i คือ ระดับความยากของข้อสอบข้อที่ i

c_i คือ ค่าโอกาสการเดาถูก

e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

D คือ scaling factor เพื่อปรับเปลี่ยนค่าของฟังก์ชันโลจิสติกมีค่าเท่ากับ 1.7

3. ประมาณค่าสถิติบ่งชี้ความเหมาะสมของบุคคล (person-fit statistics)

$$l_0 = \sum_{i=1}^{n} \left\{ U_i\left[\ln P_i(\theta)\right] + (1 - U_i) * \left\langle \ln\left[1 - P_i(\theta)\right] \right\rangle \right\}$$

4. แปลงค่าสถิติบ่งชี้ความเหมาะสมของบุคคล (person-fit statistics) เป็นค่าดัชนีมาตรฐาน (standardized person-fit indices)

$$l_z = \frac{l_0 - E(l_0)}{\left[Var(l_0)\right]^{1/2}}$$

เมื่อ E (L_0) คือ ค่าที่คาดหวังของ L_0

Var (L_0) คือ ความแปรปรวนของ L_0

โดย $$E(l_0) = \sum_{i=1}^{n}\left\{P_i\left(\hat{\theta}\right)lnP_i\left(\hat{\theta}\right)+\left[1-P_i\left(\hat{\theta}\right)\right]\ln\left[1-P_i\left(\hat{\theta}\right)\right]\right\}$$

$$Var(L_0) = \sum_{i=1}^{n}P_i(\theta)\left[1-P_i(\theta)\right]\left\{\ln\left[\frac{P_i(\theta)}{1-P_i(\theta)}\right]\right\}^2$$

เมื่อ $P_i(\theta)$ คือ โอกาสที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$1-P_i$ คือ โอกาสที่ผู้สอบจะตอบข้อสอบข้อที่ i ผิด

U_i คือ พฤติกรรมการตอบที่สังเกตได้ (O,1)

θ คือ ค่าประมาณความสามารถของผู้สอบ

5. ทดสอบสมมติฐานโดยใช้สถิติทดสอบ ซี (z-test) แบบสองทาง

5.1 บุคคลที่ไม่เหมาะกับโมเดลการตอบ (person-misfit) คือ ผู้สอบที่มี $|L_z|$ มากกว่า 1.96 : α = .05 โดยพิจารณาเฉพาะผู้สอบคนที่ j ข้อสอบ i ข้อ

2) การประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลจากคะแนนส่วนที่เหลือ (The Score Residual Approach) พัฒนาโดย ไรท์ และ สโตน(Wright & Stone,1979) ไรท์ (Wright,1980) และ ราช์ส (Rasch,1980) วิธีนี้สามารถตรวจสอบแบบแผนการตอบข้อสอบได้ทั้งการให้คะแนนแบบ 2 ค่าและหลายค่า ดัชนีในกลุ่มนี้คือ W_1 (Wright,1979) และ W_3 (Rudner,1983) ประมาณค่าดัชนีโดยประมาณค่าสัดส่วนระหว่างผลรวมของกำลังสองของผลต่างของคะแนนที่สังเกตได้กับคะแนนที่คาดหวัง กับค่าที่คาดหวังจากโมเดลซึ่งคะแนนสองชุดนี้เป็น

สัดส่วนของความแปรปรวนของคะแนนส่วนที่เหลือซึ่งไม่เป็นอิสระจากกันมีการแจกแจงเป็นแบบที (t-distribution) ทดสอบความแตกต่างแบบสองทาง (two-tailed test) บุคคลซึ่งมีค่าสถิติบ่งชี้ความเหมาะสมของบุคคลแตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 คือ มี $|W_i|$ มากกว่า 2.00 เป็นบุคคลซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบ วิธีนี้มีข้อดี คือ ประมาณค่าโดยไม่มีความลำเอียง สามารถตรวจสอบแบบแผนการตอบได้ดีในทุกระดับความสามารถ การวิเคราะห์ผู้ตอบที่เหมาะสม (person-fit) เป็นการวิเคราะห์เพื่อคัดเลือกผู้ตอบที่มีแบบแผนการตอบเหมาะสมนำไปใช้ในการวิเคราะห์ข้อสอบที่เหมาะสม คือ ผู้ที่มีความสามารถสูงควรจะตอบข้อกระทงที่ง่ายและยากได้ถูกต้องมากกว่าผู้ที่มีความสามารถต่ำ ผู้ตอบที่ไม่เหมาะสมก็คือผู้ตอบที่มีแบบแผนการตอบที่ไม่เป็นไปตามแนวคิดนี้สามารถพิจารณาคัดเลือกบุคคลที่ไม่เหมาะสมได้จากการพิจารณาค่าสถิติ Person Infit , Person Outfit และค่า Mean Square Outfit (MNSQ) โดยผู้ตอบที่เหมาะสมควรจะมี |ค่าสถิติ INFIT| และ |ค่าสถิติ OUTFIT| ไม่เกิน 2.00 และค่า MNSQ ไม่เกิน 1.20 (Linacre and Wright, 1994) การวิเคราะห์ข้อสอบที่เหมาะสม (item-fit) เป็นการวิเคราะห์เพื่อคัดเลือกข้อสอบที่สามารถจำแนกคนที่มีระดับคุณลักษณะสูงและต่ำออกจากกันได้ซึ่งมีหลักการในการคัดเลือกค่าสถิติ INFIT,OUTFIT และ MNSQ เหมือนกับการวิเคราะห์ผู้ตอบที่เหมาะสม โดยการวิเคราะห์ข้อกระทงที่เหมาะสม จะวิเคราะห์หลังจากได้คัดเลือกผู้ตอบที่ไม่เหมาะสมออกจากการวิเคราะห์แล้ว เพราะถ้าเครื่องมือที่สร้างมีคุณภาพดี ความคลาดเคลื่อนจากตัวผู้ตอบน่าจะมากกว่าเครื่องมือ (ธนวัฒน์ แสนสุข, 2539)

วิธีการประมาณค่าดัชนี W_i

1. ประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบ โดยโมเดลโลจิสติกแบบหนึ่งพารามิเตอร์

2. ประมาณค่าความน่าจะเป็นที่ผู้สอบคนที่ j จะตอบข้อสอบข้อที่ i ได้ถูกต้องโดยใช้ค่าพารามิเตอร์ที่ประมาณได้จากข้อ 1

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

เมื่อ

$P_i(\theta)$ คือ โอกาสที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ ถูกต้อง

θ คือ ระดับความสามารถที่แท้จริงของผู้สอบ

b_i คือ ระดับความยากของข้อสอบข้อที่ i

e คือ ค่าคงที่มีค่าเท่ากับ 2.7182818

D คือ scaling factor เพื่อปรับเปลี่ยนค่าของฟังก์ชั่นโลจิสติก มีค่าเท่ากับ 1.7

3. ประมาณค่าสถิติบ่งชี้ความเหมาะสมของบุคคล (person-fit statistics) :

Unweight total mean square : MNSQ - outfit

$$W_1 = \frac{1}{N}\sum_{j=1}^{N}\frac{(U_{ij} - P_{ij})^2}{P_{ij}(1-P_{ij})}$$

เมื่อ

U_{ij} คือ พฤติกรรมการตอบข้อสอบที่สังเกตได้ (0 , 1) ของผู้สอบคนที่ j ข้อสอบข้อที่ i

P_{ij} คือ โอกาสที่ผู้สอบคนที่ j จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

บุคคลที่มีค่า MNSQ-outfit มากกว่า 1.20 เป็นผู้สอบที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบ

4. ประมาณค่าสถิติบ่งชี้ความเหมาะสมของบุคคล (person-fit statistics)

Weighted total mean square : MNSQ - infit

$$W_1 = \frac{\sum_{i=1}^{n}(U_{ij} - P_{ij})^2}{\sum_{j=1}^{N}P_{ij}(1-P_{ij})}$$

(Smith, 1994; Linacre & Wright, 1994)

เมื่อ N คือ จำนวนผู้เข้าสอบ

บุคคลที่มีค่า Wighted total mean square มากกว่า 1.20 เป็นผู้สอบที่ไม่เหมาะกับโมเดลการตอบข้อสอบ

. แปลงค่าสถิติบ่งชี้ความเหมาะสมของบุคคล (person-fit statistics) ค่าดัชนีมาตรฐาน (standardized mean square outfit statistic) : PERSON OUTFIT

5. $t_i = (v_i^{1/3} - 3/q_i) + (q_i/3)$

เมื่อ t_i คือ ค่าดัชนีความเหมาะสมของคนที่ j

v_i คือ ค่า unweighted mean square ของคนที่ j

q_i คือ ส่วนเบี่ยงเบนมาตรฐานของ unweighted mean square

6. แปลงค่าสถิติบ่งชี้ความเหมาะสมของบุคคล (person-fit statistics) เป็นค่าดัชนีมาตรฐาน (standardized mean square infit statistic) : PERSON INFIT

$$t_i = (v_i^{1/3} - 3/q_i) + (q_i/3)$$

เมื่อ t_i คือ ค่าดัชนีความเหมาะสมของคนที่ j

v_i คือ ค่า Weighted mean square ของคนที่ j

q_i คือ ส่วนเบี่ยงเบนมาตรฐานของ Weighted mean square

บุคคลที่มี |ค่าสถิติ OUTFIT| และ |ค่าสถิติ INFIT| มากกว่า 2.00 เป็นผู้สอบที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบ

3). การทดสอบไค-สแควร์ (The Chi-Square Test)

พัฒนาโดย ไรท์ และปัญจปากีสาน (Wright & Panchapakesan,1969) บอค (Bock,1972) และ วอลเลนเบอร์ก (Wollenberg, 1979) โดยเปรียบเทียบสัดส่วนของความถี่ของการตอบที่สังเกตได้กับความถี่ของการตอบที่คาดหวังของกลุ่มผู้ตอบข้อสอบ ทดสอบความแตกต่างด้วยสถิติทดสอบไค-สแควร์ (Chi-square) χ^2- test แบบทางเดียว (one-tailed test) ถ้า

ดัชนีบ่งชี้ความเหมาะสมของบุคคลใดมีค่าแตกต่างจาก 0 อย่างมีนัยสำคัญที่ระดับ 0.05 คือ มีค่าสถิติ χ^2 มากกว่า 9.48 บุคคลนั้นมีแบบแผนการตอบไม่เหมาะสมกับโมเดลการตอบข้อสอบ (person misfit) หรือผู้สอบคนนั้นมีแบบแผนการตอบข้อสอบที่ผิดปกติ (aberrant response pattern) ซึ่งวิธีนี้มีข้อดีคือ จะไม่มีข้อจำกัดในการวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ มีความไวต่อความคลาดเคลื่อนในการตอบข้อสอบ แต่จะมีจุดด้อยเนื่องจากมีความลำเอียงในการปฏิเสธสมมติฐานเกี่ยวกับความเหมาะสมของบุคคลและข้อสอบในโมเดล (Reise,1990)

การวิเคราะห์ข้อสอบที่เหมาะสมกับโมเดลการตอบข้อสอบ (item-fit) โดยการทดสอบค่าสถิติไค-สแควร์ ในโปรแกรม BILOG โปรแกรมจะจัดกลุ่มผู้สอบออกเป็นกลุ่ม ตามค่าโอกาสในการตอบถูกซึ่งประมาณค่าได้ โดยโปรแกรมจะสามารถวิเคราะห์ได้เมื่อมีจำนวนข้อสอบตั้งแต่ 20 ข้อ ขึ้นไป คำนวณสัดส่วนของข้อสอบซึ่งมีโอกาสที่จะทำให้ผู้สอบแต่ละคนในแต่ละกลุ่มตอบได้ถูกต้อง เปรียบเทียบกับพฤติกรรมการตอบข้อสอบที่สังเกตได้จริง

มีขั้นตอนการคำนวณค่าสถิติ ดังนี้

7.1 ประมาณค่าพารามิเตอร์ของผู้สอบ และข้อสอบด้วยโมเดลโลจิสติกแบบ 2 พารามิเตอร์

7.2 วิเคราะห์ข้อสอบที่เหมาะสมกับโมเดลการตอบข้อสอบ ด้วยการทดสอบค่าสถิติไค-สแควร์ ดังนี้

$$\chi^2 = \sum_{j=1}^{N} \frac{n_{ij}(O_{ij} - E_{ij})^2}{E_{ij}(1-E_{ij})}$$

(Rost and Davier,1994)

เมื่อ O_{ij} คือ พฤติกรรมการตอบข้อสอบที่สังเกตได้ (0,1)

E_i คือ โอกาสที่ผู้สอบแต่ละคนในกลุ่ม j จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

n_{ij} คือ จำนวนผู้สอบในกลุ่ม j เมื่อตอบข้อสอบข้อที่ i

j คือ จำนวนกลุ่มของผู้สอบที่อยู่ในช่วงของโอกาสที่จะตอบข้อสอบข้อที่ i ได้ถูกต้องเดียวกัน

ข้อสอบที่ไม่เหมาะกับโมเดลการตอบ (item-misfit) คือ

ข้อสอบที่มีค่า χ^2 มากกว่า ค่า χ^2 ที่มี df เท่ากับ j - 1

3. <u>ทฤษฎีการตอบสนองข้อสอบ</u> (Item Response Theory)

ผู้วิจัยนำเสนอหลักการและแนวคิดทฤษฎีการวัดผลการศึกษา เป็น 2 หัวข้อ คือ แนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ และ แนวคิดเกี่ยวกับราส์ชโมเดล

3.1 แนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบเป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่มีอยู่ภายในตัวบุคคล กับพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น (Lord and Novick ,1968: 358) โดยทฤษฎีนี้มีความเชื่อว่า พฤติกรรมการตอบสนองต่อข้อสอบของผู้เข้าสอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรง จะถูกกำหนดโดยคุณลักษณะ (trait) หรือความสามารถ (ability) ที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้ ทฤษฎีนี้พยายามที่จะอธิบายความสัมพันธ์ดังกล่าว ให้แสดงออกมาในรูปแบบจำลองทางคณิตศาสตร์โดยให้คะแนนที่ได้รับจากการตอบข้อสอบ y แทนพฤติกรรมการตอบสนองต่อข้อสอบและ θ แทนคุณลักษณะ หรือความสามารถภายในตัวบุคคล ฟังก์ชันการถดถอยของ y บน θ จะเป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างพฤติกรรมการตอบสนองต่อข้อสอบ กับระดับความสามารถ เรียกว่า ฟังก์ชันการตอบสนองข้อสอบ (Item Response Function) ซึ่งจะกำหนดได้หลายรูปแบบ ทั้งนี้ขึ้นอยู่กับข้อตกลงเบื้องต้นเกี่ยวกับข้อมูลจากการทดสอบ

3.1.1 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบมีข้อตกลงเบื้องต้นที่สำคัญ ดังนี้ (ศิริชัย กาญจนวาสี,2534)

1) แบบสอบวัดคุณลักษณะเด่นเพียงลักษณะเดียว (Unidimensionality : One trait) คุณลักษณะ หรือความสามารถที่เป็นตัวกำหนดพฤติกรรมการตอบข้อสอบแต่ละข้อมี ลักษณะเด่นและสำคัญเพียงลักษณะเดียว นั่นคือ แบบสอบวัดคุณลักษณะสำคัญเพียงลักษณะเดียวหรือแบบสอบมีความเป็นเอกพันธ์ การตรวจสอบความเป็นไปตามข้อตกลงเบื้องต้นข้อนี้สามารถทำได้โดยการวิเคราะห์องค์ประกอบ (Hambleton and Sawaminathran, 1985)

2) การตอบข้อสอบเป็นไปอย่างอิสระ (Independence : Local independent) การตอบข้อสอบแต่ละข้อของผู้ตอบแต่ละคนมีความเป็นอิสระจากกัน ซึ่งประกอบด้วย

2.1) ความเป็นอิสระระหว่างข้อสอบ หมายความว่า ข้อสอบแต่ละข้อเป็นอิสระจากกัน กล่าวคือ การตอบสนองต่อข้อสอบข้อหนึ่งไม่มีผลต่อการตอบสนองต่อข้อสอบข้อนั้นๆ ในแบบสอบ เช่น เนื้อหาของคำถามข้อหนึ่งจะต้องไม่จบของข้ออื่นและตำแหน่งข้อสอบแต่ละข้อจะอยู่ที่ใดก็ได้โดยจะไม่มีผลต่อการตอบข้อสอบ ทั้งนี้ถ้าสามารถทดสอบได้ว่าแบบสอบนั้นมีความเป็นเอกมิติ ก็สามารถสรุปได้ว่าแบบสอบนั้นมีความเป็นอิสระระหว่างข้อสอบด้วย

2.2) ความเป็นอิสระระหว่างผู้สอบ หมายความว่า ผู้ตอบแต่ละคนตอบข้อสอบแต่ละข้อ อย่างเป็นอิสระจากกัน กล่าวคือผลการตอบสนองข้อสอบของผู้สอบคนหนึ่งจะต้องไม่มีผลต่อการตอบสนองข้อสอบของผู้สอบคนอื่น

3) โค้งคุณลักษณะข้อสอบสามารถใช้อธิบายพฤติกรรมการตอบสนองข้อสอบ (Item Characteristic Curves : Item Response Models) ความสัมพันธ์ระหว่างระดับความสามารถของผู้ตอบ (θ_p) กับโอกาสที่จะตอบข้อสอบถูก [$P_i(\theta)$] สามารถแสดงได้ด้วยโค้งคุณลักษณะข้อสอบ (ICC_S) โดย ICC_S เป็นฟังก์ชันคณิตศาสตร์ที่แสดงด้วยโมเดลการตอบสนองข้อสอบ ที่นิยมใช้กันมี 3 รูปแบบคือ 1,2, หรือ 3 พารามิเตอร์ ในการเลือกใช้ขึ้นอยู่กับลักษณะข้อมูลและความเชื่อในโมเดลที่ใช้อธิบายความสัมพันธ์ดังกล่าว

4) ข้อสอบที่ใช้ต้องไม่เป็นข้อสอบประเภทความเร็ว ผู้สอบทุกคนควรมีโอกาสทำข้อสอบทุกข้อ เพื่อให้คะแนนรวมจากการสอบประกอบกับลักษณะของข้อสอบเป็นตัวประมาณค่าความสามารถที่แท้จริงของผู้สอบโดยไม่มีข้อจำกัด เกี่ยวกับปัจจัยด้านเวลาที่ใช้ในการสอบ

3.2 แบบจำลองที่ใช้ในทฤษฎีการตอบสนองข้อสอบ (Item Response Model) ทฤษฎีการตอบสนองข้อสอบได้มีการพัฒนาแบบจำลองขึ้นมาหลายแบบด้วยกัน โดยแบบจำลองแต่ละแบบนั้นจะต่างกันในรูปของฟังก์ชันทางคณิตศาสตร์ และจำนวนพารามิเตอร์ที่กล่าวถึง ซึ่งสามารถจำแนกได้เป็น 3 ประเภทใหญ่ ๆ คือ (Hambleton and Swaminathan, 1985)

1) แบบจำลองที่ประยุกต์ใช้กับการให้คะแนนแบบ 2 ค่า (Binary Item) คือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน สามารถแบ่งออกตามลำดับขั้นการพัฒนาได้ 2 ระยะ

1.1) แบบจำลองในระยะเริ่มแรก (ค.ศ. 1943-1968) ได้แก่ Guttman Perfect Scale, Latent Distance Model และ Linear Model แบบจำลองเหล่านี้ เป็นแบบจำลองที่ไม่ค่อยจะมี

ความสอดคล้องกับข้อมูลจากการสอบในสถานการณ์จริงมากนัก แต่ก็เป็นแบบจำลองที่มีคุณค่าต่อการพัฒนาแบบจำลองที่เหมาะสมในระยะเวลาต่อมา

1.2) แบบจำลองที่ได้รับการพัฒนาในระยะหลัง (ค.ศ. 1952-1982) ได้แก่ One, Two, Three-Parameter Normal Ogive Model และ One, Two,Three ,Four-Parameter Logistic Model เป็นแบบจำลองที่ได้รับการปรับปรุงให้มีความสอดคล้องกับข้อมูลจากการสอบในสถานการณ์จริงมากขึ้นสามารถนำไปปฏิบัติได้ง่าย และมีผู้สนใจนำไปใช้กันอย่างแพร่หลาย

2) แบบจำลองที่ประยุกต์ใช้กับการให้คะแนนที่มากกว่า 2 ค่า (Multichoto-mously Scored) ได้แก่ Norminal Reponse Model, Grade Response Model และ Partial Credit Model ซึ่งแบบจำลองเหล่านี้ มีจุดหมายเพื่อเพิ่มความแม่นยำในการประมาณค่าความสามารถของผู้เข้าสอบ โดยใช้ประโยชน์จากสารสนเทศ ทั้งจากการตอบข้อสอบถูก หรือผิด

3) แบบจำลองที่ประยุกต์ใช้กับการให้คะแนนแบบต่อเนื่อง (Continuous) ได้แก่ Continuous Response Model ที่พัฒนาโดย Samejima ในปี 1972 อันจะเป็นประโยชน์ต่อการประมาณค่าความสามารถของผู้เข้าสอบทางจิตวิทยา หรือผู้สนใจศึกษาทางด้านที่เกี่ยวกับเจตคติ (Attitude) ซึ่งต้องมีการตอบสนองต่อข้อสอบในมาตราที่ต่อเนื่อง (Continuous Scale)

3.3 ลักษณะเด่นของทฤษฎีการตอบสนองข้อสอบ

เมื่อแบบจำลองการตอบสนองต่อข้อสอบและข้อมูลจากผลการสอบ มีความเหมาะสมกันแล้ว ทฤษฎีการตอบสนองข้อสอบ จะมีความเหนือกว่า (Superiority) ทฤษฎีการทดสอบแบบดั้งเดิม ดังนี้ (Hambleton and Swaminathan 1985:10-13) คือ

3.3.1 ความเป็นอิสระจากกลุ่มผู้เข้าสอบ (Sample-free) นั่นคือไม่ว่าจะนำข้อสอบไปใช้สอบกับบุคคลใด โค้งลักษณะข้อสอบก็จะคงเดิม

3.3.2 ความเป็นอิสระจากกลุ่มข้อสอบ (Item-free) ในการประมาณค่าความสามารถ (θ) ของผู้เข้าสอบ จะใช้ข้อสอบชุดใดก็ได้ จำนวนเท่าใดก็ได้ ซึ่งบางครั้งอาจใช้ข้อสอบเพียง 3-5 ข้อ ก็สามารถประมาณค่าความสามารถที่แท้จริงของผู้เข้าสอบได้แล้ว ทั้งนี้ข้อสอบต้องได้รับการคัดเลือกมาจากคลังข้อสอบขนาดใหญ่ที่ข้อสอบแต่ละข้อระบุคุณลักษณะที่วัดได้

3.3.3 ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (Invariant of item Parameter) กล่าวคือ ไม่ว่าจะประมาณค่ากับกลุ่มตัวอย่างที่มีความสามารถระดับใดก็ตาม ค่าพารามิเตอร์จะไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ทั้งนี้เพราะ ฟังก์ชันการตอบสนองต่อข้อสอบ ซึ่งเป็นฟังก์ชันที่ใช้อธิบายความสัมพันธ์ระหว่างพฤติกรรมการตอบสนองต่อข้อสอบกับ ระดับความสามารถนั้น

เป็นฟังก์ชันการถดถอย (Regression function) ของคะแนนจากการสอบบนความสามารถ ซึ่งตามทฤษฎีสถิติ ฟังก์ชันการถดถอยจะไม่แปรเปลี่ยนไปตามการแจกแจงความถี่ของตัวแปรทำนาย ดังนั้น ค่าพารามิเตอร์ต่างๆ ในสมการถดถอย ซึ่งจะเป็นตัวกำหนดรูปร่างของเส้นถดถอยจึงไม่เปลี่ยนแปลงด้วย (Lord,1980)

ในประเด็นของความไม่แปรเปลี่ยนของค่าพารามิเตอร์นี้ ขึ้นอยู่กับเงื่อนไข 2 ประการ ประการแรกคือ ความสามารถ (Ability) ที่กล่าวถึง ต้องสามารถนิยามได้ชัดเจนและวัดได้ด้วยข้อสอบ อีกประการหนึ่งความสามารถที่วัดนั้นจะต้องมีความคงที่ (statics) ภายในช่วงระยะเวลาหนึ่ง นอกจากนี้ ลอร์ด (Lord,1980) ยังได้อธิบายเพิ่มเติมว่าการไม่แปรเปลี่ยนของค่าพารามิเตอร์นั้นมิได้หมายความว่า ค่าพารามิเตอร์ที่ได้จากการประมาณค่า โดยใช้กลุ่มตัวอย่างที่แตกต่างกันจะมีค่าเท่ากันเสมอ ค่าพารามิเตอร์ของข้อสอบจะเท่ากันหรือไม่ ขึ้นอยู่กับลักษณะของแบบสอบบางประการ เช่น ถ้าเลือกมาตรวัดที่มีจุดเริ่มต้นเดียวกัน และหน่วยในการวัดหน่วยเดียวกันแล้ว ค่าพารามิเตอร์ที่ประมาณได้จากกลุ่มตัวอย่างที่แตกต่างกัน จะมีค่าเท่ากัน ในทางตรงกันข้าม ถ้าหากเลือกมาตรวัดที่มีจุดเริ่มต้น และมีหน่วยในการวัดแตกต่างกันแล้ว การไม่แปรเปลี่ยนของค่าพารามิเตอร์จะหมายความว่า ค่าพารามิเตอร์ที่ประมาณได้จากกลุ่มตัวอย่างที่ต่างกัน ของข้อสอบชุดหนึ่งจะมีความสัมพันธ์กันในเชิงเส้นตรง

<u>วิธีประมาณค่าพารามิเตอร์</u>

การประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบมีหลายวิธี เช่น วิธีการของเบย์ วิธีฮิวริสติก วิธีแมกซิมัมไลค์ลิฮูด วิธีมาร์จินอลแมกซิมัมไลค์ลิฮูด เป็นต้น เนื่องจากผู้วิจัยศึกษาวิธีประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคล โดยการประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด จึงขอเสนอรายละเอียดเฉพาะวิธีการประมาณค่าแบบนี้ ดังนี้

<u>วิธีแมกซิมัมไลค์ลิฮูด</u> (Maximum Likelihood Estimation)

แนวคิดของวิธีแมกซิมัมไลค์ลิฮูด มีว่า การประมาณค่าพารามิเตอร์ ทำได้โดยอาศัยผลที่ได้จากตัวอย่างสุ่ม ที่เลือกมาจากการแจกแจงที่ทราบรูปแบบของฟังก์ชันความหนาแน่น แต่ไม่ทราบค่าพารามิเตอร์ จึงนำโอกาสที่เราจะเลือกตัวอย่างและวัดค่าได้มาใช้ในการประมาณค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูดเป็นการหาค่าประมาณของค่าพารามิเตอร์ที่จะทำให้ฟังก์ชันไลค์ลิฮูด มีค่าสูงสุด ซึ่งโดยปกติจะทำการหาค่าพารามิเตอร์ที่ทำให้ลอการิทึม (Logorithm) ของฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด ทั้งนี้เพราะค่าประมาณของพารามิเตอร์ที่ทำให้ฟังก์ชันทั้ง 2 ฟังก์ชัน มีค่าสูงสุดเป็นค่าเดียวกัน แต่การหาค่าประมาณที่ทำให้ลอการิทึมของฟังก์ชันไลค์ลิฮูดมีค่าสูงสุดนั้นทำได้ง่ายกว่า

โดยทั่วไปสมการไลค์ลิฮูดนี้ ไม่เป็นสมการเส้นตรง การหาค่า ที่ทำให้ มีค่าสูงสุดจึงไม่สามารถหาได้ด้วยการใช้วิธีการอย่างง่ายแต่หาได้โดยใช้วิธีของ Newton - Raphson (Newton - Raphson) (Lord, 1980) ซึ่งเป็นการหาค่าประมาณโดยการทำซ้ำ ๆ (Iterative) จนมีค่าคงที่ และมีขั้นตอน ดังต่อไปนี้

ขั้นที่ 1 กำหนดค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถและค่าพารามิเตอร์ของข้อสอบ

ขั้นที่ 2 ประมาณค่าความสามารถของผู้เข้าสอบแต่ละคนโดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่ ประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณความสามารถ จะมีค่าเข้าใกล้ค่าใดค่าหนึ่ง (convergence) หรือการประมาณค่าในครั้งที่ติดต่อกันมีค่าแตกต่างกันน้อยกว่าค่าคงที่ที่กำหนดไว้ เช่น 0.001

ขั้นที่ 3 ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ โดยกำหนดให้ค่าความสามารถของผู้เข้าสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่ และประมาณค่าซ้ำ (Iterative) จนกว่าค่าประมาณจะลู่เข้าใกล้ค่าใดค่าหนึ่ง (Convergence)

ขั้นที่ 4 ประมาณค่าซ้ำในขั้นที่ 2 และขั้นที่ 3 จนกว่าค่าประมาณ จะมีค่าคงที่ และมีความถูกต้องเพียงพอหรือทำให้ค่าลอการิทึมของฟังก์ชันไลค์ลิฮูดมีค่าสูงที่สุด

ค่าอนุพันธ์อันดับที่ 1 และค่าอนุพันธ์อันดับที่ 2 ของค่าพารามิเตอร์ ที่ได้จากการประมาณโดยวิธีแมกซิมัมไลค์ลิฮูด เป็นการประมาณค่าซ้ำ ๆ หลายครั้ง และเพื่อให้ได้ค่าคงที่เร็วขึ้น ค่าความสามารถจะถูกปรับให้มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ซึ่งจะเป็นผลให้ค่า a, b และ c ต้องถูกปรับตามไปด้วย (Hambleton and Swaminathan, 1985)

ปัจจุบันในการคำนวณเพื่อประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด สามารถทำได้โดยสะดวก เนื่องจาก ได้มีการพัฒนาโปรแกรมสำหรับวิเคราะห์ข้อสอบไว้หลายโปรแกรม เช่น โปรแกรมโลจีสต์ 5.0 โปรแกรม BILOG โปรแกรม MULTILOG โปรแกรม PARSCALE

คุณสมบัติที่สำคัญของวิธีแมกซิมัมไลค์ลิฮูด

1. มีความคงที่ เมื่อจำนวนของกลุ่มผู้เข้าสอบและข้อสอบเพิ่มขึ้น การประมาณก็จะมีความคงที่ไปสู่ค่าที่แท้จริง
2. ฟังก์ชันของค่าสถิติให้สารสนเทศสารสนเทศทั้งหมดเกี่ยวกับค่าพารามิเตอร์
3. มีประสิทธิภาพซึ่งวิธีการประมาณค่าพารามิเตอร์จะมีความแปรปรวนน้อยที่สุด
4. มีลักษณะของการกระจายเข้าใกล้แบบปกติ

ในกรณีที่กลุ่มตัวอย่างผู้เข้าสอบและข้อสอบมีขนาดเล็ก คุณสมบัติข้างต้นจะเป็นจริงเมื่อทำการประมาณค่าความสามารถของบุคคล หรือ ประมาณค่าพารามิเตอร์ของข้อสอบเพียงอย่างใดอย่างหนึ่งเท่านั้น แต่ถ้าผู้เข้าสอบ และจำนวนข้อสอบเพิ่มขึ้น ในการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้เข้าสอบพร้อมๆ กัน ค่าประมาณที่ได้จะมีค่าเข้าใกล้ค่าพารามิเตอร์จริง

ข้อจำกัดของกระบวนการประมาณค่าพารามิเตอร์ ของวิธีแมกซิมัมไลค์ลิฮูด

1. ในการประมาณค่าพารามิเตอร์ในขั้นที่ 2, 3 โดยใช้ค่าอนุพันธ์อันดับที่ 2 เป็นตัวหารในกระบวนการนิวตัน ราพสัน นั้น มีโอกาสที่ค่าประมาณที่ได้จะไม่ลู่เข้าสู่ค่าคงที่
2. เมื่อสมการไลค์ลิฮูดเป็นระบบสมการที่ไม่ใช่เชิงเส้นตรง ทำให้สามารถหารากของสมการที่ทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุดได้หลายค่า แต่ค่าเหล่านี้ไม่สามารถนำไปใช้เป็นค่าพารามิเตอร์ที่แท้จริงได้
3. บางครั้งค่าพารามิเตอร์ หรือค่าที่ได้จากการประมาณ ไม่ตกอยู่ในขอบเขตของค่าพารามิเตอร์ กล่าวคือ อาจมีค่าใดค่าหนึ่งอยู่ภายนอกขอบเขตที่ยอมรับได้ ในกรณีเช่นนี้ต้องมีการกำหนดขอบเขตจำกัดของค่าประมาณไว้ เพื่อให้ค่าประมาณที่ได้ไม่สูงหรือต่ำเกินไปนัก แต่การกระทำเช่นนี้ เป็นจุดอ่อนของกระบวนการประมาณค่าด้วยวิธีแมกซิมัมไลค์ลิฮูด โดยเฉพาะในแบบจำลอง 2-3 พารามิเตอร์ ซึ่งทำให้เกิดปัญหาตามมา เกี่ยวกับความตรง (Validity) ของค่าที่ประมาณได้

4. เนื่องจากการประมาณค่าพารามิเตอร์ในแบบจำลอง 3 พารามิเตอร์ มีสมการหลายสมการที่ต้องหารากที่ทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด วิธีของนิวตัน ราฟสัน จึงจำเป็นต้องใช้เครื่องคอมพิวเตอร์ และโปรแกรม ที่มีประสิทธิภาพในการคำนวณ

3.2 แนวคิดเกี่ยวกับราส์ชโมเดล

เป็นโมเดลโลจิสติกแบบ 1 พารามิเตอร์ (one-parameter logistic mode) จอร์จ ราส์ช (Georg Rasch) นักคณิตศาสตร์ชาวเดนมาร์กได้คิดขึ้นและเสนอแนวคิดนี้ในปี ค.ศ. 1960 (Wright 1977) วิธีการประมาณค่าพารามิเตอร์ของราส์ชมีข้อดีคือ คุณภาพของเครื่องมือไม่ได้ขึ้นอยู่กับลักษณะของเครื่องมือและกลุ่มบุคคลที่วัด ตามแนวคิดของราส์ชโมเดลนั้น โอกาสที่ผู้สอบจะตอบข้อสอบถูกหรือผิดขึ้นอยู่กับปัจจัย 2 อย่าง คือ ระดับความสามารถ (ability parameter, θ) และระดับความยากของข้อทดสอบ (difficulty parameter,β) เช่น ถ้า $\theta = 0.5$ และ $\beta = 0.5$ โอกาสที่คนผู้นั้นจะสามารถตอบข้อสอบนั้นได้ถูกต้องมี 50 % ถ้าหากความสามารถของบุคคล (θ) น้อยกว่า ระดับความยาก (β) ของข้อสอบ และในทำนองเดียวกัน ถ้าหากว่าความสามารถของบุคคล (θ) มากกว่าระดับความยาก (β) แล้ว โอกาสที่คนผู้นั้นจะตอบข้อสอบได้ถูกก็มีมากกว่า 50% ดังนั้นตามแนวความคิดของแบบราส์ชนั้นค่าพารามิเตอร์ที่เกี่ยวข้องมีเพียง 2 ค่าเท่านั้น ไม่มีค่าอำนาจจำแนกหรือโอกาสของการเดาเข้ามาเกี่ยวข้องด้วย เพราะเป็นข้อตกลงเบื้องต้นว่าข้อสอบที่นำมาใช้กับราส์ชโมเดลนั้นจะต้องมีอำนาจจำแนกเท่า ๆ กัน และมีลักษณะที่ทำให้เกิดการเดาได้น้อยที่สุด

4.1 วิธีการประมาณค่าพารามิเตอร์ในรูปแบบของราส์ช

การประมาณค่าพารามิเตอร์ในรูปแบบของราส์ชเพื่อการวิเคราะห์ข้อสอบสามารถประมาณค่าได้หลายวิธี คือ (Wright,1979)

4.1.1 The Log Method ราส์ชได้พัฒนาขึ้นในปี ค.ศ. 1953 เป็นวิธีที่ง่ายใช้ได้โดยประมาณค่าพารามิเตอร์ในรูป log ของความสำเร็จ ซึ่งเป็นวิธีที่เหมาะสมกับแบบสอบที่มีค่าอำนาจจำแนกเท่ากับ 1 และใช้จำนวนคนในการประมาณค่าพารามิเตอร์เป็นจำนวนมาก

4.1.2 The PAIR Method เป็นวิธีการประมาณค่าพารามิเตอร์โดยใช้ข้อสอบจับคู่เป็นคู่ ๆ เช่น 1-2 , 1-3 , 1-4 , 2-4 เป็นต้น โดยเอาจำนวนคนที่พยายามตอบ 2 ข้อ แต่ทำถูก

เพียงข้อเดียวมา วิเคราะห์ เพื่อหา sample free ดังนั้นจึงเหมาะที่จะใช้ประมาณค่าแบบสอบที่มีจำนวนข้อสอบไม่มากนัก

4.1.3 The FCON Method เป็นวิธีประมาณค่าพารามิเตอร์โดยใช้ข้อสอบทุกข้อและผู้สอบทุกคนมาวิเคราะห์ เหมาะสำหรับแบบสอบที่มีจำนวนข้อสอบน้อยกว่า 30 ข้อ ถ้ามากกว่า 30 ข้อ จะทำให้การประมาณค่าพารามิเตอร์มีความคลาดเคลื่อน แต่ภายหลังได้มีการปรับปรุงใหม่สามารถใช้กับแบบสอบที่มีความยาว 60-70 ข้อ ได้

4.1.4 The UCON Method เป็นวิธีประมาณค่าพารามิเตอร์ด้วยวิธี Unconditional Maximum Likelihood Method ใช้กับแบบสอบที่การกระจายของความสามารถของบุคคลและความยากของแบบสอบเป็นการกระจายแบบโค้งปกติ คำนวนค่าพารามิเตอร์ของผู้สอบและข้อสอบ โดยไม่มีการกำหนดค่าใดไว้ล่วงหน้า

4.1.5 The PROX Method เป็นวิธีประมาณค่าพารามิเตอร์ด้วยวิธีการของ Cohen's Approximation ใช้กับแบบสอบที่มีการกระจายของความสามารถของบุคคลและความยากของแบบสอบเป็นลักษณะเบ้ หรือมีแนวโน้มที่จะเป็นโค้งปกติ โดยมากการประมาณค่าด้วยวิธี PROX และ UCON จะให้ผลเหมือนกันต่างกันที่ค่าความคลาดเคลื่อนมาตรฐานเพียงเล็กน้อย

4.1.6 The UFORM Method ใช้กับแบบสอบที่มีการกระจายของค่าความยากของข้อสอบเป็นแบบ Uniform

5.วิธีการตรวจสอบคุณภาพของแบบสอบ

ทฤษฎีการทดสอบ มี 2 ทฤษฎี คือ ทฤษฎีการตอบสนองข้อสอบ (IRT) และ ทฤษฎีการทดสอบแบบดั้งเดิม (CTT) ดังนั้นการตรวจสอบคุณภาพของแบบสอบจึงสามารถตรวจสอบได้ ตามแนวทฤษฎีทั้งสองทฤษฎี ดังนี้

5.1 คุณภาพของแบบสอบตามทฤษฎีการตอบสนองข้อสอบ

ค่าพารามิเตอร์ของข้อสอบตามทฤษฎีการตอบสนองข้อสอบ สำหรับโมเดลแบบ 3 พารามิเตอร์ มี 3 ตัว คือ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c)

ค่าสถิติที่บ่งบอกคุณภาพของแบบสอบตามทฤษฎีการตอบสนองข้อสอบ มี 4 ค่า คือ

5.1.1 ค่าฟังก์ชันสารสนเทศของข้อสอบ (IIF)

5.1.2 ค่าฟังก์ชันสารสนเทศของแบบสอบ (TIF)

5.1.3 ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ (SE)

5.1.4 ค่าประสิทธิภาพสัมพัทธ์ (Relative Efficiency)

5.1.1 ค่าฟังก์ชันสารสนเทศของข้อสอบ (Item information funetion) เป็นดัชนีผสมที่สร้างจากดัชนีคุณลักษณะของข้อสอบ เช่น ค่าความยาก ค่าอำนาจจำแนก และค่าความแปรปรวนของคะแนนรายข้อ มีสูตรการคำนวณ ดังนี้

$$I_i(\theta) = \frac{(P_i')^2}{P_i Q_i}$$

เมื่อ P_i คือ โอกาสที่ผู้สอบที่มีระดับความสามารถ θ จะสอบข้อสอบได้ถูก

Q_i คือ โอกาสที่ผู้สอบที่มีระดับความสามารถ θ จะตอบข้อสอบผิด

P_i' คือ ความชันของฟังก์ชันการตอบสนองข้อสอบ ณ ระดับความสามารถ θ

5.1.2 ค่าฟังก์ชันสารสนเทศของแบบสอบ (Test information function) เป็นคุณภาพของแบบสอบตามทฤษฎีการตอบสนองข้อสอบ เนื่องจากการวิเคราะห์ข้อสอบตามทฤษฎีการตอบสนองข้อสอบ จะใช้แบบแผนการตอบข้อสอบเป็นรายข้อในการประมาณค่าความสามารถของผู้สอบ คุณภาพของแบบสอบจึงสามารถพิจารณาจากความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ ซึ่งดัชนีที่ใช้ คือ ค่าฟังก์ชันสารสนเทศของแบบสอบ ซึ่งเกิดจากผลรวมของค่าฟังก์ชันสารสนเทศของข้อสอบทุกข้อในแบบสอบชุดนั้น มีสูตรการคำนวณ ดังนี้

$$I(\theta) = \sum_{i=1}^{k} I_i(\theta)$$

เมื่อ $I_i(\theta)$ คือ ค่าสารสนเทศของข้อสอบข้อที่ i

5.1.3 ค่าความคลาดเคลื่อนมาตราฐานของการประมาณค่าความสามารถ (Standard error of estimates : SE) ความคลาดเคลื่อนมาตราฐานของการประมาณค่าความสามารถ

I17558876

จะมีค่าเท่ากับ ส่วนกลับของรากที่สองของค่าฟังก์ชันสารสนเทศของแบบสอบ มีสูตรการคำนวณ ดังนี้

$$SE(\theta) = \frac{1}{\sqrt{I(\theta)}}$$

5.1.4 ค่าประสิทธิภาพสัมพัทธ์ของแบบสอบ (Relative Efficiency)

การคำนวณประสิทธิภาพสัมพัทธ์ของแบบสอบเป็นการหาอัตราส่วนระหว่างค่าฟังก์ชันสารสนเทศของแบบสอบต่างฉบับที่มุ่งวัดคุณลักษณะเดียวกัน ณ ระดับหรือช่วงความสามารถเดียวกันของผู้สอบของแบบสอบทั้งสองฉบับ มีสูตรการคำนวณ ดังนี้ (Lord,1980)

$$RE\ (Y,X) = \frac{I_Y(\theta)}{I_X(\theta)}$$

เมื่อ RE (Y,X) คือ ประสิทธิภาพสัมพัทธ์ของแบบสอบ Y เมื่อเทียบกับ X

$I_Y(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของแบบสอบ Y เมื่อผู้สอบมีระดับความสามารถ θ

$I_X(\theta)$ คือ ค่าฟังก์ชันสารสนเทศของแบบสอบ X เมื่อผู้สอบมีระดับความสามารถ θ

5.2 การตรวจสอบคุณภาพของแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิม

ในการศึกษาคุณภาพของแบบสอบตามทฤษฎีการทดสอบแบบดั้งเดิม มีการตรวจสอบคุณภาพในสองด้าน คือ ตรวจสอบความเที่ยง (reliality) และความตรง (validity) (Rescoe,1969)

5.2.1 ความเที่ยงของแบบสอบ มี 3 แบบ คือ

1) แบบสอบซ้ำ (Measure of Stability) การสอบซ้ำ (Test-Retest) เป็นการนำข้อสอบชุดเดียวกันไปสอบเด็กกลุ่มเดียวกัน 2 ครั้งในเวลาห่างกันพอสมควร แล้วนำคะแนนทั้ง 2 ชุดนั้นมาหาความสัมพันธ์กัน โดยการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน ค่าที่ได้คือความเที่ยงของแบบสอบ

2) การใช้แบบสอบคู่ขนาน (Measure of Equivalence)

แบบสอบคู่ขนาน (Parallel tests หรือ Equivalence tests) หมายถึงแบบสอบ 2 ชุด ที่มีลักษณะและคุณภาพใกล้เคียงกันมากที่สุด ทั้งด้านเนื้อหา ความยากง่าย อำนาจจำแนก ลักษณะคำถาม และจำนวนข้อคำถาม จนอาจกล่าวได้ว่าเป็นแบบสอบฉบับเดียวกัน สามารถใช้ทดแทนกันได้ การใช้แบบสอบคู่ขนานนี้เป็นการแก้ปัญหาข้อจำกัดต่าง ๆ ของการหาค่าความเที่ยงของข้อสอบโดยการสอบซ้ำ การหาความเที่ยงของข้อสอบทำได้โดย นำข้อสอบคู่ขนานไปทดสอบนักเรียนกลุ่มเดียวกันทั้ง 2 ฉบับในเวลาเดียว แล้วนำคะแนนจากการทำข้อสอบ 2 ชุดนี้มาหาความสัมพันธ์กัน ก็จะได้ค่าความเที่ยงของข้อสอบทั้ง 2 ฉบับ

3) แบบวัดความคงที่ภายใน (Measure of Internal Consistency) เป็นการตรวจสอบว่าข้อสอบในแบบสอบนั้นวัดในเรื่องเดียวกันหรือไม่ ถ้าวัดในเรื่องเดียวกันก็น่าจะมีความคงที่ในการวัดสูง โดยมีการสอบเพื่อหาคุณภาพของแบบสอบเพียงครั้งเดียว

3.1) วิธีแบ่งครึ่งข้อสอบ (Split-half) โดยการนำแบบสอบที่ต้องการหาค่าความเที่ยงไปทดสอบกับกลุ่มผู้สอบ ตรวจให้คะแนน แบ่งคะแนนรวมของแต่ละคนออกเป็นสองส่วนหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน ซึ่งจะได้ค่าความเที่ยงเพียงครึ่งฉบับ ปรับให้เป็นค่าความเที่ยงของแบบสอบทั้งฉบับโดยสูตรของสเปียร์แมน บราวน์

3.2) แบบคูเดอร์ริชาร์ดสัน (Kuder-Richardson Estimation) การหาคุณภาพด้านความเที่ยงของ คูเดอร์ - ริชาร์ดสัน (Kuder-Richardson method) เป็นการหาความคงที่ภายในของข้อสอบฉบับหนึ่ง ๆ เฉพาะแบบสอบซึ่งตรวจให้คะแนนแบบ 2 ค่า และแบบสอบจะต้องมีความเป็นเอกมิติ สูตรการคำนวนมี 2 สูตร คือ KR-20 และ KR-21

3.3) แบบสัมประสิทธิ์แอลฟา (Alfa Coefficent) วิธีนี้พัฒนาโดยจากครอนบาค (Cronbach) โดยได้พัฒนาสูตรคูเดอร์ริชาดสัน 20 เป็นสัมประสิทธิ์แอลฟา เพื่อให้ใช้ได้กับการให้คะแนนที่ไม่เป็นระบบ 0-1 เช่น แบบสอบอัตนัย มาตรประมาณค่า(Rating Scale) เป็นต้น

3.4) สูตรของสเปียร์แมน-บราวน์ (The Spearman-Brown Formula) การคำนวณความเที่ยงของแบบสอบด้วยสูตรของ สเปียร์แมน บราวน์ จะใช้สารสนเทศเกี่ยวกับความเที่ยงขององค์ประกอบที่คู่ขนานกัน นอกจากนี้ เปียร์แมน บราวน์ ยังใช้ในการตีความหมาย ผลของการเปลี่ยนแปลง ความยาวของแบบสอบที่มีต่อความเที่ยง สามารถปรับความยาวของแบบสอบที่ต่างกันให้สามารถเปรียบเทียบค่าความเที่ยงกันได้ (Allen and Yen,1979)

3.5) วิธีวิเคราะห์ความแปรปรวนของฮอยท์ วิธีนี้ได้รับการพัฒนาจากฮอยท์ ซึ่งได้ดัดสูตรหาความเที่ยงโดยวิเคราะห์ความแปรปรวน (Analysis of Variance) แหล่ง

ความแปรปรวนมาจากตัวข้อสอบ ตัวผู้สอบ และความคลาดเคลื่อนในการวัดวิธีนี้กับการให้คะแนน 0-1 หรือแบบอื่นๆ เช่น แบบสอบอัตนัย มาตรประมาณค่า (Rating Scale) โดยที่การคำนวณจะให้ค่าความเที่ยงเช่นเดียวกับสูตร KR-20

ในการศึกษาครั้งนี้ ผู้วิจัยตรวจสอบความเที่ยงตามทฤษฎีการทดสอบแบบดั้งเดิม โดยใช้สูตร KR-20 เนื่องจากเป็นการให้คะแนนแบบ 2 ค่า และทดสอบความแตกต่างของค่าความเที่ยงโดยใช้ค่าฟิชเชอร์ - ซี

5.2.2 ความตรงของแบบสอบ (Validity)

ความตรง หมายถึง ความแม่นยำของเครื่องมือในการวัดสิ่งที่ต้องการจะวัด และคะแนนที่ได้จากแบบสอบที่มีความตรงสูง สามารถบอกถึงสภาพที่แท้จริง และพยากรณ์ได้ถูกต้องแม่นยำ

ความตรงของแบบสอบมีหลายชนิด ซึ่งแต่ละชนิดแบ่งตามธรรมชาติและจุดมุ่งหมายของการวัด แต่โดยทั่วไปแล้วแบ่งออกได้เป็น 3 ชนิด คือ

1) ความตรงตามเนื้อหา (Content Validity)

2) ความตรงตามเกณฑ์สัมพันธ์ (Criterion-Related Validity)

3) ความตรงเชิงทฤษฎี (Construct Validity)

1) ความตรงตามเนื้อหา (Content Validity)

ความตรงตามเนื้อหาจะเกี่ยวข้องกับเนื้อหาสาระของข้อสอบในแบบสอบ ซึ่งข้อสอบนั้นต้องเป็นตัวแทนของเนื้อหาสาระและครอบคลุมขอบเขตของพฤติกรรมที่ต้องการวัด ความตรงตามเนื้อหามีความสำคัญอย่างยิ่งในการสอบวัดผลสัมฤทธิ์ทางการเรียน การตัดสินว่าแบบสอบนั้นมีความตรงตามเนื้อหามากน้อยเพียงใด จะต้องพิจารณาทั้งหัวข้อเนื้อหาวิชา และชนิดของพฤติกรรมที่วัดไปพร้อม ๆ กัน การหาค่าความตรงตามเนื้อหา ไม่สามารถคำนวณเป็นตัวเลขได้ โดยทั่วไปจะพิจารณาจากข้อสอบในแบบสอบเปรียบเทียบกับตารางวิเคราะห์เนื้อหา (Table of Speacification) แล้วประมาณว่ามีความตรงตามเนื้อหามากน้อยเพียงใด ผู้ที่จะตัดสินความตรงตามเนื้อหาได้ก็คือ ผู้เชี่ยวชาญในสาขานั้น ๆ

2) ความตรงตามเกณฑ์สัมพันธ์ (Criterion-Related Validity)

ความตรงตามเกณฑ์สัมพันธ์ จะเกี่ยวข้องกับการพิจารณาผลของการวัดจากแบบสอบว่าสามารถทำนายได้ดีเพียงใด หรือให้ผลในการวัดได้ตรงตามสภาพความเป็นจริงใน

ปัจจุบันได้ดีเพียงใด โดยการนำผลการวัดที่ได้ไปเปรียบเทียบกับการกระทำที่เกิดขึ้นในอนาคต หรือเทียบกับการกระทำอย่างอื่นที่วัดได้ในเวลาใกล้เคียงกัน ความตรงตามเกณฑ์สัมพันธ์ หาได้โดยการปรียบเทียบคะแนนจากแบบทดสอบกับตัวแปรที่เป็นตัวแปรภายนอกหนึ่งตัวหรือหลายตัว ซึ่งเป็นคุณลักษณะที่กล่าวถึงการเปรียบเทียบค่าสหสัมพันธ์ระหว่างคะแนนทดสอบกับคะแนนเกณฑ์ ความตรงตามเกณฑ์สัมพันธ์แบ่งได้เป็น 2 ชนิด คือ

ก. ความตรงเชิงทำนาย (Predictive Validity) เป็นการหาค่าสหสัมพันธ์ระหว่างคะแนนสอบกับคะแนนเกณฑ์ในอนาคตที่เกิดขึ้น ความตรงชนิดนี้ใช้แบบสอบเพื่อทำนายผลบางอย่างที่จะเกิดขึ้นในเวลาต่อมา เช่น แบบสอบวัดความถนัดจะทำนายผลการเรียนที่เกิดขึ้นภายหลัง ถ้าผู้สอบทำคะแนนได้สูงหมายความว่า ผู้สอบคนนั้นจะประสบความสำเร็จในการเรียน

ข. ความตรงตามสภาพ (Concurent Validity) เป็นการหาค่าสหสัมพันธ์ระหว่างคะแนนสอบกับเกณฑ์ที่เกิดขึ้นในระยะเวลาใกล้เคียงกัน ความตรงชนิดนี้ใช้แบบสอบเพื่อจะดูสถานภาพที่เป็นอยู่ในปัจจุบัน เช่น การต้องการทราบผลสัมฤทธิ์ในการเรียนในปัจจุบันของเด็กในวิชาคณิตศาสตร์ ว่าถูกต้องเพียงใดกับสถานภาพปัจจุบันโดยการหาความสัมพันธ์ระหว่างคะแนน คณิตศาสตร์ กับเกณฑ์ที่เกิดขึ้นในเวลาเดียวกันหรือใกล้เคียงกันว่าสอดคล้องกันเพียงใด ความตรงตามสภาพใช้เพื่อดูว่าแบบสอบที่สร้างขึ้นใหม่นั้นสามารถใช้ได้ดีเพียงใดเมื่อเทียบกับเกณฑ์

3) ความตรงตามทฤษฎี (Construct Validity)

ความตรงตามทฤษฎี หมายถึง ความสามารถในการวัดได้ตรงตามคุณลักษณะทางจิตวิทยาของการเรียนรู้นั้น วิธีการประมาณค่า โดยการศึกษาความสัมพันธ์ระหว่างผลการวัดที่ได้จากเครื่องมือกับคุณลักษณะที่มุ่งวัดโดยอาศัยข้อสนับสนุนจากการวิเคราะห์ด้วยวิธีต่าง ๆ เช่น

3.1) ศึกษาเปรียบเทียบผลการเรียนรู้ระหว่างกลุ่มที่ทราบคุณลักษณะการเรียนรู้แล้ว (Known Group)

3.2) การวิเคราะห์องค์ประกอบ (Factor Analysis)

3.3) ใช้เทคนิคคุณลักษณะหลากวิธีหลาย (Multi-trait Multi-method)

โดยทั่วไป การตรวจสอบคุณภาพของแบบสอบในด้านความตรงตามทฤษฎี จะใช้วิธีการวิเคราะห์องค์ประกอบ ซึ่งผู้วิจัยได้ตรวจสอบความเป็นเอกมิติของแบบสอบ และตรวจสอบความตรงตามทฤษฎี ของแบบสอบที่ศึกษา ด้วยการวิเคราะห์องค์ประกอบ จึงขอกล่าวถึงรายละเอียด เกี่ยวกับการวิเคราะห์องค์ประกอบ

การวิเคราะห์องค์ประกอบ (Factor Analysis)

การวิเคราะห์องค์ประกอบ เป็นวิธีการทางสถิติที่นักวิจัยใช้ในการสร้างองค์ประกอบจากตัวแปรหลายๆที่เกี่ยวข้องสัมพันธ์กัน กล่าวได้ว่า เป็นความสัมพันธ์ระหว่างชุดของตัวแปรที่สังเกตได้ (Observed Variable) กับตัวแปรแฝง (Latent Variable) ซึ่งเป็นคุณลักษณะที่ต้องการศึกษา ผลจากการวิเคราะห์องค์ประกอบจะทำให้ลดจำนวนตัวแปรลง ได้องค์ประกอบ ซึ่งมีลักษณะโครงสร้าง(Structure) และแบบแผน (Pattern) ชัดเจน แปลความหมายและทำความเข้าใจได้ง่าย (นงลักษณ์ วิรัชชัย,2538)

วัตถุประสงค์ในการวิเคราะห์องค์ประกอบ

ในการวิเคราะห์องค์ประกอบนั้น มีจุดมุ่งหมายหลัก 2 ประการ คือ

1. เพื่อบรรยาย หรือค้นหาความสัมพันธ์ระหว่างตัวแปร หรือที่เรียกกันว่า การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis)
2. เพื่อพิสูจน์ สนับสนุน หรือตรวจสอบสมมติฐาน ที่ผู้วิจัยได้ศึกษาไว้ หรือการวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis)

ประเภทของการวิเคราะห์องค์ประกอบ

สามารถแบ่งประเภทของการวิเคราะห์องค์ประกอบ ตามวัตถุประสงค์ในการวิเคราะห์ได้เป็น 2 ประเภท คือ

1. การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory Factor Analysis)
2. การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis)

การวิเคราะห์องค์ประกอบเชิงสำรวจ (Exploratory factor analysis) หมายถึง การวิเคราะห์ว่าตัวแปรที่สังเกตได้มีองค์ประกอบย่อยอะไรบ้าง โดยที่ผู้วิจัยไม่มีสมมติฐานวางไว้ล่วงหน้า ซึ่งมีข้อตกลงเบื้องต้น ดังนี้

1. ตัวประกอบร่วมทุกตัวมีความสัมพันธ์กัน (oblique roatation) หรือ ตัวประกอบร่วมทุกตัวเป็นอิสระจากกัน (orthogonal rotation)
2. ตัวแปรที่สังเกตได้ทุกตัวได้รับอิทธิพลโดยตรงจากทุกตัวประกอบ
3. ตัวแปรที่สังเกตได้ทุกตัวได้รับอิทธิพลจากความคลาดเคลื่อนเพียงตัวเดียว

4. ความคลาดเคลื่อนทุกตัวเป็นอิสระจากกัน และเป็นอิสระจากตัวประกอบทุกตัว

การวิเคราะห์องค์ประกอบเชิงสำรวจเป็นวิธีการที่ใช้ในการตรวจสอบความเป็นเอกมิติของแบบสอบที่สำคัญ (วรนุช แหยมแสง, 2537 ; สุวิมล ติระกานันท์,2537) ทั้งนี้เกณฑ์ในการพิจารณาความเป็นเอกมิติของแบบสอบมีหลายเกณฑ์ ที่ใช้กันอยู่ทั่วไป คือ

1. การใช้ค่าสถิติ ค่าสถิติที่ใช้ในการบ่งบอกความเป็นเอกมิติของแบบสอบ คือ อัตราส่วนของค่าไอเกนขององค์ประกอบที่ 1 และค่าไอเกนขององค์ประกอบที่ 2 (Eigen Ratio: ER) ซึ่ง วรนุช แหยมแสง (2537) ได้ศึกษาพบว่าควรมีค่าประมาณ 3 เท่า

อัตราส่วนของค่าไอเกน (Eiqen Ratio : ER)

$$ER = \frac{E_1}{E_2}$$

เมื่อ ER คือ อัตราส่วนระหว่างค่าไอเกน

E_1 คือ ค่าไอเกนของตัวประกอบที่ 1

E_2 คือ ค่าไอเกนของตัวประกอบที่ 2

2. การทดสอบสกรี (Scree test criterion) องค์ประกอบที่ได้จากการวิเคราะห์องค์ประกอบครั้งหนึ่งๆ จะมีหลายองค์ประกอบ การทดสอบสกรี คือ การนำค่าไอเกนขององค์ประกอบมาพล็อตกราฟ จะได้กราฟสกรี ซึ่งแสดงความแตกต่างของค่าไอเกน เส้นกราฟจะมีความชันในองค์ประกอบแรกๆ และจะค่อยๆลาดลงในองค์ประกอบหลังๆ การพิจารณาความเป็นเอกมิติของแบบสอบ ให้พิจารณา ค่าไอเกนขององค์ประกอบที่ 1 เปรียบเทียบกับค่าไอเกน ขององค์ประกอบอื่นๆ ถ้าค่าไอเกนขององค์ประกอบที่ 1 ต่างจากค่าไอเกนขององค์ประกอบที่เหลือมากๆ และค่าไอเกนขององค์ประกอบที่เหลือมีลักษณะเหมือนๆ กัน จัดว่าแบบสอบนั้นมีความเป็นเอกมิติ (Lord & Novick,1968)

นอกจากสองวิธีการที่กล่าวข้างต้นแล้ว การวิเคราะห์องค์ประกอบเชิงสำรวจ ยังมีวิธีการในการตรวจสอบความเป็นเอกมิติของแบบสอบอีกหลายวิธี เช่น การทดสอบนัยสำคัญขององค์

ประกอบ การทดสอบค่าสหสัมพันธ์ไบซีเรียลของคะแนนรายข้อกับคะแนนรวม การทดสอบค่าการร่วม เป็นต้น เนื่องจาก 2 วิธีการที่กล่าวมา หาความสะดวกและง่ายต่อการปฏิบัติ ผู้วิจัยจึงตรวจสอบความเป็นเอกมิติของแบบสอบที่จำลอง ขั้นด้วยวิธีทั้ง 2 วิธี เพื่อตรวจสอบว่าแบบสอบซึ่งได้จากการจำลองข้อมูลมีความเป็นเอกมิติ ตามข้อตกลงเบื้องต้น ของทฤษฎีการตอบสนองข้อสอบ (IRT) หรือไม่

การวิเคราะห์องค์ประกอบเชิงยืนยัน (Confirmatory Factor Analysis) หมายถึง การวิเคราะห์องค์ประกอบ ที่ผู้วิจัยมีสมมติฐานที่แน่นอนเกี่ยวกับตัวแปรแฝง (Latent Variable) ของตัวแปรที่ศึกษาว่ามีตัวแปรสังเกตได้ (Observed Variable) อะไรบ้าง และทำการวิเคราะห์เพื่อจะได้ยืนยันองค์ประกอบนั้น

การวิเคราะห์องค์ประกอบเชิงยืนยัน เป็นการวิเคราะห์องค์ประกอบที่มีการปรับปรุงจุดอ่อนของการวิเคราะห์องค์ประกอบเชิงสำรวจ และมีความสมเหตุสมผลมากกว่า เช่น ยอมให้ความคลาดเคลื่อนมีความสัมพันธ์กันได้ การวิเคราะห์องค์ประกอบมีพื้นฐานทางทฤษฎี หรือสมมติฐานที่ชัดเจน มีกระบวนการตรวจสอบความตรงของโมเดลที่ชัดเจน แปลผลได้ง่ายและสะดวก ผลการวิเคราะห์ข้อมูลให้ค่าประมาณพารามิเตอร์ และการทดสอบนัยสำคัญ ดังนั้น การวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) จึงเป็นวิธีการในการตรวจสอบความตรงเชิงทฤษฎีของแบบสอบได้เป็นอย่างดี (อุทุมพร จามรมาน, 2532 ; นงลักษณ์ วิรัชชัย, 2538)

ขั้นตอนในการวิเคราะห์องค์ประกอบ

การวิเคราะห์องค์ประกอบแบบ EFA และ CFA มีขั้นตอนในการวิเคราะห์เหมือนกัน คือ

1. การเตรียมเมตริกซ์สหสัมพันธ์
2. การสะกัดองค์ประกอบขั้นต้น
3. การหมุนแกน
4. การสร้างสเกลองค์ประกอบ

จากการศึกษาของ สุวิมล ติรกานันท์ (2537) พบว่าในการวิเคราะห์องค์ประกอบเชิงยืนยันเพื่อตรวจสอบความเป็นเอกมิติของแบบสอบนั้น การเตรียมเมตริกซ์สหสัมพันธ์ด้วย เมตริกซ์ความแปรปรวน-ความแปรปรวนร่วม จะทำให้ดัชนี χ^2 , AGFI , NFI และ CN มีความคงที่ดีกว่าการเตรียมเมตริกซ์สหสัมพันธ์ด้วยวิธีอื่นๆ

ผู้วิจัยจึงตรวจสอบคุณภาพของแบบสอบที่ได้จากการวิเคราะห์ความเหมาะสมของบุคคล ในด้านความตรงทางทฤษฎี โดยการวิเคราะห์องค์ประกอบเชิงยืนยัน (CFA) ด้วยโปรแกรม LISREL 8.10 ด้วยเมตริกซ์ความแปรปรวน-ความแปรปรวนร่วม เพื่อเปรียบเทียบความตรงของโมเดลด้วยดัชนีซึ่งประมาณค่าพารามิเตอร์ และทดสอบนัยสำคัญไว้แล้ว

5.รายงานการวิจัยที่เกี่ยวข้อง

แนวความคิดเกี่ยวกับการวิเคราะห์ บุคคลที่เหมาะสมกับโมเดลการตอบข้อสอบหรือการวิเคราะห์แบบแผนการตอบข้อสอบ มีการวิจัยศึกษากันตั้งแต่ ค.ศ. 1940 (Lumsden, 1976 ; cited in Smith, 1986) โดยเริ่มตั้งแต่การวิเคราะห์ตามแนวทฤษฎีการทดสอบแบบดั้งเดิมตามแนวความคิดของกัตแมน และมีการพัฒนาดัชนีเพื่อตรวจสอบความเหมาะสมของบุคคลตามทฤษฎีการตอบสนองข้อสอบ ซึ่งผู้วิจัยได้ศึกษาค้นคว้ารายงานการวิจัยเกี่ยวกับดัชนี L_z และ W_1 ที่มีผู้ศึกษาวิจัยไว้นำเสนอดังนี้

อวยพร วิบูลย์กาญจน์ (2526) ได้ศึกษาเปรียบเทียบ ผลการวิเคราะห์ข้อสอบตามทฤษฎีการทดสอบแบบดั้งเดิม และการวิเคราะห์ด้วยราส์ชโมเดล (W_1 index) โดยใช้แบบทดสอบมาตรฐาน ด้านอุปมาอุปมัย และวิชาวิทยาศาสตร์ เก็บข้อมูลจริงในการศึกษา ปรากฏว่า เมื่อวิเคราะห์ด้วยราส์ชโมเดลมีจำนวนข้อสอบประมาณ 65% ที่เหมาะสมกับโมเดลการตอบข้อสอบ ส่วนการวิเคราะห์ด้วยทฤษฎีแบบดั้งเดิมมีข้อสอบจำนวน 85% ที่เหมาะสมกับโมเดล เมื่อตรวจสอบคุณภาพของแบบสอบ แบบสอบที่คัดเลือกด้วยทฤษฎีการทดสอบแบบดั้งเดิมมีความเที่ยง (KR-20) สูงกว่า ส่วนค่าความตรงที่คัดเลือกจากทั้ง 2 วิธี มีค่าไม่แตกต่างกัน

สุนันท์ ศลโกสุม (2530) ได้ศึกษาเปรียบเทียบผลการวิเคราะห์ข้อสอบด้วยทฤษฎีดั้งเดิม ทฤษฎีการตอบสนองข้อสอบ ดัชนีบ่งชี้ของซาโต้และดัชนีของทาทซูโอกะ โดยใช้ผลการสอบประเมินคุณภาพวิชาภาษาไทย ของนักเรียนชั้นมัธยมศึกษาปีที่ 6 จากกรมวิชาการ กระทรวงศึกษาธิการจำนวนนักเรียน 3 เขตการศึกษา คือ เขต 9, 10 และ 11 เพื่อพิจารณาถึงความบกพร่องของสภาพการเรียนการสอน จากความไม่เหมาะสมของพฤติกรรมการตอบข้อสอบกับโมเดลการตอบข้อสอบปรากฏว่าวิธีการตามทฤษฎีการตอบสนองข้อสอบ และดัชนีบ่งชี้ของซาโต้และดัชนีของทาทซูโอกะ มีประสิทธิผลใกล้เคียงกัน

สมิธ (Smith, 1986) ทำการศึกษาผลการวิเคราะห์บุคคลที่เหมาะสมกับโมเดลการตอบข้อสอบด้วยดัชนี W_1 โดยการจำลองข้อมูลจำนวน 1,800 คน ผู้สอบมีระดับความสามารถแตกต่างกัน 6 ระดับ คือ ระดับความสามารถระหว่าง 0.0 - 0.5, 0.0 - 1.0, 0.0 - 1.5, 0.0 - 2.0, 0.0 - 2.5 และ -0.5 - 2.5 ข้อมูลมีลักษณะการแจกแจงเป็นแบบเดียว (Uniform) ผลการศึกษาพบว่าอัตราการตรวจค้นจะเพิ่มขึ้นเมื่อระดับความสามารถของผู้สอบมีการกระจายมากขึ้น

ดราสโกว์และคณะ (Drasgow et.al., 1987) ได้จำลองข้อมูลให้เป็นคะแนนที่ผู้สอบตอบได้สูงกว่าความเป็นจริง (Spuriously high) และต่ำกว่าความเป็นจริง (Spuriously low) ในระดับที่แตกต่างกัน คือ 15% และ 30% โดยใช้ดัชนีตามทฤษฎีการตอบสนองข้อสอบจำนวน 10 ดัชนี พบว่าดัชนี L_z สามารถตรวจค้นผู้สอบที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบได้ร้อยละ 75 และ 46 ส่วนดัชนี W_1 มีอัตราการตรวจค้น ร้อยละ 73 และ 21 เมื่อจำลองให้ผู้ตอบมีคะแนนสูงและต่ำกว่าความเป็นจริง ตามลำดับ

ดราสโกว์และกัทเลอร์ (Drasgow and Guertler, 1987) ทำการวิจัยเพื่อศึกษาแนวทางในการเลือกใช้ดัชนีบ่งชี้ความเหมาะสมของบุคคล เพื่อตรวจค้นผู้สอนที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบ โดยการจำลองข้อมูลให้เป็นแบบแผนการตอบข้อสอบที่ปกติจำนวน 4,000 คน และข้อมูลอีก 1 ชุด จำนวน 2,000 คน ซึ่งมีผู้ตอบซึ่งตอบได้คะแนนต่ำกว่าความสามารถที่แท้จริง จำนวน 30% วิเคราะห์แบบแผนการตอนโดยใช้ดัชนี L_z เปรียบผลการศึกษาจากข้อมูลที่จำลองขึ้น กับผลการสอบของแบบสอบ ASVAB (The Armied Services Vocational Aptitude Battery) พบว่าดัชนี L_z สามารถตรวจค้นบุคคลที่ไม่เหมาะกับโมเดลการตอบข้อสอบได้ ทั้งในข้อมูลที่จำลองขึ้นและข้อมูลจริง

ไรส์ (Reise, 1990) ได้ศึกษาเปรียบเทียบวิธีการตรวจค้นข้อสอบและผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบ ตามทฤษฎีการตอบสนองข้อสอบ โดยเปรียบเทียบดัชนี L_z และ χ^2 ด้วยการจำลองข้อมูลให้เป็นบุคคลและข้อสอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบ (Missfitting data) การแจกแจงของข้อมูลเป็นแบบเดียว (Uniform) เพื่อจะศึกษาว่า ดัชนีที่ประมาณค่าด้วยวิธีใดมีอัตราการตรวจค้นได้ดีกว่า ผลการศึกษาพบว่าดัชนี χ^2 มีอัตราการตรวจค้นสูงกว่า แต่เมื่อพิจารณาการปฏิเสธสมมติฐานศูนย์ (Type I and Type II error) พบว่า χ^2 มีความลำเอียงในการปฏิเสธสมมติฐาน กล่าวคือ มีความไวมากกว่า ไรส์ ให้ข้อเสนอแนะว่าดัชนี L_z มีความเหมาะสมที่จะใช้วิเคราะห์ความเหมาะสมของบุคคลและข้อสอบมากกว่า χ^2

ไรส์ และ ดิวส์ (Reise & Due, 1991) ทำการศึกษาปัจจัยเกี่ยวกับคุณลักษณะของแบบสอบต่อประสิทธิผลของดัชนีบ่งชี้ความเหมาะสมของบุคคล ซึ่งประมาณค่าด้วยฟังก์ชันไลค์ลิฮูด

(L_z index) เมื่อผู้สอบมีระดับความสามารถแตกต่างกัน 11 ระดับ แต่ละระดับมีช่วงความสามารถห่างกัน 0.5 โดยมีระดับความสามารถระหว่าง -2.5 ถึง +2.5 จำลองข้อมูลเงื่อนไขละ 1,000 คน ความยาวของแบบสอบ 5 ระดับ คือ 7, 14, 21, 35 และ 42 และช่วงความยากของแบบสอบที่แตกต่างกัน ผลการศึกษาพบว่าอัตราการตรวจค้นผู้สอบที่ไม่เหมาะกับโมเดลการตอบข้อสอบ จะเพิ่มขึ้นเมื่อระดับความสามารถของผู้สอบสูงกว่า 1.5 หรือต่ำกว่า -1.5 หรือช่วงความยากของแบบสอบเพิ่มขึ้น โดยความยาวของแบบสอบที่เหมาะสมสำหรับการวิเคราะห์ความเหมาะสมของบุคคลกับโมเดลการตอบข้อสอบจะต้องมีความยาวตั้งแต่ 20 ข้อ ขึ้นไป

นูแนน และคณะ (Noonan et. al., 1993) ศึกษาผลของความยาวของแบบสอบ แบบจำลองโมเดลตามทฤษฎีการตอบสนองข้อสอบ ต่อลักษณะการแจกแจง และความคงที่ของจุดตัดของดัชนี L_z , W_3 และ ECIZ4 โดยศึกษาความยาวของแบบสอบ 40 ข้อ และ 80 ข้อ โมเดลแบบ 2 พารามิเตอร์ และ 3 พารามิเตอร์ พบว่าดัชนี L_z และ ECIZ4 มีลักษณะการแจกแจงเป็นโค้งปกติ ส่วนดัชนี W_3 มีค่าเฉลี่ยเท่ากับ 1.00 แต่ค่าส่วนเบี่ยงเบนมาตรฐานจะขึ้นอยู่กับความยาวของแบบสอบ สำหรับจุดตัดของดัชนีนั้น ดัชนี ECIZ4 จะมีความคงที่ ส่วนดัชนี L_z , W_3 จะขึ้นอยู่กับโมเดลของทฤษฎีการตอบสนองข้อสอบและความยาวของแบบสอบ

ชมิท และคณะ (Schmitt et. al., 1993) ได้ศึกษาความตรงตามเกณฑ์สัมพันธ์ (Criterion Related Validity) ของแบบวัดความสามารถทางเครื่องกล โดยวิเคราะห์แบบแผนการตอบข้อสอบด้วยดัชนี L_{zm} ซึ่งเป็นดัชนี L_z ที่พัฒนาสำหรับใช้วิเคราะห์แบบแผนการตอบข้อสอบที่เป็นแบบสอบชุด (Battries test) คัดเลือกบุคคลที่เหมาะสมกับโมเดลการตอบข้อสอบ (Person-fit) แล้ววิเคราะห์ความตรง พบว่าความตรงของแบบสอบซึ่งวิเคราะห์เฉพาะบุคคลซึ่งเหมาะกับโมเดลการตอบข้อสอบ สูงกว่า ค่าความตรงซึ่งวิเคราะห์ด้วยผู้สอบทั้งหมด

สมิช (Smith, 1994) ทำการศึกษาเปรียบเทียบประสิทธิผลในการตรวจผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบของดัชนี W_1 ในระดับความสามารถของผู้สอบที่แตกต่างกันคือ ความสามารถเฉลี่ยเท่ากับ 0.0, 0.5, 1.0, 1.5 และค่าความยากของข้อสอบในระดับเดียวกัน กับระดับความสามารถของผู้สอบ โดยใช้ความยาวของแบบสอบ 5 ระดับ คือ 10, 20, 30, 40 และ 50 ข้อ ผลการศึกษา พบว่า ระดับความสามารถของผู้สอบและค่าความยากของแบบสอบส่งผลอย่างไม่คงที่ต่อการแจกแจงของค่าสถิติบ่งชี้ความเหมาะสมของข้อสอบ และความยาวของแบบสอบตั้งแต่ 20 ข้อขึ้นไป ส่งผลให้ประสิทธิผลในการตรวจค้นเพิ่มขึ้น

เนอร์ริง (Nering, 1995) ได้ศึกษาลักษณะการแจกแจงของค่าสถิติบ่งชี้ความเหมาะสมของบุคคล ซึ่งประมาณค่าด้วยฟังก์ชันไลด์ ลิฮูด (L_z index) โดยการจำลองข้อมูลให้เป็นแบบ

เดียว (Uniform) และการแจกแจงเป็นโค้งปกติ (Normal) ศึกษาวิธีประมาณค่าพารามิเตอร์ 2 วิธีคือ วิธีแมกซิมัมไลค์ ลิฮูด และวิธีของเบย์ ค่าความสามารถของผู้สอบ (θ) แบบที่ทราบค่าแล้ว และค่าที่ต้องประมาณ และค่าการเดา ผลการศึกษาพบว่า ลักษณะการแจกแจงของค่าสถิติ L_z มีการแจกแจงเป็นโค้งปกติ และการประมาณค่าระดับความสามารถของผู้สอบ ส่งผลต่อดัชนีมากกว่าระดับความสามารถที่ทราบค่าแล้ว วิธีการประมาณค่าพารามิเตอร์ด้วยวิธีการของเบย์ ให้ผลดีกว่า ทั้งนี้เนอร์ริ่ง ไม่ได้เสนอแนะให้ใช้วิธีนี้เนื่องจากในความเป็นจริง ไม่สามารถทราบค่าการแจกแจงเริ่มแรกของกลุ่มตัวอย่างได้ และการลดค่าการเดา จาก 0.20 เป็น 0.00 ไม่ส่งผลต่อลักษณะการแจกแจงของค่าสถิติ

โคราชี (Korashy, 1995) ทำการวิจัยเพื่อหาวิธีการในการคัดเลือกข้อสอบแทนวิธีการตามทฤษฎีการทดสอบแบบดั้งเดิม โดยทำการคัดเลือก ด้วยการวิเคราะห์ข้อสอบที่เหมาะสมกับโมเดลการตอบข้อสอบ ตามวิธีการประมาณค่าจากคะแนนส่วนที่เหลือ (W_1 index) ผลการศึกษาพบว่าค่า Infit จะมีความไวต่อข้อสอบซึ่งมีค่าความยากระดับปานกลาง ส่วนค่า Outfit จะไวต่อข้อสอบซึ่งยากหรือง่าย โดยค่าความยากของข้อสอบซึ่งศึกษาในครั้งนี้มีจำนวน 80 ข้อ ร้อยละ 10 เป็นข้อสอบที่ง่าย ร้อยละ 34 เป็นข้อสอบที่ยาก และร้อยละ 56 เป็นข้อสอบที่มีค่าความยากปานกลาง ได้ผลการศึกษาว่าการใช้สถิติ Infit และค่าสถิติ Outfit คัดเลือกมีข้อสอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบ 8 ข้อ และ 5 ข้อ ตามลำดับ ทั้งนี้เนื่องจากระดับความสามารถของผู้สอบใกล้เคียงกับค่าความยากของข้อสอบ ค่าสถิติ Infit จึงสามารถตรวจค้นข้อสอบที่ไม่เหมาะสมได้มากกว่า

รัดเนอร์ (Ruder, 1996) ได้วิเคราะห์ผลการสอบเพื่อประเมินคุณภาพของนักเรียนระดับมัธยมศึกษาของสหรัฐอเมริกา โดยใช้การประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลจากคะแนนส่วนที่เหลือ ตามโมเอลแบบ 3 พารามิเตอร์ คือดัชนี W_3 สามารถวิเคราะห์ผลการสอบได้ในทุกระดับความสามารถ และไม่มีความแตกต่างของผลการศึกษาระหว่างรัฐ

จากรายงานการวิจัยที่ศึกษา ส่วนใหญ่เป็นการศึกษาโดยการจำลองข้อมูล เป็นข้อมูลซึ่งผู้สอบตอบข้อสอบได้สูงหรือต่ำกว่าความเป็นจริง มีจำนวนบุคคลซึ่งไม่เหมาะกับโมเดลการตอบข้อสอบ ใช้ดัชนีซึ่งประมาณค่าด้วยวิธีการต่างกันตรวจค้น เพื่อเปรียบเทียบอัตราในการตรวจค้นและความถูกต้องในการตรวจค้น (Type I and Type II error rate) มีการประยุกต์ใช้ในข้อมูลจริงเฉพาะดัชนี W_1 เนื่องจากมีการพัฒนาโปรแกรมคอมพิวเตอร์สำหรับใช้ในการวิเคราะห์อย่างแพร่หลาย

ส่วนการศึกษาถึงปัจจัยที่ส่งผลต่อประสิทธิผลของดัชนีนั้น มีการศึกษาค้นคว้าในแต่ละดัชนี ไม่สามารถตอบคำถามได้ว่า เมื่อกำหนดสถานการณ์หรือเงื่อนไขการสอบที่เหมือนหรือแตกต่างกัน ดัชนีใดจะมีประสิทธิผลดีกว่ากัน ผู้วิจัยจึงสนใจที่จะศึกษาว่า เมื่อจำลองแบบแผนการตอบข้อสอบโดยกำหนดเมตริกซ์คำตอบซึ่งไม่สามารถระบุผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบได้ใช้ดัชนี L_z และดัชนี W_1 ซึ่งเป็นดัชนีที่มีประสิทธิภาพในการตรวจค้นผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบ แต่มีวิธีประมาณค่าที่แตกต่างกัน ตรวจค้นผู้สอบที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบ นำผู้สอบเหล่านั้นออกจากการวิเคราะห์ แล้วตรวจสอบคุณภาพของแบบสอบที่ได้จากการวิเคราะห์ในด้านค่าสารสนเทศของแบบสอบความเที่ยงและความตรงตามทฤษฎี เพื่อเป็นแนวทางในการศึกษาค้นคว้าเกี่ยวกับการวิเคราะห์แบบแผนการตอบข้อสอบ ด้วยดัชนีบ่งชี้ความเหมาะสมของบุคคลต่อไป

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย