หอสมุดกลาง สถาบันวิทยบริการ จุฬาลงกรณ์มหาวิทยาลัย

บทที่ 2
แนวคิดและทฤษฎีที่เกียวข้อง

สาระสำคัญในส่วนนี้ จะนำเสนอเป็น 3 ตอน คือ ตอนที่ 1 กล่าวถึงแนวคิดเกี่ยวกับสเกลของกัตแมน (Guttman Perfect Scale) ตอนที่ 2 กล่าวถึงดัชนีบ่งชี้ความเหมาะสมของบุคคล (person-fit indexes) และตอนที่ 3 กล่าวถึงทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

ตอนที่ 1 แนวคิดเกี่ยวกับสเกลของกัตแมน (Guttman Perfect Scale)

กัตแมน (Guttman, 1940) นับว่าเป็นผู้บุกเบิกแนวคิดเกี่ยวกับการวิเคราะห์แบบแผนการตอบข้อสอบรายบุคคล โดยเสนอดัชนีที่บ่งชี้ความสอดคล้องของคะแนนรายข้อของผู้สอบทั้งกลุ่มกับค่าความยากของแบบสอบทั้งฉบับและของข้อสอบรายข้อ โดยการพิจารณาจากคะแนนรายข้อของผู้สอบ (พนิชา สังข์เพชร,2535)

1.1 ข้อตกลงเบื้องต้นของกัตแมน (Allen and Yen, 1979 อ้างถึงในพนิชา สังข์เพชร,2535)

1.1.1 ข้อสอบที่ผู้สอบที่มีความสามารถระดับหนึ่งตอบได้ถูกต้องน่าจะเป็นข้อสอบที่ผู้สอบทุกคนที่มีความสามารถสูงกว่าตอบได้ถูกต้อง ในทางตรงกันข้ามข้อสอบข้อที่ผู้ที่มีความสามารถระดับหนึ่งตอบผิดย่อมทำให้ผู้สอบทุกคนที่มีความสามารถต่ำกว่าผู้สอบคนนั้นตอบผิดด้วย

1.1.2 ผู้สอบที่ตอบข้อสอบที่มีความยากระดับหนึ่งได้ถูกต้อง น่าจะตอบข้อสอบทุกข้อที่ง่ายกว่าข้อดังกล่าวได้ถูกต้องด้วย ในทางตรงกันข้ามเมื่อผู้สอบตอบข้อสอบที่มีความยากระดับหนึ่งผิด เขาย่อมจะตอบข้อสอบทุกข้อที่ยากกว่าข้อดังกล่าวผิดด้วย

กัตแมน เสนอความคิดเกี่ยวกับแบบแผนการตอบข้อสอบ (item response pattern) โดยจัดเรียงลำดับข้อสอบตามแนวนอนจากข้อที่ง่ายที่สุดซึ่งเป็นข้อที่มีจำนวนผู้สอบภายในกลุ่มตอบได้ถูกต้องมากที่สุด เรื่อยไปจนถึงข้อที่ยากที่สุดซึ่งเป็นข้อที่มีจำนวนผู้สอบภายในกลุ่มตอบได้ถูกต้องน้อยที่สุด และนำคะแนนรายข้อของผู้สอบแต่ละคนเรียงลำดับจากผู้สอบที่มีคะแนนรวมสูงสุดถึงผู้สอบที่มีคะแนนรวมต่ำสุดตามแนวตั้งจากบนลงล่าง จะได้เมทริกซ์ของคะแนนรายข้อของผู้สอบ N คน ที่ตอบข้อสอบ K ข้อ ซึ่งลักษณะเมทริกซ์ของกัตแมนนี้จะมีความสอดคล้องสมบูรณ์เป็นไปตามข้อตกลงเบื้องต้น เมื่อคะแนนรายข้อที่อยู่มุมบนด้านซ้ายมือของเมทริกซ์มีค่าเป็น 1

หมดและคะแนนรายข้อที่อยู่มุมล่างด้านขวามือมีค่าเป็น 0 หมด และเรียกเมทริกซ์ที่มีลักษณะเช่นนี้ว่า กัตแมนสมบูรณ์ (Perfect Guttman) ดังตัวอย่างในตารางที่ 1

ตารางที่ 1 แบบแผนการตอบข้อสอบ 5 ข้อ ของผู้สอบ 5 คน ที่มีลักษณะเป็นเมทริกซ์กัตแมนสมบูรณ์

ข้อสอบ / ผู้สอบ	5	3	1	4	2	รวม
a	1	1	1	1	1	5
b	1	1	1	1	0	4
c	1	1	1	0	0	3
d	1	1	0	0	0	2
e	1	0	0	0	0	1

จากตารางที่ 1 แสดงแบบแผนการตอบของผู้สอบ 5 คนที่ตอบข้อสอบ 5 ข้อ ที่มีลักษณะเป็นเมทริกซ์กัตแมนสมบูรณ์ (Perfect Guttman) ทั้งนี้เนื่องจากเมื่อเรียงลำดับคะแนนรวมของผู้สอบจากมากไปน้อยตามแนวตั้งจากบนลงล่าง คือ ผู้สอบ *a* ซึ่งได้คะแนน 5 คะแนน จนถึงคนสุดท้าย คือ e ซึ่งได้คะแนนรวมเพียง 1 คะแนน ในขณะเดียวกันก็เรียงลำดับข้อสอบจากข้อที่ง่ายที่สุดตามแนวนอนจากซ้ายไปขวา คือข้อ 5 ซึ่งมีคนตอบถูก 5 คน จนถึงข้อที่ยากที่สุดคือข้อ 2 ซึ่งมีคนตอบถูกเพียง 1 คน จะพบว่า รูปแบบของตารางจะได้รูปแบบที่มีเลข 1 อยู่ในสามเหลี่ยมด้านบน และมีเมทริกซ์เลข 0 อยู่ในสามเหลี่ยมด้านล่าง โดยในเมทริกซ์ด้านบนไม่มีเลข 0 ปะปนอยู่เลย และในเมทริกซ์รูปสามเหลี่ยมด้านล่างก็ไม่มีเลข 1 ปนอยู่เช่นกัน ซึ่งเป็นไปตามข้อตกลงเบื้องต้นของกัตแมน

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

ตอนที่ 2. ดัชนีบ่งชี้ความเหมาะสมของบุคคล (person-fit indexes)

การวิเคราะห์แบบแผนการตอบข้อสอบด้วยดัชนีบ่งชี้ความเหมาะสมของบุคคล (person-fit indexes) เป็นการใช้ค่าสถิติบ่งบอกว่าบุคคลนั้นมีแบบแผนการตอบที่สอดคล้องกับโมเดลการตอบข้อสอบตามแนวความคิดของกัตแมนที่เรียกว่า กัตแมนสมบูรณ์ (Perfect Guttman) มากน้อยเพียงใด ซึ่งผู้สอบที่มีแบบแผนการตอบที่ปกติ (nonaberrant response pattern) นั้นผู้สอบต้อง

ตอบถูกในข้อที่มีค่าความยากต่ำกว่าหรือเท่ากับความสามารถของตน หรือตอบผิดในข้อที่มีค่าความยากสูงกว่าระดับความสามารถของตน ผู้ที่มีแบบแผนการตอบข้อสอบที่ปกตินี้แสดงว่าเป็นบุคคลที่เหมาะสมกับโมเดล (person-fit) ส่วนแบบแผนการตอบข้อสอบที่ผิดปกติ (aberrant response pattern) ผู้สอบจะตอบผิดในข้อที่มีค่าความยากต่ำกว่าหรือเท่ากับความสามารถของตน หรือตอบถูกในข้อที่มีค่าความยากสูงกว่าระดับความสามารถของตน ซึ่งบุคคลที่มีแบบแผนการตอบที่ผิดปกติแสดงว่าเป็นบุคคลที่ไม่เหมาะสมกับโมเดล (person-misfit) ตามแนวคิดของกัตแมน

2.1 วัตถุประสงค์ในการวิเคราะห์บุคคลที่เหมาะสม

นักวัดผลการศึกษาวิเคราะห์บุคคลที่เหมาะสม (person-fit) เพื่อจุดมุ่งหมายที่แตกต่างกันซึ่งสามารถกล่าวโดยสรุปได้ดังนี้

2.1.1 เพื่อบ่งชี้ความผิดพลาดในการแปลผลของคะแนนจากแบบสอบชุดนั้นๆ เนื่องจากความสามารถ (ability) ของผู้สอบ ที่ตอบข้อสอบแตกต่างไปจากที่คาดหวังไว้ในโมเดล ได้จากการประมาณค่าที่คลาดเคลื่อน (Drasgow&Levine,1987; Drasgow,Levine& Williams,1985; Levine & Rubin,1979 cited in Drasgow,1996)

2.1.2 เพื่อบ่งชี้บุคคลที่มีแบบแผนการตอบที่ผิดปกติเนื่องจากมีมโนทัศน์ที่คลาดเคลื่อน (misconception) ซึ่งทาทซูโอกะได้พัฒนา วิธีการวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนโดยใช้วิธีการ "rule space" วิเคราะห์แบบการคิดที่นักเรียนใช้ในการแก้ปัญหาจากแบบแผนคะแนนการตอบของนักเรียนแต่ละคน (Birnbaum,Kelly & Tatsuoka,1992; Tatsuoka & Tatsuoka,1992; cited in Meijer,1996; ศิริเดช สุชีวะ ,2537)

2.1.3 เพื่อบ่งชี้แบบแผนการตอบที่ผิดปกติของผู้สอบซึ่งอาจเกิดจากพฤติกรรมการตอบข้อสอบของผู้สอบ การบริหารการสอบ หรือความบกพร่องของข้อสอบ (Hanisch & Linn,1981; Rudner,1983; Birenbaum,1986; Meijer,1996)

2.2 ประเภทของดัชนีบ่งชี้ความเหมาะสมสำหรับบุคคล

ดัชนีที่ใช้วิเคราะห์แบบแผนการตอบข้อสอบรายบุคคลนั้น มีผู้ศึกษาและพัฒนาไว้หลายดัชนีด้วยกัน สามารถแบ่งตามเทคนิคของการวิเคราะห์ได้ 2 กลุ่ม คือ (Harnisch and Linn, 1981; สุนันท์ ศลโกสุม, 2530; กุสุมา สุวรรณแก้ว, 2540)

กลุ่มแรก วิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) โดยการหาความสัมพันธ์ของแบบแผนการตอบถูกหรือผิดเป็นรายบุคคลกับคะแนนรวม ซึ่งแบบแผนการตอบข้อสอบที่ถูกต้อง และตรงตามลักษณะของข้อสอบที่มีคุณภาพดี ผู้สอบควรจะตอบข้อสอบที่ง่ายได้ และตอบข้อสอบที่ยากไม่ได้ และคนที่ได้คะแนนรวมเท่ากันควรตอบคำถามได้เหมือนกัน กล่าวคือ ต้องมีแบบแผนการตอบถูกหรือผิดเหมือนกันในกลุ่มที่มีคะแนนเท่ากัน แต่ถ้าในกลุ่มที่ได้คะแนนเท่ากันมีรูปแบบการตอบถูกและผิดไม่เหมือนกัน แสดงว่ามีความผิดปกติเกิดขึ้น อาจเนื่องมาจากข้อสอบหรือตัวผู้สอบ ดัชนีกลุ่มนี้ได้แก่ ดัชนีชี้นำของซาโต้(Sato,1970) ดัชนียู(U-index) ของเวนเดอร์(Vender,1977) ดัชนีความสามารถของบุคคลภายในกลุ่ม(Norm conformity index) ของทาทซูโอกะและทาทซูโอกะ(Tatsuoka and Tatsuoka,1980) ดัชนีที่แสดงความสอดคล้องไม่สอดคล้อง(Agreement and disagreement indices) ของเคนและแบรนแนน(Kane and Brennan,1980) และดัชนีความสอดคล้องของแบบแผนการตอบข้อสอบรายบุคคลของพนิชา สังข์เพ็ชร (2535)

กลุ่มที่สอง ใช้หลักการของทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) ดัชนีกลุ่มนี้คำนวณโดยการประมาณค่าความน่าจะเป็นที่ผู้สอบแต่ละคนจะตอบข้อสอบแต่ละข้อได้ถูกต้อง เปรียบเทียบกับผลการตอบในข้อนั้นๆ เพื่อดูว่าจะผิดไปจากแบบแผนการตอบข้อสอบที่ปกติหรือไม่ ถ้าผิดปกติแสดงว่าผู้สอบคนนั้นไม่เหมาะสมกับโมเดลการตอบข้อสอบ (person misfit) หรือผู้สอบคนนั้นมีแบบแผนการตอบที่ผิดปกติ (aberrant response pattern) ดัชนีกลุ่มนี้ได้แก่ ดัชนีของเลวินและรูบิน (Levine and Rubin, 1979) ดัชนีที่พัฒนาโดยดราสโกว์ (Drasgow,1978) การทดสอบไค-สแควร์ของบุคคล(Chi-Square Test of Person-fit) สถิติ U_3 (U_3 Statistic)ที่พัฒนาโดยโมแคนและเลวิส (Mokken and Lewis,1982) ดัชนี DFC_1 และ DFC_2 ของสำราญ มีแจ้ง(2533) และดัชนีชี้เตือนมาตรฐานหมายเลข4 (Standardized Extended Caution Index 4, SECI4) ของทาทซูโอกะ (Tatsuoka,1984)

2.3 ลักษณะของดัชนีบ่งชี้ความเหมาะสมของบุคคลที่มีประสิทธิภาพ

คุณลักษณะที่พึงประสงค์ของดัชนีบ่งชี้ความเหมาะสมของบุคคลมี 5 ประการ ดังนี้ (พนิชา สังข์เพชร,2535)

ประการที่ 1 ดัชนีต้องมีค่าคงที่เสมอ เมื่อผู้สอบมีแบบแผนการตอบข้อสอบเป็นกัตแมนสมบูรณ์

ประการที่ 2 ดัชนีต้องมีค่าเป็นอิสระจากการเปลี่ยนแปลงของแบบแผนการตอบข้อสอบของผู้สอบคนอื่นๆภายในกลุ่ม โดยเฉพาะอย่างยิ่งในกรณีที่การเปลี่ยนแปลงนั้นไม่มีผล ให้ลำดับที่ของข้อสอบเปลี่ยนไป

ประการที่ 3 ดัชนีต้องมีค่าที่ตรงกับความหมายที่แท้จริงของแบบแผนการตอบข้อสอบของผู้สอบในทุกกรณี

ประการที่ 4 ดัชนีต้องมีช่วงพิสัยที่แน่นอนไม่ขึ้นอยู่กับค่าความยากของข้อสอบหรือจำนวนข้อสอบ

ประการที่ 5 ดัชนีต้องมีความไวต่อความคลาดเคลื่อน ซึ่งความไวต่อความคลาดเคลื่อนหมาย หมายถึง การที่ผู้สอบตอบข้อสอบที่จัดว่าเป็นข้อที่ง่ายมากๆ ผิดจะมีความคลาดเคลื่อนที่รุนแรงกว่าการที่ผู้สอบตอบข้อสอบที่จัดว่ามีความยากใกล้เคียงกับความสามารถของเขาผิด และการที่ผู้สอบตอบข้อสอบที่จัดว่าเป็นข้อที่ยากมากๆ ได้ถูกต้องจะมีความคลาดเคลื่อนที่รุนแรงกว่าการที่ผู้สอบตอบข้อสอบที่จัดว่ามีความยากมากกว่าความสามารถของเขาเล็กน้อยได้ถูกต้อง

2.4 วิธีประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลด้วยฟังก์ชันไลด์ลิฮูด (The Likelihood-based Approach)

การประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลด้วยฟังก์ชันไลด์ลิฮูด (The Likelihood-based Approach) พัฒนาโดย เลวินและรูบิน (Leven & Rubin,1979) และดราสโกว์ (Drasgow,1985) คำนวณค่าดัชนีโดยการประมาณค่าความน่าจะเป็นที่ผู้สอบแต่ละคนจะตอบข้อสอบแต่ละข้อได้ถูกต้อง เปรียบเทียบกับผลการตอบข้อสอบในข้อนั้นๆ แปลงค่าสถิติความเหมาะสมของบุคคลเป็นค่ามาตรฐาน (Z- Value) ทดสอบความแตกต่างแบบสองทาง (two-tailed test) ถ้าดัชนีบ่งชี้ความเหมาะสมของบุคคลใดมีค่าแตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 คือมี ค่าสถิติ $|L_z|$ มากกว่า 1.96 แสดงว่าบุคคลนั้นมีแบบแผนการตอบไม่เหมาะสมกับโมเดลการตอบข้อสอบ (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 คำนวณได้ง่าย ไม่มีความลำเอียง มีการแจกแจงเป็นโค้งปกติ ซึ่งวิธีการประมาณค่าดัชนี L_z มีขั้นตอน ดังนี้ (กุสุมา สุวรรณแก้ว, 2540)

2.4.1 ประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบโดยโมเดลโลจิสติกแบบ 3 พารามิเตอร์

2.4.2 ประมาณค่าความน่าจะเป็นที่ผู้สอบคนที่ i จะตอบข้อสอบข้อที่ j ได้ถูกต้อง โดยใช้ค่าพารามิเตอร์ที่ประมาณได้จากข้อ 1 มีสูตรคำนวณคือ

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

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

2.4.3 ประมาณค่าสถิติบ่งชี้ความเหมาะสมของบุคคล (person-fit statistics) มีสูตรคำนวณคือ

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

2.4.4 แปลงค่าสถิติบ่งชี้ความเหมาะสมของบุคคล (person-fit statistics) เป็นค่าดัชนีมาตรฐาน (standardized person-fit indices) มีสูตรคำนวณคือ

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

เมื่อ $E(L_0)$ คือ ค่าที่คาดหวังของ L_0

$\mathrm{Var}(L_0)$ คือ ความแปรปรวนของ L_0

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

$$\mathrm{Var}(L_0) = \sum_{i=1}^{n}P_i(\theta)\left[1-P_i(\theta)\right]\left\{\ln\left[\frac{P_i(\theta)}{1-P_i(\theta)}\right]\right\}^2 1 \quad 1\text{-}21 \quad 1\text{-}2$$

เมื่อ

$P_i(\theta)$ คือ โอกาสที่ผู้สอบที่มีความสามารถ θจะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$1-P_i$ คือ โอกาสที่ผู้สอบจะตอบข้อสอบข้อที่ i ผิด

U_i คือ พฤติกรรมการตอบที่สังเกตได้ (0,1)

θ คือ ค่าประมาณความสามารถของผู้สอบ

2.4.5 ทดสอบสมมติฐานโดยใช้สถิติทดสอบ ซี (z-test) แบบสองทางซึ่งบุคคลที่ไม่เหมาะกับโมเดลการตอบ(person-misfit) คือ ผู้สอบที่มี ค่าสถิติ $|L_z|$ มากกว่า 1.96 : α = .05 โดยพิจารณาเฉพาะผู้สอบคนที่ j ข้อสอบ i ข้อ

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

3.1 แนวคิดเกี่ยวกับทฤษฎีการตอบสนองข้อสอบ

สำหรับหลักการของทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่อยู่ภายในตัวบุคคล กับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้น โดยมีความเชื่อว่าพฤติกรรมการตอบสนองข้อสอบของผู้สอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรงจะถูกกำหนดโดยคุณลักษณะ (trait) หรือความสามารถ (ability) ที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้และยังเชื่อว่าค่าพารามิเตอร์ต่างๆ ของข้อสอบไม่ว่าจะเป็นค่าความยาก (b) ค่าอำนาจจำแนก (a) หรือค่าการเดา (c) ของข้อสอบแต่ละข้อเป็นคุณลักษณะที่มีอยู่และคงที่พอสมควรในตัวข้อสอบนั้นจริง ฉะนั้นค่าพารามิเตอร์ของข้อสอบจึงไม่แปรเปลี่ยนไปตามกลุ่ม

ตัวอย่าง และค่าความสามารถของผู้สอบซึ่งเชื่อว่าเป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบนั้นจริงจึงไม่ควรแปรเปลี่ยนไปตามค่าความยากของข้อสอบ เนื่องจากความสามารถของผู้สอบเป็นคุณลักษณะแฝงซึ่งไม่สามารถสังเกตได้โดยตรง นักวัดผลทางการศึกษาจึงได้พยายามหาความสัมพันธ์ระหว่างปริมาณการกระทำข้อสอบหรือคะแนน (test performence or score) กับปริมาณความสามารถ (ability) โดยแสดงความสัมพันธ์ดังกล่าวในลักษณะโมเดลทางคณิตศาสตร์ที่กำหนดด้วยฟังก์ชั่นทางคณิตศาสตร์ เรียกว่า ฟังก์ชั่นลักษณะข้อสอบ (Item Characteristic Function) หรือเรียกว่าโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve) หรือฟังก์ชั่นการตอบสนองข้อสอบ (Item Response Function) ซึ่งจะกำหนดได้หลายรูปแบบ ทั้งนี้ขึ้นอยู่กับข้อตกลงเบื้องต้นเกี่ยวกับข้อมูลจากการทดสอบ (ศิริชัย กาญจนวาสี,2538)

3.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบมีข้อตกลงเบื้องต้นที่สำคัญ ดังนี้ (ศิริชัย กาญจนวาสี,2538)

3.2.1 แบบสอบวัดคุณลักษณะเด่นเพียงลักษณะเดียว (unidimensionality : one trait) คุณลักษณะหรือความสามารถที่เป็นตัวกำหนดพฤติกรรมการตอบข้อสอบแต่ละข้อมีลักษณะเด่นและสำคัญเพียงลักษณะเดียว นั่นคือ แบบสอบวัดคุณลักษณะสำคัญเพียงลักษณะเดียว หรือแบบสอบมีความเป็นเอกพันธ์ การตรวจสอบความเป็นไปตามข้อตกลงเบื้องต้นข้อนี้สามารถทำได้โดยการวิเคราะห์องค์ประกอบ (Hambleton and.Sawaminathran, 1985)

3.2.2 การตอบข้อสอบเป็นไปอย่างอิสระ (independence : local independent) การตอบข้อสอบแต่ละข้อของผู้ตอบแต่ละคนมีความเป็นอิสระจากกัน ซึ่งประกอบด้วย

1) ความเป็นอิสระระหว่างข้อสอบ หมายความว่า ข้อสอบแต่ละข้อเป็นอิสระจากกัน กล่าวคือ การตอบสนองต่อข้อสอบข้อหนึ่งไม่มีผลต่อการตอบสนองต่อข้อสอบข้อนั้นๆ ในแบบสอบ เละตำแหน่งข้อสอบแต่ละข้อจะอยู่ที่ใดก็ได้โดยจะไม่มีผลต่อการตอบข้อสอบ ทั้งนี้ถ้าสามารถทดสอบได้ว่าแบบสอบนั้นมีความเป็นเอกมิติ ก็สามารถสรุปได้ว่าแบบสอบนั้นมีความเป็นอิสระระหว่างข้อสอบด้วย

2) ความเป็นอิสระระหว่างผู้สอบ หมายความว่า ผู้ตอบแต่ละคนตอบข้อสอบแต่ละข้อ อย่างเป็นอิสระจากกัน กล่าวคือผลการตอบสนองข้อสอบของผู้สอบคนหนึ่งจะต้องไม่มีผลต่อการตอบสนองข้อสอบของผู้สอบคนอื่น

3.2.3 โค้งคุณลักษณะข้อสอบสามารถใช้อธิบายพฤติกรรมการตอบสนองข้อสอบ (Item Characteristic Curves : Item Response Models) ความสัมพันธ์ระหว่างระดับความสามารถ

ของผู้ตอบ (θ_a) กับโอกาสที่จะตอบข้อสอบถูก [$P_i(\theta)$] สามารถแสดงได้ด้วยโค้งคุณลักษณะข้อสอบ (ICC_s) โดย ICC_s เป็นฟังก์ชันคณิตศาสตร์ที่แสดงด้วยโมเดลการตอบสนองข้อสอบ ที่นิยมใช้กันมี 3 รูปแบบคือ 1,2 หรือ 3 พารามิเตอร์ ในการเลือกใช้ขึ้นอยู่กับลักษณะข้อมูลและความเชื่อในโมเดลที่ใช้อธิบายความสัมพันธ์ดังกล่าว

3.2.4 ข้อสอบที่ใช้ต้องไม่เป็นข้อสอบประเภทความเร็ว ผู้สอบทุกคนควรมีโอกาสทำข้อสอบทุกข้อ เพื่อให้คะแนนรวมจากการสอบประกอบกับลักษณะของข้อสอบเป็นตัวประมาณค่าความสามารถที่แท้จริงของผู้สอบโดยไม่มีข้อจำกัดเกี่ยวกับปัจจัยด้านเวลาที่ใช้ในการสอบ

3.3 แบบจำลองที่ใช้ในทฤษฎีการตอบสนองข้อสอบ (Item Response Model)

ทฤษฎีการตอบสนองข้อสอบได้มีการพัฒนาแบบจำลองขึ้นมาหลายแบบด้วยกัน โดยแบบจำลองแต่ละแบบนั้นจะต่างกันในรูปของฟังก์ชันทางคณิตศาสตร์ และจำนวนพารามิเตอร์ที่กล่าวถึง ซึ่งสามารถจำแนกได้เป็น 3 ประเภทใหญ่ ๆ คือ (Hambleton and Swaminathan, 1985)

3.3.1 แบบจำลองที่ประยุกต์ใช้กับการให้คะแนนแบบ 2 ค่า (binary item) คือ ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน สามารถแบ่งออกตามลำดับขั้นการพัฒนาได้ 2 ระยะดังนี้

1) แบบจำลองในระยะเริ่มแรก (ค.ศ. 1943-1968) เป็นแบบจำลองที่ไม่ค่อยจะมีความสอดคล้องกับข้อมูลจากการสอบในสถานการณ์จริงมากนัก แต่ก็เป็นแบบจำลองที่มีคุณค่าต่อการพัฒนาแบบจำลองที่เหมาะสมในระยะเวลาต่อมา ตัวอย่างแบบจำลองประเภทนี้ เช่น Guttman Perfect Scale, Latent Distance Model และ Linear Model

2) แบบจำลองที่ได้รับการพัฒนาในระยะหลัง (ค.ศ. 1952-1982) เป็นแบบจำลองที่ได้รับการปรับปรุงให้มีความสอดคล้องกับข้อมูลจากการสอบในสถานการณ์จริงมากขึ้น สามารถนำไปปฏิบัติได้ง่าย และมีผู้สนใจนำไปใช้กันอย่างแพร่หลาย ตัวอย่างแบบจำลองประเภทนี้ เช่น One, Two, Three-Parameter Normal Ogive Model และ One, Two,Three ,Four-Parameter Logistic Model

3.3.2 แบบจำลองที่ประยุกต์ใช้กับการให้คะแนนที่มากกว่า 2 ค่า (multichotomously scored) ซึ่งแบบจำลองเหล่านี้มีจุดหมายเพื่อเพิ่มความแม่นยำในการประมาณค่าความสามารถของผู้เข้าสอบ โดยใช้ประโยชน์จากสารสนเทศทั้งจากการตอบข้อสอบถูกหรือผิด ตัวอย่างแบบจำลองประเภทนี้ เช่น Norminal Reponse Model, Grade Response Model และ Partial Credit Model

ของขวัญกลาง สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย

3.3.3 แบบจำลองที่ประยุกต์ใช้กับการให้คะแนนแบบต่อเนื่อง (continuous) ได้แก่ Continuous Response Model ที่พัฒนาโดย Samejima ในปี 1972 อันจะเป็นประโยชน์ต่อการประมาณค่าความสามารถของผู้เข้าสอบทางจิตวิทยา หรือผู้สนใจศึกษาทางด้านที่เกี่ยวกับเจตคติ (attitude) ซึ่งต้องมีการตอบสนองต่อข้อสอบในมาตราที่ต่อเนื่อง (continuous scale)

3.4 ลักษณะเด่นของทฤษฎีการตอบสนองข้อสอบ

เมื่อแบบจำลองการตอบสนองต่อข้อสอบและข้อมูลจากผลการสอบ มีความเหมาะสมกันแล้ว ทฤษฎีการตอบสนองข้อสอบจะมีความเหนือกว่า (superiority) ทฤษฎีการทดสอบแบบดั้งเดิม ดังนี้ (Hambleton and Swaminathan 1985:10-13) คือ

3.4.1 ความเป็นอิสระจากกลุ่มผู้เข้าสอบ (sample-free) นั่นคือไม่ว่าจะนำข้อสอบไปใช้สอบกับบุคคลใด โค้งลักษณะข้อสอบก็จะคงเดิม

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

3.4.3 ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ (invariant of item parameter) กล่าวคือ ไม่ว่าจะประมาณค่ากับกลุ่มตัวอย่างที่มีความสามารถระดับใดก็ตาม ค่าพารามิเตอร์จะไม่ แปรเปลี่ยนไปตามกลุ่มผู้สอบ ทั้งนี้เพราะ ฟังก์ชันการตอบสนองต่อข้อสอบ ซึ่งเป็นฟังก์ชันที่ใช้อธิบายความสัมพันธ์ระหว่างพฤติกรรมการตอบสนองต่อข้อสอบกับ ระดับความสามารถนั้นเป็นฟังก์ชันการถดถอย (Regression function) ของคะแนนจากการสอบบนความสามารถ ซึ่งตามทฤษฎีสถิติ ฟังก์ชันการถดถอยจะไม่แปรเปลี่ยนไปตามการแจกแจงความถี่ของตัวแปรทำนาย ดังนั้น ค่าพารามิเตอร์ต่างๆ ในสมการถดถอย ซึ่งจะเป็นตัวกำหนดรูปร่างของเส้นถดถอยจึงไม่เปลี่ยนแปลงด้วย (Lord,1980)

เบเกอร์ (Baker,1977 อ้างถึงใน รัตนา ศรีเหรัญ, 2539) กล่าวว่า ความไม่แปรเปลี่ยนของค่าพารามิเตอร์ขึ้นอยู่กับเงื่อนไข 2 ประการ คือ ประการแรก ความสามารถ (ability) ที่กล่าวถึงต้องสามารถนิยามได้ชัดเจนและวัดได้ด้วยข้อสอบ อีกประการหนึ่งความสามารถที่วัดนั้นจะต้องมีความคงที่ (statics) ภายในช่วงระยะเวลาหนึ่ง นอกจากนี้ ลอร์ด (Lord,1980) ยังได้อธิบายเพิ่มเติมว่าการไม่แปรเปลี่ยนของค่าพารามิเตอร์นั้นมิได้หมายความว่า ค่าพารามิเตอร์ที่ได้จากการ

ประมาณค่าโดยใช้กลุ่มตัวอย่างที่แตกต่างกันจะมีค่าเท่ากันเสมอ ค่าพารามิเตอร์ของข้อสอบจะเท่ากันหรือไม่ ขึ้นอยู่กับลักษณะของแบบสอบบางประการ เช่น ถ้าเลือกมาตรวัดที่มีจุดเริ่มต้นเดียวกันและหน่วยในการวัดหน่วยเดียวกันแล้ว ค่าพารามิเตอร์ที่ประมาณได้จากกลุ่มตัวอย่างที่แตกต่างกันจะมีค่าเท่ากัน ในทางตรงกันข้าม ถ้าหากเลือกมาตรวัดที่มีจุดเริ่มต้นและมีหน่วยในการวัดแตกต่างกันแล้ว การไม่แปรเปลี่ยนของค่าพารามิเตอร์จะหมายความว่า ค่าพารามิเตอร์ที่ประมาณได้จากกลุ่มตัวอย่างที่ต่างกันของข้อสอบชุดหนึ่งจะมีความสัมพันธ์กันในเชิงเส้นตรง

3.5 วิธีประมาณค่าพารามิเตอร์

การประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบมีหลายวิธี เช่น วิธีการของเบย์ วิธีฮิวริสติก วิธีแมกซิมัมไลค์ลิฮูด วิธีมาร์จินอลแมกซิมัมไลค์ลิฮูด เป็นต้น เนื่องจากการวิจัยครั้งนี้ศึกษาวิธีประมาณค่าดัชนีบ่งชี้ความเหมาะสมของบุคคลโดยการประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด จึงขอเสนอรายละเอียดเฉพาะวิธีการประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด(Maximum Likelihood Estimation) ดังนี้

3.5.1 แนวคิดของวิธีประมาณค่าพารามิเตอร์วิธีแมกซิมัมไลค์ลิฮูด

แนวคิดของวิธีแมกซิมัมไลค์ลิฮูด มีว่า การประมาณค่าพารามิเตอร์ ทำได้โดยอาศัยผลที่ได้จากตัวอย่างสุ่ม ที่เลือกมาจากการแจกแจงที่ทราบรูปแบบของฟังก์ชันความหนาแน่น แต่ไม่ทราบค่าพารามิเตอร์ จึงใช้ความน่าจะเป็นในการเลือกตัวอย่างและวัดค่า ($U_1 = u_1, U_2 = u_2, \ldots, U_n = u_n$) มาพิจารณาค่าประมาณของค่าพารามิเตอร์ ทั้งนี้การประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูดเป็นการหาค่าประมาณของค่าพารามิเตอร์ที่จะทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด ซึ่งโดยปกติจะทำการหาค่าพารามิเตอร์ที่ทำให้ลอการิทึม (Logorithm) ของฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด เนื่องจากค่าประมาณของพารามิเตอร์ที่ทำให้ฟังก์ชันทั้ง 2 ฟังก์ชันมีค่าสูงสุดเป็นค่าเดียวกัน แต่การหาค่าประมาณที่ทำให้ลอการิทึมของฟังก์ชันไลค์ลิฮูดมีค่าสูงสุดนั้นทำได้ง่ายกว่า ซึ่งโดยทั่วไปสมการไลค์ลิฮูดนี้ไม่เป็นสมการเส้นตรง การหาค่าที่ทำให้มีค่าสูงสุดจึงไม่สามารถหาได้ด้วยการใช้วิธีการอย่างง่าย แต่หาได้โดยใช้วิธีของนิวตัน-ราพสัน (Newton - Raphson) (Lord, 1980) ซึ่งเป็นการหาค่าประมาณโดยการทำซ้ำ ๆ (iterative) จนมีค่าคงที่ โดยมีขั้นตอนดังนี้

ขั้นที่ 1 กำหนดค่าเริ่มต้นสำหรับใช้ในการประมาณค่าความสามารถ (θ) และค่าพารามิเตอร์ของข้อสอบ (a, b, c)

ขั้นที่ 2 ประมาณค่าความสามารถของผู้เข้าสอบแต่ละคนโดยกำหนดให้ค่าพารามิเตอร์ของข้อสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่ และประมาณค่าซ้ำ (iterative) จนกว่าค่าประมาณความสามารถ จะมีค่าเข้าใกล้ค่าใดค่าหนึ่ง (convergence) หรือการประมาณค่าในครั้งที่ติดต่อกันมีค่าแตกต่างกันน้อยกว่าค่าคงที่ที่กำหนดไว้

ขั้นที่ 3 ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อ โดยกำหนดให้ค่าความสามารถของผู้เข้าสอบที่ประมาณได้ในครั้งก่อนเป็นค่าคงที่ และประมาณค่าซ้ำ (iterative) จนกว่าค่าประมาณจะลู่เข้าใกล้ค่าใดค่าหนึ่ง (convergence)

ขั้นที่ 4 ประมาณค่าซ้ำในขั้นที่ 2 และขั้นที่ 3 จนกว่าค่าประมาณจะมีค่าคงที่และมีความถูกต้องเพียงพอหรือทำให้ค่าลอการิทึมของฟังก์ชันไลค์ลิฮูดมีค่าสูงที่สุด

ทั้งนี้ค่าอนุพันธ์อันดับที่ 1 และค่าอนุพันธ์อันดับที่ 2 ของค่าพารามิเตอร์ ที่ได้จากการประมาณโดยวิธีแมกซิมัมไลค์ลิฮูดเป็นการประมาณค่าซ้ำ ๆ หลายครั้ง และเพื่อให้ได้ค่าคงที่เร็วขึ้น ดังนั้นค่าความสามารถจะถูกปรับให้มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ซึ่งจะเป็นผลให้ค่า a, b และ c ต้องถูกปรับตามไปด้วย (Hambleton and Swaminathan, 1985) ซึ่งปัจจุบันในการคำนวณเพื่อประมาณค่าพารามิเตอร์ด้วยวิธีแมกซิมัมไลค์ลิฮูด สามารถทำได้โดยสะดวก เนื่องจาก ได้มีการพัฒนาโปรแกรมสำหรับวิเคราะห์ข้อสอบไว้หลายโปรแกรม เช่น โปรแกรมโลจีสต์ 5.0 โปรแกรม BILOG โปรแกรม MULTILOG โปรแกรม PARSCALE เป็นต้น ทั้งนี้หากต้องการใช้แบบจำลองโลจิสติก 2 พารามิเตอร์ ให้กำหนดค่าการเดา (c) เป็น 0 และเมื่อต้องการใช้แบบจำลองโลจิสติก 1 พารามิเตอร์ ให้กำหนดค่าการเดา (c) เป็น 0 และค่าอำนาจจำแนก (a) เป็น 1

3.5.2 คุณสมบัติที่สำคัญของวิธีแมกซิมัมไลค์ลิฮูด (Hambleton and Swaminathan, 1985)

1) มีความคงที่ เมื่อจำนวนของกลุ่มผู้เข้าสอบและข้อสอบเพิ่มขึ้น การประมาณก็จะมีความคงที่ไปสู่ค่าที่แท้จริง

2) ฟังก์ชันของค่าสถิติให้สารสนเทศทั้งหมดเกี่ยวกับค่าพารามิเตอร์

3) มีประสิทธิภาพซึ่งวิธีการประมาณค่าพารามิเตอร์จะมีความแปรปรวนน้อยที่สุด

4) มีลักษณะของการกระจายเข้าใกล้แบบปกติ

เมื่อกลุ่มตัวอย่างผู้เข้าสอบและข้อสอบมีขนาดเล็ก คุณสมบัติข้างต้นจะเป็นจริงหากทำการประมาณค่าความสามารถของบุคคล หรือ ประมาณค่าพารามิเตอร์ของข้อสอบเพียงอย่างใดอย่างหนึ่งเท่านั้น แต่ถ้าผู้เข้าสอบและจำนวนข้อสอบเพิ่มขึ้น ในการประมาณค่าพารามิเตอร์ข้อสอบและความสามารถของผู้เข้าสอบพร้อมๆ กัน ค่าประมาณที่ได้จะมีค่าเข้าใกล้ค่าพารามิเตอร์จริง

3.5.3 ข้อจำกัดและปัญหาของกระบวนการประมาณค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบไปพร้อมๆกันของวิธีแมกซิมัมไลค์ลิฮูด (Hambleton and Swaminathan, 1985)

1) ในการประมาณค่าพารามิเตอร์ในขั้นที่ 2 และ 3 โดยใช้ค่าอนุพันธ์อันดับที่ 2 เป็นตัวหารในกระบวนการนิวตัน ราฟสัน นั้น มีโอกาสที่ค่าประมาณที่ได้จะไม่ลู่เข้าสู่ค่าคงที่

2) เมื่อสมการไลค์ลิฮูดเป็นระบบสมการที่ไม่ใช่เชิงเส้นตรง ทำให้สามารถหารากของสมการที่ทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุดได้หลายค่า แต่ค่าเหล่านี้ไม่สามารถนำไปใช้เป็นค่าพารามิเตอร์ที่แท้จริงได้

3) บางครั้งค่าพารามิเตอร์หรือค่าที่ได้จากการประมาณ ไม่ตกอยู่ในขอบเขตของค่าพารามิเตอร์ กล่าวคือ อาจมีค่าใดค่าหนึ่งอยู่ภายนอกขอบเขตที่ยอมรับได้ ในกรณีเช่นนี้ต้องมีการกำหนดขอบเขตจำกัดของค่าประมาณไว้ เพื่อให้ค่าประมาณที่ได้ไม่สูงหรือต่ำเกินไปนัก แต่การกระทำเช่นนี้ เป็นจุดอ่อนของกระบวนการประมาณค่าด้วยวิธีแมกซิมัมไลค์ลิฮูด โดยเฉพาะในแบบจำลอง 2 และ 3 พารามิเตอร์ ซึ่งทำให้เกิดปัญหาตามมาเกี่ยวกับความตรง (validity) ของค่าที่ประมาณได้

4) เนื่องจากการประมาณค่าพารามิเตอร์ในแบบจำลอง 3 พารามิเตอร์ มีสมการหลายสมการที่ต้องหารากที่ทำให้ฟังก์ชันไลค์ลิฮูดมีค่าสูงสุด วิธีของนิวตัน ราฟสัน จึงจำเป็นต้องใช้เครื่องคอมพิวเตอร์ขนาดใหญ่และโปรแกรมที่มีประสิทธิภาพในการคำนวณ

<u>รายงานวิจัยที่เกี่ยวข้อง</u>

การศึกษางานวิจัยและวรรณคดีที่เกี่ยวข้อง พบว่า ในการวิเคราะห์บุคคลที่เหมาะสม (person-fit) ด้วยดัชนีบ่งชี้ความเหมาะสมของบุคคล (person-fit indices) นั้น มีข้อค้นพบที่น่าสนใจหลายประเด็นด้วยกัน อาทิเช่น รัดเนอร์ (Rudner, 1983) ได้ศึกษาเปรียบเทียบประสิทธิผลของการวิเคราะห์แบบแผนการตอบข้อสอบด้วยดัชนี L_z ซึ่งประมาณค่าด้วยฟังก์ชันไลด์ลิฮูด และดัชนี W_1 ซึ่งประมาณค่าด้วยคะแนนส่วนที่เหลือ โดยจำลองคะแนนของผู้สอบให้สูงกว่าความเป็นจริง (spuriosly high) และต่ำกว่าความเป็นจริง (spuriosly low) พบว่า ดัชนี L_z สามารถตรวจค้นผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบได้ดีกว่าดัชนี W_1 นอกจากนี้ดราสโกว์และคณะ (Drasgow et. al., 1987) ได้จำลองข้อมูลคะแนนผู้สอบให้สูงกว่าความเป็นจริง (spuriosly high) และต่ำกว่าความเป็นจริง (spuriosly low) ในระดับที่แตกต่างกัน คือ 15 % และ 30 % แล้วใช้ดัชนีตามทฤษฎีการตอบสนองข้อสอบตรวจค้นแบบแผนการตอบที่ผิดปกติ พบว่า เมื่อจำลองให้ผู้สอบมีคะแนนสูงกว่าความเป็นจริง ดัชนี L_z และดัชนี W_1 สามารถตรวจค้นผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบได้ร้อยละ 75 และ 46 ตามลำดับ และเมื่อจำลองให้ผู้สอบมีคะแนนต่ำกว่าความเป็นจริง ดัชนี L_z และดัชนี W_1 สามารถตรวจค้นผู้สอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบได้ร้อยละ 73 และ 21 ตามลำดับ

เบิอร์นบอม (Birnbaum, 1986) ได้ศึกษาเปรียบเทียบดัชนีบ่งชี้ความเหมาะสมของบุคคลซึ่งใช้หลักการของทฤษฎีการตอบสนองข้อสอบ ทั้งหมด 9 ดัชนี ได้แก่ ดัชนีชี้เตือนของทาทสุโอกะที่ยังไม่ได้ปรับความเป็นมาตรฐาน (The Extended Caution Index : ECI) คือ ดัชนี ECI_1 , ECI_2 ECI_4 ดัชนีชี้เตือนมาตรฐานของทาทสุโอกะ (Standardized Extended Caution Index , SECI) คือ ดัชนี ECI_z , $ECI2_z$ $ECI4_z$ ดัชนีของเลวิฟและรูบิน (L_0) ดัชนีของเลวินและดราสโกว์ (L_z) และดัชนีที่เสนอโดยไรท์ (U_z) โดยได้แบ่งกลุ่มตัวอย่างเป็น 3 กลุ่ม คือ กลุ่มนักเรียนที่ให้ความร่วมมือในการทำข้อสอบ (coperative examines) กลุ่มที่ไม่ให้ความร่วมมือ (uncoperative examines) และกลุ่มที่สุ่มทั่วๆไป ใช้การพิจารณาการเขียนชื่อบนกระดาษคำตอบของนักเรียนเป็นเกณฑ์ตัดสินว่านักเรียนให้ความร่วมมือหรือไม่ ถ้าเขียนชื่อแสดงว่าให้ความร่วมมือ ถ้าไม่เขียนแสดงว่าไม่ให้ความร่วมมือ ผลการเปรียบเทียบความสัมพันธ์พบว่าดัชนี L_z มีความสัมพันธ์กับดัชนีอื่นๆ สูงมากที่สุดและเป็นดัชนีที่มีความสัมพันธ์กับคะแนนรวมน้อยที่สุด

ไรส์ (Reise, 1990) ได้ศึกษาเพื่อเปรียบเทียบประสิทธิภาพของดัชนี L_z และการทดสอบ χ^2 ว่าดัชนีที่ประมาณค่าด้วยวิธีใดมีอัตราการตรวจค้นผู้สอบที่ไม่เหมาะสมกับโมเดลการตอบข้อ

สอบได้ดีกว่ากัน ด้วยการจำลองข้อมูลให้เป็นการตอบซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบ (misfitting data) การแจกแจงของข้อมูลเป็นแบบเดียว (uniform) พบว่า ดัชนี L_z มีความเหมาะสมในการวิเคราะห์ความเหมาะสมของบุคคล (person fit) มากกว่าการทดสอบ χ^2

ไรส์และดิวส์ (Reise & Due, 1991) ได้ศึกษาคุณลักษณะของแบบสอบที่มีผลต่อประสิทธิผลของดัชนีบ่งชี้ความเหมาะสมซึ่งประมาณค่าด้วยฟังก์ชั่นไลค์ลิฮูด (L_z index) โดยศึกษาผลของค่าการเดา ความยาวของแบบสอบและช่วงความยากของแบบสอบ ต่อการตรวจค้นบุคคลซึ่งไม่เหมาะสมกับโมเดลการตอบข้อสอบ เมื่อระดับความสามารถของผู้สอบแตกต่างกัน โดยได้ศึกษาค่าการเดาที่ระดับ 0.00, 0.05, 0.10, 0.15และ0.20 ความยาวของแบบสอบ 7, 21, 35และ 49 ข้อ และช่วงความยากของแบบสอบมี 5 ช่วง คือ -3.0ถึง0.3, -2.5ถึง2.5, -2.0ถึง2.0, -1.5ถึง1.5 และ -1.0ถึง1.0 ผลการศึกษาพบว่า ดัชนี L_z จะมีปัญหาในการตรวจค้นบุคคลที่ไม่เหมาะสมกับโมเดลในแบบสอบที่มีความยาวน้อยกว่า 20 ข้อ ส่วนอัตราการตรวจค้นจะเพิ่มขึ้นเมื่อช่วงความยากของแบบสอบเพิ่มขึ้น หรือเมื่อความสามารถของผู้สอบสูงกว่า 1.5 หรือ ต่ำกว่า -1.5 และอัตราการตรวจค้นจะลดลงเมื่อค่าการเดาสูงขึ้น

ชมิทและคณะ (Schmitt et al. 1993) ได้ศึกษาความตรงตามเกณฑ์สัมพันธ์ (criterion-ralate validity) ของแบบสอบวัดความสามารถทางเครื่องกล ซึ่งเป็นแบบสอบชุด (Battries Test) โดยวิเคราะห์แบบแผนการตอบข้อสอบรายบุคคลด้วยดัชนี L_{zm} ซึ่งเป็นดัชนี L_z ที่พัฒนาขึ้นให้สามารถใช้วิเคราะห์แบบสอบชุด เมื่อวิเคราะห์หาคุณภาพของแบบสอบ พบว่าแบบสอบซึ่งวิเคราะห์โดยใช้แบบแผนการตอบข้อสอบรายบุคคล มีความตรงตามเกณฑ์สัมพันธ์สูงกว่าแบบสอบซึ่งวิเคราะห์โดยไม่พิจารณาแบบแผนการตอบ การศึกษาครั้งนี้ของชมิทและคณะได้สรุปสอดคล้องกับดราสโกว์ (Drasgow, 1987) กล่าวคือเมื่อวิเคราะห์แบบแผนการตอบข้อสอบรายบุคคลเพื่อวิเคราะห์บุคคลที่เหมาะสมกับโมเดลการตอบข้อสอบแล้วแยกบุคคลที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบออกจากการวิเคราะห์ จะทำให้ค่าความตรงของแบบสอบที่วิเคราะห์เฉพาะบุคคลที่มีแบบแผนการตอบปกติเพิ่มขึ้น ดังนั้น การวิเคราะห์ความเหมาะสมของบุคคลกับโมเดล (analysis of person-fit) จึงมีบทบาทสำคัญในการสร้างและศึกษาคุณภาพแบบสอบ ซึ่งดัชนีบ่งชี้ความเหมาะสมของบุคคล(person-fit index) ที่มีคุณภาพจะต้องสามารถบ่งชี้บุคคลซึ่งมีแบบแผนการตอบผิดปกติได้อย่างถูกต้องแม่นยำ เป็นผลให้แบบสอบที่วิเคราะห์หลังจากนำบุคคลที่ไม่เหมาะสมกับโมเดลการตอบข้อสอบออกจากการวิเคราะห์มีคุณภาพเพิ่มขึ้น (กุสุมา สุวรรณแก้ว, 2540)

จากการศึกษาของเนอริง (Nering, 1995) ซึ่งได้จำลองข้อมูลการตอบข้อสอบของผู้สอบให้เป็นแบบทางเดียว (uniform) และมีการแจงแจกเป็นโค้งปกติ เพื่อศึกษาลักษณะการแจกแจงของค่าสถิติบ่งชี้ความเหมาะสมของบุคคลซึ่งประมาณค่าด้วยฟังก์ชั่นไลด์ลิฮูด (L_z index) พบว่า ลักษณะการแจกแจงของสถิติ L_z มีการแจกแจงเป็นโค้งปกติ และการประมาณค่าความสามารถของผู้สอบส่งผลต่อดัชนีมากกว่าระดับความสามารถที่ทราบค่าแล้ว วิธีการประมาณค่าพารามิเตอร์โดยวิธีของเบย์ให้ผลดีกว่าวิธีแมกซิมัมไลด์ลิฮูด แต่เนื่องจากในความเป็นจริงไม่สามารถทราบค่าการแจกแจงเริ่มแรกของกลุ่มตัวอย่างได้ ดังนั้นเนอริงจึงไม่ได้เสนอแนะให้ใช้วิธีการประมาณค่าด้วยวิธีนี้ และการเปลี่ยนค่าการเดาจาก 0.20 เป็น 0.00 ไม่มีผลต่อลักษณะการแจกแจงของค่าสถิติ L_z นอกจากนี้ เนอริง (Nering, 1996) ยังได้ศึกษาเปรียบเทียบการกระจายของดัชนีบ่งชี้ความเหมาะสมของบุคคล คือดัชนี L_z กับ ดัชนี $ECI4_z$ ในบริบทการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (Computerized Adaptive Testing) โดยจำลองข้อมูลการตอบข้อสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAT) เมื่อจำนวนข้อสอบในกลุ่มข้อสอบ (item pool) การกระจายของค่าอำนาจจำแนกของข้อสอบ (a) และเกณฑ์การยุติการสอบซึ่งใช้ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์มีค่าต่างกัน พบว่า ในบริบทการทดสอบแบบ CAT ทั้งดัชนี L_z และ ดัชนี $ECI4_z$ มีการกระจายเข้าใกล้โค้งปกติเมื่อจำนวนข้อสอบในกลุ่มข้อสอบ (item pool) มีจำนวนน้อย ค่าอำนาจจำแนกของข้อสอบ (a) มีค่าน้อย และเกณฑ์การยุติการสอบซึ่งใช้ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์มีค่ามาก

กุสุมา สุวรรณแก้ว (2540) ได้ศึกษาเปรียบเทียบผลการตรวจสอบความเหมาะสมของบุคคลระหว่างดัชนีซึ่งประมาณค่าด้วยวิธีที่แตกต่างกัน ได้แก่ ดัชนี L_z ประมาณค่าด้วยฟังก์ชั่นไลด์ลิฮูด และ ดัชนี W_1 ประมาณค่าดัชนีจากคะแนนส่วนที่เหลือ เมื่อระดับความสามารถของผู้สอบ และความยาวของแบบสอบแตกต่างกัน ดำเนินการวิจัยโดย จำลองข้อมูลด้วยโปรแกรม IRTDATA ตรวจสอบความเป็นเอกมิติของแบบสอบทุกเงื่อนไขด้วยโปรแกรม SPSSX และ SPSSPC$^+$ ผลการตรวจสอบความเหมาะสมของบุคคลวัดจากคุณภาพของแบบสอบหลังการตรวจสอบความเหมาะสมของบุคคลกับโมเดลการตอบข้อสอบในด้าน ค่าสารสนเทศ ความเที่ยง และความตรงตามทฤษฎีของแบบสอบ ผลการวิจัยพบว่า ผลการตรวจสอบความเหมาะสมของบุคคลด้วยดัชนี L_z ด้านค่าสารสนเทศของแบบสอบสูงกว่าดัชนี W_1 ในทุกระดับความสามารถของผู้สอบและทุกระดับความยาวของแบบสอบ และผลการตรวจสอบความเหมาะสมของบุคคลด้วยดัชนี L_z และ ดัชนี W_1 ด้านความเที่ยงและความตรงตามทฤษฎีไม่แตกต่างกัน ในทุกระดับความสามารถของผู้สอบ และทุกระดับความยาวของแบบสอบ

จากรายงานการวิจัยที่ได้ศึกษา ส่วนใหญ่เป็นการศึกษาเปรียบเทียบประสิทธิภาพของดัชนีปรงชี้ความเหมาะสมของบุคคลกับโมเดลการตอบข้อสอบซึ่งประมาณค่าแตกต่างกัน ว่าแต่ละดัชนีมีอัตราการตรวจค้นและมีความถูกต้องในการตรวจค้นมากน้อยเพียงใด โดยมักจะกำหนดสถานการณ์ในการสอบให้ผู้สอบมีความสามารถเหมือนกัน หรือข้อสอบมีค่าพารามิเตอร์ที่เหมือนกัน การศึกษาโดยใช้ดัชนีปรงชี้ความเหมาะสมดัชนีใดดัชนีหนึ่งโดยเฉพาะยังมีน้อย

ผู้วิจัยจึงสนใจศึกษาเปรียบเทียบสัดส่วนของบุคคลที่มีแบบแผนการตอบที่ผิดปกติ เมื่อให้ผู้สอบที่มีระดับความสามารถแตกต่างกันทำข้อสอบที่มีค่าอำนาจจำแนกของข้อสอบ (a), ค่าความยากของข้อสอบ (b) และค่าการเดา (c) ที่แตกต่างกัน โดยเลือกใช้ดัชนีปรงชี้ความเหมาะสมที่มีประสิทธิภาพ ซึ่งดัชนี L_z ก็เป็นอีกดัชนีหนึ่งที่มีรายงานการวิจัยที่ผ่านมาสนับสนุนว่ามีประสิทธิภาพในการตรวจค้นบุคคลที่มีแบบแผนการตอบที่ผิดปกติ และปัจจุบันได้มีการพัฒนาโปรแกรมสำเร็จรูปที่สามารถวิเคราะห์ได้โดยสะดวกและมีประสิทธิภาพ ดังนั้น ในการวิจัยครั้งนี้ผู้วิจัยจึงใช้ดัชนี L_z ในการตรวจค้นบุคคลที่มีแบบแผนการตอบผิดปกติ เพื่อให้ทราบว่าสัดส่วนของผู้ที่มีแบบแผนการตอบที่ผิดปกติในแต่ละเงื่อนไขมากน้อยเพียงใด อันจะเป็นประโยชน์ในการเลือกข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบเพื่อให้การวัดผลการศึกษามีความคลาดเคลื่อนต่ำ โดยพิจารณาจากสัดส่วนของแบบแผนการตอบผิดปกติในแต่ละเงื่อนไข

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย