บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ทฤษฎีการวัดเป็นการประยุกต์วิธีการทางสถิติมาใช้อธิบายลักษณะของผู้สอบและลักษณะของข้อสอบ เพื่อประเมินคุณภาพการวัด เพื่อประโยชน์ในการปรับปรุงและเพื่อเสนอแนะวิธีการพัฒนาการสอบวัด โดยหลักการสอบวัดมุ่งเน้นไปที่ความถูกต้อง ความแน่นอนและความแท้จริงของการสอบวัด ซึ่งกล่าวได้ว่าทฤษฎีการสอบวัดนั้นมุ่งเน้นให้ผลที่ได้บอกค่าที่แท้จริง โดยเฉพาะอย่างยิ่งค่าความสามารถจริงหรือคะแนนจริงของการสอบวัดนั้น

ทฤษฎีการสอบวัดถ้าพิจารณาจากหลักการของทฤษฎี สามารถจำแนกได้เป็นสองลักษณะ คือ แบบที่มีข้อตกลงคะแนนจริงอ่อนหรือไม่เข้มงวด กับแบบที่มีข้อตกลงคะแนนจริงแกร่งหรือเข้มงวด โดยแบบที่มีข้อตกลงคะแนนจริงอ่อน ได้แก่ทฤษฎีแบบดั้งเดิม ส่วนแบบที่มีข้อตกลงคะแนนจริงแกร่ง ได้แก่ทฤษฎีแบบคุณลักษณะแฝงและอื่น ๆ ดังนั้นในการนำเสนอทฤษฎีและงานวิจัยที่เกี่ยวข้องในการศึกษาครั้งนี้จึงแยกเสนอตามลำดับดังนี้ 1. ทฤษฎีการสอบวัดแบบดั้งเดิม 2. ทฤษฎีการสอบวัดแบบคุณลักษณะแฝง และ 3. ผลงานวิจัยที่เกี่ยวข้อง

1. <u>ทฤษฎีการสอบวัดแบบดั้งเดิม</u>

แอเลนและเยน (Allen and Yen , 1979 : 3) ได้สรุปประวัติของการสอบวัดไว้ว่า เริ่มมีบทบาทในหลายประเทศตั้งแต่ศตวรรษที่ 12 - 16 แต่ได้รับการพัฒนาอย่างจริงจังในศตวรรษที่ 18 บุคคลที่นับว่ามีบทบาทที่ก่อให้เกิดการเคลื่อนไหวในการสอบวัดมีอยู่หลายคน เช่น แกลตัน (Sir Francis Galton , 1822 - 1911) ชาวอังกฤษ นับเป็นบุคคลแรกที่ให้ข้อคิดด้านการวัดความแตกต่างระหว่างบุคคล และเป็นบุคคลที่วางรากฐานในวิชาสถิติและการทำแบบสอบมาตรฐาน แคทเทล (Cattell , 1879) ชาวอเมริกัน ได้สร้างข้อสอบวัดความสามารถในด้านต่าง ๆ ของมนุษย์ขึ้นหลายฉบับ และได้รับการยกย่องว่าเป็นบุคคลที่บุกเบิกทางให้แก่การวัดความสามารถด้านเชาวน์ปัญญา (Intelligence test) และเป็นบุคคลแรกที่ใช้คำว่า "การวัดทางสมอง" (Mental test) เพียร์สัน (Karl Pearson, 1857 - 1936) ชาวอังกฤษ เป็นผู้วางรากฐานทางสถิติผู้หนึ่งและได้นำเทคนิควิธีการทางสถิติมาใช้เป็นพื้นฐานของการสอบวัด บิเนท์ (Alfred Binet , 1857 - 1911) ชาวฝรั่งเศส เป็นผู้พัฒนาแบบสอบวัดเชาวน์ปัญญา และนับเป็นจุดเริ่มต้นที่ทำให้เกิดความสนใจในการสร้างแบบสอบวัดความสามารถทางสมองเป็นอย่างมาก สเตม (William Stem, 1871 - 1938) ชาวเยอรมัน ได้พัฒนาแบบวัด IQ ขึ้น สเปียร์แมน (Charles Spearman, 1863 - 1945) ชาวอังกฤษ เป็นผู้พัฒนาแนวคิดของแกลตันและเพียร์สันในด้านสถิติและความสามารถทางสมอง และนับว่าเป็นบุคคลที่สร้างมโนทัศน์เกี่ยวกับความเที่ยงของแบบสอบ และการวิเคราะห์องค์ประกอบ ธอร์นไดค์ (Edward L.Thorndike, 1904) ชาวอเมริกัน เป็นบุคคลที่สำคัญยิ่งที่ได้เผยแพร่วิทยาการด้านการวัดด้วยวิธีการทางวิทยาศาสตร์ โดยประยุกต์ทฤษฎีการวัดและวิชาทางสถิติมาใช้ในการสอบวัด และนับเป็นผู้บุกเบิกการวัดผลสัมฤทธิ์ทางการเรียนและการวัดสมรรถภาพทางสมอง แต่อย่างไรก็ตามบุคคลที่นับว่ามีความสำคัญในการเสนอหลักการและแนวคิดทฤษฎีการวัดแบบดั้งเดิมก็คือ กัลลิคเซน (Gulliksen, 1950)

1.1 ความหมายของคะแนนจริง

กัลลิคเซน (Gulliksen , 1967 : 4 - 38) นับเป็นบุคคลสำคัญที่ได้เสนอหลักการและแนวคิดเกี่ยวกับทฤษฎีการสอบวัดแบบดั้งเดิม โดยได้อธิบายคะแนนจริง (True score) เป็นสองแนวทาง

แนวทางที่ 1 ได้อธิบายถึงคะแนนดิบ (Raw score) เพื่อนำไปสู่คะแนนจริง โดยอธิบายถึงความสัมพันธ์ของคะแนนดิบ คะแนนจริงและคะแนนความคลาดเคลื่อน ดังนี้

1. คะแนนดิบ (Raw score) คือคะแนนที่สังเกตมาได้โดยตรงจากแบบสอบหรือเครื่องมือที่ใช้ในการสอบวัด โดยคะแนนดิบจะประกอบด้วยส่วนสำคัญสองส่วนคือ คะแนนจริงกับคะแนนความคลาดเคลื่อน ซึ่งเขียนเป็นสมการแสดงความสัมพันธ์ได้ว่า

$$X_i = T_i + E_i$$

เมื่อ X_i เป็นคะแนนดิบหรือคะแนนที่สังเกตมาได้ของบุคคลที่ i

T_i เป็นคะแนนจริงของบุคคลที่ i

E_i เป็นค่าความคลาดเคลื่อนในการสอบวัดแต่ละคน

2. ความคลาดเคลื่อน หมายถึงค่าที่เบี่ยงเบนไปจากค่าความจริง และเป็นความคลาดเคลื่อนเชิงสุ่มที่เกิดขึ้นในการสอบของแต่ละบุคคล โดยความคลาดเคลื่อนมีคุณสมบัติดังนี้

1) ค่าเฉลี่ยของความคลาดเคลื่อนจากการสอบมีค่าเท่ากับศูนย์ ทั้งนี้เพราะผลรวมของความคลาดเคลื่อนเชิงสุ่มของแต่ละคนมีค่าเป็นศูนย์ นั่นคือ

$$M_E = 0 \quad \text{และ} \quad (E_i) = 0$$

เมื่อ M_E เป็นค่าเฉลี่ยของความคลาดเคลื่อน

E_i เป็นค่าความคลาดเคลื่อนของคนที่ i

2) ความคลาดเคลื่อนในแบบสอบแต่ละฉบับมีความเป็นอิสระต่อกัน (Independence) กล่าวคือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างความคลาดเคลื่อนจากแบบสอบใด ๆ จะมีค่าเป็นศูนย์ นั่นคือ

$$r_{E_g E_h} = 0$$

เมื่อ E_g, E_h เป็นค่าความคลาดเคลื่อนของแบบสอบ g , h

3) ค่าความคลาดเคลื่อนจะแปรผันอิสระ ไม่ขึ้นกับคะแนนจริง กล่าวคือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างความคลาดเคลื่อนของแบบสอบหนึ่ง กับคะแนนจริงของอีกแบบสอบหนึ่ง มีค่าเป็นศูนย์

$$r_{TE} = 0$$

4) ค่าความคลาดเคลื่อนของบุคคลมีการแจกแจงเป็นปกติ (Normal curve) กล่าวคือค่าความคลาดเคลื่อนในการวัดมีลักษณะเชิงสุ่ม

5) ค่าความคลาดเคลื่อนมาตรฐานในการวัดมีลักษณะเป็นเอกพันธ์ (Homomgeneity) ในทุกค่าของคะแนนจริง กล่าวคือทุก ๆ ค่าของคะแนนจริงใด ๆ ที่นำมาประมาณค่าคะแนนดิบจะใช้ค่า S_E ค่าเดียวตลอด

3. คะแนนจริงหมายถึงคะแนนดิบหรือคะแนนที่สังเกตมาได้ที่ปราศจากความคลาดเคลื่อนหรือได้หักค่าความคลาดเคลื่อนแล้ว โดยคะแนนจริงมีคุณสมบัติดังนี้

1) คะแนนจริงมีการแจกแจงเป็นโค้งปกติ โดยค่าเฉลี่ยของคะแนนจริงเท่ากับค่าเฉลี่ยของคะแนนดิบ ($M_T = M_X$) และค่าความแปรปรวนของคะแนนจริงเท่ากับผลคูณระหว่างความแปรปรวนของคะแนนดิบกับค่าความเที่ยงของแบบสอบ คือ

$$S_t^2 = S_X^2 r_{XX}$$

2) สัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนดิบกับคะแนนจริง เท่ากับอัตราส่วนระหว่างค่าความเบี่ยงเบนมาตรฐานของคะแนนจริงและความเบี่ยงเบนมาตรฐานของคะแนนดิบ คือ

$$r_{xt} = \frac{S_t}{S_x}$$

หรือเท่ากับดัชนีความเที่ยงของแบบสอบ (Index of reliability) คือ

$$r_{xt} = \sqrt{r_{XX}}$$

<u>แนวทางที่ 2</u> ได้อธิบายถึงคะแนนจริง (True score) ที่นำไปสู่คะแนนดิบ โดยอธิบายความสัมพันธ์ของคะแนนจริงกับคะแนนดิบ ดังนี้

1. คะแนนจริง เป็นค่าเฉลี่ยของคะแนนดิบของบุคคลที่สอบแบบสอบฉบับเดียวกันหลาย ๆ ครั้ง หรือสอบแบบสอบคู่ขนานหลาย ๆ ฉบับ คือ

$$T_i = \lim_{k \to \infty} \frac{\sum_{g=1}^{n} X_{ig}}{k}$$

เมื่อ T_i คือ คะแนนจริงของบุคคลที่ i

X_i คือ คะแนนของบุคคลที่ i

n คือ จำนวนแบบสอบ

g คือ แบบสอบชุดที่ g

k คือ จำนวนครั้งที่สอบ

นอกจากนั้น ได้สรุปคุณสมบัติของคะแนนจริงไว้ว่า

$$M_T = M_X \quad ; \quad S_t^2 = S_{X_g}^2 \; r_{X_g X_h}$$

เมื่อ M_T คือ ค่าเฉลี่ยของคะแนนจริง

M_X คือ ค่าเฉลี่ยของคะแนนดิบ

S_t^2 คือ ความแปรปรวนของคะแนนจริง

S_X^2 คือ ความแปรปรวนของคะแนนจากแบบสอบ

X_g คือ คะแนนจากแบบสอบชุดที่ g

$r_{X_g X_h}$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนจากแบบสอบชุดที่ g และ h

1.2 การประมาณค่าคะแนนจริงตามวิธีแบบดั้งเดิม

ในการสอบวัด ผลของการสอบวัดที่ได้รับมาสิ่งแรกคือคะแนนดิบ แต่สิ่งที่ต้องการคือคะแนนจริงหรือคะแนนความสามารถที่แท้จริง ดังนั้นจึงได้มีการพัฒนาแนวทฤษฎีสอบวัด โดยเฉพาะวิธีการประมาณค่าความสามารถจริงขึ้นมาใช้

1.2.1 แนวการประมาณค่าตามวิธีแบบดั้งเดิม ตามแนวความคิดของกัลลิคเซน (Gulliksen , 1967 : 43) นั้นมีอยู่ 2 วิธี คือ

1) การประมาณค่าในรูปของคะแนนเบี่ยงเบนมาตรฐาน

$$t' = r_{xt}\left(\frac{S_t}{S_x}\right) X$$

เมื่อ t' คือ ค่าการประมาณคะแนนจริง

$$r_{xt} = \sqrt{r_{XX}}$$

$$S_t = S_X \sqrt{r_{XX}}$$

โดยความคลาดเคลื่อนในการประมาณมีค่าดังนี้

$$S_e = S_X \sqrt{r_{XX}} \sqrt{1 - r_{XX}}$$

2) การประมาณค่าในรูปของคะแนนจริง

$$T_i = r_{X_g X_h} X_i + (1 - r_{X_g X_h}) M_X$$

เมื่อ T_i คือ คะแนนจริงที่ประมาณได้ของคนที่ i

X_i คือ คะแนนดิบของคนที่ i

$M_X = \overline{X}$

$r_{X_g X_h} = r_{XX}$

โดยความคลาดเคลื่อนในการประมาณมีค่าดังนี้

$$S_E = S_t \sqrt{(1 - r_{X_g X_h})}$$

เมื่อ $S_t = S_x \sqrt{r_{XX}}$

$r_{X_g X_h} = r_{XX}$

1.2.2 แนวการประมาณค่าตามวิธีของลอร์ดและโนวิค

(Lord and Novick , 1974 : 64 - 69) ลอร์ดและโนวิคได้พัฒนาวิธีการประมาณค่าพารามิเตอร์ตามทฤษฎีการสอบวัดแบบดั้งเดิมขึ้น โดยให้เหตุผลว่าสามารถประมาณค่าได้จากคะแนนที่สังเกตมาได้ โดยเฉพาะอย่างยิ่งเมื่อมีจำนวนกลุ่มตัวอย่างมาก โดยอาศัยฟังก์ชั่นถดถอยเชิงเส้นตรง ดังนี้

$$R(T \mid X) = \rho_{XX'} X_i + (1 - \rho_{XX'}) M_X$$

เมื่อ R คือ ค่าการประมาณคะแนนจริงจากคะแนนดิบ

X_i คือ คะแนนดิบของคนที่ i

$\rho_{XX'} = r_{XX}$

$M_X = \bar{X}$

โดยค่าความคลาดเคลื่อนหาได้จาก

$$R(E \mid X) = (1 - \rho_{XX'})(X_i - M_X)$$

ค่าความคลาดเคลื่อนในการประมาณค่า หาได้จาก

$$6_E = 6_X \sqrt{\rho_{XX'}} \sqrt{(1 - \rho_{XX'})}$$

ค่าความคลาดเคลื่อนในการพยากรณ์ หาได้จาก

$$6_\Delta = 6_X \sqrt{1 - \rho^2_{XX'}}$$

เมื่อพิจารณาวิธีการในการประมาณค่าความสามารถจริงที่ กัลลิคเซน (Gulliksen) ลอร์ดและโนวิค (Lord and Novick) ได้เสนอไว้นั้น เป็นสมการเส้นตรงที่มีแนวคิดเหมือนกันและสอดคล้องกันกับแนวการประมาณค่าของแอเลนและเยน (Allen and Yen , 1979 : 232) โดยสมการนี้ใช้ค่าความเที่ยงของแบบสอบและค่าเฉลี่ยของคะแนนในกลุ่มเป็นตัวแปรในการกำหนดค่าความสามารถจริง ซึ่งการศึกษาค่าการประมาณความสามารถจริงตามแบบดั้งเดิมในครั้งนี้ ได้เลือกใช้สมการเชิงเส้นตรงนี้

2. ทฤษฎีการวัดแบบคุณลักษณะแฝง (Latent trait theory)

วอร์ม (Warm , 1978 : 19) ได้สรุปประวัติพัฒนาการของทฤษฎีคุณลักษณะแฝงหรือทฤษฎีการตอบสนองข้อสอบว่า ได้พัฒนามาตั้งแต่ปี 1942 - 1943 ซึ่งกล่าวถึงเส้นแสดงความสัมพันธ์ระหว่างคุณลักษณะกับความน่าจะเป็นในการตอบถูกหรือตอบผิด (Trace line) เฟอร์กูซัน (Ferguson , 1942) และลอว์เลย์ (Lawlay , 1943) โดยมีหลักการว่า ผลการสอบจากแบบสอบใด ๆ ของผู้สอบได้มาจากความสามารถที่แท้จริงของผู้สอบที่ตอบแบบสอบนั้น และสามารถเขียนเป็นคุณลักษณะข้อสอบแต่ละข้อได้โดยฟังก์ชันทางคณิตศาสตร์ที่แสดงการแจกแจงเป็นแบบปกติ (Normal ogive model) ต่อมา ลอร์ด (Lord , 1952) ได้พัฒนาขึ้นใหม่โดยใช้ชื่อว่า ทฤษฎีโค้งลักษณะข้อสอบ (Item characteristic curve) และต่อมาได้เรียกชื่อใหม่ว่า ทฤษฎีการตอบสนองข้อสอบ (Item response theory) จนกระทั่งในปี 1960 ราสช์ (Rasch) ได้พัฒนารูปแบบคุณลักษณะแฝงประเภท 1 พารามิเตอร์หรือเรียกกันว่ารูปแบบของราสช์ (Rasch model) และในปี 1968 เบิร์นบอม (Birnbaum) ได้พัฒนารูปแบบคุณลักษณะแฝง แบบ 2 และ 3 พารามิเตอร์ โดยใช้ฟังก์ชันการแจกแจงโลจิสติค (Logistic model) ขึ้นแทนฟังก์ชันเดิม ทั้งนี้เนื่องจากลักษณะโค้งคล้ายกันแต่มีความยุ่งยากน้อยกว่า ประกอบกับในปัจจุบันได้มีพัฒนาด้านเทคโนโลยีทางคอมพิวเตอร์ จึงทำให้การประมาณค่าพารามิเตอร์ง่ายและชัดเจนขึ้น ซึ่งเป็นเหตุทำให้รูปแบบทฤษฎีคุณลักษณะแฝงเป็นที่สนใจและพัฒนาให้มีความแจ่มชัดในรายละเอียดมากขึ้น

2.1 ความหมายของคะแนนจริง

ทฤษฎีคุณลักษณะแฝงเป็นทฤษฎีการตอบสนองข้อสอบที่พยายามอธิบายลักษณะของข้อสอบด้วยพารามิเตอร์ของข้อสอบ และอธิบายความสามารถของผู้สอบด้วยพารามิเตอร์ของผู้สอบ โดยกล่าวถึงลักษณะ (Trait) ที่ซ่อนเร้นอยู่ในตัว

บุคคลซึ่งไม่สามารถวัดได้โดยตรง แต่สามารถอ้างอิงไปสู่คุณลักษณะแฝงหรือความสามารถนั้นได้ โดยอาศัยความสัมพันธ์ระหว่างโอกาสการตอบสนองข้อสอบของผู้ตอบข้อสอบนั้นกับความสามารถของผู้สอบ ดังนั้นจึงอาจกล่าวได้ว่า ในการอธิบายคะแนนจริงหรือความสามารถแท้จริงของผู้สอบตามทฤษฎีคุณลักษณะแฝงนั้น อาศัยความสัมพันธ์ระหว่างความสามารถที่แท้จริงกับการตอบข้อสอบ โดยการตอบข้อสอบแต่ละข้อแสดงได้โดยโค้งลักษณะข้อสอบ

สงบ ลักษณะ (สงบ ลักษณะ , 2525 : 49) ได้กล่าวถึงทฤษฎีคุณลักษณะแฝงว่า มุ่งหาความสัมพันธ์ระหว่างความสามารถที่แท้จริงกับการตอบข้อสอบภายใต้รากฐานของความเชื่อที่ว่า

1. ผู้สอบแบบสอบใด ๆ ย่อมมีความสามารถ (Ability) หรือคุณลักษณะแฝงที่แบบสอบมุ่งวัด ในการตอบข้อสอบผู้สอบจะต้องใช้ความสามารถที่มีอยู่ตอบสนองข้อสอบเหล่านั้น

2. ผู้สอบที่มีความสามารถในระดับสูงย่อมตอบข้อสอบได้ถูกต้องเป็นจำนวนมากกว่าผู้ที่มีความสามารถต่ำ หรือกล่าวได้ว่า ผลการสอบแบบสอบใด ๆ นั้นขึ้นอยู่กับความสามารถของผู้สอบนั้น ๆ

3. ถ้าข้อตกลงเบื้องต้นของการสอบตรงตามทฤษฎี และถ้าสามารถคำนวณค่าพารามิเตอร์ได้เหมาะกับรูปแบบแล้ว ค่าพารามิเตอร์ของข้อสอบ ซึ่งได้แก่ ค่าความยาก ค่าอำนาจจำแนก หรือค่าสัมประสิทธิ์การเดาจะเป็นค่าที่ไม่เปลี่ยนแปลงไปตามกลุ่มความสามารถของผู้สอบ

4. เมื่อทราบลักษณะการตอบข้อสอบแต่ละข้อของผู้ใดผู้หนึ่งแล้ว สามารถคำนวณหาค่าความสามารถที่แท้จริงของบุคคลนั้น ๆ ได้ โดยค่าความสามารถจริงมีความสัมพันธ์โดยตรงกับคะแนนจริง และในการคำนวณค่าความสามารถที่แท้จริงอาจใช้ข้อสอบข้อใดก็ได้ที่วัดในสิ่งเดียวกันซึ่งเป็นลักษณะของความเป็นอิสระของข้อสอบ

2.2 การประมาณค่าคะแนนจริงตามวิธีแบบคุณลักษณะแฝง

ทฤษฎีแบบคุณลักษณะแฝงพยายามหาความสามารถที่แท้จริงของผู้สอบแต่ละคนโดยใช้ความสามารถที่แท้จริงนี้เชื่อมโยงไปสู่คะแนนจริงของผู้สอบแต่ละคนด้วยฟังก์ชั่นทางคณิตศาสตร์ (Mathematical function) ที่แสดงโค้งลักษณะเฉพาะของข้อสอบ (Item characteristic curve)

ทฤษฎีคุณลักษณะแฝงหรือทฤษฎีตอบสนองข้อสอบ ได้มีการพัฒนารูปแบบต่าง ๆ หลายรูปแบบ โดยแต่ละรูปแบบจะแตกต่างกันไปตามฟังก์ชั่นทางคณิตศาสตร์และจำนวนพารามิเตอร์ที่ใช้อธิบายโค้งลักษณะข้อสอบของแต่ละรูปแบบ (Warm , 1978 : 23 - 25 ; Hambleton and Cook , 1977 : 81 - 82)คือ

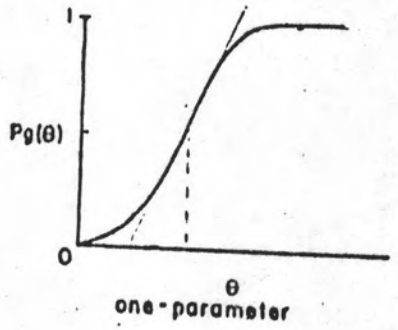
2.2.1 รูปแบบการแจกแจงปกติ (Normal ogive model) มีสามรูปแบบ คือ

1) แบบ 1-พารามิเตอร์ (One-rarameter) เป็นรูปแบบที่มีค่าพารามิเตอร์เพียงตัวเดียว คือ ค่า b ที่แสดงลักษณะข้อสอบ เมื่อ θ คือความสามารถของผู้สอบ แสดงได้ด้วยฟังก์ชั่นนี้

$$P_g(\theta) = \int_{-\infty}^{(\theta - b_g)} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{t^2}{2}} \, dt$$

เมื่อ g = 1, 2, 3, ...

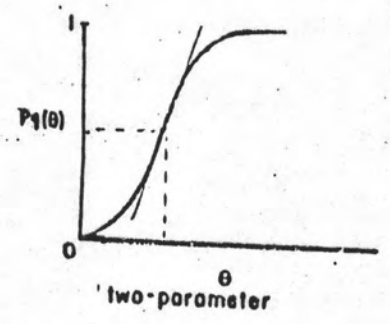
และแสดงโค้งลักษณะข้อสอบได้ดังนี้



2) แบบ 2-พารามิเตอร์ (Two-parameters) เป็นรูปแบบที่ประกอบด้วย 2 พารามิเตอร์ คือ a และ b ที่แสดงลักษณะข้อสอบ เมื่อ θ คือความสามารถของผู้สอบ มีฟังก์ชั่นดังนี้

$$P_g(\theta) = \int_{-\infty}^{a(\theta-b_g)} \frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2}\, dt$$

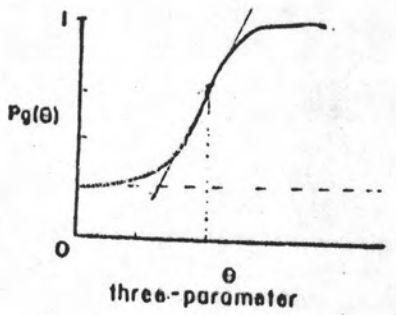
และมีโค้งลักษณะข้อสอบดังนี้



3) แบบ 3-พารามิเตอร์ (Three-parameters) เป็นรูปแบบที่มีค่าพารามิเตอร์ลักษณะข้อสอบ 3 ค่า คือ a , b และ c เมื่อ θ คือความสามารถของผู้สอบ มีฟังก์ชั่น ดังนี้

$$P_{g.}(\theta) = C_{g.} + (1 - C_{g.}) \int_{-o}^{a_{g.}(\theta - b_{g.})} \frac{1}{\sqrt{2\pi}} \cdot e^{-t^2/2} \, dt$$

และมีโค้งลักษณะข้อสอบ ดังนี้



เมื่อ a เป็นค่าอำนาจจำแนกหรือสัดส่วนของความชัน ณ จุดเปลี่ยนโค้ง

b เป็นค่าความยากของข้อสอบ หรือตำแหน่งของข้อสอบบนเส้นความสามารถ

c เป็นค่าโอกาสการเดาถูก

2.2.2 แบบโลจิสติค (Logistic model) เป็นรูปแบบที่พัฒนาขึ้นเพื่อใช้แทนแบบการแจกแจงปกติ เพราะมีลักษณะโค้งใกล้เคียงกันแต่มีความง่ายในการประมาณค่ามากกว่า รูปแบบโลจิสติคมีสามรูปแบบ คือ

1. แบบ 1 - พารามิเตอร์ (1-parameter) หรือรูปแบบของราสช์ (Rasch model) เป็นรูปแบบที่ราสช์ได้พัฒนาขึ้นตั้งแต่ปี 1960 โดยมีค่าพารามิเตอร์แสดงลักษณะข้อสอบเพียงตัวเดียวคือค่าความยาก (b) เขียนเป็นฟังก์ชันได้ดังนี้

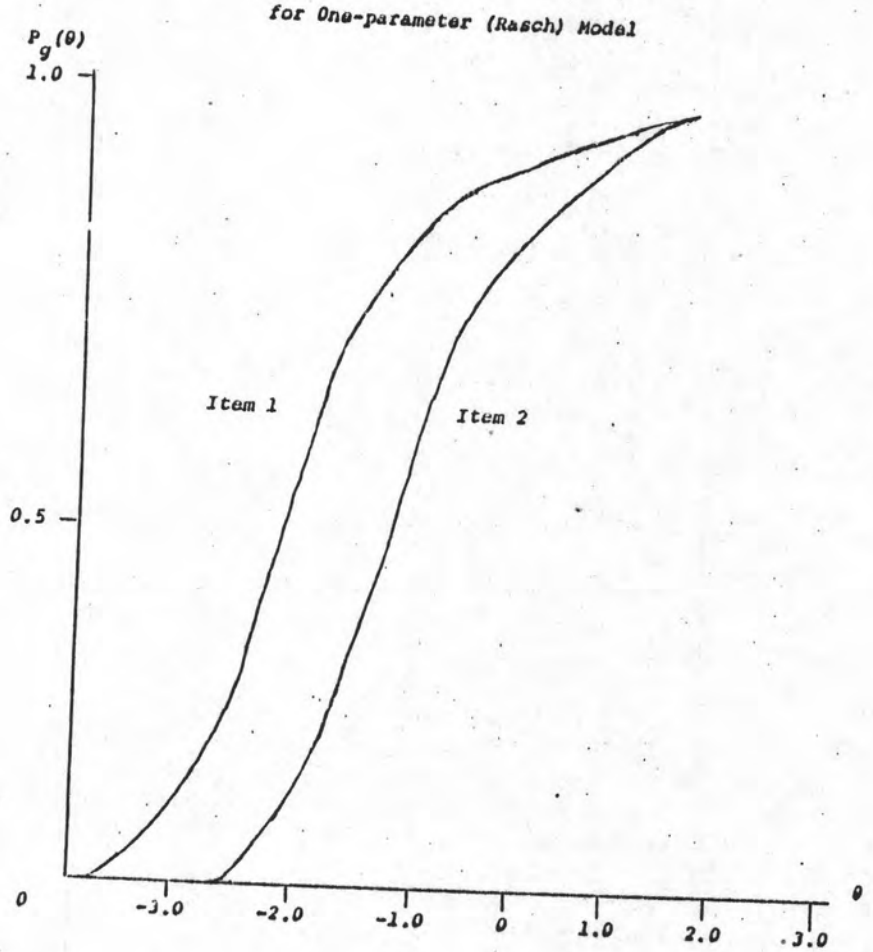
$$P_g(\theta) = \frac{e^{D\bar{a}(\theta - b_g)}}{1 + e^{D\bar{a}(\theta - b_g)}}$$

หรือ

$$= \frac{e^{(\theta - b_g)}}{1 + e^{(\theta - b_g)}} = \frac{EXP(\theta - b_g)}{1 + EXP(\theta - b_g)}$$

Typical Item Characteristic Curves

for One-parameter (Rasch) Model



2) แบบ 2-พารามิเตอร์ (2 - parameters) เป็นรูปแบบที่เบิร์นบอม (Birnbaum) ได้พัฒนาขึ้นในปี 1968 มีค่าพารามิเตอร์แสดงลักษณะข้อสอบ 2 ค่า คือ ค่าอำนาจจำแนก (a) และค่าความยาก (b) แสดงได้โดยฟังก์ชันนี้

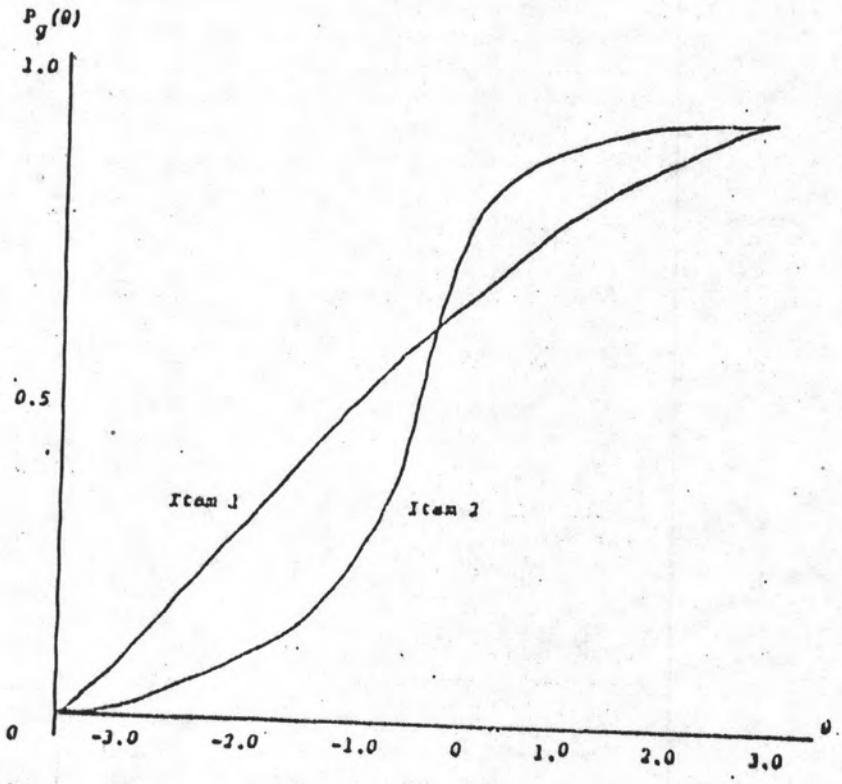
$$P_g(\theta) = \frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}}$$

หรือ

$$= \frac{\text{EXP}\,(Da_g(\theta - b_g))}{1 + \text{EXP}\,(Da_g(\theta - b_g))}$$

TYPICAL ITEM CHARACTERISTIC CURVES

FOR TWO-PARAMETER MODEL

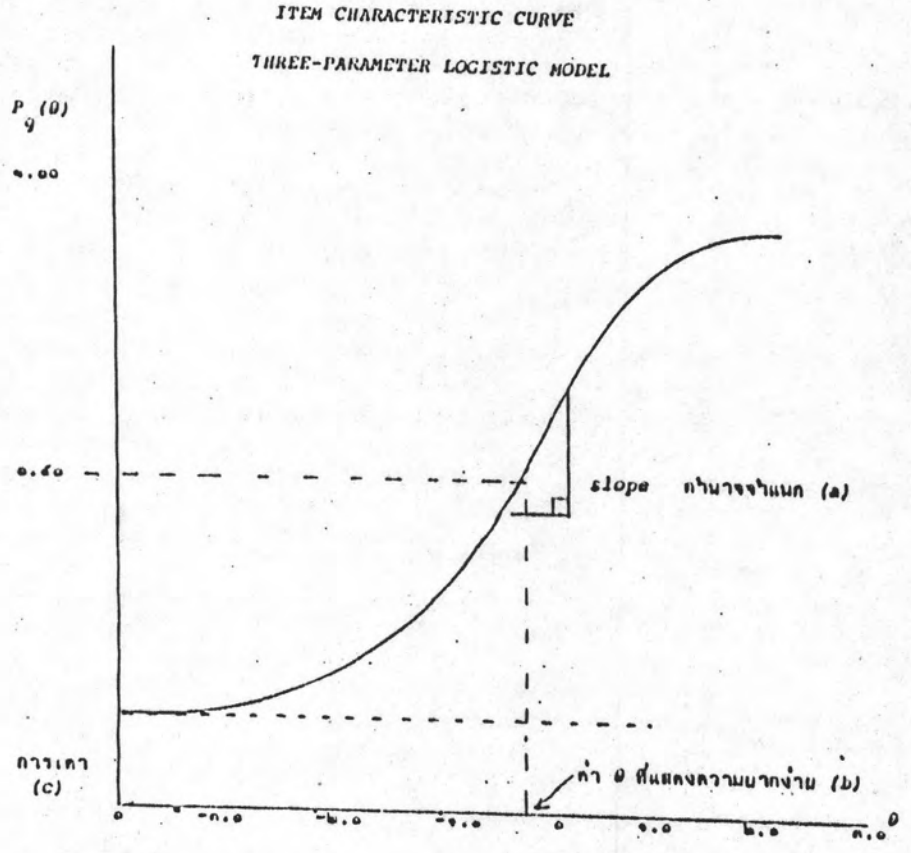


3) แบบ 3-พารามิเตอร์ (3 - parameters) เป็นรูปแบบที่เบิร์นบอมได้พัฒนาขึ้น โดยมีค่าพารามิเตอร์แสดงคุณลักษณะข้อสอบ 3 ค่า โดยเพิ่มค่าการเดา (c) ใน 2-พารามิเตอร์ เขียนเป็นฟังก์ชันได้ดังนี้

$$P_g(\theta) = c_g + (1 - c_g)\frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}}$$

หรือ

$$= c_g + (1 - c_g)\frac{EXP(Da_g(\theta - b_g)}{1 + EXP(Da_g(\theta - b_g)}$$



เมื่อ θ คือระดับความสามารถแท้จริงที่คำนวณจากคะแนนจริง แต่ปรับหน่วยให้เป็นคะแนนมาตรฐาน

$P_g(\theta)$ คือโอกาสที่ผู้สอบที่มีความสามารถ θ จะทำข้อสอบ g ได้ถูกต้อง

D คือค่าคงที่ มีค่าเท่ากับ 1.7

a_g คือค่าอำนาจจำแนก ซึ่งมีค่าเป็นสัดส่วนโดยตรงกับค่าความชัน (Slope) ของโค้ง ณ จุดเปลี่ยนโค้ง โดยความชันมีค่าเท่ากับ a/2 และมีค่าเป็นบวกไม่เกิน 2.00

b_g คือค่าความยากที่แสดงระดับความสามารถ θ ที่จุดโค้งความชันที่สูงสุด หรือในกรณีที่ไม่มีการเดา ค่า b_g คือ θ ณ จุดโอกาสการตอบมีค่า 0.50 ดังนั้น มีค่าเป็นหน่วยเดียวกับ θ ซึ่งมีค่าเป็นบวกหรือลบ ถ้าเป็นลบแสดงว่าข้อสอบง่าย และถ้าเป็นบวกแสดงว่าข้อสอบยาก

c_g คือโอกาสการเดา ซึ่งเป็นค่าโอกาสที่ผู้สอบที่มีความสามารถต่ำมีโอกาสจะทำข้อสอบถูก

ค่าพารามิเตอร์ในทฤษฎีการตอบสนองข้อสอบ แบ่งเป็นค่าพารามิเตอร์ของข้อสอบ (Item parameter) ซึ่งได้แก่ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าสัมประสิทธิ์การเดา (c) ส่วนค่าพารามิเตอร์ของผู้สอบ (Examinee parameter) ซึ่งได้แก่ระดับความสามารถของผู้สอบ (θ) นั้น มีพิสัยของค่าพารามิเตอร์ดังนี้ (Hambleton and Cook , 1977 : 81)

i 10310733

ค่า b_g มีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติจะมีค่าอยู่ระหว่าง - 2 ถึง +2 โดยค่า - 2 แสดงลักษณะข้อสอบว่างง่ายมาก และค่า +2 แสดงลักษณะข้อสอบว่ายากมาก

ค่า a_g มีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติจะมีค่าตั้งแต่ 0 ถึง + 2 ทั้งนี้เพราะค่าที่เป็นลบเป็นข้อสอบที่ไม่ดี ใช้ไม่ได้ ต้องตัดทิ้ง โดยค่า 0 แสดงว่าข้อสอบไม่มีค่าอำนาจจำแนกหรือค่าอำนาจจำแนกต่ำ และค่า + 2 แสดงว่าข้อสอบมีค่าอำนาจจำแนกสูง

ค่า θ มีค่าอยู่ระหว่าง - 3 ถึง +3 โดยค่า - 3 แสดงว่ามีระดับความสามารถต่ำ และค่า +3 แสดงว่ามีระดับความสามารถสูง

ในการประมาณค่าพารามิเตอร์ของข้อสอบและของผู้สอบ ผจงจิต อินทสุวรรณ (ผจงจิต อินทสุวรรณ , 2525 : 61-62) สรุปว่า สำหรับแบบโลจิสติคสามพารามิเตอร์นั้น จะประมาณด้วยเทคนิควิธีโอกาสการเกิดสูงสุด (Maximum likelihood) นอกจากนี้ยังได้มีการดัดแปลงฟังก์ชั่นในเทคนิควิธีการประมาณค่าพารามิเตอร์ให้ใช้ได้กับข้อสอบที่เว้นไม่ตอบได้ จึงมีการเสนอวิธีการประมาณค่าที่ง่ายกว่า โดยการพัฒนาโปรแกรมคอมพิวเตอร์ชื่อ LOGIST เพื่อประมาณค่าพารามิเตอร์ของความสามารถของข้อสอบในเวลาเดียวกัน โดยใช้วิธีการ iterative เมื่อไม่ทราบค่าพารามิเตอร์ทั้งความสามารถและข้อสอบมาก่อน

ส่วนพารามิเตอร์ในรูปแบบของราสช์ก็สามารถประมาณได้โดยวิธีการเดียวกัน และเมื่อความสามารถของกลุ่มตัวอย่างมีการกระจายเป็นแบบปกติ วิธีการประมาณก็ยังนำมาใช้ได้ ทั้งนี้เพราะนอกจากจะประหยัดกว่าแล้วยังให้ผลใกล้เคียงอีกด้วย ซึ่งแสดงให้เห็นค่าการประมาณว่ามีคุณสมบัติที่ดี นอกจากนั้นวิธีการประมาณค่าพารามิเตอร์ยังได้มีการพัฒนาโปรแกรมคอมพิวเตอร์ชื่อ BICAL ขึ้นมาใช้

คะแนนจริงตามทฤษฎีคุณลักษณะแฝงประมาณค่าได้จากสมการ

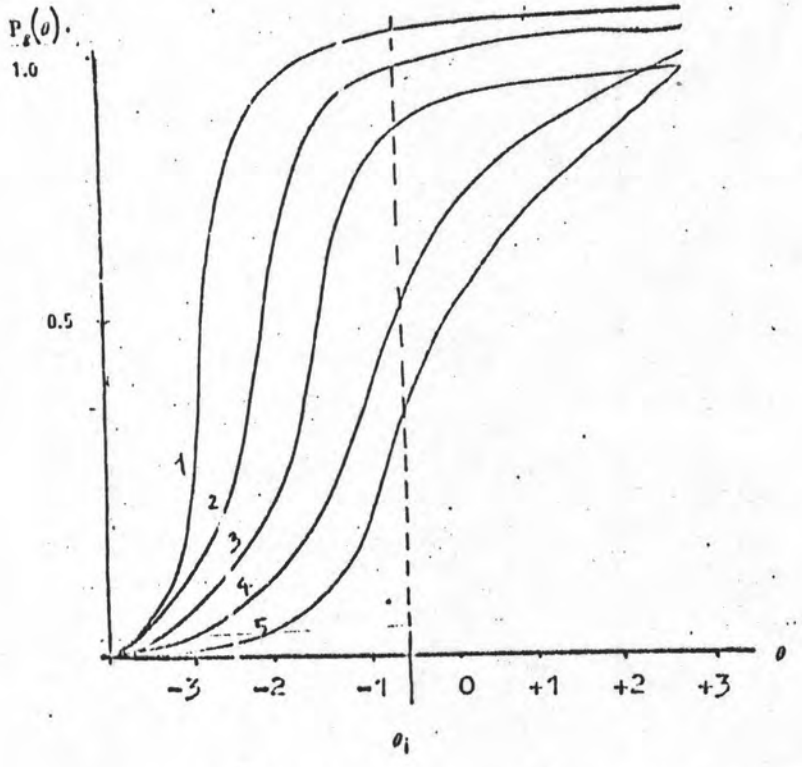
$$T_i = \sum_{g=1}^{n} P_g(\theta_i)$$

เมื่อ T_i เป็นคะแนนจริงของบุคคลที่มีความสามารถ θ

P_g เป็นความน่าจะเป็นในการตอบข้อสอบ g ได้ถูก

n เป็นจำนวนข้อสอบทั้งหมด

นั่นคือ ทฤษฎีคุณลักษณะแฝง นิยามคะแนนจริงไว้ว่า คะแนนจริงของบุคคลที่มีความสามารถ θ คือผลรวมของความน่าจะเป็นในการทำรายข้อ i ได้ถูกต้อง ซึ่งความน่าจะเป็นดังกล่าวพิจารณาได้จากโค้งแสดงลักษณะข้อสอบทั้งฉบับ ดังตัวอย่าง (สงบ ลักษณะ , 2525 : 25)



คะแนนจริงของผู้ที่มีความสามารถ θ_i ก็คือ ผลรวมของ $P_1(\theta_i)$, $P_2(\theta_i)$, $P_3(\theta_i)$, $P_4(\theta_i)$, $P_5(\theta_i)$.

3. งานการวิจัยที่เกี่ยวข้อง

เมื่อพิจารณาจากเอกสารและผลงานการวิจัยในด้านการประมาณค่าความสามารถจริงที่เกี่ยวข้องกับลักษณะการแจกแจงของค่าประมาณความสามารถจริงจากแบบดั้งเดิมกับแบบคุณลักษณะแฝง 1 พารามิเตอร์ 2 พารามิเตอร์และ 3 พารามิเตอร์นั้น ยังไม่มีผลงานการวิจัยโดยตรง จะมีก็แต่เพียงข้อเสนอแนะและความคิดเห็นเกี่ยวกับหลักการของทฤษฎีในการประมาณค่าเท่านั้น ผลงานการวิจัยส่วนใหญ่เป็นเรื่องของการประมาณค่าพารามิเตอร์ เพื่อนำไปสู่การอธิบายค่าความสามารถจริง และวิธีการเทียบคะแนนซึ่งอาศัยเทคนิควิธีที่คล้ายกับการประมาณค่า เช่น แบบดั้งเดิมมีการเทียบคะแนนโดยใช้ อีควิเปอร์เซนไทล์ สมการเชิงเส้นตรง เป็นต้น ส่วนแบบคุณลักษณะแฝงจะศึกษาเกี่ยวกับการเทียบคะแนนหรือการประมาณค่าในแบบใดแบบหนึ่งเพื่อเปรียบเทียบค่าที่ประมาณได้ และการประมาณค่าพารามิเตอร์ที่เกี่ยวข้องกับลักษณะของแบบสอบ เช่นค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ในแต่ละแบบคุณลักษณะแฝง ดังงานการวิจัยนี้

คุกและไอเนอร์ (Cook and Eignor , 1981) (อ้างจาก ภาวิณี ศรีสุขวัฒนานันท์ , 2529 : 74) ได้ศึกษารูปแบบทฤษฎีการตอบสนองข้อสอบ (IRT) เพื่อนำไปสู่การแก้ปัญหาการทดสอบในการปฏิบัติ โดยการเปรียบเทียบผลการใช้รูปแบบดั้งเดิมและแบบ IRT รวม 4 วิธี คือ วิธีเชิงเส้นตรง วิธีอีควิเปอร์เซนไทล์ธรรมดา วิธีอีควิเปอร์เซนไทล์ที่ใช้ประมาณความถี่ และวิธี IRT ประมาณค่าคะแนนจริง โดยเปรียบเทียบความแตกต่างของค่าที่ได้ในสองลักษณะคือ หาความสอดคล้องสัมพันธ์โดยพิจารณาจากภาพการแจกแจงของคะแนนที่ใช้วิธีเทียบแบบดั้งเดิมในแต่ละวิธีกับการแจกแจงของคะแนนที่ได้จากแบบ IRT และหาดัชนีความแตกต่างระหว่างค่าที่ประมาณได้กับเกณฑ์ที่กำหนด ผลการศึกษาพบว่า วิธีการแบบดั้งเดิมทั้งสามวิธีให้ผลสอดคล้องกับแบบ IRT 3 พารามิเตอร์ นอกจากนั้นยังค้นพบว่า

แบบสอบที่มีความยากต่างกัน กลุ่มตัวอย่างที่ไม่สมมูลและไม่ได้สุ่มนั้น การแจกแจงของคะแนนแต่ละชุดของกลุ่มตัวอย่างจะมีลักษณะเหมือนกัน

โคเลนและวิทนีย์ (Kolen and Whitney , 1982 : 279) ได้ศึกษาเปรียบเทียบความถูกต้องของการประมาณค่าในการเทียบคะแนนแบบดั้งเดิมกับแบบคุณลักษณะแฝง รวม 4 วิธีคือ สมการเชิงเส้นตรง เปอร์เซนไทล์ โค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์และ 3 พารามิเตอร์ โดยใช้แบบสอบ 12 ฉบับ สอบกลุ่มตัวอย่างฉบับละ 200 คน พบว่า วิธีการแบบเปอร์เซนไทล์และแบบ 3 พารามิเตอร์ให้ผลยังไม่คงที่ ส่วนวิธีการแบบสมการเชิงเส้นตรงและแบบ 1 พารามิเตอร์ได้ผลคงที่ และจากการศึกษาสรุปได้ว่า ความถูกต้องของวิธีการเทียบคะแนนนั้นขึ้นอยู่กับลักษณะของแบบสอบ รูปแบบการเทียบ และขนาดของกลุ่มตัวอย่าง นอกจากนั้นยังเสนอแนะจากผลการวิจัยว่า ถ้าขนาดกลุ่มตัวอย่างเล็กควรใช้วิธีการเชิงเส้นตรง

เบเกอร์ (Baker , 1987 . 111 - 138) ได้สรุปประเด็นและแนวคิดจากการศึกษาของผู้อื่นที่เกี่ยวกับทฤษฎีการตอบสนองข้อสอบไว้หลายประเด็น คือ

1. จากการศึกษาเอกสารทางทฤษฎีที่เกี่ยวกับคุณสมบัติของค่าประมาณค่าพารามิเตอร์ สรุปได้ว่า แบบ 1 พารามิเตอร์ และ 2 พารามิเตอร์มีโค้งลักษณะข้อสอบดีกว่าแบบ 3 พารามิเตอร์ นอกจากนั้นทั้งทางด้านทฤษฎีและข้อมูลเชิงประจักษ์ที่เกี่ยวกับค่าความคลาดเคลื่อนมาตรฐานนั้น แบบ 1 พารามิเตอร์และ 2 พารามิเตอร์มีค่าน้อยกว่าแบบ 3 พารามิเตอร์

2. เทคนิคในการประมาณค่าพารามิเตอร์สามารถกระทำได้หลายรูปแบบ ทั้งนี้ขึ้นอยู่กับการเลือกใช้โปรแกรมคอมพิวเตอร์ กล่าวคือในการประมาณค่าพารามิเตอร์ตามทฤษฎีคุณลักษณะแฝงนั้นมีโปรแกรมคอมพิวเตอร์หลายโปรแกรมตามรูปแบบหรือฟังก์ชั่นทางคณิตศาสตร์ที่ใช้เป็นเกณฑ์ในการพิจารณา เช่น โปรแกรม BICAL

(Wright and Mead , 1976) ใช้ประมาณค่าแบบ 1 พารามิเตอร์ โปรแกรม LOGIST (Wingersky , Baton and Lord,1982) ใช้ประมาณค่าแบบ 1 - 2 - 3 พารามิเตอร์ โปรแกรม PML (Gustafsson , 1980) ใช้ประมาณค่าแบบ 1 พารามิเตอร์ โปรแกรม BILOG (Mislevy and Bock , 1982) ใช้ประมาณค่าแบบ 1 และ 2 พารามิเตอร์ โปรแกรม OGIVIA และ ANCILLES (Urry , 1976) ใช้ประมาณค่าแบบ 3 พารามิเตอร์ แต่ โปรแกรม BICAL และ LOGIST เป็นโปรแกรมที่นิยมใช้และได้รับการพัฒนาในการใช้มานานกว่าชนิดอื่น

3. รี (Ree, 1979) ได้ศึกษาเปรียบเทียบค่าประมาณพารามิเตอร์ โดยใช้โปรแกรม ANCILLES , OGIVIA , LOGISTและ การประมาณจากค่าความสัมพันธ์แบบ biserial โดยจัดกระทำกับกลุ่มตัวอย่าง 2,000 คน ปรากฏว่าแบบ ANCILLESการแจกแจงของค่า θ เป็นแบบ uniform และ θ มีค่า -2.5 ถึง 2.5 แบบ OGIVIA มีการแจกแจงเป็นแบบ uniform และแบบ LOGIST มีการแจกแจงเป็นโค้งปกติ และจากการศึกษาครั้งนี้สรุปว่า ค่าที่ได้จากโปรแกรม OGIVIA และ ANCILLES ค่อนข้างจะเหมาะกว่าแบบ LOGIST โดยค่าความสัมพันธ์ของคะแนนดิบกับค่า θ มีค่า 0.936 ถึง 0.977 ซึ่งสูงกว่าค่าความสัมพันธ์ของค่า θ ที่ได้จากโปรแกรมทั้งสาม (0.935 ถึง 0.974) นอกจากนั้นยังสรุปเกี่ยวกับเวลาที่ใช้ในการวิเคราะห์แต่ละโปรแกรมว่า แบบ LOGIST ใช้เวลา 2,061 วินาที แบบ ANCILLES ใช้เวลา 296 วินาที และแบบ OGIVIAใช้เวลา 180 วินาที ส่วนการแปลงคะแนนแบบดั้งเดิมนั้นใช้เวลาประมาณ 38 นาที

4. จากการศึกษาของสวามินาธานและจิฟฟอร์ด (Swaminathan and Gifford , 1983) ในการทดลองข้อมูล (simulation) ในการตรวจสอบความคงที่ (consistency) และความลำเอียง (bias) ของการประมาณค่าพารามิเตอร์แบบ 3 พารามิเตอร์โดยใช้โปรแกรม LOGIST จากแบบสอบ

ที่มีจำนวนข้อต่างกัน คือ 10 ข้อ 15 ข้อ และ 20 ข้อ และจำนวนกลุ่มตัวอย่างที่ต่างกันคือ 50 คน 200 คน และ 1,000 คน พบว่า จำนวนข้อสอบและจำนวนตัวอย่างมากขึ้นเส้นถดถอย (regression line) ของความสัมพันธ์ของค่าความยากและค่าอำนาจจำแนกกับค่าพารามิเตอร์ที่ประมาณเอียงเป็นมุม 45 องศา ซึ่งแสดงว่าการประมาณค่าความยากและค่าอำนาจจำแนกแบบ 3 พารามิเตอร์นั้นมีความคงที่ ส่วนการตรวจสอบความลำเอียงของค่าพารามิเตอร์ของข้อสอบนั้น จำนวนข้อสอบ 20 ข้อ จำนวนตัวอย่าง 200 คน ให้ค่าน้อยกว่าปกติ (over estimate of small value)

ลอร์ด (Lord , 1983) ได้ศึกษาข้อมูลในการตรวจสอบความลำเอียงของฐานโค้ง (asymptotic biases) ในการประมาณค่าพารามิเตอร์แบบ 3 พารามิเตอร์ โดยใช้จำนวนข้อสอบ 90 ข้อ จำนวนตัวอย่าง 2,995 คน จากแบบสอบความถนัดด้านภาษาพบว่า ในกรณีความยากง่ายของข้อสอบนั้น ถ้าข้อสอบง่ายและยากง่ายปานกลางนั้นความลำเอียงมีค่าเป็นลบ แต่ถ้าข้อสอบยากความลำเอียงมีค่าเป็นบวก ส่วนในการประมาณค่าอำนาจจำแนกนั้นความลำเอียงมีค่าเป็นบวก และการประมาณค่าโอกาสการเดานั้นความลำเอียงมีค่าเป็นลบ โดยลอร์ดสรุปว่าในลักษณะทั่วไปถ้าค่าพารามิเตอร์ของข้อสอบมีค่าความคลาดเคลื่อนมาตรฐาน (Standard error) มากแล้วจะมีความลำเอียงมากด้วย

ส่วนในการศึกษาของไรท์และดักลาส (Wright and Douglas , 1977) ในการตรวจสอบความลำเอียงในการประมาณค่าความยากโดยใช้แบบ 1 พารามิเตอร์นั้น สรุปว่ามีความลำเอียงในการประมาณค่า

5. ในการศึกษาของ ฮูลินและคณะ (Hulin and others, 1982) ในการประมาณค่าพารามิเตอร์ของข้อสอบโดยใช้แบบ 2 และ 3 พารามิเตอร์ พบว่า ถ้าจำนวนข้อสอบและจำนวนตัวอย่างน้อย การประมาณค่าพารามิเตอร์นั้นขึ้นอยู่กับแบบประมาณค่าที่ใช้ กล่าวคือแบบ 2 พารามิเตอร์ให้ค่าการประมาณที่เหมาะเมื่อจำนวนข้อสอบ 30 ข้อ จำนวนตัวอย่าง 500 คน แต่แบบ 3 พารามิเตอร์

ให้ค่าเหมาะเมื่อจำนวนข้อสอบ 60 ข้อ จำนวนตัวอย่าง 1,000 คน นอกจากนั้น ค่าความสัมพันธ์ระหว่างค่าที่ประมาณได้กับค่าพารามิเตอร์ของข้อสอบนั้นแบบ 2 พารามิเตอร์มีค่าสูงกว่าแบบ 3 พารามิเตอร์

6. จากการศึกษาของไทซ์เซนและไวเนอร์ (Tissen and Wainer, 1982) เกี่ยวกับความคลาดเคลื่อนมาตรฐานในการประมาณค่าพารามิเตอร์โดยใช้แบบ 1 พารามิเตอร์ 2 พารามิเตอร์และ 3 พารามิเตอร์ ภายใต้ข้อตกลงที่ว่า ค่าความสามารถจริง(θ) ของผู้สอบมีการแจกแจงเป็นแบบปกติ สรุปว่า ถ้าข้อสอบง่ายและค่าอำนาจจำแนกต่ำแล้วค่าความคลาดเคลื่อนมาตรฐานแบบ 2 พารามิเตอร์ไม่เท่ากับ 3 พารามิเตอร์ ค่าความคลาดเคลื่อนมาตรฐานของโค้งแบบ 1 พารามิเตอร์มีความคงที่น้อยกว่าแบบ 2 พารามิเตอร์และ 3 พารามิเตอร์ ส่วนแบบ 3 พารามิเตอร์นั้นค่าความคลาดเคลื่อนมาตรฐานจะเหมาะสมในกรณีที่พิสัยของค่าความยากและค่าอำนาจจำแนกน้อยและค่าโอกาสการเดาต่ำเท่านั้น

ในการศึกษาการประมาณค่าความสามารถจริงจากแบบสอบที่มีความยากง่ายต่างกันนั้น โคเลน (Kolen , 1981 : 3) ได้สรุปผลงานวิจัยต่าง ๆ ว่า วิธีแบบเชิงเส้นตรง และแบบของราส์ชเป็นวิธีที่ยังไม่ดีพอ แต่วิธีแบบ 3 พารามิเตอร์เหมาะกับแบบสอบที่มีความยากต่างกัน และจากการศึกษาวิธีการประมาณค่าแบบประเพณีนิยมหรือแบบดั้งเดิมกับแบบคุณลักษณะแฝง 1 และ 3 พารามิเตอร์โดยใช้แบบสอบที่มีความยากเท่าเทียมกันและความยากที่ต่างกันนั้น แบบ 1 พารามิเตอร์ไม่เหมาะกับแบบสอบที่มีความยากต่างกัน และแบบ 3 พารามิเตอร์ยังมีปัญหาในการประมาณค่าเกี่ยวกับปลายฐานโค้ง (Lower asymptote parameter) ส่วนแบบดั้งเดิมตามวิธีแบบเปอร์เซนไทล์นั้นเหมาะกับแบบสอบที่มีความยากต่างกัน