

บทที่ 2

แนวความคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ในเนื้อหาต่าง ๆ ดังนี้

- 2.1 ความเป็นมาของทฤษฎีคุณลักษณะแฝง
- 2.2 หลักการของทฤษฎีคุณลักษณะแฝง
- 2.3 ชนิดของแบบทดสอบปรัมาตร
- 2.4 วิธีการให้คะแนนในแบบทดสอบปรัมาตร
- 2.5 ข้อดีและข้อจำกัดของแบบทดสอบปรัมาตร
- 2.6 ลักษณะของการวิจัยที่เกี่ยวกับแบบทดสอบเทเลอร์
- 2.7 เกณฑ์ที่ใช้ในการประเมินแบบทดสอบเทเลอร์
- 2.8 งานวิจัยที่เกี่ยวข้องกับแบบทดสอบปรัมาตร

2.1 ความเป็นมาของทฤษฎีคุณลักษณะแฝง

จุดเริ่มต้นของทฤษฎีคุณลักษณะแฝง เริ่มจากการเสนอผลงานของ Ferguson และ Lawley ในปี ค.ศ.1942 และ ค.ศ.1943 (อ้างอิงใน Lord and Novick, 1968) ซึ่งถือว่าผลงานของทั้งสองท่านนี้เป็นแนวคิดและหลักการเบื้องต้นที่สำคัญของทฤษฎีคุณลักษณะแฝง ในระยะต่อมาได้มีอีกหลายท่านเสนอผลงานที่มีแนวคิดเดียวกันนี้ เช่น Brogden(1946), Tucker(1946), Carroll (1950), และ Cronbach and Warrington (1952) ต่อมาในปี ค.ศ.1952 Lord ได้เสนอทฤษฎีในรูปโค้งคุณลักษณะข้อสอบ (Item Characteristic Curve : ICC) ของข้อสอบแต่ละข้อ Lord กล่าวว่า พฤติกรรมการตอบข้อสอบแต่ละข้อของผู้สอบสามารถอธิบายได้ด้วยโมเดลทางคณิตศาสตร์ที่เราคุ้นเคยกันอยู่แล้ว คือ โค้งคุณลักษณะข้อสอบของแต่ละข้อมีลักษณะเป็นโค้งปกติสะสม หรือที่เรียกว่าโมเดลโค้งปกติสะสม (Normal Ogive Model) ซึ่งโมเดลนี้จะกล่าวถึงค่าพารามิเตอร์ 2 ตัว คือ ค่าความยากและค่าอำนาจจำแนก แต่เนื่องจากโมเดลนี้ต้องใช้การคำนวณที่ยุ่งยาก ซับซ้อน และขาดโปรแกรมคอมพิวเตอร์ที่จำเป็นต้องใช้ในการวิเคราะห์ข้อมูลตามทฤษฎีจึงทำให้ Lord ต้องใช้ระยะเวลาช่วงหนึ่งในการพิสูจน์ความเหมาะสมและความเป็นไปได้ของทฤษฎีดังกล่าว

ต่อมาในปี ค.ศ.1960 Rasch ได้ศึกษาและเสนอแนวคิดในรูปนารามิเตอร์เดียว คือใช้ค่าความยากของข้อสอบเพียงค่าเดียว เรียกกันว่า ราสชโมเดล (Rasch Model) จากการเสนอแนวคิดของ Rasch ทำให้ Lord กลับมาสนใจทฤษฎีนี้ใหม่ในปี ค.ศ.1965 (Warm, 1979)

และในปี ค.ศ.1968 Birnbaum ได้ศึกษาและเสนอโมเดลโลจิสติก (Logistic Model) ที่ใช้นารามิเตอร์ 2 ตัวคือค่าความยากและค่าอำนาจจำแนก ซึ่งเป็นโมเดลที่ง่ายกว่าของ Lord จึงเป็นที่นิยมใช้กันแพร่หลายและได้รับการพัฒนาขึ้นเรื่อย ๆ จนสามารถใช้ได้กับนารามิเตอร์ตัวเดียวและนารามิเตอร์ 3 ตัว (Warm, 1979)

2.2 หลักการของทฤษฎีคุณลักษณะแฝง

2.2.1 ลักษณะสำคัญของทฤษฎีคุณลักษณะแฝง

สราเริง (2525) กล่าวถึงใจความสำคัญของทฤษฎีคุณลักษณะแฝงไว้ว่า ผลการสอบของผู้สอบจากแบบทดสอบใด ๆ ขึ้นอยู่กับความสามารถของผู้สอบ หรือคะแนนของผู้สอบจากแบบทดสอบใด ๆ นั้น สามารถพยากรณ์หรืออธิบายได้จากความสามารถของคน ๆ นั้น ดังนั้นทฤษฎีนี้จึงเป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างปริมาณของการกระทำ หรือคะแนนของผู้สอบกับปริมาณความสามารถของผู้สอบ ซึ่งแสดงได้ดังสมการ (1)

$$P = f(\theta) \quad \dots\dots\dots(1)$$

เมื่อ P แทน ผลการสอบ (Performance)

θ แทน ความสามารถ (Ability หรือ Trait)

f แทน ความสัมพันธ์ (Function)

สงบ (2525) กล่าวถึงใจความสำคัญของทฤษฎีคุณลักษณะแฝงไว้ว่า เป็นการมุ่งหาความสัมพันธ์ ระหว่างความสามารถที่แท้จริงกับการตอบข้อสอบ ซึ่งในที่นี้ความสามารถที่แท้จริงนั้นคือคุณลักษณะที่วัดได้ด้วยแบบทดสอบ เช่น ความรู้ ความเข้าใจ หรือทักษะในด้านต่าง ๆ ส่วนการตอบข้อสอบ (Response) หมายถึง การตอบข้อสอบแบบปรนัยที่ให้คะแนนแบบตอบถูกให้ 1 คะแนน และตอบผิดให้ 0 คะแนน

จากความสัมพันธ์ในสมการ (1) เป็นการแสดงความสัมพันธ์ทั่ว ๆ ไปยังไม่เฉพาะเจาะจง จึงมีผู้สนใจศึกษาและพัฒนา เพื่อให้ได้สมการที่สามารถอธิบายลักษณะเฉพาะของข้อสอบแต่ละข้อที่สา-

มารถอธิบายความสัมพันธ์ของสองตัวแปรนี้ได้เด่นชัดขึ้น โดยอาศัยข้อตกลงบางอย่าง และวิธีการทางคณิตศาสตร์

นอกจากนี้ สงบ (2525) ยังกล่าวว่า ถ้าข้อตกลงเบื้องต้นของการสอบตรงตามที่ทฤษฎีกำหนด และเราสามารถคำนวณค่าการมีเตอร์ได้เหมาะสมกับโมเดลแล้วผลที่ได้จะได้รับคือ

1. ค่าการมีเตอร์ของข้อสอบ คือ ค่าความยาก ค่าอำนาจจำแนก และค่าสัมประสิทธิ์การแตก จะเป็นค่าที่ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ไม่ว่าจะนำไปสอบกับผู้ใดก็ตาม

2. เมื่อทราบลักษณะการตอบข้อสอบแต่ละข้อของผู้สอบคนใด เราจะสามารถคำนวณหาความสามารถที่แท้จริงของบุคคลนั้นได้ ค่าความสามารถที่แท้จริงนี้มีความสัมพันธ์โดยตรงกับคะแนนจริง การคำนวณความสามารถที่แท้จริงนั้นอาจใช้ข้อสอบข้อใดก็ได้ที่วัดสิ่งเดียวกัน ลักษณะเช่นนี้ถือว่าเป็นลักษณะของความเป็นอิสระของข้อสอบ

2.2.2 ข้อตกลงเบื้องต้นของทฤษฎีคุณลักษณะแฝง

ทฤษฎีคุณลักษณะแฝงมีข้อตกลงเบื้องต้นที่สำคัญอยู่ 4 ประการ ดังนี้ (Warm, 1978)

1. การรู้ค่าตอบที่ถูกต้อง (The Know-Correct Assumption) หมายความว่า ในการตอบข้อสอบข้อใดข้อหนึ่งนั้น ถ้าผู้สอบรู้ค่าตอบที่ถูกต้องแล้วจะตอบข้อสอบข้อนั้นถูก หรือถ้าผู้สอบตอบข้อสอบข้อใดข้อหนึ่งผิดแสดงว่าผู้สอบไม่ทราบค่าตอบที่ถูกต้องของข้อสอบข้อนั้น

2. แบบทดสอบที่มีมิติเดียว (Unidimensional Test) หมายความว่าข้อสอบแต่ละข้อในแบบทดสอบฉบับนั้นจะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน หรือมีความเป็นเอกพันธ์กันทั้งนี้ เพื่อให้ง่ายต่อการรวมคะแนนและการแปลความหมายของคะแนนผลการสอบ การตรวจสอบว่าแบบทดสอบฉบับใดฉบับหนึ่งใช้วัดความสามารถด้านเดียวหรือไม่นั้น มีวิธีดำเนินการดังรายละเอียดต่อไปนี้ (Warm, 1978)

ขั้นที่ 1 แปลงผลการตอบข้อสอบแต่ละข้อของผู้สอบแต่ละคนให้เป็นคะแนน 0 เมื่อตอบผิดและให้คะแนนเป็น 1 เมื่อตอบถูก

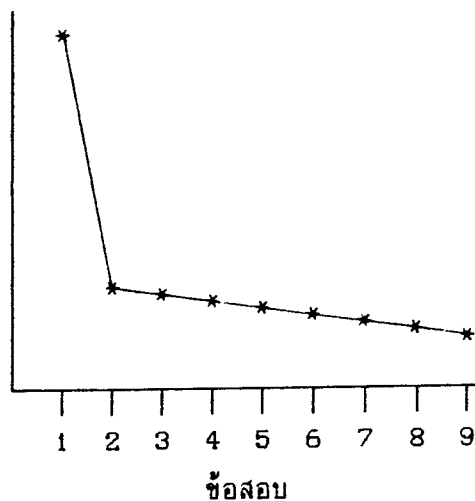
ขั้นที่ 2 คำนวณค่าสหสัมพันธ์ระหว่างข้อแบบ Tetachoric แล้วจัดค่าสหสัมพันธ์ที่ได้ให้อยู่ในรูป Matrix

ขั้นที่ 3 แทนที่ค่าสหสัมพันธ์ในแนวทแยง ด้วยค่าสหสัมพันธ์ที่มีค่าสูงสุดในแต่ละแถว

ขั้นที่ 4 วิเคราะห์ตัวประกอบจาก Correlation Matrix ที่เตรียมไว้โดยใช้วิธีแยกตัวประกอบแบบ Principal Component หรือ Principal Axis

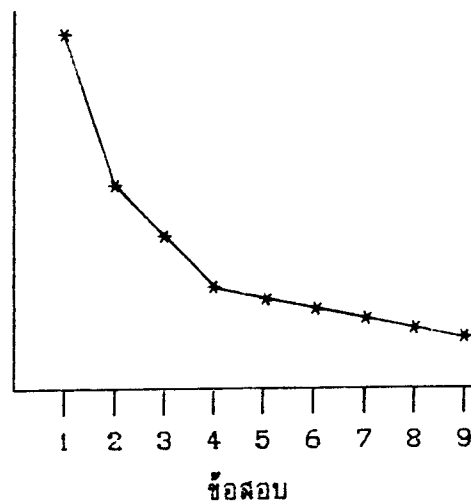
เมื่อวิเคราะห์ตัวประกอบเสร็จแล้ว นำผลการวิเคราะห์ที่ได้มาพิจารณาว่าแบบทดสอบฉบับนั้นใช้วัดความสามารถด้านเดียวกันหรือไม่ จากตัวบ่งชี้ดังต่อไปนี้

Eigenvalue



A. UNIDIMENSIONAL

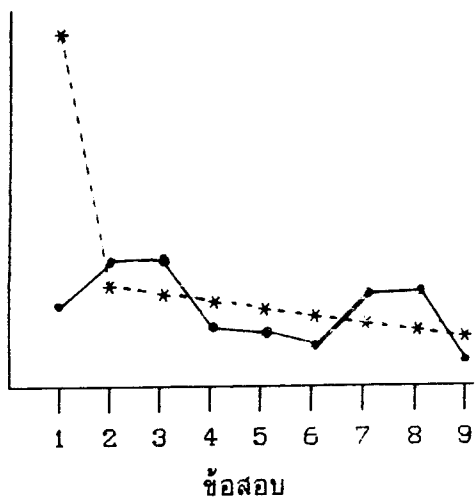
Eigenvalue



B. NONUNIDIMENSIONAL

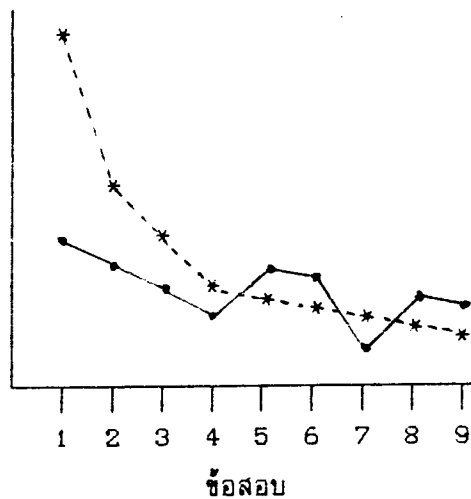
พิจารณาจากกราฟของ Eigenvalue

Eigenvalue



C. UNIDIMENSIONAL

Eigenvalue



D. NONUNIDIMENSIONAL

พิจารณาจาก Random Baseline

ภาพที่ 2 กราฟแสดงค่า Eigenvalue กับตัวประกอบ(ข้อสอบ) สำหรับการพิจารณาความเป็น Unidimensionality ของแบบทดสอบ (Warm, 1978)

(1) The Eigenvalue Test นำค่า Eigenvalue ของแต่ละตัวประกอบมาเขียนกราฟ ถ้าพบว่ากราฟมีลักษณะดังภาพที่ 2 A ก็แสดงว่าแบบทดสอบนั้นใช้วัดความสามารถด้านเดียวกัน

(2) The Random Baseline Test วิธีการนี้จะสุ่มตัวอย่างข้อมูลมาเรียงบางส่วน แล้วดำเนินการวิเคราะห์ตัวประกอบตามวิธีการดังกล่าวข้างต้น แล้วนำค่า Eigenvalue ของแต่ละตัวประกอบที่ได้ชุดใหม่นี้มาเขียนกราฟบนแกนเดียวกันกับข้อ (1) ถ้ากราฟที่ได้มีลักษณะดังภาพ 2 C ก็แสดงว่าแบบทดสอบฉบับนั้นใช้วัดความสามารถด้านเดียว

(3) The Baseline Test วิธีการนี้จะคำนวณค่าสหสัมพันธ์ระหว่างข้อสอบแต่ละข้อ กับคะแนนรวมแบบ Biserial แล้วนำค่าสหสัมพันธ์ที่ได้ไปหาค่าสหสัมพันธ์กับ Factor Loading ของข้อสอบแต่ละข้อที่มีต่อ Factor แรก ถ้าค่าสหสัมพันธ์ที่ได้นี้มีค่าสูงตั้งแต่ 0.80 ขึ้นไปก็แสดงว่าแบบทดสอบนั้นใช้วัดความสามารถด้านเดียว

(4) The Factor Loading Test วิธีการนี้จะพิจารณาจากค่า Factor Loading ของข้อสอบแต่ละข้อที่มีต่อ Factor แรก ถ้าพบว่า Factor Loading แต่ละค่ามีค่าสูงอย่างมีนัยสำคัญและมีเครื่องหมายเหมือนกัน (+ หรือ -) แสดงว่าแบบทดสอบนั้นใช้วัดความสามารถด้านเดียว

(5) The Congruence Test วิธีการนี้จะแบ่งผู้สอบออกเป็น 2 กลุ่มตามตัวแปรใดตัวแปรหนึ่ง เช่น กลุ่มเพศชาย/หญิง แล้วดำเนินการวิเคราะห์ตัวประกอบในแต่ละกลุ่ม แล้วคำนวณหาค่าสัมประสิทธิ์ความสอดคล้อง (C_{AB}) ของ Factor Loading บน Factor แรก ระหว่างสองกลุ่มโดยใช้สูตรดังนี้

$$C_{AB} = \sqrt{\frac{\sum_{i=1}^n (L_{ia} - L_{ib})^2}{n}} \quad \dots\dots\dots(2)$$

เมื่อ L_{ia} = ค่า Factor Loading ของข้อสอบข้อที่ i บน Factor แรกในกลุ่ม A

L_{ib} = ค่า Factor Loading ของข้อสอบข้อที่ i บน Factor แรกในกลุ่ม B

n = จำนวนข้อสอบในแบบทดสอบ

C_{AB} = สัมประสิทธิ์ความสอดคล้องระหว่างกลุ่ม A และ B

ถ้าพบว่าค่า C_{AB} เข้าใกล้ 0 แสดงว่าแบบทดสอบนั้นใช้วัดความสามารถด้านเดียว

(6) The Commuality Test วิธีการนี้จะคำนวณค่า G จากสูตร

$$G = \sum \frac{r_{ij}}{\sqrt{\frac{h_i h_j}{n}}} \quad , i > j \dots\dots\dots (3)$$

เมื่อ r_{ij} = ค่าสหสัมพันธ์แบบ interitem tetrachoric

h_i = ค่าความแปรปรวนร่วมของข้อสอบข้อที่ i

h_j = ค่าความแปรปรวนร่วมของข้อสอบข้อที่ j

n = จำนวนของค่าสหสัมพันธ์

ถ้าพบว่าค่า G เข้าใกล้ 1.00 แสดงว่าแบบทดสอบนี้ใช้วัดความสามารถด้านเดียว

(7) The Part/Whole Test วิธีการนี้เริ่มจากนำแบบทดสอบทั้งฉบับ (whole) ซึ่งสมมุติว่ามี 20 ข้อมาวิเคราะห์หาค่านารามิเตอร์ของข้อสอบแต่ละข้อ แล้วแยกบางส่วน (Part) มาสมมุติว่านำมา 10 ข้อเพื่อวิเคราะห์หาค่านารามิเตอร์ของข้อสอบแต่ละข้อเช่นเดียวกัน หลังจากนั้นนำค่านารามิเตอร์ของข้อสอบ 10 ข้อที่ได้ 2 ชุดนั้นมาหาค่าสหสัมพันธ์กัน ถ้าพบว่ามีความสัมพันธ์กันสูงแสดงว่าแบบทดสอบ 10 ข้อดังกล่าวใช้วัดความสามารถด้านเดียวกัน แล้วดำเนินการเช่นเดิมกับข้อสอบข้ออื่น ๆ ที่ยังไม่ได้แยกส่วนมาวิเคราะห์เพื่อวิเคราะห์ให้ครบทั้งฉบับต่อไป

(8) The Vector Frequency Test วิธีการนี้ใช้การทดสอบความแตกต่างระหว่างแบบแผนการตอบ (Answer Patterns) จริงของผู้สอบกับแบบแผนการตอบที่คาดหวังว่าแตกต่างกันหรือไม่ ซึ่งวิธีการนี้มีข้อตกลงเบื้องต้นว่าความสามารถของกลุ่มผู้สอบ (θ) จะต้องมีการกระจายแบบปกติ และต้องทราบค่านารามิเตอร์ของข้อสอบแต่ละข้อแล้ว เมื่อนำแบบแผนการตอบจริงของผู้สอบกับแบบแผนการตอบที่คาดหวังมาทดสอบความแตกต่างโดยใช้ Chi-Square แล้วพบว่าไม่แตกต่างกันอย่างมีนัยสำคัญ ก็แสดงว่าแบบทดสอบนี้ใช้วัดความสามารถด้านเดียว

3. ข้อสอบแต่ละข้อจะต้องเป็นอิสระจากกัน (Local Independence) หมายความว่า โอกาสที่ผู้สอบตอบข้อสอบข้อใดข้อหนึ่งถูกนั้นไม่มีผลต่อการตอบข้อสอบข้ออื่น ๆ ในแบบทดสอบนั้น

ถ้าให้ U_g แทนผลการตอบข้อสอบข้อที่ g ของผู้สอบคนหนึ่ง

$$g = 1, 2, 3, \dots, n$$

และให้คะแนนแบบตอบถูกให้ 1 คะแนนตอบผิดให้ 0 คะแนน

P_g แทน โอกาสที่ผู้สอบตอบข้อสอบข้อ g ได้ถูกต้อง

$$Q_g = 1 - P_g$$

จากข้อตกลงเกี่ยวกับ Local Independence จะได้ว่า

$$\text{Prob} \{U_1=U_1, U_2=U_2, \dots, U_n=U_n\} = \prod_{g=1}^n P_g^{u_g} Q_g^{1-u_g} \dots\dots\dots(4)$$

จากสมการที่(4)อ่านว่า โอกาสของการตอบแบบทดสอบได้ถูกต้องทั้งหมดทุกข้อคือผลคูณ (Π) ของโอกาสของการตอบข้อสอบแต่ละข้อได้ถูกต้อง

จากข้อตกลงดังกล่าวนี้ ทำให้สามารถสร้างสมการแสดงความถี่ของคะแนนของกลุ่มผู้สอบที่มีความสามารถ θ ได้ดังสมการ (5)

$$f(x|\theta) = \sum \prod_{g=1}^n P_g^{u_g} Q_g^{1-u_g} \dots\dots\dots(5)$$

เมื่อ x เป็นคะแนนของผู้สอบที่มีค่าตั้งแต่ 0 ถึง n

การทดสอบว่าข้อสอบแต่ละข้อเป็นไปตามข้อตกลงของ Local Independence หรือไม่นั้น ให้ตรวจสอบดูแบบแผนการตอบข้อสอบของผู้สอบที่มีความสามารถใกล้เคียงกันว่าข้อสอบแต่ละข้อมีลักษณะการตอบแตกต่างกันหรือไม่ โดยใช้ Chi-Square ทดสอบความเป็นอิสระของการตอบข้อสอบเหล่านั้น และในการทดสอบต้องให้เวลาผู้สอบได้ทำข้อสอบอย่างเต็มที่เพื่อให้ผู้สอบแสดงความสามารถที่แท้จริงออกมา โดยที่โอกาสของการตอบข้อสอบข้อหนึ่ง ๆ ถูกของผู้สอบคนหนึ่งนั้น ต้องไม่ขึ้นอยู่กับความสามารถของบุคคลอื่น

4. โค้งคุณลักษณะข้อสอบ (Item Characteristic Curve : ICC) หมายความว่า โอกาสที่ผู้สอบจะตอบข้อสอบข้อหนึ่ง ๆ ถูกนั้นขึ้นอยู่กับโค้งคุณลักษณะข้อสอบของแต่ละโมเดลที่ใช้ และไม่ขึ้นกับการแจกแจงความสามารถของกลุ่มประชากร นั่นคือรูปร่างเส้นโค้งคุณลักษณะของข้อสอบไม่

ได้ขึ้นอยู่กับ การกระจายความสามารถของกลุ่มผู้สอบ

ถ้าเป็นไปตามข้อตกลงดังกล่าวข้างต้น การวิเคราะห์ข้อสอบโดยใช้ทฤษฎีคุณลักษณะแฝงจะให้ ข้อดีหรือข้อได้เปรียบดังนี้

(1) การประมาณค่าความสามารถของผู้สอบแต่ละคน จะขึ้นอยู่กับ การตอบข้อสอบที่ถูกเลือกมา จากกลุ่มข้อสอบขนาดใหญ่ที่วัดความสามารถด้านเดียวกัน

(2) ค่าพารามิเตอร์ของข้อสอบจะไม่ผันแปรไปตามกลุ่มตัวอย่างผู้สอบ

(3) ผลการทดสอบจะให้การประมาณค่าความสามารถที่ถูกต้องแน่นอน ในแต่ละระดับความสามารถ และสามารถประมาณความคลาดเคลื่อนมาตรฐานของการวัดของแต่ละบุคคลได้ด้วย

2.2.3 ค่าพารามิเตอร์ในทฤษฎีคุณลักษณะแฝง

ค่าพารามิเตอร์ในทฤษฎีคุณลักษณะแฝง แบ่งค่าพารามิเตอร์ออกเป็น 2 ลักษณะคือ ค่าพารามิเตอร์ของข้อสอบ (Item Parameters) ได้แก่ ค่าความยาก (b_u) ค่าอำนาจจำแนก (a_u) ค่าสัมประสิทธิ์การเดา (c_u) และค่าพารามิเตอร์ของผู้สอบ (Examinee Parameter) ได้แก่ ระดับความสามารถของผู้สอบ (θ) นิสัยของค่าพารามิเตอร์ต่าง ๆ เหล่านี้ มีดังนี้ (Hambleton and Cook, 1977)

ค่าความยาก (b_u) คือค่าที่แสดงระดับความสามารถที่จุดเปลี่ยนโค้งของโค้งคุณลักษณะข้อสอบ ในกรณีที่ไม่มี การเดา ค่าความยากก็คือค่า θ ณ จุดที่ $P_u = 0.50$ ดังนั้นค่าความยากจึงมีมาตรชนิดเดียวกับ θ โดยทางทฤษฎีแล้ว b_u มีค่าตั้งแต่ $-\alpha$ ถึง $+\alpha$ ถ้ามีค่าลบมาก แสดงว่าข้อสอบข้อนั้น ง่ายมาก และถ้ามีค่าบวกมาก แสดงว่าข้อสอบข้อนั้นยากมาก

ค่าอำนาจจำแนก (a_u) คือค่าที่เป็นสัดส่วนโดยตรงกับค่าความชัน (Slope) ของโค้งลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง หรือจุดบนโค้งที่ $\theta = b_u$ ในทางทฤษฎีค่า a_u มีค่าตั้งแต่ $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัติมันจะมีค่ามากกว่า 0 เพราะค่า a_u ที่มีค่าเป็นลบแสดงว่าข้อสอบไม่ดีใช้ไม่ได้จึงคัด ข้อสอบข้อนั้นทิ้งไป ถ้ามีค่า 0 แสดงว่าข้อสอบข้อนั้นไม่มีอำนาจจำแนกหรือมีอำนาจจำแนกต่ำ ถ้า มีค่าเป็นบวกมาก แสดงว่าข้อสอบมีค่าอำนาจจำแนกสูง

ค่าสัมประสิทธิ์การเดา (c_u) คือค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำมาก มีโอกาส ทำข้อสอบข้อ u ถูก

ค่าความสามารถของผู้สอบ (θ) คือระดับความสามารถของผู้สอบที่คำนวณจากคะแนนจริง (True Score) ในทางทฤษฎีมีค่าตั้งแต่ $-\alpha$ ถึง $+\alpha$ ถ้ามีค่าเป็นลบ หมายถึง ผู้สอบมีความสามารถต่ำ ถ้ามีค่าเป็นบวก หมายถึง ผู้สอบมีความสามารถสูง

2.2.4 โมเดลต่าง ๆ ในทฤษฎีคุณลักษณะแฝง

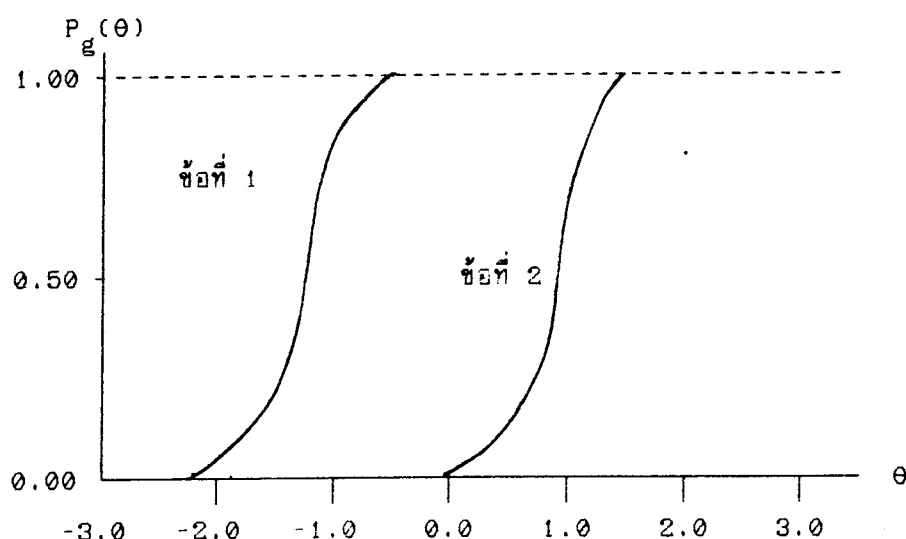
ทฤษฎีคุณลักษณะแฝงมีการพัฒนาในรูปแบบต่าง ๆ ทำให้เกิดเป็นโมเดลเฉพาะขึ้นหลายโมเดล โดยแต่ละ โมเดลจะแตกต่างกันที่ฟังก์ชันทางคณิตศาสตร์และจำนวนพารามิเตอร์ที่ใช้อธิบายโค้งลักษณะข้อสอบ เช่น โมเดลโค้งปกติสะสม (Normal Ogive Model) โมเดลเส้นตรง (The Linear Model) และโมเดลโลจิสติก (The Logistic Model) ในที่นี้จะกล่าวถึงโมเดลโลจิสติกที่มีอยู่ 3 โมเดลย่อย คือ (Hambleton and Cook, 1977)

1. โมเดลที่ใช้พารามิเตอร์ตัวเดียว (One-Parameter Logistic Model)

Rasch ได้พัฒนาทฤษฎีคุณลักษณะแฝง และเสนอโมเดลนี้ในปี ค.ศ. 1966 โดยที่ฟังก์ชันของโมเดลนี้สามารถอธิบายได้ด้วยพารามิเตอร์ของข้อสอบเพียงตัวเดียว คือค่าความยาก นิยมเรียกกันว่าราสชโมเดล(Rasch Model)ซึ่งตรงกับโมเดลที่ใช้พารามิเตอร์ตัวเดียวที่ Birnbaum พัฒนาขึ้น สามารถเขียนเป็นสมการดังแสดงในสมการที่ (6)

$$P_g(\theta) = \frac{e^{D a_g(\theta - b_g)}}{1 + e^{D a_g(\theta - b_g)}} \quad ; \quad g = 1, 2, \dots, n \quad \dots (6)$$

โมเดลนี้มีข้อตกลงว่าข้อสอบทุกข้อไม่มีโอกาสของการเดาถูก ($c_g = 0$) และค่าอำนาจจำแนกของข้อสอบ (a_g) ทุกข้อเท่ากันหมด โค้งคุณลักษณะข้อสอบของโมเดลนี้แสดงไว้ในภาพที่ 3



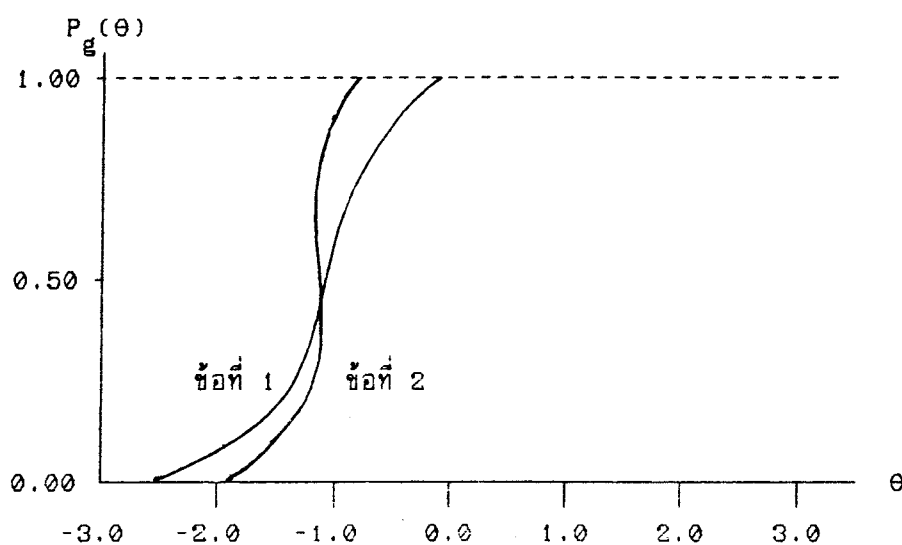
ภาพที่ 3 แสดงโค้งลักษณะข้อสอบโมเดลโลจิสติกที่มีพารามิเตอร์ตัวเดียว (b_g)

2. โมเดลที่ใช้พารามิเตอร์สองตัว (Two-Parameter Logistic Model)

ในปี ค.ศ. 1968 Birnbaum ได้เสนอโมเดลโลจิสติกที่มีพารามิเตอร์ 2 ตัว คือ ค่าความยากและค่าอำนาจจำแนก ซึ่งมีสมการดังแสดงในสมการที่ (7)

$$P_g(\theta) = \frac{e^{Da_g(\theta - b_g)}}{1 + e^{Da_g(\theta - b_g)}} \quad ; \quad g = 1, 2, \dots, n \quad \dots (7)$$

โมเดลนี้มีข้อตกลงว่า ข้อสอบทุกข้อไม่มีการเดาถูก คือค่า $c_g = 0$ โค้งคุณลักษณะข้อสอบของโมเดลนี้แสดงไว้ในภาพที่ 4



ภาพที่ 4 แสดงโค้งลักษณะข้อสอบโมเดลโลจิสติกที่มีพารามิเตอร์สองตัว (a_g และ b_g)

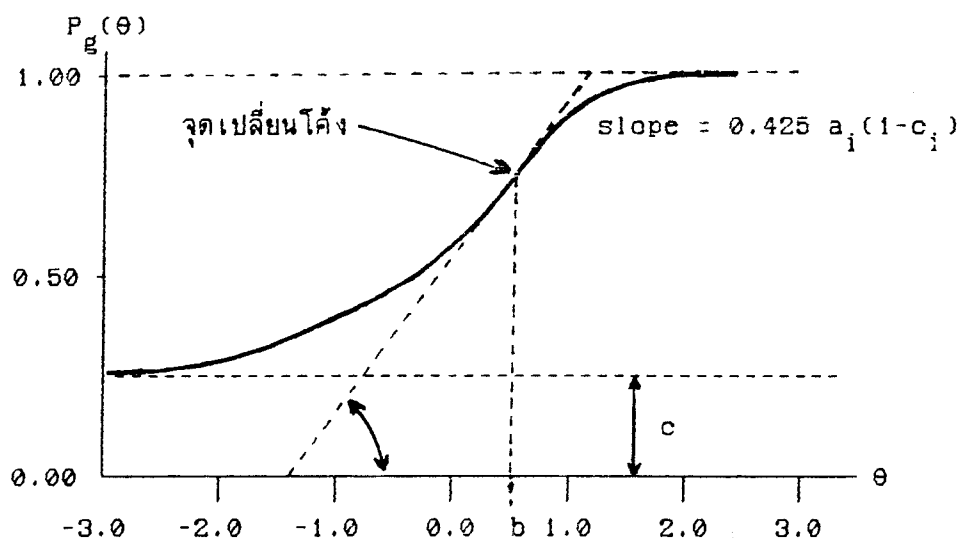
3. โมเดลที่ใช้พารามิเตอร์สามตัว (Three-Parameter Model)

โมเดลนี้ดัดแปลงมาจาก โมเดลที่ใช้พารามิเตอร์สองตัว พารามิเตอร์ที่เพิ่มขึ้น คือค่าสัมประสิทธิ์การเดา (c_g) มีสมการดังแสดงในสมการที่ (8) ดังนี้

$$P_g(\theta) = C_g + (1-C_g) \frac{e^{\frac{Da_g(\theta-b_g)}{1+e^{Da_g(\theta-b_g)}}}}{1+e^{Da_g(\theta-b_g)}} \quad ; \quad g = 1, 2, \dots, n \quad \dots (8)$$

- เมื่อ θ คือ ระดับความสามารถที่แท้จริงที่คำนวณจากคะแนนจริงแต่ปรับหน่วยให้เป็นมาตรฐาน
- $P_g(\theta)$ คือ ค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะทำข้อสอบข้อ g ได้ถูกต้อง
- a_g คือ ค่าอำนาจจำแนกซึ่งมีค่าเป็นสัดส่วนโดยตรงกับค่าความชันของโค้งลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง หรือจุดที่ชันที่สุด
- b_g คือ ค่าความยากที่แสดงระดับความสามารถที่แท้จริง ที่จุดโค้งชันที่สุด หรือในกรณีไม่มีการเดาค่า b_g คือ θ ณ จุดความน่าจะเป็น .50
- D คือ ค่าคงที่ของ Scaling factor มีค่าเท่ากับ 1.7
- e คือ ค่าคงที่มีค่าประมาณ 2.71828

โค้งคุณลักษณะข้อสอบของโมเดลนี้แสดงไว้ในภาพที่ 5



ภาพที่ 5 แสดงโค้งลักษณะข้อสอบโมเดลโลจิสติกที่มีพารามิเตอร์สามตัว (a_g , b_g และ c_g)

2.2.5 คุณภาพของแบบทดสอบ

การวิเคราะห์คุณภาพของแบบทดสอบ หรือการนิยามความแน่นอนของการประมาณค่าความสามารถที่แท้จริง ใช้วิธีการนิยามจากฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) ซึ่งได้จากผลรวมของฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อ (Item Information Function) (Lord, 1980)

ฟังก์ชันสารสนเทศของข้อสอบ (Item Information Function) แต่ละข้อกำหนดได้ดังแสดงในสมการที่ (9)

$$I(\theta, U_g) = (P'_g)^2 / P_g Q_g \quad ; g = 1, 2, \dots, n \quad \dots (9)$$

เมื่อ $I(\theta, U_g)$ คือ ฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อ
 P_g คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบข้อ g ถูก
 Q_g คือ $(1 - P_g)$
 P'_g คือ ความชันของโค้งลักษณะข้อสอบที่ระดับความสามารถ θ

จากสมการที่ (9) จะเห็นได้ว่า ฟังก์ชันสารสนเทศของข้อสอบแต่ละข้อขึ้นอยู่กับความชันของโค้งลักษณะข้อสอบ ถ้าโค้งลักษณะข้อสอบชันมากขึ้นในขณะที่ความแปรปรวนของการตอบข้อสอบถูกย่นย่อลง โค้งสารสนเทศของข้อสอบที่ความสามารถนั้น ๆ จะยิ่งสูงขึ้น ซึ่งความสูงของโค้งสารสนเทศของข้อสอบอยู่ที่ระดับความสามารถใดแสดงว่าสามารถจำแนกจาแนกจาแนกระดับความสามารถของผู้สอบได้ดี ณ ระดับความสามารถนั้น (ผจงจิต, 2525)

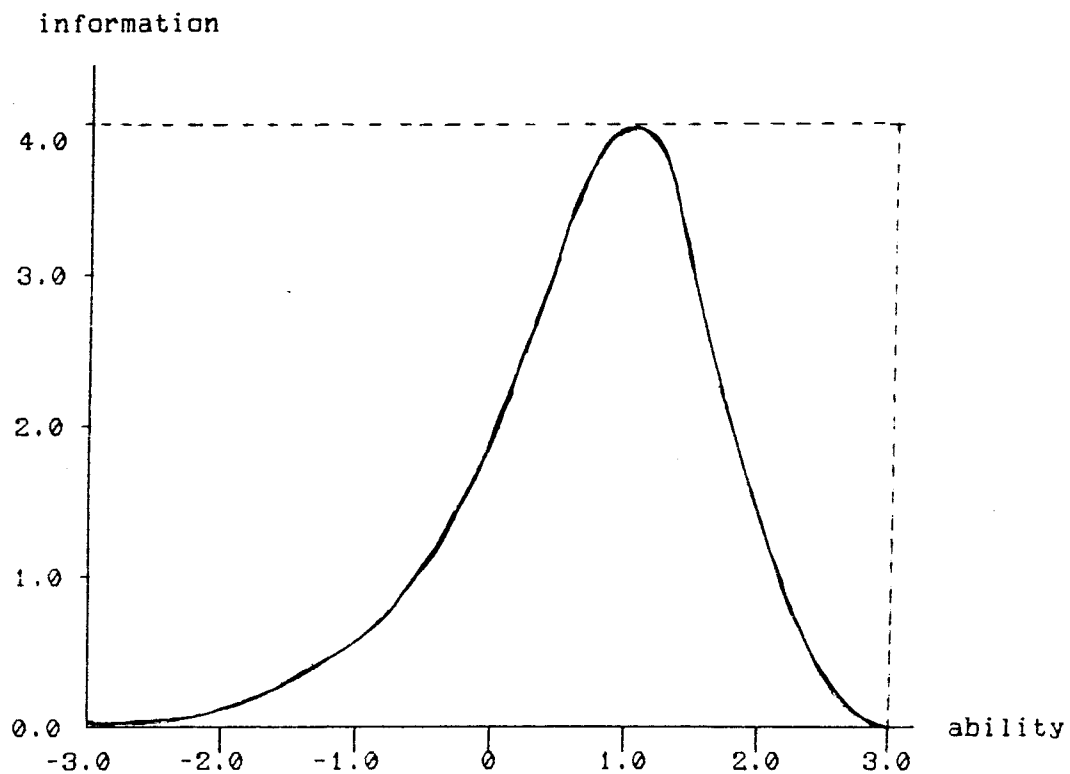
ฟังก์ชันสารสนเทศของแบบทดสอบ (Test Information Function) เป็นสัดส่วนกลับกับกำลังสองของความยาวของช่วงความเชื่อมั่น ที่เป็นผลมาจากการประมาณค่าความสามารถของผู้สอบจากการทำข้อสอบของเขา ดังนั้นโค้งสารสนเทศของแบบทดสอบจึงเป็นเครื่องแสดงถึงความถูกต้องแน่นอนของค่าความสามารถที่ประมาณได้ ซึ่งแสดงในรูปสมการได้ดังในสมการที่ (10)

$$I(\theta) = \sum_{g=1}^n I(\theta, U_g) = \sum_{g=1}^n \frac{(P'_g)^2}{P_g Q_g} \quad \dots (10)$$

เมื่อ $I(\theta)$ คือ นิ่งก์ชั้นสารสนเทศของแบบทดสอบ
 $I(\theta, \psi)$ คือ นิ่งก์ชั้นสารสนเทศของข้อสอบ

ถ้าเรามีกลุ่มข้อสอบที่ทราบโค้งสารสนเทศ เราสามารถสร้างแบบทดสอบให้มีโค้งสารสนเทศของแบบทดสอบ ณ ระดับความสามารถระดับใดระดับหนึ่งที่เราต้องการได้ เช่น การสร้างแบบทดสอบเพื่อคัดเลือกนักเรียนเข้าเรียนทุนเรียนดี ก็ต้องสร้างแบบทดสอบที่มีประสิทธิภาพสูงสุดที่ระดับความสามารถสูง ๆ นั่นคือ ให้มีโค้งสารสนเทศของแบบทดสอบสูง ณ ระดับความสามารถสูง ๆ นั่นเอง

ตัวอย่างโค้งสารสนเทศของแบบทดสอบที่เหมาะสมสำหรับวัดความสามารถที่ระดับความสามารถ 1.0 จะมีลักษณะดังแสดงไว้ในภาพที่ 6



ภาพที่ 6 โค้งสารสนเทศของแบบทดสอบสำหรับวัดความสามารถที่ระดับ 1.0 (Lord, 1980)

2.3 ชนิดของแบบทดสอบปิรามิด

แบบทดสอบปิรามิด (Pyramidal Test) สามารถจัดรูปแบบของแบบทดสอบได้โดยอาศัยโครงสร้างของแบบจำลองรูปปิรามิด (Pyramidal Models) หรือแบบจำลองโครงสร้างต้นไม้ (Tree Structure Models) ซึ่งเป็นรูปแบบหนึ่งของแบบทดสอบเทเลอร์ (Tailored Test) ประเภทวิธีการหลายขั้นที่มีแบบจำลองการแยกทางคงที่ แบบทดสอบปิรามิดเป็นแบบจำลองแบบแรกของการทดสอบเทเลอร์ที่ได้มีการศึกษาและก่อให้เกิดการวิจัยอย่างกว้างขวางในปัจจุบัน งานวิจัยเกี่ยวกับแบบจำลองรูปปิรามิดเริ่มต้นจากการศึกษาของ Krathwohl และ Huyser ในปี ค.ศ. 1956 ในระยะต่อมาก็มีรายงานการศึกษาของ Bayroff และคณะ (1960, 1967) Paterson (1962) Hansen (1969) และ Lord (1970, 1971) ผลการศึกษาของงานวิจัยที่ผ่านมาเหล่านี้ส่วนใหญ่สอดคล้องกับผลงานวิจัยของ Mussio (1972) และ Larkin and Weiss (1974)

ในปัจจุบันแบบทดสอบปิรามิดได้มีการศึกษากันใน 5 รูปแบบ คือ รูปแบบที่ใช้ขนาดขั้นคงที่ รูปแบบที่ใช้ขนาดขั้นแปรผัน รูปปิรามิดข้างตัด รูปจำลองที่มีข้อสอบหลายข้อในแต่ละขั้น และรูปแบบการให้น้ำหนักตัวเลือกของข้อสอบเพื่อการแยกทาง Weiss (1974) ได้สรุปแบบทดสอบปิรามิดทั้ง 5 ชนิดดังกล่าวไว้ ดังนี้

1. แบบทดสอบปิรามิดที่ใช้ขนาดขั้นคงที่ (Constant Step Size Pyramids)

แบบทดสอบปิรามิดที่ใช้ขนาดขั้นคงที่ จะมีจำนวนข้อสอบในแต่ละขั้นเท่ากับลำดับที่ของขั้น คือ ในขั้นที่ 1 จะมีข้อสอบ 1 ข้อ ในขั้นที่ 5 ก็จะมีข้อสอบ 5 ข้อ หรือในขั้นที่ 8 ก็จะมีข้อสอบ 8 ข้อ ดังนั้นถ้าเป็นแบบทดสอบปิรามิดชนิด 10 ขั้น ก็จะมีข้อสอบทั้งหมด 55 ข้อ ดังภาพที่ 7 แยกแสดงถึงค่าความยากของข้อสอบที่สัมพันธ์กับข้อสอบในแบบทดสอบปิรามิด ค่าความยากของข้อสอบอยู่ในช่วงตั้งแต่ .05 ถึง .95 ข้อสอบที่อยู่ในแนวตั้งเดียวกันจะเป็นข้อสอบที่มีระดับความยากเท่ากัน และข้อสอบที่ผู้สอบจะตอบในขั้นที่อยู่ติดกันจะมีความแตกต่างระหว่างค่าความยากเท่ากัน คือ เท่ากับ .05 หรือมีขนาดขั้นคงที่นั่นเอง

จากภาพที่ 7 เส้นทาง การตอบข้อสอบตามโครงสร้างแบบทดสอบปิรามิดชนิดนี้ ผู้สอบทุกคนจะเริ่มทำข้อสอบหมายเลข 1 ในขั้นที่ 1 คำตอบของผู้สอบในขั้นนี้อาจจะผิดหรือถูกก็ตามจะถูกนำไปใช้ในการพิจารณาเพื่อการแยกทางในการตอบข้อสอบข้อต่อไปในขั้นที่ 2 ภายใต้กฎการแยกทางแบบเต็ม 1/ลด 1 จากกฎการแยกทางลักษณะนี้ ถ้าผู้สอบตอบข้อสอบถูกต้อง ผู้สอบจะได้ตอบข้อสอบข้อที่มีความยากสูงขึ้น หรือเป็นข้อสอบที่ยากมากกว่าข้อเดิมที่ได้ทำเป็นข้อต่อไป แต่ถ้าผู้

พ. ๕๕๒๘๔

สอบตอบข้อสอบผิด ก็จะถูกแยกทางให้ไปตอบข้อสอบข้อที่มีความยากลดลง หรือเป็นข้อสอบที่ง่ายกว่าข้อเดิมเป็นข้อต่อไป จากตัวอย่างผู้สอบ 3 คน เริ่มตอบข้อสอบหมายเลข 1 ในขั้นที่ 1 เหมือนกัน เส้นทึบแสดงเส้นทางการตอบของคนที่มีความสามารถปานกลาง เมื่อเขาตอบข้อสอบหมายเลข 1 ($p=.50$) ถูก จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 3 ($p=.45$) ในขั้นที่ 2 ที่ยากขึ้น ซึ่งเขาตอบถูกอีก ในขั้นที่ 3 เขาจึงได้ตอบข้อสอบหมายเลข 6 ($p=.40$) และก็ตอบถูกอีกจึงได้ทำข้อสอบหมายเลข 10 ($p=.35$) ในขั้นที่ 4 ซึ่งเขาตอบผิดเป็นข้อแรกดังนั้นในขั้นที่ 5 เขาจึงได้ทำข้อสอบหมายเลข 14 ($p=.40$) ซึ่งข้อที่มีความยากลดลงและเขาตอบข้อสอบในขั้นที่ 5 นี้ผิดอีกจึงถูกแยกทางให้ไปทำข้อสอบหมายเลข 19 ($p=.45$) ในขั้นที่ 6 ซึ่งเป็นข้อสอบที่ง่ายกว่าเดิมอีก ผลการตอบในขั้นที่ 6 นี้ปรากฏว่าเขาตอบถูกจึงได้ทำข้อสอบที่ยากขึ้นในขั้นที่ 7 คือได้ข้อสอบหมายเลข 26 ($p=.40$) ซึ่งเขาตอบถูกอีกจึงได้ทำข้อสอบหมายเลข 34 ($p=.35$) ในขั้นที่ 8 ซึ่งเป็นข้อที่ยากขึ้นและเขาตอบผิดจึงได้ทำข้อสอบหมายเลข 42 ($p=.40$) ในขั้นที่ 9 ซึ่งมีความยากลดลงและเขาตอบข้อสอบข้อนี้ถูกจึงได้ตอบข้อสอบข้อที่ยากขึ้นในขั้นที่ 10 คือได้ข้อสอบหมายเลข 52 ($p=.35$) ซึ่งเป็นข้อสุดท้าย รวมทำข้อสอบทั้งหมด 10 ข้อ

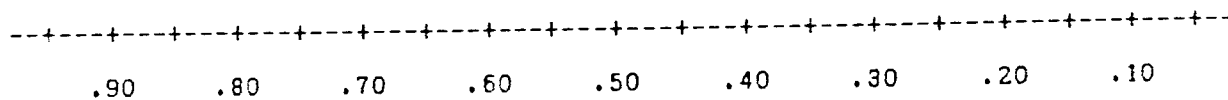
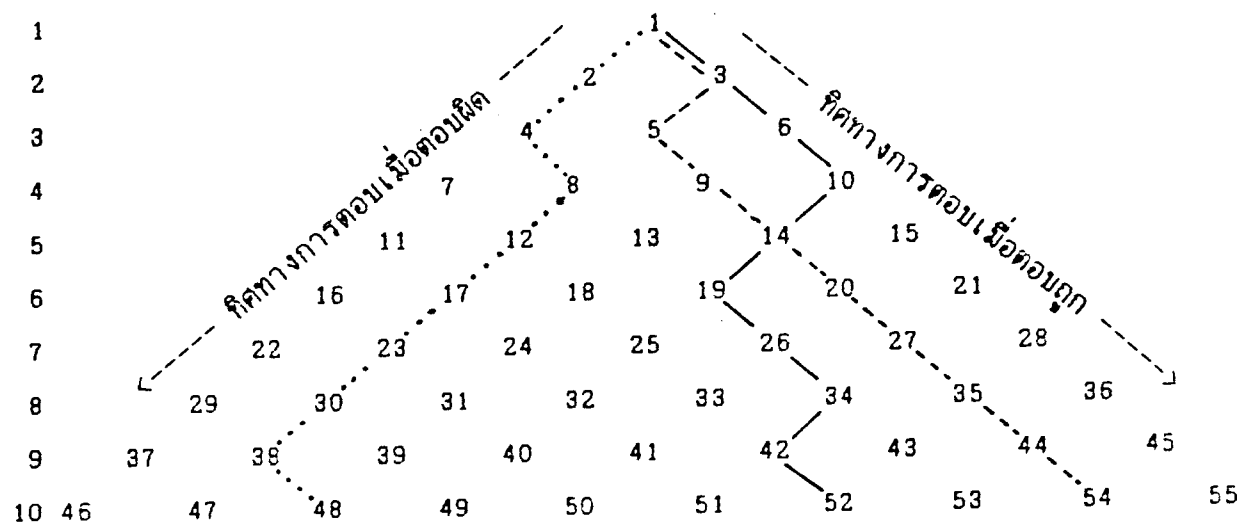
เส้นทางการตอบข้อสอบของผู้สอบคนที่ 2 ในภาพที่ 7 แสดงด้วยเส้นประเป็นคนที่มีความสามารถค่อนข้างสูง เขาตอบข้อสอบทั้งหมด 10 ข้อ ตอบขึ้นละ 1 ข้อเช่นเดียวกัน แต่เขาตอบผิดเพียง 1 ข้อเท่านั้นคือข้อสอบหมายเลข 3 ($p=.45$) เส้นประจึงแสดงเส้นทางการตอบข้อสอบที่มีความยากสูงขึ้นเรื่อย ๆ จนถึงข้อสอบหมายเลข 54 ($p=.15$) ในขั้นที่ 10 ซึ่งเป็นขั้นสุดท้าย และแนวเส้นจุดไข่ปลาเป็นเส้นทางการตอบของผู้สอบคนที่ 3 จะเห็นว่าเขาตอบข้อสอบ 10 ข้อถูกเพียง 2 ข้อเท่านั้น คือข้อสอบหมายเลข 4 ($p=.60$) และข้อสอบหมายเลข 38 ($p=.80$) และข้อสอบข้อสุดท้ายที่เขาได้ตอบคือข้อสอบหมายเลข 48 ($p=.75$) ในขั้นที่ 10

โดยสรุปแล้ว จากภาพที่ 7 จะเห็นได้ว่าค่าความยากของข้อสอบในขั้นที่อยู่ติดกันจะแตกต่างกันเท่ากับ .05 คงที่ในกฎการแยกทางแบบเพิ่ม 1/ลด 1 โดยมีจำนวนขั้นที่แยกทางไปทำข้อที่ยากขึ้นเท่ากับจำนวนขั้นที่แยกไปทำข้อที่ง่ายลง (Equal Offset) ผู้สอบทุกคนจะเริ่มทำข้อสอบข้อแรก คือข้อสอบหมายเลข 1 ในขั้นที่ 1 ถ้าตอบถูกก็จะได้ทำข้อสอบข้อที่ยากขึ้น แต่ถ้าตอบผิดก็จะได้ทำข้อสอบข้อที่ง่ายลง จะดำเนินการสอบไปจนกว่าจะครบทั้ง 10 ขั้น หรือ n ขั้นเมื่อแบบทดสอบมี n ขั้น และผู้สอบจะได้ทำข้อสอบ 1 ข้อในแต่ละขั้น

เนื่องจากข้อสอบชนิดเลือกตอบ (Multiple-Choice Items) ผู้สอบอาจใช้วิธีการเดาตอบข้อสอบได้ จึงมีการศึกษาจัดรูปแบบการแยกทางการตอบข้อสอบเพื่อขจัดผลที่เกิดจาก

ชั้นที่

หมายเลขข้อสอบ



ข้อสอบง่าย

ความยาก/ความสามารถ

ข้อสอบยาก

(ความสามารถต่ำ)

(สัดส่วนการตอบถูก)

(ความสามารถสูง)

---- ความสามารถสูง

— ความสามารถปานกลาง

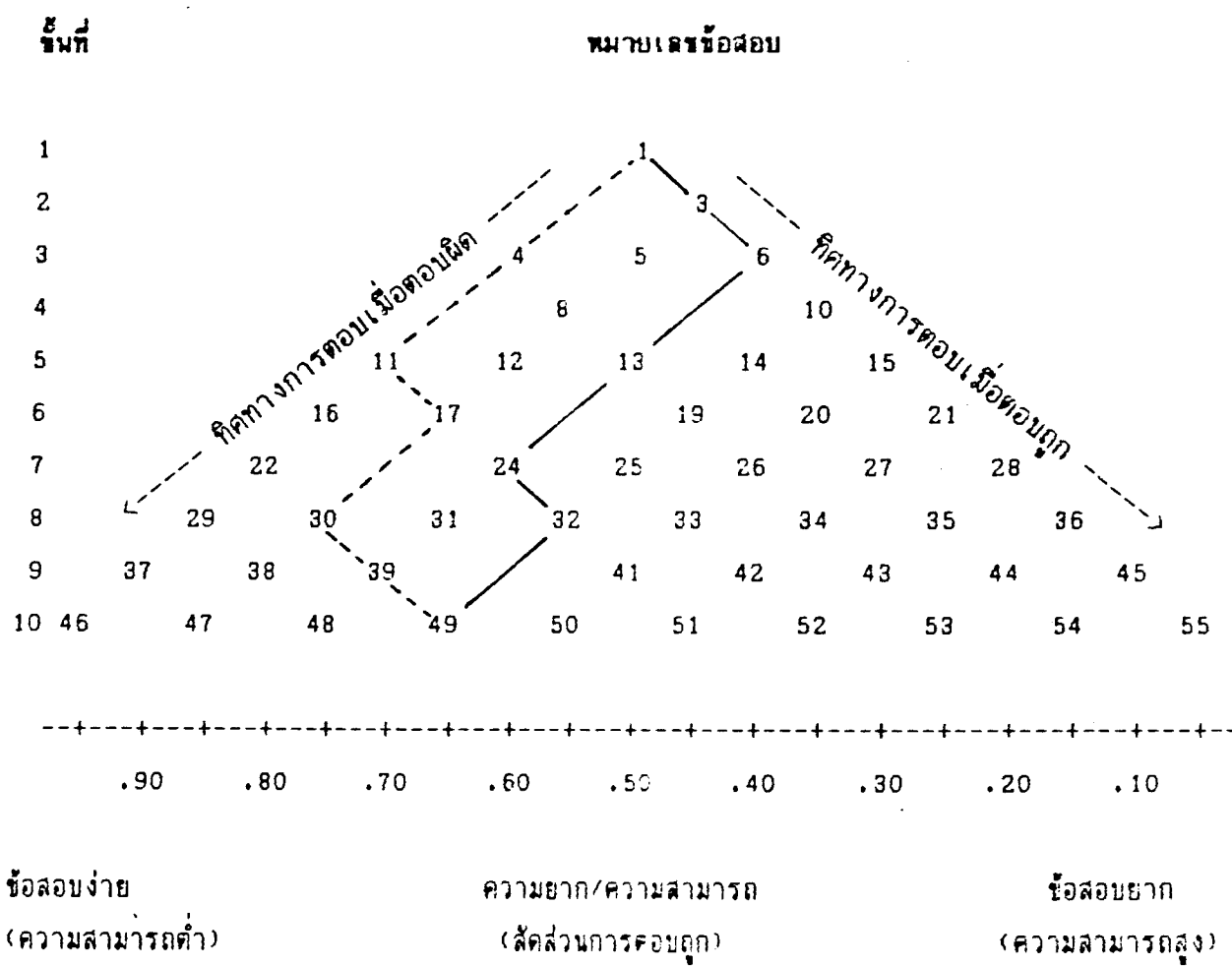
.... ความสามารถต่ำ

ภาพที่ 7 แสดงโครงสร้างของแบบทดสอบปิรามิดที่ใช้ขนาดชั้นคงที่ ภายใต้กฎการแยกทางแบบ
 เน้ม1/ลต1 (Weiss, 1974)

การแตกตอบ โดยใช้กฎการแยกทางแบบเพิ่ม 1/ลด 2 (Up-One/Down-Two Branching Rule) ซึ่งมีจำนวนขั้นที่แยกทางไม่เท่ากัน (Unequal Offset) ดังแสดงในภาพที่ 8 การแยกทางในการตอบข้อสอบตามรูปแบบนี้กำหนดไว้ดังนี้ ถ้าตอบข้อสอบถูกจะได้ทำข้อสอบข้อต่อไปที่มีความยากเพิ่มขึ้นจากเดิมเท่ากับ .05 แต่ถ้าตอบผิดจะได้ทำข้อสอบข้อต่อไปที่มีความยากลดลงจากเดิม .10 เสมอ ตัวอย่างเส้นทางการตอบในภาพที่ 8 เส้นทึบแสดงการตอบข้อสอบหมายเลข 1 ($p=.50$) และหมายเลข 3 ($p=.45$) ถูก จึงแยกทางไปทำข้อสอบหมายเลข 6 ($p=.40$) ซึ่งมีระดับความยากเพิ่มขึ้น .05 เขาตอบข้อสอบหมายเลข 6 ผิด จึงได้ทำข้อสอบหมายเลข 13 ($p=.50$) เป็นข้อต่อไป ซึ่งมีความยากลดลงจากเดิม .10 และเขาตอบข้อสอบหมายเลข 13 ผิดอีกจึงได้ทำข้อสอบหมายเลข 24 ($p=.60$) เขาตอบข้อสอบข้อนี้ถูกจึงได้ทำข้อสอบหมายเลข 32 ($p=.55$) เป็นข้อต่อไป และข้อสอบข้อสุดท้ายที่เขาได้ทำคือข้อสอบหมายเลข 49 ($p=.65$) ส่วนแนวเส้นประแสดงเส้นทางการตอบข้อสอบของผู้สอบที่ตอบข้อสอบหมายเลข 11 ($p=.70$) หมายเลข 30 ($p=.75$) และหมายเลข 39 ($p=.70$) ถูกต้อง ในแต่ละครั้งที่เขาตอบข้อสอบถูกเขาจะได้ตอบข้อสอบข้อที่มีความยากเพิ่มขึ้น .05 ในขั้นถัดไปเป็นข้อต่อไป แต่ถ้าเขาตอบข้อสอบผิดเขาจะได้ตอบข้อสอบข้อที่มีความยากลดลงจากเดิม .10 ในอีก 2 ขั้นถัดไปเป็นข้อต่อไป

ข้อแตกต่างของโครงสร้างแบบทดสอบปิรามิตทั้งสองลักษณะนี้อีกประการหนึ่งคือ ผู้สอบที่มีเส้นทางการตอบต่างกันจะได้ทำข้อสอบจำนวนไม่เท่ากัน นั่นคือถ้าเป็นโครงสร้างปิรามิตที่มีขนาดขั้นคงที่ ภายใต้กฎการแยกทางแบบเพิ่ม 1/ลด 1 ผู้สอบแต่ละคนจะได้ทำข้อสอบจำนวนเท่ากันคือเท่ากับจำนวนขั้นของโครงสร้างแบบทดสอบพอดี แต่ถ้าเป็นโครงสร้างปิรามิตที่มีขนาดขั้นคงที่ ภายใต้กฎการแยกทางแบบเพิ่ม 1/ลด 2 จำนวนข้อสอบที่ผู้สอบแต่ละคนจะต้องทำจะแตกต่างกันขึ้นอยู่กับจำนวนข้อที่เขาตอบถูก เช่นผู้สอบที่ตอบตามแนวเส้นทึบและเส้นประในภาพที่ 8 จะได้ตอบข้อสอบเพียง 7 ข้อในแบบทดสอบปิรามิตชนิด 10 ขั้น ถ้าผู้สอบคนใดตอบข้อสอบถูกทุกข้อเขาจะได้ทำข้อสอบทั้งหมด 10 ข้อ และถ้าผู้สอบคนใดตอบข้อสอบผิดหมดทุกข้อเขาจะได้ตอบข้อสอบเพียง 5 ข้อเท่านั้น

การศึกษาแบบทดสอบปิรามิตที่มีขนาดขั้นคงที่ นอกจากศึกษากฎการแยกทางแบบ เพิ่ม 1/ลด 1 และแบบเพิ่ม 1/ลด 2 แล้วยังมีการศึกษาภายใต้กฎการแยกทางแบบเพิ่ม 1/ลด 3 และการแยกทางแบบเพิ่ม 2/ลด 3 ด้วยซึ่งวิธีการดำเนินการก็คล้ายกันกับที่ใช้ใน 2 วิธีดังกล่าวนั่นเอง

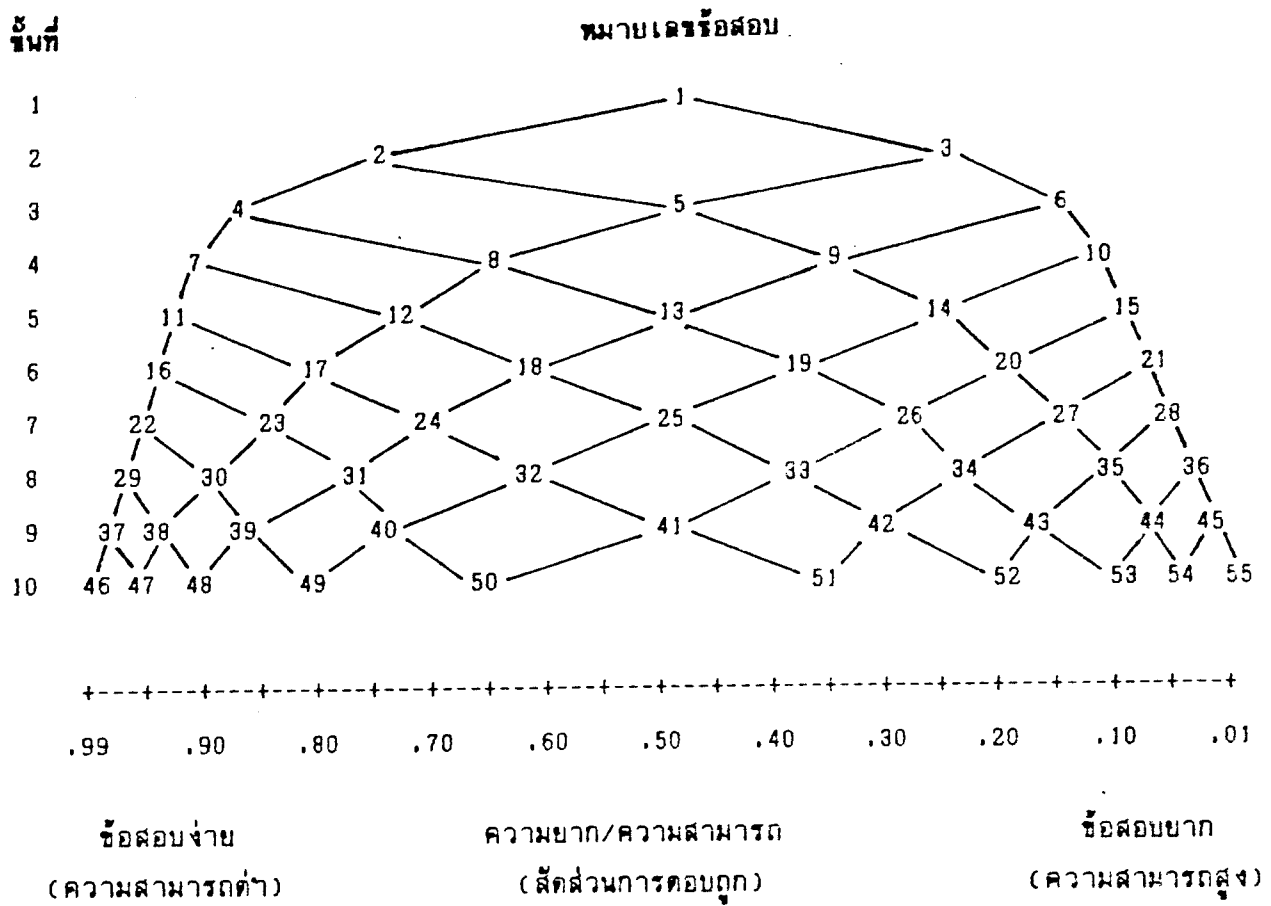


ภาพที่ 8 แสดงโครงสร้างของแบบทดสอบปียามักที่ใช้ขนาดชั้นคงที่ ภายใต้กฎการแยกทางแบบ
เน็ม1/ลด2 (Weiss, 1974)

2. แบบทดสอบปิรามิดที่ใช้ขนาดขั้นแปรผัน (Variable Step Size Pyramids)

Paterson และ Lord (อ้างถึงใน Weiss, 1974) ได้ให้ข้อเสนอแนะว่าแบบทดสอบปิรามิดที่ใช้ขนาดขั้นคงที่ ไม่มีประสิทธิภาพเกี่ยวกับความไวด้านการกำหนดเส้นทางในการตอบข้อสอบ จึงได้เสนอแนะว่าขั้นที่อยู่ติดกันในขั้นแรก ๆ ของโครงสร้างรูปปิรามิดนั้น ควรให้มีค่าความแตกต่างระหว่างค่าความยากของข้อสอบมาก ๆ และขั้นที่อยู่ติดกันในขั้นหลัง ๆ ของโครงสร้างรูปปิรามิดให้มีค่าความแตกต่างของค่าความยากของข้อสอบน้อยลง วิธีดำเนินการในลักษณะเช่นนี้จะช่วยให้สามารถวัดระดับความสามารถของผู้สอบแต่ละคนได้เที่ยงตรงมากขึ้น

ในปี ค.ศ.1962 Paterson (อ้างถึงใน Weiss, 1974) ได้กำหนดโครงสร้างของแบบทดสอบปิรามิดที่ใช้วิธีการแยกทางโดยลดค่าความยากลงขั้นละครึ่งอย่างคงที่ ซึ่งแสดงไว้ในภาพที่ 9 จะเห็นว่าข้อสอบข้อแรกจะมีค่าความยากอยู่ตรงมีชัยฐาน ($p=.50$) ข้อสอบในขั้นที่ 2 คือข้อสอบหมายเลข 2 และหมายเลข 3 จะมีค่าความยากอยู่กึ่งกลางระหว่างค่าความยากของข้อสอบในขั้นที่ 1 กับค่าความยากของข้อสอบที่มากที่สุดและน้อยที่สุดในแบบทดสอบ นั่นคือข้อสอบหมายเลข 2 มีค่าความยากเท่ากับ .75 และข้อสอบหมายเลข 3 มีค่าความยากมีค่าความยากเท่ากับ .25 ตามลำดับ ดังนั้นความแตกต่างระหว่างค่าความยากของข้อสอบจากขั้น 1 ไปยังขั้นที่ 2 เท่ากับ .25 ซึ่งผลทำให้ค่าความแตกต่างระหว่างค่าความยากของข้อสอบในขั้นที่ 1 และขั้นที่ 2 มีค่ามาก ทำให้การประมาณระดับความสามารถของผู้สอบ ในการตอบข้อสอบขั้นแรกนั้น เป็นการประมาณไปสู่จุดกึ่งกลางระหว่างระดับความสามารถที่สูงกว่า และที่ต่ำกว่า ในระดับความสามารถที่กระจายกันอยู่ทั้งหมด ในขั้นที่ 3 จะมีข้อสอบอยู่ 3 ข้อ ซึ่งจะมีค่าความยากของข้อสอบเท่ากับค่ากึ่งกลางระหว่างค่าความยากของข้อสอบในขั้นที่ 2 หรือระหว่างค่าความยากของข้อสอบในขั้นที่ 2 กับระดับความสามารถที่มีค่ามากที่สุดหรือน้อยที่สุด ดังนั้นข้อสอบหมายเลข 4 หมายเลข 5 และหมายเลข 6 จะมีค่าความยากเท่ากับ .875 .50 และ .125 ตามลำดับ นั่นคือค่าความแตกต่างระหว่างค่าความยากของข้อสอบในขั้นที่ 2 กับขั้นที่ 3 จึงมีค่าความยากเท่ากับ .125 (จากหมายเลข 2 ถึงหมายเลข 4 หรือจากหมายเลข 3 ถึงหมายเลข 6) หรือมีค่าเป็นครึ่งหนึ่งของค่าความแตกต่างระหว่างค่าความยากของข้อสอบในขั้นที่ 1 กับขั้นที่ 2 สำหรับค่าความแตกต่างระหว่างค่าความยากของข้อสอบในขั้นที่อยู่ติดกันอื่น ๆ ก็คำนวณหาได้โดยวิธีการเดียวกันนี้ กล่าวคือ สำหรับข้อสอบที่อยู่ในตำแหน่งปลายบนสุดหรือปลายล่างสุดของขั้น ค่าความยากของข้อสอบจะเป็นค่าที่อยู่กึ่งกลางระหว่างค่าความยากของข้อสอบที่อยู่ปลายสุดของขั้นที่อยู่ติดกันที่มาก่อนกับค่าความยากที่สูงสุด หรือต่ำสุดที่เป็นไปได้ของแบบทดสอบฉบับนั้น สำหรับข้อสอบที่อยู่ระหว่างข้อที่มีค่าความยากสูงสุดและต่ำสุดในขั้นนั้น ค่าความยากของข้อสอบจะประมาณจากครึ่งหนึ่งของค่าความยากของข้อสอบข้อที่อยู่เหนือข้อสอบข้อนั้น ๆ ในขั้นที่อยู่ติดกันที่มาก่อน



ภาพที่ 9 แสดงโครงสร้างของแบบทดสอบปิรามิดที่ใช้กฎการแยกทางโดยลดค่าความยากลง
ชั้นละครึ่งอย่างคงที่ (Weiss, 1974)

โครงสร้างแบบทดสอบปิรามิดลักษณะนี้ ค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันจะลดลงเร็วมากเมื่อข้อสอบข้อนั้นอยู่ใกล้ ๆ กับค่าความยากที่สูงสุดหรือต่ำสุดในแบบทดสอบฉบับนั้น และลดลงมากกว่าข้อสอบข้อที่อยู่ใกล้ ๆ กับค่าความยากตรงกลาง ๆ นอกจากนี้ โครงสร้างปิรามิดลักษณะนี้ยังมีผลให้บริเวณระดับค่าความยากสูงสุดหรือต่ำสุดมีจำนวนข้อสอบมากกว่าบริเวณที่มีค่าระดับความยากตรงกลาง ซึ่งแตกต่างกับโครงสร้างปิรามิดที่ใช้ขนาดขั้นคงที่ (ภาพที่ 7) ที่มีข้อสอบจำนวนมากข้อในบริเวณที่มีระดับความยากตรงกลาง ๆ จากลักษณะที่สำคัญของโครงสร้างรูปปิรามิดแบบที่ใช้ขนาดขั้นแปรผันนี้ จะทำให้มีความเที่ยงตรงในการให้คะแนนมากเท่า ๆ กันในทุกระดับความสามารถของผู้สอบโดยที่จำนวนข้อสอบก็ไม่มากไปกว่าโครงสร้างปิรามิดที่ใช้ขนาดขั้นคงที่

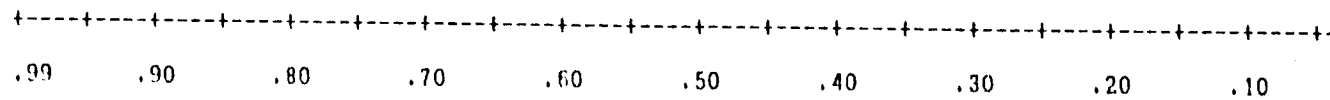
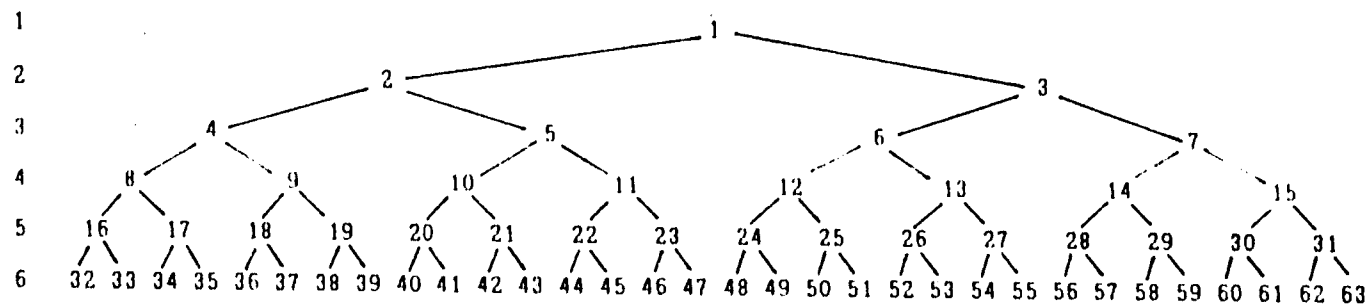
ในปี ค.ศ.1971 Lord (อ้างถึงใน Weiss,1974) ได้เสนอวิธีการกำหนดค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันในโครงสร้างปิรามิดแบบใช้ขนาดขั้นแปรผันอีกวิธีหนึ่ง ซึ่งคาดว่าจะมีความเที่ยงตรงต่อการวัดระดับความสามารถของผู้สอบแต่ละคน วิธีการดังกล่าวนี้อาศัยกระบวนการ Robbins-Monro โครงสร้างปิรามิดลักษณะนี้แสดงไว้ในภาพที่ 10 เป็นชนิด 6 ชั้น ในชั้นที่ 1 มีข้อสอบหมายเลข 1 มีค่าความยากเท่ากับ .50 ในชั้นที่ 2 มีข้อสอบหมายเลข 2 และ หมายเลข 3 ซึ่งมีค่าความยากเท่ากับ .75 และ .25 ตามลำดับ จะเห็นว่าค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่ 1 กับ ชั้นที่ 2 เท่ากับ .25 ค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่ 2 กับชั้นที่ 3 จะเท่ากับครึ่งหนึ่งของค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่ 1 กับชั้นที่ 2 คือเท่ากับ .125 ดังนั้นข้อสอบในชั้นที่ 3 คือข้อสอบหมายเลข 4, 5, 6 และ 7 จะมีค่าความยากของข้อสอบเป็น .875, .625, .375 และ .125 ตามลำดับ ค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่ 3 กับชั้นที่ 4 ก็จะเป็นเท่ากับครึ่งหนึ่งของค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่ 2 กับชั้นที่ 3 ซึ่งมีค่าเท่ากับ .0625 ค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกัน จะลดลงครึ่งหนึ่งเรื่อยไป ดังนั้นค่าความยากของข้อสอบจะเพิ่มขึ้นทุก ๆ ชั้น ครึ่งละครึ่งหนึ่งของค่าความยากในชั้นที่มีมาก่อน

จากกระบวนการของ Robbins-Monro จำนวนข้อสอบในแต่ละชั้นที่เพิ่มขึ้นจะเป็น 2 เท่าของจำนวนข้อสอบในชั้นที่มีมาก่อนเสมอ กล่าวคือในชั้นที่ 2 จะมีข้อสอบ 2 ข้อ ในชั้นที่ 3 จะมีข้อสอบ 4 ข้อ และในชั้นที่ 4 จะมีข้อสอบ 8 ข้อ ซึ่งจะเป็นดังนี้เรื่อยไป หรือกล่าวอีกนัยหนึ่งได้ว่า จะมีข้อสอบจำนวน 2^{n-1} ข้อ ในชั้นที่ n ดังนั้นแบบทดสอบรูปปิรามิดชนิด 6 ชั้น ถ้าใช้วิธีการ Robbins-Monro จะมีข้อสอบทั้งหมด 63 ข้อ โดยชั้นที่ 6 จะมีข้อสอบ 2^{6-1} คือ 32 ข้อ

กระบวนการที่ใช้การแยกทางคงที่มีจุดอ่อนคือ วัดระดับความสามารถได้ถูกต้องมากเฉพาะระดับความสามารถที่อยู่ใกล้ ๆ กับค่าเฉลี่ยเท่านั้น ส่วนโครงสร้างปิรามิดชนิดการแยกทางแปรผันที่

ชั้นที่

หมายเลขข้อสอบ



ข้อสอบง่าย

(ความสามารถต่ำ)

ข้อสอบยาก/ความสามารถ

(สัดส่วนการตอบถูก)

ข้อสอบยาก

(ความสามารถสูง)

ภาพที่ 10 แสดงโครงสร้างของแบบทดสอบปิรามิดชนิด 6 ชั้นที่ใช้กฎการแยกทางโดยลดค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่ติดกันของ Robbins-Monro (Weiser, 1974)

ใช้กระบวนการ Robbins-Monro ก็มีจุดอ่อนตรงที่ต้องใช้ข้อสอบจำนวนมากเพื่อจัดเข้าโครงสร้างแบบทดสอบปิรามิด เช่นโครงสร้างของ Robbins-Monro ชนิด 6 ขึ้นต้องใช้ข้อสอบจำนวน 63 ข้อ แต่แบบทดสอบปิรามิดชนิด 10 ขึ้นที่ใช้ขนาดขั้นคงที่ใช้ข้อสอบเพียง 55 ข้อเท่านั้น นอกจากนี้โครงสร้างของ Robbins-Monro ยังมีจุดอ่อนที่น่าสนใจอีกประการหนึ่งคือ มีความไวต่อโอกาสการตอบถูกมากกว่า ตัวอย่างเช่น กรณีที่ผู้สอบมีความสามารถต่ำ แต่โชคดีที่เดาข้อสอบหมายเลข 1 ถูก ก็จะได้ทำข้อสอบหมายเลข 3 ($p=.25$) และเขาตอบผิดจึงถูกแยกทางให้ตอบข้อสอบหมายเลข 6 ซึ่งมีค่าความยากลดลงเป็น .375 ถ้าสมมติว่าต่อไปเขาตอบข้อสอบทุกข้อตามแนวการแยกทางผิดหมด ข้อสอบข้อสุดท้ายที่เขาได้ตอบ คือหมายเลข 48 ที่มีความยากเท่ากับ .05 ซึ่งมีความหมายที่ดีกว่ากว่าผู้สอบที่ตอบข้อสอบที่มีความยาก .80 ได้ถูก แม้ว่าจะมีการกำหนดจำนวนขั้นมากขึ้น แต่ก็ยังคงขึ้นอยู่กับโอกาสการตอบข้อสอบข้อแรกอยู่นั่นเอง ดังนั้นกระบวนการของ Robbins-Monro ที่ใช้กับข้อสอบที่ให้ตอบอย่างอิสระจะต้องไม่มีการตอบโดยการเดาเกิดขึ้นเลย

แบบทดสอบปิรามิดที่ใช้ขนาดขั้นลดลงครั้งละครึ่งอย่างคงที่ในลักษณะที่แสดงในภาพที่ 9 จะมีความไวเกี่ยวกับผลของการเดาในลักษณะนี้น้อยกว่าที่ใช้กระบวนการ Robbins-Monro เช่น ผู้สอบที่เดาตอบข้อสอบหมายเลข 1 ถูกแต่เดาตอบข้อสอบข้ออื่น ๆ ผิดหมด ถ้าโครงสร้างแบบทดสอบนี้มี 6 ขั้น เขาจะได้ตอบข้อสอบหมายเลข 1, 3, 5, 8, 12 และ 17 ซึ่งข้อสอบหมายเลข 17 จะมีค่าความยาก .80 แต่การเดาตอบในลักษณะเช่นเดียวกันนี้ ถ้าเป็นแบบทดสอบปิรามิดที่ใช้ขนาดขั้นคงที่ดังภาพที่ 7 จะยังมีความไวต่อการเดาน้อยกว่า คือการที่ผู้สอบจะถูกประมาณระดับความสามารถโดยค่าความยากของข้อสอบข้อสุดท้ายที่ระดับความยาก .80 ได้ นั่นเมื่อเขาเดาตอบถูกในขั้นที่ 1 เขาจะต้องตอบผิดอีก 7 ข้อ ดังนั้นแบบทดสอบเลือกตอบที่จะนำมาใช้กับแบบทดสอบปิรามิด หากผู้สอบมีโอกาสเดาตอบถูกสูงแล้ว กระบวนการของ Robbins-Monro จะไม่เหมาะสมที่จะนำมาใช้ ควรพิจารณาใช้วิธีการอื่นแทน

รูปแบบโครงสร้างแบบทดสอบปิรามิดที่ใช้ขนาดขั้นแปรผัน หรือขนาดขั้นลดลงดังที่แสดงในภาพที่ 9 และภาพที่ 10 นี้เป็นเพียง 2 รูปแบบในหลาย ๆ รูปแบบ ในทางปฏิบัติถ้าต้องการแบบทดสอบปิรามิดที่สามารถวัดในทุกระดับความสามารถได้ถูกต้องมากที่สุด เราก็สามารถใช้วิธีการที่ใช้ขนาดขั้นคงที่ร่วมกับวิธีการที่ใช้ขนาดขั้นแปรผันร่วมกันในขั้นต่าง ๆ ได้

3. แบบทดสอบปิรามิดข้างคัต (Truncated Pyramids)

เนื่องจากแบบทดสอบปิรามิดที่ใช้กระบวนการของ Robbins-Monro แบบทดสอบปิรามิดที่ใช้ขนาดขั้นลดลง และแบบทดสอบปิรามิดที่ใช้ขนาดขั้นคงที่ จำเป็นต้องใช้ข้อสอบเป็นจำนวนมากในกลุ่ม

ข้อสอบ (Item Pool) ดังนั้น Mussio (อ้างถึงใน Weiss, 1974) จึงได้เสนอวิธีการลดจำนวนข้อสอบในโครงสร้างปิรามิดลง โดย Mussio อาศัย Markov Chain Stochastic Model ซึ่งใช้วิธีการสะท้อนกลับ หรือวิธีการเก็บรักษา (Reflecting or Retraining Barriers) Mussio ให้ข้อเสนอแนะว่าให้ตัดหางของแบบทดสอบปิรามิดออก กล่าวคือตัดข้อสอบที่มีระดับความยากบริเวณสูงสุดและต่ำสุดทิ้งไป จะทำให้ได้แบบทดสอบปิรามิดข้างตัด ซึ่งมีวิธีการแยกทางในการตอบข้อสอบอยู่ 2 วิธี คือ วิธีสังเกตการสะท้อนกลับและวิธีสังเกตการเก็บรักษา

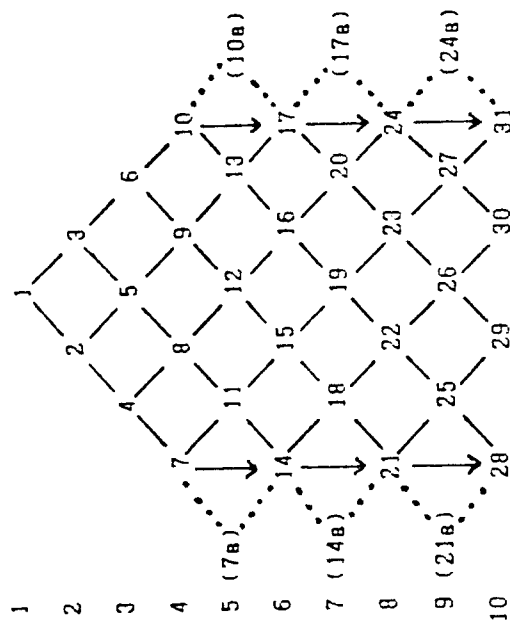
ภาพที่ 11 แสดงโครงสร้างแบบทดสอบปิรามิดชนิด 10 ชั้นที่ใช้วิธีสังเกตการสะท้อนกลับและวิธีสังเกตการเก็บรักษา ในภาพที่ 11a เป็นวิธีสังเกตการสะท้อนกลับ ข้อสอบจะมีความยากอยู่ในช่วง 0.65 ถึง 0.35 เท่านั้น ข้อสอบที่มีระดับความยากสูงหรือต่ำกว่านี้จะไม่จัดเข้าในโครงสร้างข้อสอบ จึงทำให้มีข้อสอบเพียง 31 ข้อใน 10 ชั้น วิธีการแยกทางการทำข้อสอบใน 4 ชั้นแรกจะดำเนินการตามปกติ คือถ้าผู้สอบตอบข้อสอบถูกต้องได้ตอบข้อสอบข้อที่ยากขึ้นเป็นข้อต่อไป และถ้าตอบผิดก็จะได้ทำข้อสอบข้อที่ง่ายลงเป็นข้อต่อไป เมื่อผู้สอบตอบข้อสอบผิดถึงข้อสอบหมายเลข 7 หรือตอบข้อสอบถูกต้องถึงข้อสอบหมายเลข 10 แล้วต่อไปจะใช้วิธีการของวิธีสังเกตการสะท้อนกลับ ซึ่งมีลักษณะดังนี้คือ ถ้าผู้สอบตอบข้อสอบหมายเลข 7, 14 หรือ 21 ผิด ข้อสอบที่ควรจะได้ตอบเป็นข้อต่อไปคือข้อสอบหมายเลข 7a, 14a หรือ 21a แต่ตามวิธีการสังเกตการสะท้อนกลับในแบบทดสอบปิรามิดข้างตัดจะให้ตอบข้อสอบหมายเลข 14, 21 หรือ 28 แทน ดังนั้นการใช้วิธีการสังเกตการสะท้อนกลับเมื่อผู้สอบตอบข้อสอบหมายเลข 7 ผิดจะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 14 แต่ถ้าผู้สอบตอบข้อสอบหมายเลข 7 ถูกจะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 11 ลักษณะสำคัญของวิธีการสังเกตการสะท้อนกลับก็คือ ข้อสอบ ณ จุดสังเกต หรือข้อสอบ ณ จุดตัดของรูปปิรามิดเป็นข้อสอบที่สมมติเอาว่าไม่มีความแตกต่างในการแยกทางเพื่อทำข้อสอบ จึงสามารถตัดออกจากโครงสร้างข้อสอบได้ จึงทำให้มีข้อสอบในโครงสร้างของแบบทดสอบน้อยลง คือ แบบทดสอบปิรามิดชนิด 10 ชั้นปกติจะมีข้อสอบ 55 ข้อแต่ถ้าเป็นแบบทดสอบปิรามิดชนิดข้างตัดตามวิธีการสังเกตการสะท้อนกลับจะมีข้อสอบเพียง 31 ข้อเท่านั้นซึ่งช่วยแก้ปัญหาการสร้างแบบทดสอบปิรามิด ที่ต้องใช้กลุ่มข้อสอบขนาดใหญ่ในการสร้างได้ แต่อย่างไรก็ตามการตัดข้อสอบที่มีค่าความยากมากกว่า .65 และน้อยกว่า .35 ออกจากโครงสร้างของแบบทดสอบตามวิธีการนี้ จะมีผลทำให้แบบทดสอบใช้วัดความสามารถของผู้สอบได้ในช่วงที่แคบลง

วิธีการสังเกตการเก็บรักษา จากภาพที่ 11b แสดงรูปแบบการแยกทางของแบบทดสอบปิรามิดข้างตัดโดยวิธีการสังเกตการเก็บรักษา ซึ่งถือว่าข้อสอบที่มีระดับความยากที่ปลายทั้งสองข้างของรูปปิรามิดจะถูกสะกิดกันโดยใช้ข้อสอบที่มีความยากเท่ากับข้อสอบข้อเดิมที่ได้ตอบมา ดังนั้นโครงสร้างของแบบทดสอบลักษณะนี้จะมีข้อสอบเพิ่มเข้ามา ณ จุดสะกิดในโครงสร้าง เช่นที่ระดับความยาก .70 จะมีข้อสอบเพิ่มเข้ามา 3 ข้อ คือข้อสอบหมายเลข 16, 27 และ 38 และที่ระดับความยาก .30 ก็จะมีข้อสอบเพิ่มเข้ามา 3 ข้อเช่นกัน คือข้อสอบหมายเลข 21, 32 และ 43

a

วิธีสกัดการสะท้อนกลับ

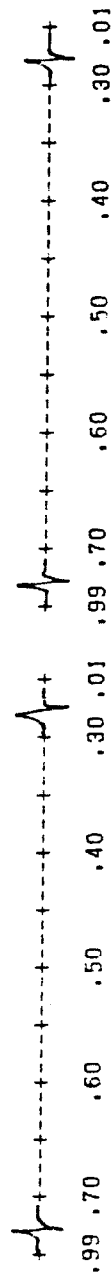
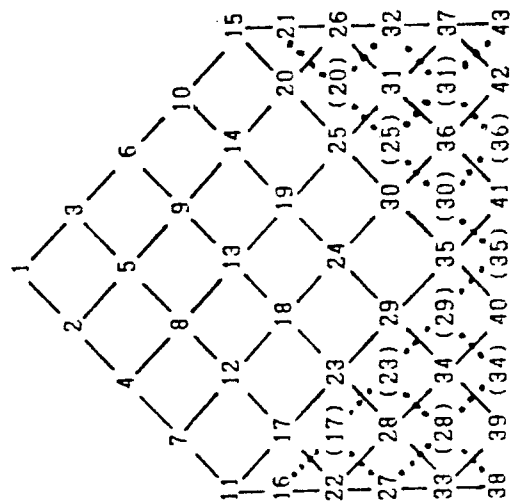
ชั้นที่



b

วิธีสกัดการเก็บรักษา

หมายเลขข้อสอบ



ค่าความยาก/ความสามารถ
(สัดส่วนการตอบถูก)

ค่าความยาก/ความสามารถ
(สัดส่วนการตอบถูก)

ภาพที่ 11 แสดงโครงสร้างของแบบทดสอบปิรามิดข้างต้น (a) วิธีสกัดการสะท้อนกลับ
(b) วิธีสกัดการเก็บรักษา (Weiss, 1974)

การแยกทางในการตอบข้อสอบตามวิธีการสะกิดการเก็บรักษาในช่วงแรก จะเป็นไปตามปกติของการตอบแบบทดสอบปิรามิดจนถึงจุดสะกิดกัน คือผู้สอบที่ตอบข้อสอบ 4 ข้อแรกใน 4 ชั้นแรกผิดก็จะได้ทำข้อสอบที่ง่ายจากข้อสอบข้อ 1, 2, 4 และ 7 จนถึงข้อสอบหมายเลข 11 ถ้าตอบข้อสอบหมายเลข 11 ถูกก็จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 17 เป็นข้อต่อไป แต่ถ้าตอบข้อสอบหมายเลข 11 ผิดจะต้องตอบข้อสอบหมายเลข 16 ซึ่งมีค่าความยากเท่ากับข้อสอบหมายเลข 11 เป็นข้อต่อไป และในทำนองเดียวกันถ้าตอบข้อสอบหมายเลข 16 ผิดอีกก็จะได้ตอบข้อสอบที่มีระดับความยากเท่าเดิมจนถึงข้อสอบหมายเลข 38

ในกรณีที่ผู้สอบตอบข้อสอบข้อใดข้อหนึ่งในจำนวนที่เพิ่มนี้ถูก ตามวิธีการสะกิดการเก็บรักษาก็จะให้ทำข้อสอบข้อที่ยากขึ้นตามรูปแบบการแยกทางที่แสดงไว้โดยเส้นประในภาพที่ 11b กล่าวคือถ้าตอบข้อสอบหมายเลข 16 ในชั้นที่ 6 ถูกจะได้ตอบข้อสอบหมายเลข (17) ในชั้นที่ 7 ซึ่งจะเห็นว่าการแยกทางในการตอบข้อสอบในแนวเส้นประนี้จะมีข้อสอบที่มีหมายเลขซ้ำกับหมายเลขในโครงสร้างหลัก แต่ให้วงเล็บไว้ให้ทราบว่าข้อสอบเหล่านี้ เป็นข้อสอบที่จะให้ตอบในการแยกทางตามวิธีการสะกิดการเก็บรักษาในกรณีที่ตอบข้อสอบ ณ จุดสะกิดถูก และเมื่อผู้สอบตอบข้อสอบที่มีหมายเลขซ้ำกับในโครงสร้างหลักเหล่านี้ที่อยู่ในวงเล็บครบหมดทุกข้อแล้วจะถือว่าถึงจุดยุติการสอบ

Mussio (อ้างถึงใน Weiss, 1974) ได้เสนอไว้ว่าจำนวนข้อสอบที่ใช้ในโครงสร้างของแบบทดสอบปิรามิดข้างตัดชนิด 10 ชั้นตามวิธีการสะกิดการสะท้อนกลับจะมี 31 ข้อ ส่วนวิธีการสะกิดการเก็บรักษาจะมี 43 ข้อระดับความยากของข้อสอบในวิธีการสะกิดการสะท้อนกลับจะมีเพียง 7 ระดับ ส่วนวิธีการสะกิดการเก็บรักษาจะมีถึง 9 ระดับ ซึ่งเมื่อเปรียบเทียบกับโครงสร้างของแบบทดสอบปิรามิดที่ใช้ขนาดขั้นคงที่ในภาพที่ 7 แล้วจะต้องใช้ข้อสอบจำนวน 55 ข้อที่มีระดับความยากของข้อสอบถึง 19 ระดับ จะเห็นได้ว่าโครงสร้างแบบทดสอบปิรามิดข้างตัดทั้งวิธีการสะกิดการสะท้อนกลับและวิธีการสะกิดการเก็บรักษาที่ Mussio เสนอไว้นั้นนอกจากสามารถใช้ในการกำหนดขนาดขั้นได้ทั้งขนาดขั้นคงที่และขนาดขั้นแปรผันแล้ว ประโยชน์ที่สำคัญของแบบทดสอบปิรามิดข้างตัดอีกประการหนึ่งคือใช้ข้อสอบน้อยกว่าแบบทดสอบปิรามิดที่ใช้ขนาดขั้นคงที่ และใช้ข้อสอบที่มีช่วงระดับความยากแคบกว่าจึงสามารถสร้างได้สะดวกกว่า อย่างไรก็ตามจุดอ่อนที่สำคัญของแบบทดสอบปิรามิดข้างตัดคือเมื่อมีช่วงความยากของข้อสอบแคบลงก็จะทำให้สามารถวัดระดับความสามารถของผู้สอบในช่วงที่แคบลงด้วย

4. แบบทดสอบปิรามิดที่มีข้อสอบหลายข้อในแต่ละชั้น (Multi Item Per Stage)

ได้มีผู้ศึกษาและเสนอรูปแบบการแยกทางในการดำเนินการสอบแบบทดสอบปิรามิดชั้นอีกลักษณะหนึ่ง คือใช้ข้อสอบมากกว่าหนึ่งข้อในแต่ละชั้นของโครงสร้างแบบทดสอบปิรามิด ทั้งนี้เนื่องจากการ

ปรับปรุงความเชื่อมั่นของการแยกทางให้สูงขึ้นและยังเป็นการลดจำนวนขั้นของแบบทดสอบลงด้วย ซึ่งวิธีการโดยทั่วไปของแบบจำลองนี้ จะใช้ผลการตอบข้อสอบทุกข้อในแต่ละชั้นเป็นเกณฑ์ในการกำหนดการแยกทางเพื่อหาข้อสอบในชั้นถัดไป ภาวนที่ 12 แสดงโครงสร้างแบบทดสอบปิรามิดที่มีข้อสอบชุดละ 3 ข้อในแต่ละชั้น ข้อสอบ 3 ข้อในชุดแรกในชั้นที่ 1 เป็นข้อสอบที่มีความยากระหว่าง .55 กับ .45 ผู้สอบทุกคนจะเริ่มทำข้อสอบ 3 ข้อนี้เหมือนกัน เมื่อตอบครบ 3 ข้อแรกนี้แล้ว จะตรวจผลการตอบข้อสอบแต่ละข้อโดยถ้าตอบถูกจะได้ 1 ตอบผิดได้ 0 จากการตอบ 3 ข้อนี้ ถ้าผลรวมคะแนนเป็น 0 หรือ 1 จะถูกแยกทางให้ตอบข้อสอบหมายเลข 4, 5 และ 6 ในชุดที่ 2 ของชั้นที่ 2 ซึ่งมีระดับความยากลดลงคืออยู่ระหว่าง .56 กับ .65 ส่วนผู้ที่ตอบข้อสอบ 3 ข้อแรกในชั้นที่ 1 ได้คะแนน 2 หรือ 3 จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 7, 8 และ 9 ในชุดที่ 3 ของชั้นที่ 2 ซึ่งมีระดับความยากสูงขึ้นคือมีค่าอยู่ระหว่าง .44 กับ .35 จำนวนข้อสอบที่ผู้สอบจะต้องตอบทั้งหมดในแบบทดสอบปิรามิดชนิดที่มีข้อสอบชุดละ 3 ข้อในแต่ละชั้นนี้ มีจำนวน 15 ข้อ ซึ่งจะมากกว่าแบบทดสอบปิรามิดที่มีข้อสอบ 1 ข้อในแต่ละชั้นที่มีจำนวนชั้น 5 ชั้นเท่ากันซึ่งจะหาข้อสอบเพียง 5 ข้อ

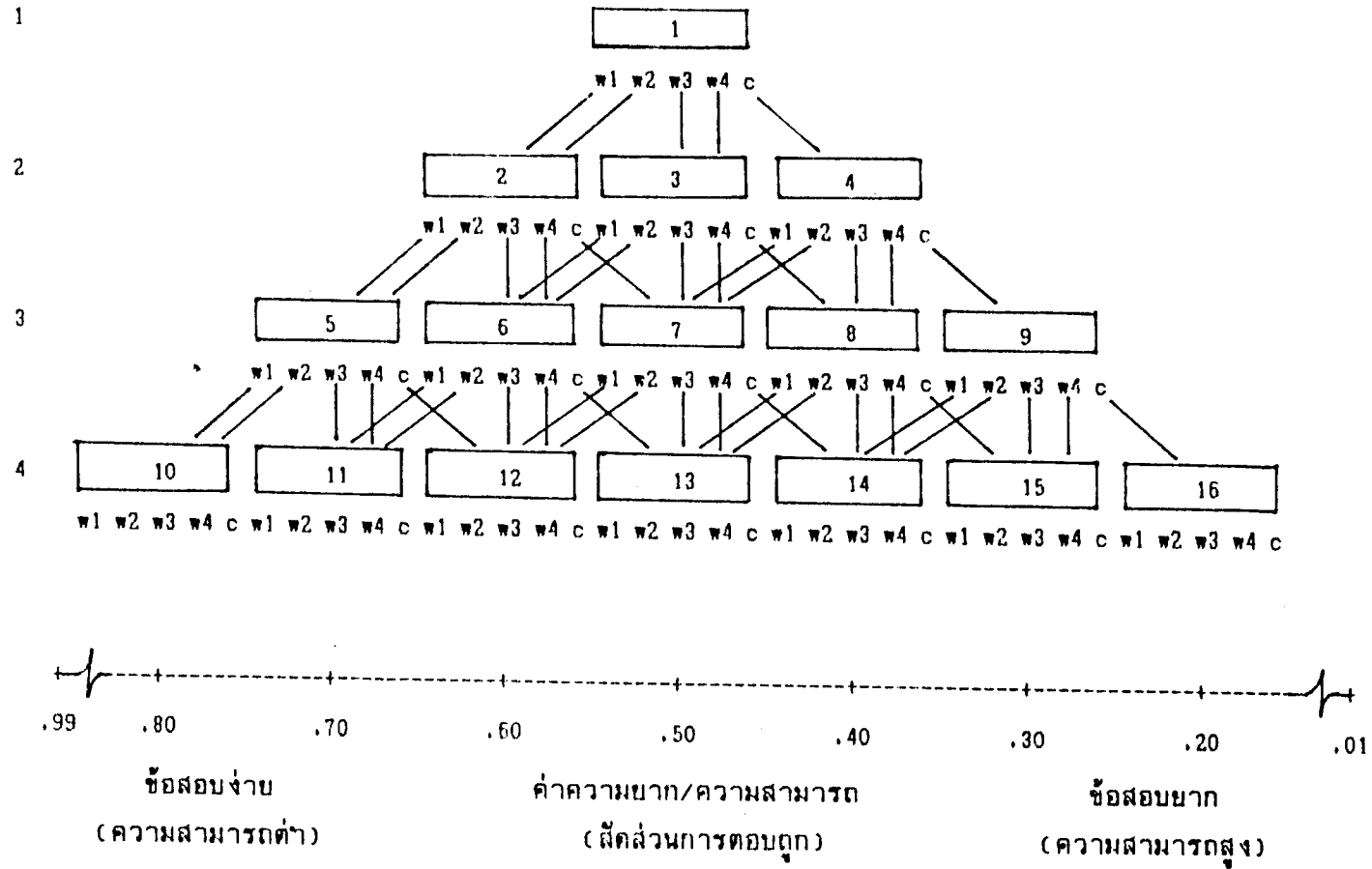
โครงสร้างของแบบทดสอบปิรามิดที่มีข้อสอบหลายข้อในแต่ละชั้นสามารถกำหนดได้หลายลักษณะ คือจะใช้ขนาดชั้นคงที่ ขนาดชั้นแปรผัน หรือจำนวนชั้นที่ให้แยกทางไปทำข้อสอบที่ยากกับจำนวนชั้นที่ให้แยกทางไปทำข้อสอบที่ง่ายไม่เท่ากันก็ได้ นอกจากนี้ยังสามารถกำหนดจำนวนข้อสอบในแต่ละชั้นไม่เท่ากันก็ได้ คือในชั้นแรก ๆ อาจใช้จำนวนข้อสอบมากกว่าในชั้นหลัง ๆ ทั้งนี้เพื่อให้สามารถจำแนกระดับความสามารถของผู้สอบได้ดีขึ้นในช่วงแรก ๆ ซึ่งกล่าวได้ว่าเป็นการเพิ่มค่าอำนาจจำแนกในช่วงแรกของแบบทดสอบให้สูงขึ้นนั่นเอง จากการที่ใช้คะแนนผลการตอบในแต่ละชั้นเป็นเกณฑ์ในการแยกทางแบบทดสอบปิรามิดชนิดนี้จึงคล้ายกับเป็นการรวมเอาแบบทดสอบชนิดสองชั้นมารวมเข้ากับแบบทดสอบปิรามิด ซึ่งประโยชน์ที่สำคัญของแบบทดสอบปิรามิดที่มีข้อสอบหลายข้อในแต่ละชั้นนี้คือช่วยทำให้ลดผลจากการเตาถูกที่มีต่อการวัดได้ นั่นคือการแยกทางที่อาศัยผลการตอบข้อสอบมากกว่า 1 ข้อในแต่ละชั้นนั้นจะมีความถูกต้องหรือมีความเชื่อมั่นสูงขึ้น ซึ่งถือว่าเป็นการปรับปรุงการวัดผลที่ถูกต้องมากยิ่งขึ้น

5. แบบทดสอบปิรามิดที่ให้นั้หน้านักตัวเลือกของข้อสอบเพื่อการแยกทาง (Differential Response Option Branching)

แบบทดสอบปิรามิดชนิดนี้จะใช้สารสนเทศทั้งหมดจากคำตอบของผู้สอบที่ตอบข้อสอบเพื่อกำหนดการแยกทางให้ทำข้อสอบข้อต่อไปในชั้นถัดไป ซึ่งจะใช้ระดับของการตอบผิด (Degree of Wrongness) ของคำตอบที่ผิดในแต่ละข้อเป็นเกณฑ์ในการแยกทาง รูปแบบการแยกทางลักษณะนี้สามารถใช้ได้กับข้อสอบที่เป็นแบบเลือกตอบ และข้อสอบที่ให้ตอบอย่างอิสระ (Free Response

ชั้นที่

หมายเลขข้อสอบ



ภาพที่ 13 แสดงโครงสร้างแบบทดสอบปราคิให้ให้นักตัวเลือกของข้อสอบ
เพื่อการแยกทาง (Weiss, 1974)

Items) ที่สามารถกำหนดค่าความยากของแต่ละคำตอบได้ วิธีการนี้เสนอโดย Bayroff (อ้างถึงใน Weiss, 1974) ซึ่ง Bayroff ได้ศึกษาโดยใช้วิธีนี้ในชั้นที่ 1 ของแบบทดสอบชนิด 5 ชั้นเท่านั้น จากภาพที่ 13 แสดงรูปแบบหนึ่งของแบบทดสอบปิรามิดที่ให้นักหน้าตัวเลือกของข้อสอบเพื่อการแยกทาง ในชั้นที่ 1 จะมีข้อสอบ 1 ข้อ ชั้นที่ 2 จะมีข้อสอบ 3 ข้อ ชั้นที่ 3 มีข้อสอบ 5 ข้อ และในชั้นที่ 4 มีข้อสอบ 7 ข้อ จะเห็นได้ว่าจำนวนข้อสอบในชั้นถัดไปจะเพิ่มขึ้นจากชั้นเดิม 2 ข้อ ข้อสอบที่แสดงในภาพที่ 13 เป็นข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก มีคำตอบที่ถูกอยู่ 1 ตัวเลือก (C) และคำตอบที่ผิดอีก 4 ตัวเลือก (w_1, w_2, w_3 และ w_4) ตัวเลือกที่ผิดมากที่สุดแทนด้วย w_1 และตัวเลือกที่ผิดน้อยที่สุดหรือเกือบจะถูกต้อง แทนด้วย w_4 โดยถือเอาคำตอบของกลุ่มคนปกติในการกำหนดน้ำหนักของคำตอบ

ในชั้นที่ 1 เมื่อผู้สอบตอบข้อสอบหมายเลข 1 ถูกหรือเลือกตัวเลือก C จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 4 ในชั้นที่ 2 ซึ่งเป็นข้อสอบที่มีความยากมากที่สุด ในชั้นที่ 2 ถ้าตอบตัวเลือกที่มีระดับความผิด w_1 หรือ w_2 จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 2 ในชั้นที่ 2 ซึ่งมีความยากน้อยกว่าข้อสอบหมายเลข 1 ในชั้นที่ 1 ส่วนผู้สอบที่เลือกตอบตัวเลือก w_3 หรือ w_4 ซึ่งมีระดับความผิดน้อยกว่า w_1 และ w_2 แต่ยังไม่ใช้คำตอบที่ถูกต้อง จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 3 ในชั้นที่ 2 ซึ่งมีความยากเท่ากับข้อสอบหมายเลข 1 ในชั้นที่ 1 ทั้งนี้เพราะเหตุผลว่า ผู้ที่จะเลือกตัวเลือก w_3 และ w_4 ส่วนใหญ่จะมีความสามารถในระดับเฉลี่ยของกลุ่มคนปกติ ส่วนผู้ที่มีระดับความสามารถที่สูงหรือต่ำกว่านี้จะถูกแยกทางให้ไปตอบข้อสอบที่มีความยากสูงหรือต่ำกว่า และจะดำเนินการเช่นนี้จนถึงขั้นสุดท้ายของโครงสร้างของแบบทดสอบลักษณะนี้

แบบทดสอบที่ให้นักหน้าตัวเลือกของข้อสอบนี้ จะให้สารสนเทศของผู้สอบจากคำตอบที่เลือกได้มากกว่าแบบทดสอบปิรามิดที่ให้คำตอบถูกเป็น 1 และผิดเป็น 0 ทั้งนี้เพราะการเลือกตอบในแต่ละตัวเลือกที่มีระดับความไม่ถูกต้องแตกต่างกันนี้ จะทำให้ทราบว่าผู้ตอบเข้าใจผิดในลักษณะใดในแต่ละเนื้อหาที่วัดของข้อสอบแต่ละข้อ ซึ่งจะเป็ประโยชน์มากสำหรับการสร้างแบบทดสอบเพื่อการวินิจฉัยข้อบกพร่องในการเรียน อย่างไรก็ตามแบบทดสอบปิรามิดที่ให้นักหน้าตัวเลือกของข้อสอบเพื่อการแยกทางที่แสดงในภาพที่ 13 เป็นเพียงรูปแบบหนึ่งเท่านั้น ซึ่งถ้าจะให้การแยกทางตามน้ำหนักตัวเลือกที่สมบูรณ์แล้วควรจะให้แยกทางไปหาข้อสอบข้อที่แตกต่างกันในแต่ละตัวเลือกที่ผิด คือถ้าผู้สอบที่ตอบตัวเลือกที่มีระดับความผิดเป็น w_1 จะได้ตอบข้อสอบที่มีความยากน้อยกว่าข้อสอบที่ผู้เลือกตอบตัวเลือก w_2 ในชั้นถัดไป วิธีการนี้เมื่อใช้กับข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก การตอบข้อสอบในชั้นหนึ่ง ๆ ก็จะมีทางเลือกในการแยกทางเนื้อหาข้อสอบในชั้นถัดไป 5 ทาง ซึ่งมีผลทำให้โครงสร้างแบบทดสอบไม่สมมาตรโดยมีทางเลือกที่มีโอกาสไปทางซ้ายมากกว่าทางขวา และจำเป็นต้องมีกลุ่ม

ข้อสอบขนาดใหญ่ขึ้น ซึ่งเป็นข้อจำกัดที่ทำให้สร้างได้ยาก แต่อย่างไรก็ตามแบบทดสอบปิรามิดรูปแบบนี้มีความถูกต้องในการวัดมากขึ้น จำนวนข้อสอบที่ผู้สอบแต่ละคนทำมีจำนวนคงที่ และผลการตอบข้อสอบแต่ละข้อของผู้สอบแต่ละคน จะให้สารสนเทศเกี่ยวกับผู้สอบมากในด้านการวินิจฉัยข้อบกพร่องในการเรียนรู้ แบบทดสอบปิรามิดรูปแบบนี้จึงได้รับความสนใจอย่างกว้างขวาง

2.4 วิธีการให้คะแนนแบบทดสอบปิรามิด

แบบทดสอบปิรามิดทั้ง 5 รูปแบบที่กล่าวมานั้น สามารถให้คะแนนวิธีเดียวกันหรือต่างกันได้ได้ การให้คะแนนมีอยู่หลายวิธี ผลการให้คะแนนแต่ละวิธีขึ้นอยู่กับแนวทางในการแยกทางเนื้อหาข้อสอบในชั้นถัดไปและค่าความยากของข้อสอบที่ผู้สอบได้ตอบ ดังนั้นจึงมีข้อตกลงว่ากลุ่มข้อสอบที่นำมาสร้างแบบทดสอบปิรามิดจะต้องวัดในเนื้อหาเดียวกัน หรือมีมิติเดียว และมาตรวัดระดับความยากของข้อสอบกับมาตรวัดระดับความสามารถของผู้สอบเป็นมาตรวัดเดียวกัน การให้คะแนนแบบทดสอบปิรามิดมีอยู่ 6 วิธี ดังต่อไปนี้

2.4.1 ให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบ (Difficulty of the Final Item)

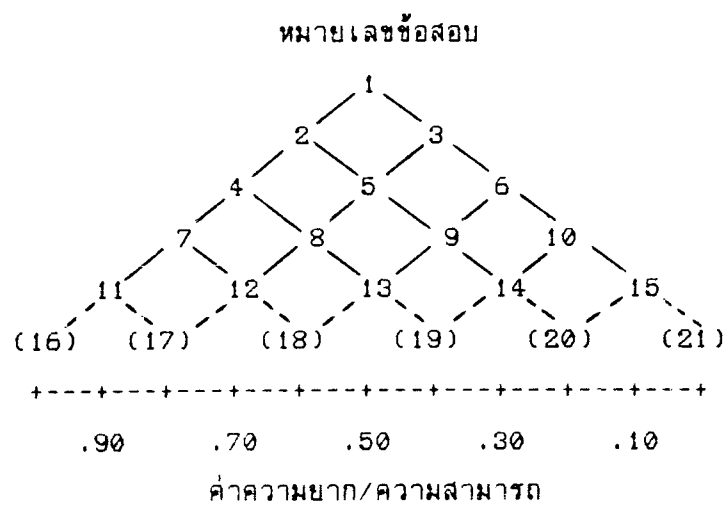
การให้คะแนนวิธีนี้มีข้อสมมติว่าเป็นการให้คะแนนตามความสามารถในการสอบที่ปฏิบัติได้สูงสุด (Maximum Performance) คะแนนที่ให้จะเชื่อถือได้น้อยลงเมื่อมีการตอบข้อสอบโดยการเดา และถ้าจำนวนข้อในแบบทดสอบมีน้อยคะแนนที่ให้อาจจะน้อยเกินไป มีผลให้ค่าสหสัมพันธ์ระหว่างคะแนนแบบทดสอบปิรามิดที่ให้โดยวิธีนี้กับแบบทดสอบอื่น ๆ มีค่าต่ำ กล่าวคือความตรงเชิงสัมพัทธ์กับเกณฑ์มีค่าต่ำ ดังนั้นในทางปฏิบัติแล้วการให้คะแนนวิธีนี้จึงให้ประโยชน์น้อย

2.4.2 ให้คะแนนตามค่าความยากของข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ (Difficulty of the $N+1^{th}$ Item)

วิธีการให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบนั้น ไม่ได้พิจารณาถึงคำตอบข้อสุดท้ายว่าถูกหรือผิด คะแนนของผู้สอบที่ทำข้อสุดท้ายข้อเดียวกันจึงเท่ากันไม่ว่าจะตอบถูกหรือผิด การให้คะแนนตามค่าความยากของข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบ ($N+1^{th}$) ซึ่งสมมติว่าเป็นข้อสอบที่ควรจะได้ทำต่อไปจากผลการตอบข้อสอบข้อสุดท้าย การให้คะแนนโดยวิธีนี้จึงมีข้อสมมติว่ามีข้อสอบขั้นสุดท้ายที่ต่อจากขั้นที่ N ของแบบทดสอบจริง ดังแสดงไว้ในภาพที่ 14a ซึ่งเป็นแบบทดสอบปิรามิดชนิด 5 ขั้น

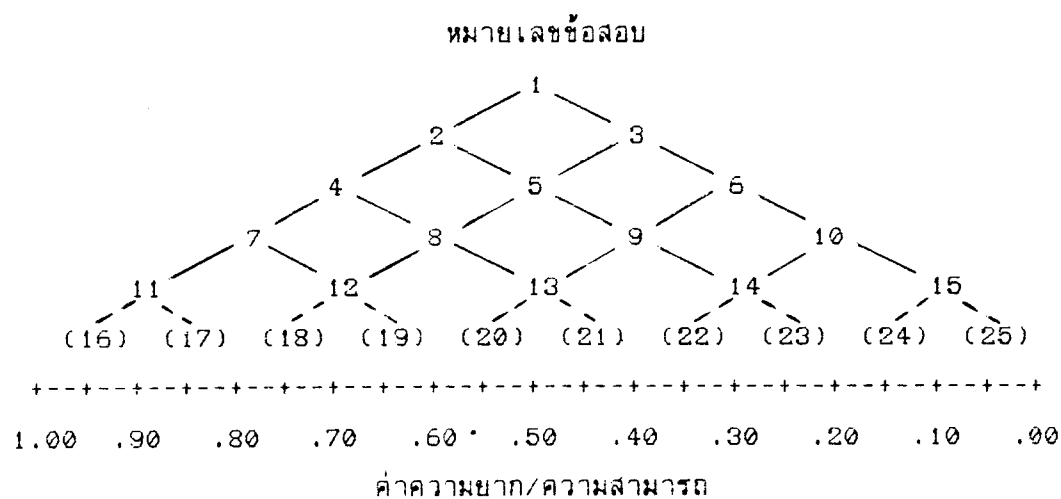
a

ให้คะแนนตามค่าความยากของข้อสอบ
ที่คัดจากข้อสุดท้ายที่ได้ตอบ



b

ให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย



ภาพที่ 14 แสดงการให้คะแนนแบบทดสอบปิรามิต 2 วิธี ที่ใช้ผลการตอบข้อสอบข้อสุดท้าย

จากภาพที่ 14a ในขั้นที่ 5 จะมีข้อสอบหมายเลข 11 ถึง 15 และมีข้อสอบหมายเลข 16 ถึง 21 ในขั้นที่ 6 ซึ่งเป็นข้อสอบที่สมมติขึ้นสำหรับนิจาร์ณาให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย เช่น ถ้าผู้สอบทำข้อสอบหมายเลข 12 ซึ่งมีค่าความยากเท่ากับ .70 เป็นข้อสุดท้าย ถ้าให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายผู้สอบคนนี้ก็จะได้คะแนน .70 ไม่ว่าจะตอบข้อสอบข้อนี้ถูกหรือผิดก็ตาม แต่ถ้านิจาร์ณาให้คะแนนตามค่าความยากของข้อสอบที่ถัดจากข้อสุดท้ายที่ได้ตอบนั้น จะต้องนิจาร์ณาว่าผลการตอบข้อสอบหมายเลข 12 นั้นถูกหรือผิด ถ้าตอบถูกก็จะได้คะแนนเท่ากับ .60 ซึ่งเป็นค่าความยากของข้อสอบหมายเลข 18 แต่ถ้าตอบผิดก็จะได้คะแนนเท่ากับ .80 ซึ่งเป็นค่าความยากของข้อสอบหมายเลข 17 เป็นต้น การให้คะแนนวิธีนี้จะให้ช่วงของคะแนนแคบกว่าแบบทดสอบปิรามิดนั้นมีจำนวนชั้นน้อย

2.4.3 ให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย (Terminal Right-Wrong)

จากวิธีการให้คะแนนตามค่าความยากของข้อสอบข้อที่ถัดจากข้อสอบข้อสุดท้ายที่ได้ตอบนั้น มีปัญหาอยู่ว่า ผู้สอบที่ทำข้อสอบที่ง่ายกว่าถูกกับผู้สอบที่ทำข้อสอบที่ยากกว่าในชั้นเดียวกันผิดจะได้คะแนนเท่ากันหรือตีความหมายว่ามีระดับความสามารถเท่ากัน ซึ่งวิธีการให้คะแนนตามผลการตอบข้อสุดท้ายถือว่าความสามารถของผู้สอบทั้งสองคนดังกล่าวน่าจะแตกต่างกัน จากภาพที่ 14b แสดงให้เห็นว่าข้อสอบแต่ละข้อในขั้นสุดท้ายหรือขั้นที่ 5 นั้น สามารถแยกทางไปยังข้อสอบที่สมมติขึ้นในขั้นที่ 6 ได้อีก 2 ข้อ เช่น ข้อสอบหมายเลข 12 จะแยกทางไปยังข้อสอบหมายเลข 18 และ 19 ซึ่งความยากของข้อสอบหมายเลข 18 ($p=.75$) จะน้อยกว่าข้อสอบหมายเลข 12 ($p=.70$) และข้อสอบหมายเลข 19 ($p=.65$) จะมีความยากมากกว่าข้อสอบหมายเลข 12 ดังนั้นการให้คะแนนโดยวิธีนี้ ถ้าผู้สอบตอบข้อสอบหมายเลข 12 ซึ่งเป็นข้อสอบข้อสุดท้ายผิดก็จะได้คะแนนเท่ากับ .75 ซึ่งเป็นค่าความยากของข้อสอบหมายเลข 18 แต่ถ้าตอบข้อสอบหมายเลข 12 ถูกก็จะได้คะแนนเท่ากับ .65 ซึ่งเป็นค่าความยากของข้อสอบหมายเลข 19

2.4.4 ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ (Average Difficulty)

Lord (อ้างถึงใน Weiss, 1974) ได้เสนอการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบขึ้น วิธีการนี้จะนำค่าความยากของข้อสอบทุกข้อที่ผู้สอบได้ตอบโดยไม่คำนึงว่าจะตอบถูกหรือตอบผิดมาหาค่าเฉลี่ยเป็นคะแนน Lord ไม่ได้นำค่าความยากของข้อสอบข้อแรกมาคิดด้วยแต่จะใช้ค่าความยากของข้อสอบข้อที่ถัดจากข้อสุดท้าย ($N+1^{th}$ Item) ที่ได้ตอบที่สมมติขึ้นมาคิดแทน วิธีการให้คะแนนโดยวิธีนี้ จะให้สารสนเทศของคะแนนเช่นเดียวกับวิธีการให้คะแนนตามจำนวนข้อสอบที่ตอบถูกในแบบทดสอบปิรามิดที่ใช้ขนาดชั้นคงที่

ถึงแม้ว่าคะแนนที่ได้จากการให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบจะสูงกว่าการให้โดยวิธีอื่นบางวิธี เช่น มากกว่าคะแนนที่ให้ตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบก็ตาม การให้คะแนนโดยวิธีนี้ก็ยังมีข้อจำกัดในการวัดระดับความสามารถของผู้สอบได้ไม่กว้างมากนัก เพราะวิธีการนี้จะให้คะแนนโดยการรวมค่าความยากของข้อสอบที่มีค่าความยากใกล้เคียง ๆ กับความสามารถเฉลี่ยของผู้สอบแล้วนำมาคิดเป็นค่าเฉลี่ยเพื่อให้เป็นคะแนน คะแนนที่ได้จึงไม่สามารถกระจายไปให้ครอบคลุมถึงระดับความสามารถที่สูงและต่ำมาก ๆ ได้ ตัวอย่างจากโครงสร้างแบบทดสอบปียามิตชนิด 5 ชั้นในภาพที่ 14a ถ้าผู้สอบที่มีความสามารถสูงสุดตอบข้อสอบถูกหมดทุกข้อคือได้ตอบข้อสอบหมายเลข 1, 3, 6, 10 และ 15 ถูกทั้งหมด ถ้าเป็นการให้คะแนนตามวิธีให้ตามค่าความยากของข้อที่ติดจากข้อสุดท้ายจะได้คะแนนตามค่าความยากของข้อสอบหมายเลข 21 ในชั้นที่สมมุติขึ้นซึ่งจะมีค่าความยากสูงกว่า .10 อยู่ 10 % หรือกล่าวได้ว่ามีความสามารถสูงกว่า .10 แต่การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบซึ่งจะได้คะแนนเท่ากับค่าเฉลี่ยของค่าความยากของข้อสอบหมายเลข 3, 6, 10, 15 และ 21 มีค่าเท่ากับ .20 ซึ่งหมายความว่ามีความสามารถต่ำกว่าระดับความสามารถ .10 ดังนั้นจึงสรุปได้ว่า ถึงแม้ว่าการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบจะให้ค่าคะแนนสูงกว่าการให้คะแนนโดยวิธีอื่น และเป็นการพิจารณาจากคำตอบทั้งหมดของผู้สอบซึ่งถือว่าเป็นการประมาณความสามารถของผู้สอบหลายครั้งหรือจากข้อสอบหลายข้อ ซึ่งแตกต่างจากวิธีอื่นที่ประมาณความสามารถของผู้สอบจากข้อสอบเพียงข้อเดียวก็ตาม การให้คะแนนโดยวิธีนี้ก็ยังคงให้ความหมายที่ไม่ดีกว่าการให้คะแนนโดยวิธีอื่น ๆ

2.4.5 ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก (Average Difficulty of All Item Answered Correctly or Average Difficulty Correct)

การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก เป็นวิธีที่ซับซ้อนบ่งชี้ของการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ วิธีนี้จะนำเอาค่าความยากของข้อสอบที่ตอบถูกเท่านั้นมาหาค่าเฉลี่ย ซึ่งคะแนนที่ได้จากค่าเฉลี่ยนี้ จะมีความหมายในการบอกระดับความสามารถของผู้สอบได้มากขึ้น คะแนนที่ได้โดยวิธีนี้จะมากน้อย ๆ กับการให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย และการให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบ การให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก จะมีความคงที่ของคะแนน และให้สารสนเทศของคะแนนมากกว่าการให้คะแนนตามผลการตอบข้อสอบข้อสุดท้าย และการให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบ อย่างไรก็ตามการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูกนี้ จะมีความไวต่อการเดาตอบถูกมากกว่าการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ ซึ่งอาจมีผลให้ความเชื่อมั่นต่ำลง

2.4.6 ให้คะแนนข้อสอบทุกข้อ (All Item Score)

Hansen (อ้างถึงใน Weiss, 1974) ได้เสนอวิธีการให้คะแนนข้อสอบทุกข้อ ซึ่งสามารถใช้ได้กับการประมาณค่าความเชื่อมั่นชนิดความคงที่ภายใน (Internal Consistency) ของแบบทดสอบปรัมาติที่ทดสอบกับผู้สอบเป็นกลุ่ม และผู้สอบทุกคนจะต้องตอบข้อสอบตามเส้นทางการตอบทุกข้อ Hansen มีข้อตกลงว่าค่าความยากของข้อสอบและระดับความสามารถของผู้สอบถือเป็นเรื่องเดียวกัน วิธีการให้คะแนนตามวิธีนี้ข้อสอบที่ตอบถูกทุกข้อจะให้ 2 คะแนน ให้คะแนนข้อที่ง่ายกว่าข้อที่ตอบถูกในชั้นเดียวกันอีกข้อละ 2 คะแนน ให้คะแนนข้อสอบข้อถัดไปในชั้นเดียวกันที่มีความยากมากกว่าข้อที่ตอบถูก 1 คะแนน ส่วนข้อที่มีค่าความยากมากกว่าข้อที่ตอบถูกในชั้นเดียวกันที่เหลือจะให้ 0 คะแนน การให้คะแนนข้อที่ตอบผิดจะให้ 0 คะแนน ให้คะแนนข้อสอบที่มีค่าความยากน้อยกว่าและอยู่ติดกับข้อสอบที่ตอบผิดที่อยู่ในชั้นเดียวกัน 1 คะแนน และให้คะแนนข้อสอบที่มีความยากน้อยกว่าข้อที่ตอบผิดที่เหลือทุกข้อข้อละ 2 คะแนน Hansen ได้เสนอวิธีการให้คะแนนไว้เท่านั้นไม่ได้ศึกษาประสิทธิภาพของวิธีการนี้ ต่อมา Larkin และ Weiss ได้วิจัยเพื่อศึกษาวิธีการให้คะแนนวิธีนี้ในปี ค.ศ. 1974 ผลการวิจัยพบว่า คะแนนที่ให้โดยวิธีนี้มีความสัมพันธ์กับคะแนนที่ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบซึ่งเสนอโดย Lord สูงถึง .99

2.4.7 ให้คะแนนตามจำนวนข้อสอบที่ตอบถูก (Number Correct)

วิธีการนี้จะให้คะแนนตามจำนวนข้อที่ตอบถูก คือให้ข้อที่ตอบถูกข้อละ 1 คะแนน การให้คะแนนวิธีนี้จะมีผลต่อการเดาตอบของผู้สอบมาก และถ้าแบบทดสอบปรัมาติที่มีจำนวนขึ้นน้อย คะแนนที่ได้ก็จะน้อยเกินไปทำให้ความตรงเชิงเปรียบเทียบกับเกณฑ์มีค่าต่ำ การให้คะแนนวิธีนี้จึงให้ประโยชน์ในทางปฏิบัติน้อย

การให้คะแนนวิธีต่าง ๆ ที่เสนอไปนั้น ทำให้การวิจัยที่เกี่ยวกับแบบทดสอบปรัมาติมีความน่าสนใจมากขึ้น วิธีการให้คะแนนอาจเปลี่ยนแปลงไปตามลักษณะของเนื้อหาวิชาที่วัด ลักษณะทางจิตวิทยา หรือความสัมพันธ์กับคะแนนจากการทดสอบแบบอื่น ๆ วิธีการให้คะแนนอาจเปลี่ยนแปลงไปตามความตรง และวัตถุประสงค์ของการวัดผล ว่าต้องการนำผลการวัดไปใช้ประโยชน์อะไร ซึ่งวิธีการให้คะแนนบางวิธีอาจดีที่สุดในการทำนายกับเกณฑ์ภายนอก หรือใช้ได้ดีในสถานการณ์ที่ต่างกัน นอกจากนี้วิธีการให้คะแนนยังสามารถเปลี่ยนแปลงไปตามลักษณะของการวิจัยได้ด้วย เช่น การวิจัยเชิงประจักษ์ การวิจัยในสถานการณ์จำลอง หรือ การวิจัยเชิงทฤษฎี ดังนั้นการที่จะกล่าวว่าวิธีการให้คะแนนวิธีใดจึงจะเหมาะสมกับสถานการณ์ใดนั้นจึงจำเป็นต้องอาศัยจากผลการศึกษาวิจัย

2.5 ข้อดีและข้อจำกัดของแบบทดสอบปรัมาตร

เมื่อเปรียบเทียบกับแบบทดสอบชนิด 2 ขั้น (Two-Stage Test) ซึ่งเป็นรูปแบบหนึ่งของการทดสอบแบบเพเลอร์แล้ว แบบทดสอบปรัมาตรมีศักยภาพในการวัดระดับความสามารถของผู้สอบได้กว้างกว่าและใช้จำนวนข้อสอบน้อยกว่า เพราะว่าแบบทดสอบชนิด 2 ขั้นต้องมีแบบทดสอบชุดแรกสำหรับใช้วัดเพื่อการแยกทางหรือจัดแบบทดสอบเพื่อการวัดผลชุดที่เหมาะสมกับผู้สอบแต่ละคน และจำเป็นต้องมีแบบทดสอบสำหรับการวัดผลหลายชุดตามความกว้างของระดับความสามารถที่ต้องการวัด ดังนั้นถ้าต้องการวัดระดับความสามารถของผู้สอบในช่วงที่กว้างเท่ากันแล้ว แบบทดสอบชนิด 2 ขั้นจะต้องใช้จำนวนข้อสอบมากกว่าแบบทดสอบปรัมาตร เพราะต้องสร้างแบบทดสอบเพื่อการวัดความสามารถหลายชุดตามจำนวนระดับความสามารถที่ต้องการวัด ส่วนแบบทดสอบปรัมาตรใช้แบบทดสอบเพียงชุดเดียวในการทดสอบซึ่งการตอบข้อสอบแต่ละข้อของผู้สอบนั้น จะใช้ผลการตอบเป็นข้อมูลสำหรับการพิจารณาในการแยกทางหรือจัดข้อสอบที่มีความยากเหมาะสมกับผู้สอบแต่ละคนและใช้เป็นข้อสอบเพื่อการวัดความสามารถไปพร้อม ๆ กันด้วย จึงถือได้ว่าเป็นข้อดีของแบบทดสอบปรัมาตรที่ใช้จำนวนข้อสอบในการวัดน้อยข้อ ทำให้ใช้เวลาในการดำเนินการสอบน้อย และยังสามารถวัดระดับความสามารถของผู้สอบในช่วงที่กว้างได้ดี ตัวอย่างเช่น แบบทดสอบปรัมาตรชนิด 10 ขั้นจะใช้ข้อสอบจำนวน 55 ข้อ ส่วนแบบทดสอบชนิด 2 ขั้นที่มีแบบทดสอบสำหรับการแยกทางชนิด 10 ข้อและแบบทดสอบเพื่อการวัดผล 4 ฉบับ ๆ ละ 20 ข้อ จะต้องใช้ข้อสอบทั้งหมดถึง 90 ข้อ ทำให้ต้องใช้กลุ่มข้อสอบที่ใหญ่มากในการสร้างแบบทดสอบชนิด 2 ขั้น

อย่างไรก็ตามแบบทดสอบปรัมาตรก็มีข้อจำกัดในการวัดที่สำคัญอยู่ 2 ประการคือ ประการแรกคะแนนที่ได้จากแบบทดสอบปรัมาตรจะมีผลกระทบจากการเดาตอบข้อสอบถูกเมื่อข้อสอบมีค่าการเดาสูงหรือข้อสอบข้อนั้นเปิดโอกาสให้มีการเดา และคะแนนที่ได้ยังอาจมีผลกระทบจากการตอบผิดที่เนื่องมาจากข้อสอบไม่เหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งถือว่าเป็นความคลาดเคลื่อนของการตอบข้อสอบ ดังนั้นถ้าผู้สอบที่มีระดับความสามารถสูงมากตอบข้อสอบข้อใดข้อหนึ่งผิดโดยไม่ตั้งใจก็จะถูกแยกทางให้ไปตอบข้อสอบที่ง่ายขึ้น ทำให้คะแนนที่ได้น้อยกว่าที่ควรจะเป็น หรือไม่มีโอกาสที่จะได้คะแนนสูงตรงกับระดับความสามารถที่เป็นจริงของตน ส่วนในกรณีที่ผู้สอบที่มีความสามารถต่ำมาก ๆ ที่ตอบข้อสอบถูกโดยการเดาแล้วก็จะทำให้ได้คะแนนที่สูงกว่าที่ควรจะเป็น หรือได้คะแนนไม่ตรงกับระดับความสามารถที่แท้จริงของผู้สอบเช่นกัน ผลที่เกิดขึ้นในทั้งสองลักษณะดังกล่าวนี้จึงเป็นข้อจำกัดของแบบทดสอบปรัมาตรที่จะส่งผลต่อความเชื่อมั่นในการวัดของแบบทดสอบปรัมาตร ดังนั้นถ้าต้องการวัดให้ได้ใกล้เคียงกับความเป็นจริงแล้วจึงจำเป็นต้องสร้างแบบทดสอบให้มีค่าความเชื่อมั่นสูง ๆ

ข้อจำกัดประการที่สอง จากลักษณะที่สำคัญประการหนึ่งของแบบทดสอบเทเลอร์คือ การให้ผู้สอบแต่ละคนได้ทำข้อสอบที่มีความยากเหมาะสมกับความสามารถของตน ดังนั้นเพื่อให้แบบทดสอบปิรามิดซึ่งเป็นรูปแบบหนึ่งของการทดสอบแบบเทเลอร์ เป็นไปตามลักษณะดังกล่าวจึงจำเป็นต้องจัดโครงสร้างของแบบทดสอบให้มีข้อสอบที่มีความยากที่ผู้สอบในแต่ละระดับความสามารถตอบถูกประมาณ 50 % ของข้อสอบทั้งหมด ซึ่งถือว่าเป็นข้อสอบที่ให้สารสนเทศสูงสุดในแต่ละระดับความสามารถของผู้สอบ แต่จากโครงสร้างของแบบทดสอบปิรามิดช่วงระดับความสามารถที่สูงและต่ำมาก ๆ ซึ่งอยู่ที่ฐานของโครงสร้างปิรามิดจะมีข้อสอบลดลงตามจำนวนชั้นของแบบทดสอบ นั่นคือที่ระดับความสามารถที่ปลายทั้งสองข้างของฐานจะมีข้อสอบจำนวนน้อยข้อ ถ้าต้องการให้มีจำนวนข้อมากขึ้นก็ต้องขยายช่วงระดับความสามารถที่จะวัดให้กว้างขึ้น ซึ่งก็จะสัมพันธ์กับจำนวนชั้นของแบบทดสอบที่ต้องมีจำนวนชั้นมากขึ้นด้วย จากลักษณะดังกล่าวนี้ทำให้ต้องใช้กลุ่มข้อสอบที่มีขนาดใหญ่ในการสร้างแบบทดสอบปิรามิดซึ่งในทางปฏิบัติแล้วถ้าสร้างให้มีจำนวนชั้นมากขึ้นเท่าไร ผู้สอบก็ต้องทำข้อสอบมากขึ้นและใช้เวลาในการสอบมากขึ้นด้วย อันก่อให้เกิดปัญหาในการทดสอบที่ไม่ต้องการให้เกิดขึ้น ลักษณะของแบบทดสอบปิรามิดที่ศึกษากันส่วนใหญ่ จึงพยายามลดจำนวนชั้นลงให้เหมาะสมที่สุดที่จะดำเนินการสอบและมีความเชื่อมั่นในการประมาณความสามารถของผู้สอบได้สูงที่สุด

2.6 ลักษณะของการวิจัยที่เกี่ยวกับแบบทดสอบเทเลอร์

Weiss และ Betz (1973) ได้สรุปลักษณะของการวิจัยที่เกี่ยวกับการทดสอบแบบทดสอบเทเลอร์ไว้ 3 ลักษณะดังนี้

2.6.1 การศึกษาเชิงประจักษ์ (Empirical Study)

เป็นการศึกษาที่ใช้ผู้สอบและกลุ่มของข้อสอบจริง กำหนดวิธีการดำเนินการสอบไว้อย่างเฉพาะและชัดเจน ซึ่งวิธีดำเนินการสอบสามารถใช้ได้ทั้งการให้ผู้สอบเขียนตอบหรือให้ผู้สอบสอบกับเครื่องคอมพิวเตอร์ก็ได้

ข้อสรุปที่ได้จากการศึกษาเชิงประจักษ์นี้จำเป็นต้องพิจารณาด้วยความระมัดระวัง เพราะการสอบในสถานการณ์จริงนั้นมีปัจจัยที่เกี่ยวข้องมากมาย เช่น ข้อจำกัดในด้านคุณลักษณะของผู้สอบ ความเหมาะสมของกลุ่มข้อสอบที่ใช้ในการศึกษา รวมถึงปฏิสัมพันธ์ (Interaction) ของผู้สอบกับวิธีดำเนินการทดสอบ

2.6.2 การศึกษาในสถานการณ์จำลอง (Simulation Study)

การศึกษาในสถานการณ์จำลองเท่าที่ศึกษากันมีอยู่ 2 ลักษณะ คือ

ลักษณะที่ 1 การใช้ข้อมูลจริง (Real Data) เป็นการศึกษาที่ใช้ข้อมูลจากผลการทดสอบที่ใช้แบบทดสอบดั้งเดิม ซึ่งข้อมูลเหล่านี้จะประกอบด้วย ผลการตอบข้อสอบแต่ละข้อของผู้สอบแต่ละคน คำนวณความยากและค่าอำนาจจำแนกของข้อสอบ การจำลองการทดสอบแบบเทเลอร์กับข้อมูลที่ได้จากแบบทดสอบดั้งเดิมและกลุ่มผู้สอบนี้ อาจใช้วิธีการใดวิธีการหนึ่งหรือหลายวิธีการของการทดสอบแบบเทเลอร์ โดยนำกระดาษคำตอบของผู้สอบแต่ละคนนั้นมาพิจารณาตรวจให้คะแนนใหม่ โดยถือเสมือนว่ากระดาษคำตอบเหล่านั้นได้รับการดำเนินการทดสอบแบบเทเลอร์

ลักษณะที่ 2 การศึกษาในสถานการณ์จำลองแบบมอนติคาร์โล (Monte Carlo Simulation Study) เป็นการศึกษาที่สร้างกลุ่มข้อสอบโดยเน้นที่ค่าความยากและค่าอำนาจจำแนกของข้อสอบและความสามารถของผู้สอบให้เป็นไปตามทฤษฎีอย่างเข้มงวด แล้วใช้ทฤษฎีโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve Theory, ICC) และตัวเลขสุ่มที่ได้จากเครื่องคอมพิวเตอร์ (Computer Generated Random Numbers) ตลอดจนแมทริกซ์ของคำตอบ คำนวณรวมและระดับความสามารถของผู้สอบมาสร้างกลุ่มข้อสอบ แล้วใช้วิธีการทดสอบแบบเทเลอร์อย่างใดอย่างหนึ่งกับผู้สอบที่ต้องการศึกษา ซึ่งวิธีการนี้จะไม่ส่งผลกระทบต่อผลการทดสอบจริงในตัวผู้สอบนั้น

2.6.3 การศึกษาเชิงทฤษฎี (Theoretical Study)

เป็นการศึกษาที่ตั้งอยู่บนพื้นฐานของทฤษฎีโค้งคุณลักษณะข้อสอบ ซึ่งข้อสอบต้องมีค่าอำนาจจำแนกที่คงที่ การศึกษาในลักษณะนี้เป็นการศึกษาเพื่อตรวจสอบตลอดช่วงความสามารถที่ต่อเนื่องกันของผู้สอบมากกว่าที่จะศึกษาในช่วงความสามารถช่วงใดช่วงหนึ่ง สำหรับกลุ่มข้อสอบในเชิงทฤษฎีที่จะให้ผลการวัดในระดับสูงสุดโดยไม่มีข้อจำกัดเกี่ยวกับจำนวนผู้สอบในสถานการณ์จำลอง ผลการศึกษาที่ได้จากวิธีนี้จะอยู่ในลักษณะของการหาคำตอบของชุดสมการทางคณิตศาสตร์ที่เสนอโดย Lord (อ้างถึงใน Weiss, 1973), Lord และ Novick (1968) และคนอื่น ๆ ซึ่งเป็นการสร้างโค้งการกระจายของคะแนนผลการสอบจากโค้งลักษณะข้อสอบที่สมมุติขึ้น สำหรับผู้สอบที่มีความสามารถผันแปรไปภายใต้วิธีการทดสอบเฉพาะ การศึกษาในลักษณะนี้นับว่ามีประโยชน์มากเพราะทำได้รวดเร็วในการประเมินวิธีการทดสอบแบบต่าง ๆ ของแบบทดสอบเทเลอร์ ตลอดจนค่าอำนาจจำแนกของข้อสอบ อย่างไรก็ตามการศึกษาในเชิงทฤษฎีนี้ไม่สามารถศึกษาถึงผลกระทบที่เกิดจากการทดสอบแบบเทเลอร์ในสถานการณ์จริงได้

2.7 เกณฑ์ที่ใช้ในการประเมินแบบทดสอบเทเลอร์

Symson (1975) ได้ให้ข้อคิดเห็นเกี่ยวกับการประเมินกระบวนการทดสอบไว้ว่า ไม่ควรจะกระทำในลักษณะรวม ๆ แต่น่าจะประเมินผลกระทบแต่ละอย่างในองค์ประกอบใดองค์ประกอบหนึ่ง

การประเมินองค์ประกอบเหล่านี้จะเป็นการเปรียบเทียบกระบวนการทดสอบ 2 กระบวนการที่มีความแตกต่างกันเพียงองค์ประกอบเดียวเท่านั้น ถ้านับว่ากระบวนการทดสอบหนึ่งดีกว่าอีกกระบวนการหนึ่งแล้ว ก็จะบอกได้ว่าองค์ประกอบที่แตกต่างกันนั้นเป็นปัจจัยที่ทำให้กระบวนการทดสอบแตกต่างกัน แต่โดยสภาพทั่ว ๆ ไปแล้วการประเมินกระบวนการทดสอบในด้านขององค์ประกอบดังกล่าวข้างต้นมักจะเป็นไปได้ยาก เนื่องจากข้อจำกัดต่าง ๆ หลายประการ เช่น วิธีการเลือกข้อสอบบางแบบไม่อาจจะใช้วิธีดำเนินการสอบโดยวิธีอื่นได้นอกจากจะต้องใช้เครื่องคอมพิวเตอร์เท่านั้น หรือการให้คะแนนบางวิธีเหมาะสมกับวิธีการเลือกข้อสอบให้สอบในลักษณะเดียวกันตลอดการดำเนินการสอบเป็นต้น ดังนั้นการศึกษาเพื่อประเมินแบบทดสอบจึงจำเป็นต้องเปรียบเทียบกระบวนการทดสอบซึ่งแตกต่างกันในสององค์ประกอบหรือมากกว่า

Symson (1975) ได้จำแนกประเภทเกณฑ์ที่ใช้ในการประเมินกระบวนการทดสอบทั่ว ๆ ไป ซึ่งไม่ได้จำกัดเฉพาะกับแบบทดสอบเทเลอร์ไว้ ดังนี้

2.7.1 เกณฑ์เชิงความเป็นไปได้ (Possibility Criteria)

เป็นการพิจารณาถึงคุณลักษณะของคะแนนที่ได้จากกระบวนการทดสอบว่า ได้แสดงให้เห็นถึงลักษณะที่ต้องการวัดเพียงใด เกณฑ์เชิงความเป็นไปได้ประกอบด้วยความเชื่อมั่น ซึ่งอาจจะหาค่าสัมประสิทธิ์ความคงที่ (Stability Coefficient) สัมประสิทธิ์ความคงที่ภายใน (Internal Consistency Coefficient) และ ค่าสหสัมพันธ์จากแบบทดสอบคู่ขนาน (Alternate Form Correlation) ค่าสหสัมพันธ์กับแบบทดสอบอื่น ๆ ตลอดจนค่าสหสัมพันธ์กับตัวแปรอื่น ๆ ที่นอกเหนือจากคะแนนที่ได้จากการทดสอบ รวมทั้งลักษณะการกระจายของคะแนนผลการสอบในกลุ่มผู้สอบกลุ่มใดกลุ่มหนึ่ง ความแตกต่างระหว่างการกระจายของคะแนนที่ได้จากผู้สอบต่างกลุ่มกัน และการใช้วิธีการทางสถิติ หรือการานเพื่อการประเมินความสอดคล้อง (Goodness of Fit) กับรูปแบบตามทฤษฎีที่มีอยู่เดิม

สำหรับการทดสอบแบบเทเลอร์ซึ่งผู้สอบแต่ละคนจะได้ตอบข้อสอบที่แตกต่างกัน ดังนั้นจึงไม่สามารถคำนวณค่าสัมประสิทธิ์ความคงที่ภายในได้ การประมาณค่าความเชื่อมั่นของคะแนนผลการสอบแบบทดสอบเทเลอร์จึงจำเป็นต้องใช้ค่าสหสัมพันธ์จากแบบทดสอบคู่ขนาน หรือค่าสัมประสิทธิ์ความคงที่ ซึ่งจะเหมาะสมกว่า

2.7.2 เกณฑ์เชิงทฤษฎี (Theoretical Criteria)

การเปรียบเทียบโดยเกณฑ์นี้จะเหมาะสมกับการตรวจสอบว่า กระบวนการทดสอบหนึ่งดีกว่าอีกกระบวนการหนึ่งหรือไม่ เมื่อทราบมาแล้วว่ากระบวนการทดสอบทั้งสองนั้น น่าจะมีความตรงเชิง

โครงสร้างเท่ากัน เกณฑ์เชิงทฤษฎีประกอบด้วย การกระจายของค่าประมาณคุณลักษณะแฝง โค้งของสารสนเทศ (Information Curve) โค้งประสิทธิภาพสัมพัทธ์ (Relative Efficiency Curve) โค้งความลำเอียง (Bias Curve) โค้งความคลาดเคลื่อนมาตรฐานของการวัด ตลอดจนการคงสภาพ (Robustness) ในลักษณะต่าง ๆ ซึ่งได้แก่ความคลาดเคลื่อนในการประมาณค่า นารามิเตอร์ของข้อสอบหรือโค้งคุณลักษณะของข้อสอบที่ได้จากกระบวนการทดสอบที่แตกต่างกัน เป็นต้น

การศึกษาโดยใช้เกณฑ์เชิงทฤษฎีนี้ ต้องมีการกำหนดรูปแบบของโค้งคุณลักษณะของข้อสอบที่ต้องการขึ้นและต้องทราบค่านารามิเตอร์ของข้อสอบตั้งแต่แรก แล้วใช้วิธีการทางคณิตศาสตร์หรือคอมพิวเตอร์ในการคำนวณ ดังนั้นเกณฑ์เชิงทฤษฎีนี้จึงไม่สามารถนำมาใช้ในการศึกษาในสถานการณ์การทดสอบจริงได้ ในรายเราไม่ทราบสถานที่แท้จริงของผู้สอบแต่ละคนบนมาตราคุณลักษณะแฝง

2.7.3 เกณฑ์เชิงจิต-สังคม (Psycho-Social Criteria)

การศึกษาโดยใช้เกณฑ์นี้ จะเป็นการศึกษาถึงผลกระทบที่เกิดจากกระบวนการทดสอบที่มีต่อคุณลักษณะทางจิตวิทยาในตัวผู้สอบ เช่น การจูงใจ (Motivation) และความวิตกกังวล (Anxiety) ในการทดสอบ ตลอดจนผลกระทบที่มีต่อสังคม เช่น การทดสอบแบบเทเลอร์นั้นผู้สอบแต่ละคนจะได้ตอบข้อสอบที่แตกต่างกันตามระดับความสามารถของแต่ละคน ซึ่งอาจทำให้บุคคลทั่วไปเกิดปัญหาขึ้นว่าจะเปรียบเทียบความสามารถระหว่างผู้สอบได้อย่างไร เมื่อไม่ได้ตอบข้อสอบข้อที่เหมือนกัน เป็นต้น

2.7.4 เกณฑ์เชิงค่าใช้จ่าย (Cost Criteria)

การใช้เกณฑ์นี้ในการศึกษา เป็นการศึกษาเพื่อประเมินค่าใช้จ่ายในระบบการสร้างแบบทดสอบ การสื่อสารกับผู้สอบด้วยเครื่องคอมพิวเตอร์ รวมถึงมูลค่าวัสดุต่าง ๆ ที่ต้องใช้ในกระบวนการทดสอบ ค่าใช้จ่ายในการสร้างและการหาเกณฑ์ปกติของกลุ่มข้อสอบ ที่มีขนาดตามที่กระบวนการทดสอบนั้นต้องการ การประเมินค่าใช้จ่ายนี้จะต้องประเมินควบคู่ไปกับผลที่ได้จากกระบวนการทดสอบที่สร้างขึ้นด้วยว่า การลงทุนนั้นเมื่อเทียบกับจุดอ่อนต่าง ๆ ที่อาจเกิดขึ้นแล้วเป็นการคุ้มค่าหรือเหมาะสมหรือไม่ เช่น กระบวนการทดสอบนั้นผู้สอบมีโอกาสเกิดข้อผิดพลาดในการตอบได้มากน้อยเพียงไร ข้อผิดพลาดในการให้คะแนน การเสียเวลาของผู้สอบต้องติดต่อกับระบบคอมพิวเตอร์ เวลาที่ต้องใช้ในการดำเนินการสอบ รวมถึงการให้คะแนนและการตีความหมายของคะแนน

Weiss และ Betz (1973) ได้กล่าวถึงการประเมินแบบทดสอบเทเลอร์ที่สัมพันธ์กับลักษณะของการวิจัยที่เกี่ยวกับการทดสอบแบบเทเลอร์ไว้ ซึ่งมีสาระสำคัญดังนี้

1. การประเมินประสิทธิภาพของแบบทดสอบเทเลอร์ ส่วนใหญ่จะเป็นการเปรียบเทียบกับแบบ

ทดสอบดั้งเดิม โดยใช้เกณฑ์เชิงปฏิบัติซึ่งประกอบด้วยเวลาที่ใช้ในการดำเนินการสอบ ค่าใช้จ่ายในการสร้างและดำเนินการสอบ รวมถึงความยุ่งยากซับซ้อนในการดำเนินการสอบ

2. การศึกษาบางครั้งใช้ค่าสหสัมพันธ์ของคะแนนที่ได้จากแบบทดสอบเทเลอร์กับคะแนนจากแบบทดสอบดั้งเดิมเป็นเกณฑ์ในการประเมิน การศึกษาในลักษณะนี้โดยปกติจะให้กลุ่มผู้สอบทำแบบทดสอบทั้งแบบทดสอบดั้งเดิมและแบบทดสอบเทเลอร์ ซึ่งอาจจะให้สอบในสถานการณ์จริงหรือสถานการณ์จำลองก็ได้ แล้วนำคะแนนที่ได้มาศึกษา ซึ่งจุดมุ่งหมายของการศึกษาในลักษณะนี้ส่วนใหญ่จะศึกษาว่าแบบทดสอบเทเลอร์รูปแบบใดที่สามารถประมาณคะแนนรวมในแบบทดสอบดั้งเดิมได้ดีที่สุด

3. การศึกษาเชิงทฤษฎี จะใช้ค่าสหสัมพันธ์ของคะแนนที่ได้จากแบบทดสอบกับความสามารถของผู้สอบเป็นเกณฑ์ในการประเมิน เช่น การศึกษาในสถานการณ์จำลองแบบมอนติคาร์โลใช้ค่าสหสัมพันธ์ระหว่างความสามารถของผู้สอบกับค่าประมาณความสามารถที่ประมาณจากรูปแบบของคำตอบของผู้สอบแต่ละคนในสถานการณ์ที่สมมติขึ้น การศึกษาในลักษณะนี้ผู้ศึกษามักสนใจในเรื่องความตรงของรูปแบบการทดสอบแบบเทเลอร์ว่าคะแนนที่ได้สามารถใช้ทำนายความสามารถของผู้สอบที่สมมติขึ้นได้มากน้อยเพียงใด

นอกจากนี้ในการศึกษาบางครั้งอาจใช้น้ำหนักของสารสนเทศ (Information Function) เป็นเกณฑ์ในการประเมินแบบทดสอบเทเลอร์ได้ (Birnbau, 1968) เกณฑ์นี้จะเหมาะสมกับการศึกษาที่ต้องการเปรียบเทียบรูปแบบการทดสอบตั้งแต่สองรูปแบบขึ้นไป โดยการเปรียบเทียบในเทอมของสารสนเทศที่ได้จากการทดสอบแต่ละรูปแบบที่ระดับความสามารถต่าง ๆ ตลอดช่วงความสามารถของผู้สอบ (Underlying Ability Continuum) การตีความหมายจากฟังก์ชันสารสนเทศจะมีอยู่ 3 ลักษณะ คือ ค่าของสารสนเทศจะบอกให้ทราบถึงประสิทธิภาพสัมพัทธ์ของจำนวนข้อสอบในแบบทดสอบต่าง ๆ ที่จะส่งผลให้แบบทดสอบเหล่านั้นมีระดับความเชื่อมั่นเท่ากัน ตลอดจนสัมพันธภาพของความคลาดเคลื่อนมาตรฐานของการวัด (Relative Error of Measurement) และสัมพันธภาพของความสามารถ (Relative Capacity) ของการทดสอบแบบเทเลอร์รูปแบบต่าง ๆ ที่จะจำแนกผู้สอบตามระดับความสามารถ

2.8 งานวิจัยที่เกี่ยวข้องกับแบบทดสอบปิรามิด

Larkin และ Weiss (1974) ได้สรุปไว้ว่า การศึกษาวิจัยที่เกี่ยวกับแบบทดสอบปิรามิดนั้นแบ่งออกเป็น 3 ลักษณะ คือ การศึกษาเชิงประจักษ์นิยม (Empirical Studies) การศึกษาในสถานการณ์จำลอง (Simulation Studies) และ การศึกษาเชิงทฤษฎี (Theoretical Studies) การศึกษาวิจัยใน 3 ลักษณะดังกล่าวมีการศึกษากันอย่างกว้างขวาง ซึ่งสามารถสรุป

สาระสำคัญได้ดังต่อไปนี้ (Larkin และ Weiss, 1974)

2.8.1 การศึกษาเชิงประจักษ์นิยม (Empirical Studies)

Krathwohl และ Huyser (1956) ได้ศึกษาแบบทดสอบปิรามิดชนิด 8 ชั้นที่มีข้อสอบชั้นละ 1 ข้อ และแบบทดสอบปิรามิดชนิด 4 ชั้นที่มีข้อสอบชั้นละ 2 ข้อ แบบทดสอบปิรามิดทั้ง 2 ฉบับดำเนินการสอบโดยการเขียนตอบ (Paper and Pencil) เปรียบเทียบกับแบบทดสอบดั้งเดิมที่มีข้อสอบจำนวน 60 ข้อ ข้อสอบในแบบทดสอบปิรามิดและแบบทดสอบดั้งเดิมมีลักษณะเหมือนกันคือวัดในเนื้อหาเดียวกัน จากการศึกษาแก่นักศึกษาในระดับวิทยาลัย 100 คน พบว่าแบบทดสอบปิรามิดต้องใช้คำชี้แจงประกอบมาก แต่เวลาที่ใช้ในการสอบจะน้อยกว่าแบบทดสอบดั้งเดิม ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบปิรามิดชนิด 8 ชั้นกับแบบทดสอบดั้งเดิมมีค่าเท่ากับ .78 และค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบปิรามิดชนิด 4 ชั้นกับแบบทดสอบดั้งเดิมมีค่าเท่ากับ .68

Bayroff, Thomas และ Anderson (อ้างถึงใน Larkin and Weiss, 1974) ได้ศึกษาในลักษณะใกล้เคียงกับวิธีการของ Krathwohl โดยสร้างแบบทดสอบปิรามิดชนิด 6 ชั้นซึ่งมีค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันลดลงเรื่อย ๆ ข้อสอบเป็นแบบเลือกตอบชนิด 4 ตัวเลือก ได้กำหนดเกณฑ์การเลือกข้อสอบข้อที่ 2 โดยอาศัยผลการตอบข้อสอบข้อที่ 1 เป็นหลักในการพิจารณา ดังนี้ ถ้าตอบข้อสอบข้อแรกถูกข้อสอบข้อที่ 2 ที่จะได้ทำจะมีความยากมากขึ้น ถ้าตอบผิดโดยตอบตัวเลือกที่มีคนเลือกตอบน้อย ๆ กัน 2 ตัวเลือกจะได้ทำข้อสอบในชั้นที่ 2 ข้อที่มีความยากเท่าเดิม ถ้าตอบผิดโดยเลือกตอบตัวเลือกที่มีคนเลือกตอบน้อยที่สุด จะให้ทำข้อสอบในชั้นที่สองข้อที่มีความยากลดลง ส่วนในชั้นต่อ ๆ ไปใช้รูปแบบการเพิ่ม 1/ลด 1 Seeley, Morton และ Anderson (1962) ได้นำแบบทดสอบดังกล่าวไปทดลองสอบกับกลุ่มตัวอย่างจำนวน 327 คน แล้วมาหาค่าสหสัมพันธ์กับแบบทดสอบชนิดดั้งเดิมได้เท่ากับ .63 และพบว่าคะแนนสอบด้วยแบบทดสอบปิรามิดมีความแม่นยำ คือส่วนใหญ่จะได้คะแนนสูง ผู้วิจัยให้ข้อคิดเห็นว่าเป็นเพราะผู้ที่ได้คะแนนต่ำส่วนใหญ่ไม่ได้ทำข้อสอบตามคำชี้แจง ดังนั้นคะแนนของผู้ที่ได้คะแนนต่ำจึงไม่สามารถนำมาวิเคราะห์ได้

Wood (1969) ได้ทดลองใช้แบบทดสอบปิรามิดชนิด 4, 5 และ 6 ชั้นกับตัวอย่างจำนวน 91 คน แบบทดสอบดังกล่าวใช้ขนาดของชั้นแบบคงที่ คือ $p = .05$ ข้อสอบข้อแรกมีความยากปานกลาง คือ $p = .50$ ใช้เทคนิคการแยกทางแบบเพิ่ม 1/ลด 1 และให้คะแนนโดยใช้จำนวนข้อที่ตอบถูก ได้ศึกษาความเที่ยงตรงของแบบทดสอบโดยหาค่าสหสัมพันธ์ระหว่างคะแนนที่สอบได้ กับระดับผลการเรียนที่สอบโดยแบบทดสอบดั้งเดิมจำนวน 46 ข้อ ซึ่งผลได้ค่าสหสัมพันธ์เท่ากับ .35 ทั้ง 3 ฉบับ เมื่อรวม

คะแนนทั้ง 3 ฉบับแล้วมาหาค่าสหสัมพันธ์กับระดับผลการเรียนพบว่ามีค่าเพิ่มขึ้นเป็น .51 เมื่อเปรียบเทียบกับค่าสหสัมพันธ์ระหว่างคะแนนจากแบบทดสอบดั้งเดิมกับระดับผลการเรียนซึ่งมีค่าเท่ากับ 0.68 แล้ว Wood สรุปผลการวิจัยว่า แบบทดสอบดั้งเดิมมีประสิทธิภาพดีเท่ากับแบบทดสอบปิรามิดที่มีจำนวนข้อสอบเท่ากัน

ระยะต่อมา การวิจัยเชิงประจักษ์เกี่ยวกับแบบทดสอบปิรามิดที่ควบคุมการดำเนินการสอบด้วยโปรแกรมคอมพิวเตอร์ เริ่มเป็นที่สนใจกันมากขึ้น ซึ่งมีงานวิจัยที่สำคัญ ๆ สรุปได้ดังนี้

Bayroff และ Seeley (อ้างถึงใน Larkin and Weiss, 1974) ได้ทดลองใช้แบบทดสอบปิรามิดชนิด 8 ชั้น กับตัวอย่างจำนวน 102 คน ขนาดของชั้นของแบบทดสอบเท่ากับ .05 ($p=.05$) ข้อสอบในชั้นสุดท้ายคือชั้นที่ 8 มีความยากกระจายตั้งแต่ $p=.95$ ถึง $p=.20$ การให้คะแนนขึ้นอยู่กับผลการตอบข้อสอบข้อสุดท้าย ว่าถูกหรือผิดซึ่งช่วงคะแนนที่จะได้เท่ากับ 17 คะแนน กลุ่มตัวอย่างสอบแบบทดสอบปิรามิดพร้อมกับแบบทดสอบดั้งเดิมอีก 2 ฉบับ เป็นข้อสอบทางด้านจำนวนจำนวน 40 ข้อ และเป็นข้อสอบทางด้านภาษาจำนวน 50 ข้อ ผลการวิจัยพบว่าค่าสหสัมพันธ์ระหว่างแบบทดสอบปิรามิดกับแบบทดสอบดั้งเดิม 2 ฉบับมีค่าเท่ากับ .83 และ .79 ตามลำดับซึ่งเมื่อเปรียบเทียบกับค่าสหสัมพันธ์ระหว่างคะแนนข้อสอบแบบดั้งเดิมจำนวน 8 ข้อกับคะแนนผลการสอบจากแบบทดสอบดั้งเดิมทั้ง 2 ฉบับมีค่าเท่ากับ .75 และ .67 ตามลำดับ จึงสรุปได้ว่าแบบทดสอบปิรามิดมีค่าสหสัมพันธ์กับแบบทดสอบดั้งเดิมสูงกว่าแบบทดสอบดั้งเดิมที่มีจำนวนข้อสอบเท่ากันกับแบบทดสอบดั้งเดิมฉบับเดียวกัน นอกจากนี้จากการใช้สูตรของ Spearman-Brown พบว่าจะต้องใช้จำนวนข้อสอบในแบบทดสอบดั้งเดิมเป็น 2 เท่าของจำนวนข้อสอบในแบบทดสอบปิรามิด เพื่อให้ได้ค่าสหสัมพันธ์กับแบบทดสอบที่ใช้เป็นเกณฑ์เท่ากัน

Hansen (อ้างถึงใน Larkin and Weiss, 1974) ได้ทดลองดำเนินการสอบด้วยคอมพิวเตอร์กับแบบทดสอบปิรามิดจำนวน 5 ฉบับกับนักศึกษาระดับวิทยาลัยปีที่ 1 จำนวน 56 คน ข้อสอบทั้ง 5 ฉบับ มีทั้งชนิด 3 และ 5 ชั้น ซึ่งนักศึกษาแต่ละคนจะได้ทำข้อสอบจำนวน 17 ข้อ ขนาดของชั้นเท่ากับ .10 ($p=.10$) การให้คะแนนใช้วิธีที่แตกต่างกัน 4 วิธี ก่อนให้ทำแบบทดสอบปิรามิด 1 สัปดาห์ให้กลุ่มตัวอย่างสอบแบบทดสอบดั้งเดิมในเนื้อหาเดียวกัน ผู้วิจัยได้ศึกษาความสัมพันธ์ระหว่างคะแนนที่ได้จากแบบทดสอบปิรามิดกับคะแนนจากแบบทดสอบดั้งเดิม คะแนนจากแบบทดสอบผลสัมฤทธิ์อีก 1 ฉบับ และระดับผลการเรียนปลายภาคการศึกษา ผลการศึกษพบว่าแบบทดสอบดั้งเดิม แม้จะปรับให้มีความยาวเท่ากับแบบทดสอบปิรามิดแล้วก็ตาม จะมีค่าความเชื่อมั่นชนิดความคงที่ภายใน (Internal Consistency) ต่ำกว่าแบบทดสอบปิรามิดทั้ง 5 ฉบับ คะแนนผลการสอบจากแบบ

ทดสอบปริมาตรมีการกระจายเป็นปกติมากกว่าจากแบบทดสอบดั้งเดิม ซึ่งการกระจายของคะแนนมีความเบ้ทางลบ นอกจากนี้ผลการวิจัยยังพบว่า โดยเฉลี่ยแล้วผู้สอบจะใช้เวลาในการสอบแบบทดสอบปริมาตรทั้ง 5 ฉบับน้อยกว่าแบบทดสอบดั้งเดิมประมาณ 5 นาที การให้คะแนนแบบทดสอบปริมาตร 2 วิธีมีค่าสหสัมพันธ์กับคะแนนแบบทดสอบผลสัมฤทธิ์ และระดับผลการเรียนปลายภาคสูงกว่าค่าสหสัมพันธ์ระหว่างแบบทดสอบดั้งเดิม กับคะแนนแบบทดสอบผลสัมฤทธิ์และระดับผลการเรียน

Bryson (อ้างถึงใน Larkin and Weiss, 1974) ได้ศึกษาเปรียบเทียบค่าสหสัมพันธ์ระหว่างแบบทดสอบปริมาตรชนิด 5 ชั้นจำนวน 2 ฉบับและแบบทดสอบดั้งเดิมที่มีข้อสอบจำนวน 5 ข้อจำนวน 2 ฉบับ กับคะแนนผลการสอบจากแบบทดสอบฉบับหนึ่งที่มีข้อสอบ 100 ข้อ แบบทดสอบดั้งเดิมดำเนินการสอบโดยการเขียนตอบ ส่วนแบบทดสอบปริมาตรดำเนินการสอบด้วยคอมพิวเตอร์ แบบทดสอบปริมาตรฉบับหนึ่งการเลือกข้อสอบให้ผู้สอบใช้ค่าอำนาจจำแนกเป็นเกณฑ์ ส่วนอีกฉบับหนึ่งนั้นใช้การพยากรณ์คะแนนสูงสุดเป็นเกณฑ์ แบบทดสอบปริมาตรทั้งสองฉบับจึงมีขนาดของชั้นแปรผัน แบบทดสอบปริมาตรทั้งสองฉบับนำไปสอบกลุ่มตัวอย่างทั้งสองกลุ่มจำนวน 263 คนและแบบทดสอบดั้งเดิมนำไปสอบกลุ่มตัวอย่างอีกกลุ่มหนึ่งจำนวน 250 คน ผลการวิจัยพบว่าแบบทดสอบดั้งเดิมฉบับหนึ่งมีค่าสหสัมพันธ์กับคะแนนรวมสูงกว่าแบบทดสอบปริมาตรทั้งสองฉบับ และแบบทดสอบปริมาตรฉบับหนึ่งมีค่าสหสัมพันธ์กับคะแนนรวมต่ำกว่าแบบทดสอบดั้งเดิมทั้งสองฉบับ

Larkin และ Weiss (1974) ได้ศึกษาแบบทดสอบปริมาตร 3 ฉบับกับแบบทดสอบดั้งเดิม 1 ฉบับซึ่งมีข้อสอบที่รับค่าการเดาแล้ว 40 ข้อ มีค่าความยากตั้งแต่ $-.957$ ถึง 1.156 แบบทดสอบปริมาตรแต่ละฉบับเป็นแบบสมมาตรชนิด 15 ชั้น มีค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันคงที่ ใช้กฎการแยกทางแบบ เนิม1/ลด1 แต่ละฉบับจะมีข้อสอบ 120 ข้อ โดยจัดระดับความยากของข้อสอบออกเป็น 29 ระดับ ช่วงระดับความยากกระจายตั้งแต่ -2.86 ถึง 2.95 ให้คะแนนแบบทดสอบปริมาตร 6 วิธี คือ ให้ตามจำนวนข้อที่ตอบถูก ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก ให้ตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบ ให้ตามค่าความยากของข้อสอบข้อที่ถัดจากข้อสุดท้ายที่ได้ตอบ และให้คะแนนข้อสอบทุกข้อตามวิธีการของ Hansen ข้อสอบเป็นแบบเลือกตอบชนิด 5 ตัวเลือกเป็นคำถามที่เกี่ยวกับเนื้อหาด้านคำศัพท์ (Vocabulary) ที่อยู่ในระดับต่ำกว่าระดับวิทยาลัย มีกลุ่มข้อสอบ (Item Pool) จำนวน 369 ข้อ ใช้ดำเนินการสอบด้วยคอมพิวเตอร์ โดยแบ่งการทดสอบเป็น 2 ระยะ ดังต่อไปนี้

การทดสอบระยะที่ 1 แบ่งกลุ่มตัวอย่างออกเป็น 3 กลุ่ม กลุ่มที่ 1 และ 2 จำนวน 125 คน เท่ากันให้ทำแบบทดสอบปริมาตรฉบับที่ 1 และแบบทดสอบดั้งเดิมเหมือนกัน กลุ่มที่ 3 จำนวน 142 คน ให้ทำแบบทดสอบปริมาตรฉบับที่ 3 และแบบทดสอบปรับระดับขึ้น

การทดสอบระยะที่ 2 ทั้งช่วงห่างจากการทดสอบระยะที่ 1 7 สัปดาห์ กลุ่มตัวอย่างกลุ่มที่ 1 และกลุ่มที่ 2 กลุ่มละจำนวน 125 คน ได้มาโดยการสุ่มจากกลุ่มที่ 1 และกลุ่มที่ 2 ของระยะที่ 1 กลุ่มที่ 1 ให้ทำแบบทดสอบปิรามิดฉบับที่ 1 และแบบทดสอบดั้งเดิม กลุ่มที่ 2 ให้ทำแบบทดสอบปิรามิดฉบับที่ 2 และแบบทดสอบดั้งเดิม ส่วนกลุ่มที่ 3 สุ่มมาจำนวน 138 คนให้ทำแบบทดสอบปิรามิดฉบับที่ 3 และแบบทดสอบชนิดสองชั้น

ผลการศึกษาพบว่าลำดับที่ของการทำแบบทดสอบมีผลต่อคะแนนผลการสอบทั้งแบบทดสอบปิรามิด และแบบทดสอบดั้งเดิมอย่างไม่มีนัยสำคัญทางสถิติ คือถ้าสอบแบบทดสอบปิรามิดหลังจากการสอบแบบทดสอบดั้งเดิมคะแนนที่ได้จะมีแนวโน้มลดลง ลักษณะการกระจายของคะแนนผลการสอบแบบทดสอบปิรามิดฉบับที่ 1 และฉบับที่ 2 มีลักษณะใกล้เคียงกับการกระจายแบบปกติ ส่วนแบบทดสอบปิรามิดฉบับที่ 3 มีลักษณะการกระจายแบบเบ้ไปทางบวกอย่างมีนัยสำคัญทางสถิติ การศึกษาผลของวิธีการให้คะแนนกับลักษณะการกระจายของคะแนน พบว่าการให้คะแนนทั้ง 6 วิธีส่งผลให้การกระจายของคะแนนในแบบทดสอบปิรามิดฉบับที่ 1 และฉบับที่ 2 มีความโค้งต่ำกว่าการกระจายแบบปกติอย่างมีนัยสำคัญทางสถิติ ส่วนในแบบทดสอบปิรามิดฉบับที่ 3 การให้คะแนนทั้ง 6 วิธีไม่ส่งผลให้การกระจายของคะแนนมีความโค้งต่ำกว่าการกระจายแบบปกติ ส่วนคะแนนที่ได้จากแบบทดสอบดั้งเดิมมีการกระจายของคะแนนเป็นแบบปกติอย่างมีนัยสำคัญ และจากการใช้ร้อยละของช่วงคะแนนในการประมาณระดับความสามารถของผู้สอบ พบว่า การให้คะแนนโดยวิธีให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ทำสามารถประมาณระดับความสามารถของผู้สอบได้ในช่วงกว้างเท่ากับแบบทดสอบดั้งเดิม และการให้คะแนนข้อสอบทุกข้อตามวิธีการของ Hansen สามารถประมาณระดับความสามารถของผู้สอบได้ในช่วงกว้างมากกว่าแบบทดสอบดั้งเดิมเล็กน้อย และการให้คะแนนทั้งสองวิธีนี้มีค่าสหสัมพันธ์สูงเท่ากับ .99 จากการเปรียบเทียบวิธีการให้คะแนนแบบทดสอบปิรามิดทั้ง 6 วิธีพบว่า การให้คะแนนโดยวิธีให้ตามค่าความยากเฉลี่ยของข้อที่ได้ทำ และการให้คะแนนข้อสอบทุกข้อมีความคงที่ของคะแนนมากที่สุด จากการศึกษาผลของช่วงเวลาห่างระหว่างการสอบครั้งที่ 1 และครั้งที่ 2 ที่มีต่อคะแนนผลการสอบ พบว่า ช่วงเวลาห่างระหว่างการสอบแต่ละครั้ง ไม่มีผลต่อความคงที่ของคะแนนทั้งในแบบทดสอบปิรามิด และแบบทดสอบดั้งเดิม ผลการศึกษาผลของการจำแนก การจำไม่มีผลต่อการทำแบบทดสอบปิรามิดและแบบทดสอบดั้งเดิม จากการศึกษาผลของวิธีการให้คะแนนที่มีต่อการเพิ่มของคะแนนผลการสอบในครั้งที่ 2 พบว่าการให้คะแนนทั้ง 6 วิธีไม่มีผลต่อการเพิ่มของคะแนนผลการสอบในครั้งที่ 2 ของแบบทดสอบปิรามิดฉบับที่ 1 และฉบับที่ 2 ส่วนในแบบทดสอบปิรามิดฉบับที่ 3 มีการให้คะแนนอยู่ 4 วิธี ที่มีผลให้คะแนนผลการสอบครั้งที่ 2 เพิ่มขึ้นอย่างมีนัยสำคัญทางสถิติ แต่มีอยู่ 2 วิธีคือการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ได้ทำ และการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูกไม่มีผลต่อการเพิ่มของคะแนนผลการสอบครั้งที่ 2 จากการศึกษาผลของ

การให้คะแนนที่มีต่อความเชื่อมั่นชนิดความคงที่ภายในของแบบทดสอบพบว่า เมื่อใช้ค่าดัชนีของ Hoyt (Hoyt Index) แบบทดสอบดั้งเดิมชนิด 40 ข้อจะมีค่าความเชื่อมั่นเท่ากับ 0.89 และ 0.90 เมื่อดำเนินการสอบครั้งที่ 1 และครั้งที่ 2 ตามลำดับ และเมื่อเพิ่มความยาวของแบบทดสอบดั้งเดิมเป็น 120 ข้อตามวิธีการของ Spearman Brown พบว่าความเชื่อมั่นเพิ่มขึ้นเป็น 0.96 ในการทดสอบทั้งสองครั้งพบว่า ค่าความเชื่อมั่นชนิดความคงที่ภายในของแบบทดสอบปิรามิดที่ให้คะแนนตามวิธีการให้คะแนนข้อสอบทุกข้อจะมีค่าสูงสุดถึง 0.99 และจากการศึกษาเปรียบเทียบวิธีการให้คะแนนทั้ง 6 วิธี พบว่า การให้คะแนนโดยวิธีการให้คะแนนข้อสอบทุกข้อในแบบทดสอบปิรามิดฉบับที่ 1 มีความสัมพันธ์กับแบบทดสอบดั้งเดิมสูงกว่าแบบทดสอบปิรามิดฉบับอื่น ๆ คือให้ค่าสหสัมพันธ์สูงถึง 0.86 คะแนนจากแบบทดสอบปิรามิดฉบับที่ 1 และฉบับที่ 2 ที่ให้คะแนนตามจำนวนข้อที่ตอบถูกกับการให้คะแนนตามค่าความยากของข้อสอบข้อที่ตัดจากข้อสอบข้อสุดท้ายที่ตอบ มีความสัมพันธ์กับคะแนนจากแบบทดสอบดั้งเดิมต่ำที่สุด เมื่อพิจารณาความสัมพันธ์ของคะแนนที่ได้จากการให้คะแนนวิธีต่าง ๆ จากผลการสอบทั้งสองครั้งพบว่า การให้คะแนนโดยวิธีให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ทำการให้คะแนนวิธีให้คะแนนข้อสอบทุกข้อมีความสัมพันธ์กันสูงมาก ส่วนการให้คะแนนวิธีให้ตามจำนวนข้อที่ตอบถูกมีความสัมพันธ์กับการให้คะแนนตามค่าความยากของข้อสอบข้อที่ตัดจากข้อสอบข้อสุดท้ายที่ตอบสูงที่สุด คือให้ค่าสหสัมพันธ์เท่ากับ 1.00 และการให้คะแนนวิธีให้คะแนนข้อสอบทุกข้อจะมีความสัมพันธ์กับคะแนนที่ให้โดยวิธีอื่น ๆ

จากผลการศึกษาของ Larkin และ Weiss ดังกล่าวนี้นี้สรุปได้ว่า แบบทดสอบปิรามิดมีความคงที่ในการประมาณความสามารถของผู้สอบได้เท่ากับแบบทดสอบดั้งเดิม ที่มีจำนวนข้อสอบมากกว่าและประมาณความสามารถของผู้สอบได้กว้างกว่าข้อสอบดั้งเดิมที่มีจำนวนข้อสอบเท่ากัน

Larkin และ Weiss (1975) ได้ศึกษาเปรียบเทียบแบบทดสอบชนิด 2 ชั้น กับแบบทดสอบปิรามิดที่สร้างมาจากกลุ่มข้อสอบ จำนวน 369 ข้อ เป็นข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก ซึ่งถามเนื้อหาเกี่ยวกับคำศัพท์ที่อยู่ในระดับต่ำกว่าระดับวิทยาลัย ข้อสอบได้วิเคราะห์หาค่าอำนาจจำแนก a_g (Discrimination) และ b_g (Difficulty) ตามทฤษฎีคุณลักษณะแฝง ข้อสอบแต่ละข้อมีค่าสัมประสิทธิ์สหสัมพันธ์แบบ Biserial ตั้งแต่ .30 ขึ้นไป แบบทดสอบชนิดสองชั้นที่สร้างขึ้นนี้ประกอบด้วยแบบทดสอบแยกทาง (Routing Test) และแบบทดสอบวัดผล (Measurement Tests) แบบทดสอบแยกทางประกอบด้วยข้อสอบ 10 ข้อ มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบ (b_g) เท่ากับ -0.232 และ 0.050 มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบ (a_g) เท่ากับ 0.702 และ 0.163 ตามลำดับ ผู้สอบที่ทำแบบทดสอบแยกทางได้คะแนน 0-4, 5-6, 7-8 และ 9-10 คะแนนจะได้ตอบแบบทดสอบวัดผลฉบับที่ง่ายที่สุดหรือ

ยากขึ้นชุดใดชุดหนึ่งตามลำดับของคะแนน แบบทดสอบวัดผลซึ่งมีอยู่ 4 ฉบับ ๆ ละ 30 ข้อ มีค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบในฉบับที่ 1 ถึง 4 เท่ากับ 1.825 และ 0.558, 0.350 และ 0.297, -0.709 และ 0.189, -1.603 และ 0.373 ตามลำดับ มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบในฉบับที่ 1 ถึง 4 เท่ากับ 0.530 และ 0.126, 0.684 และ 0.214, 0.611 และ 0.122, 0.693 และ 0.213 ตามลำดับ การให้คะแนนแบบทดสอบชนิดสองชั้นใช้วิธีประมาณความสามารถ (Ability Estimate) ของผู้สอบตามวิธีของ Lord

แบบทดสอบปิรามิดที่ใช้ในการศึกษามี 3 ฉบับ เป็นแบบทดสอบเช่นเดียวกับที่ใช้ในการศึกษาในปี 1974 (Larkin และ Weiss, 1974) แต่ละฉบับมี 15 ชิ้น มีค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชิ้นที่อยู่ติดกันคงที่ ใช้กฎการแยก ทางแบบเพิ่ม 1/ลด 1 มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบเท่ากับ -1.094 และ 1.256 มีค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบเท่ากับ .799 และ .457 ตามลำดับ ใช้วิธีการให้คะแนน 4 วิธีคือ ให้ตามจำนวนข้อที่ตอบถูก ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก และให้ตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบ ดำเนินการสอบโดยใช้เครื่องคอมพิวเตอร์ ทดสอบกับนักศึกษาจำนวน 111 คน ผลการศึกษานพบว่า จากการใช้ One-Way Multivariate Analysis of Variance ทดสอบคะแนนเฉลี่ยที่ได้จากการจัดลำดับการทดสอบก่อนและหลังระหว่างแบบทดสอบปิรามิดกับแบบทดสอบสองชั้น พบว่าคะแนนไม่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ จึงถือได้ว่าการจัดให้ผู้สอบทำแบบทดสอบใดก่อนก็จะไม่ส่งผลกระทบต่อคะแนนที่ได้รับ จากการศึกษาการกระจายของคะแนนแบบทดสอบปิรามิดที่ได้จากการให้คะแนน 4 วิธีพบว่าคะแนนที่ได้จากทั้ง 4 วิธีมีการกระจายแบบเบ้ทางบวกเล็กน้อย (Slightly Positive Skewness) โดยคะแนนที่ให้ตามจำนวนข้อที่ตอบถูกจะมีความเบ้มากที่สุด และแตกต่างจากการกระจายแบบโค้งปกติ (Zero Skewness) อย่างมีนัยสำคัญทางสถิติ ส่วนคะแนนที่ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูกเกือบจะไม่มี ความเบ้เลย คะแนนที่ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ และให้ตามค่าความยากของข้อสอบข้อสุดท้ายที่ได้ตอบจะมีการกระจายแบบโค้งน้อย แต่ไม่แตกต่างจากโค้งปกติอย่างมีนัยสำคัญทางสถิติ คะแนนที่ให้ตามจำนวนข้อที่ตอบถูกและให้ตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก จะมีโค้งการกระจายโค้งค่อนข้างมาก คะแนนที่ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ได้ตอบ จะมีโค้งการกระจายแบบโค้งน้อย เมื่อพิจารณาทั้งความเบ้และความโค้งของการกระจายของคะแนนแล้ว คะแนนที่ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูกจะแตกต่างไปจากโค้งการกระจายปกติเล็กน้อยที่สุด การกระจายของคะแนนที่ได้จากแบบทดสอบชนิดสองชั้น จะมีการกระจายในลักษณะโค้งเบ้ทางบวกเล็กน้อยและความโค้งเล็กน้อย ซึ่งแตกต่างจากโค้งการกระจาย

จ่ายแบบปกติอย่างมีนัยสำคัญทางสถิติ นอกจากนี้ทั้งความเบ้และความโด่งของโค้งการกระจายของคะแนนจากแบบทดสอบชนิดสองชั้น ยังมีความเบ้และความโด่งมากกว่า คะแนนที่ได้จากแบบทดสอบปิรามิตด้วย เมื่อพิจารณาแบบทดสอบย่อยของแบบทดสอบชนิดสองชั้นพบว่า แบบทดสอบแยกทางมีความยากของข้อสอบที่เหมาะสม คือมีคะแนนเฉลี่ยของข้อสอบเท่ากับ 5.58 จาก 10 คะแนนเมื่อคิดจาก 111 คน ส่วนเบี่ยงเบนมาตรฐานของคะแนนข้อสอบมีค่ามากคือ เท่ากับ 2.61 แสดงว่าแบบทดสอบแยกทางสามารถจำแนกผู้สอบออกตามความสามารถได้ดี คะแนนเฉลี่ยข้อสอบของแบบทดสอบวัดผลรวมทั้ง 4 ฉบับเท่ากับ 18.56 คะแนน มีค่ามากกว่าครึ่งหนึ่งของจำนวนข้อในแบบทดสอบวัดผล เมื่อพิจารณาแยกในแบบทดสอบวัดผลแต่ละฉบับ พบว่าคะแนนเฉลี่ยข้อสอบจะอยู่ระหว่าง 17.00 ถึง 19.93 หรือประมาณ ร้อยละ 57 ถึง ร้อยละ 66 ของจำนวนข้อสอบทั้ง 30 ข้อ แสดงว่าความยากของแบบทดสอบวัดผลเหมาะสมกับกลุ่มผู้สอบในแต่ละกลุ่มทางเลือก อย่างไรก็ตามพบว่ามีการเดาเกิดขึ้นในการทำแบบทดสอบชนิดสองชั้นมากกว่าแบบทดสอบปิรามิต

จากการพิจารณาความสัมพันธ์ของคะแนนแบบทดสอบปิรามิตที่ให้คะแนนทั้ง 4 วิธีกับคะแนนจากแบบทดสอบชนิดสองชั้น พบว่ามีค่าสหสัมพันธ์แบบ Pearson Product Moment เท่ากับ .84, .81, .79 และ .83 ตามลำดับ และจากการพิจารณาค่าความเชื่อมั่นชนิดความคงที่ภายในของแบบทดสอบชนิดสองชั้น จากค่าดัชนีของ Hoyt พบว่า แบบทดสอบแยกทางมีค่า .72 (10 ข้อ 111 คน) แบบทดสอบวัดผลฉบับที่ 1 ถึง 4 มีค่า .84, .66, .75 และ .72 ตามลำดับ (ทุกฉบับมี 30 ข้อ และมีจำนวนคน 21, 20, 27 และ 43 คนตามลำดับ)

เมื่อพิจารณาความสัมพันธ์ภายใน (Intercorrelations) ของคะแนนที่ได้จากแบบทดสอบปิรามิต โดยใช้ค่าสหสัมพันธ์แบบ Pearson Product Moment พบว่าคะแนนที่ให้ตามวิธีที่ 2 กับวิธีที่ 3 มีความสัมพันธ์กันสูงที่สุดคือมีค่าเท่ากับ .99 คะแนนที่ให้ตามวิธีที่ 1 มีความสัมพันธ์กับคะแนนที่ให้ตามวิธีที่ 2 และวิธีที่ 3 ต่ำที่สุดคือมีค่าเท่ากับ .93 เท่ากัน และคะแนนที่ให้ตามวิธีที่ 4 กับวิธีที่ 2 และวิธีที่ 3 มีความสัมพันธ์กับปานกลางคือมีค่าเท่ากับ .95 เท่ากัน

2.8.2 การศึกษาจากสถานการณ์จำลอง (Simulation Studies)

การศึกษาจากสถานการณ์จำลองมีอยู่สองลักษณะคือ การใช้ข้อมูลจริง (Real Data Simulation) เป็นการทดลองให้คะแนนจากผลการสอบด้วยแบบทดสอบดั้งเดิมในลักษณะของแบบทดสอบปิรามิต อีกลักษณะหนึ่งเป็นการใช้คอมพิวเตอร์ในการสุ่มเลือกข้อสอบ และผลการสอบในลักษณะของแบบทดสอบปิรามิต แล้วนำผลที่ได้มาวิเคราะห์ต่อไป

Bryson (อ้างถึงใน Larkin and Weiss, 1974) ได้ศึกษาเปรียบเทียบผลการวิจัยเชิงประจักษ์ กับการใช้ข้อมูลจริงในการเปรียบเทียบแบบทดสอบปิรามิดและแบบทดสอบดั้งเดิมทั้ง 4 ฉบับ เดิมที่กล่าวมาแล้วนั้น โดยใช้กลุ่มตัวอย่างจำนวน 100 คน ผลการวิเคราะห์ พบว่าแบบทดสอบปิรามิดฉบับหนึ่งมีค่าสหสัมพันธ์กับคะแนนรวมสูงที่สุด แบบทดสอบปิรามิดอีกฉบับหนึ่งมีค่าสหสัมพันธ์เท่ากับแบบทดสอบดั้งเดิมฉบับหนึ่งและสูงกว่าแบบทดสอบดั้งเดิมอีกฉบับหนึ่ง ผลการศึกษาดังนี้ทำให้มองเห็นจุดเด่นของแบบทดสอบแบบปิรามิดชัดเจนมากขึ้นกว่าการศึกษาเชิงประจักษ์ ที่ผ่านมาของเขามากยิ่งขึ้น

Linn, Rock และ Cleary (1969) ได้ศึกษาจากสถานการณ์จำลองในลักษณะข้อมูลจริงจากผลการทำข้อสอบแบบดั้งเดิมจำนวน 190 ข้อของกลุ่มตัวอย่าง จำนวน 4,885 คนเพื่อศึกษาแบบทดสอบที่ปรับความยากให้เหมาะสมกับผู้สอบที่ใช้การแยกทางต่าง ๆ กันจำนวน 7 วิธี ในแต่ละวิธีจะคัดเลือกข้อสอบที่เหมาะสม จากแบบทดสอบดั้งเดิม แล้วมาให้คะแนนตามวิธีการของแบบทดสอบชนิด Two-stage จำนวน 5 ฉบับ (Betz & Weiss, 1973) และตามวิธีการสอบแบบทดสอบปิรามิด 2 ฉบับ ฉบับหนึ่งเป็นแบบ 10 ชั้น มีขนาดของชั้นประมาณ .02 ($p=.02$) อีกฉบับหนึ่งเป็นแบบ 5 ชั้นที่ในแต่ละชั้นมีข้อสอบ 5 ข้อ ดังนั้นผู้สอบแต่ละคนจะได้ทำข้อสอบคนละ 25 ข้อ แบบทดสอบปิรามิดทั้งสองฉบับมีการแยกทางที่เหมือนกันทุกประการ ได้ศึกษาเปรียบเทียบแบบทดสอบปิรามิดทั้งสองฉบับกับแบบทดสอบดั้งเดิมจำนวน 5 ฉบับที่มีข้อสอบตั้งแต่ 10-50 ข้อ ผลการวิจัยสรุปได้ว่าแบบทดสอบปิรามิดใช้ข้อสอบน้อยกว่าแบบทดสอบดั้งเดิมเพื่อให้ค่าความเที่ยงตรง ในการทดสอบเท่ากัน

Paterson (อ้างถึงใน Larkin and Weiss, 1974) ใช้คอมพิวเตอร์สร้างสถานการณ์จำลองในลักษณะมอนติคาโร เพื่อศึกษาแบบทดสอบปิรามิดที่สร้างโดยยึดค่าความยากเป็นหลักและจัดเรียงตามค่าอำนาจจำแนก ในแต่ละระดับความยากข้อสอบข้อแรกจะมีค่าอำนาจจำแนกสูงที่สุดแล้วค่อย ๆ ลดลงเรื่อย ๆ ดังนั้นขนาดของชั้นจึงเปลี่ยนไป เพราะเป็นผลมาจากการยึดค่าอำนาจจำแนกเป็นหลัก เมื่อการตอบข้อสอบข้อที่มีค่าอำนาจจำแนกต่ำถูกการเพิ่มความยากของข้อสอบข้อต่อไปจะกว้างมาก แต่ถ้าตอบข้อสอบข้อที่มีค่าอำนาจจำแนกต่ำผิดการเพิ่มความยากของข้อสอบข้อต่อไปจะเพิ่มน้อยมาก เช่นเดียวกันการลดขนาดของชั้นก็ขึ้นอยู่กับคำตอบข้อสอบว่าถูกหรือผิด ผลการวิเคราะห์ข้อมูลพบว่าการจัดโครงสร้างของแบบทดสอบปิรามิด เมื่อเกิดความผิดพลาดของค่าพารามิเตอร์ของข้อสอบเกิดขึ้นจะมีผลต่อการแจกแจงของคะแนนผลการสอบน้อยมาก แบบทดสอบปิรามิดจะใช้ประมาณความสามารถของผู้สอบได้ดีกว่าแบบทดสอบดั้งเดิม เมื่อการแจกแจงความสามารถของผู้สอบเป็นแบบรูปตัวยู หรือรูปสี่เหลี่ยม โดยที่คะแนนจากแบบทดสอบจะมีความแม่นยำมากกว่าในแบบทดสอบดั้งเดิม โดยเฉพาะในส่วนปลาย ๆ ของความสามารถทั้งทางด้านต่ำและสูง และสามารถให้คะแนนผลการ

สอบในการพยากรณ์ความสามารถได้ดีเท่า ๆ กันทั้งแบบทดสอบปิรามิดและแบบทดสอบดั้งเดิม

2.8.3 การศึกษาเชิงทฤษฎี (Theoretical Studies)

Waters (อ้างถึงใน Larkin and Weiss, 1974) ได้ศึกษาเปรียบเทียบเชิงทฤษฎีระหว่างแบบทดสอบปิรามิดชนิด 5 ชั้นกับแบบทดสอบดั้งเดิมจำนวน 4 ฉบับ แต่ละฉบับมีข้อสอบ จำนวน 5 ข้อ โดยใช้โมเดลของลอว์รีต เพื่อหาค่าสหสัมพันธ์ระหว่างคะแนนของแบบทดสอบกับระดับความสามารถในแบบทดสอบแต่ละฉบับ แบบทดสอบปิรามิดใช้ขนาดของชั้นเท่ากับ .10 ($p=.10$) ใช้การแยกทางแบบเพิ่ม1/ลด1 และให้คะแนน 2 วิธี ผลการศึกษาการให้คะแนนทั้ง 2 วิธีของแบบทดสอบปิรามิดได้ค่าสหสัมพันธ์ระหว่างคะแนนรวมกับระดับความสามารถสูงกว่าแบบทดสอบดั้งเดิมทุกฉบับ ถึงแม้ว่าการตอบจะเป็นลักษณะอิสระหรือเป็นแบบปรนัยก็ตาม คะแนนผลการสอบของแบบทดสอบปิรามิดจะมีการกระจายเป็นลักษณะรูปลิ่มเหลี่ยม และมีช่วงกว้างกว่าแบบทดสอบดั้งเดิม

Waters & Bayroff (1971) ได้ศึกษาเปรียบเทียบแบบทดสอบปิรามิดชนิด 5, 10 และ 15 ชั้นที่มีข้อสอบชั้นละ 1 ข้อ และชนิด 10 ชั้นที่มีข้อสอบชั้นละ 2 ข้อ กับแบบทดสอบดั้งเดิมที่มีจำนวนข้อสอบเท่ากัน ทั้งแบบทดสอบปิรามิดและแบบทดสอบดั้งเดิมมีข้อสอบที่มีระดับความยากและค่าอำนาจจำแนกแตกต่างกันอย่างมีระบบ และมีข้อสมมติว่าการแจกแจงของความสามารถเป็นแบบปกติ ผลการศึกษานพบว่า คะแนนผลการสอบและระดับความสามารถมีความสัมพันธ์กับการแจกแจงของความยากและค่าอำนาจจำแนกของแบบทดสอบ ซึ่งแบบทดสอบปิรามิดจะมีความสัมพันธ์สูงกว่าในแบบทดสอบดั้งเดิมโดยเฉพาะอย่างยิ่งข้อที่มีค่าอำนาจจำแนกสูง ๆ และในแบบทดสอบปิรามิดชนิดที่มีข้อสอบ 1 ข้อในแต่ละชั้นจะมีความสัมพันธ์ระหว่างคะแนนรวมกับระดับความสามารถสูงมาก

Lord ได้รายงานผลการศึกษาในเชิงทฤษฎี เกี่ยวกับแบบทดสอบปิรามิดไว้สองสามเรื่อง (Weiss & Betz, 1973) โดยอาศัย Item Characteristic Curve Theory และ Theory of Markov Chain ในการวิเคราะห์ เพื่อเปรียบเทียบแบบทดสอบปิรามิดชนิด 10 ชั้น 15 ชั้น และ 60 ชั้น กับแบบทดสอบดั้งเดิมที่มีจำนวน 60 ข้อ แบบทดสอบแต่ละฉบับใช้ขนาดของชั้นแตกต่างกันแต่เป็นแบบคงที่เหมือนกัน การแยกทางที่ศึกษาเป็นแบบ เพิ่ม1/ลด1 เพิ่ม1/ลด2 เพิ่ม1/ลด3 และเพิ่ม2/ลด3 และวิธีการให้คะแนนที่แตกต่างกัน ผลการศึกษานพบว่า ในแบบทดสอบดั้งเดิมข้อมูลที่ได้จะมีลักษณะการแจกแจงเป็นรูประฆังคว่ำมีความโด่งสูงมาก (Leptokurtic) และมีลักษณะสมมาตรตรงค่ากึ่งกลางของระดับความสามารถ การประมาณความสามารถจะแม่นยำในช่วงความสามารถที่ใกล้ระดับความสามารถตรงกลางมากที่สุด ส่วนแบบทดสอบปิรามิดลักษณะข้อมูลจะมีการแจกแจงที่มี

ความโค้งแบน (Platykurtic) ในบางกรณีอาจพบว่ามีลักษณะเป็นเส้นตรงได้ ซึ่งหมายความว่าคะแนนผลการสอบกระจายในแต่ละระดับความสามารถเท่า ๆ กัน ที่ระดับความสามารถตรงกลางแบบทดสอบดั้งเดิมมีความแม่นยำในการวัดสูงกว่าแบบทดสอบปิรามิต ส่วนในช่วงระดับความสามารถตั้งแต่ $+1.5$ SD. ถึง $+1.0$ SD. แบบทดสอบปิรามิตจะมีความแม่นยำในการวัดสูงกว่าแบบทดสอบดั้งเดิม

Mussio (อ้างถึงใน Larkin and Weiss, 1974) ได้ศึกษาเนื้อหาแนวทางการลดจำนวนข้อสอบในแบบทดสอบปิรามิตลงจึงได้เสนอวิธีการ 2 วิธี คือ Reflecting Barrier และ Retraining Barrier ทั้งสองวิธีเป็นเทคนิคการตัดข้อสอบที่อยู่ในช่วงปลายทั้งสองข้างของระดับความสามารถสูงและต่ำของแบบทดสอบปิรามิตออกไป ซึ่งเรียกรูปแบบนี้ว่าแบบทดสอบปิรามิตข้างตัด ซึ่งเป็นการตัดข้อสอบที่มีค่าความยากต่ำและสูงมาก ๆ ออกไปนั่นเอง ผลการศึกษาได้ผลเช่นเดียวกับของ Lord คือในช่วงระดับความสามารถเฉลี่ยแบบทดสอบดั้งเดิมจะให้ข้อมูลได้ดีกว่าแบบทดสอบปิรามิต แต่ในช่วงระดับความสามารถที่อยู่ห่างจากความสามารถเฉลี่ยแล้วแบบทดสอบปิรามิตให้ข้อมูลได้ดีกว่า และพบว่าวิธี Retraining Barrier ใช้ประมาณความสามารถในระดับความสามารถต่าง ๆ ได้แม่นยำกว่าวิธี Reflecting Barrier ถึงแม้ว่าแบบทดสอบปิรามิตทั้งสองรูปแบบนี้จะขาดในช่วงปลายสุดของระดับความสามารถต่ำและสูงไปก็ยังไม่พบว่า การวัดในช่วงความสามารถปลายสุดทั้งสองด้านนั้น แบบทดสอบปิรามิตมีความแม่นยำมากกว่าแบบทดสอบดั้งเดิม

จากผลการศึกษาวิจัยที่เกี่ยวกับแบบทดสอบปิรามิตดังกล่าวข้างต้น พบว่ามีการศึกษาในหลายลักษณะคือ ชนิดของแบบจำลองปิรามิต ประเภทเนื้อหาวิชาของกลุ่มข้อสอบ จำนวนข้อสอบในแต่ละชั้นลักษณะของแบบทดสอบ จำนวนชั้นในโครงสร้างปิรามิต กฎการแยกทาง ความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกัน วิธีการให้คะแนน การดำเนินการสอบ และการเปรียบเทียบผลที่ได้จากแบบทดสอบปิรามิตกับแบบทดสอบดั้งเดิม หรือกับเกณฑ์ภายนอกต่าง ๆ ด้วย อย่างไรก็ตามผลการวิจัยดังกล่าวยังได้ข้อสรุปที่ขัดแย้งกันอยู่บ้าง ดังเช่น ผลการวิจัยของ Wood ในปี ค.ศ. 1969 สรุปว่า แบบทดสอบดั้งเดิมจะดีกว่าแบบทดสอบปิรามิตที่มีค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันมีค่าคงที่และใช้กฎการแยกทางแบบเพิ่ม $1/\log$ ซึ่งให้ผลตรงข้ามกับผลการวิจัยของ Waters ในปี ค.ศ. 1964 ที่พบว่าแบบทดสอบปิรามิตที่ใช้ความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันคงที่และใช้กฎการแยกทางแบบเพิ่ม $1/\log$ มีความสัมพันธ์กับเกณฑ์และให้ศักยภาพการกระจายของคะแนนดีกว่าแบบทดสอบดั้งเดิม นอกจากนี้การศึกษาวิจัยที่ดำเนินการสอบด้วยคอมพิวเตอร์ส่วนใหญ่จะเป็นการศึกษาในสถานการณ์จำลอง ซึ่งสามารถดำเนินการสอบได้สะดวก คะแนนผลการสอบก็ไม่มีผลกระทบที่มาจากผู้ดำเนินการสอบ ตัวผู้สอบ วิธีดำเนินการสอบ ตลอดจน

สถานแวดล้อมในขณะดำเนินการสอบก็ตาม การศึกษาในลักษณะดังกล่าวจึงยังมีข้อจำกัดของการศึกษาในสถานการณ์ทดสอบจริง โดยเฉพาอย่างยิ่งการศึกษากับแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนในชั้นเรียน

จากแนวความคิด ทฤษฎี และผลงานวิจัยที่ผ่านมาดังกล่าว ตลอดจนการศึกษาแนวทางการพัฒนารูปแบบการทดสอบในประเทศไทยของเรา ที่ควรมุ่งพัฒนาการทดสอบที่สามารถวัดระดับความสามารถของผู้สอบได้อย่างถูกต้องและเหมาะสม โดยการนำเทคโนโลยีทางด้านคอมพิวเตอร์ซึ่งมีความเจริญอย่างรวดเร็วมาใช้เป็นเครื่องมือในการพัฒนาการทดสอบให้มีประสิทธิภาพสูงสุด ผู้วิจัยจึงต้องการศึกษาการดำเนินการสอบแบบทดสอบปริมาตรด้วยคอมพิวเตอร์ 2 ฉบับคือแบบทดสอบปริมาตรที่ใช้ขนาดชั้นคงที่ชนิด 8 ชั้นและใช้กฎการแยกทางแบบเต็ม 1/ลด 1 กับแบบทดสอบปริมาตรชนิด 5 ชั้นที่มีข้อสอบ 3 ข้อในแต่ละชั้น เปรียบเทียบกับแบบทดสอบดั้งเดิมที่ดำเนินการสอบโดยการเขียนตอบ ในสถานการณ์ทดสอบจริงในชั้นเรียน นอกจากนี้ผู้วิจัยยังได้ศึกษาประสิทธิภาพของแบบทดสอบปริมาตรที่ใช้ในการวิจัยทั้งทางด้านความตรงเชิงสัมพันธ์กับเกณฑ์และ เชิงปฏิบัติการ และได้ประเมินผลโปรแกรมคอมพิวเตอร์สำหรับดำเนินการสอบแบบทดสอบปริมาตรที่พัฒนาขึ้นในการวิจัยครั้งนี้ด้วย