บทที่ 2

ทฤษฎี แนวคิด และงานวิจัยที่เกี่ยวข้อง

ในการวิจัยนี้ ผู้วิจัยได้ศึกษาทฤษฎี แนวคิดและงานวิจัยที่เกี่ยวข้อง สามารถสรุปเนื้อหาได้เป็น 4 หัวข้อคือ

1. ทฤษฎีคุณลักษณะแฝง (Latent Traits Theory) กับทฤษฎีโค้งลักษณะข้อสอบ (Item Characteristic Curve Theory)

2. การวิเคราะห์ข้อสอบ และ การประมาณค่าความสามารถตามทฤษฎีโค้งลักษณะข้อสอบรูปแบบโลจิสติคที่มีพารามิเตอร์ 3 ตัว

3. การทดสอบแบบเทเลอร์

4. ผลการวิจัยที่เกี่ยวข้อง

รายละเอียดของเนื้อหาทั้ง 4 หัวข้อมีดังนี้

1. ทฤษฎีคุณลักษณะแฝงกับทฤษฎีโค้งลักษณะข้อสอบ

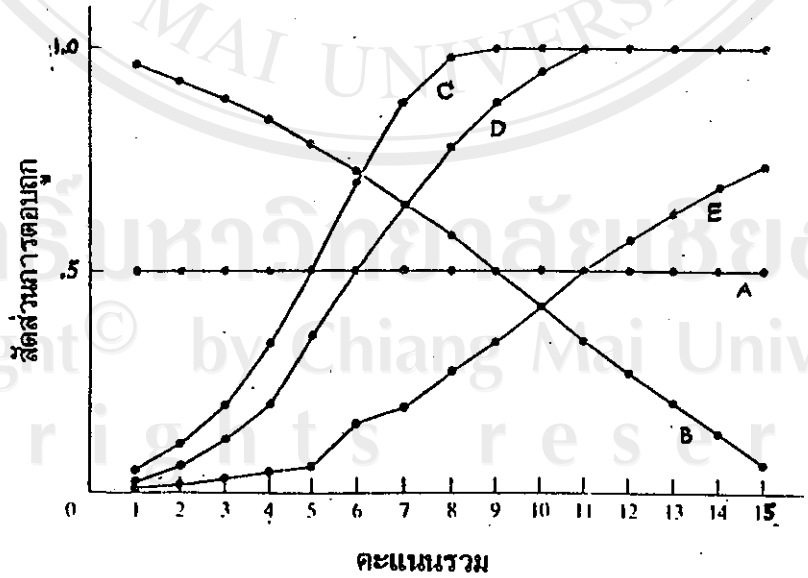
ทฤษฎีคุณลักษณะแฝง เป็นทฤษฎีทางจิตวิทยาซึ่งถือว่า คุณลักษณะภายใน (Traits) ของบุคคลคนหนึ่งเป็นตัวกำหนดพฤติกรรมของเขา และเนื่องจากคุณลักษณะภายในของบุคคลเป็นสิ่งที่ไม่สามารถสังเกตหรือวัดได้โดยตรง แต่พฤติกรรมของเขาสามารถวัดได้โดยตรง จึงสามารถใช้ผลจากการสังเกตพฤติกรรมของบุคคลมาเป็นแนวทาง เพื่อสรุปคุณลักษณะภายในอันนั้นของเขา (ผจงจิต อินทสุวรรณ 2525 หน้า 55) นักจิตวิทยาที่จะใช้ผลการตอบข้อสอบวัดสติปัญญาไปอธิบายอ้างอิงเป็นคุณลักษณะทางจิตวิทยาของผู้ตอบข้อสอบ จะต้องมีความรู้ว่าคุณลักษณะทางจิตวิทยาเหล่านั้นสามารถกำหนดหรือสัมพันธ์กับคำตอบของผู้ตอบข้อสอบอย่างไรและทฤษฎีโค้งลักษณะข้อสอบก็คือทฤษฎีหนึ่งที่ใช้รูปแบบทางคณิตศาสตร์ (Mathematical model) ชนิดหนึ่งมาอธิบายถึงความสัมพันธ์ดังกล่าว (Lord and Novick, 1968 : 358)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

แอลเลนและเยน (Allen and Yen, 1979 : 127) อธิบายว่า " โค้งลักษณะข้อสอบ (Item-characteristic Curve : ICC) คือ กราฟแสดงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อหนึ่งได้ถูกต้องกับตำแหน่งของผู้เข้าสอบตามคุณลักษณะที่วัดในแบบทดสอบนั้น ผลการวิเคราะห์ข้อสอบโดยใช้ ICC สามารถอธิบายได้คล้าย ๆ กับการวิเคราะห์ข้อสอบโดยใช้ค่าความยากง่ายและค่าอำนาจจำแนก โดยที่ความชัน (slope) ของ ICC จะใช้บอกอำนาจจำแนก ส่วนความยากของข้อสอบจะพิจารณาจากความสัมพันธ์ของคะแนนรวมกับสัดส่วน .50 ของโอกาสในการตอบข้อสอบข้อนั้นได้ถูกต้อง"

ลักษณะของ ICC แสดงไว้ในภาพที่ 1

ภาพที่ 1 แสดงตัวอย่าง ICC ของข้อสอบ 5 ข้อ



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

จากภาพ ข้อ A แสดงลักษณะข้อสอบที่ไม่มีอำนาจจำแนกเพราะทุกระดับคะแนนรวมจะมีคนตอบถูกเท่า ๆ กัน

ข้อ B แสดงลักษณะข้อสอบที่มีอำนาจจำแนกเป็นลบ

ข้อ C,D แสดงลักษณะข้อสอบที่มีอำนาจจำแนกดีโดยที่ข้อ C มีความยากเท่ากับ 5 คะแนน ข้อ D มีความยากเท่ากับ 6 คะแนน

ข้อ E แสดงลักษณะข้อสอบที่มีอำนาจจำแนกพอใช้ได้และเป็นข้อที่ยากกว่าข้อ C และ D โดยมีความยากเป็น 11 คะแนน

ในระยะแรก ๆ ทฤษฎีโค้งลักษณะข้อสอบ เรียกกันว่าทฤษฎีการตอบข้อสอบ (Item Response Theory) ซึ่งริเริ่มแนวคิดโดยเฟอร์กุสัน (Ferguson, 1942 Cited by Warm, 1979 :19) โดยใช้รูปแบบโค้งปกติสะสม (Normal Ogive Model) ส่วนผู้พัฒนาวิธีการให้สามารถใช้โค้งปกติสะสมกับข้อสอบ n ข้อ (n > 2) เป็นคนแรกคือลอว์เลย์ (Lawlew, 1943, 1944 cited by Lord and Novick, 1968 :369) และมีการนำรูปแบบนี้ไปใช้ในงานทางทฤษฎีโดยผู้เขียนหลาย ๆ คนได้แก่ บรอกเดน (Brogden, 1946) ทัคเกอร์ (Tucker, 1946) ครอนบัคและวาร์ริงตัน (Cronbach and Warrington, 1952) ลอร์ด (Lord, 1952, 1953) ครอนบัคและเมอร์วิน (Cronbach and Mervin, 1955) ครอนบัคและอาซุมา (Cronbach and Azuma, 1962) และปีเตอร์สัน (Perterson, 1962) มีการนำไปใช้ในการสร้างแบบทดสอบจริงโดยอินโดว์และซาเมจิมา (Indow and Samejima, 1962, 1966)

อย่างไรก็ตามรูปแบบโค้งปกติสะสมเป็นรูปแบบที่มีความยุ่งยากซับซ้อนและคำนวณได้ยากจึงไม่ได้รับความนิยมเท่าที่ควรจนกระทั่งในปี 1960 ราชค์ (Rasch) ได้เสนอรูปแบบการวิเคราะห์ข้อสอบที่เรียกว่า Rasch's Model และลอร์ด (Lord) ได้หันกลับมาสนใจ

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ทฤษฎีนี้อีกในปี 1965 และพัฒนาทฤษฎีจนสามารถใช้ได้ดี ในปี 1968 เบิร์นบอม (Birnbaum) ได้เสนอรูปแบบการวิเคราะห์ที่คล้ายกับรูปแบบโค้งปกติสะสมเรียกว่า รูปแบบโลจิสติค (Logistic Model) ที่มีค่าพารามิเตอร์ 2 ตัว ซึ่งง่ายกว่ารูปแบบและวิธีการที่ลอร์ดพัฒนาจึงได้รับความนิยมอย่างแพร่หลาย และมีการพัฒนาให้สามารถใช้ได้กับพารามิเตอร์เดียวและพารามิเตอร์ 3 ตัวด้วย (Warm, 1979 : 19-21)

ข้อตกลงเบื้องต้นตามทฤษฎีโค้งลักษณะข้อสอบ

สำเริง บุญเรืองรัตน์ (2527 : 99-101) ไอรอนสัน (Ironson, 1982 : 112) กล่าวถึงข้อตกลงเบื้องต้นในการวิเคราะห์ข้อสอบตามทฤษฎีโค้งลักษณะข้อสอบ สรุปได้ว่ามี 3 ประการ คือ

1. แบบทดสอบต้องมีความเป็นมิติเดียว (Unidimension Test) หมายความว่า ข้อสอบแต่ละข้อในแบบทดสอบจะต้องวัดคุณลักษณะหรือความสามารถเดียวกันหรือมีความเป็นเอกพันธุ์นั่นเอง

2. ข้อสอบแต่ละข้อจะต้องเป็นอิสระจากกัน (Local Independence) หมายความว่า การตอบข้อสอบข้อใดข้อหนึ่งถูก จะไม่มีผลต่อการตอบข้อสอบข้ออื่น ๆ ในแบบทดสอบนั้น

3. โอกาสที่ผู้สอบจะตอบข้อสอบถูก ขึ้นอยู่กับโค้งลักษณะข้อสอบของแต่ละรูปแบบ ไม่ขึ้นกับการแจกแจงความสามารถของกลุ่มตัวอย่าง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2. การวิเคราะห์ข้อสอบและการประมาณค่าความสามารถตามทฤษฎีโค้งลักษณะข้อสอบรูปแบบโลจิสติคที่มีพารามิเตอร์ 3 ตัว

2.1. การวิเคราะห์ข้อสอบ

สมการแสดงความสัมพันธ์ของพารามิเตอร์ต่าง ๆ ในรูปของความน่าจะเป็นคือสมการ (Ironson, 1982 : 142)

$$P_g(\Theta) = c_g + (1 - c_g)\frac{e^{Da_g(\Theta - b_g)}}{1 + e^{Da_g(\Theta - b_g)}}, \quad g = 1,2,3,\ldots,n$$

ในที่นี้ $P_g(\Theta)$ คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ Θ จะสามารถตอบข้อสอบข้อ g ได้ถูกต้อง

D เป็นค่าคงที่มีค่าเป็น 1.7 ซึ่งเป็นสเกลลิ่งแฟคเตอร์ (Scaling Factor) ได้จากค่าความเบี่ยงเบนมาตราฐานของฟังก์ชัน ซี ดี เอฟ ของ โลจิสติค (Birnbaum, 1968, in Lord and Novick, 1968 : 399)

e คือ ค่าเอกซ์โพเนนเชี่ยล (exponential) เป็นค่าคงที่มีค่าเป็น 2.71828

a_g คือ ค่าพารามิเตอร์อำนาจจำแนกของข้อสอบข้อ g

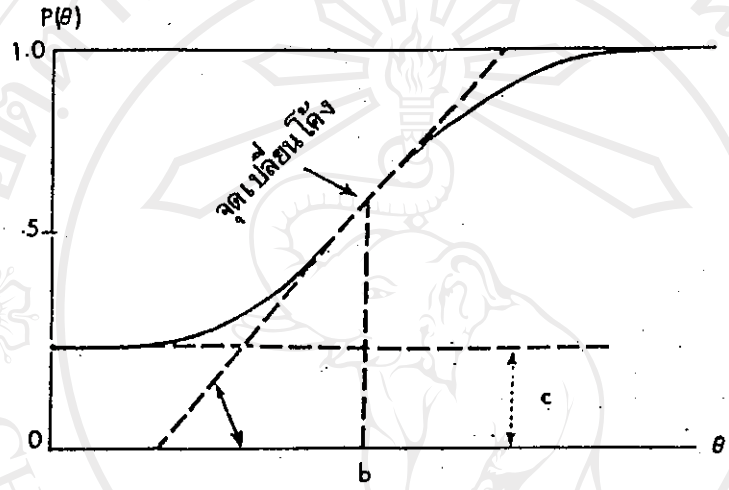
b_g คือ ค่าพารามิเตอร์ความยากของข้อสอบข้อ g

c_g คือ ค่าพารามิเตอร์การเดาของข้อสอบข้อ g

ความหมายของค่าพารามิเตอร์ดังกล่าวแสดงไว้ในภาพที่ 2

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 2 แสดงความหมายของค่าพารามิเตอร์ของข้อสอบตามรูปแบบโลจิสติค ที่มีพารามิเตอร 3 ตัว



แฮมเบิลตันและคุ๊ก (Hambleton and Cook, 1977 : 81) กล่าวถึงค่าพารามิเตอร์ต่าง ๆ ตามรูปแบบพารามิเตอร์ 3 ตัว สรุปได้ว่า ค่าพารามิเตอร์ของข้อสอบ (Item Parameter) ได้แก่ ค่าความยาก (b_g) ค่าอำนาจจำแนก (a_g) และค่าสัมประสิทธิการเดา (c_g) ส่วนค่าพารามิเตอร์ของผู้สอบ (Examinee Parameter) ได้แก่ ค่าความสามารถของผู้สอบ (θ) พิสัยของค่าพารามิเตอร์ต่าง ๆ มีดังนี้

b_g มีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติจะมีค่าระหว่าง -2 ถึง +2 ค่า -2 แสดงว่าข้อสอบง่ายมาก ค่า +2 แสดงว่าข้อสอบยากมาก

a_g มีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติ ค่า a_g ที่เป็นลบ แสดงว่าข้อสอบใช้ไม่ได้ จึงตัดข้อนั้นทิ้งไป จึงมักพบ a_g ที่มีค่า 0 ถึง +2 ค่า $a_g = 0$ แสดงว่าข้อสอบไม่มีอำนาจจำแนกหรืออำนาจจำแนกต่ำ ค่า +2 แสดงว่าข้อสอบมีอำนาจจำแนกสูง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

θ ในทางปฏิบัติมีค่าอยู่ระหว่าง -3 ถึง +3 ถ้าค่อนมาทาง -3 แสดงว่ามีความสามารถต่ำ และค่อนไปทาง +3 แสดงว่ามีความสามารถสูง

c_g คือค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำมาก มีโอกาสจะทำข้อสอบข้อนี้ได้ถูกต้อง

สำเริง บุญเรืองรัตน์ (2527 หน้า 104 - 109) อธิบายการประมาณค่าพารามิเตอร์ของรูปแบบโลจิสติคไว้ซึ่งผู้วิจัยเรียบเรียงข้อความและขั้นตอนใหม่ สรุปได้ดังนี้

1. คัดเลือกข้อสอบที่ทุกคนตอบถูกและข้อสอบที่ทุกคนตอบผิดออกจากแบบทดสอบ

2. ตรวจดูว่าข้อสอบที่เหลือมีคนทำถูกทุกข้อหรือผิดทุกข้ออยู่หรือไม่ ถ้ามีให้คัดออกจากกลุ่มตัวอย่าง

ข้อสอบที่เหลือ เรียกว่า เป็นข้อสอบที่สามารถให้ข้อมูลได้ ซึ่งก็คือ ข้อสอบที่มีความแปรปรวนนั่นเอง

3. ตรวจให้คะแนนแล้วจัดกลุ่มผู้เข้าสอบที่ได้คะแนนรวมเท่ากันไว้เป็นกลุ่มความสามารถเดียวกัน ถ้ากลุ่มใดมีคนน้อยกว่า 20 คน ควรรวมเข้ากับคะแนนถัดไปให้เป็นกลุ่มเดียวกัน ให้จำนวนคนในกลุ่ม i ทั้งหมดมี m_i คน

4. แจงนับจำนวนคนแต่ละกลุ่มที่ตอบถูกในข้อสอบที่จะวิเคราะห์ให้ n_i เป็นจำนวนคนที่ตอบถูกในกลุ่ม i

5. คำนวณหาสัดส่วนคนที่ตอบถูกต่อคนทั้งหมดในกลุ่ม สมมุติให้เป็น P_i

$$P_i = n_i / m_i$$

6. นำค่า P_i ของทุกกลุ่มในข้อสอบข้อเดียวกัน ไปพล็อตกราฟ โดยให้แกนนอน (แกน x) ใช้อ่านค่าคะแนนรวม และแกนตั้ง (แกน y) ใช้อ่านค่า P_i ลากเส้นเชื่อมจุดต่าง ๆ จากคะแนนรวมน้อยไปยังคะแนนรวมมากตามลำดับ เส้นกราฟที่ได้จะเป็นโค้งลักษณะข้อสอบข้อนั้น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

7. ลากเส้นตรงจากจุด P = .5 บนแกนตั้ง ให้ขนานกับแกนนอนตัดเส้นกราฟที่จุดใด ลากเส้นจากจุดตัดไปตั้งฉากกับแกนนอน ตำแหน่งจุดตัดบนแกนนอน คือ ค่าความยากของข้อสอบเป็นคะแนนซึ่งถ้าปรับคะแนนนี้ให้เป็นมาตราฐานตามค่า θ ก็จะเป็นค่า b_g ของข้อสอบข้อนี้

8. ค่าความชันของกราฟ ณ จุดเปลี่ยนโค้งคูณกับ $\sqrt{2\pi}$ คือ ค่าอำนาจจำแนกของข้อสอบที่จำแนกกลุ่มบุคคลได้ระหว่างผู้มีความสามารถต่ำกว่าค่าความยากง่ายและผู้มีความสามารถสูงกว่าค่าความยากง่าย

ในการวิเคราะห์ข้อสอบจริง ๆ จะใช้คอมพิวเตอร์ในการคำนวณโดยใช้โปรแกรมสำเร็จรูปชื่อ LOGIST ซึ่งพัฒนาโปรแกรมโดย วู้ด วิงเกอร์สกี้ และ ลอร์ด กระบวนการประมาณค่าพารามิเตอร์ในโปรแกรม LOGIST สรุปได้เป็น 4 ขั้นดังนี้ (Wood, Wingersky and Lord, 1976 : 4)

ขั้นที่ 1 ประมาณค่าความสามารถของผู้สอบแต่ละคนและค่าความยากของข้อสอบแต่ละข้อโดยกำหนดให้ค่าอำนาจจำแนกและค่าสัมประสิทธิ์การเดาเป็นค่าคงที่ตามที่กำหนดค่าเริ่มต้นให้

ขั้นที่ 2 ประมาณค่าพารามิเตอร์ของข้อสอบแต่ละข้อทั้งสามพารามิเตอร์ โดยให้ค่าความสามารถของแต่ละคนเท่ากับค่าความสามารถที่ได้จากขั้นที่ 1

ขั้นที่ 3 กระทำซ้ำในขั้นที่ 1 ด้วยการกำหนดให้ค่าอำนาจจำแนกและค่าสัมประสิทธิ์การเดาเป็นค่าคงที่ตามที่คำนวณได้จากขั้นที่ 2

ขั้นที่ 4 ประมาณค่าอำนาจจำแนกและค่าความยากของข้อสอบโดยให้ค่าความสามารถและค่าสัมประสิทธิ์การเดามีค่าคงที่

กระบวนการประมาณค่าทั้ง 4 ขั้นใช้วิธีการแก้สมการเชิงเส้นของนิวตัน (Newton's method) ในแต่ละขั้นจะยุติลงก็ต่อเมื่อค่าพารามิเตอร์ทั้งหลายที่ได้จากการคำนวณแต่ละข้อแต่ละคนเป็นคอนเวอร์เจนซ์ (Convergence) ตามเกณฑ์ที่กำหนดในแต่ละรอบ จากนั้นกระบวนการนี้จะถูกกระทำซ้ำโดยลดเปอร์เซนต์ความแตกต่างที่จะยอมรับว่าเป็นคอนเวอร์เจนซ์ลง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

การวิเคราะห์ข้อสอบถ้าต้องการผลที่ดีควรใช้กลุ่มตัวอย่างไม่ต่ำกว่า 1000 คน และมีข้อสอบในฉบับที่วิเคราะห์ไม่น้อยกว่า 40 ข้อ ผลการจากวิเคราะห์ข้อสอบโดยใช้โปรแกรม LOGIST จะได้ทั้งค่าพารามิเตอร์ของข้อสอบ คือ ค่าความยาก ค่าอำนาจจำแนก และค่าสัมประสิทธิ์การเดา และค่าพารามิเตอร์ของผู้สอบ คือ ค่าความสามารถของผู้สอบ

การพิจารณาคุณภาพของแบบทดสอบจะเป็นการพิจารณาความแน่นอนในการประมาณค่าความสามารถที่แท้จริงซึ่งพิจารณาได้ฟังก์ชั่นสาระประโยชน์ของแบบทดสอบ (Test Information Function) ที่ได้จากผลรวมของฟังก์ชั่นสาระประโยชน์ของข้อสอบแต่ละข้อ (Item Information Function)

ฟังก์ชั่นสาระประโยชน์ของข้อสอบแต่ละข้อ กำหนดได้จากสมการ

$$I_g\{\theta,U_g\} = (P'_g)^2 / P_g Q_g , \quad g = 1,2,\ldots,n$$

ในที่นี้ $I_g\{\theta,U_g\}$ คือ ฟังก์ชั่นสาระประโยชน์ของข้อสอบแต่ละข้อ

P_g คือ ความน่าจะเป็นที่ผู้สอบที่มีความสามารถ θ จะตอบข้อสอบข้อ g ได้ถูกต้อง

Q_g คือ $(1-P_g)$

P'_g คือ ความชันของโค้งลักษณะข้อสอบที่ระดับความสามารถ θ

$$= \frac{1.7a_g(1-c_g)\exp(1.7a_g(\theta-b_g))}{[1+\exp(1.7a_g(\theta-b_g))]^2}$$

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

และถ้าให้ $I(\theta)$ เป็นฟังก์ชั่นสาระประโยชน์ของแบบทดสอบ จะได้ว่า

$$I(\theta) = \sum I_i(\theta, U_i)$$

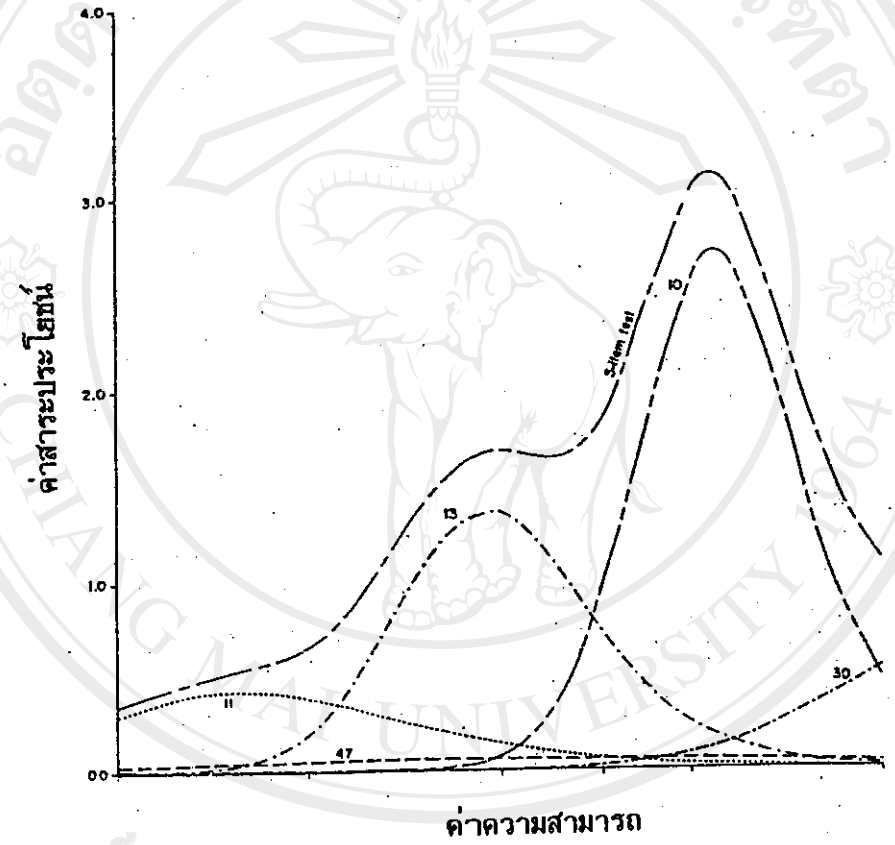
$$= \sum \frac{(P_i')^2}{P_i Q_i}$$

(Birnbuam in Lord and Novick, 1968 : 454)

ถ้าเรามีกลุ่มข้อสอบที่ทราบโค้งสาระประโยชน์ เราสามารถสร้างแบบทดสอบให้มีโค้งสาระประโยชน์ ณ ระดับของความสามารถที่เราต้องการได้โดยคัดเลือกข้อสอบที่เหมาะสมมาแล้วคำนวณหาค่าฟังก์ชั่นสาระประโยชน์ ณ ระดับความสามารถนั้น ดังภาพที่ 3

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 3 แสดงฟังก์ชั่นค่าสาระประโยชน์ของข้อสอบ 5 ข้อและแบบทดสอบที่ประกอบด้วยข้อสอบ 5 ข้อนั้น



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.2 การประมาณค่าความสามารถ

การประมาณค่าความสามารถของผู้สอบจากสมการแสดงความสัมพันธ์ของพารามิเตอร์ในรูปความน่าจะเป็น ที่อ้างถึงในตอนต้นของหัวข้อ 2.1 และทราบค่า a_g, b_g และ c_g ของข้อสอบ ถ้าคัดเลือกข้อสอบที่สามารถให้ค่าสาระประโยชน์สูงสุด ณ ระดับความสามารถ θ แล้วค่า θ คำนวณได้จากสมการ (Birnbuam, 1968 in Lord and Novick, 1968 : 463)

$$\theta = b_g + \frac{1}{Da_g} \ln \frac{1 + (1 + 8c_g)^{1/2}}{2}$$

เนื่องจากโอกาสที่จะคัดเลือกข้อสอบให้เป็นข้อสอบที่สามารถให้ค่าสาระประโยชน์สูงสุด ณ ระดับ θ มีน้อยมากและจำนวนข้อสอบในกลุ่มข้อสอบ (Item pool) หนึ่ง ๆ จะมีจำกัด ดังนั้น การเลือกข้อสอบไปใช้จึงทำได้เพียงเลือกข้อที่ดีที่สุดเท่าที่จะหาได้จากกลุ่มข้อสอบเท่านั้น ซึ่งไม่จำเป็นว่าจะต้องเป็นข้อสอบที่ให้ค่าสาระประโยชน์สูงสุด การประมาณค่าความสามารถจึงไม่สามารถใช้สูตรดังกล่าวข้างต้นได้ ในตอนต่อไปนี้จะได้นำเสนอวิธีการประมาณค่าความสามารถ 3 วิธีโดยสรุป ดังนี้

ก. **การประมาณค่าความสามารถโดยวิธีความเป็นไปได้สูงสุด** (Maximum Likelihood)

การประมาณค่าโดยวิธีความเป็นไปได้สูงสุด เป็นวิธีการทางสถิติวิธีหนึ่งที่นักสถิติจะนำมาใช้ประมาณค่าพารามิเตอร์ ในกรณีที่ ไม่มีค่าสถิติที่พอเพียง (no sufficient statistic exist) การวิเคราะห์ข้อสอบตามรูปแบบโลจิสติคที่มี 3 พารามิเตอร์ก็เช่นกัน ถ้าเราเชื่อว่ามีการเดาข้อสอบได้ ($c_g \neq 0$) แล้วก็จะไม่มีค่าสถิติที่พอเพียงสำหรับการประมาณค่าความสามารถ (θ) (Lord, 1980 : 58) ดังนั้นจึงต้องประมาณค่าความสามารถของผู้ตอบข้อสอบด้วยวิธีความเป็นไปได้สูงสุด

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ถ้าให้ $\hat{\theta}$ เป็นค่าความสามารถของผู้ตอบข้อสอบ

a_g, b_g, c_g ($g = 1,2,3,\ldots,n$) เป็นค่าอำนาจจำแนก ค่าความยาก และค่าสัมประสิทธิ์การเดาของข้อสอบข้อที่ g

P_g คือ ความน่าจะเป็นที่ผู้ตอบข้อสอบจะตอบข้อสอบข้อ g ได้ถูกต้อง

Q_g คือ $(1-P_g)$

U_g คือ ผลการตอบข้อสอบข้อที่ g ของผู้ตอบข้อสอบค่า $U_g = 1$ หรือ 0

$$P'_g = \frac{\partial P_g}{\partial \hat{\theta}} = \frac{Da_g}{1 - c_g} Q_g (P_g - c_g) \quad \text{และ}$$

$$\frac{P'_g}{Q_g} = \frac{Da_g}{1 + e_g^{-DL}} \quad , \quad \text{เมื่อ} \quad L_g = a_g (\hat{\theta} - b_g)$$

จะสามารถหาค่า $\hat{\theta}$ ได้จากการแก้สมการ

$$\sum_{g=1}^{n} \frac{U_g - P_g}{P_g Q_g} \cdot \frac{\partial P_g}{\partial \hat{\theta}} = 0 \qquad \text{หรือ}$$

$$\sum_{g=1}^{n} \frac{U_g - P_g}{P_g} \cdot \frac{P'_g}{Q_g} = 0$$

(Lord, 1980 : 60 - 61 , 179 - 180)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ในกรณีที่จำนวนข้อสอบมีน้อยข้อ การประมาณค่าความสามารถโดยแก้สมการดังกล่าวข้างต้นจะได้ค่า $\hat{\theta}$ มากกว่า 1 ค่า (Samejima, 1973) แต่ก็ไม่เคยพบว่ามีมากกว่า 1 ค่าถ้ามีจำนวนข้อสอบมากกว่าหรือเท่ากับ 20 ข้อ ซึ่งก็เป็นไปตามทฤษฎีของฟูทซ์ (Foutz, 1977) ที่ว่า "ถ้าจำนวนข้อสอบมีมากข้อพอแล้ว สมการความเป็นไปได้สูงสุดจะให้ค่ารากของสมการเพียงค่าเดียวและเป็นค่าพารามิเตอร์ที่แท้จริง" (Lord, 1980 : 59)

ข. <u>การประมาณค่าความสามารถโดยวิธีของเบย์</u> (Baysian Updating)

การประมาณค่าความสามารถวิธีของเบย์ เสนอโดย โอเวน (Roger J. Owen, 1975) เป็นวิธีที่เหมาะสมกับกระบวนการทดสอบแบบเทเลอร์ เพราะสามารถประมาณค่าความสามารถจากการตอบข้อสอบทีละข้อ และยังสามารถคำนวณหาค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถได้อีกด้วย ถ้าข้อสอบแต่ละข้อมีความยากใกล้เคียงกับระดับความสามารถของผู้สอบค่าความคลาดเคลื่อนในการประมาณค่าความสามารถจะลดลง ซึ่งถ้าเรากำหนดระดับความคลาดเคลื่อนไว้ ณ ระดับใดระดับหนึ่งก็สามารถใช้เป็นจุดยุติการสอบได้ โดยที่เราสามารถเชื่อมั่นในผลการประมาณค่าได้ว่ามีความคลาดเคลื่อนไม่เกินระดับที่คาดหวังไว้ก่อน

กระบวนการตามวิธีของเบย์ มีดังนี้

1. ประมาณค่าความสามารถเริ่มต้นและความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ ถ้ามีข้อมูลมาก่อน แต่ถ้าไม่มีข้อมูล อาจกำหนดเป็นข้อตกลงว่าผู้สอบมีค่าความสามารถเริ่มต้นเป็นเท่าไร และความคลาดเคลื่อนมาตรฐานเป็นเท่าไร เช่น ใช้ค่าความสามารถเฉลี่ย ($\theta = 0$) หรือค่าความสามารถที่เป็นเกณฑ์เป็นค่าเริ่มต้น และความคลาดเคลื่อนมาตรฐานเมื่อเริ่มต้นมีค่า = 1.00 (Urry, 1977 : 181)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

2. คัดเลือกข้อสอบที่มีค่าความยากใกล้เคียงกับระดับความสามารถของผู้สอบที่ทราบมาก่อนหรือที่กำหนดไว้ก่อนไปให้ผู้สอบตอบ นำผลการตอบข้อสอบ (ถูก หรือ ผิด) ไปประมาณค่าความสามารถ และค่าความคลาดเคลื่อนมาตราฐาน โดยใช้สูตรดังนี้ (Owen, 1975 : 353 and Throndike, 1982 :303)

ในกรณีที่ตอบข้อสอบข้อนั้นได้ถูกต้อง

$$\Theta_{m+1} = \Theta_m + (1 - c_g) \left[\frac{\sigma_m^2}{\sqrt{1/a^2 + \sigma_m^2}} \right] \left[\frac{\phi (D)}{c_g + (1 - c_g) \Phi(-D)} \right]$$

และ

$$\sigma_{m+1}^2 = \sigma_m^2 \left[1 - \left[\frac{1 - c_g}{1+1/a_g^2 \sigma_m^2} \right] \left[\frac{\phi (D)}{A} \right] \left[\frac{(1 - c_g) \phi (D)}{A} - D \right] \right]$$

ในกรณีที่ตอบข้อสอบข้อนั้นผิด

$$\Theta_{m+1} = \Theta_m - \left[\frac{\sigma_m^2}{\sqrt{1/a_g^2 + \sigma_m^2}} \right] \left[\frac{\phi (D)}{\Phi (D)} \right]$$

และ

$$\sigma_{m+1}^2 = \sigma_m^2 \left[1 - \left[\frac{\phi (D)}{1 + 1/a_g^2 \sigma_m^2} \right] \left[\frac{\phi (D)}{\Phi (D)} + D \right] \Big/ \Phi (D) \right]$$

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

เมื่อ θ_m คือ ค่าประมาณความสามารถก่อนตอบข้อสอบข้อที่ m+1

θ_{m+1} คือ ค่าประมาณความสามารถจากผลการตอบข้อสอบข้อที่ m+1

a_g คือ ค่าอำนาจจำแนกแบบโลจิสติค ของข้อที่ m+1

b_g คือ ค่าความยากแบบโลจิสติค ของข้อที่ m+1

c_g คือ ค่าสัมประสิทธิ์การเดาแบบโลจิสติค ของข้อที่ m+1

σ_m^2 คือ ค่าความแปรปรวนในการประมาณค่าความสามารถก่อนตอบข้อสอบข้อที่ m+1

σ_{m+1}^2 คือ ค่าความแปรปรวนในการประมาณค่าความสามารถเมื่อตอบข้อสอบข้อที่ m+1 แล้ว

$\phi(D)$ คือ ค่า Ordinate ของโค้งปกติ ณ จุด D (pdf. at point D)

$\Phi(D)$ คือ พื้นที่ใต้โค้งปกติถึงจุด D (cdf. up to point D)

$$D = \frac{b_g - \theta_g}{\sqrt{1 / a_g^2 + \sigma_m^2}}$$ และ

$$A = c_g + (1 - c_g) \Phi (-D)$$

ค่า σ_{m+1} คือ ค่าความเบี่ยงเบนมาตรฐาน (Standard deviation) ของการประมาณค่าความสามารถซึ่งได้จากการถอดกรณฑ์ที่สองของความแปรปรวน ธอร์นไดค์ (Thorndike, 1982 : 305) เรียกว่า เป็นค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard error of estimate) หรือพิสัยของความไม่แน่นอน (range of uncertainty) และเออรี่ (Urry, 1977 : 183) อธิบายว่าค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถนี้ก็คือ ค่าความเบี่ยงเบนมาตรฐานของการประมาณค่าความสามารถรอบเส้นภาระถดถอย ณ ระดับความสามารถจริงที่คงที่ค่าหนึ่ง ไม่ใช่ค่าความคลาดเคลื่อนมาตรฐานของการวัด

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

ค. การประมาณค่าความสามารถโดยวิธีของเบทซ์และไวส์

(Betz and Weiss's Method)

เบทซ์และไวส์ (Betz and Weiss 1973, 1974) ได้ปรับปรุงวิธีการจากการศึกษาของลอร์ดที่ใช้วิธีความเป็นไปได้สูงสุด เพื่อให้สามารถประมาณค่าความสามารถได้จากการตอบแบบทดสอบทั้งฉบับง่ายขึ้น (Larkin and Weiss, 1975 :10) โดยมีสูตรดังนี้

$$\hat{\theta} = \frac{1}{\bar{a}_i} \Phi^{-1}\left[\frac{(x/m - c)}{1 - c}\right] + \bar{b}_i$$

เมื่อ $\bar{a}_i$ คือ ค่าอำนาจจำแนกเฉลี่ยของข้อสอบที่ให้ตอบ

$\bar{b}_i$ คือ ค่าความยากเฉลี่ยของข้อสอบที่ให้ตอบ

c คือ โอกาสในการเดาข้อสอบได้ถูกซึ่งขึ้นอยู่กับจำนวนตัวเลือกในข้อสอบ เช่น ถ้ามี 5 ตัวเลือก c = .20

x คือ จำนวนข้อที่ตอบถูก

m คือ จำนวนข้อของข้อสอบที่นักเรียนได้ตอบ

Φ^{-1} คือ อินเวอร์สของฟังก์ชั่นการกระจายปรกติสะสมที่สัมพันธ์กับสัดส่วนการตอบถูก

$\hat{\theta}$ คือ ค่าความสามารถ

ในกรณีที่ผู้สอบตอบถูกทุกข้อ (x = m) ค่าความสามารถ o ไม่สามารถคำนวณได้ ในกรณีนี้จะต้องแทนค่า x ด้วย x = m - .5 สำหรับผู้สอบได้คะแนนต่ำกว่าหรือเท่ากับโอกาสในการเดาถูกคือ $x \leq cm$ ค่า $\hat{\theta}$ จะเป็น indeterminate ต้องแทนค่า x ด้วย cm + .5

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

3. การทดสอบแบบเทเลอร์

การทดสอบแบบเทเลอร์เป็นกระบวนการใช้แบบทดสอบซึ่งข้อสอบแต่ละข้อถูกเลือกมาใช้บนพื้นฐานของคำตอบของผู้เข้าสอบ (Weiss and Kingsbury, 1984 : 361) ความคิดพื้นฐานสำหรับการทดสอบแบบเทเลอร์ ก็คือ การคัดเลือกข้อสอบที่เหมาะสมกับระดับความสามารถของผู้สอบมากกว่าที่จะให้เหมาะสมกับความสามารถเฉลี่ยของกลุ่ม การจัดข้อสอบจะยึดผลการตอบของผู้สอบแต่ละคนโดยเลือกข้อที่มีค่าความยากในช่วง P = .50 ที่สัมพันธ์กับระดับความสามารถที่ประมาณได้ของผู้เข้าสอบแต่ละคน (Weiss, 1974 :2)

ไวส์และคิงส์เบอรี่ (Weiss and Kingsbury, 1984 : 362 - 364) กล่าวถึง ส่วนประกอบของกระบวนการการทดสอบแบบเทเลอร์ไว้ สรุปได้ดังนี้

1. เป็นข้อสอบตามทฤษฎีการตอบข้อคำถาม (Item Response Theory)

2. กลุ่มข้อสอบ (item pool) มีจำนวนข้อมากพอและมีค่าพารามิเตอร์กระจายเต็มพิสัยของระดับคุณลักษณะในประชากร ผลการวัดจะมีประสิทธิภาพมากถ้าใช้ข้อสอบที่มีอำนาจจำแนกสูง

3. ข้อสอบข้อแรกที่จะสอบต้องเลือกให้เหมาะสมกับระดับความสามารถของผู้เข้าสอบ เป็นการช่วยลดจำนวนข้อของการทดสอบให้น้อยลง

4. กฎการเลือกข้อสอบจะต้องเลือกใช้ข้อที่สามารถให้สาระประโยชน์สูงสุด ณ ระดับความสามารถของผู้เข้าสอบ

5. วิธีให้คะแนนเพื่อกำหนดระดับคุณลักษณะของผู้เข้าสอบแต่ละคน สามารถประมาณได้ด้วยวิธีความเป็นไปได้สูงสุด (Maximum Likelihood) หรือวิธีของเบย์ (Bayesian Methods)

6. เกณฑ์การสิ้นสุดการทดสอบขึ้นอยู่กับจุดมุ่งหมายของการทดสอบ โดยพิจารณาว่าการวัดผลแต่ละบุคคลนั้น ช่วยให้สามารถกำหนดระดับความสามารถของเขาได้อย่างแม่นยำอย่างพอเพียงแล้ว

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ไวส์ (Weiss, 1974 : 1 - 67) กล่าวถึงยุทธวิธีในการจัดแบบทดสอบแบบเทเลอร์ว่ามี 2 ยุทธวิธีใหญ่ ๆ คือ

1. ยุทธวิธีสองขั้น (Two - Stage Strategies)

2. ยุทธวิธีหลายขั้น (Multi - Stage Strategies) ซึ่งแบ่งเป็นแบบจำลอง 2 กลุ่มได้แก่

2.1. แบบจำลองการแยกทางคงที่ (Fixed - Branching Models) ประกอบด้วยแบบจำลองย่อย 3 รูปแบบ ได้แก่ แบบจำลองรูปปิรามิด (Pyramidal Models) แบบทดสอบเปลี่ยนระดับ (Flexilevel Test) และแบบทดสอบปรับระดับขั้น (Stradative Test) ในแบบจำลองรูปแบบปิรามิดยังแบ่งเป็นแบบจำลองย่อยได้อีก 5 ชนิด ได้แก่ แบบที่ใช้ขนาดขั้นคงที่ (Constant Step Size Pyramid) แบบที่ใช้ขนาดขั้นแปรผัน (Variable or Decreasing Step Size Pyramids) แบบรูปปิรามิดข้างตัด (Truncated Pyramids) แบบที่มีข้อสอบหลายข้อในแต่ละขั้น (Multiple Item Pyramids) และแบบการให้น้ำหนักแก่ตัวเลือกของข้อสอบแยกทาง (Differential Response Option Branching)

2.2 แบบจำลองการแยกทางแปรผัน (Variable Branching Models) ประกอบด้วยแบบจำลองย่อย 2 แบบ ได้แก่ ยุทธวิธีของเบย์ (Baysian Strategies) กับยุทธวิธีความเป็นไปได้สูงสุด (Maximum Likelihood Strategies)

ไรท์และสโตน (Wright and Stone, 1979 : 151) กล่าวถึงการทดสอบแบบเทเลอร์ไว้ 3 รูปแบบคือ การทดสอบแบบเทเลอร์ที่จัดตามสภาพเดิม (Status Tailoring) การทดสอบแบบเทเลอร์ที่ให้มีการทดลองทำ (Performance Tailoring) และการทดสอบแบบเทเลอร์ที่ให้ตัดสินใจเลือกด้วยตัวเอง (Self - Tailoring) ซึ่งเมื่อพิจารณาในรายละเอียดแล้ว ก็คือการทดสอบแบบเทเลอร์ตามยุทธวิธีสองขั้นที่นำมาแยกแยะตามวิธีการกำหนดความสามารถเริ่มต้นของผู้เข้าสอบนั่นเอง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

การทดสอบแบบเทเลอร์ดังกล่าวสรุปได้ดังนี้

1. **การทดสอบแบบเทเลอร์ยุทธวิธีสองขั้น**

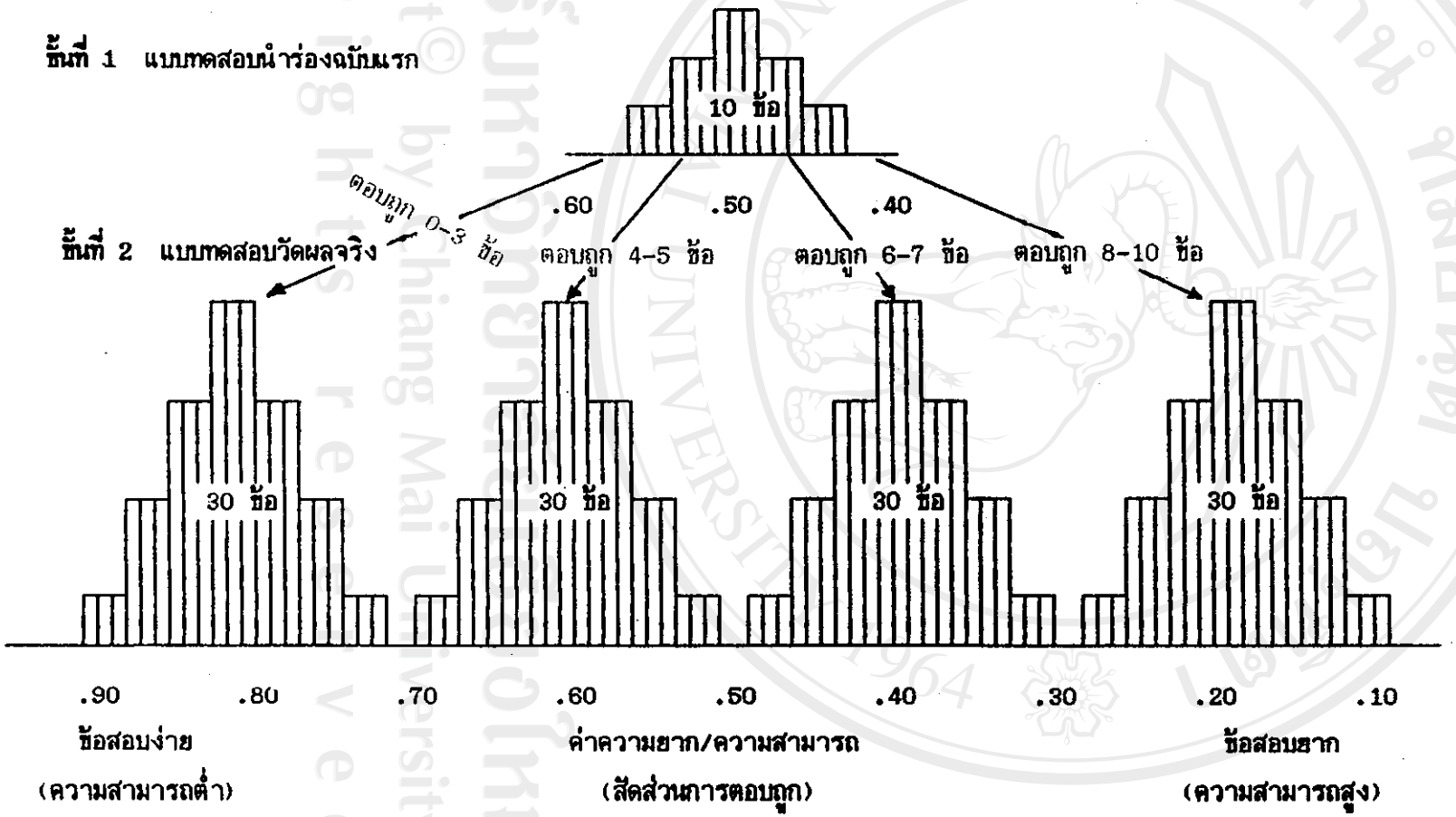
เมื่อยึดหลักการเบื้องต้นของการทดสอบแบบเทเลอร์ที่ว่า "ผู้ดำเนินการสอบต้องทราบระดับความสามารถเดิมของผู้เข้าสอบก่อน จึงจะสามารถจัดข้อสอบให้เหมาะสมกับระดับความสามารถได้" จะสามารถแบ่งการทดสอบแบบเทเลอร์ยุทธวิธีสองขั้นได้เป็น 3 รูปแบบ คือ

1.1 **ยุทธวิธีสองขั้นที่จัดตามสภาพเดิม** (Status Tailoring) เป็นการจัดแบบทดสอบที่อาศัยข้อมูลความสามารถเดิมของผู้เข้าสอบมาพิจารณาว่าควรให้ผู้เข้าสอบได้ตอบข้อสอบที่มีความยากระดับใด ข้อมูลดังกล่าว เช่น ระดับชั้น ลำดับที่ความสามารถในกลุ่ม อายุ ผลการสอบเก่า เป็นต้น จากข้อมูลเดิมที่ช่วยชี้นำ ผู้ดำเนินการสอบจะคัดเลือกข้อสอบที่เหมาะสมกับระดับความสามารถให้ผู้เข้าสอบทำ

1.2 **ยุทธวิธีสองขั้นที่ให้มีการทดลองทำ** (Performance Tailoring) เป็นแบบที่เหมาะกับกรณีที่มีข้อมูลเดิมเกี่ยวกับตัวผู้เข้าสอบไม่เพียงพอ แต่เนื่องจากต้องกำหนดระดับความสามารถเริ่มต้นจึงจัดให้มีการทดสอบนำร่อง (Routing or Pilot Test) ด้วยแบบทดสอบจำนวนหนึ่งแล้วใช้ผลการสอบนำร่องนี้ไปช่วยกำหนดแบบทดสอบฉบับที่จะใช้วัดผลจริง (Measurement Test) การจัดแบบทดสอบนำร่องอาจกระทำได้หลายลักษณะ ได้แก่ สอบนำร่องด้วยแบบทดสอบที่มีความยากปานกลางทั้งฉบับ แล้วนำผลการสอบนำร่องไปกำหนดระดับความยากของแบบทดสอบฉบับวัดผลจริง ดังภาพที่ 4 หรือ สอบนำร่อง ด้วยแบบทดสอบที่มีความยากกระจายจากน้อยไปหามาก แล้วนำผลการตอบไปกำหนดระดับความยากของแบบทดสอบฉบับวัดผลจริง ดังภาพที่ 5 หรือ สอบนำร่องด้วยแบบทดสอบนำร่องฉบับที่ 1 นำผลไปกำหนดความยากของแบบทดสอบนำร่องฉบับที่ 2 แล้วนำผลการสอบนำร่องฉบับที่ 2 ไปกำหนดความยากของแบบทดสอบฉบับวัดผลจริงดังภาพที่ 6

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

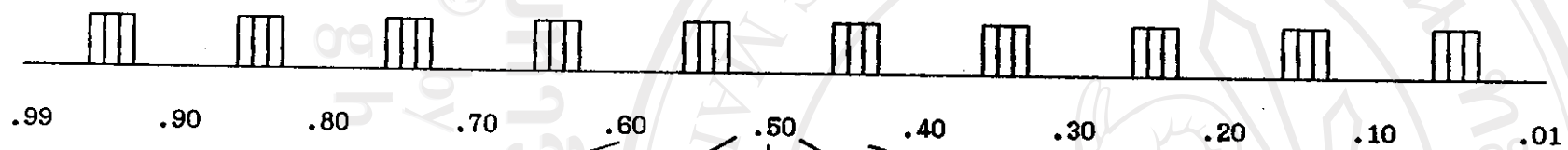
ภาพที่ 4 แสดงการจัดแบบทดสอบแบบเทเลอร์ยุทธวิธีสองขั้นที่สอบนำร่องด้วยแบบทดสอบที่มีความยากปานกลาง



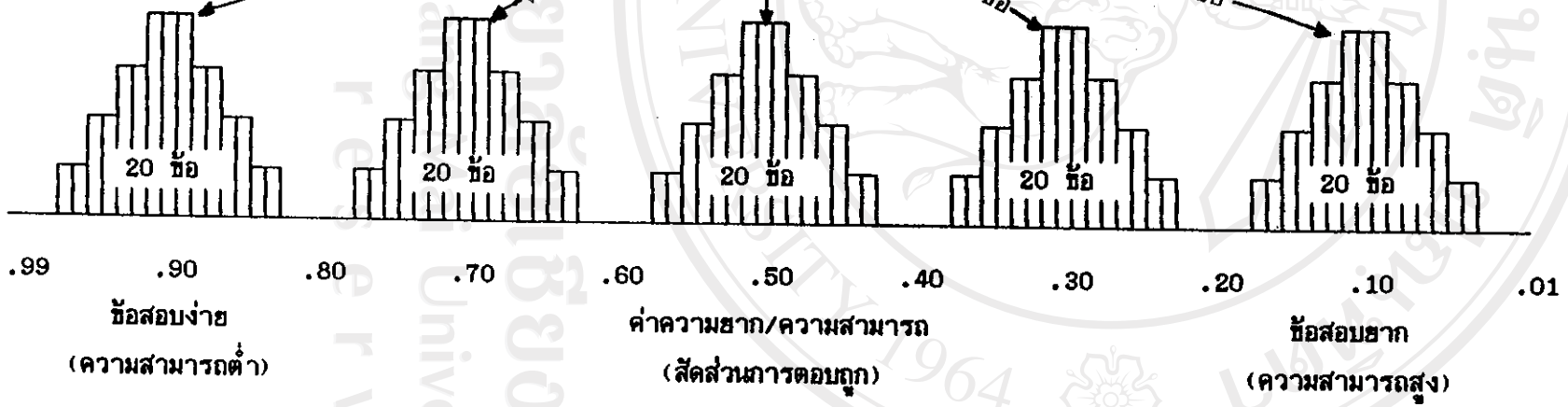
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 5 แสดงการจัดแบบทดสอบแบบเทเลอร์ยุทธวิธีสองขั้นที่สอบนำร่องด้วยแบบทดสอบที่มีความยากกระจาย

ขั้นที่ 1 แบบทดสอบนำร่อง

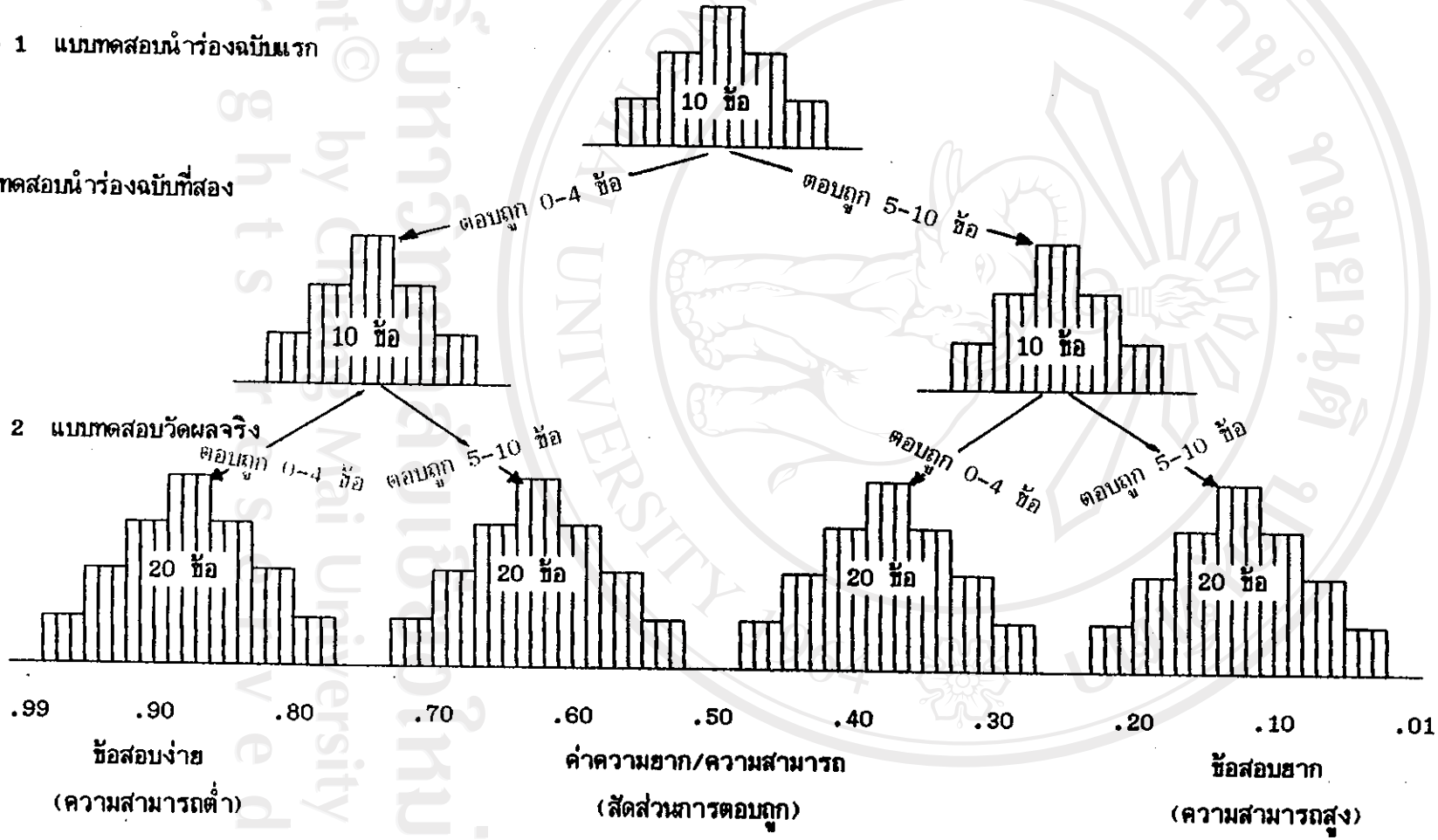


ขั้นที่ 2 แบบทดสอบวัดผลจริง



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 6 แสดงการจัดแบบทดสอบแบบเทเลอร์ยุทธวิธีสองขั้นที่ใช้การสอบนำร่องสองครั้ง



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ลอร์ด (Lord, 1977 : 126) กล่าวถึง การทดสอบแบบนี้ว่าจะต้องคำนึงถึงสิ่งต่าง ๆ ดังนี้

1. ความยาวของแบบทดสอบนำร่อง
2. ความยาวของแบบทดสอบฉบับวัดผลจริง
3. จำนวนฉบับของแบบทดสอบในขั้นวัดผลจริง
4. ระดับความยากของแบบทดสอบนำร่อง
5. ระดับความยากของแบบทดสอบแต่ละฉบับในขั้นวัดผลจริง
6. วิธีให้คะแนนการทดสอบนำร่อง
7. คะแนนจุดตัดจากการทดสอบนำร่องเพื่อจัดการสอบในขั้นวัดผลจริง
8. วิธีการให้คะแนนการทดสอบขั้นวัดผลจริง
9. วิธีรวมคะแนนจากการสอบนำร่องกับการสอบวัดผลจริง

1.3 <u>ยุทธวิธีสองขั้นที่ให้ตัดสินใจเลือกด้วยตนเอง</u> (Self - Tailoring) เป็นแบบที่ให้ผู้เข้าสอบพิจารณาว่าแบบทดสอบที่จัดเรียงลำดับข้อไว้จากง่ายไปหายากนั้นตนเองสนใจหรือเหมาะที่จะเริ่มทำข้อสอบข้อใดจึงจะใกล้เคียงกับความสามารถของตน แต่ง่ายพอที่ตนจะทำได้ เมื่อเลือกได้แล้วการทดสอบก็จะเริ่มต้น ณ ข้อสอบข้อนั้น และจัดข้อสอบข้อที่ยากขึ้นตามลำดับให้เป็นข้อต่อ ๆ ไป การทดสอบจะยุติเมื่อมั่นใจว่าผู้เข้าสอบจะไม่สามารถตอบได้ถูกต้องอีกในข้อต่อไป เช่น ยุติการสอบเมื่อตอบผิด 3 ข้อติดต่อกันเป็นต้น จากนั้นจึงนำผลการตอบไปประมาณค่าความสามารถต่อไป

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2. <u>การทดสอบแบบเทเลอร์ยุทธวิธีหลายขั้น</u>

มีรูปแบบต่าง ๆ จำแนกได้ดังนี้

2.1 แบบจำลองการแยกทางคงที่ (Fixed - Branching Model)
มี 3 รูปแบบ

2.1.1 แบบจำลองรูปปิรามิด (Pyramidal Models) มี 5 รูปแบบย่อย

2.1.1.1 แบบปิรามิดที่ใช้ขนาดขั้นคงที่
(Constant Step Size Pyramids)

2.1.1.2 แบบปิรามิดที่ใช้ขนาดขั้นแปรผัน
(Variable or Decreasing Step Size Pyramids)

2.1.1.3 แบบปิรามิดช้างตัด
(Truncated pyramids)

2.1.1.4 แบบปิรามิดที่มีหลายข้อในแต่ละขั้น
(Multiple - Items Pyramids)

2.1.1.5 แบบปิรามิดที่ให้น้ำหนักแก่ตัวเลือกของข้อสอบแยกทาง
(Differential Response Option Branching)

2.1.2 แบบจำลองเปลี่ยนระดับ (Flexilevel Test)

2.1.3 แบบจำลองปรับระดับขั้น (Stradative Test)

2.2 แบบจำลองการแยกทางแปรผัน มี 2 รูปแบบ

2.2.1 ยุทธวิธีของเบย์ (Baysian Strategies)

2.2.2 ยุทธวิธีความเป็นไปได้สูงสุด (Maximum Likelihood Strategies)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

แบบจำลองรูปปิรามิดที่ใช้ขนาดขั้นคงที่

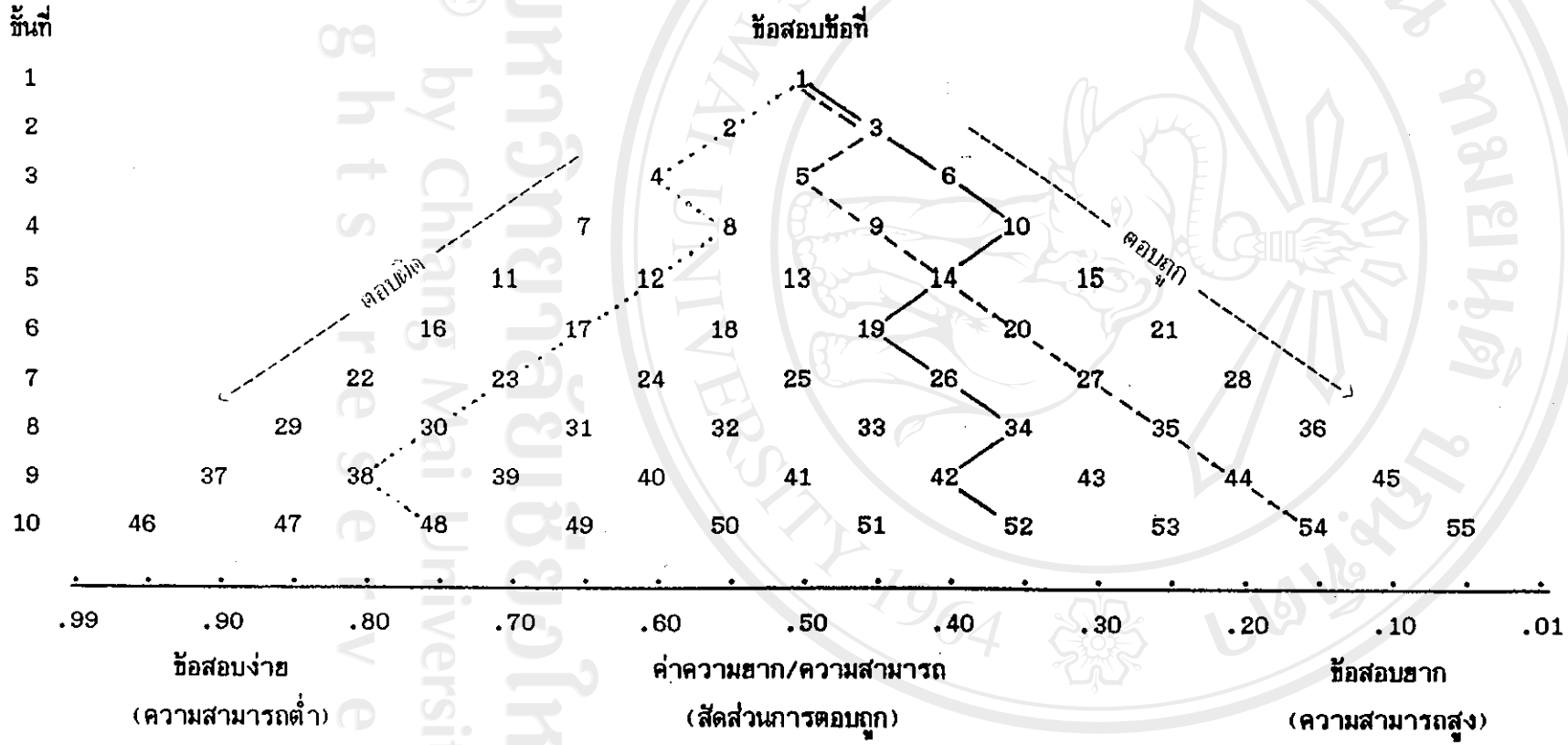
แบบทดสอบที่จัดตามแบบจำลองรูปปิรามิดที่ใช้ขนาดขั้นคงที่ จะมีจำนวนข้อสอบในแต่ละขั้นเท่ากับลำดับที่ของขั้น ดังภาพที่ 7 เป็นตัวอย่างขนาด 10 ขั้น บนแกนนอนของรูปปิรามิดแสดงค่าความยากของข้อสอบที่สัมพันธ์กับข้อสอบในแบบทดสอบ โดยมีค่าความยากอยู่ในช่วงตั้งแต่ .05 ถึง .95 ข้อสอบที่อยู่ในแนวตั้งเดียวกันจะมีระดับความยากเท่ากัน และข้อสอบที่ผู้เข้าสอบจะต้องตอบในขั้นที่อยู่ติดกัน จะมีค่าความยากแตกต่างกันเท่ากับ .05

ในการดำเนินการสอบ ผู้สอบทุกคนจะเริ่มทำข้อสอบหมายเลข 1 ในขั้นที่ 1 คำตอบของผู้เข้าสอบข้อนี้อาจถูกหรือผิดจะใช้พิจารณาเพื่อแยกทางในการตอบข้อต่อไปในขั้นที่ 2 โดยใช้กฎการแยกทางแบบเพิ่ม 1 / ลด 1 ถ้าเขาตอบถูกจะถูกแยกทางให้ไปตอบข้อที่มีความยากสูงขึ้นในขั้นต่อไป แต่ถ้าตอบผิด ก็จะถูกแยกทางให้ไปตอบข้อที่มีความยากลดลงในขั้นต่อไป ดังตัวอย่างในภาพที่ 7 เป็นเส้นทางการตอบข้อสอบของผู้เข้าสอบ 3 คน เส้นทึบแสดงเส้นทางการตอบข้อสอบของคนที่ 1 ที่มีความสามารถปานกลาง เส้นทางการตอบข้อสอบของคนที่ 2แสดงด้วยเส้นประ เขาตอบข้อสอบผิดเพียงข้อเดียวในขั้นที่ 2 คือข้อสอบหมายเลข 3 (p = .45) ส่วนเส้นทางการตอบข้อสอบของคนที่ 3 แสดงด้วยเส้นจุดไข่ปลา จะเห็นได้ว่าเขาตอบข้อสอบถูกเพียง 2 ข้อ คือ ข้อสอบหมายเลข 4 (p = .60) และข้อสอบหมายเลข 38 (p = .80)

กฎการแยกทางในแบบทดสอบที่จัดตามแบบจำลองรูปปิรามิดอาจใช้วิธีเพิ่ม / ลดไม่เท่ากันก็ได้ เช่น เพิ่ม 1/ ลด 2 เพิ่ม 2/ ลด 3 หรือเพิ่ม 1 / ลด 3 เป็นต้น แต่การเพิ่ม / ลดไม่เท่ากันนี้จะมีผลทำให้ผู้เข้าสอบได้ตอบข้อสอบมีจำนวนข้อไม่เท่ากัน

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 7 แสดงโครงสร้างและเส้นทางการตอบข้อสอบตามแบบจำลองรูปปิรามิดที่ใช้ขนาดขั้นคงที่ 10 ขั้น และใช้กฎการแยกทางแบบเพิ่ม 1/ลด 1



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

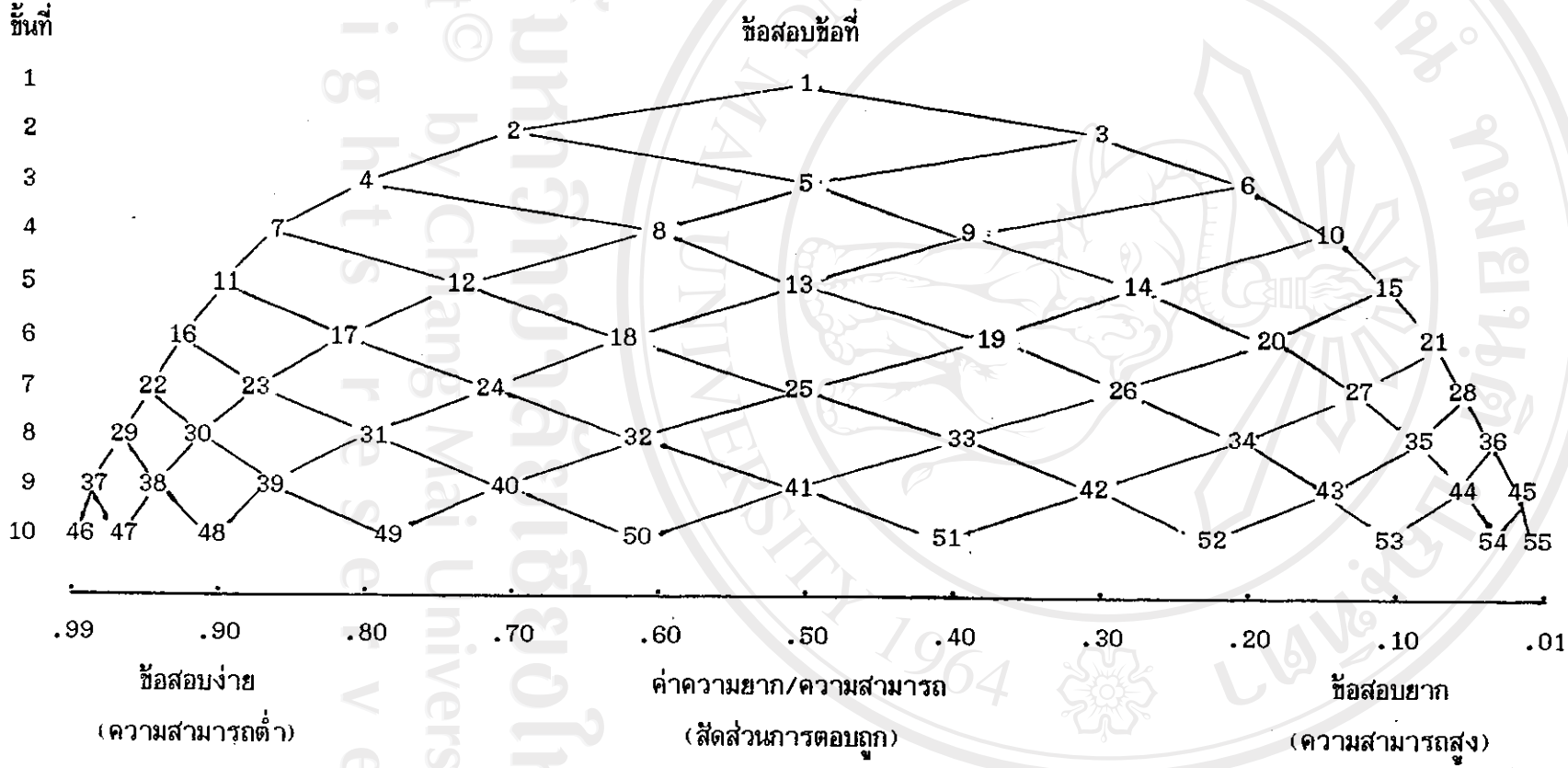
แบบจำลองรูปปิรามิดที่ใช้ขนาดขั้นแปรผัน

แบบทดสอบที่จัดตามแบบรูปจำลองปิรามิดที่ใช้ขนาดขั้นแปรผัน เป็นแบบที่มุ่งแก้ไขข้อบกพร่องของแบบที่ใช้ขนาดขั้นคงที่ตามข้อสังเกตของแพทเตอร์สัน และ ลอร์ด (Paterson, 1962 and Lord, 1971 cited by Weiss, 1974 : 18) ที่ว่า " แบบทดสอบรูปปิรามิดที่ใช้ขนาดขั้นคงที่ขาดประสิทธิภาพด้านความไวในการแยกทางการตอบในขั้นต่อ ๆ ไป เพื่อแก้ปัญหาดังกล่าวควรให้ค่าความแตกต่างระหว่างค่าความยากของข้อสอบในขั้นที่อยู่ติดกันมีค่าน้อยลง จะช่วยให้การวัดระดับความสามารถมีความเที่ยงตรงสูงขึ้น " พร้อมกับข้อสังเกตดังกล่าว แพทเตอร์สัน ได้เสนอแบบจำลองรูปปิรามิดที่มีจำนวนข้อเท่ากับลำดับขั้น แต่การเพิ่มลดค่าความยากของข้อสอบจากขั้นหนึ่งไปยังขั้นถัดไป จะเป็นการเพิ่มลดไปยังตำแหน่งกึ่งกลางระหว่าง 2 ข้อในขั้นก่อนที่ใกล้เคียงกันที่สุด หรือกึ่งกลางระหว่างข้อที่ตอบกับข้อที่ยากสูงสุดในกลุ่มข้อสอบ (ถ้าตอบถูกทุกข้อที่ผ่านมา) หรือกึ่งกลางระหว่างข้อที่ตอบกับข้อที่ง่ายที่สุดในกลุ่มข้อสอบ (ถ้าตอบผิดทุกข้อที่กล่าวมา) ตัวอย่างโครงสร้างตามแนวคิดของแพทเตอร์สัน แสดงไว้ในภาพที่ 8

ส่วนลอร์ดได้เสนอวิธีการจัดรูปแบบข้อสอบโดยอาศัยแนวความคิดตามกระบวนการโรบิน-มอนโร คอนเวอร์เจนซ์ (Robin - Monro Convergence Procedure) รูปแบบนี้มี 6 ขั้น การเพิ่มลดความยากจะลดลงทีละครึ่งหนึ่งของขั้นก่อน กล่าวคือ ถ้าขั้นที่ 1. ข้อแรกความยาก (p) = .50 จากขั้นที่ 1 ไปขั้นที่ 2 เพิ่มลดความยาก 0.25 ดังนั้น ถ้าตอบถูกในขั้นที่ 1 จะต้องไปทำข้อสอบที่มีค่า p = .25 และถ้าตอบผิดจะต้องไปทำข้อสอบที่มีค่า p = .75 ในขั้นที่ 3 การเพิ่มลดความยากข้อสอบจะเป็นขนาด .125 ดังนั้น ถ้าตอบถูกในข้อที่มีค่า p = .25 ในขั้นที่ 2 แล้ว ข้อต่อไปที่ต้องตอบในขั้นที่ 3 จะมีค่า p = (.25 - .125) = .125 และถ้าตอบผิดในขั้น 2 ข้อ p = .25 จะต้องไปตอบข้อที่มีค่า p = (.25 + .125) = .375 ในขั้นที่ 3 ดังนี้เป็นต้น ตัวอย่างโครงสร้างข้อสอบตามกระบวนการโรบิน - มอนโร แสดงไว้ในภาพที่ 9

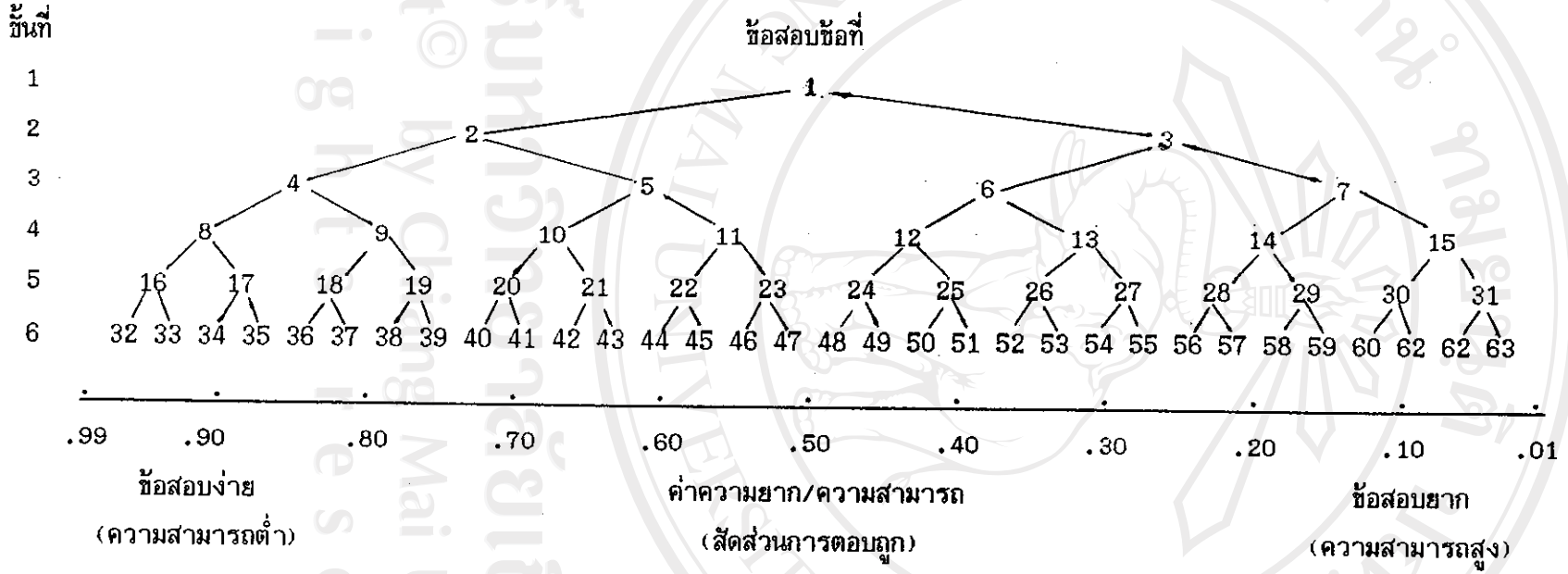
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 8 แสดงโครงสร้างการจัดแบบทดสอบรูปปิรามิดที่ใช้ขนาดขั้นแปรผันตามแนวของแพทเตอร์สัน



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 9 แสดงโครงสร้างการจัดแบบทดสอบรูปปิรามิดที่ใช้ขนาดชั้นแปรผันตามกระบวนการโรบิน-มอนโร



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

แบบจำลองรูปปิรามิดข้างตัด

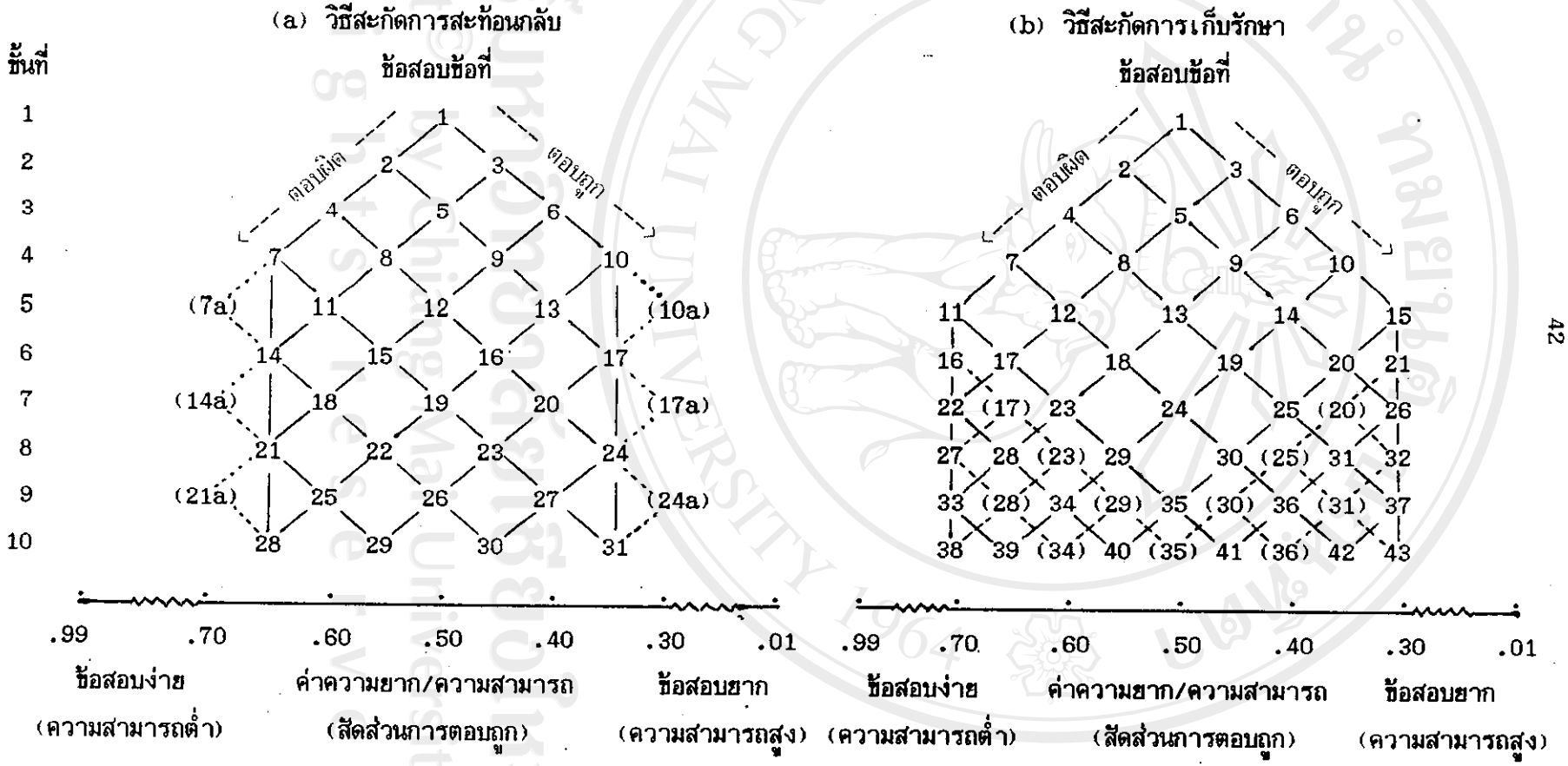
มุสสิโอ (Mussio, 1973 cited by Weiss, 1974 : 22) เสนอรูปแบบปิรามิดข้างตัดเพื่อลดจำนวนข้อสอบที่ต้องใช้ในแบบจำลอง โดยตัดหางของปิรามิดบริเวณยากสุดและง่ายสุดทิ้งไป แล้วใช้กฎการแยกทางแบบใดแบบหนึ่งใน 2 แบบ เมื่อการสอบดำเนินไปจนถึงชั้นที่เป็นข้อสอบส่วนข้างที่ตัดออก ดังภาพที่ 10

ภาพที่ 10 แสดงโครงสร้างของแบบทดสอบรูปปิรามิดข้างตัดชนิด 10 ชั้น ที่ใช้วิธีสะกัดการสะท้อนกลับ และ วิธีสะกัดการเก็บรักษา (Reflecting and Retainning Barrier) ข้อสอบจะมีความยากอยู่ในช่วง .65 ถึง .35 เท่านั้น ข้อสอบที่มีระดับความยากสูงหรือต่ำกว่านี้ไม่ถือว่าอยู่ในโครงสร้างนี้ จึงเหลือข้อสอบทั้งหมด 31 ข้อ ใน 10 ชั้น กระบวนการแยกทางจะยังคงดำเนินไปตามปกติใน 4 ชั้นแรก กล่าวคือ ถ้าตอบข้อสอบถูกจะถูกกำหนดให้ตอบข้อสอบข้อที่ยากขึ้น และถ้าตอบผิดก็จะถูกกำหนดให้ตอบข้อสอบที่มีความยากลดลง เมื่อผู้สอบตอบข้อสอบผิดถึงข้อสอบหมายเลข 7 หรือตอบข้อสอบถูกถึงหมายเลข 10 แล้วต่อไปจะเป็นผลของวิธีสะกัดการสะท้อนกลับ

ข้อสอบหมายเลข 7a 14a และ 21a จะแสดงถึงข้อสอบข้อที่จะตามมาหลังจากที่ผู้สอบตอบข้อสอบหมายเลข 7, 10 หรือ 21 ผิด อย่างไรก็ตาม เนื่องจากแบบทดสอบรูปปิรามิดข้างตัดที่ใช้วิธีสะกัดการสะท้อนกลับ ถ้าผู้สอบตอบข้อสอบข้อดังกล่าวผิดจะถูกกำหนดให้ตอบข้อสอบข้อถัดไปคล้ายกัน เช่น แทนที่จะตอบข้อ 7a ในการแยกทาง ก็จะมาตอบข้อ 14a แทน กรณีเช่นนี้ ข้อสอบเหล่านี้จึงไม่จำเป็นต้องจัดเอาไว้ ดังนั้นในการใช้วิธีสะกัดการสะท้อนกลับ เมื่อผู้สอบตอบข้อสอบหมายเลข 7 ผิด จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 14 ขณะที่ผู้สอบตอบข้อสอบหมายเลข 7 ถูก จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 11 ลักษณะที่สำคัญของวิธีสะกัดการสะท้อนกลับก็คือ ข้อสอบ ณ จุดสะกัด หรือ ข้อสอบ ณ จุดตัดของรูปปิรามิด เป็นข้อสอบที่สมมุติเอาว่าถูกจัดให้ไม่มีความแตกต่างในการแยกทาง สามารถตัดทิ้งจากข้อสอบในโครงสร้างได้ ซึ่งมีผลต่างจากการใช้ข้อสอบ 55 ข้อ ในแบบทดสอบชนิด 10 ชั้นเพียงเล็กน้อยสำหรับผู้สอบบางคนเท่านั้น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 10 แสดงโครงสร้างการจัดแบบทดสอบรูปปิรามิดแบบข้างตัด



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ส่วนวิธีสะกัดการเก็บรักษานั้น ข้อสอบที่มีระดับความยากที่ปลายทั้งสองของรูปปิรามิดจะถูกสะกัดกั้น วิธีนี้จะมีการเพิ่มข้อสอบที่ตำแหน่งสะกัดกั้นโดยมีระดับความยากของข้อสอบเท่ากับค่าความยากของข้อสอบที่มีอยู่เดิม ดังภาพที่ 10 จะมีข้อสอบ 3 ข้อ ที่เพิ่มเข้าไปที่ระดับความยาก .70 คือข้อสอบหมายเลข 16, 27 และ 28 และเพิ่มเข้าไปอีก 3 ข้อ ที่ระดับความยาก .30 คือข้อสอบหมายเลข 21, 32 และ 43

การแยกทางในวิธีสะกัดการเก็บรักษายังคงเป็นไปตามปกติจนถึงจุดสะกัดกั้นผู้สอบที่ตอบข้อสอบ 4 ข้อแรกใน 4 ขั้นแรกผิด จะได้ทำข้อสอบที่ง่ายจากข้อสอบหมายเลข 1, 2, 4 และ 7 จนถึงข้อสอบหมายเลข 11 ถ้าตอบข้อสอบหมายเลข 11 ถูก จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 17 เป็นข้อต่อไป แต่ถ้าตอบข้อสอบหมายเลข 11 ผิด จะถูกกำหนดให้ตอบข้อสอบหมายเลข 16 ซึ่งมีค่าความยากเท่ากับข้อสอบหมายเลข 11 ต่อไปก็เช่นกัน ถ้าผู้สอบตอบข้อสอบหมายเลข 16 ผิดอีกก็จะได้ตอบข้อสอบที่มีความยากเท่าเดิมจนถึงข้อที่ 38

เมื่อผู้สอบตอบข้อสอบข้อหนึ่งข้อใดในจำนวนข้อสอบที่เพิ่มเข้าไว้ตามวิธีสะกัดการเก็บรักษา (อาจเป็นข้อสอบหมายเลข 16, 27 หรือ 38 ที่มีระดับความยาก .70) จะได้ตอบข้อสอบข้อที่ยากขึ้นไปอีกตามรูปแบบการแยกทางที่แสดงไว้ กล่าวคือ ถ้าตอบข้อสอบหมายเลข 16 ในขั้นที่ 6 ถูก จะได้ตอบข้อสอบหมายเลข 17 ในขั้นที่ 7 ที่แสดงโดยเส้นประในภาพประกอบ 10b การแยกทางไปตอบข้อสอบข้อต่อไปก็เป็นไปตามเส้นประนั่นเอง โดยใช้ข้อสอบหมายเลขในวงเล็บ โดยวิธีดำเนินการเช่นนี้ผู้สอบแต่ละคน จะได้ตอบข้อสอบเป็นจำนวนข้อรวมเท่ากันทุกคน

ตามข้อเสนอของ มุสสิโอ ที่เกี่ยวกับโครงสร้างการสะกัดการสะท้อนกลับและโครงสร้างสะกัดการเก็บรักษานั้น สามารถนำไปใช้ได้กับวิธีการใช้ขนาดขั้นคงที่และขนาดขั้นแปรผัน ประโยชน์ที่สำคัญของวิธีการนี้คือ ประหยัดข้อสอบ เช่น แบบทดสอบรูปปิรามิดข้างตัดชนิด 60 ขั้น

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

จะใช้ข้อสอบน้อยกว่า 25% ของจำนวนข้อสอบที่ต้องใช้กับโครงสร้างรูปปิรามิดที่ใช้ขนาดขั้นคงที่ เมื่อโครงสร้างรูปปิรามิดข้างตัดใช้ระดับความยากข้อสอบ 7 ระดับ ข้อบกพร่องของวิธีสะกัดการย้อนกลับ และวิธีสะกัดการเก็บรักษาก็คือ เป็นการวัดความสามารถรอบ ๆ ค่าเฉลี่ยของระดับความยากของข้อสอบในแบบทดสอบ และ ลดความสามารถของแบบทดสอบที่จะจำแนกความสามารถของผู้สอบที่มีความสามารถอยู่บริเวณสูงสุดหรือต่ำสุด

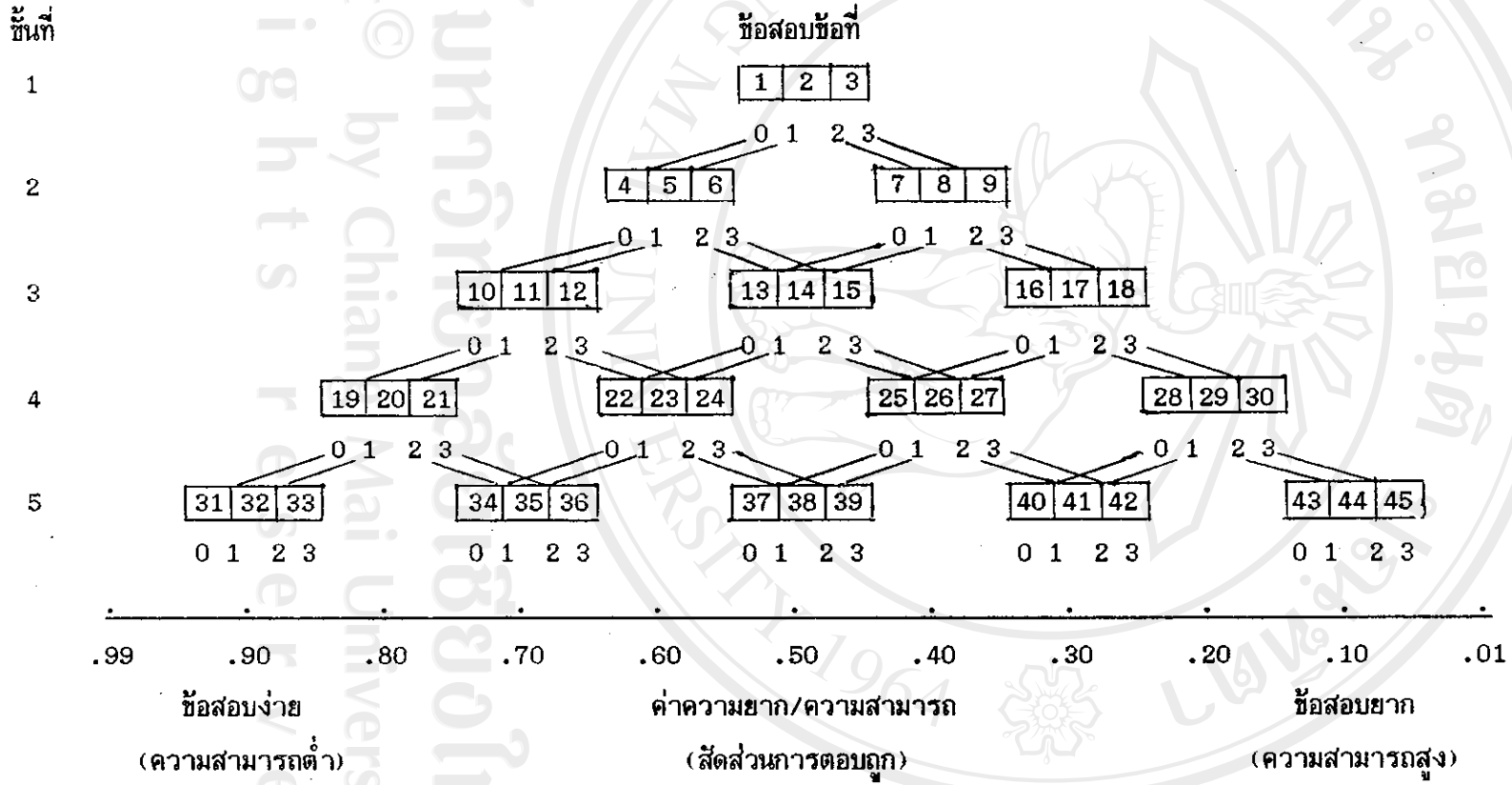
แบบจำลองรูปปิรามิดที่มีข้อสอบหลายข้อในแต่ละขั้น

แบบจำลองรูปปิรามิดที่มีข้อสอบหลายข้อในแต่ละขั้น มีจุดประสงค์เพื่อปรับปรุงความเที่ยงตรงของการแยกทางและลดจำนวนขั้นในแบบจำลองรูปปิรามิดลง ภาพที่ 11 แสดงตัวอย่างแบบจำลองรูปปิรามิดที่มีข้อ 3 ข้อในแต่ละขั้น ผู้เข้าสอบจะตอบข้อสอบทั้ง 3 ข้อในแต่ละขั้น ก่อนที่จะแยกทางต้องนับคะแนนของการตอบข้อสอบในขั้นนั้น ๆ ก่อน โดยแต่ละข้อให้คะแนนแบบ 1 - 0 ผู้ที่ได้คะแนน 0 หรือ 1 จะถูกแยกทางให้ไปตอบข้อสอบที่มีความยากลดลงในขั้นถัดไป ผู้ที่ได้คะแนน 2 หรือ 3 จะถูกแยกทางให้ไปตอบข้อสอบที่มีความยากสูงขึ้นในขั้นถัดไป ตามตัวอย่างในภาพที่ 11 ข้อสอบ 3 ข้อแรก มีค่าความยากระหว่าง .55 ถึง .45 ถ้าได้คะแนน 0 - 1 จะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 4, 5 และ 6 ในขั้นที่ 2 ซึ่งมีค่าความยากระหว่าง .56 ถึง .65 ผู้ได้คะแนน 2 - 3 จะถูกแยกทางไปตอบข้อสอบหมายเลข 7, 8 และ 9 ในขั้นที่ 2 ซึ่งมีค่าความยากระหว่าง .44 ถึง .35 กระบวนการตอบข้อสอบและการแยกทางในแต่ละขั้นจะดำเนินการเช่นนี้ไปจนถึงขั้นสุดท้าย

ผู้เสนอแนวคิดตามแบบจำลองนี้ได้แก่ แครธโวห์ล และ ฮายเซอร์ (Krathwohl and Huyser, 1956) ลินน์ (Linn, 1969) (Weiss, 1974 : 25) โครงสร้างของแบบทดสอบตามแบบจำลองรูปปิรามิดที่มีข้อสอบหลายข้อในแต่ละขั้น สามารถกำหนดได้หลายลักษณะ ทั้งขนาดขั้นคงที่หรือขนาดขั้นแปรผัน และ สามารถใช้กฎการแยกทางได้หลายลักษณะ กล่าวได้ว่า รูปแบบนี้เป็นรูปแบบผสมระหว่างยุทธวิธีสองขั้นกับแบบจำลองรูปปิรามิด

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 11 แสดงโครงสร้างแบบทดสอบรูปปิรามิดที่มีข้อสอบ 3 ข้อ ในแต่ละชั้น



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

แบบจำลองปิรามิดที่ให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง

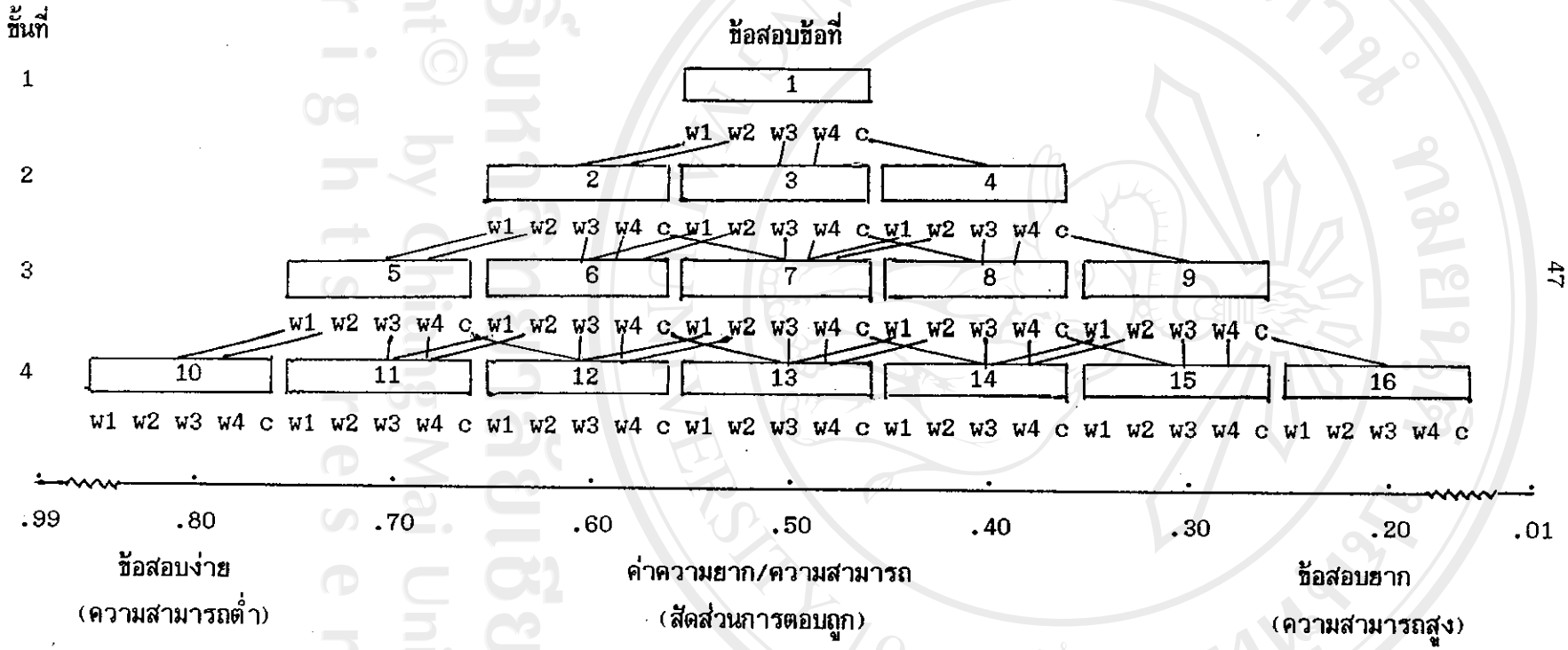
จุดประสงค์ของกระบวนการนี้ก็คือ ต้องการใช้รายละเอียดทั้งหมดในคำตอบของผู้สอบเพื่อแยกทางไปทำข้อสอบข้อต่อไป โดยให้ขึ้นอยู่กับความหนักเบาของคำตอบที่ตอบผิดแต่ละข้อ แบบจำลองนี้จึงออกแบบเพื่อใช้กับข้อสอบแบบเลือกตอบ หรืออาจดัดแปลงเพื่อใช้กับข้อสอบที่ให้ตอบอย่างอิสระที่ซึ่งสามารถกำหนดค่าความยากให้กับแต่ละคำตอบได้ รูปแบบการให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทางนี้เสนอโดย เบย์รอฟฟ์และซีเลย์ (Bayroff and Seeley, 1967 cited by Weiss, 1974 ; 28) โดยที่เขาใช้วิธีนี้กับขั้นที่ 1 เท่านั้น ในแบบทดสอบรูปปิรามิดชนิด 5 ขั้น ที่เขาทำการศึกษา

ภาพที่ 12 มีข้อสอบ 1 ข้อ ในขั้นที่ 1 มี 3 ข้อ ในขั้นที่ 2 มี 5 ข้อ ในขั้นที่ 3 มี 7 ข้อ ในขั้นที่ 4 จำนวนข้อสอบจะเพิ่มขึ้น 2 ข้อ ในแต่ละขั้นที่อยู่ถัดกันไป ข้อสอบดังภาพประกอบเป็นข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก จะมีคำตอบที่ถูกอยู่ 1 ตัวเลือก (c) และคำตอบที่ผิดอีก 4 ตัวเลือก (w1, w2, w3, w4) แทนตัวเลือกที่ผิดมากที่สุดด้วย w1 และ w4 แทนตัวเลือกที่ผิดน้อยที่สุดหรือเกือบจะถูกต้อง โดยถือเอากลุ่มคนปกติช่วยพิจารณา

ในขั้นที่ 1 ผู้สอบจะได้ตอบข้อสอบหมายเลข 1 ถ้าคำตอบถูก หรือ c ผู้สอบจะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 4 ในขั้นที่ 2 ซึ่งเป็นข้อสอบที่ยากที่สุดในขั้นที่ 2 ถ้าตอบตัวเลือกที่มีระดับความผิด w1 และ w2 แต่ยังคงไม่ถูกต้อง ผู้สอบจะถูกแยกทางให้ไปตอบข้อสอบหมายเลข 3 ซึ่งมีระดับความยากเท่ากับข้อสอบหมายเลข 1 ในขั้นที่ 1 ที่จัดแยกทางเช่นนี้มีเหตุผลว่า ผู้ที่จะเลือกตัวเลือกที่มีระดับความผิด w3 และ w4 ส่วนใหญ่จะมีความสามารถอยู่ในระดับเฉลี่ยของกลุ่มปกติ ส่วนผู้ที่มีระดับความสามารถที่สูงหรือต่ำกว่านี้ก็จะถูกจัดให้ไปตอบข้อสอบที่มีความยากสูงหรือต่ำกว่า การแยกทางในขั้นต่อ ๆ ไปก็ยังคงอาศัยวิธีการเช่นเดียวกันนี้

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 12 แสดงโครงสร้างของแบบทดสอบรูปปิรามิดที่ใช้การกำหนดน้ำหนักแก่ตัวเลือกเพื่อแยกทาง



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ถ้าให้คะแนนผู้สอบจากการตอบแบบทดสอบนี้ โดยถือเอาตามตัวเลือกที่กำหนดน้ำหนักของความผิดในขั้นสุดท้ายแล้ว แบบทดสอบรูปปิรามิดชนิด 4 ชั้น ที่มีโครงสร้างตามรูปแบบนี้จะมีคะแนนที่แตกต่างกัน 35 คะแนน ขณะที่แบบทดสอบรูปปิรามิดที่ใช้ขนาดขั้นคงที่ 4 ชั้น แบบภาพที่ 7 จะมีคะแนนที่จะได้รับต่างกันเพียง 8 คะแนน แบบทดสอบรูปปิรามิดที่ให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง จึงให้ประโยชน์เกี่ยวกับรายละเอียดของคำตอบของผู้สอบมากกว่าแบบทดสอบรูปปิรามิดที่มีการรวมตัวเลือกที่ผิดเป็นคะแนนเดียว เช่น ตอบข้อสอบถูกให้ 1 ตอบข้อสอบผิดให้ 0

ภาพที่ 12 เป็นเพียงรูปแบบหนึ่งของการให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง การแยกทางจะสมบูรณ์ที่สุด เมื่อทุก ๆ ตัวเลือกที่ผู้สอบได้ตอบ สามารถแยกทางโดยเฉพาะของมันเอง กล่าวคือ เดิมทีการแยกทางผู้สอบเมื่อผู้สอบเลือกตอบตัวเลือกที่มีระดับความผิด w1 และ w2 ในขั้นที่ 1 จะถูกจัดให้ไปตอบข้อสอบหมายเลข 2 ในขั้นที่ 2 ข้อเดียวกันก็จะดำเนินทางแยกทางใหม่ โดยผู้สอบที่ตอบตัวเลือกที่มีระดับความผิด w1 จะได้ตอบข้อสอบที่มีระดับความยากน้อยกว่า ข้อสอบที่ผู้ตอบตัวเลือกที่มีระดับความยาก w2 ในขั้นต่อไป ดังนี้เป็นต้น วิธีการเช่นนี้เมื่อใช้กับข้อสอบแบบเลือกตอบชนิด 5 ตัวเลือก ก็จะมีทางให้เลือกเพื่อแยกทางอยู่ถึง 5 วิธี แต่ที่แสดงทางแยกเพียง 3 ทาง ในภาพประกอบ 11 นับเป็นวิธีการแยกทางที่ง่ายที่สุด ซึ่งทำให้ได้รูปปิรามิดที่ไม่สมมาตร โดยมีทางเลือกที่มีโอกาสจะไปทางซ้ายได้มากกว่าทางขวา ดังนั้นจำนวนข้อสอบในกลุ่มข้อสอบจึงจำเป็นต้องใช้มากกว่า กระบวนการแยกทางต่าง ๆ ที่ได้กล่าวมา แต่อย่างไรก็ตาม รูปแบบนี้ก็มีความถูกต้องในภารวัด มีจำนวนข้อสอบที่ผู้สอบจะได้ตอบคงที่และเป็นข้อสอบที่ได้รับการให้น้ำหนักตัวเลือกไว้ล่วงหน้า ด้วยเหตุนี้รูปแบบการให้น้ำหนักแก่ตัวเลือกของข้อสอบแยกทาง จึงได้รับความสนใจอย่างกว้างขวาง

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

แบบทดสอบเปลี่ยนระดับ

แบบทดสอบเปลี่ยนระดับ เป็นรูปแบบที่ปรับจากรูปแบบปิรามิดเพื่อให้มีจำนวนข้อในกลุ่มข้อสอบน้อยลง โดยที่โครงสร้างข้อสอบจะต่างจากแบบปิรามิดในข้อที่ว่า จะไม่มีข้อใดที่มีระดับความยาก (difficulty) เท่ากันเลยในแบบทดสอบเปลี่ยนระดับในขณะที่แบบปิรามิดจะมีซ้ำกันมากกว่า 1 ข้อ

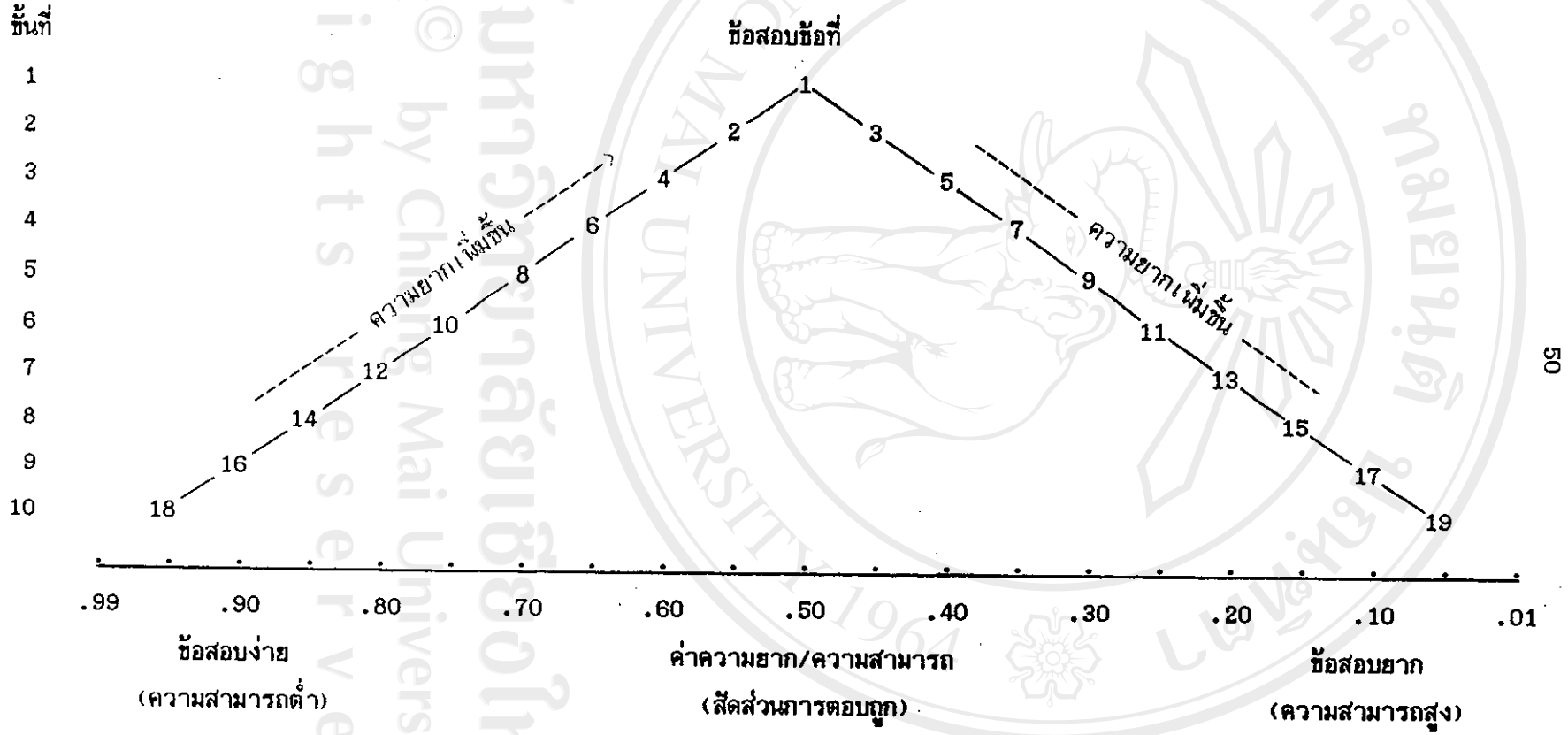
เนื่องจากแบบทดสอบเปลี่ยนระดับจะมีข้อสอบเพียงข้อเดียวในแต่ละระดับความยากดังนั้นกฎการแยกทางจึงต่างจากแบบปิรามิด กฎการแยกทางในแบบทดสอบเปลี่ยนระดับระบุว่า "เมื่อคำตอบของผู้เข้าสอบถูกต้อง ข้อสอบข้อต่อไปจะยากขึ้นและค่าความยากจะต้องแตกต่างจากข้อที่ตอบไปแล้ว และถ้าคำตอบของเขาผิดผู้เข้าสอบจะได้ตอบข้อสอบที่ง่ายลงในข้อต่อไปและค่าความยากก็จะต้องแตกต่างจากข้อที่ตอบไปแล้ว" (Weiss, 1974 : 37)

จากภาพที่ 13 แสดงให้เห็นโครงสร้างการจัดข้อสอบในแบบทดสอบเปลี่ยนระดับ ภาพ 14a แสดงให้เห็นการตอบของผู้เข้าสอบที่มีความสามารถสูง ผู้เข้าสอบทุกคนเริ่มทำข้อสอบข้อที่ 1 ซึ่งเป็นข้อที่มีความยากในระดับกลาง ๆ เมื่อตอบถูกก็จะได้รับข้อสอบที่ยากขึ้นเช่นเดียวกับรูปแบบปิรามิด

จากภาพที่ 14b สรุปได้ว่า ผู้เข้าสอบมีความสามารถอยู่ในระดับกลาง ๆ เพราะข้อที่มีความยากเกิน .50 เขาจะตอบผิดในขณะที่ข้อที่ง่ายกว่า .50 เขาจะตอบถูก ส่วนภาพ 14c แสดงว่าผู้เข้าสอบมีความสามารถอยู่ในระดับต่ำ

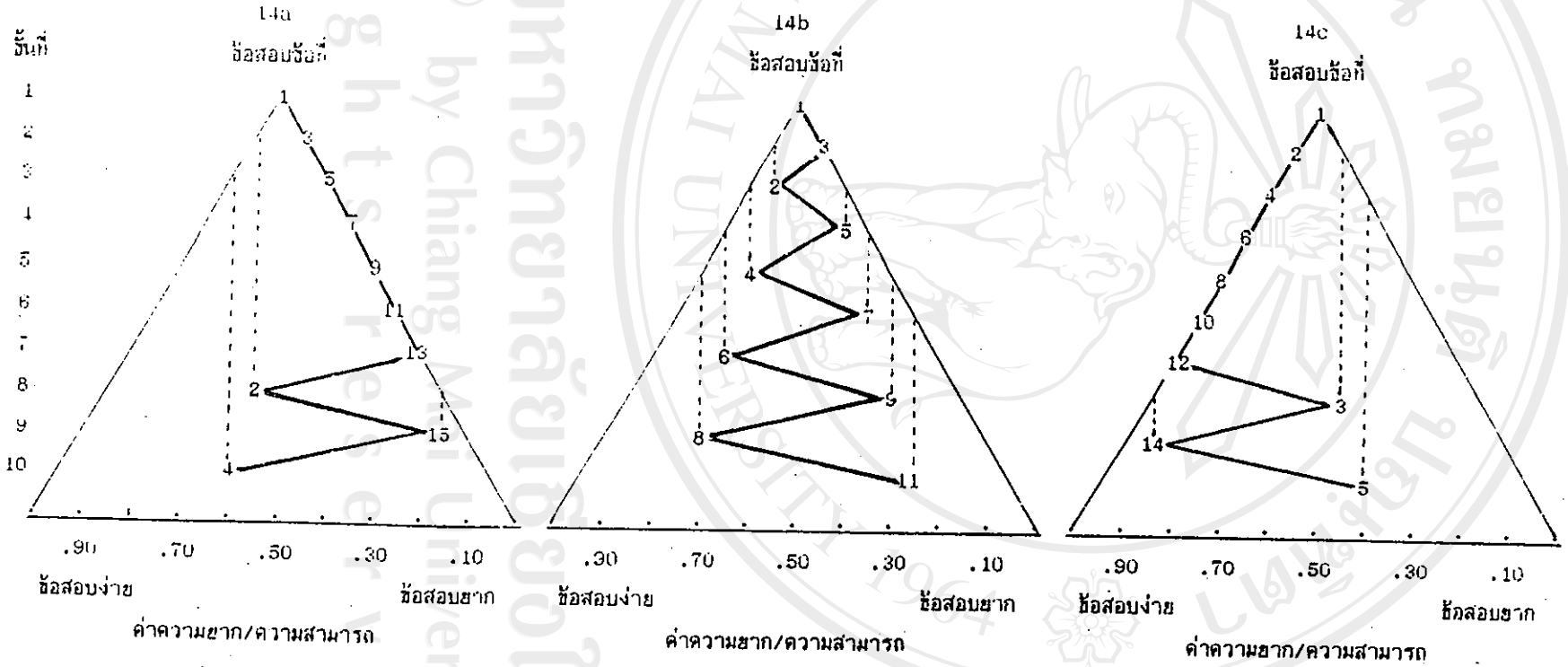
ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 13 แสดงโครงสร้างการจัดแบบทดสอบแบบเปลี่ยนระดับ



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 14 แสดงเส้นทางการตอบแบบทดสอบเปลี่ยนระดับ



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

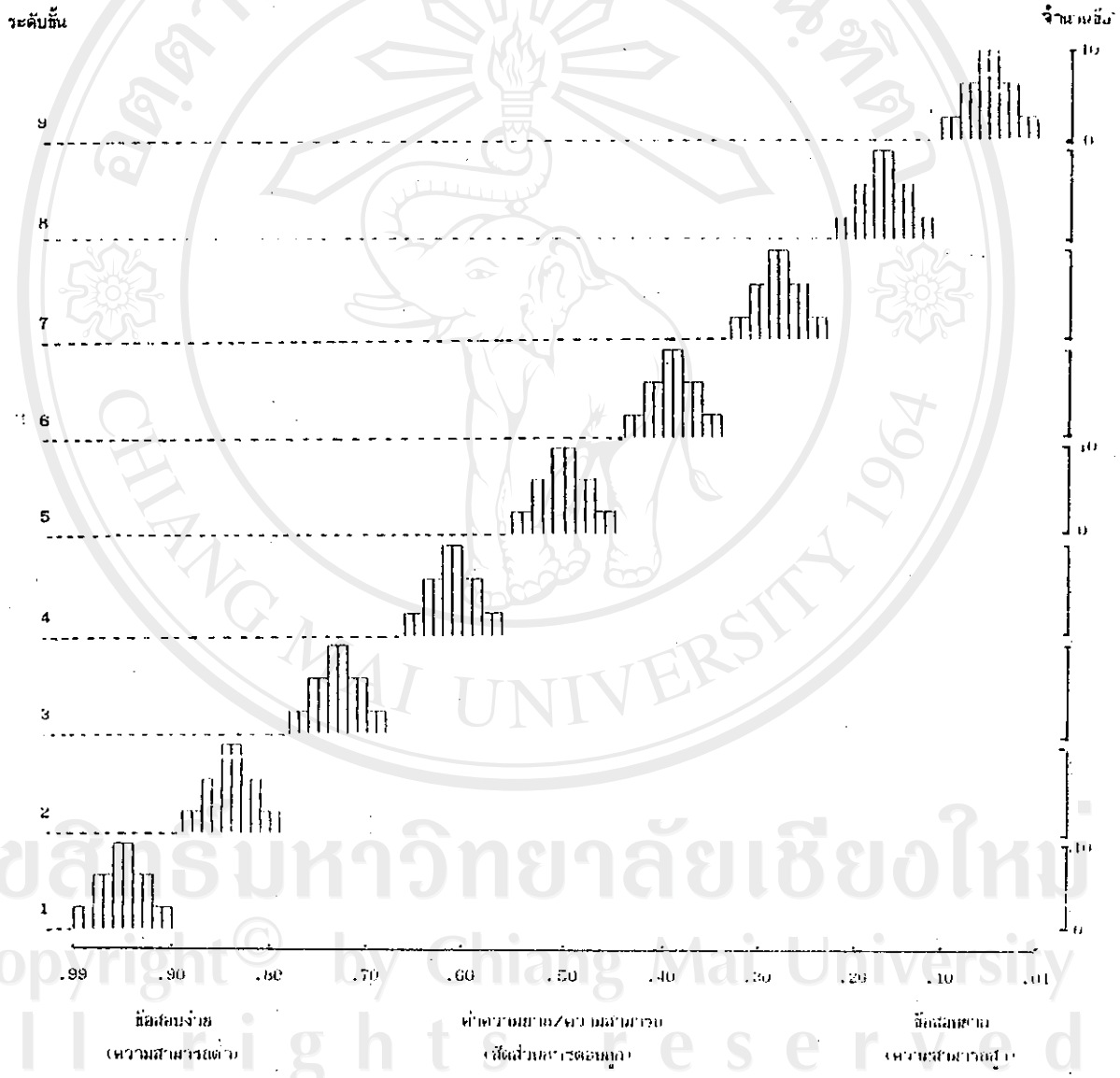
แบบจำลองปรับระดับขั้น (The Stradative Test)

แบบจำลองนี้เสนอโดย ไวส์ (Weiss, 1972 cited by Weiss, 1974 : 44) คำว่า Stradative มาจากคำว่า Stratified + adaptive การจัดแบบทดสอบเริ่มต้นด้วยการจัดกลุ่มข้อสอบตามระดับความยาก เป็น 9 ระดับขั้น (Stratum) แต่ละระดับขั้นประกอบด้วยข้อสอบที่มีค่าความยากใกล้เคียงกัน โดยมีค่าความยากกระจายรอบค่าความยากเฉลี่ยของระดับขั้นนั้น ๆ และมีจำนวนข้อมากที่สุดในช่วงความยากเท่า ๆ กับความยากเฉลี่ยของกลุ่มข้อสอบในระดับขั้นเดียวกัน ทำให้การแจกแจงของค่าความยากในแต่ละระดับขั้นมีลักษณะโด่งแหลมตรงกลาง ไวส์ เรียกกลุ่มข้อสอบในแต่ละระดับขั้นว่า Peaked test ภาพที่ 15 ได้แสดงให้เห็นลักษณะการจัดข้อสอบตามแบบจำลองปรับระดับขั้น 9 ระดับ ระดับขั้นที่ 1 ซึ่งเป็นกลุ่มข้อสอบที่ง่ายที่สุดมีค่าความยากระหว่าง .89 และ .99 โดยมีค่าความยากเฉลี่ย .94 ข้อสอบในกลุ่มส่วนใหญ่จะมีความยากใกล้ ๆ .94 มีจำนวนน้อยที่มีค่าความยากใกล้ ๆ .89 และ .99 ในระดับขั้นที่ 2 เป็นกลุ่มข้อสอบที่ยากกว่าระดับขั้นที่ 1 โดยมีค่าความยากระหว่าง .78 ถึง .88 โดยมีค่าความยากเฉลี่ย .83 ข้อสอบส่วนใหญ่มีค่าความยากใกล้ ๆ .83 มีจำนวนน้อยที่มีค่าความยากใกล้ .78 และ .88 ในระดับขั้นต่อ ๆ ไป กลุ่มข้อสอบจะมีความยากยิ่งขึ้นจนถึงระดับขั้นที่ 9 เป็นกลุ่มข้อสอบที่ยากที่สุดมีค่าความยากระหว่าง .01 ถึง .11 และค่าความยากเฉลี่ย .06

การใช้แบบจำลองปรับระดับขั้นในการสอบ จะเริ่มต้นด้วยการประมาณค่าความสามารถของผู้สอบแต่ละคนจากข้อมูลเดิมของผู้สอบแต่ละคนหรือจากการให้เขารายงานตัวเอง ผู้สอบที่ได้รับการประมาณความสามารถเริ่มต้นต่ำ จะได้เริ่มตอบข้อสอบในระดับขั้นต่ำ ส่วนผู้ที่ได้รับการประมาณค่าความสามารถสูงจะได้เริ่มตอบข้อสอบในระดับขั้นที่สูงกว่า

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 15 แสดงโครงสร้างการจัดแบบทดสอบเทเลอร์แบบปรับระดับชั้น



ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

การแยกทางในแบบจำลองนี้จะกระทำระหว่างระดับขั้นตามกฎการแยกทางที่กำหนด เช่น ขึ้น 1 / ลง 1 ขึ้น 1 / ลง 2 การตอบข้อสอบที่ให้ทำ 1 ข้อในระดับขั้นหนึ่งถ้าตอบถูกจะให้ไปทำข้อสอบที่ยากที่สุดในระดับขั้นสูงขึ้นตามกฎการแยกทาง และถ้าตอบผิดจะให้ไปทำข้อสอบที่มีความยากเท่ากับความยากเฉลี่ยในระดับขั้นที่ต่ำลงตามกฎการแยกทาง ข้อที่ผู้สอบทำไปแล้วจะไม่นำมาใช้อีก การกำหนดระดับความสามารถของผู้สอบ จะกระทำเมื่อได้สาระข้อมูลเพียงพอ นั่นคือ ผู้สอบแต่ละคนตอบข้อสอบในระดับขั้นหนึ่งถูกทุกข้อหรือเกือบทุกข้อ เรียกระดับขั้นนี้ว่า "ขั้นฐาน" (Basal Stratum) ในอีกระดับขั้นหนึ่งผู้สอบจะตอบข้อสอบผิดหมดทุกข้อหรือเกือบทุกข้อ เรียกระดับขั้นนี้ว่า "ขั้นเพดาน" (Ceiling Stratum) และระหว่าง 2 ระดับขั้นนี้มีกลุ่มข้อสอบในระดับขั้นใดที่ผู้สอบตอบถูกประมาณ 50% ของข้อสอบในขั้นนั้น

แบบจำลองการแยกทางแปรผันยุทธวิธีของเบย์

การทดสอบแบบเทเลอร์ที่ประยุกต์ทฤษฎีของเบย์มาใช้ในการจัดลำดับข้อสอบให้ตอบและในการประมาณค่าความสามารถ เสนอโดย โนวิค (Novick, cited by Weiss, 1974 : 57) และปรับปรุงใหม่โดย โอเวน (Owen) มีกระบวนการดังนี้

1. ประมาณค่าความสามารถและความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถในแต่ละขั้นของการทดสอบ โดยอาศัยข้อมูลพื้นฐานเกี่ยวกับผู้สอบ และค่าพารามิเตอร์ต่าง ๆ จากประชากร ถ้าไม่ทราบข้อมูล อาจกำหนดเป็นข้อตกลงเบื้องต้น ให้ค่าความสามารถเบื้องต้น = 0.00 และค่าความแปรปรวน = 1.00

2. คัดเลือกข้อสอบจากกลุ่มข้อสอบให้เป็นข้อที่มีความยากใกล้เคียงที่สุดกับความสามารถของผู้สอบที่ได้จากการประมาณตามข้อ 1 และมีอำนาจจำแนกดีที่สุดในบรรดาข้อสอบที่ยากเท่ากัน ข้อสอบข้อใดที่นำมาให้ผู้สอบตอบไปแล้ว จะไม่นำมาพิจารณาเลือกให้ผู้สอบคนเดิมตอบอีก คำตอบของผู้สอบจะจำแนกเป็น ถูก หรือ ผิด แล้วนำไปประมาณค่าความสามารถและความคลาดเคลื่อน

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

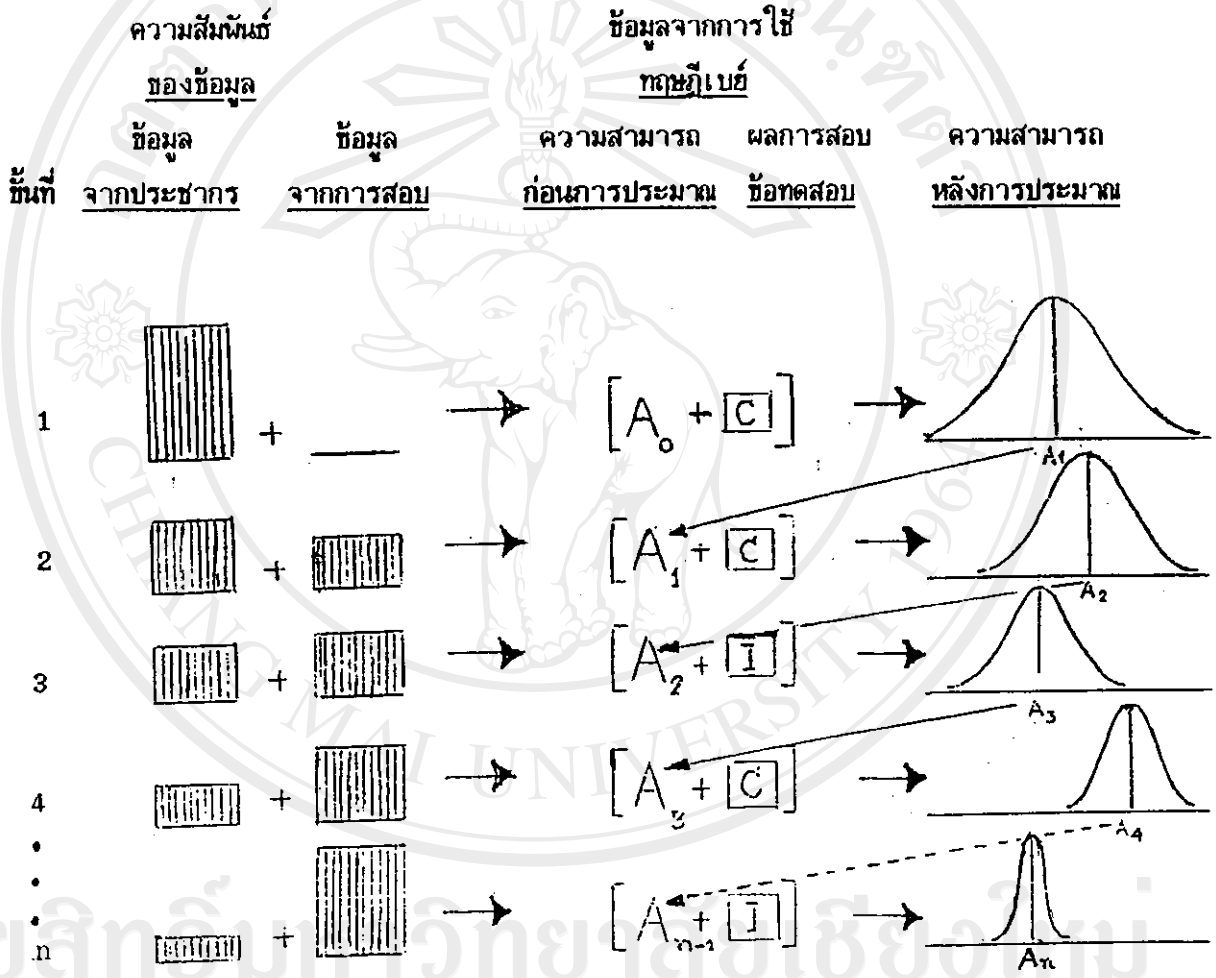
มาตรฐานของการประมาณค่าความสามารถ จากนั้นจึงเลือกข้อสอบข้อใหม่ที่มีความยากเท่า ๆ กับค่าความสามารถที่ประมาณได้ใหม่ กระบวนการนี้ จะทำให้ความคลาดเคลื่อนมาตราฐานของการประมาณค่าความสามารถมีค่าลดลงเรื่อย ๆ ในขณะที่ค่าความสามารถจะมีการเปลี่ยนแปลงไป กล่าวคือ ถ้าตอบข้อสอบถูก ค่าความสามารถที่ประมาณได้ใหม่จะมีค่าสูงขึ้น และถ้าตอบผิดค่าความสามารถที่ประมาณได้ใหม่จะมีค่าลดลง กระบวนดังกล่าวนี้ แสดงให้เห็นได้ดังภาพที่ 16 กระบวนการนี้จะยุติลง เมื่อผลการทดสอบความสามารถลดความคลาดเคลื่อนมาตราฐานลงได้ถึงจุดหนึ่งที่กำหนดไว้

กระบวนการนี้ต้องใช้คอมพิวเตอร์ในการดำเนินการสอบและประมาณค่าต่าง ๆ เพราะต้องกระทำในทันทีและรวดเร็ว จึงนิยมเรียกว่า Computerized Adaptive Mastery Testing

ตัวอย่างค่าตัวเลขต่าง ๆ ที่แสดงไว้ในตารางที่ 1 เป็นบันทึกผลการตอบแบบทดสอบตามยุทธวิธีของเบย์ด้วยคอมพิวเตอร์จำนวน 15 ข้อ วัดความสามารถทางภาษา ค่าประมาณความสามารถจะรายงานในรูปของคะแนนมาตรฐาน ถ้าเราไม่ทราบอะไรเลยเกี่ยวกับผู้เข้าสอบ ค่าประมาณความสามารถก่อนการทดสอบที่ดีที่สุดคือ ค่าเฉลี่ยของคะแนนมาตรฐานหรือก็คือ 0.0 ก่อนการทดสอบเริ่มต้น ค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถ คือ ค่าความเบี่ยงเบนมาตราฐานของคะแนนมาตรฐาน หรือก็คือ 1.00 กำหนดให้ค่าดังกล่าวเป็นค่าก่อนการทดสอบ คอมพิวเตอร์จะดำเนินการคำนวณตามอนุกรมที่กำหนดไว้เป็นรายข้อ ผลการคำนวณนี้จะเป็นตัวชี้ว่าข้อสอบข้อต่อไปจะเป็นข้อใดในกลุ่มข้อสอบที่ใช้ ผู้เข้าสอบจะตอบข้อสอบ คอมพิวเตอร์ให้คะแนน ค่าความคลาดเคลื่อนมาตรฐานจะลดลง ในกรณีนี้คอมพิวเตอร์ระบุให้ข้อสอบหมายเลข 43 จากกลุ่มข้อสอบเป็นข้อเริ่มต้น ผู้เข้าสอบตอบถูกค่าความสามารถใหม่ที่ประมาณได้เป็น .469 ความคลาดเคลื่อนมาตราฐานเป็น .8568 จากค่าที่ประมาณได้ใหม่นี้ คอมพิวเตอร์จะดำเนินการคำนวณใหม่ตามอนุกรมการคำนวณกับข้อสอบที่เหลือในกลุ่ม ผลการคำนวณนี้จะชี้ว่าข้อสอบข้อต่อไปเป็นข้อใด โดยคาดหวังว่าเป็นข้อที่ให้สาระประโยชน์สูงสุด ในกรณีนี้ คอมพิวเตอร์กำหนดว่าเป็นข้อสอบหมายเลข 57 จากกลุ่มข้อสอบเป็นข้อที่เหมาะสมข้อที่ 2 ผู้เข้าสอบตอบถูกอีก ค่าประมาณความสามารถใหม่เป็น .929 และค่าความคลาดเคลื่อนมาตรฐานเป็น .7527

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภาพที่ 16 แสดงการทดสอบตามยุทธวิธีของเบย์



A หมายถึง ความสามารถ (Ability)

C หมายถึง ผลการตอบถูก (Correct)

I หมายถึง ผลการตอบผิด (Incorrect)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ตารางที่ 1 แสดงผลการตอบข้อสอบและผลการประมาณค่าความสามารถทางภาษา

ข้อสอบข้อที่	หมายเลขข้อสอบ	ผลการตอบ	ค่าประมาณความสามารถ	ค่าความคลาดเคลื่อนมาตราฐาน
1	43	ถูก	.469	.8568
2	57	ถูก	.929	.7527
3	55	ถูก	1.271	.6440
4	12	ถูก	1.443	.5703
5	13	ถูก	1.590	.5281
6	54	ถูก	1.772	.5009
7	114	ถูก	1.879	.4741
8	26	ถูก	1.975	.4525
9	103	ผิด	1.805	.3910
10	79	ถูก	1.872	.3768
11	78	ถูก	1.950	.3651
12	149	ผิด	1.802	.3378
13	15	ถูก	1.848	.3282
14	76	ถูก	1.882	.3208
15	74	ถูก	1.941	.3161

จาก Vern W. Urry "Tailored Testing : A Successful Application of Latent Trait Theory," Journal of Educational Measurement. Vol. 14. No.2, 1977 : 182.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภายหลังจากที่ข้อสอบแต่ละข้อถูกเสนอ ผู้เข้าสอบตอบ คอมพิวเตอร์ให้คะแนนค่าความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถจะลดลง ค่าความคลาดเคลื่อนของการประมาณนี้ เป็นค่าความเบี่ยงเบนมาตรฐานของการประมาณค่าความสามารถรอบ ๆ เส้นถดถอย (regression line) ณ ระดับของความสามารถจริงที่ตายตัว ไม่ใช่ความคลาดเคลื่อนมาตรฐานของการวัด ข้อสอบที่ได้รับการคัดเลือกออกมาให้ตอบ จะเป็นข้อที่ช่วยลดความคลาดเคลื่อนมาตรฐานได้เร็วที่สุด

ความเชื่อมั่นของแบบทดสอบแบบเทเลอร์ก็คือ ค่ากำลังสองของสหสัมพันธ์ระหว่างความสามารถที่ประมาณได้ ($\hat{\theta}$) กับความสามารถจริง (θ) และมีความสัมพันธ์กับค่าความคลาดเคลื่อนมาตรฐานดังสมการ

$$\rho^2_{\hat{\theta}\theta} = 1 - \sigma_e^2$$

ภายใต้เงื่อนไขว่าการทดสอบแบบเทเลอร์สำหรับผู้เข้าสอบแต่ละคนจะถึงจุดสิ้นสุดเมื่อค่าความคลาดเคลื่อนมาตรฐาน (σ_e) มีค่าลดลงถึงค่าคงที่ค่าหนึ่งที่กำหนดไว้ก่อน ดังนั้นความเที่ยงตรงในความเป็นมิติเดียว (Unidimensional Validity) ก็คือ ค่าสหสัมพันธ์ระหว่างค่าความสามารถที่ประมาณได้ ($\hat{\theta}$) กับค่าความสามารถจริง (θ)

$$\rho_{\hat{\theta}\theta} = \sqrt{1 - \sigma_e^2}$$

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ภายใต้เงื่อนไขเดียวกันที่ว่า การทดสอบแบบเทเลอร์จะสิ้นสุดลงสำหรับผู้เข้าสอบทุกคน เมื่อค่าความคลาดเคลื่อนมาตรฐาน (σ_e) มีค่าลดลงจนถึงค่าคงที่ค่าหนึ่งที่กำหนดไว้ก่อน ในบริบทนี้ความเที่ยงตรงตามโครงสร้างนิยามได้ว่าเป็นสหสัมพันธ์ระหว่างคุณลักษณะแฝงที่ประมาณได้กับคุณลักษณะแฝงจริง ค่าความเชื่อมั่นของแบบทดสอบแบบเทเลอร์ จะเท่ากับค่าความเที่ยงตรงตามโครงสร้าง (ความเป็นมิติเดียว) ยกกำลังสอง จะเห็นได้ว่าค่าความเชื่อมั่นและค่าความเที่ยงตรงตามโครงสร้าง (ความเป็นมิติเดียว) ของการทดสอบแบบเทเลอร์ยุทธวิธีเบย์เป็นสิ่งที่สามารถควบคุมได้อย่างแน่นอน โดยที่เมื่อกำหนดค่าความเชื่อมั่นที่ต้องการไว้แล้ว ค่าความคลาดเคลื่อนมาตรฐานสุดท้ายก็คือ

$$\sigma_e = \sqrt{1 - \rho^2_{\theta\theta}}$$

ดังนั้น ถ้าให้การทดสอบแบบเทเลอร์ตามยุทธวิธีของเบย์มีค่า $\rho^2_{\theta\theta} = 0.90$ ค่าความคลาดเคลื่อนมาตรฐานสุดท้ายจะมีค่าเป็น .3162 ในตารางที่ 1 จะเห็นได้ว่าค่าความคลาดเคลื่อนมาตรฐานสุดท้ายจะสัมฤทธิ์เมื่อผู้เข้าสอบตอบข้อสอบไปเป็นจำนวน 15 ข้อ ณ จุดนี้ ผู้เข้าสอบมีค่าความสามารถประมาณ 1.941 ค่าความคลาดเคลื่อนมาตรฐานของการประมาณเป็น .3161 ซึ่งสัมฤทธิ์ตามค่าสุดท้าย ถ้าการทดสอบแบบเทเลอร์สำหรับผู้เข้าสอบทุกคนถึงจุดสิ้นสุดในลักษณะคล้ายคลึงกัน ค่าความเชื่อมั่นจะเป็น .90 และความเที่ยงตรงตามโครงสร้าง (ความเป็นมิติเดียว) จะเป็น .95

<u>แบบจำลองการแยกทางแปรผันยุทธวิธีความเป็นไปได้สูงสุด</u>

เออร์รี่ (Urry, 1970 cited by Weiss, 1974 : 62) เป็นผู้พัฒนากระบวนการทดสอบแบบเทเลอร์ยุทธวิธีความเป็นไปได้สูงสุดขึ้น กระบวนการนี้คล้ายคลึงกับยุทธวิธีของเบย์ แต่ใช้ทฤษฎีทางคณิตศาสตร์ที่ต่างกัน โดยที่ในกระบวนการประมาณค่าความสามารถและความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถโดยวิธีความเป็นไปได้สูงสุดนั้นจะกระทำได้ก็ต่อเมื่อผลการตอบข้อสอบของผู้เข้าสอบมีทั้งข้อถูกและข้อผิด

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

กระบวนการทดสอบตามยุทธวิธีนี้ การจัดข้อสอบข้อแรกให้ผู้เข้าสอบทำมักจะใช้ข้อสอบที่มีความยากเท่า ๆ กับความยากเฉลี่ยของข้อสอบในกลุ่มข้อสอบที่ใช้ ถ้าผู้เข้าสอบตอบถูก จะถูกจัดให้ทำข้อสอบข้อที่ยากที่สุดในกลุ่มข้อสอบ และถ้าตอบผิด จะถูกจัดให้ทำข้อสอบที่ง่ายที่สุดในกลุ่มข้อสอบ ทั้งนี้เพื่อให้สามารถประมาณค่าความสามารถและความคลาดเคลื่อนมาตรฐานของการประมาณค่าความสามารถโดยวิธีความเป็นไปได้สูงสุดนั่นเอง เมื่อประมาณค่าครั้งแรกได้แล้ว ข้อสอบข้อต่อไปจะถูกเลือกมาให้ตอบโดยเลือกข้อที่มีความยากใกล้เคียงกับความสามารถที่ประมาณได้เช่นเดียวกับยุทธวิธีของเบย์

ข้อแตกต่างของยุทธวิธีทั้งสองนี้อีกประการหนึ่งก็คือ กระบวนการประมาณค่าความ - สามารถยุทธวิธีของเบย์จะประมาณจากข้อสอบเป็นรายข้อ ในขณะที่ยุทธวิธีความเป็นไปได้สูงสุดจะนำผลตั้งแต่ข้อแรกจนถึงข้อสุดท้ายที่ทำมาใช้ในการประมาณค่า แต่อย่างไรก็ตามเมื่อเพิ่มข้อสอบให้สอบ 1 ข้อแล้วประมาณค่าใหม่นั้น จะมีผลทำให้ค่าความคลาดเคลื่อนมาตราฐานลดลงเช่นเดียวกัน

เรคเคส (Reckase, 1973, 1974, cited by Weiss, 1974 : 64) ได้พัฒนายุทธวิธีความเป็นไปได้สูงสุดที่แตกต่างจากเออร์รี่ กล่าวคือ แทนที่จะเริ่มต้นด้วยข้อสอบที่มีความยากเท่ากับความยากเฉลี่ย เรคเคสจะเริ่มต้นด้วยข้อใดก็ได้ที่สอดคล้องกับข้อมูลเบื้องต้นเกี่ยวกับตัวผู้เข้าสอบแต่ละคน ถ้าผู้เข้าสอบตอบถูก ข้อสอบข้อต่อไปที่นำมาให้ตอบจะยากขึ้นเป็น 2 เท่าของข้อก่อน ถ้าตอบถูกอีก ข้อต่อไปก็จะยากขึ้นเป็น 2 เท่าอีกจนกว่าจะได้ผลการตอบที่ผิด และในทางตรงกันข้าม ถ้าตอบข้อสอบข้อแรกผิด ข้อต่อไปก็จะง่ายลงในอัตราเดียวกันกับการตอบถูก

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

4. <u>ผลการวิจัยที่เกี่ยวข้อง</u>

วูด (Wood, 1969 cited by Larkin and Weiss, 1974 : 5) ได้ศึกษาเปรียบเทียบแบบทดสอบรูปปิรามิดชนิด 4, 5 และ 6 ขั้น กับแบบทดสอบดั้งเดิมโดยดำเนินการสอบแบบเขียนตอบกับนักเรียน 91 คน ค่าความยากของข้อสอบที่อยู่ติดกันในชั้นเดียวกันมีค่าคงที่เป็น .05 กฎการแยกทางใช้วิธีเพิ่ม 1 / ลด 1 ให้คะแนนตามจำนวนข้อที่ตอบถูก พบว่าค่าความเที่ยงตรงของแบบทดสอบที่หาได้จากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากแบบทดสอบรูปปิรามิดกับเกรดในชั้นเรียน มีค่า ต่ำกว่า .35 เมื่อรวมคะแนนจากแบบทดสอบทั้ง 3 ฉบับเข้าด้วยกัน ค่าสัมประสิทธิ์สหสัมพันธ์เพิ่มขึ้นเป็น .51 ส่วนค่าสัมประสิทธิ์ระหว่างคะแนนที่ได้จากแบบทดสอบดั้งเดิมจำนวน 46 ข้อ กับเกรดในชั้นเรียนเป็น .68 และเมื่อใช้แบบทดสอบดั้งเดิมที่มีอำนาจจำแนกดีที่สุดเพียง 15 ข้อ จะได้ค่าสัมประสิทธิ์เป็น .52 วูด สรุปว่า แบบทดสอบดั้งเดิมจะดีกว่าแบบทดสอบรูปปิรามิดที่มีจำนวนข้อเท่า ๆ กันซึ่งขัดแย้งกับผลการวิจัยของแฮนเซน (Hansen, 1969) ซึ่งใช้แบบทดสอบรูปปิรามิด 5 แบบ ผ่านทางเครื่องรับโทรเลขที่พิมพ์เป็นตัวอักษรกับนักศึกษาชั้นปีที่ 1 ระดับวิทยาลัย จำนวนชั้นของแบบทดสอบแต่ละฉบับเป็น 3 หรือ 4 ชั้น รวมแล้วนักศึกษาจะต้องตอบข้อสอบทั้งสิ้น 17 ข้อ แฮนเซน ใช้ขนาดขั้น (step size) ที่มีค่า p = .10 ให้คะแนนด้วยวิธีต่างกัน 4 วิธี พบว่า คะแนนจากแบบทดสอบรูปปิรามิดมีความสัมพันธ์กับคะแนนจากการทดสอบในชั้นเรียนที่สอบในเนื้อหาเดียวกันโดยใช้เวลา 1 ชั่วโมง ดำเนินการสอบก่อนแบบทดสอบรูปปิรามิด 1 สัปดาห์ และยังมีความสัมพันธ์กับคะแนนวัดผลสัมฤทธิ์ที่ได้จากแบบทดสอบฉบับอื่นและเกรดในรายวิชานั้น ส่วนแบบทดสอบดั้งเดิมที่มีความยาวของแบบทดสอบเท่ากัน พบว่า มีความเชื่อมั่นชนิดความคงที่ภายในต่ำกว่าแบบทดสอบรูปปิรามิดทุกฉบับ คะแนนจากแบบทดสอบรูปปิรามิดมีการกระจายเป็นรูปสี่เหลี่ยมผืนผ้ามากกว่าแบบทดสอบดั้งเดิมซึ่งมีการกระจายเป็นแบบเบ้ทางลบ ผลการศึกษาแสดงให้เห็นว่า แบบทดสอบรูปปิรามิดใช้เวลาในการสอบน้อยกว่าเวลาที่ใช้กับแบบทดสอบดั้งเดิม 5 นาที แบบทดสอบรูปปิรามิดที่ให้คะแนนโดยวิธีการ 2 วิธี กับเกรดและแบบทดสอบวัดผลสัมฤทธิ์ที่ใช้เป็นเกณฑ์ มีความสัมพันธ์กันสูงกว่าแบบทดสอบดั้งเดิม การศึกษาครั้งที่สองของแฮนเซนยังคงได้รับผลคล้ายคลึงกัน

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ผลการวิจัยที่สนับสนุนแฮนเซนคือผลการวิจัยของ วอเตอร์และเบย์รอฟฟ์ (Waters and Bayroff, 1971 cited by Larkin and Weiss, 1974 : 8) ที่ศึกษาเปรียบเทียบแบบทดสอบรูปปิรามิดขนาด 5 10 15 ชั้นและขนาด 10 ชั้น ที่มีข้อสอบ 2 ข้อในแต่ละชั้นกับแบบทดสอบดั้งเดิมที่มีความยาวเท่ากัน ทั้งแบบทดสอบรูปปิรามิดและแบบทดสอบดั้งเดิมแตกต่างกันในด้านความแปรเปลี่ยนได้ของค่าความยาก ค่าอำนาจจำแนกของข้อสอบแปรผันไปตามระบบความสามารถได้รับการสมมุติว่ามีการกระจายเป็นปกติ ผลการศึกษาแสดงให้เห็นว่าคะแนนจากแบบทดสอบและความสามารถ ต่างก็สัมพันธ์กับการกระจายของค่าความยากและค่าอำนาจจำแนกของข้อสอบ แต่ค่าสหสัมพันธ์จากแบบทดสอบรูปปิรามิด มีสูงกว่าแบบทดสอบดั้งเดิม โดยเฉพาะอย่างยิ่งกับข้อสอบที่มีค่าอำนาจจำแนกสูง ๆ แบบทดสอบรูปปิรามิดที่มีข้อสอบ 1 ข้อในแต่ละชั้นมีความสัมพันธ์ของคะแนนจากแบบทดสอบกับความสามารถ สูงกว่าชนิดที่มี 2 ข้อในแต่ละชั้น

คิงส์เบอรี่ และไวส์ (Kingsbury and Weiss, 1979 cited by Weiss and Kingsbury, 1984 : 365) ศึกษาเปรียบเทียบกระบวนการทดสอบความรอบรู้แบบเทเลอร์ (Adaptive Mastery Testiny : AMT) กับกระบวนการทดสอบความรอบรู้แบบดั้งเดิม (Conventional Mastery Testing) แบบทดสอบ AMT แบ่งเป็น 4 แบบตามลักษณะกลุ่มข้อสอบ (Item Pool) แบบแรกประกอบด้วยข้อสอบที่มีค่าพารามิเตอร์ทุกตัวเท่ากันหรือใกล้เคียงกันที่สุดเรียกว่า Uniform Item Pool แบบที่ 2 ประกอบด้วยข้อสอบที่มีค่าพารามิเตอร์ต่างกันเฉพาะค่าความยาก เรียกว่า b - Variable Item Pool แบบที่ 3 ประกอบด้วยข้อสอบที่มีค่าพารามิเตอร์ต่างกัน 2 ตัวคือ ค่าความยากกับค่าอำนาจจำแนกเรียกว่า a-and b-Variable Item Pool และแบบที่ 4 ประกอบด้วยข้อสอบที่มีค่าพารามิเตอร์ต่างกัน 3 ตัว คือ ค่าความยาก ค่าอำนาจจำแนก และค่าสัมประสิทธิ์การเดา เรียกว่า a-b-and c-Variable Item Pool ทุกฉบับให้คะแนนตามวิธีของเบย์ (Bayesian Scoring) แบบทดสอบความรอบรู้แบบดั้งเดิม เป็นแบบทดสอบอิงเกณฑ์ (Critertion Test : CT) ใช้วิธีให้คะแนน 2 แบบคือ แบบสัดส่วนของคำตอบที่ถูกต้อง (Proportion of correct answers : PC) ใช้สัญญลักษณ์ CT / PC กับให้คะแนน

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ตามวิธีของเบย์ ใช้สัญญลักษณ์ CT / B แบบทดสอบแบบดั้งเดิมทั้ง 2 ฉบับคือ CT / PC และ CT / B จัดให้มีความยาว 3 ขนาดคือ 10 ข้อ 25 ข้อ และ 50 ข้อ ส่วนแบบทดสอบ AMT จะใช้ค่าความยาวเฉลี่ย (mean test lengths : MTL) ที่ใช้สอบเพื่อจำแนกการเป็นผู้รอบรู้ ใช้ค่าสหสัมพันธ์แบบฟาย (Phi correlation) ระหว่างนักเรียนที่รู้ก่อนแล้วว่าเป็นผู้รอบรู้ กับผลการประมาณความรอบรู้โดยแบบทดสอบ เป็นค่าความเที่ยงตรงในการจำแนก (classification validity) ยุทธวิธีการทดสอบจะเรียกได้ว่ามีประสิทธิภาพมากที่สุดก็ต่อเมื่อมันให้ค่าสหสัมพันธ์สูงสุดในทุก ๆ ความยาวของแบบทดสอบ หรือถ้ามันสามารถสร้างค่าสหสัมพันธ์สูงกว่ายุทธวิธีอื่น ๆ และเป็นแบบทดสอบที่สั้นที่สุด ผลการศึกษาพบว่า AMT ทั้ง 4 แบบ มีประสิทธิภาพสูงกว่าแบบทดสอบดั้งเดิม รวมทั้งมีค่าความเที่ยงตรงในการจำแนกสูงกว่าด้วย

ลาร์คิน และ ไวส์ (Larkin and Weiss, 1975 : 1 - 27) ศึกษาเปรียบเทียบผลการใช้แบบทดสอบรูปปิรามิดขนาดขั้นคงที่ 15 ชั้น ใช้กฎการแยกทางแบบเพิ่ม 1 ลด 1 จำนวน 3 ฉบับ ตรวจให้คะแนน 4 วิธีคือ 1) ให้ตามจำนวนข้อที่ตอบถูก 2) ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ทำ 3) ให้ตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก 4) ให้ตามค่าความยากของข้อสอบข้อสุดท้ายที่ทำ เปรียบเทียบกับแบบทดสอบยุทธวิธี 2 ชั้นที่มีข้อสอบนำร่อง 10 ข้อ แบบทดสอบวัดผล 4 ชุด ๆ ละ 30 ข้อ ให้คะแนนด้วยวิธีประมาณค่าความสามารถตามวิธีของลอร์ด ผลการศึกษาพบว่า การจัดลำดับก่อนหลังในการทดสอบไม่มีผลต่อคะแนนที่ได้ การกระจายของคะแนนจากแบบทดสอบรูปปิรามิดให้คะแนนตามจำนวนข้อที่ตอบถูก มีลักษณะเบ้ทางบวกและแตกต่างจากโค้งปกติอย่างมีนัยสำคัญ แบบทดสอบรูปปิรามิดที่ให้คะแนนวิธีอื่น มีการกระจายไม่ต่างจากโค้งปกติ ส่วนแบบทดสอบยุทธวิธีสองขั้น ให้ผลที่มีการกระจายเป็นรูปเบ้ทางบวกและโด่งแตกต่างจากโค้งปกติอย่างมีนัยสำคัญ อย่างไรก็ตาม คะแนนจากแบบทดสอบรูปปิรามิดที่ให้คะแนน 4 แบบ กับคะแนนจากแบบทดสอบยุทธวิธีสองขั้นก็มีความสัมพันธ์กันสูงถึง .84 .81 .79 และ .83 โดยวิธีหาค่าสหสัมพันธ์แบบเพียร์สันโปรดัคโมเมนต์

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

ประเวศ เวชชะ (2530) ศึกษาความเที่ยงตรงในการจำแนกความรอบรู้ของแบบทดสอบแบบเทเลอร์ แบบแตกแขนง แบบตามสภาพ แบบทดลองทำ และแบบให้ตัดสินใจเลือกเอง เปรียบเทียบกับการใช้แบบทดลองอิงเกณฑ์ขนาดความยาว 10, 20 และ 30 ข้อ โดยใช้กลุ่มข้อสอบที่วิเคราะห์ตามวิธีราชช์ (Rasch) พบว่าความเที่ยงตรงในการจำแนกความรอบรู้ไม่แตกต่างกันทางสถิติ ยกเว้นระหว่างแบบทดสอบอิงเกณฑ์ 10 ข้อ กับแบบตามสภาพเท่านั้นที่ต่างกันอย่างมีนัยสำคัญ โดยที่การทดสอบแบบเทเลอร์ให้ความถูกต้องในการจำแนกความรอบรู้สูงกว่า

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved