

วรรณคดีและงานวิจัยที่เกี่ยวข้อง

จากการศึกษาบทความ งานวิจัย และเอกสารอื่นๆ ที่เกี่ยวข้องกับ
แบบสอบเฉพาะบุคคล ผู้วิจัยขอเสนอสาระต่างๆ ในหัวข้อ ต่อไปนี้

1. การทดสอบเฉพาะบุคคล
 - 1.1 หลักการทดสอบเฉพาะบุคคล
 - 1.2 ส่วนประกอบของการทดสอบเฉพาะบุคคล
 - 1.3 การดำเนินการสอบ
 - 1.4 ประเภทของการทดสอบเฉพาะบุคคล
 - 1.5 วิธีการให้คะแนนการทดสอบเฉพาะบุคคล
2. ทฤษฎีการตอบสนองข้อสอบ
 - 2.1 คุณสมบัติของรูปแบบการตอบข้อสอบ
 - 2.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ
 - 2.3 ความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ
 - 2.4 ความคลาดเคลื่อนมาตรฐานในการประมาณค่า
 - 2.5 การนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้
3. งานวิจัยที่เกี่ยวข้องกับแบบสอบเฉพาะบุคคล
 - 3.1 งานวิจัยต่างประเทศ
 - 3.2 งานวิจัยในประเทศไทย

1. การทดสอบเฉพาะบุคคล (Tailored Testing)

1.1 ความหมายและหลักการในการทดสอบเฉพาะบุคคล

การทดสอบเฉพาะบุคคล (Tailored Testing) เป็นการทดสอบ

ที่จัดข้อสอบให้มีความยากเหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน (Lord, 1971)

และผู้สอบแต่ละคนก็ไม่จำเป็นต้องทำข้อสอบเหมือนกันทุกข้อ จำนวนข้อก็ไม่จำเป็นต้องเท่ากัน ทั้งนี้ขึ้นอยู่กับรูปแบบและผลการตอบข้อสอบในการทดสอบเฉพาะบุคคลแบบนั้นๆ (Weiss and Kingsbury, 1984)

หลักการทั่วไป จะให้ผู้สอบเริ่มทำข้อสอบที่มีความยากปานกลางก่อน ถ้าผลการตอบข้อสอบถ่วงข้อต่อไปจะมีความยากมากขึ้น หากตอบข้อสอบข้อนั้นผิดข้อต่อไปที่ได้รับก็จะง่ายขึ้น การทำข้อสอบขั้นต่อไปจะดำเนินการเช่นนั้นไปจนสิ้นสุดการทดสอบ (Green and others, 1984)

1.2 ส่วนประกอบของการทดสอบเฉพาะบุคคล

การทดสอบเฉพาะบุคคล เป็นวิธีการทดสอบที่มีรูปแบบแตกต่างไปจากวิธีการทดสอบแบบประเพณีนิยม เนื่องจากการทดสอบเฉพาะบุคคลจะต้องกำหนดส่วนประกอบต่างๆ ที่ใช้ในการดำเนินการสอบไว้ล่วงหน้าทั้งสิ้น ซึ่งส่วนประกอบที่กล่าวนี้ประกอบด้วย (Weiss, 1984)

1. ข้อสอบ ตามทฤษฎีการตอบสนองข้อสอบ ซึ่งมีข้อตกลงเบื้องต้นว่าคุณลักษณะที่สอบวัดต้องเป็นมิติเดียวกัน ค่าพารามิเตอร์ของข้อสอบมีความคงที่ข้ามกลุ่ม และมีความเป็นอิสระในการตอบข้อสอบ (Lord and Novick, cited in Koch, B.R. and Reckase, M.D. 1978) ดังนั้นการทดสอบเฉพาะบุคคลจึงสามารถประยุกต์ใช้รูปแบบที่มี 1, 2 หรือ 3 พารามิเตอร์ และอาจใช้รูปแบบนอร์มอล โอจีฟ (Normal Ogive Models) ก็ได้ การเลือกใช้รูปแบบใดก็ควรพิจารณาธรรมชาติของข้อสอบและต้องเหมาะสมกับข้อมูลการตอบข้อสอบด้วย แต่จากงานวิจัยของแมคคินเลย์และเรคเคส (Mc Kinley, R.L. and Reckase, M.D. 1980) และคอคกับเรคเคส (Koch, B.R. and Reckase, M.D. 1978) เปรียบเทียบการทดสอบเฉพาะบุคคลโดยใช้รูปแบบการตอบข้อสอบที่มี 1 พารามิเตอร์และ 3 พารามิเตอร์ พบว่ารูปแบบที่มี 3 พารามิเตอร์ มีความเหมาะสมมากกว่ารูปแบบที่มี 1 พารามิเตอร์

และจากการศึกษาของธอร์นไคค์ (Thronkide, 1982) เกี่ยวกับแบบสอบเฉพาะบุคคลชี้ให้เห็นว่า ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของการทดสอบเฉพาะบุคคลขึ้นอยู่กับจำนวนข้อและค่าอำนาจจำแนก กล่าวคือ ข้อสอบที่มีค่าอำนาจจำแนกสูงขึ้นค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าก็จะลดลง โดยปกติแล้วถ้าข้อสอบที่มีค่าอำนาจจำแนกโดยเฉลี่ย 0.7 ควรจะใช้ข้อสอบประมาณ 10 ถึง 15 ข้อ

2. กลุ่มข้อสอบ (Item Pool) จะต้องประกอบด้วยข้อสอบที่ประมาณค่าตามทฤษฎีการตอบสนองข้อสอบ พารามิเตอร์ของข้อสอบทุกข้อต้องได้รับการจัดให้อยู่ใน

มาตรฐานวัดเดียวกันโดยใช้กระบวนการที่เหมาะสม ข้อสอบในกลุ่มข้อสอบจะต้องครอบคลุมช่วงของระดับความสามารถในประชากรและวัดความสามารถมิติเดียวกัน เออร์รี่ (Urry, 1984) กล่าวว่า กลุ่มข้อสอบควรจะประกอบด้วยข้อสอบอย่างน้อย 100 ข้อ ค่าอำนาจจำแนกของข้อสอบเกิน .8 ความยากกระจายอยู่ในช่วง -3.00 ถึง $+3.00$ และค่าการแตกต่องน้อยกว่า .3

3. ระดับในการเริ่มต้น (Entry Level) ในการทดสอบเฉพาะบุคคล ผู้สอบแต่ละคนจะได้รับข้อสอบต่างกัน ดังนั้นจึงสามารถที่จะให้ผู้สอบเริ่มต้นทำข้อสอบที่มีระดับความยากที่ต่างกันได้ การเริ่มทำข้อสอบที่คลาดเคลื่อนไปจากความสามารถนั้นไม่มีผลกระทบต่อการประมาณความสามารถเท่าใดนัก แต่ถ้าให้ผู้สอบเริ่มทำข้อสอบที่ตรงกับความสามารถของเขาก็จะทำให้ลดจำนวนข้อที่จะใช้ในการทดสอบลง

4. กฎการเลือกข้อสอบ (Item selection rule) จะเลือกใช้ข้อสอบที่สามารถให้สารสนเทศสูงสุด ณ ระดับความสามารถของผู้สอบ

การเลือกข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบ นอกจากจะพิจารณาจากข้อที่ให้สารสนเทศสูงสุดแล้ว ผลการศึกษาเชิงประจักษ์ของลอร์ด (Lord, 1970; 1971) ก็ให้เห็นว่าคะแนนจากแบบสอบจะสะท้อนให้เห็นความสามารถของแต่ละบุคคลได้แม่นยำ เมื่อความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของแต่ละบุคคลเป็น .50 ซึ่งสอดคล้องกับผลการศึกษาของฮิค (Hick, 1951; อ้างถึงในนันทิยา พิงคำ, 2531) จึงสรุปได้ว่าแบบสอบจะมีความแม่นยำของการวัดต่ำ เช่น มีความคลาดเคลื่อนมาตรฐานสูง (Green, 1970) เมื่อความน่าจะเป็นของการตอบข้อสอบถูกของแต่ละบุคคลมีค่ามากหรือน้อยกว่า .50 (Weiss, 1974)

5. เกณฑ์การยุติการสอบ (Terminal criterion) โดยพิจารณาว่าการวัดผลแต่ละบุคคลนั้น ช่วยให้เราสามารถกำหนดระดับความสามารถของผู้สอบได้อย่างแม่นยำ โดยขึ้นอยู่กับรูปแบบของการทดสอบเฉพาะบุคคลที่เลือกใช้ เช่น รูปแบบปิรามิดขนาดชั้นคงที่ 10 ชั้น จะยุติการสอบเมื่อผู้สอบทำข้อสอบครบ 10 ข้อ เป็นต้น

1.3 การดำเนินการสอบ

ผู้สอบจะสอบทีละข้อ ผลการตอบข้อสอบแต่ละข้อจะถูกประมาณความสามารถเพื่อคัดเลือกข้อสอบให้เหมาะสมกับระดับความสามารถในข้อที่จะตอบต่อไป กระบวนการจะดำเนินเช่นนี้จนสิ้นสุดการทดสอบ วิธีการนี้เหมาะที่จะใช้กับเครื่องคอมพิวเตอร์ (Urry, 1977; Weiss and Kingsbury, 1984)

การดำเนินการสอบ เริ่มต้นโดยการให้ผู้สอบนั่งหน้าจอคอมพิวเตอร์



ซึ่งอาจเป็นเครื่องไมโครคอมพิวเตอร์หรือเครื่องมินิคอมพิวเตอร์ ที่มีจานขับแม่เหล็ก (disk drive) ที่สามารถนำแผ่นบันทึกข้อมูล (diskette) ใส่หรือถอดออกมาได้ ในแผ่นบันทึกข้อมูลจะมีโปรแกรมการทดสอบที่สร้างขึ้นตามองค์ประกอบต่างๆ ของการทดสอบเฉพาะบุคคล ดังกล่าวข้างต้น ผู้ดำเนินการสอบจัดเตรียมเครื่องให้อยู่ในสภาพที่พร้อมจะทดสอบ คอมพิวเตอร์จะให้คำแนะนำในการทำและรับข้อมูลส่วนตัวของนักเรียน เช่น ชื่อ-ชื่อสกุล เลขประจำตัว ชั้น/ห้อง เกรดเฉลี่ย เป็นต้น หลังจากนั้น คอมพิวเตอร์จะแสดงข้อสอบข้อแรกบนจอภาพเพื่อให้ผู้สอบได้ตอบ ข้อสอบข้อแรกนี้เป็นข้อสอบที่มีค่าความยากปานกลาง หรือเป็นข้อสอบที่คัดเลือกตามกฎในการคัดเลือกข้อสอบก็ได้ ถ้าผู้สอบตอบข้อสอบข้อแรกถูก ข้อสอบข้อถัดไปก็จะเป็นข้อสอบที่มีความยากเพิ่มขึ้น หรือถ้าผู้สอบตอบข้อสอบข้อแรกผิดข้อสอบข้อถัดไปก็จะเป็นข้อสอบที่มีความยากลดลง จากนั้นก็จะประมาณค่าความสามารถ การทดสอบจะดำเนินต่อไปตามแนวทางต่อไปนี้ คือ

1. หลังจากประมาณค่าความสามารถของผู้สอบแล้ว คอมพิวเตอร์ก็จะคัดเลือกข้อสอบข้อใหม่ที่มีประโยชน์ในการกำหนดความสามารถจริง
2. แสดงข้อสอบข้อใหม่บนจอภาพและผู้สอบตอบข้อสอบโดยการเลือกตัวเลือกที่ต้องการ แล้วพิมพ์ตัวเลือกที่ต้องการ คอมพิวเตอร์จะรับคำตอบแล้วตรวจสอบ
3. ประมาณค่าความสามารถของผู้สอบอีกครั้งหนึ่งตามการตอบข้อสอบข้อที่ผ่านมา
4. พิจารณาว่าผู้สอบควรจะทำข้อสอบข้อต่อไปหรือไม่ตามเกณฑ์การยุติการสอบที่กำหนดไว้
5. ถ้าผู้สอบจะต้องทำข้อสอบข้อต่อไปก็ย้อนกลับไปขั้นที่ 1 อีกครั้ง เมื่อสิ้นสุดการทดสอบสารสนเทศเกี่ยวกับการตอบข้อสอบของผู้สอบก็จะถูกบันทึกข้อมูลและแสดงผลการทดสอบให้ผู้สอบทราบ

1.4 ประเภทของการทดสอบเฉพาะบุคคล

การทดสอบเฉพาะบุคคล จัดได้เป็น 2 ประเภทใหญ่ๆ (Weiss, 1974 ; Hambleton and Swaminathan, 1985) ในแต่ละประเภทยังแบ่งย่อยได้อีกดังนี้

- (1). กลยุทธ์สองขั้นตอน (Two - Stage Strategies)
- (2). กลยุทธ์หลายขั้นตอน (Multi-stage Strategies)

ซึ่งแบ่งเป็น

- 2.1 รูปแบบแยกทางคงที่ (Fixed Branching Model) ได้แก่

2.1.1 รูปแบบปิรามิด (Pyramidal Model) ได้แก่

- ก. รูปแบบปิรามิดขนาดขั้นคงที่
(Constant Step Size Pyramidal Model)
- ข. รูปแบบปิรามิดขนาดขั้นแปรผัน
(Variable Step Size Pyramidal Model)
- ค. รูปแบบปิรามิดข้างตัด
(Truncated Pyramidal Model)
- ง. รูปแบบปิรามิดแบบมีหลายข้อในแต่ละขั้น
(Multiple-Item Pyramidal Model)
- จ. รูปแบบปิรามิดแบบให้น้ำหนักแก่ตัวเลือกเพื่อแยกทาง
(Differential Response Option
Branching Pyramidal Model)

2.1.2 รูปแบบเฟล็กซีเลเวล (Flexilevel Model)

2.1.3 รูปแบบปรับระดับขั้น (Stradaptive Model)

2.2 รูปแบบแยกทางแปรผัน (Variable Branching Model) ได้แก่

2.2.1 กลยุทธ์เบย์ (Bayesian Strategies)

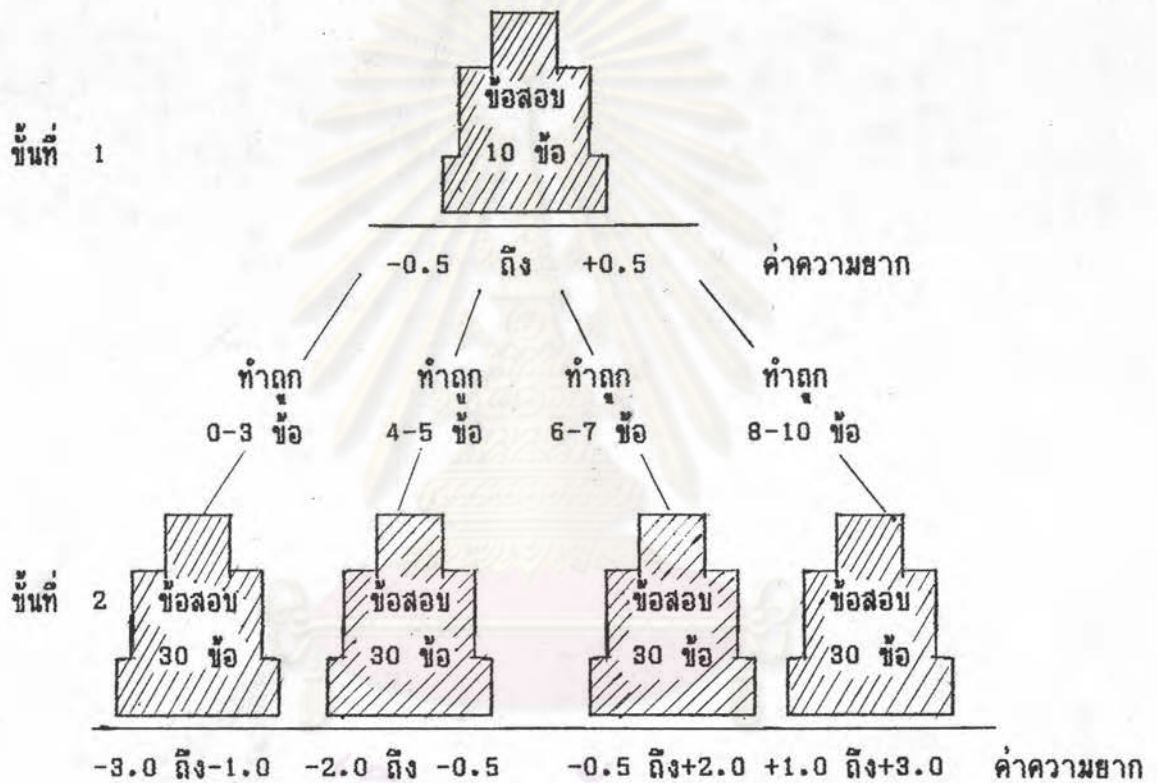
2.2.2 กลยุทธ์ความเป็นไปได้สูงสุด (Maximum Likelihood Strategies)

(1). การทดสอบเฉพาะบุคคลโดยใช้กลยุทธ์สองขั้นตอน

การทดสอบเฉพาะบุคคลโดยใช้กลยุทธ์สองขั้นตอน จะใช้ข้อสอบ 2 ชุด ข้อสอบชุดแรก (Routing Test) ซึ่งผู้สอบทุกคนต้องสอบเหมือนกัน สอบเพื่อแยกระดับความสามารถเพื่อที่จะไปรับข้อสอบในชุดที่สอง โดยปกติข้อสอบชุดแรกจะมีประมาณ 10 ข้อ ส่วนข้อสอบชุดที่สอง (Main Test) จะแบ่งเป็นข้อสอบชุดย่อยๆ หลายชุดตามระดับความยาก โดยปกติจะมีชุดละประมาณ 20-30 ข้อ ผู้สอบแต่ละคนจะได้รับข้อสอบชุดที่สองไม่เหมือนกัน ผู้สอบที่ได้รับการประเมินจากข้อสอบชุดแรกว่ามีความสามารถสูงจะได้รับข้อสอบชุดที่สองที่ยาก ส่วนผู้สอบที่มีความสามารถปานกลางจะได้รับข้อสอบชุดที่สองที่มีความยากปานกลาง และผู้สอบที่มีความสามารถต่ำจะได้รับข้อสอบชุดที่สองที่ง่าย (Lord, 1971 ; Betz and Weiss, 1973)

ลอร์ด (Lord, 1980) ได้แนะนำว่าข้อสอบชุดย่อย (Subtest) แต่ละฉบับ
 ในขั้นที่สอง (Main Test) ควรมีข้อสอบบางส่วนที่คาบเกี่ยวกันในเรื่องของความยาก
 ของข้อสอบรายชื่อ (Overlap) ถ้าให้ m คือ จำนวนข้อสอบที่คาบเกี่ยวกัน L คือ
 จำนวนฉบับชุดย่อยในขั้นที่สอง n คือ จำนวนข้อสอบทั้งหมด n_1 คือ จำนวนข้อสอบใน
 ชุดย่อยแต่ละฉบับในขั้นที่สอง จะได้ $m = (n_1L - n)/(L-1)$

โครงสร้างของการทดสอบเฉพาะบุคคลแบบสองขั้นตอน ดังตัวอย่าง ภาพที่ 1



ภาพที่ 1 โครงสร้างของการทดสอบเฉพาะบุคคลแบบสองขั้นตอน

จากภาพที่ 1 ผู้สอบทุกคนต้องทำข้อสอบในขั้นที่ 1 ซึ่งมีอยู่ 10 ข้อ ผู้สอบที่ทำ
 ข้อสอบถูก 0-3 ข้อ จะไปทำข้อสอบขั้นที่ 2 ที่ง่ายที่สุด ส่วนผู้สอบที่ทำข้อสอบขั้นที่ 1 ถูก
 4-5 ข้อ, 6-7 ข้อ หรือ 8-10 ข้อ จะแยกไปทำข้อสอบในขั้นที่ 2 ที่ยากขึ้นตามลำดับ

(2). การทดสอบเฉพาะบุคคลโดยใช้กลยุทธ์หลายขั้นตอน

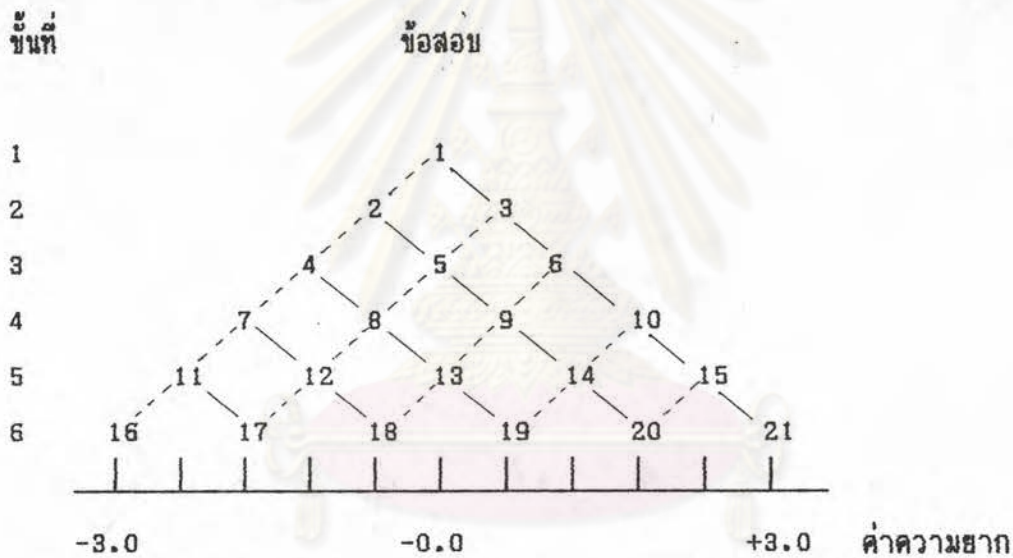
2.1 รูปแบบแยกทางคงที่

การทดสอบเฉพาะบุคคลโดยใช้กลยุทธ์หลายขั้นตอน (Multi-stage

Strategies) แบบแยกทางคงที่ (Fixed Branching Model) หมายถึง การทดสอบเฉพาะบุคคลที่มีหลายขั้นตอน ในแต่ละขั้นตอนอาจจะมีข้อสอบเพียงข้อเดียวหรือหลายข้อก็ได้ โดยมีรูปแบบหรือเส้นทางในการตอบข้อสอบที่แน่นอน ซึ่งได้แก่

2.1.1 รูปแบบปิรามิด

ก. การทดสอบเฉพาะบุคคลรูปแบบปิรามิดขนาดขั้นคงที่ (Constant Step Size Pyramidal Models) ได้แก่ การทดสอบเฉพาะบุคคลที่จัดเรียงข้อสอบเป็นรูปสามเหลี่ยม โดยมีจำนวนข้อสอบในแต่ละขั้นเท่ากับลำดับที่ของขั้น กล่าวคือ ขั้นที่ 1 จะมีข้อสอบหนึ่งข้อ ขั้นที่ 2 จะมีข้อสอบสองข้อ ในขั้นที่ 5 จะมีข้อสอบห้าข้อ ดังนั้น การทดสอบเฉพาะบุคคลรูปแบบปิรามิดที่มี 6 ขั้น จะมีข้อสอบทั้งหมด 21 ข้อ ดังภาพที่ 2

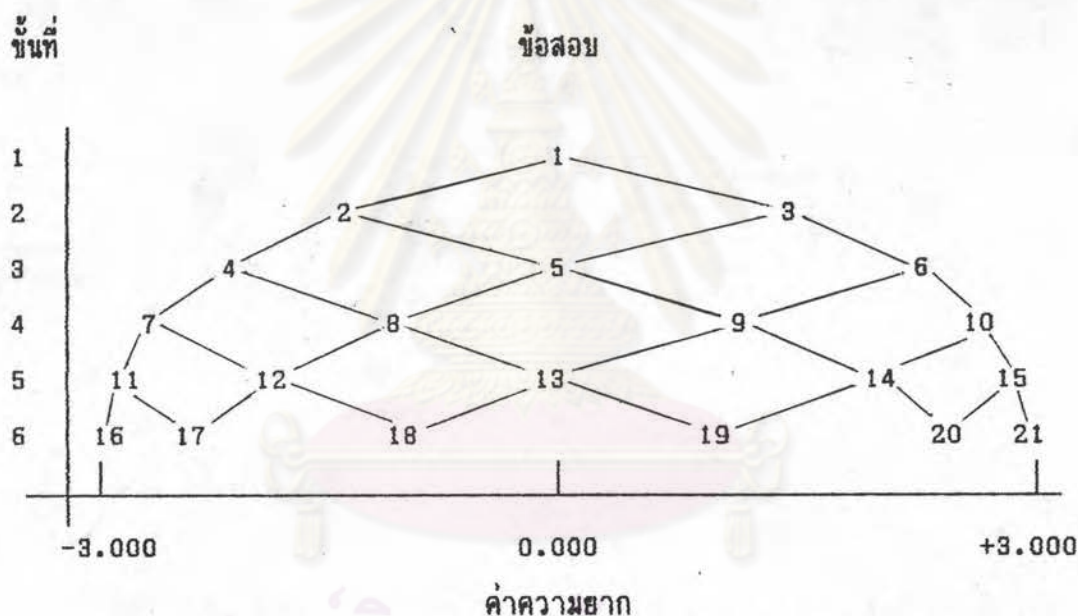


ภาพที่ 2 โครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบปิรามิดขนาดขั้นคงที่

จากภาพที่ 2 แกนนอนเป็นค่าความยากของข้อสอบ (b) ที่สัมพันธ์กับข้อสอบในการทดสอบเฉพาะบุคคลรูปแบบปิรามิด ซึ่งจะเห็นว่าความยากมีค่าอยู่ระหว่าง -3.0 ถึง +3.0 โดยแบ่งออกเป็น 10 ช่วงเท่าๆ กัน ฉะนั้นค่าความยากจะห่างกันช่วงละ 0.6 ข้อสอบที่อยู่ในแนวตั้งแนวเดียวกัน จะมีค่าความยากเท่ากัน และช่วงห่างค่าความยากของข้อ ที่อยู่ติดกันภายในขั้นมีค่าเท่ากันตลอด คือ 1.2

ในการทำแบบสอบ ผู้สอบจะต้องทำข้อสอบขั้นละหนึ่งข้อ โดยเริ่มทำข้อที่อยู่บนยอดสามเหลี่ยมซึ่งมีความยากปานกลางก่อน ถ้าทำถูกขั้นต่อไปจะแยกไปทำข้อที่ยากขึ้น แต่ถ้าทำผิดขั้นต่อไปจะไปทำข้อที่ง่ายขึ้น กระบวนการทดสอบจะเป็นเช่นนี้จนถึงขั้นสุดท้าย

ข. การทดสอบเฉพาะบุคคลรูปแบบปิรามิดขนาดขั้นแปรผัน
(Variable Step Size Pyramidal Models) โดยที่แพตเตอร์สัน และลอร์ด
(Weiss, 1974 ; Citing Paterson, 1962 ; Lord, 1971) มีความเห็นว่า
รูปปิรามิดแบบขนาดขั้นคงที่ขาดประสิทธิภาพเกี่ยวกับความไวในการกำหนดเส้นทางการตอบ
ข้อสอบ จึงได้กำหนดช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในขั้นให้มีความไม่เท่ากันโดยให้
ข้อที่อยู่กลางๆ มีช่วงห่างของค่าความยากมาก แล้วค่อยๆ ลดลงในข้อที่ค่อนข้างง่ายหรือ
ยาก แพตเตอร์สันได้กำหนดการเพิ่มหรือลดนั้น จะเพิ่มหรือลดไปยังตำแหน่งกึ่งกลางระหว่าง
2 ข้อ ในขั้นที่มาก่อน หรือระหว่างข้อในขั้นที่มาก่อนกับค่าความยากสูงสุดหรือต่ำสุด ดังตัวอย่าง
ภาพที่ 3

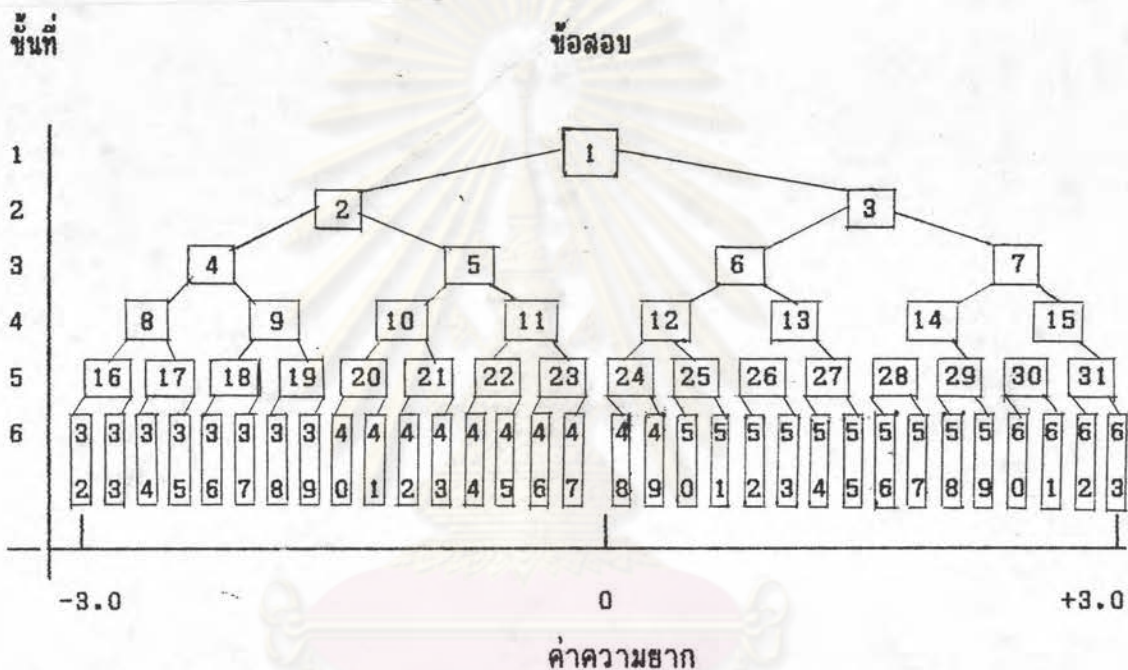


ภาพที่ 3 โครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบปิรามิดขนาดขั้นแปรผัน

จากภาพที่ 3 ในขั้นที่ 1 มีข้อสอบเพียงข้อเดียว คือ ข้อ 1 ซึ่งมีค่าความยาก
เท่ากับ 0.000 ในขั้นที่ 2 มีข้อสอบ 2 ข้อ คือข้อ 2 มีค่าความยากอยู่กึ่งกลาง
ระหว่าง -3.000 (ความยากต่ำสุด) กับ 0.000 (ความยากของข้อที่ 1) ซึ่งก็คือ -1.500
ข้อ 3 มีความยากอยู่กึ่งกลางระหว่าง +3.000 (ความยากสูงสุด) กับ 0.000 (ความยาก
ของข้อที่ 1) ซึ่งก็คือ 1.500 ในขั้นที่ 3 มีข้อสอบสามข้อ คือข้อ 4 มีความยากอยู่
กึ่งกลางระหว่าง -3.000 (ความยากต่ำสุด) กับ -1.500 (ความยากของข้อที่ 2) ซึ่ง
ก็คือ -2.250 ข้อ 5 มีความยากเท่ากับ 0.000 ข้อ 6 มีความยากอยู่กึ่งกลางระหว่าง

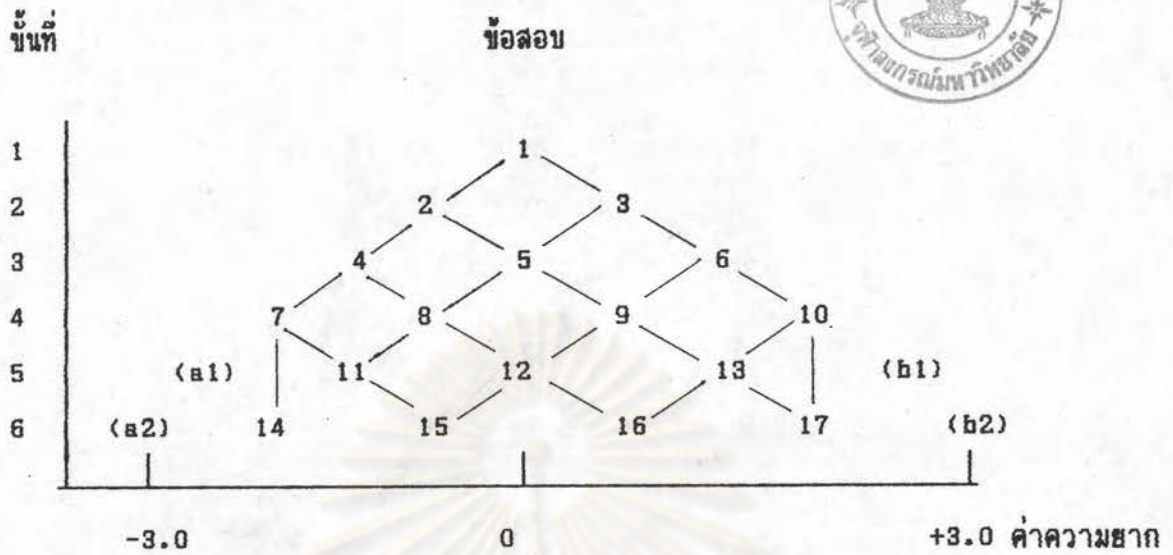
1.500 (ความยากของข้อที่ 3) กับ 3.000 (ความยากสูงสุด) ซึ่งก็คือ 2.250 สำหรับในข้ออื่นๆ ก็ในทำนองเดียวกัน

จากรูปแบบปิรามิดขนาดขั้นแปดขั้นนี้ ต่อมา ลอร์ด (Weiss, 1974 citing Lord, 1971) ได้ใช้วิธีการเพิ่มจำนวนข้อในแต่ละขั้นให้เป็นสองเท่าของจำนวนข้อสอบในขั้นที่มาก่อน ซึ่งให้ชื่อรูปแบบใหม่นี้ว่า แบบรอบบิน-มอนโร (Robin-Monro) ดังตัวอย่างภาพที่ 4



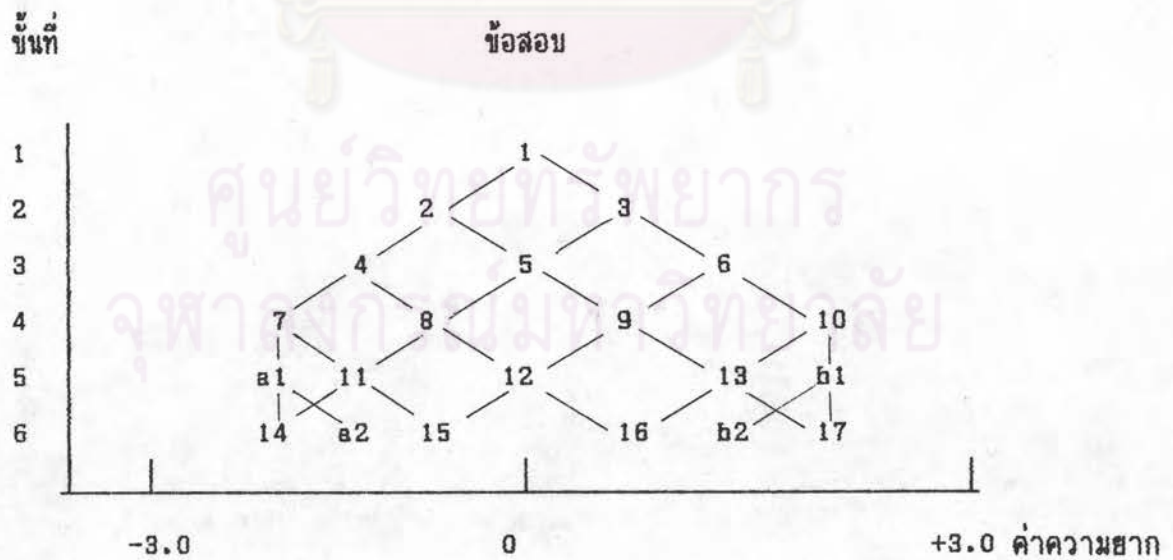
ภาพที่ 4 โครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบปิรามิดแบบรอบบิน-มอนโร

ค. การทดสอบเฉพาะบุคคลรูปแบบปิรามิดแบบข้างตัด (Truncated Pyramidal Models) เป็นรูปแบบที่มัสซิโอ (Weiss, 1974; citing Mussio, 1973) เป็นผู้เสนอรูปแบบนี้ขึ้นมา เพื่อให้ลดจำนวนข้อสอบลงจากรูปแบบปิรามิดแบบขนาดขั้นคงที่ จึงใช้วิธีสกัดการสะท้อนกลับ หรือวิธีการรักษาการสะท้อนกลับ (Reflecting or Retaining Barrier) ดังตัวอย่าง ภาพที่ 5 และ 6



ภาพประกอบ 5 โครงสร้างของการทดสอบเฉพาะบุคคลรูปปิรามิดแบบข้างตัด (Truncated Pyramid) สกักการสะท้อนกลับ (Reflecting Barrier)

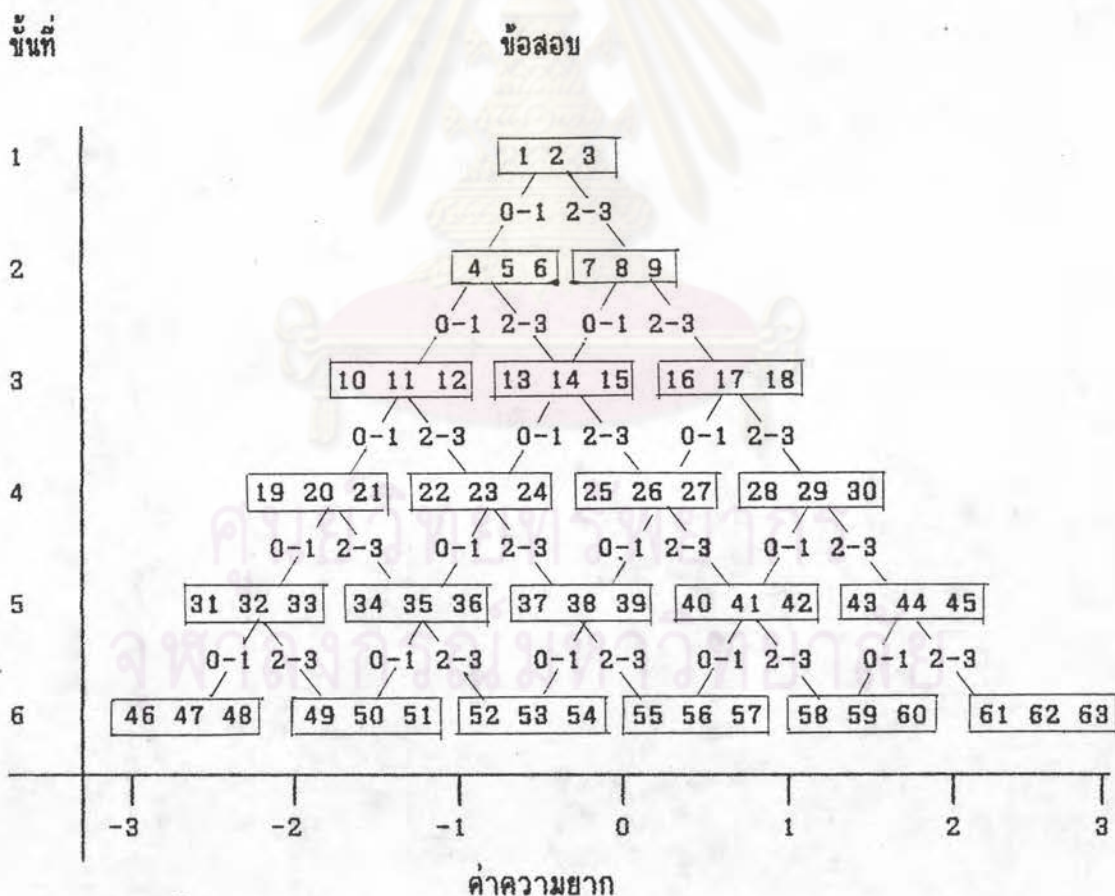
จากภาพที่ 5 การตอบข้อสอบในชั้นที่ 1 ถึงชั้นที่ 3 มีวิธีตอบเช่นเดียวกับรูปแบบปิรามิดขนาดชั้นคงที่ จนถึงจุดสกัดในชั้นที่ 4 ถ้าตอบข้อสอบข้อที่ 7 ถูก ก็จะแยกไปตอบข้อ 11 แต่ถ้าตอบข้อ 7 ผิด ก็จะลงไปตอบข้อ 14 และในทำนองเดียวกัน ถ้าตอบข้อ 10 ผิด ก็จะลงไปตอบข้อ 17 เลข



ภาพที่ 6 โครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบปิรามิดแบบข้างตัด (Truncated Pyramid) รักษาการสะท้อนกลับ (Retaining Barrier)

จากภาพที่ 6 การตอบข้อสอบในชั้นที่ 1 ถึงชั้นที่ 3 วิธีตอบเช่นเดียวกับรูปแบบปิรามิดขนาดชั้นคงที่ จนถึงจุดสกัดในชั้นที่ 4 ถ้าตอบข้อสอบข้อ 7 ถูก ก็จะแยกไปตอบข้อ 11 แต่ถ้าตอบข้อ 7 ผิด ก็จะลงไปตอบข้อ a1 ซึ่งมีความยากเท่ากับข้อ 7 และถ้าตอบข้อ a1 ผิดอีกก็จะลงไปตอบข้อ 14 แต่ถ้าตอบข้อ a1 ถูก ก็จะแยกไปตอบข้อ a2 ซึ่งมีความยากเท่ากับข้อ 11

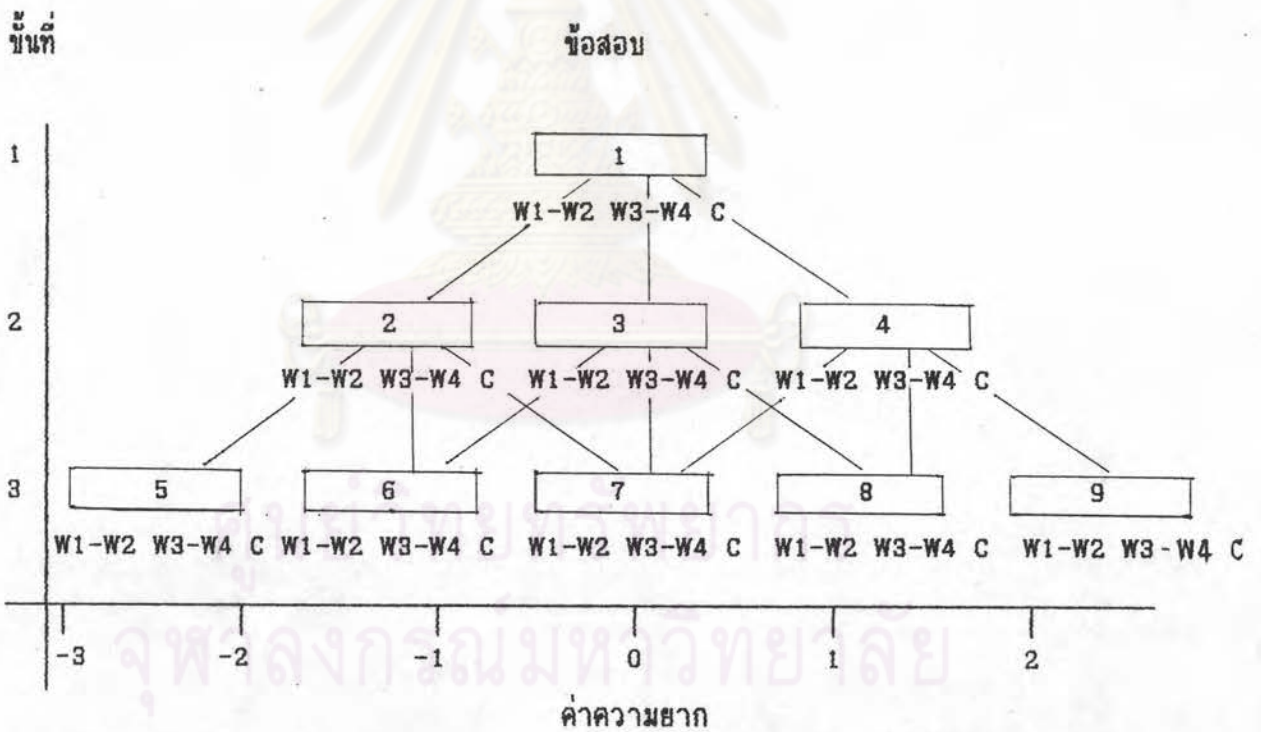
ง. การทดสอบเฉพาะบุคคลรูปแบบปิรามิดแบบมีหลายข้อในแต่ละชั้น (Multiple Item Pyramid) เป็นรูปแบบที่เครธไวสกับฮายเซอร์ และลินน์ (Weiss, 1974 ; citing Krathworth and Huyser, 1956 ; Linn, 1969) เป็นผู้เสนอรูปแบบนี้ขึ้นมา โดยให้มีจำนวนข้อในแต่ละชั้นมากกว่าหนึ่งข้อ ทั้งนี้เพื่อให้การตอบเพื่อแยกทางได้ถูกต้องมากยิ่งขึ้น ดังตัวอย่าง ภาพที่ 7



ภาพที่ 7 โครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบปิรามิดแบบมีหลายข้อในแต่ละชั้น (Multi-Item Model)

จากภาพที่ 7 ในขั้นที่ 1 ถ้าผู้สอบตอบข้อสอบถูก 0 ถึง 1 ข้อ จะแยกไปทำข้อสอบข้อที่ 4, 5 และ 6 ถ้าตอบข้อสอบถูก 2 ถึง 3 ข้อ ก็จะแยกไปทำข้อสอบข้อที่ 7, 8 และ 9 สำหรับในขั้นอื่นๆ ก็ในทำนองเดียวกัน

จ. การทดสอบเฉพาะบุคคลรูปแบบปิรามิดแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง (Differential Response Option Branching Model) เป็นรูปแบบที่เบย์รอฟฟ์และซีเลีย (Weiss, 1974 citing Bayroff and Seeley, 1967) เป็นผู้เสนอขึ้น ซึ่งมีรูปแบบคล้ายกับแบบที่ใช้ข้อสอบหลายๆ ข้อในแต่ละชั้น กล่าวคือ แทนที่จะใช้ข้อสอบหลายๆ ข้อในแต่ละชั้น ก็ใช้ตัวเลือกของข้อสอบแทน รูปแบบนี้เหมาะสำหรับข้อสอบเลือกตอบ ที่สามารถกำหนดน้ำหนักคะแนนของตัวเลือก ดังตัวอย่าง ภาพที่ 8



ภาพที่ 8 โครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบปิรามิดแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง (Differential Response Option Branching)

2.1.2 รูปแบบเฟล็กซ์ลีเวล

การทดสอบเฉพาะบุคคลรูปแบบเฟล็กซ์ลีเวล (Flexilevel Models) เป็นรูปแบบการทดสอบเฉพาะบุคคล ที่ลอร์ด ได้คิดขึ้นมาในครั้งแรกเพื่อใช้กับการทดสอบที่ใช้กระดาษกับดินสอ (Paper and Pencil Test) วิธีการ คือ แบ่งข้อสอบออกเป็น 2 กลุ่ม คือ ข้อสอบกลุ่มง่ายกับข้อสอบกลุ่มยาก และคัดเลือกข้อที่มีความยากอยู่กึ่งกลางระหว่างข้อสอบกลุ่มง่ายกับกลุ่มยาก ไว้ต่างหากอีกหนึ่งข้อ เพื่อเป็นข้อเริ่มต้นสำหรับผู้สอบทุกคน ภายในข้อสอบกลุ่มง่ายจะเรียงลำดับข้อสอบจากข้อที่ยากที่สุดภายในกลุ่มนี้ไปหาข้อที่ง่าย ข้อสอบกลุ่มง่ายนี้จะพิมพ์ด้วยหมึกสีแดง ส่วนข้อสอบในกลุ่มยากจะเรียงลำดับจากข้อที่ง่ายที่สุดภายในกลุ่มนี้ไปหาข้อที่ยาก และจะพิมพ์ด้วยหมึกสีน้ำเงิน ดังตัวอย่างภาพที่ 9 (Lord, 1971 ; Lord, 1980)

ข้อ 0. (มีความยากอยู่กึ่งกลางระหว่างข้อสอบกลุ่มง่ายกับกลุ่มยาก)	
(ข้อสอบกลุ่มง่าย) (พิมพ์สีแดง)	(ข้อสอบกลุ่มยาก) (พิมพ์สีน้ำเงิน)
ข้อ 1. (ข้อที่ยากที่สุดในกลุ่มนี้)	ข้อ 1. (ข้อที่ง่ายที่สุดในกลุ่มนี้)
ข้อ 2.	ข้อ 2.
ข้อ 3.	ข้อ 3.
⋮	⋮
⋮	⋮
ข้อ 37. (ข้อที่ง่ายที่สุดในกลุ่มนี้)	ข้อ 37. (ข้อที่ยากที่สุดในกลุ่มนี้)

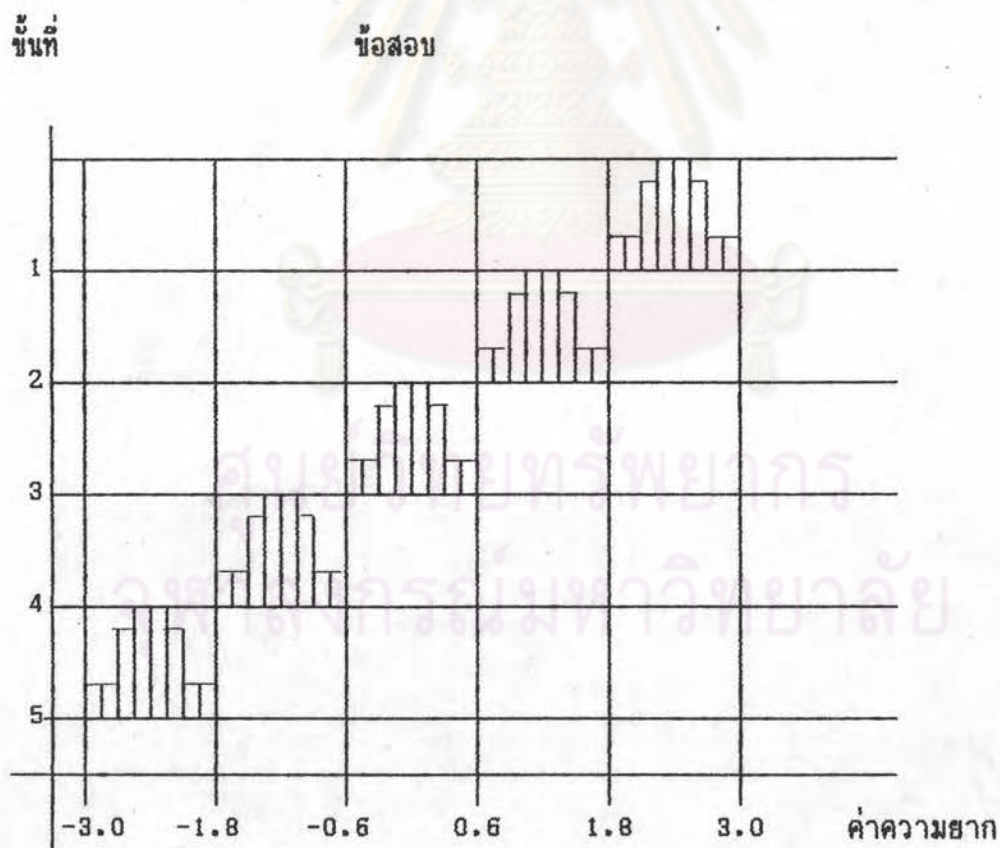
ภาพที่ 9 ตัวอย่างข้อสอบรูปแบบเฟล็กซ์ลีเวล (Flexilevel Test)

สำหรับกระดาษคำตอบที่ได้รับการออกแบบมาเป็นพิเศษ คือ เมื่อนักเรียนตอบถูกจุดสีน้ำเงินก็จะปรากฏขึ้น ข้อต่อไปผู้สอบจะต้องไปเลือกทำข้อสอบในกลุ่มสีน้ำเงิน (กลุ่มยาก) ในข้อที่มีหมายเลขต่ำสุดที่ยังไม่ได้ทำ และถ้าตอบผิดจุดสีแดงก็จะปรากฏขึ้น ผู้สอบจะต้อง

ไปเลือกทำข้อสอบในกลุ่มสีแดง (กลุ่มง่าย) ในข้อที่มีหมายเลขต่ำสุดที่ยังไม่ได้ทำ ถ้าให้ k เป็นจำนวนข้อสอบเดิม ผู้สอบในข้อสอบเฟลิกซีเลเวลจะต้องทำข้อสอบทั้งหมด $(k+1)/2$ ข้อ (Lord, 1971) ข้อสอบเฟลิกซีเลเวล ก็คือ ข้อสอบเฉพาะบุคคลรูปแบบปิรามิดที่มีข้อสอบเพียงข้อเดียวในแต่ละระดับความยากนั่นเอง (Hambleton and Swaminathan, 1985; citing Betz and Weiss, 1975) แต่มีข้อแตกต่างกันที่การใช้เส้นทางในการเลือกตอบข้อสอบ

2.1.3 รูปแบบปรับระดับขั้น

การทดสอบเฉพาะบุคคลรูปแบบปรับระดับขั้น (Stradaptive Models) เป็นรูปแบบของการทดสอบเฉพาะบุคคลที่ปรับปรุงจากรูปแบบเฟลิกซีเลเวล โดยการแบ่งความยากของข้อสอบออกเป็นช่วงๆ (Stratified) ในแต่ละช่วงความยาก (Stratum) จะมีข้อสอบหลายข้อ (Hambleton and Swaminathan, 1985 ; citing Weiss, 1973 ; Waters, 1977) ดังตัวอย่างภาพประกอบ 10



ภาพที่ 10 โครงสร้างของการทดสอบเฉพาะบุคคลรูปแบบปรับระดับขั้น

2.2 การทดสอบเฉพาะบุคคลโดยใช้กลยุทธ์หลายขั้นตอนรูปแบบแยกทางแปรผัน
 การทดสอบเฉพาะบุคคลโดยใช้กลยุทธ์หลายขั้นตอน (Multi-Stage Strategies) รูปแบบแยกทางแปรผัน (Variable Branching Models) หมายถึง รูปแบบการตอบข้อสอบหลายขั้นตอนที่ไม่ได้กำหนดข้อสอบ และเส้นทางในการตอบข้อสอบไว้ล่วงหน้าว่าถ้าผู้สอบตอบถูกจะต้องไปทำข้อใดหรือถ้าตอบผิดจะต้องไปทำข้อสอบข้อใด ในรูปแบบแยกทางแปรผันนี้จะไม่มีขนาดของขั้น แต่จะดำเนินการสอบจากกลุ่มข้อสอบที่คำนวณค่าความยากและค่าอำนาจจำแนกของข้อสอบไว้แล้ว กฎการคัดเลือกข้อสอบข้อถัดไปเพื่อให้เหมาะสมกับความสามารถของผู้สอบนั้นจะใช้รูปแบบทางคณิตศาสตร์เข้ามาช่วย ได้แก่ รูปแบบเบย์และรูปแบบความเป็นไปได้สูงสุด

2.2.1 กลยุทธ์แบบเบย์ (Bayesian Strategies) กลยุทธ์นี้ได้ประยุกต์ทฤษฎีของเบย์ในกระบวนการตอบข้อสอบ โดยใช้วิธีการทดสอบแบบใช้เครื่องคอมพิวเตอร์ในการทดสอบ (Computerized Adaptive Testing) มีขั้นตอนดังนี้

ก. ประมาณค่าความสามารถของผู้สอบ และความคลาดเคลื่อนมาตรฐานการประมาณค่าไว้ก่อน ในแต่ละขั้นของการทดสอบจากสารสนเทศที่มีอยู่เกี่ยวกับผู้สอบ

ข. คัดเลือกข้อสอบจากกลุ่มข้อสอบที่ได้คำนวณค่ารามิเตอร์ของข้อสอบไว้แล้ว ข้อสอบในกลุ่มข้อสอบทุกข้อที่ยังไม่ได้นำมาให้ผู้สอบคนนั้นๆ สอบ ก็จะถูกถือว่าเป็นข้อสอบที่มีโอกาสนำมาใช้สอบได้ กระบวนการนี้แสดงให้เห็นว่าข้อสอบข้อใดๆ ในกลุ่มข้อสอบที่นำมาใช้ในการทดสอบกับผู้สอบคนใดๆ ก็ตาม จะเป็นข้อสอบที่ลดความไม่แน่นอนในการประมาณความสามารถของผู้สอบมากที่สุด ข้อสอบที่นำมาใช้กับผู้สอบจะเป็นข้อสอบที่มีระดับความยากใกล้เคียงกับระดับความสามารถของผู้สอบ หลังจากดำเนินการสอบโดยใช้ข้อสอบที่คัดเลือกไว้ก็จะประมาณความสามารถก่อน และสารสนเทศที่ได้จากการตอบข้อสอบนั้น เช่น การตอบถูก หรือการตอบผิด นำมารวมกันโดยวิธีทฤษฎีเบย์ก็จะได้การประมาณค่าความสามารถภายหลังการประมาณค่าครั้งหลังนี้ เป็นการประมาณค่าที่ปรับจากสิ่งที่รู้เกี่ยวกับข้อมูลที่อธิบายประชากรผู้สอบก็สามารถนำมาใช้เป็นสารสนเทศก่อนได้ กระบวนการนี้จะสิ้นสุดลงเมื่อความแปรปรวนภายหลังมีค่าน้อยกว่าที่กำหนดไว้ล่วงหน้า

2.2.2 กลยุทธ์ความเป็นไปได้สูงสุด (Maximum Likelihood Estimation) ยูรี (Urry, 1970 ; cited in Weiss, 1975) ได้พัฒนาการทดสอบแบบซี เอ ที (Computerized Adaptive Testing) โดยใช้วิธีการความเป็นไปได้สูงสุดตามแนวทฤษฎีการตอบสนองข้อสอบ (IRT) กลยุทธ์ความเป็นไปได้สูงสุดดำเนินการใน

แนวทางที่คล้ายคลึงกับกระบวนการเบย์ แม้ว่าเหตุผลทางคณิตศาสตร์จะต่างกัน หลังจากที่มี
 สอบตอบข้อสอบข้อหนึ่งๆ ถูกต้อง และอีกข้อหนึ่งผิดก็จะสามารถนำไปแก้สมการความเป็นไปได้
 สูงสุดและจะได้ค่าการประมาณความสามารถและความคลาดเคลื่อนมาตรฐาน ข้อสอบข้อ
 ถัดไปที่เลือกมาใช้สอบเป็นข้อสอบจากกลุ่มข้อสอบ เป็นข้อสอบที่มีค่าความยากเหมาะสมกับความ
 สามารถของผู้สอบมากที่สุด เมื่อผู้สอบตอบข้อสอบข้อนั้นแล้วก็จะมีการประเมินคำตอบทันที
 จากบันทึกการตอบทั้งหมดของผู้สอบซึ่งรวมถึงข้อสอบขั้นสุดท้ายที่ผู้สอบตอบ จากนั้นก็ประมาณค่า
 ความสามารถใหม่และความคลาดเคลื่อนมาตรฐานใหม่โดยใช้สมการความเป็นไปได้สูงสุด ถ้า
 ผู้สอบตอบข้อสุดท้ายผิดการประมาณความสามารถก็จะลดต่ำลง ในขั้นสุดท้ายของกระบวนการ
 ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถลงเป็นผลจากสารสนเทศใหม่ที่
 ได้จากข้อสุดท้ายร่วมกับรูปแบบการตอบข้อสอบที่ทำผ่านมาของผู้สอบ

1.5 วิธีการให้คะแนนการทดสอบเฉพาะบุคคล (Scoring)

ได้มีผู้เสนอวิธีการให้คะแนนการทดสอบเฉพาะบุคคล ไว้หลายวิธี สรุป
 ได้ดังนี้

(1) ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก (The Most
 difficulty of Item answered correctly) เป็นการให้คะแนน โดยดูจาก
 ข้อสอบการทดสอบเฉพาะบุคคลที่ผู้สอบทำมาทั้งหมด ข้อใดที่มีค่าความยากสูงสุดที่ผู้สอบทำถูก
 ถือว่าเป็นค่าความสามารถของผู้สอบคนนั้น ค่าความยากที่ว่านี้ ควรจะเป็นค่าความยากที่
 วิเคราะห์มาจากทฤษฎีการตอบสนองข้อสอบ

(2) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก (Average
 difficulty of all Item answered correctly) เป็นการให้คะแนนโดยการ
 เอาค่าความยากของข้อสอบเฉพาะบุคคลทุกข้อที่ผู้สอบทำถูกมาหาค่าเฉลี่ย ค่าเฉลี่ยความยากนี้
 ถือว่าเป็นค่าความสามารถของผู้สอบคนนั้น

(3) ให้คะแนนตามค่าเฉลี่ยความยากของทุกข้อที่ทำ (Average
 difficulty) วิธีการให้คะแนนแบบนี้เสนอโดย ลอร์ด (Weiss, 1974 ; citing
 Lord, 1970) วิธีการ คือ นำค่าความยากของข้อสอบเฉพาะบุคคลทุกข้อที่ผู้สอบได้ทำ
 โดยไม่คำนึงว่าจะตอบถูกหรือไม่ มาหาค่าเฉลี่ย ค่าเฉลี่ยนี้จะถือเป็นความสามารถของผู้สอบ
 ผู้นั้น

(4) ให้คะแนนตามค่าความยากของข้อสุดท้ายที่ทำ (Difficulty of
 the final term) นั่นคือ ไม่คำนึงถึงว่าขั้นสุดท้ายจะตอบถูกหรือไม่ก็ตาม จะถือว่าค่า
 ความยากของการทดสอบเฉพาะบุคคลขั้นสุดท้ายที่ผู้สอบได้ทำคือค่าความสามารถของผู้สอบผู้นั้น

(5) ให้คะแนนตามค่าความยากของข้อสอบที่ต่อจากขั้นสุดท้ายที่ทำ (Difficulty of (N+1)th Item) ซึ่งมีลักษณะคล้ายกับวิธีที่ 4 แต่เป็นการตรวจสอบต่อไปอีกชั้นหนึ่ง กล่าวคือ ถ้าขั้นสุดท้ายที่ทำถูก คะแนนในขั้นต่อไปก็จะเพิ่มขึ้น และถ้าขั้นสุดท้ายที่ทำผิดขั้นต่อไปคะแนนก็จะลดลง

(6) ให้คะแนนโดยใช้หลักการของเบย์ (Bayesian Strategies) การใช้หลักการของเบย์นี้มีวิธีการย่อยหลายวิธี วิธีที่นิยมใช้กันมาก คือ วิธีที่นักสถิติโรเจอร์ เจ โอเวน (Owen, 1969 ; 1975 ; Thorndike, 1982 ; citing Owen, 1975) เป็นผู้เสนอขึ้น เพื่อใช้ในการประมาณค่าความสามารถของผู้สอบในการทดสอบเฉพาะบุคคลโดยเฉพาะ ที่ให้ชื่อว่า เบย์เซียน อัปเดตติง (Bayesian Updating) ซึ่งสามารถประมาณค่าความสามารถของผู้สอบได้ค่อนข้างคงที่ โดยมีสูตร ดังนี้

ในกรณีที่ตอบข้อสอบถูก

$$e_{m+1} = e_m + (1-c_t) \left[\frac{\sigma_m^2}{\sqrt{\frac{1}{B_t^2} + \sigma_m^2}} \right] \left[\frac{\phi(D)}{c_t + (1-c_t)\Phi(-D)} \right]$$

และ

$$\sigma_{m+1}^2 = \sigma_m^2 \left[1 - \frac{1-c_t}{1 + \frac{1}{B_t^2 \sigma_m^2}} \left| \frac{\phi(D)}{A} \right| \left| \frac{(1-c_t)\phi(D) - D}{A} \right| \right]$$

ในกรณีที่ตอบข้อสอบผิด

$$e_{m+1} = e_m - \left[\frac{\sigma_m^2}{\sqrt{\frac{1}{B_t^2} + \sigma_m^2}} \right] \left[\frac{\phi(D)}{\Phi(D)} \right]$$

และ

$$\sigma_m^2 + 1 = \sigma_m^2 \left[1 - \frac{\left| \frac{\phi(D)}{1 + \frac{1}{a_1^2 \sigma_m^2}} \right| \left| \frac{\phi(D) + D}{\phi(D)} \right|}{\left| \phi(D) \right|} \right]$$

เมื่อ $\phi(D)$ คือ ค่าออร์ดิเนต ของโค้งปกติ ณ จุด D
 $\psi(D)$ คือ พื้นที่ใต้โค้งปกติจากค่า D ต่ำสุดจนถึงจุด D

$$D = \frac{b_1 - e_m}{\sqrt{\frac{1}{a_1^2} + \sigma_m^2}}$$

$$A = c_1 + (1 - c_1)\psi(-D)$$

e_m คือ ความสามารถของผู้สอบที่ประมาณค่าได้ก่อนตอบข้อสอบข้อที่ $m+1$ โดยปกติถ้าเป็นการประมาณค่าครั้งแรก ที่ยังไม่ทราบค่าความสามารถของผู้สอบ จะให้ $e_m = 0.0000$

σ_m^2 คือ ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบ ก่อนตอบข้อสอบข้อที่ $m+1$ โดยปกติถ้าเป็นการประมาณค่าครั้งแรกที่ยังไม่ทราบค่าความแปรปรวนจะให้ $\sigma_m^2 = 1.0000$

e_{m+1} คือ ความสามารถของผู้สอบที่ประมาณค่าได้เมื่อตอบข้อสอบข้อที่ $m+1$ แล้ว

σ_{m+1}^2 คือ ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบ เมื่อตอบข้อสอบข้อที่ $m+1$ แล้ว

a_1 คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ $m+1$

b_1 คือ ค่าความยากของข้อสอบข้อที่ $m+1$

c_1 คือ ค่าการเดาของข้อสอบข้อที่ $m+1$

(7) ให้คะแนนโดยใช้หลักการความเป็นไปได้สูงสุด (Maximum Likelihood) การให้คะแนนโดยใช้หลักการนี้ยังมีวิธีย่อยหลายวิธี วิธีที่นิยมใช้มากคือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) การประมาณค่าโดยวิธีนี้มีข้อจำกัดในกรณีที่ผู้สอบตอบข้อสอบถูกหมดหรือผิดหมด จะไม่สามารถประมาณค่าได้ นอกจากนี้ แฮมมิงตัน (Hambleton and Swaminathan, 1985 ; citing Samejima, 1973a) ยังพบว่าในกรณีที่ใช้ 3-พารามิเตอร์บางครั้งอาจจะมีค่าสูงสุดหลายค่า ในกรณีนี้ลอร์ดได้ให้ข้อสังเกตว่าปัญหาจะไม่เกิดขึ้นถ้าใช้ข้อสอบมากข้อ ($k > 20$) ซึ่งมีขั้นตอนในการประมาณค่า ดังนี้ (Hambleton and Swaminathan, 1985)

ขั้นที่ 1 ประมาณค่าความสามารถเริ่มต้น ($\theta_m ; m=0$) โดยใช้สูตร ดังนี้

$$\theta_0 = \ln[r_u / (k - r_u)]$$

เมื่อ $r_u = \sum a_i U_i ; (U_i = 1 \text{ เมื่อตอบข้อสอบถูก,}$
 $U_i = 0 \text{ เมื่อตอบข้อสอบผิด,}$
 a_i คือ อำนาจจำแนกของข้อสอบข้อที่ i
 k คือ จำนวนข้อสอบทั้งหมดที่ผู้สอบได้ทำในการทดสอบเฉพาะบุคคล

ขั้นที่ 2 หาค่า $P_i(\theta_m)$ และ $Q_i(\theta_m)$ โดยใช้สูตร ดังนี้

$$P_i(\theta_m) = c_i + (1-c_i) \frac{Da_i(\theta_m - b_i)}{e}$$

$$Q_i(\theta_m) = \frac{Da_i(\theta_m - b_i)}{1+e}$$

$$Q_i(\theta_m) = 1 - P_i(\theta_m)$$

$$= \frac{1 - c_i}{1+e}$$

$$\frac{Da_i(\theta_m - b_i)}{1+e}$$

$$1+e$$

เมื่อ	e_m	คือ	ความสามารถของผู้สอบที่ประมาณค่าได้ในครั้งที่ m
	a_i	คือ	ค่าอำนาจจำแนกของข้อสอบข้อที่ i
	b_i	คือ	ค่าความยากของข้อสอบข้อที่ i
	c_i	คือ	ค่าการเตาของข้อสอบข้อที่ i
	D	คือ	ค่าคงที่ ซึ่งมีค่าเท่ากับ 1.7
	e	คือ	ค่าคงที่ ซึ่งมีค่าเท่ากับ 2.7182

ขั้นที่ 3 หาค่าปรับแก้ (h_m) โดยใช้สูตร

$$h_m = \frac{D[r_m - \sum P_i(e_m)]}{-D^2 \sum P_i(e_m) Q(e_m)}$$

ขั้นที่ 4 ประมาณค่าความสามารถของผู้สอบใหม่ (e_{m+1}) โดยใช้สูตร

$$e_{m+1} = e_m - h_m$$

ขั้นที่ 5 ทำการคำนวณซ้ำในขั้นที่ 2, 3 และ 4 จนกระทั่ง h_m เข้าใกล้ศูนย์ ($h_m < 0.001$)

2. ทฤษฎีการตอบสนองข้อสอบ (Item Response Theory)

ทฤษฎีการทดสอบแบบประเพณีนิยม (Classical Test Theory) เป็นทฤษฎีที่ไม่เหมาะสมกับการทดสอบเฉพาะบุคคล เนื่องจากดัชนีความเที่ยง ความตรงและคุณภาพของข้อสอบแบบประเพณีนิยมนั้นจะสอดคล้องกันเมื่อผู้สอบทุกคนได้ทำข้อสอบชุดเดียวกัน กล่าวคือ ค่าความยาก และค่าอำนาจจำแนกของข้อสอบไม่คงที่ แต่จะแปรเปลี่ยนไปตามกลุ่มตัวอย่าง (Hambleton and Swaminathan, 1985) และคะแนนของผู้สอบก็แปรเปลี่ยนไปตามค่าความยากของข้อสอบ จึงจำเป็นต้องใช้ทฤษฎีที่เหมาะสมกับการทดสอบเฉพาะบุคคล (Tailored Testing) ที่มีการจัดข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบ โดยที่ผู้สอบที่ต่างความสามารถกันจะได้รับชุดของข้อสอบที่แตกต่างกัน และมีค่าความยากหรือง่ายไม่เท่ากัน ทฤษฎีที่เหมาะสมและสามารถนำมาใช้กับการทดสอบเฉพาะบุคคลได้อย่างมีประสิทธิภาพ ก็คือ ทฤษฎีการตอบสนองข้อสอบ (IRT) (Urry, 1977; Green and



others, 1984) ซึ่งจะกล่าวถึงทฤษฎีนี้เฉพาะในส่วนที่เกี่ยวกับการทดสอบเฉพาะบุคคล ดังนี้

2.1 คุณสมบัติของรูปแบบการตอบข้อสอบ

แอมเบิลตันและสวามินาธาน (Hambleton and Swaminathan, 1985) ได้สรุปไว้ดังนี้

1. เป็นรูปแบบที่อธิบายเกี่ยวกับปฏิบัติการของผู้สอบในการทำแบบสอบ ว่าสามารถอธิบายได้ในลักษณะของคุณลักษณะหนึ่งหรือหลายคุณลักษณะ ซึ่งเรียกว่าคุณลักษณะแฝง
2. รูปแบบการตอบข้อสอบเป็นความสัมพันธ์ระหว่างปฏิบัติการของผู้สอบ ในการตอบข้อสอบทั้งที่สังเกตได้และคุณลักษณะแฝงหรือความสามารถที่ถือว่าเป็นปฏิบัติการหลักของพฤติกรรมกรรมการทำแบบสอบ
3. รูปแบบการตอบข้อสอบที่จะต้องให้วิถีทางในการประมาณคะแนนของผู้สอบ ได้ตรงกับคุณลักษณะแฝงนั้น
4. คุณลักษณะแฝงอธิบายได้จากปฏิบัติการที่สังเกตได้ของผู้สอบในการตอบข้อสอบ

ความสัมพันธ์ระหว่างปริมาณของการทำข้อสอบหรือคะแนนของผู้สอบ (Test Performance) กับปริมาณความสามารถของผู้สอบ (Ability) สามารถเขียนในรูปของความสัมพันธ์ต่างๆ ไปดังนี้ (Lord, 1982)

$$P = f(U_1 | \theta_1, \theta_2, \theta_3, \dots, \theta_k ; \beta_j)$$

เมื่อ	P	แทน	ผลการสอบ	
	f	แทน	ฟังก์ชัน	
	U ₁	แทน	ผลการตอบข้อสอบ	ตอบถูก U ₁ = 1 ตอบผิด U ₁ = 0
	θ ₁ , θ ₂ , θ ₃ , θ _k	แทน	ความสามารถ (Ability หรือ Trait) ที่ 1, 2, 3, k	
	β _j	แทน	ค่าพารามิเตอร์ของข้อสอบข้อที่ j	

จากความสัมพันธ์ดังกล่าวนี้เป็นฟังก์ชันความสัมพันธ์ในลักษณะต่างๆ ไป นักวัดผลการศึกษาจะต้องหา โมเดลทางคณิตศาสตร์ (Mathematical Model) ที่เหมาะสมเพื่อใช้แทนฟังก์ชันความสัมพันธ์ดังกล่าว โดยอาศัยข้อตกลงเบื้องต้นต่างๆ ดังนี้

2.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ

1. ความเป็นมิติเดียว (Unidimension) หมายถึง ข้อสอบแต่ละข้อ ในแบบสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียวกัน (Unidimensionality) หรือ ข้อสอบเหล่านี้มีลักษณะเป็นเอกพันธ์ (Homogeneous) การกำหนดเช่นนี้ก็เพื่อให้รูปแบบของ ทฤษฎีนี้มีความซับซ้อนน้อยลงและง่ายแก่การแปลความหมายของคะแนนที่ได้จากแบบสอบ

วิธีการตรวจสอบว่าแบบสอบนั้นวัดในมิติเดียวหรือไม่นั้น ทำได้หลายวิธี เช่น โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกตค่าไอเก้น (Eigen value) ค่าสูงสุด ว่าแตกต่างจากค่าอื่นๆ อย่างชัดเจนหรือไม่ (Hambleton and Thum, 1973; Reckase, 1979) สำหรับการทดสอบเฉพาะบุคคล กรีนและ คนอื่นๆ (Green and others, 1984) ได้เสนอแนะว่าการที่จะทำให้เกิดความเป็น มิติเดียว (Unidimension) ได้ง่ายวิธีหนึ่งคือ การเลือกข้อสอบที่มีค่าอำนาจจำแนกสูง เนื่องจากข้อสอบที่มีค่าอำนาจจำแนกสูงจะมีพหุสัมพันธ์กับคะแนนแบบสอบทั้งหมดสูง ยูรี (Urry, 1981) เสนอไว้ว่าควรเลือกข้อสอบที่มีค่าอำนาจจำแนกตั้งแต่ .8 ขึ้นไป

2. ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือ การตอบ ข้อสอบข้อใดข้อหนึ่งได้ถูกหรือผิดจะไม่มีผลต่อการตอบข้ออื่นๆ ด้วย (Lord, 1980; Hambleton and Swaminathan, 1985) ซึ่งแมกซ์โดนาลด์ (Mc Donald, 1982) และแฮมเบิลตันกับสวามินาธาน (Hambleton and Swaminathan, 1985) มีความเห็น ตรงกันว่า ถ้ามีความเป็นมิติเดียวกันแล้ว ความเป็นอิสระในการตอบข้อสอบก็จะมีเอง

3. โค้งลักษณะข้อสอบ (Item Characteristic Curve) เป็น ฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบนั้นได้ถูกต้องกับ ระดับความสามารถที่วัดได้โดยชุดของข้อสอบหรือแบบสอบนั้น (Hambleton and Swaminathan, 1985) ซึ่งจะเห็นว่าโอกาสที่ผู้สอบตอบข้อสอบถูก (Probability) จะขึ้นอยู่กับโค้งลักษณะข้อสอบ (ICC) ในแต่ละโมเดลที่ใช้ โดยที่รูปร่าง (Shape) ของ โค้งลักษณะข้อสอบ (ICC) ในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยน (Invariant) ไปตาม กลุ่มตัวอย่างหรือตัวผู้สอบ (Hambleton and Cook, 1977) ฉะนั้นจึงทำให้โอกาส ในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยนด้วย

เนื่องจากมีความเชื่อหลายความเชื่อ เกี่ยวกับความสัมพันธ์ระหว่าง โอกาสในการตอบข้อสอบได้ถูกต้องกับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบ หรือ แบบสอบนั้น จึงทำให้เกิดมีรูปแบบของโค้งลักษณะข้อสอบ (Item Response Model)

ขึ้นมาหลายรูปแบบ เช่น Guttman Perfect Scale, Latent Distance Model, Linear model, Normal orgive Model, Logistic Model, Nominal Response Model และ Grade Response Model (จักรกฤษณ์ สํารานใจ 2531) แต่รูปแบบที่นิยมใช้กันแพร่หลายคือ รูปแบบโลจิสติก (Logistic Model) รูปแบบนี้แบ่งออกเป็น 4 รูปแบบตามจำนวนพารามิเตอร์ของโค้งลักษณะข้อสอบดังนี้ (Hambleton and Swaminathan, 1985)

1. รูปแบบโลจิสติกที่มี 1 พารามิเตอร์ (One Parameter Logistic Model) เป็นรูปแบบที่มีข้อตกลงเบื้องต้นว่า ค่าอำนาจจำแนกของข้อสอบทุกข้อ เท่ากันและค่าการเดาของข้อสอบเป็น 0 ดังสมการ

$$P_i(\theta) = \frac{D a_i (\theta - b_i)}{1 + e^{D a_i (\theta - b_i)}}, \quad (i = 1, 2, 3, \dots, n)$$

2. รูปแบบโลจิสติกที่มี 2 พารามิเตอร์ (Two Parameter Logistic Model) เป็นรูปแบบที่เบิร์นบอม (Birnbbaum, 1957, 1958a, 1958b, 1968) เสนอไว้แทนรูปแบบนอร์มอลโอโจว เนื่องจาก รูปแบบโลจิสติกเป็นรูปแบบเชิงคณิตศาสตร์ ที่นำมาใช้ในทางปฏิบัติได้สะดวกกว่า รูปแบบนี้ถือว่าไม่มีการเดา ดังสมการ

$$P_i(\theta) = \frac{D a_i (\theta - b_i)}{1 + e^{D a_i (\theta - b_i)}}, \quad (i = 1, 2, 3, \dots, n)$$

3. รูปแบบโลจิสติกที่มี 3 พารามิเตอร์ (Three Parameter Logistic Model) โดยเพิ่มพารามิเตอร์ ค่าการเดา (c_i) เข้าไปในรูปแบบโลจิสติก ที่มี 2 พารามิเตอร์ มีสมการดังนี้

$$P_i(e) = c_i + (1-c_i) \frac{e^{Da_i(e-b_i)}}{1+e^{Da_i(e-b_i)}}, \quad (i = 1, 2, 3, \dots, n)$$

- เมื่อ $P_i(e)$ คือ โอกาสที่ผู้สอบซึ่งมีระดับความสามารถ (e) ตอบข้อสอบข้อที่ i ได้ถูกต้อง
- a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i โดยทฤษฎีแล้วจะมีค่าอยู่ระหว่าง $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัติจะใช้ค่าที่อยู่ระหว่าง 0 ถึง 2
- b_i คือ ค่าความยากของข้อสอบข้อที่ i โดยทฤษฎีแล้ว b จะมีค่าอยู่ระหว่าง $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัติ b จะมีค่าอยู่ระหว่าง -3.00 ถึง $+3.00$ ข้อที่มีค่า b ใกล้ -3.00 เป็นข้อที่ง่ายมาก และข้อที่มีค่า b ใกล้ $+3.00$ เป็นข้อที่ยากมาก (Hambleton and Cook, 1977)
- c_i คือ ค่าการเดาของข้อสอบข้อ i ซึ่งได้จากเส้นกำกับโค้งลักษณะข้อสอบและแทนความน่าจะเป็นของผู้สอบที่มีความสามารถต่ำตอบข้อสอบได้ถูกต้อง ในทางปฏิบัติค่า c_i มีค่าระหว่าง 0 ถึง 0.3 เมื่อข้อสอบมี 4 ตัวเลือก

4. รูปแบบโลจิสติกที่มี 4 พารามิเตอร์ (Four Parameter Logistic Model) รูปแบบนี้คำนึงถึงผู้สอบที่มีความสามารถสูง เนื่องจากผู้สอบที่มีความสามารถสูงบางครั้งอาจมีความสับสนในการทำข้อสอบ จึงทำให้ตอบข้อสอบข้อนั้นผิดพลาดไป แมคโคเนล (McDonald อ้างถึงใน Hambleton and Swaminathan, 1985) จึงเสนอรูปแบบนี้ขึ้นดังสมการ

$$P_i(e) = c_i + (Y_i - c_i) \frac{e^{Da_i(e-b_i)}}{1+e^{Da_i(e-b_i)}}, \quad (i = 1, 2, 3, \dots, n)$$

รูปแบบนี้ต่างจากรูปแบบ 3 พารามิเตอร์ คือ ค่า Y ให้มีค่าต่ำกว่า 1 เล็กน้อย แต่อย่างไรก็ตาม รูปแบบนี้ยังไม่สามารถนำมาใช้ในทางปฏิบัติได้

2.3 ความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบ

(Accuracy) โดยปกติไม่ว่าจะเป็นการวัดหรือการประมาณค่าใดๆ ก็ตาม จะต้องมีความคลาดเคลื่อนอยู่เสมอ โดยเฉพาะอย่างยิ่งในการวัดหรือการประมาณค่าความสามารถในการศึกษาและจิตวิทยา จะมีความคลาดเคลื่อนค่อนข้างสูง ถ้าความคลาดเคลื่อนในการประมาณค่ามีค่าสูง ความถูกต้องแม่นยำในการประมาณค่าก็จะต่ำ ซึ่งสามารถเขียนเป็นสมการแสดงความสัมพันธ์ระหว่างความสามารถจริง (θ) ความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}$) และค่าความคลาดเคลื่อน (E) ได้ดังนี้ (Lord and Novick, 1968)

$$SEE = \hat{\theta} - \theta$$

จากสมการนี้จะเห็นว่า ถ้าค่าความคลาดเคลื่อน (E) มีค่าน้อยหรือเข้าใกล้ศูนย์ ความสามารถที่ได้จากการประมาณค่า ($\hat{\theta}$) ก็จะเข้าใกล้ความสามารถจริง (θ) นั่นคือ มีความถูกต้องแม่นยำในการประมาณค่าความสามารถของผู้สอบสูง

2.4 ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (Standard Error of Estimate : SEE) ก็คือ ส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนในการประมาณค่า ซึ่งแตกต่างจากความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement : SEM) (Lord and Novick, 1968 ; Urry, 1977) ตรงที่ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ที่คลาดเคลื่อนไปจากความสามารถจริง ($E = \hat{\theta} - \theta$) ส่วนความคลาดเคลื่อนมาตรฐานในการวัด (SEM) จะใช้ความคลาดเคลื่อนที่เกิดจากคะแนนดิบ หรือคะแนนที่ได้จากการวัดที่คลาดเคลื่อนไปจากคะแนนจริง ($E = x - T$)

ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) มีความสัมพันธ์กับค่าสารสนเทศแบบสอบ (Test Information Function) ดังนี้ $SEE = 1/\sqrt{I(\theta)}$ (Hambleton and Cook, 1977 ; Hambleton and Swaminathan, 1985) ค่าสารสนเทศแบบสอบจะเป็นตัวบ่งชี้ถึงคุณภาพของการประมาณค่า ซึ่งยังสามารถใช้แทนค่าความเที่ยง (Reliability : $r_{xx'}$) หรือค่าความคลาดเคลื่อนมาตรฐานในการวัด (SEM) ในทฤษฎีการทดสอบแบบเดิม (Classical Theory) ได้ด้วย (Hambleton and Cook, 1977) กรีนและคนอื่นๆ (Green and others, 1984) ได้ชี้ให้เห็นว่าความคลาดเคลื่อน

มาตรฐานในการวัด (SEM) มีข้อตกลงเบื้องต้นที่ไม่น่าจะถูกต้อง ที่ว่าความแปรปรวนของคะแนนความคลาดเคลื่อนมีค่าเท่ากันสำหรับทุกๆ คะแนน ฉะนั้น การทดสอบเฉพาะบุคคล (Tailored Testing) จึงควรใช้ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) แทน โดยเฉพาะอย่างยิ่งการทดสอบเฉพาะบุคคลที่ใช้การประมาณค่าความสามารถของผู้สอบ โดยวิธีของเบย์ส์ (Bayesian Updating) สามารถที่จะกำหนดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) ต่ำสุดสำหรับในการทดสอบเฉพาะบุคคลแต่ละครั้งได้ด้วย ส่วนค่าความเที่ยงนั้น วอม (Warm, 1978) ได้แสดงความสัมพันธ์ระหว่างความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) กับค่าความเที่ยง (r_{xx}) ไว้ดังนี้

$$r_{xx} = 1 - (\overline{SEE})^2$$

เมื่อ \overline{SEE} คือ ค่าเฉลี่ยของ SEE

ได้มีนักการศึกษาหลายท่าน (Samejima, 1977 ; Green and others, 1984) มีความเห็นที่ไม่ควรนำเอาการหาค่าความเที่ยง (Reliability) มาใช้กับการทดสอบเฉพาะบุคคล แต่ไดว์จี (Divgi, 1989) มีความเห็นว่าค่าความเที่ยงยังมีประโยชน์สำหรับการทดสอบเฉพาะบุคคลอยู่

ความคลาดเคลื่อนมาตรฐานในการประมาณค่า (SEE) ในการวิจัยครั้งนี้ หาได้จากสูตร ดังนี้

$$SEE = \sqrt{\frac{N\overline{E}^2 - (\overline{\Sigma E})^2}{N(N-1)}}$$

เมื่อ SEE คือ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

E คือ ค่าความคลาดเคลื่อนในการประมาณค่า ($E = o - \bar{o}$)

N คือ จำนวนผู้สอบ

2.5 การนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้

จากคุณสมบัติที่สำคัญของทฤษฎีการตอบสนองข้อสอบ (IRT) ในเรื่องค่าพารามิเตอร์ของข้อสอบไม่ขึ้นอยู่กับความสามารถของผู้สอบ และความสามารถของผู้สอบก็ไม่ขึ้นอยู่กับความยาก หรือง่ายของข้อสอบ และค่าพารามิเตอร์ไม่เปลี่ยนแปลงไปตามตำแหน่งของข้อสอบ คุณลักษณะดังกล่าวนี้ คือ ความเป็นปรนัยของการวัดผล (Objectivity of Measurement) ที่มีความสำคัญมากในทางวัดผล (Wright and Stones อ้างถึงใน สุนันท์ สุกมลสันต์, 2528) จึงสามารถนำเอาคุณสมบัติเด่นเหล่านี้มาประยุกต์ใช้ในการวัดผลการศึกษาได้อย่างมีประสิทธิภาพในหลายๆ เรื่อง ดังนี้

1. การทดสอบเฉพาะบุคคล (Tailored Testing) หมายถึง การทดสอบที่มีการจัดข้อสอบให้มีความยากที่เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล นับเป็นการนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้อย่างมีประสิทธิภาพ (Urry, 1977) โดยการนำเอาคุณสมบัติความไม่แปรเปลี่ยนของค่าพารามิเตอร์ของข้อสอบ และความสามารถของผู้สอบมาใช้ การทดสอบเฉพาะบุคคลนั้นผู้สอบแต่ละคนจะได้รับชุดของข้อสอบที่ต่างกันไปตามระดับความสามารถของผู้สอบ แต่ก็ยังสามารถนำเอาความสามารถที่ประมาณค่าได้จากการทดสอบนี้มาเปรียบเทียบกันได้ ทั้งนี้เพราะความสามารถที่ประมาณค่าได้ ยังอยู่ในมาตราวัดเดียวกัน (Common Ability Scale) (Hambleton and Cook, 1977 ; Hambleton and Swaminathan, 1985)

2. การสร้างคลังข้อสอบ (Item Bank) โดยการนำเอาทฤษฎีการตอบสนองข้อสอบมาใช้ในการวิเคราะห์และสร้างคลังข้อสอบ ทำให้ข้อสอบแต่ละข้อมีค่าสถิติต่างๆ เป็นของประจำข้อไม่ขึ้นอยู่กับข้ออื่น และกลุ่มตัวอย่าง จึงทำให้เกิดความสะดวกในการคัดเลือกข้อสอบมาใช้ และจัดชุดข้อสอบให้ได้คุณภาพตามเกณฑ์ที่ต้องการ (Hambleton and Swaminathan, 1985)

3. การหาความลำเอียงของข้อสอบ (Item Bias) สามารถใช้โค้งลักษณะข้อสอบ (ICC) มาตรวจสอบความลำเอียงของข้อคำถาม โดยปกติถ้าข้อคำถามนั้นไม่มีความลำเอียง โค้งลักษณะข้อสอบ ไม่ว่าจะได้มาจากผู้สอบกลุ่มใดก็ตาม จะมีรูปร่าง (Shape) เหมือนกัน แต่ถ้าข้อคำถามนั้นมีความลำเอียง โค้งลักษณะข้อสอบจะมีรูปร่างแตกต่างกัน (Hambleton and Swaminathan, 1985)

4. การเทียบมาตรา (Test Equating) หมายถึง กระบวนการทางสถิติเพื่อปรับคะแนนที่ได้จากการทดสอบด้วยข้อสอบที่ต่างชุดกัน ที่มีโครงสร้างเดียวกัน ให้สามารถเทียบกันได้ แต่ถ้าการทดสอบนั้นวัดความสามารถ (Ability or Trait)

ที่ต่างกันแล้วจะนำมาเทียบกันไม่ได้ (Hambleton and Swaminathan, 1985, citing Lord, 1977b; 1980b) เนื่องจากความสามารถของผู้สอบที่ประมาณค่ามาจาก ทฤษฎีการตอบสนองข้อสอบจะมีหน่วยการวัดเดียวกัน และไม่ขึ้นอยู่กับชุดของแบบสอบ จึงสามารถนำมาเทียบกันได้

3. งานวิจัยที่เกี่ยวข้องกับการทดสอบเฉพาะบุคคล

3.1 งานวิจัยในต่างประเทศ

วอเตอร์ (Larkin and Weiss, 1975; citing Water, 1964) ได้ศึกษาเปรียบเทียบแบบสอบรูปแบบปิรามิดชนิด 5 ชั้น กับแบบสอบประเพณีนิยม 4 ชุด ชุดละ 5 ข้อ โดยใช้รูปแบบจำลองของลอร์ด (Lord Models) แบบสอบรูปแบบปิรามิด ใช้ความแตกต่างระหว่างค่าความยากของแบบสอบในชั้นที่อยู่ติดกันคงที่เท่ากับ .01 ใช้กฎ การแยกทางแบบเพิ่ม 1 / ลด 1 และให้คะแนน 2 วิธี ภายใต้การให้คะแนนทั้งสองวิธี จากคำตอบแบบอิสระหรือคำตอบแบบเลือกตอบ พบว่า ความสัมพันธ์ของคะแนนและความ สามารถของผู้สอบ แบบสอบรูปแบบปิรามิดจะสูงกว่าแบบสอบประเพณีนิยมทุกชุด แบบสอบ รูปแบบปิรามิดยังให้การกระจายของคะแนนรูปสี่เหลี่ยมผืนผ้ามาก และให้ศักยภาพการกระจาย ของคะแนนมากกว่าแบบสอบประเพณีนิยม

เบย์รอฟและซีลีย์ (Bayroff and Seeley, 1967 cited in Larkin, K.C. and Weiss, D.J., 1974) ดำเนินการสอบโดยใช้แบบสอบเฉพาะบุคคล รูปแบบปิรามิด 8 ชั้น 2 ฉบับ โดยให้ผู้สอบจำนวน 102 คน ทดสอบกับคอมพิวเตอร์ ผู้สอบ จะต้องทำข้อสอบที่ทดสอบแบบประเพณีนิยม ซึ่งวัดด้านภาษา 50 ข้อ ด้านตัวเลข 40 ข้อ พบว่า สหสัมพันธ์ระหว่างการทดสอบเฉพาะบุคคลและการทดสอบแบบประเพณีนิยมมีค่าเท่ากับ 0.87 และ 0.79 เมื่อเปรียบเทียบกับค่าสหสัมพันธ์โดยประมาณระหว่างการทดสอบแบบประเพณี นิยมโดยใช้ข้อสอบ 8 ข้อ กับการทดสอบแบบประเพณีนิยมโดยใช้ข้อสอบ 40 และ 50 ข้อพบว่า ค่าสหสัมพันธ์ที่ได้มีค่าเท่ากับ .75 และ .67 ดังนั้น จึงพบว่า การทดสอบโดยใช้รูปแบบ ปิรามิดมีค่าสหสัมพันธ์สูงกว่าการทดสอบแบบประเพณีนิยมเมื่อใช้ข้อสอบจำนวนเท่ากัน

ลินน์และคนอื่นๆ (Larkin and Weiss, 1974 ; citing Linn and others, 1969) ได้ศึกษาความแตกต่างของกลยุทธ์แยกทาง 7 วิธี โดยใช้คอมพิวเตอร์ ศึกษาแก่นักเรียน 4,885 คน แบบสอบประเพณีนิยมมี 190 ข้อ ในแต่ละกลยุทธ์ข้อสอบที่ นำมาใช้จะเลือกเฉพาะข้อสอบที่เหมาะสมกับผู้สอบแต่ละคน กลยุทธ์แยกทาง 5 วิธีแรกเป็น แบบสองขั้น (Two Stages) อีก 2 วิธีเป็นแบบปิรามิด ซึ่งปิรามิดแบบแรกเป็นแบบ 10 ชั้น

ที่มีความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันประมาณ .02 และปิรามิดรูป-
แบบที่ 2 เป็นแบบ 5 ชั้น โดยที่แต่ละชั้นมีข้อสอบ 5 ข้อ นั่นคือ ผู้สอบแต่ละคนจะต้องทำข้อสอบ
ในทางแยกกิตทางแยกหนึ่งจำนวน 25 ข้อ แบบสอบรูปแบบปิรามิดทั้งสอง จะใช้วิธีแยกทาง
แบบเพิ่ม/ลดเท่ากัน แบบสอบรูปแบบปิรามิดจะถูกนำมาเปรียบเทียบกับแบบสอบประเพณีนิยม
ขนาดสั้น 5 ชุด ซึ่งมีจำนวนข้อสอบ 10 ถึง 50 ข้อ ผลการศึกษาพบว่าแบบสอบรูปแบบปิรามิด
ชนิด 10 ชั้น มีความสัมพันธ์ .87 กับคะแนนรวมของแบบสอบ .95 กับแบบสอบรูปแบบปิรามิด
25 ข้อ และ .89 ถึง .96 กับแบบสอบประเพณีนิยมขนาดสั้นทั้ง 5 ชุด แบบสอบรูปแบบ
ปิรามิดชนิด 25 ข้อ มีความสัมพันธ์กับคะแนนรวมของแบบสอบ ซึ่งสอดคล้องกับความสัมพันธ์
กับแบบสอบประเพณีนิยมที่มีข้อสอบ 35 ข้อ นอกจากนี้ลินน์และคณะได้ทำการศึกษากับแบบสอบ
วัดผลสัมฤทธิ์ 2 ชุด โดยใช้กลุ่มตัวอย่างเดียวกันเพื่อใช้เป็นเกณฑ์การวัด พบว่า แบบสอบ
รูปแบบปิรามิดชนิด 10 ชั้น มีความสัมพันธ์กับเกณฑ์การวัดสูงกว่าความสัมพันธ์กับแบบสอบ
ประเพณีนิยมที่มีความยาวเท่ากัน ในทำนองเดียวกันนี้ แบบสอบรูปแบบปิรามิดชนิด 5 ชั้น ที่มี
ข้อสอบ 25 ข้อ จะมีความสัมพันธ์กับแบบสอบอิงเกณฑ์ทั้ง 2 ชุดสูงกว่าแบบสอบประเพณีนิยมที่มี
ข้อสอบ 50 ข้อ ผลการศึกษาแสดงให้เห็นว่า แบบสอบรูปแบบปิรามิดที่มีข้อสอบจำนวนน้อยข้อ
จะมีความตรงเมื่อเปรียบเทียบกับแบบสอบประเพณีนิยม

แอนเซน (Larkin and Weiss, 1975; citing Hansen, 1969)

ได้สร้างแบบสอบรูปแบบปิรามิด 5 ชนิดเกี่ยวกับตัวนิมฟ์ของเครื่องรับโทรเลข และนำไปทดสอบ
กับนักศึกษาวิทยาลัยชั้นปีที่ 1 จำนวน 56 วิทยาลัย จำนวนชั้นของแบบสอบแต่ละชุดเป็น 3 หรือ
4 ชั้นทั้งนี้ขึ้นกับคำตอบของนักศึกษาแต่ละคนที่ตอบ ซึ่งต้องตอบทั้งสิ้น 17 ข้อใช้ค่าความแตกต่าง
ระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันมีค่าคงที่เท่ากับ .10 การให้คะแนนมี 4 วิธี
จากการศึกษาพบว่า คะแนนที่ได้จากแบบสอบรูปแบบปิรามิดมีความสัมพันธ์กับคะแนนที่ได้จากแบบ
สอบในชั้นเรียนที่ใช้เวลา 1 ชั่วโมง โดยมีเนื้อหาเดียวกันและทดสอบก่อนนำแบบสอบรูปแบบ
ปิรามิดมาใช้ 1 สัปดาห์ นอกจากนี้ยังสัมพันธ์กับคะแนนแบบสอบวัดผลสัมฤทธิ์อื่นๆ และสัมพันธ์กับ
เกรดผลการเรียนด้วย แบบสอบประเพณีนิยมแม้ว่าจะมีการปรับช่วงค่าความยากแล้วก็ตามพบว่า
มีค่าความคงที่ภายใน (Internal Consistency) เชื่อกันได้ต่ำกว่าแบบสอบรูปแบบปิรามิดทั้ง
5 ชนิด คะแนนของแบบสอบรูปแบบปิรามิดมีการกระจายเป็นรูปสี่เหลี่ยมผืนผ้ามากกว่าคะแนน
ของแบบสอบประเพณีนิยมซึ่งมีการกระจายเป็นแบบเบ้ทางลบ (Negative Skew) ผลการศึกษา
แสดงให้เห็นว่าแบบสอบรูปแบบปิรามิดมีความเหมาะสมในด้านเวลาที่ใช้ ซึ่งโดยเฉลี่ยประมาณ
5 นาที น้อยกว่าเวลาที่ใช้ทำแบบสอบประเพณีนิยม คะแนนที่ได้จากแบบสอบรูปแบบปิรามิด 2
วิธีปรากฏว่ามีค่าสัมประสิทธิ์สหสัมพันธ์สูงกว่าคะแนนที่ได้จากแบบสอบประเพณีนิยม เมื่อพิจารณา

ร่วมกับเกณฑ์ผลการเรียนและแบบสอบอิงเกณฑ์วัดผลสัมฤทธิ์ทางการเรียน (Achievement Test Criterion) การศึกษาในครั้งที่สองก็ยังคงให้ผลเช่นเดียวกับการศึกษาในครั้งแรก

วูด (Larkin and Weiss, 1975 ; citing Wood, 1969) ได้ศึกษาการทดสอบเฉพาะบุคคลรูปแบบปิรามิดที่มี 4, 5 และ 6 ชั้น โดยใช้กฎการแยกทางแบบทำถูกเพิ่มหนึ่งชั้นและทำผิดลดหนึ่งชั้น ได้ใช้ขนาดขั้นคงที่ ซึ่งเท่ากับ 0.05 และให้คะแนนตามจำนวนข้อที่ทำถูก ทำการทดสอบกับกลุ่มตัวอย่าง จำนวน 91 คน ผลการศึกษาพบว่าความสัมพันธ์ระหว่างคะแนนรวมทั้ง 3 ฉบับ ที่ได้จากการทดสอบเฉพาะบุคคลรูปแบบปิรามิดกับเกรดเฉลี่ยในชั้นเรียนเท่ากับ .51 ในขณะที่เมื่อใช้การทดสอบแบบประเพณีนิยมจำนวน 46 ข้อ คะแนนมีความสัมพันธ์กับเกรดเฉลี่ยเท่ากับ .68

ไบรสัน (Bryson, 1971) ได้ศึกษาเปรียบเทียบแบบสอบรูปแบบปิรามิดชนิด 5 ชั้น 2 ชุด กับแบบสอบประเพณีนิยม 2 ชุดๆ ละ 5 ข้อ สร้างจากข้อสอบที่สัมพันธ์กันจำนวน 100 ข้อ แบบสอบประเพณีนิยมดำเนินการสอบโดยวิธีเขียนตอบ ขณะที่แบบสอบรูปแบบปิรามิดดำเนินการสอบโดยใช้เครื่องคอมพิวเตอร์ แบบสอบรูปแบบปิรามิดชุดหนึ่งใช้วิธีเลือกลำดับข้อสอบที่ขึ้นอยู่กับความแตกต่างของข้อสอบที่มากที่สุดสำหรับผู้สอบที่ตอบได้ถึงจุดที่กำหนดไว้ในโครงสร้างของรูปแบบปิรามิด ขณะที่แบบสอบรูปแบบปิรามิดอีกชุดหนึ่งใช้กระบวนการเลือกข้อสอบโดยการทำนายค่าสูงสุดของคะแนนรวม แบบสอบรูปแบบปิรามิดทั้ง 2 ฉบับใช้ค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันแบบแปรผัน นำแบบสอบรูปแบบปิรามิดแต่ละชุดไปทดสอบกับกลุ่มตัวอย่างจำนวน 236 คน โดยจัดเป็นกลุ่มแปรผัน 2 กลุ่ม และนำแบบสอบประเพณีนิยมทั้ง 2 ชุด ไปทดสอบกับกลุ่มเปรียบเทียบจำนวน 250 คน ผลปรากฏว่าแบบสอบประเพณีนิยมขนาดสั้นชุดหนึ่ง มีความสัมพันธ์กันสูงกับคะแนนรวมในการทดสอบ และสูงกว่าความสัมพันธ์กับแบบสอบรูปแบบปิรามิด แบบสอบรูปแบบปิรามิดชุดหนึ่งมีความสัมพันธ์ต่ำกว่ากับคะแนนรวมของแบบสอบและต่ำกว่าความสัมพันธ์กับแบบสอบประเพณีนิยม

ลอร์ด (Lord, 1971) ได้ใช้คอมพิวเตอร์จำลองสถานการณ์ เพื่อศึกษาประเมินคุณภาพการทดสอบเฉพาะบุคคลแบบเฟล็กซีเลเวล โดยใช้ข้อสอบเฟล็กซีเลเวลจำนวน 119 ข้อ ซึ่งผู้สอบแต่ละคนจะต้องตอบข้อสอบแบบเฟล็กซีเลเวล จำนวน 60 ข้อ ผลการศึกษาพบว่า เมื่อขนาดของช่วงห่างความยากของข้อสอบมากขึ้นจะทำให้ค่าอินฟอร์เมชันลดลงทุกระดับความสามารถ และจะมีประสิทธิภาพสูงสุดเมื่อมีค่าอยู่ระหว่าง 0.033 ถึง 0.067 แบบสอบจะไม่มีประสิทธิภาพเมื่อใช้ช่วงห่างความยากใกล้เคียงกับ 0.1 เมื่อเปรียบเทียบกับ การทดสอบแบบประเพณีนิยม การทดสอบแบบเฟล็กซีเลเวลจะให้ค่าอินฟอร์เมชัน

สูงกว่าการทดสอบแบบประเพณีนิยม เมื่อผู้สอบมีความสามารถสูงหรือต่ำ ส่วนผู้สอบที่มีความสามารถกลางๆ การทดสอบแบบประเพณีนิยมจะให้ค่าอินฟอร์เมชันสูงกว่าการทดสอบแบบแฟล็กซีเบิลเวล

วอเตอร์และเบย์รอฟ (Larkin and Weiss, 1974 ; citing Water and Bayroff, 1971) ได้เปรียบเทียบแบบสอบรูปแบบปิรามิดชนิด 5 ชั้น, 10 ชั้น และ 15 ชั้น และแบบสอบรูปแบบปิรามิดชนิด 10 ชั้นที่มีข้อสอบ 2 ข้อในแต่ละชั้น กับแบบสอบประเพณีนิยม 2 ชุดที่มีความยาวเท่ากัน ทั้งแบบสอบประเพณีนิยม 2 ชุดและแบบสอบรูปแบบปิรามิดทุกชุดมีความแตกต่างกันในเรื่องความแปรเปลี่ยนของค่าความยากของข้อสอบ และค่าอำนาจจำแนกของข้อสอบที่แปรผันอย่างมีระบบ ผลการศึกษาพบว่า คะแนนจากแบบสอบและค่าความสามารถของผู้สอบมีความสัมพันธ์กับการกระจายของค่าความยากของข้อสอบและค่าอำนาจจำแนกของข้อสอบ ทั้งนี้ความสัมพันธ์ของคะแนนที่ได้จากแบบสอบรูปแบบปิรามิดจะมีค่าสูงกว่าความสัมพันธ์ที่ได้จากแบบสอบประเพณีนิยมทุกชุด โดยเฉพาะข้อสอบที่มีค่าอำนาจจำแนกสูงๆ ความสัมพันธ์ของคะแนนจากแบบสอบและค่าความสามารถของผู้สอบที่ได้จากแบบสอบรูปแบบปิรามิดที่มีข้อสอบให้ตอบ 1 ข้อในแต่ละชั้นจะสูงกว่าที่ได้จากแบบสอบรูปแบบปิรามิดที่มีข้อสอบให้ตอบ 2 ข้อในแต่ละชั้นด้วย

มุสลีโอ (Larkin and Weiss, 1974; citing Mussio, 1972) ได้พยายามลดจำนวนข้อสอบในแบบสอบรูปแบบปิรามิดให้น้อยลง โดยใช้ยุทธวิธีสกัดการสะท้อนกลับและยุทธวิธีรักษาการสะท้อนกลับ เพื่อการปรับปรุงรวมถึงการตัดส่วนปลายสุด และปลายต่ำสุดของโครงสร้างรูปแบบปิรามิดออกด้วย ดังนั้น จึงเป็นการจัดข้อสอบที่มีระดับความยากที่กระจัดกระจายออกไปได้หลายข้อ มุสลีโอได้รายงานผลการทดสอบทฤษฎีของเขาในรูปแบบของโค้งการแจกแจงของคะแนนและให้ผลเช่นเดียวกับวิธีของลอร์ด การปรับปรุงแบบสอบรูปแบบปิรามิด โดยวิธีการนี้ให้สารสนเทศที่ค่าเฉลี่ยของการกระจายความสามารถน้อยกว่าที่ได้รับจากแบบสอบประเพณีนิยมที่ให้กราฟเป็นยอดแหลม แต่จะให้สารสนเทศมากกว่าเมื่อความสามารถกระจายออกไปจากค่าเฉลี่ย พบว่า ยุทธวิธีรักษาการสะท้อนกลับให้ความตรงในการประมาณที่เหนือพิสัยของความสามารถมากกว่ายุทธวิธีสกัดการสะท้อนกลับ แม้ว่ายุทธวิธีทั้งสองจะสูญเสียความตรงไปบ้างที่ระดับความสามารถที่กระจายมากๆ แต่ยุทธวิธีทั้งสองยังคงให้ความตรงที่ระดับความสามารถที่กระจายออกไปจากค่าเฉลี่ยมากกว่าแบบสอบประเพณีนิยม

ลาร์คินและไวส์ (Larkin, K.C. and Weiss, D.C. 1974) ได้ทดสอบวัดความสามารถโดยใช้แบบทดสอบเฉพาะบุคคลรูปแบบปิรามิด และใช้เครื่องคอมพิวเตอร์ดำเนินการสอบ โดยใช้โครงสร้างข้อสอบแบบปิรามิด 3 ชุด และแบบสอบ

ประเพณีนิยมที่มีความยากปานกลาง ดำเนินการสอบกับนักเรียนในระดับวิทยาลัยที่เรียน
จิตวิทยา 2 กลุ่ม พบว่า แบบสอบปิรามิด 15 ชั้น ให้ค่าสหสัมพันธ์แบบสอบซ้ำค่อนข้าง
ต่ำกว่าแบบสอบประเพณีนิยมจำนวน 40 ข้อ อย่างไรก็ตามเมื่อพิจารณาตามผลกระทบด้าน
การจำแล้ว แบบสอบรูปแบบปิรามิดให้การประมาณค่าความสามารถที่คงที่มากกว่าแบบสอบ
ประเพณีนิยมที่มีความยาวเท่ากัน

วิลโนว์สกี (Wisniewski, 1986) ได้ตรวจสอบความถูกต้องแม่นยำ
ในการประมาณค่าความสามารถของผู้สอบ จากการประยุกต์ใช้วิธีโมเดลกับการทดสอบ
เอ ที (Adaptive Testing) แบบบี เอส เอ็ม (Binary Search Method :
BSM) การวิเคราะห์ข้อมูล ใช้การวิเคราะห์ถดถอยพหุคูณ (Multiple Regression)
โดยใช้คะแนนจากการทดสอบบี เอส เอ็ม เป็นตัวทำนายคะแนนการทดสอบแบบประเพณีนิยม
ที่ใช้กระดาษ-ดินสอ (Paper Pencil Testing) ผลการศึกษาพบว่า คะแนนจากการ
ทดสอบบี เอส เอ็ม สามารถทำนายคะแนนจากการทดสอบแบบประเพณีนิยมที่ใช้กระดาษ
ดินสอได้ดี นอกจากนี้การทดสอบแบบบี เอส เอ็ม ยังสามารถที่จะควบคุมเวลาในการตอบ
เป็นรายข้อ และสามารถควบคุมสภาพแวดล้อมที่เป็นตัวแปรแทรกซ้อนได้ดีกว่าการทดสอบ
ประเพณีนิยม

แบล็คมอร์ (Blackmore, 1987) ได้ศึกษาเปรียบเทียบการดำเนิน
การสอบ 3 วิธี คือ ใช้คอมพิวเตอร์ดำเนินการสอบ (Computerized) ใช้คอมพิวเตอร์ปรับ
ความยากให้เหมาะสมกับระดับความสามารถของผู้สอบ (Computerized Adaptive
Testing) และการดำเนินการสอบแบบประเพณีนิยม คือ ใช้กระดาษกับดินสอ (Paper
and Pencil) ได้ทำการศึกษากับนักเรียนเกรดสิบสองที่กำลังจะจบการศึกษาในปี ค.ศ. 1986
ของโรงเรียนประจำตำบลการ์เดนโกรฟ (Gardeen Grove High School) ในรัฐ
แคลิฟอร์เนีย จำนวน 24 คน โดยแบ่งออกเป็น 3 กลุ่มๆ ละ 8 คน เพื่อดำเนินการ
ทดสอบ คือ ความถนัดทางการเรียน จำนวน 6 ฉบับย่อย คือ การให้เหตุผลทางภาษา
(Verbal Reasoning) ความสามารถทางตัวเลข (Numerical Ability) การให้
เหตุผลเชิงนามธรรม (Abstract Reasoning) การให้เหตุผลเชิงกล (Mechanical
Reasoning) มิติสัมพันธ์ (Space Relations) และการใช้ภาษา (Language Usage)
ผลจากการศึกษาพบว่า คะแนนจากข้อสอบย่อยทุกฉบับจากการดำเนินการสอบทั้ง 3 วิธี ไม่
แตกต่างกัน ส่วนทัศนคติและความคิดเห็นต่างๆ เกี่ยวกับการดำเนินการสอบทั้ง 3 วิธี จาก
การใช้แบบสอบถามและสัมภาษณ์ปรากฏว่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ฮันกิน (Hankins, 1987) ได้ศึกษาการทดสอบเฉพาะบุคคลโดยใช้

ยุทธวิธีของเบย์ (Bayesian Adapting Test) 2 วิธี คือ แบบกำหนดความสามารถเริ่มต้นของผู้สอบคงที่ (Fixed Entry) กับแบบกำหนดความสามารถเริ่มต้นของผู้สอบแปรผัน (Variable Entry) ผลการศึกษาพบว่าทั้ง 2 วิธี ไม่มีความลำเอียง (Bias) เมื่อประมาณค่าความสามารถของผู้สอบที่มีความสามารถปานกลาง แต่จะมีความลำเอียงเมื่อผู้สอบมีความสามารถค่อนข้างเก่งหรืออ่อนมากๆ ทั้ง 2 วิธีให้ค่าอินฟอร์เมชัน (Information) มีการแจกแจงเป็นรูปสมมาตร (Symmetry) ที่ค่อนข้างโค้ง และการทดสอบแบบกำหนดความสามารถเริ่มต้นของผู้สอบแปรผัน จะต้องใช้จำนวนข้อสอบมากกว่าแบบกำหนดความสามารถเริ่มต้นของผู้สอบคงที่

ลี (Lee, 1987) ได้พัฒนาการทดสอบแบบเอ็มเอเอที (Microcomputerized Adaptive Achievement Test : MAAT) โดยใช้ยุทธวิธี 2 แบบ คือ แบบเฮ็ลล์วอร์ทท์ ของวอลท์ (Wald's Sequential Probability Ratio Test : SPRT) และยุทธวิธีของลอร์ด แบบเฟล็กซ์ลีเวล (Lord's Flexilevel Test) โดยใช้ข้อสอบที่มีอยู่ในคลังข้อสอบพีอีทีเอ (The Pitt Educational Testing Aids : PETA) ได้ทดลองศึกษากับนักศึกษาระดับบัณฑิตศึกษา จำนวน 3 ห้องเรียน ในรายวิชาการวิจัยการศึกษาเบื้องต้น ผลการศึกษาพบว่า

(1) มีความเป็นไปได้สูงในอนาคตที่จะนำเอาการทดสอบแบบเอ็มเอเอที (MAAT) มาใช้ในการทดสอบผลสัมฤทธิ์ทางการเรียน ในระดับบัณฑิตศึกษา ผู้ดำเนินการสอบไม่จำเป็นต้องมีความรู้ทางด้านจิตวิทยาหรือคอมพิวเตอร์มากนักก็สามารถดำเนินการสอบได้ เพียงแต่ดำเนินการตามขั้นตอนต่างๆ ตามคู่มือครุฑที่เขียนไว้อย่างละเอียด

(2) คณะกรรมการทดสอบแบบเอ็มเอเอที (MAAT) มีความสัมพันธ์ทางบวกกับคะแนนการสอบกลางเทอม

(3) นักศึกษามีทัศนคติที่ดีต่อการทดสอบแบบเอ็มเอเอที (MAAT) ในทางบวก
โฮ (Ho, 1989) ได้ใช้ไมโครคอมพิวเตอร์ศึกษาเปรียบเทียบยุทธวิธีการให้ข้อสอบแบบปรับได้ (Adaptive Testing) 3 วิธี คือ ยุทธวิธีของเบย์ (Bayesian) โมดัลเบย์ (Modal Bayesian) และความเป็นไปได้สูงสุด (Maximum Likelihood) ยุทธวิธีทั้ง 3 นี้ ได้เปรียบเทียบโดยใช้ขนาดของคลังข้อสอบ (Bank Sizes) 3 ขนาด คือ ขนาด 86 ข้อ, 71 ข้อ และ 56 ข้อ ชนิดของคลังข้อสอบ (Bank Type) 3 แบบ คือ แบบที่มีข้อสอบแบบสุ่ม แบบที่มีเฉพาะข้อสอบที่มีค่าอำนาจจำแนกสูง และแบบที่มีเฉพาะข้อสอบที่ง่ายๆ ส่วนระดับความสามารถของผู้สอบ มี 3 ระดับ คือ สูง ปานกลาง ต่ำ การวิเคราะห์ข้อมูลโดยใช้การวิเคราะห์ความแปรปรวน 3 ทาง (Three ways Anova

with repeated measures) ผลการศึกษาพบว่า

(1) ยุทธวิธีของโมดัลเบย์ จะมีประสิทธิภาพ (Efficient) สูงสุด ยุทธวิธีของเบย์ จะให้ค่าความเที่ยงสูงสุด และยุทธวิธีความเป็นไปได้สูงสุดจะให้ค่าไม่คงที่ (Inconsistent) ภายใต้การทดสอบในสถานการณ์ต่างๆ

(2) การเลือกขนาดคลังข้อสอบ ดเหมือนว่าจะส่งผลน้อยมาก ยกเว้นในกลุ่มผู้สอบที่มีความสามารถต่ำ เมื่อใช้ข้อสอบเริ่มต้นที่มีความยากสูงกว่าความสามารถที่แท้จริงของผู้สอบ

(3) ถ้าใช้ข้อสอบเริ่มต้นที่มีความยากเท่ากับ หรือน้อยกว่าความสามารถที่แท้จริงของผู้สอบแล้ว ความถูกต้องในการประมาณค่าจะมีมากขึ้น

(4) มีปฏิสัมพันธ์ (Interaction) ระหว่างชนิดของคลังข้อสอบ กับยุทธวิธีของอันดับดีเฟสเลตติง และระหว่างชนิดของคลังข้อสอบกับระดับความสามารถของผู้สอบ

(5) ความแตกต่างระหว่างคุณลักษณะทางสถิติของการเลือกข้อสอบกับไม่เลือกข้อสอบมีความสัมพันธ์กันต่ำ

3.2 งานวิจัยในประเทศ

งามนิตย์ ธาตทอง และปรีชา เครือวรรณ (2527) ได้สร้างและประเมินผลการใช้การทดสอบเฉพาะบุคคลรูปแบบแฟล็กชีเลเวล ในวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 2 โดยใช้การทดสอบเฉพาะบุคคลรูปแบบแฟล็กชีเลเวล 1 ฉบับที่มี 18 ข้อ มีข้อสอบจำนวน 35 ข้อ และแบบสอบประเพณีนิยม 1 ฉบับ ที่มีจำนวนข้อสอบ 35 ข้อ จากการศึกษาพบว่าความสัมพันธ์ระหว่างคะแนนจากการทดสอบเฉพาะบุคคลกับคะแนนผลสัมฤทธิ์ทางการเรียน มีค่าเท่ากับ .7238 และความสัมพันธ์ระหว่างคะแนนจากการทดสอบประเพณีนิยมกับผลสัมฤทธิ์ทางการเรียน มีค่าเท่ากับ .6109 เมื่อทดสอบความแตกต่างระหว่างค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าว ปรากฏว่า ไม่แตกต่างกัน แต่ผู้ถูกทดสอบส่วนใหญ่มีทัศนคติที่ดีต่อการสอบด้วยข้อสอบแบบแฟล็กชีเลเวล

จิราพร ไกรสรศิริเวท (2529) ได้ศึกษาความสัมพันธ์ระหว่างรูปแบบบางประการของการทดสอบเฉพาะบุคคลรูปแบบปิรามิดกับความสามารถทางการเรียน โดยใช้การทดสอบเฉพาะบุคคลรูปแบบปิรามิดที่มี 8 ชั้น 3 แบบ คือ แบบที่ใช้ขนาดขั้นคงที่ แบบขนาดขั้นแปรผัน และแบบข้างตัด และใช้วิธีการให้คะแนน 3 วิธี คือ ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ได้ทำ ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก และให้คะแนนทุกข้อ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 348 คน ผลการศึกษา

พบว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากการทดสอบเฉพาะบุคคลรูปแบบปิรามิด ทั้ง 3 แบบ โดยให้คะแนนทั้ง 3 วิธี กับคะแนนจากการทดสอบแบบประเพณีนิยม มีความสัมพันธ์กันในทางบวกอย่างมีนัยสำคัญทางสถิติทุกค่า เมื่อพิจารณาความสัมพันธ์ภายใน ระหว่างคะแนนที่ได้จากการทดสอบเฉพาะบุคคลรูปแบบปิรามิดทั้ง 3 แบบ โดยให้คะแนน ทั้ง 3 วิธี ปรากฏว่า มีค่าสหสัมพันธ์เป็นบวกอย่างมีนัยสำคัญทางสถิติทุกค่า

ในปี พ.ศ. 2531 นันทิยา นิ่งคำ ได้ศึกษาเปรียบเทียบคุณภาพของ การทดสอบแบบซี เอ ที (Computerized Adaptive Testing : CAT) กับแบบสอบ ประเพณีนิยม (Conventional Test) ในการวัดความสามารถด้านคำศัพท์ภาษาอังกฤษ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 120 คน การทดสอบแบบซี เอ ที ใช้กลุ่มข้อสอบ (Item Pool) จำนวน 361 ข้อ โดยมีค่าความยากอยู่ระหว่าง -0.90552 ถึง 3.61860 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.80334 ถึง 2.0000 และมีค่าการเดาอยู่ ระหว่าง 0.04421 ถึง 0.29886 การประมาณค่าความสามารถของผู้สอบโดยใช้วิธี ความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) ส่วน การทดสอบแบบประเพณีนิยมใช้ข้อสอบทั้งหมด 40 ข้อ โดยมีค่าความยากอยู่ระหว่าง -0.76488 ถึง 2.97884 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.94069 ถึง 2.0000 และมีค่าการเดา อยู่ระหว่าง 0.03462 ถึง 0.28620 ผลการศึกษาพบว่า

(1) การทดสอบแบบซี เอ ที กับแบบสอบประเพณีนิยม มีความตรงเชิงเกณฑ์ สัมพันธ์ไม่แตกต่างกัน แต่การทดสอบแบบซี เอ ที สามารถลดจำนวนข้อสอบลงได้มากกว่า ครั้งหนึ่ง

(2) ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จาก การทดสอบแบบซี เอ ที มีค่าต่ำกว่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ ของผู้สอบจากแบบสอบประเพณีนิยม อย่างมีนัยสำคัญทางสถิติ

และในปีเดียวกันนี้ เสรี ปรมชวลิตโรจน์ (2531) ได้ศึกษาความตรง ในการจำแนกความรอบรู้และการประมาณค่าความสามารถโดยใช้แบบสอบรูปแบบเฟล็กซีเลเวล จำนวน 2 ฉบับ และแบบสอบรูปแบบปิรามิดแบบขนาดขั้นแปรผันจำนวน 2 ฉบับ โดยทำ การศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 5,987 คน ผลการศึกษาพบว่าความตรง ในการจำแนกความรอบรู้เมื่อใช้ข้อสอบที่สร้างขึ้นเป็นเกณฑ์ ปรากฏว่าข้อสอบแบบเฟล็กซีเลเวล และแบบปิรามิดแบบขนาดขั้นแปรผันทุกฉบับ มีความตรงในการจำแนกความรอบรู้ได้อย่างมี นัยสำคัญทางสถิติ และเมื่อเปรียบเทียบค่าความตรงแต่ละแบบ พบว่า แบบสอบแบบเฟล็กซี- เลเวลทั้งสองฉบับ มีความตรงในการจำแนกความรอบรู้ได้สูงกว่าแบบสอบรูปแบบปิรามิดอยู่หนึ่ง

ฉบับ อย่างมีนัยสำคัญทางสถิติ

ต่อมาในปี พ.ศ. 2532 เกริงศักดิ์ สุวรรณภาค (2532) ได้ศึกษาความสัมพันธ์ระหว่างรูปแบบของแบบสอบแยกกลุ่มกับความสามารถทางการเรียน ในวิชาคณิตศาสตร์ โดยทำการศึกษาในตัวแปร 2 ตัว คือ ตัวแปรตัวแรก รูปแบบของแบบสอบแยกกลุ่มซึ่งมี 3 แบบ คือ แบบที่มีค่าความยากของข้อสอบใกล้เคียงกัน แบบที่มีค่าความยากของข้อสอบกระจายจากน้อยที่สุดถึงมากที่สุด และแบบที่มีการแยกกลุ่มสองครั้ง และตัวแปรที่สองคือ วิธีการให้คะแนน ซึ่งมี 2 วิธี คือ ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก และให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 2,455 คน ผลการศึกษาพบว่า

(1) ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการสอบที่ใช้แบบสอบแยกกลุ่มแต่ละฉบับ ซึ่งตรวจให้คะแนนในแต่ละวิธีกับความสามารถทางการเรียนมีค่าเป็นบวกอย่างมีนัยสำคัญทางสถิติทุกค่า และผลการทดสอบความแตกต่างของค่าความสัมพันธ์ดังกล่าวเป็นรายคู่ พบว่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติทุกคู่

(2) ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากการตรวจให้คะแนนทั้ง 2 วิธี ดังกล่าวโดยแยกวิเคราะห์ตามแบบสอบแยกกลุ่มแต่ละแบบ มีค่าเป็นบวกอย่างมีนัยสำคัญทางสถิติทุกค่า และผลการทดสอบค่าสัมประสิทธิ์สหสัมพันธ์ดังกล่าว พบว่า เมื่อพิจารณาในแบบสอบแยกกลุ่มแต่ละแบบ ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนนที่ได้จากการตรวจให้คะแนนทั้ง 2 วิธีมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติทุกคู่

งานวิจัยที่เกี่ยวกับแบบสอบเฉพาะบุคคลดังกล่าวข้างต้นเกี่ยวข้องกับการศึกษาในหลายลักษณะต่อไปนี้เป็น ชนิตของรูปแบบที่ใช้ ประเภทวิชา จำนวนชั้น กลุ่มข้อสอบ จำนวนข้อสอบในแต่ละชั้น ลักษณะของแบบสอบ สหวิธีการแยกทาง ความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกัน การให้คะแนน การดำเนินการสอบรวมทั้งได้มีการศึกษาเปรียบเทียบผลที่ได้จากการทดสอบเฉพาะบุคคลกับการทดสอบประเพณีนิยม หรือเกณฑ์ภายนอกต่างๆ ด้วย อย่างไรก็ตาม ผลการวิจัยดังกล่าวแสดงให้เห็นว่า ยังได้ข้อสรุปที่ขัดแย้งกันอยู่บ้างดังเช่น ผลการวิจัยของวูด (Wood) ในปี ค.ศ. 1969 สรุปว่าแบบสอบประเพณีนิยมจะดีกว่าแบบสอบรูปแบบปิรามิดที่มีค่าความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันมีค่าคงที่ และใช้กฎการแยกทางแบบเพิ่ม 1/ลด 1 ซึ่งให้ผลตรงกันข้ามกับผลการวิจัยของ วอเตอร์ (Water) ในปี ค.ศ. 1964 ที่พบว่าแบบสอบรูปแบบปิรามิดที่ใช้ความแตกต่างระหว่างค่าความยากของข้อสอบในชั้นที่อยู่ติดกันมีค่าคงที่ และใช้กฎการแยกทางแบบเพิ่ม 1/ลด 1 มีความสัมพันธ์กับเกณฑ์และให้ศักยภาพการกระจายของคะแนน



ดีกว่าแบบสอบประเพณีนิยม นอกจากนั้นงานวิจัยส่วนใหญ่เป็นการศึกษาในสถานการณ์จำลอง โดยใช้เครื่องคอมพิวเตอร์ ซึ่งแม้ว่าจะสามารถดำเนินการสอบได้สะดวก คณะกรรมการสอบ ก็ไม่มีผลกระทบจากผู้ดำเนินการสอบ จากตัวผู้สอบ หรือวิธีดำเนินการสอบ การศึกษาตามลักษณะดังกล่าวยังมีข้อจำกัดของการศึกษาในสถานการณ์ทดสอบจริง โดยเฉพาะการศึกษา กับแบบสอบวัดผลสัมฤทธิ์ทางการเรียนในห้องเรียน

จากแนวความคิด ทฤษฎี ผลงานวิจัยที่ผ่านมาข้างต้น และความเป็นไปได้ที่จะ นำเอาการทดสอบเฉพาะบุคคลมาใช้เพื่อการพัฒนาปรับเปลี่ยนการทดสอบที่สามารถวัดระดับความ-สามารถของผู้สอบได้ถูกต้องและเหมาะสมยิ่งขึ้น ผู้วิจัยจึงต้องการพัฒนาแบบสอบเฉพาะบุคคล ในวิชาวิทยาศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 โดยใช้รูปแบบแยกทางคงที่ รูปแบบปิรามิดขนาด ชั้นคงที่ 10 ชั้นในสถานการณ์ทดสอบจริงในห้องเรียนโดยใช้เครื่องคอมพิวเตอร์ในการทดสอบ เนื่องจากยังไม่มีผู้ทำการศึกษาในลักษณะดังกล่าวข้างต้น และให้คะแนนโดยวิธีการประมาณ ค่าความสามารถแบบเบย์เซียน อัปเดตติง (Bayesian Updating)

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย