บทที่ 2

วรรณคดีที่เกี่ยวข้อง

จากการศึกษาบทความ งานวิจัย และตำราต่าง ๆ ที่เกี่ยวข้องกับการพัฒนาโปรแกรมสำหรับการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ ผู้วิจัยสามารถสรุปและนำเสนอเนื้อหาได้ 4 ตอน ดังนี้ คือ

ตอนที่ 1 การทดสอบแบบปรับเหมาะ

ตอนที่ 2 ทฤษฎีการตอบสนองข้อสอบ

ตอนที่ 3 ระเบียบวิธีวิจัยและพัฒนาโปรแกรมคอมพิวเตอร์ทางการศึกษา

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ตอนที่ 1 การทดสอบแบบปรับเหมาะ

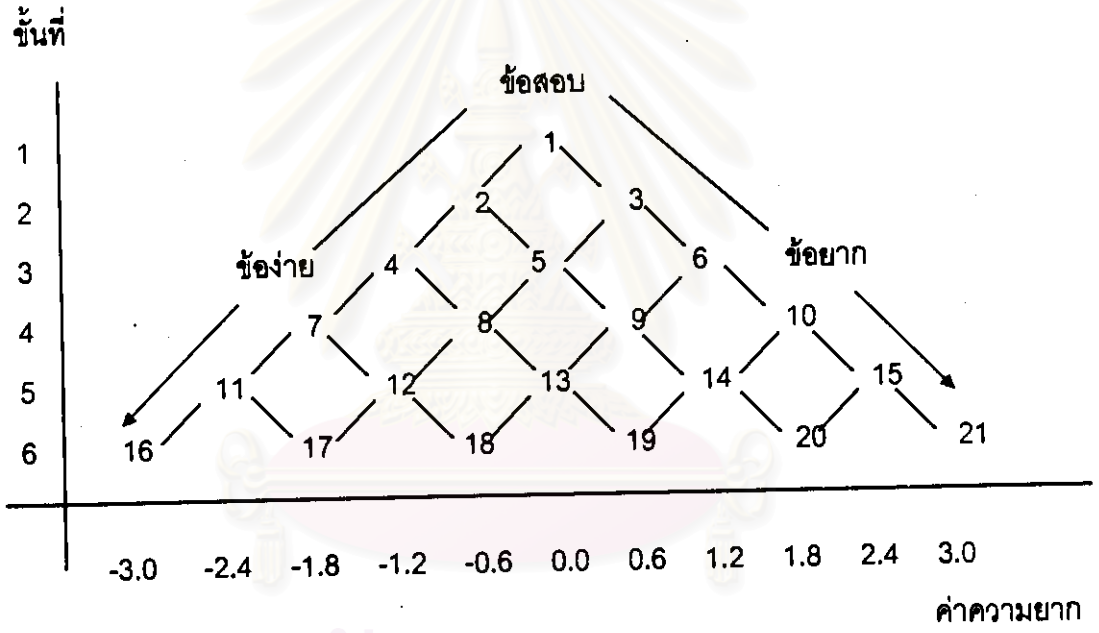
การทดสอบแบบปรับเหมาะมี ความหมาย หลักการและประเภทดังรายละเอียดต่อไปนี้

1. <u>ความหมายของการทดสอบแบบปรับเหมาะ</u>

การทดสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบ (Adaptive Testing) หมายถึง การทดสอบที่ใช้แบบสอบต่างชุดกันสำหรับผู้สอบต่างกัน โดยมีการคัดเลือกข้อสอบที่มีความยากให้เหมาะกับระดับความสามารถของผู้สอบเพื่อประมาณค่าความสามารถของผู้สอบได้โดยมีความคลาดเคลื่อนต่ำ (ศิริชัย กาญจนวาสี, 2538)

2. หลักการของการทดสอบแบบปรับเหมาะ

การทดสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบ หลักการคัดเลือกข้อสอบคือเลือกข้อคำถามที่ง่ายหรือยากที่เหมาะสมกับระดับความสามารถของผู้สอบโดยข้อสอบข้อแรกเป็นข้อสอบที่มีความยากปานกลาง เมื่อผู้สอบตอบข้อคำถามแล้วจะมีการประมาณค่าความสามารถของผู้สอบและเลือกข้อสอบข้อถัดไปที่มีความเหมาะสมกับผู้สอบมาตอบต่อไป โดยทั่ว ๆ ไป ถ้าผู้สอบตอบข้อสอบข้อที่ผ่านมาผิดข้อสอบข้อถัดไปจะง่ายลง ถ้าผู้สอบตอบข้อสอบข้อที่ผ่านมาถูกข้อถัดไปจะยากขึ้น การทดสอบจะยุติลงเมื่อสามารถประมาณค่าความสามารถของผู้สอบได้โดยมีความคลาดเคลื่อนต่ำถึงระดับที่ต้องการ (Green, 1984) ดังแผนภูมิที่ 1



แผนภูมิที่ 1 โครงสร้างของการทดสอบแบบปรับเหมาะ

3. ประเภทของการทดสอบแบบปรับเหมาะ

การทดสอบแบบปรับเหมาะซึ่ง ไวส์ (Weiss,1874 อ้างถึงใน ตรึงใจ พูลผลอำนวย,2534) มีวิธีเลือกข้อสอบที่มีระดับความยากให้เหมาะสมกับความสามารถของผู้สอบแต่ละคนนั้น สามารถแบ่งได้เป็น 2 วิธี คือ กลยุทธสองขั้นตอน (Two - Stage Strategies) และกลยุทธหลายขั้นตอน (Multi - Stage Strategies) ซึ่งกลยุทธหลายขั้นตอนมีรูปแบบต่าง ๆ ดังนี้ คือ

3.1 รูปแบบแยกทางคงที่ (Fixed Branching Model)

3.1.1 รูปแบบปิรามิด (Pyramidal Model) ได้แก่

1) รูปแบบปิรามิดขนาดขั้นคงที่ (Constant Step Size Pyramidal Model)

2) รูปแบบปิรามิดขนาดขั้นแปรผัน (Variable Step Size Pyramidal Model)

3) รูปแบบปิรามิดข้างตัด (Truncated Pyramidal Model)

4) รูปแบบปิรามิดแบบมีหลายข้อในแต่ละขั้น (Multiple -Item Pyramidal Model)

5) รูปแบบปิรามิดแบบให้น้ำหนักตัวเลือกเพื่อแยกทาง (Differential Response Option Branching Pyramidal Model)

3.1.2 รูปแบบเฟล็กซิเลเวล (Flexilevel Model)

3.2.3 รูปแบบปรับระดับขั้น (Stradaptive Model)

3.2 รูปแบบแยกทางแปรผัน (Variable Branching Model) ซึ่งแบ่งเป็นการประมาณค่าได้ 2 แบบ คือ

3 .2.1 การประมาณค่าแบบเบส์ (Bayesian Estimation)

3.2.2 การประมาณค่าแบบความเป็นไปได้สูงสุด (Maximum Likelihood Estimation)

จากรูปแบบของการทดสอบแบบปรับเหมาะซึ่งมีหลายรูปแบบนั้นจะขอนำเสนอพอเป็นสังเขปดังนี้

1. กลยุทธสองขั้นตอน (Two - Stage Strategies) กลยุทธสองขั้นตอนนี้จะประกอบด้วยแบบสอบ 2 ฉบับ แบบสอบฉบับแรกเรียกว่า แบบสอบเพื่อแยกทิศทาง (Routing test) จะประกอบด้วยข้อสอบที่มีระดับความยากกระจายช่วงกว้าง หรือประกอบด้วยข้อสอบที่มีความยากปานกลาง โดยปกติจะมีประมาณ 10 ข้อ ซึ่งผู้สอบจะต้องทำทุกข้อเหมือนกัน เพื่อจะนำผลการสอบที่ได้มาเป็นเกณฑ์ในการแยกผู้สอบเพื่อไปทำแบบสอบฉบับที่ 2 ที่เรียกว่า แบบสอบวัด (Main or Measurement Test) ซึ่งแบบสอบวัดจะประกอบด้วยแบบสอบย่อย จำนวน 3 - 5 ฉบับ แยกตามระดับความยาก ซึ่งแต่ละฉบับจะมีจำนวนข้อสอบประมาณ 20 - 30 ข้อ ผู้สอบแต่ละคนจะได้ตอบแบบสอบชุดที่สองต่างกันขึ้นอยู่กับความสามารถที่ประเมินได้จากแบบสอบฉบับแรก ถ้าผู้สอบได้รับการประเมินว่ามีความสามารถสูงก็จะได้ตอบแบบสอบย่อยฉบับที่ยาก ถ้าได้รับ

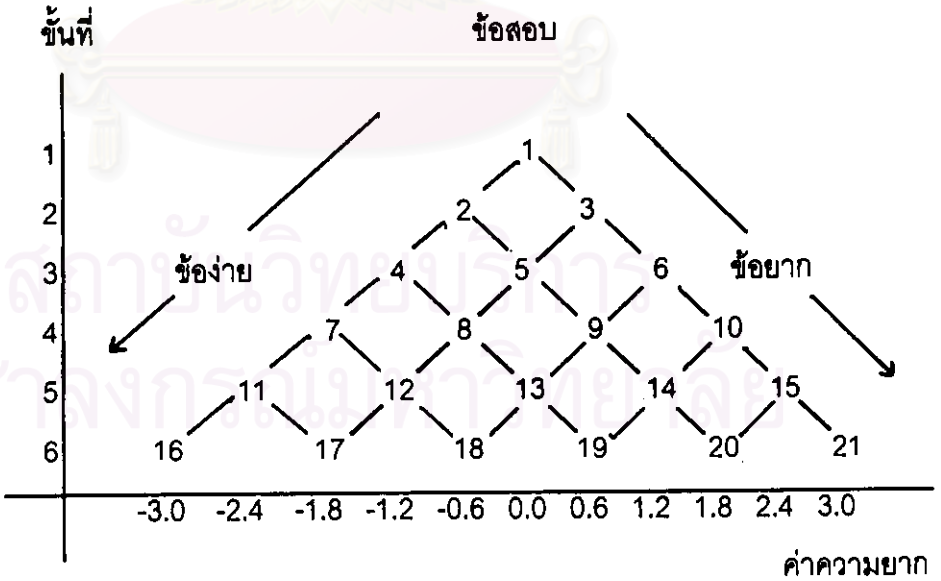
การประเมินว่ามีความสามารถปานกลาง จะได้ตอบแบบสอบย่อยฉบับที่ยากปานกลาง ถ้าได้รับการประเมินว่ามีความสามารถต่ำ จะได้ตอบแบบสอบย่อยฉบับที่ง่าย

2. กลยุทธหลายขั้นตอน (Multi - Stage Strategies)

2.1 รูปแบบแยกทางคงที่ (Fixed Branching Model) หมายถึงการทดสอบด้วยข้อสอบเฉพาะตัวที่มีหลายขั้นตอนในแต่ละขั้นตอนอาจจะมีข้อสอบเพียงข้อเดียวหรือหลายข้อก็ได้ โดยมีรูปแบบหรือเส้นทางในการตอบข้อสอบที่แน่นอน ประกอบด้วย 3 รูปแบบ ดังนี้

2.1.1 รูปแบบปิรามิด (Pyramidal Model) เป็นรูปแบบ การทดสอบด้วยข้อสอบเฉพาะตัว ที่จัดเป็นรูปสามเหลี่ยม ซึ่งการทดสอบรูปแบบนี้มีการศึกษาในหลาย ๆ ลักษณะโดยการใช้ขนาดขั้นคงที่ (Constant Step Size Pyramidal Model) ขนาดขั้นแปรผัน (Variable Step Size Pyramidal Model) ปิรามิดข้างตัด (Truncated Pyramidal Model) ปิรามิดแบบมีหลายข้อในแต่ละขั้น (Multiple -Item Pyramidal Model) ปิรามิดแบบให้น้ำหนักตัวเลือกเพื่อแยกทาง (Differential Response Option Branching Pyramidal Model) โดยแต่ละรูปแบบมีลักษณะดังนี้

1) รูปปิรามิดแบบขนาดขั้นคงที่ (Constant Step Size Pyramidal Model) ได้แก่ การทดสอบแบบปรับเหมาะที่จัดเป็นรูปสามเหลี่ยม โดยมีจำนวนข้อสอบในแต่ละขั้นเท่ากับลำดับที่ของขั้น กล่าวคือ ขั้นที่ 1 จะมีข้อสอบ 1 ข้อ ขั้นที่ 2 จะมีข้อสอบ 2 ข้อ ขั้นที่ 6 จะมีข้อสอบ 6 ข้อ ดังแผนภูมิที่ 2

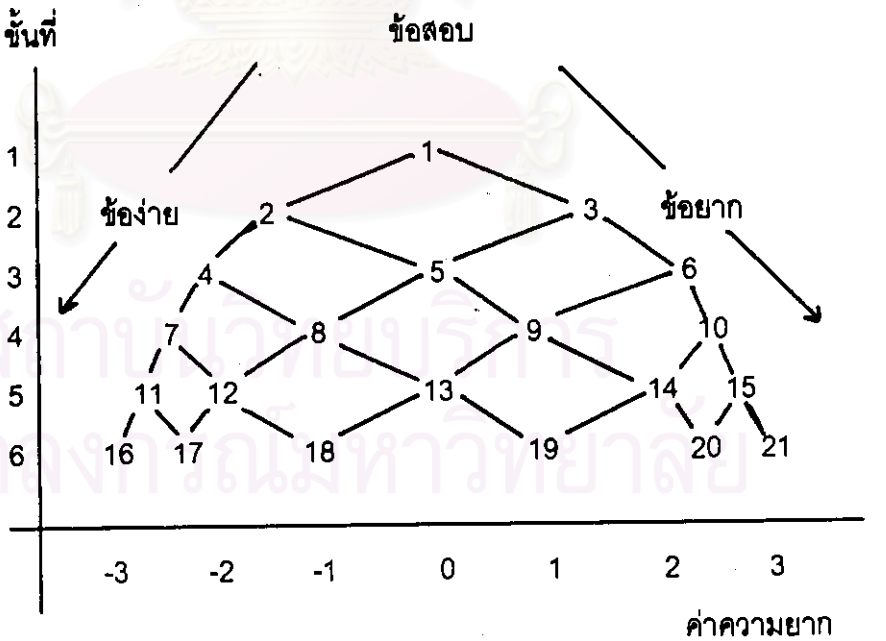


แผนภูมิที่ 2 โครงสร้างของการทดสอบแบบปรับเหมาะรูปปิรามิดแบบขนาดคงที่ ที่มี 6 ขั้น

จากภาพ แกนนอนแสดงถึงค่าความยากของข้อสอบ (b) ที่สัมพันธ์กับข้อสอบในการทดสอบแบบปรับเหมาะรูปปิรามิด ซึ่งจะเห็นว่ามีค่าความยากอยู่ระหว่าง -3.0 ถึง +3.0 โดยแบ่งออกเป็น 10 ช่วง เท่า ๆ กัน ฉะนั้นค่าความยากจะห่างกันช่วงละ 0.6 ข้อสอบที่อยู่ในแนวตั้งแนวเดียวกันจะมีค่าความยากเท่ากัน และช่วงห่างค่าความยากของข้อที่อยู่ติดกันภายในขั้นมีค่าเท่ากันตลอด คือ 1.2

ในการทำแบบทดสอบ ผู้สอบจะต้องทำข้อสอบขั้นละหนึ่งข้อข้อ โดยเริ่มทำข้อที่อยู่บนยอดสามเหลี่ยมซึ่งมีค่าความยากปานกลาง ถ้าทำถูกขั้นต่อไปจะแยกไปทำข้อที่ยากขึ้น ถ้าทำผิดขั้นต่อไปจะทำข้อที่ง่ายขึ้น กระบวนการทดสอบจะเป็นเช่นนี้จนถึงขั้นสุดท้าย

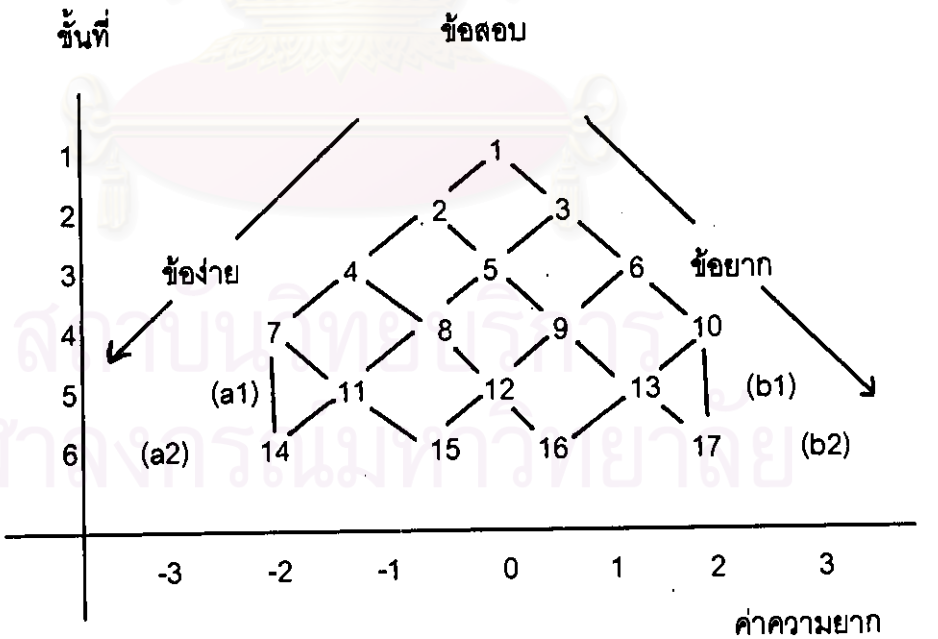
2) ขนาดขั้นแปรผัน (Variable Step Size Pyramidal Model) จากปิรามิดขนาดขั้นคงที่ ขาดประสิทธิภาพเกี่ยวกับความไวในการกำหนดเส้นทางการตอบข้อสอบ จึงได้กำหนดช่วงห่างค่าความยากของข้อสอบที่อยู่ติดกันภายในขั้น ให้มีค่าไม่เท่ากันโดยที่ข้อที่อยู่กลาง ๆ มีช่วงห่างของค่าความยากมาก แล้วค่อย ๆ ลดลงในข้อที่ค่อนไปทางง่ายและยาก (Lord,1971) ดังแผนภูมิที่ 3



แผนภูมิที่ 3 โครงสร้างของการทดสอบแบบปรับเหมาะรูปปิรามิดแบบขนาดขั้นแปรผันที่มี 6 ขั้น

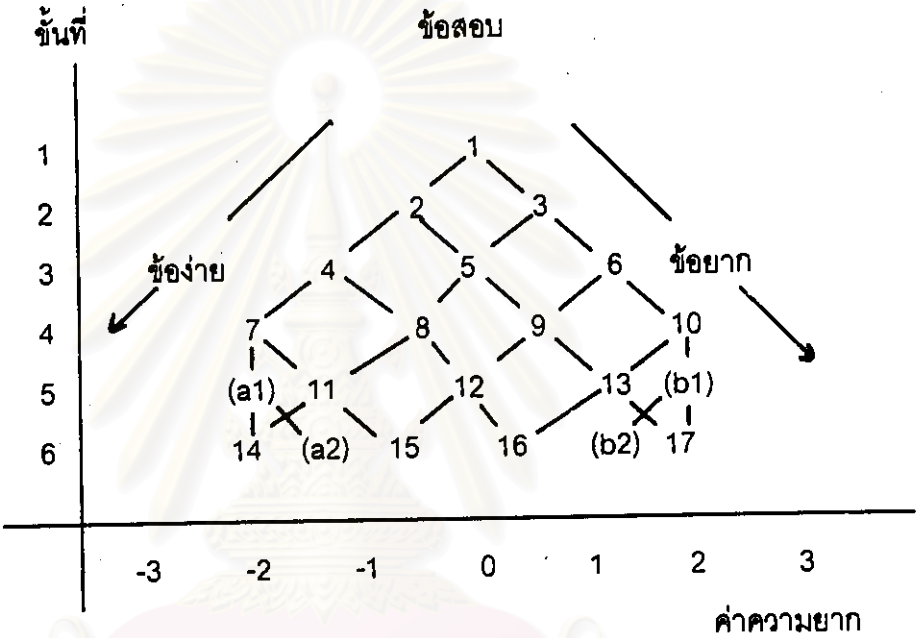
จากภาพ ในขั้นที่ 1 มีข้อสอบเพียงข้อเดียว ซึ่งมีค่าความยากเท่ากับ 0.000 ขั้นที่ 2 มีข้อสอบ 2 ข้อ คือข้อ 2 มีค่าความยากอยู่ระหว่าง -3.000 กับ 0.000และข้อ 3 มีค่าความยากอยู่ระหว่าง +3.000 กับ 0.000 ขั้นที่ 3 มีข้อสอบ 3 ข้อ คือ ข้อ 4 มีค่าความยากเท่ากับ -2.250 ซึ่งเป็นค่าที่อยู่ระหว่าง-3.000 กับ -1.500 (ค่าความยากของข้อ 2) ข้อ 5 มีค่าความยากเท่ากับ 0.000 ข้อ 6 มีค่าความยากเท่ากับ +2.250 ซึ่งเป็นค่าที่อยู่ระหว่าง +3.000 กับ 1.500 (ค่าความยากของข้อ 3) ขั้นที่ 4 มีข้อสอบ 4 ข้อ คือ ข้อ7 มีค่าความยากเท่ากับ -2.625 ซึ่งเป็นค่าที่อยู่ระหว่าง -3.000 กับ -2.250 (ค่าความยากของข้อ 4) ข้อ 8 มีค่าความยากเท่ากับ -1.125 ซึ่งเป็นค่าที่อยู่ระหว่าง -2.250 (ค่าความยากของข้อ 4) กับ -0.000 (ค่าความยากของข้อ 5) ข้อ 9 มีค่าความยากเท่ากับ 1.125 ซึ่งเป็นค่าที่อยู่ระหว่าง 0.000 (ค่าความยากของข้อ 5) กับ 2.250 (ค่าความยากของข้อ 6) ข้อ 10 มีค่าความยากเท่ากับ 2.625 ซึ่งเป็นค่าที่อยู่ระหว่าง 2.250 (ค่าความยากของข้อ 6) กับ 3.000 สำหรับขั้นต่อ ๆ ไป ก็จะใช้หลักการเดียวกัน

3) ปิรามิดข้างตัด (Truncated Pyramidal Model) เสนอขึ้นมาเพื่อต้องการลดจำนวนข้อสอบจากรูปปิรามิดขั้นคงที่ลง จึงใช้วิธีสกัดการสะท้อนกลับ หรือวิธีรักษาการสะท้อนกลับ(Re-flecting or Retaining Barrier) (Weiss,1974) ดังแผนภูมิที่ 4,5



แผนภูมิที่ 4 โครงสร้างของการทดสอบแบบปรับเหมาะรูปปิรามิดแบบข้างตัด สกัดการสะท้อนกลับ

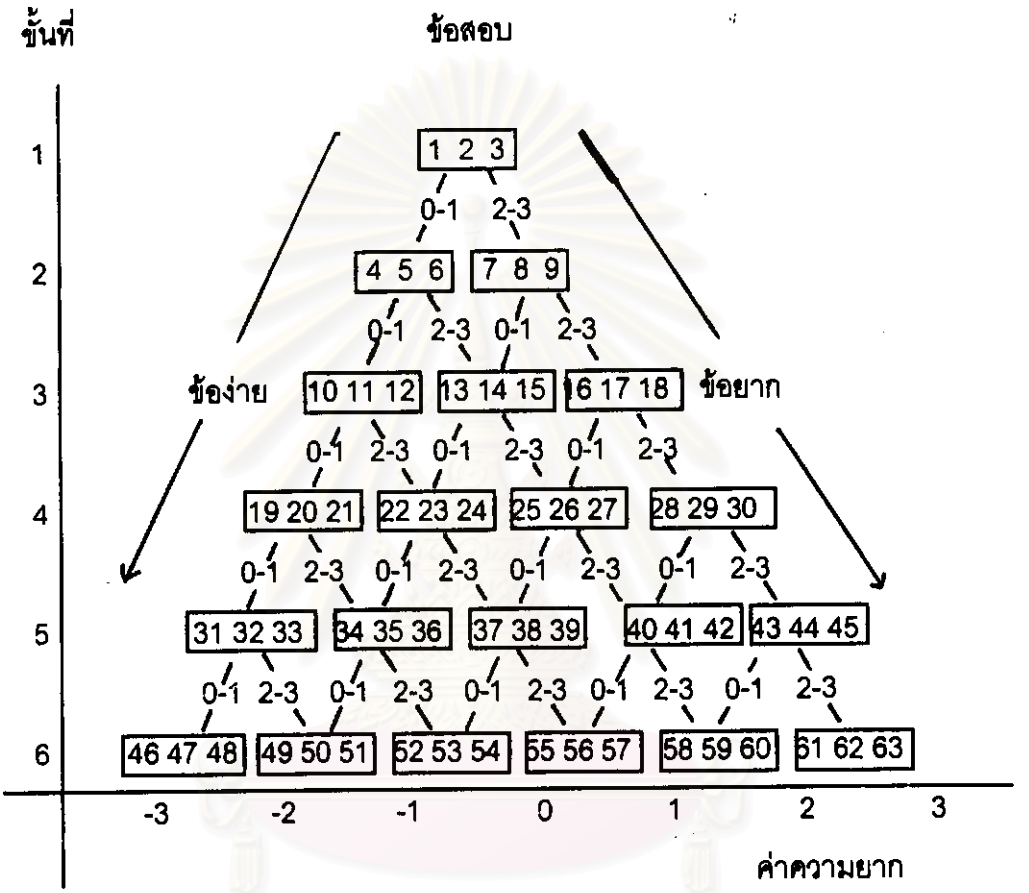
จากภาพ การตอบข้อสอบใน ขั้นที่ 1 ถึง ขั้นที่ 3 มีวิธีการตอบเช่นเดียวกับรูปปิรามิดแบบขนาดขั้นคงที่ จนถึงขั้นที่ 4 ถ้าตอบข้อสอบข้อที่ 7 ถูก ก็จะแยกไปทำข้อสอบข้อ 11 แต่ถ้าตอบข้อ 7 ผิดก็จะไปทำข้อ 14 และในทำนองเดียวกันถ้าตอบข้อสอบข้อ 10 ผิด ก็จะไปทำข้อ 13 ถ้าตอบข้อ 10 ถูก จะไปทำข้อ 17 เลย



แผนภูมิที่ 5 โครงสร้างของการทดสอบแบบปรับเหมาะรูปปิรามิดแบบข้างตัด รักษา การสะท้อนกลับ

จากภาพ การตอบข้อสอบในขั้นที่ 1 ถึงขั้นที่ 3 วิธีตอบเช่นเดียวกับรูปแบบปิรามิดขนาดขั้นคงที่ จนถึงจุดสกัดในขั้นที่ 4 ถ้าตอบข้อสอบข้อที่ 7 ถูกก็จะแยกไปตอบข้อสอบข้อ 11 ถ้าตอบข้อสอบข้อที่ 7 ผิด ก็จะลงไปตอบในข้อ a1 ซึ่งมีค่าความยากเท่ากับข้อ 7 และถ้าตอบข้อ a1 ผิดก็จะลงไปตอบข้อ 14 แต่ถ้าตอบข้อ a1 ถูก จะแยกไปตอบข้อ a2 ซึ่งมีค่าความยากเท่ากับข้อ 11

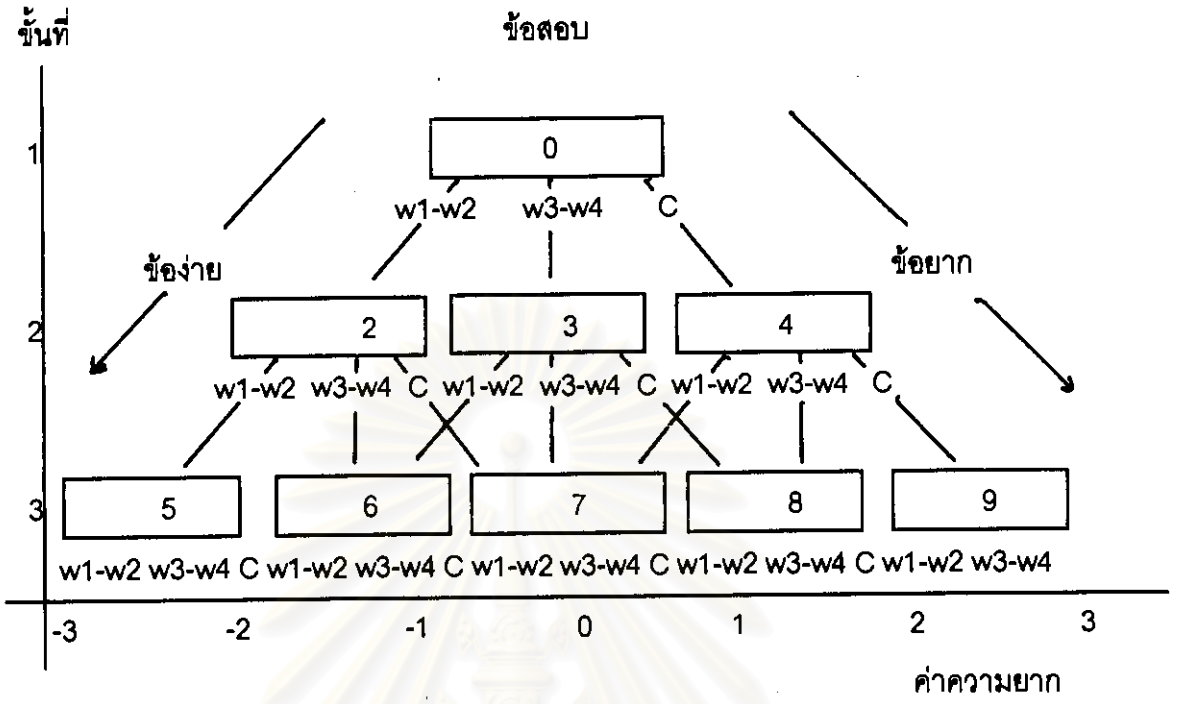
4) ปิรามิดแบบมีหลายข้อในแต่ละขั้น (Multiple -Item Pyramidal Model) พัฒนาขึ้นเพื่อให้การแยกทางเพื่อทำข้อสอบในขั้นต่อไปถูกต้องยิ่งขึ้น ดังตัวอย่าง แผนภูมิที่ 6



แผนภูมิที่ 6 โครงสร้างของการทดสอบแบบปรับเหมาะรูปปิรามิด แบบมีข้อสอบหลายข้อในแต่ละขั้น

จากภาพ ในขั้นที่ 1 ถ้าผู้สอบตอบข้อสอบถูก 0 ถึง 1 ข้อจะแยกไปทำข้อที่ 4,5,6 แต่ถ้าตอบข้อสอบถูก 2-3 ข้อ จะแยกไปทำข้อที่ 7,8,9 สำหรับขั้นต่อ ๆ ไปก็เช่นเดียวกัน

5) ปิรามิดแบบให้น้ำหนักตัวเลือกเพื่อแยกทาง (Differential Response Option Branching Pyramidal Model) พัฒนาขึ้นเพื่อประหยัดการใช้ข้อสอบ กล่าวคือ จากการที่ใช้ข้อสอบหลาย ๆ ข้อในแต่ละขั้น ก็เปลี่ยนมาดูที่ตัวเลือกของข้อสอบแทน ซึ่งรูปแบบนี้จะเหมาะกับแบบทดสอบแบบเลือกตอบเท่านั้น ดังแผนภูมิที่ 7



แผนภูมิที่ 7 โครงสร้างของการทดสอบแบบปรับเหมาะรูปปิรามิดแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบ เพื่อแยกทาง

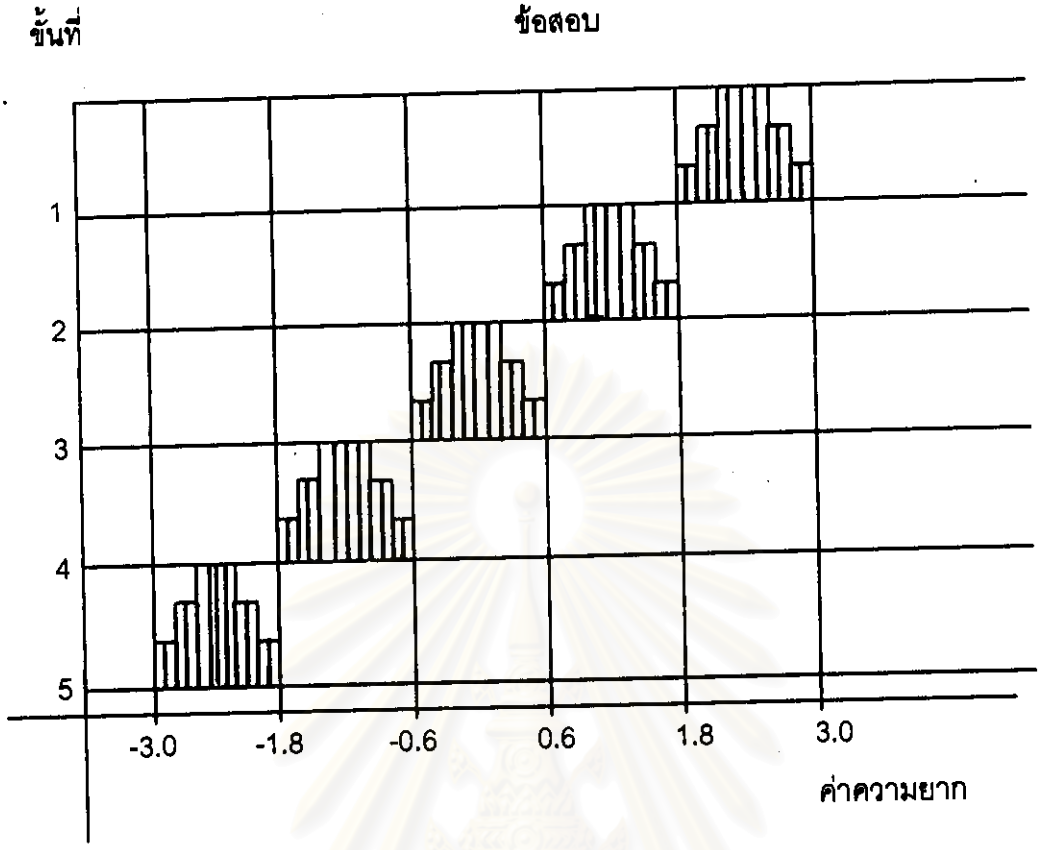
2.1.2 รูปแบบเฟล็กซิเลเวล (Flexilevel Model) เป็นรูปแบบของการทดสอบด้วยข้อสอบเฉพาะตัว ที่ลอร์ด (Lord,1971;Weiss.1982) คิดขึ้นเพื่อใช้กับการทดสอบด้วยการเขียนตอบ ลักษณะทั่วไปของแบบสอบเฟล็กซิเลเวล จะประกอบด้วยชุดของข้อสอบจำนวนหนึ่ง ซึ่งมีช่วงห่างของค่าความยากเท่า ๆ กัน โดยเรียงจากข้อที่มีง่ายที่สุดไปยังข้อที่ยากที่สุด โดยข้อสอบ กลุ่มง่ายจะพิมพ์ด้วยหมึกสีแดง ข้อสอบกลุ่มยากจะพิมพ์ด้วยหมึกสีน้ำเงิน การทดสอบจะทำโดยให้ผู้สอบทำข้อที่มีความยากปานกลางถ้าทำข้อเริ่มต้นถูกจะไปตอบข้อที่ยากขึ้น ถ้าตอบผิดจะไปทำข้อที่ง่ายลง ดังแผนภูมิที่ 8

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

แผนภูมิที่ 8 ตัวอย่างข้อสอบแบบเฟล็กซิเลเวล (Flexilevel Test)

สำหรับกระดาษคำตอบ ได้ออกแบบพิเศษ คือ เมื่อนักเรียนตอบถูก จุดสีน้ำเงินก็จะปรากฏขึ้น ข้อต่อไปผู้สอบจะต้องทำข้อสอบในกลุ่มสีน้ำเงิน (กลุ่มยาก) ในข้อที่มีหมายเลขต่ำสุดที่ยังไม่ได้ทำ และถ้าตอบผิดจุดสีแดงจะปรากฏขึ้น ผู้สอบจะต้องไปเลือกทำข้อสอบในกลุ่ม สีแดง (กลุ่มง่าย) ในข้อที่หมายเลขต่ำสุดที่ยังไม่ได้ทำ ถ้าให้ k เป็นจำนวนข้อสอบเดิม ผู้สอบในข้อสอบเฟล็กซิเลเวลจะต้องทำข้อสอบทั้งหมด (k+1) /2 ข้อ (Lord, 1971)

2.1.3 รูปแบบปรับระดับขั้น (Stradaptive Model) เป็นรูปแบบการสอบที่พัฒนามาจากแบบสอบเฟล็กซิเลเวลโดยให้มีข้อสอบในแต่ละระดับความยากมากกว่า 1 ข้อ ซึ่งเรียกว่าระดับขั้น (Stratum) โดยจะเริ่มทำข้อสอบที่ระดับความยากใดก็ได้ ถ้าตอบถูกก็ไปตอบข้อที่ยากขึ้น ถ้าตอบผิดก็ไปทำข้อที่ง่ายลง และยุติการสอบเมื่อผู้สอบไม่สามารถทำข้อสอบในระดับขั้นนั้นถูกเลย ดังตัวอย่างแผนภูมิที่ 9



แผนภูมิที่ 9 โครงสร้างของการทดสอบแบบปรับเหมาะแบบสเตร็ดเด็ฟตีฟ เทสติ้ง (Stradaptive Testing)

2.2 รูปแบบแยกทางแปรผัน (Variable Branching Model) เป็นรูปแบบการสอบที่ไม่ได้กำหนดโครงสร้างและข้อสอบไว้ล่วงหน้า ว่าถ้าผู้สอบตอบผิดหรือจะต้องไปตอบข้อสอบข้อที่ง่ายลงหรือยากขึ้นตามที่กำหนดไว้ แต่จะเป็นการสอบโดยที่ผู้สอบตอบข้อสอบข้อหนึ่งแล้วคำนวณค่าความสามารถของผู้สอบเลือกข้อสอบจากกลุ่มข้อสอบที่มีความเหมาะสมกับความสามารถของผู้สอบที่คำนวณแล้วมาให้ผู้สอบทำในข้อต่อไป การสอบจะดำเนินต่อไปจนกระทั่งผู้สอบทำข้อสอบครบตามที่กำหนด หรือ การประมาณค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ [Standard Error ; SE (θ)] ลดต่ำลงจนถึงระดับที่ยอมรับได้ก็จะยุติการสอบ

ในการคำนวณค่าความสามารถของผู้สอบ ในการทดสอบปรับเหมาะรูปแบบแยกทางแปรผันนั้น มีวิธีการคำนวณ 2 วิธี ดังนี้

2.2.1 การประมาณค่าแบบเบส์ (Bayesian Estimation) เป็นกลยุทธที่ได้จากการประยุกต์ทฤษฎีของเบส์ โดยจะใช้หลักการที่ว่า ข้อสอบทุกข้อที่ยังไม่ได้นำมาให้ผู้สอบคนนั้น ๆ

สอบ จะถือว่าเป็นข้อสอบที่มีโอกาสนำมาใช้ได้ กระบวนการนี้แสดงให้เห็นว่าข้อสอบใด ๆ ในกลุ่มข้อสอบที่นำมาใช้ในการทดสอบกับผู้สอบคนใดก็ตามจะเป็นข้อสอบที่ลดความไม่แน่นอนในการประมาณค่าความสามารถของผู้สอบมากที่สุด ข้อสอบที่นำมาใช้กับผู้สอบจะเป็นข้อสอบที่มีระดับความยากใกล้กับระดับความสามารถของผู้ตอบ หลังจากดำเนินการสอบโดยใช้ข้อสอบที่คัดเลือกไว้ก็จะประมาณความสามารถก่อน และสารสนเทศที่ได้จากการสอบนั้น ก็จะนำมารวมกันโดยวิธีทฤษฎีเบส์ จะได้การประมาณค่าความสามารถหลังการสอบ การประมาณค่าครั้งหลังนี้เป็นการประมาณค่าที่ปรับจากสิ่งที่รู้เกี่ยวกับข้อมูลของผู้สอบ และค่าที่ประมาณได้นี้ก็จะเป็นค่าสารสนเทศก่อนอีก กระบวนการนี้จะสิ้นสุดเมื่อความแปรปรวนภายหลังมีค่าน้อยกว่าค่าที่กำหนดไว้

2.2.2 การประมาณค่าแบบความเป็นไปได้สูงสุด (Maximum Likelihood Estimation) เป็นกระบวนการที่คล้ายกันกับกระบวนการเบส์ โดยมีหลักการดังนี้ เมื่อผู้สอบตอบข้อสอบข้อหนึ่ง ๆ ถูกต้องและอีกข้อหนึ่งผิดก็จะสามารถนำไปแก้สมการความเป็นไปได้สูงสุดและจะได้ค่าการประมาณความสามารถและความคลาดเคลื่อนมาตรฐาน ข้อสอบข้อถัดไปที่จะนำมาใช้สอบจะเป็นข้อสอบที่มีความยากพอเหมาะกับระดับความสามารถของผู้สอบมากที่สุด เมื่อผู้สอบตอบข้อนั้นแล้วก็จะมีการประเมินคำตอบทันทีโดยจะนำผลจากการตอบข้อสอบที่ผ่านมาแล้วมาประเมินด้วย ถ้าข้อสอบข้อสุดท้ายที่ตอบนั้นผิดการประมาณค่าความสามารถก็จะลดต่ำลง ถ้าตอบข้อสอบข้อสุดท้ายถูกการประมาณค่าความสามารถก็จะสูงขึ้น (Hambleton,1990)

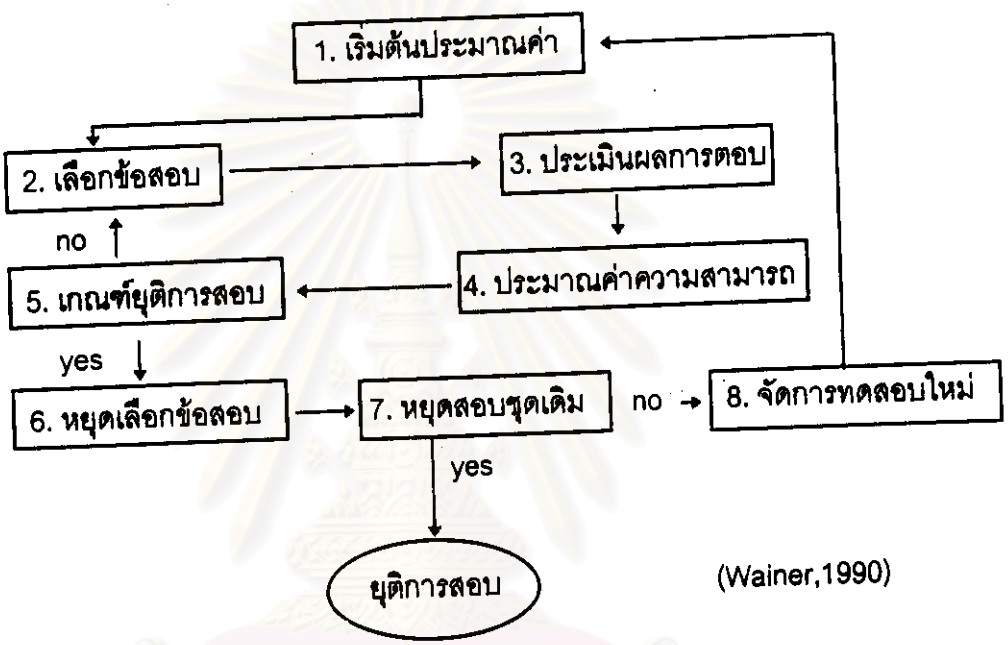
4. <u>การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (Computerized adaptive testing)</u>

การทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ มีหลักการและองค์ประกอบดังรายละเอียดต่อไปนี้

4.1<u>หลักการของการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAT)</u>

หลักการทั่วไปของการทดสอบปรับเหมาะกับระดับความสามารถโดยใช้คอมพิวเตอร์ (CAT) เริ่มจากการประมาณค่าความสามารถของผู้สอบครั้งแรก โดยทั่วไปแล้วจะกำหนดให้ความสามารถของผู้สอบมีค่าปานกลาง เมื่อประมาณค่าได้แล้วคอมพิวเตอร์จะเลือกข้อสอบที่มีระดับความยากใกล้เคียงกับระดับความสามารถของผู้สอบให้ผู้สอบทำ เมื่อผู้สอบทำข้อสอบแล้วคอมพิวเตอร์จะประเมินผลการตอบว่าตอบถูกหรือตอบผิด แล้วประมาณค่าความสามารถของผู้สอบและค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบจากผลการตอบนั้น เมื่อประมาณค่าความสามารถของผู้สอบและความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบแล้วนำค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบไปเทียบกับ

เกณฑ์ยุติการสอบ ถ้าค่าที่ได้นั้นสูงกว่าเกณฑ์ยุติการสอบก็จะทำการเลือกข้อสอบข้อต่อไปให้ผู้ทดสอบทำต่อไปจนกระทั่งค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบที่ได้จากการคำนวณนั้นต่ำกว่าเกณฑ์ยุติการสอบเครื่องคอมพิวเตอร์จะหยุดการเลือกข้อสอบและยุติการทำข้อสอบ แต่ถ้าผู้สอบยังต้องการทำข้อสอบอีกก็จะจัดการทดสอบต่อไป ถ้าไม่ต้องการก็จะยุติการสอบ (ดูแผนภูมิที่ 10)



(Wainer,1990)

แผนภูมิที่ 10 แสดงหลักการทำงานของการทดสอบแบบปรับเหมาะที่ใช้คอมพิวเตอร์

4.2 <u>องค์ประกอบของการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAT)</u>

การทดสอบแบบปรับเหมาะมีองค์ประกอบ 6 องค์ประกอบ (Hambleton,1991) ดังนี้ คือ

1. โมเดลการตอบข้อสอบ

การเลือกโมเดลการตอบข้อสอบที่จะใช้สำหรับ CAT ไวส์และยอส (Weiss and Yoes) กล่าวว่า โมเดลโลจิสติก 3 พารามิเตอร์ มีความเหมาะสมที่สุด ซึ่งสอดคล้องกับ Green,Bock,Linn, Lord (1980),Weiss (1983)ได้กล่าวว่า เพื่อให้การประมาณค่าความสามารถของผู้สอบถูกต้องมากที่สุดควรใช้โมเดลโลจิสติก 3 พารามิเตอร์

2. กลุ่มข้อสอบ

เพื่อความเหมาะสมกับการใช้ทฤษฎีการตอบสนองข้อสอบ(IRT) ข้อสอบควรมีจำนวนมาก โดยที่กลุ่มข้อสอบจะต้องมีค่าความยากครอบคลุมช่วงของระดับความสามารถของ ผู้สอบ และถ้ามีจำนวนข้อสอบมากจะไม่มีปัญหาในเรื่องข้อสอบไม่เพียงพอในการเลือกข้อสอบ และเออรี (Urry,1977 อ้างถึงใน นันทิยา พึ่งคำ,2531) กล่าวว่า กลุ่มข้อสอบควรจะประกอบด้วยข้อสอบอย่างน้อย 100 ข้อ ค่าอำนาจจำแนกของข้อสอบเกิน 0.8 ค่าความยากอยู่ระหว่าง -2.0 ถึง +2.0 และค่าการเดาน้อยกว่า 0.3 แต่ถ้าต้องการให้การทดสอบนั้นไม่มีปัญหาควรกระจายค่าความยากให้มีช่วงกว้างขึ้นคืออยู่ระหว่าง -3.0 ถึง+3.0 และค่าอำนาจจำแนกอาจลดต่ำลงให้ช่วงอยู่ระหว่าง 0 ถึง +2.5

3. การเลือกจุดเริ่มต้นการทดสอบ

ข้อสอบข้อแรกควรมีค่าความยากปานกลาง ไวเนอร์ (Wainer,1990) และแฮมเบิลตัน (Hambleton,1991) กล่าวว่า การเริ่มต้นควรจะเลือกข้อสอบให้มีค่าความยากเท่ากับหรือต่ำกว่าความสามารถของผู้สอบเพื่อให้การประมาณค่าความสามารถของผู้สอบได้ถูกต้องมากยิ่งขึ้นและยังกล่าวว่าการเลือกข้อสอบให้ใกล้เคียงกับความสามารถของผู้สอบนั้นอาจดูจากผลการเรียนของผู้สอบที่ผ่านมาได้

4. การคัดเลือกข้อสอบ

ในการทดสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบ มีรูปแบบการคัดเลือกข้อสอบหลายรูปแบบดังกล่าวมาแล้ว แต่การทดสอบปรับเหมาะกับระดับความสามารถโดยใช้คอมพิวเตอร์นี้ใช้รูปแบบทางแยกแปรผัน กลยุทธที่นิยมใช้อยู่ 3 วิธีคือ

1. วิธีการจับคู่ระหว่าง ค่าความยากของข้อสอบ (b) กับความสามารถของผู้สอบ

2. วิธีการจับคู่ระหว่าง ค่าอำนาจจำแนก,ค่าความยากและค่าการเดา ซึ่งอยู่ในรูปของสมการ m_i กับความสามารถของผู้สอบ เมื่อ

$$m_i = b_i + \frac{1}{Da_i} In\left(\frac{1+\sqrt{1+8c_i}}{2}\right)$$

(Hulin,Darsgow and Parsons, 1983)

3. วิธีใช้ค่าสารสนเทศของข้อสอบที่สูงที่สุด (Maximum item information)

จากการศึกษาพบว่า วิธีใช้ค่าสารสนเทศของข้อสอบที่สูงที่สุดจะให้ค่าความตรงเชิงสภาพสูงที่สุด (รังสรรค์ มณีเล็ก, 2539)

5. คะแนนหรือการประมาณค่าความสามารถของผู้สอบ

คะแนนหรือการประมาณค่าความสามารถของผู้สอบ เพื่อให้การประมาณค่าความสามารถของผู้สอบถูกต้องมากที่สุดโดยยึดตามทฤษฎีการตอบสนองข้อสอบ(IRT) กลยุทธที่เหมาะสมมี 2 วิธี (Hambletom,1991) คือ

5.1 การประมาณค่าโดยวิธีความเป็นไปได้สูงสุด (Maximum Likelihood) (Lord,1980)

5.2 การประมาณค่าโดยวิธีเบส์ (Bayesian updating) (Weiss,1982)

จากการศึกษาพบว่าการประมาณค่าโดยวิธีความเป็นไปได้สูงสุดและการประมาณค่าโดยวิธีเบส์ไม่แตกต่างกันเมื่อผู้สอบมากกว่า 2,000 คน แต่การประมาณค่าโดยวิธีเบส์จะให้ผลดีกว่าเมื่อจำนวนผู้สอบน้อยกว่า 500 คน

6. การยุติการทดสอบ

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ โดยใช้รูปแบบทางแยกแปรผัน นิยมใช้เกณฑ์ยุติการสอบ 2 วิธี คือ

6.1 การกำหนดจำนวนข้อสอบ เป็นการกำหนดจำนวนข้อสอบให้ผู้สอบ ถ้าผู้สอบทำข้อสอบได้ตามกำหนดที่ตั้งไว้โดยปรกติจะตั้งไว้ที่ 25 ข้อ ก็จะยุติการสอบ ซึ่งการกำหนดเกณฑ์นี้มีข้อดีในการเปรียบเทียบสารสนเทศของแบบสอบได้โดยตรงแต่มีข้อจำกัดคือ อาจทำให้การวัดผลครั้งนั้นมีความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบสูง จึงได้มีการกำหนดเกณฑ์การยุติอีกแบบหนึ่งขึ้น

6.2 การกำหนดค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ (Hambleton,1991)

จากการศึกษาโดยใช้ Partial Credit Model พบว่า การยุติการสอบโดยดูจากค่าความคลาดเคลื่อนในการวัด จะให้ผลดีกว่าการยุติการสอบโดยการกำหนดจำนวนข้อสอบ (Dodd,1993) และจากการศึกษาพบว่าการกำหนดค่าความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบน้อยกว่า หรือเท่ากับ 0.3 จะให้ค่าความตรงร่วมสมัยสูงสุด
(รังสรรค์ มณีเล็ก, 2539)

5. <u>การตรวจสอบการทดสอบแบบปรับเหมาะกับระดับความสามารถ</u>

การตรวจสอบการทดสอบแบบปรับเหมาะ แฮมเบิลตัน (Hambleton,1991) กล่าวว่า ควรจะดูจากค่าสัมประสิทธิ์สหสัมพันธ์ ระหว่างค่า θ ที่ได้จาก การทดสอบแบบ

ปรับเหมาะกับระดับความสามารถกับแบบสอบประเพณีนิยม และ Green (1984) ได้ทำการศึกษาพบว่า การตรวจสอบควรดูจาก ความคลาดเคลื่อนในการประมาณค่าความสามารถของผู้สอบ ความตรง การประมาณค่าพารามิเตอร์ การเท่าเทียมกันของการทดสอบแบบปรับเหมาะกับระดับความสามารถ กับแบบสอบเกณฑ์

ตอนที่ 2. ทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (IRT)เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างความสามารถของผู้สอบกับพฤติกรรมการตอบข้อสอบโดยมีหลักการของทฤษฎี โมเดลการตอบสนองข้อสอบและข้อตกลงของทฤษฎี ดังรายละเอียดดังต่อไปนี้

1 หลักการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบมีหลักการ ดังนี้

ทฤษฎีการตอบสนองข้อสอบ เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างลักษณะหรือความสามารถที่มีอยู่ภายในตัวบุคคล กับพฤติกรรมการตอบสนองข้อสอบของบุคคลนั้น โดยทฤษฎีนี้มีความเชื่อว่าพฤติกรรมการตอบสนองข้อสอบของผู้สอบ ซึ่งเป็นสิ่งที่สังเกตได้โดยตรง จะถูกกำหนดโดยคุณลักษณะ (Trait) หรือความสามารถ (Ability) ที่มีอยู่ในตัวบุคคลซึ่งไม่สามารถสังเกตได้และยังเชื่อว่าค่าพารามิเตอร์ต่าง ๆ ของข้อสอบไม่ว่าจะเป็นค่าความยาก (b) ค่าอำนาจจำแนก(a) หรือค่าการเดา(c) ของข้อสอบแต่ละข้อเป็นคุณลักษณะที่มีอยู่ประจำ และคงที่พอสมควรในตัวข้อสอบนั้นจริง ฉะนั้นค่าพารามิเตอร์ของข้อสอบจึงไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง และค่าความสามารถของผู้สอบซึ่งเชื่อว่าเป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบนั้นจริง จึงไม่ควรแปรเปลี่ยนไปตามค่าความยากของข้อสอบ เนื่องจากความสามารถของผู้สอบเป็นคุณลักษณะแฝงซึ่งไม่สามารถวัดหรือสังเกตได้โดยตรง นักวัดผลทางการศึกษาจึงได้พยายามหาความสัมพันธ์ระหว่างปริมาณการกระทำข้อสอบ หรือคะแนน(Test Performance or Score) กับปริมาณความสามารถ (Ability) โดยแสดงความสัมพันธ์ดังกล่าวในลักษณะโมเดลทางคณิตศาสตร์ที่กำหนดด้วยฟังก์ชันทางคณิตศาสตร์ เรียกว่า ฟังก์ชั่นลักษณะข้อสอบ (Item Characteristic Function) หรือเรียกว่าโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curve) หรือฟังก์ชั่นการตอบสนองข้อสอบ(Item Response Function) (ศิริชัย กาญจนวาสี,2538:Hambleton,R.K.,1990)

2 โมเดลการตอบสนองข้อสอบ

โมเดลทางคณิตศาสตร์ ที่แสดงความสัมพันธ์ระหว่างผลการทดสอบ (Test Performance) กับความสามารถของผู้สอบ (Ability or Trait) มีหลายโมเดลด้วยกันแต่ที่สำคัญจะนำเสนอในครั้งนี้มี 3 โมเดล คือ

2.1 นอร์มัล โอไจฟ์ โมเดล (Normal Ogive Model)

2.2 ราชค์ โมเดล (Rasch Model)

2.3 โลจิสติค โมเดล (Logistic Model)

2.1 นอร์มัล โอไจฟ์ โมเดล (Normal Ogive Model) โมเดลนี้พัฒนาขึ้นมาในปี ค.ศ. 1943-1944 โดย ลอว์รีย์ (Lowley,1943 อ้างถึงใน ต่าย เชียงฉี,2534) และลอร์ดได้ศึกษาและพัฒนาต่อ ซึ่งโมเดลนี้ไม่มีผู้นิยมมากนักเนื่องมาจากความซับซ้อนทางคณิตศาสตร์ ยากแก่การปฏิบัติหรือนำมาใช้จริง แต่โมเดลนี้ถือว่าเป็นต้นแบบในการพัฒนาโมเดลโลจิสติค

2.2 ราชค์ โมเดล (Rasch Model) ราชค์ โมเดล พัฒนาขึ้นโดย นักคณิตศาสตร์ชาวเดนมาร์ก ชื่อจอรช์ ราชค์ (Georg Rasch,1960) โมเดลนี้เชื่อว่า ผู้สอบมีความสามารถจริงแฝงอยู่ และข้อสอบก็มีความยากแฝงอยู่ในแต่ละข้อ พฤติกรรมการตอบข้อสอบจึงขึ้นอยู่กับผลต่างระหว่างความสามารถของผู้สอบ กับความยากของข้อสอบ ถ้าความสามารถของผู้สอบมีมากกว่าความยากของข้อสอบ ($\theta - b > 0$) โอกาสที่ผู้สอบจะตอบข้อสอบข้อนั้นถูกต้องมีมากกว่า 0.5 ถ้าความสามารถของผู้สอบมีค่าน้อยกว่าความยากของข้อสอบ ($\theta - b < 0$) โอกาสที่ผู้สอบจะตอบข้อสอบข้อนั้นถูกต้องจะมีค่าน้อยกว่า 0.5 หรือถ้าผู้สอบมีความสามารถพอ ๆ กับความยากของข้อสอบ ($\theta - b = 0$) โอกาสที่ผู้สอบจะตอบข้อสอบข้อนั้นถูกหรือผิดจะมีค่าเท่ากับ 0.5 สำหรับ ราชค์ โมเดลนี้ มีโมเดลทางคณิตศาสตร์ ดังนี้ (Wright and stone.1979, อ้างถึงใน ต่าย เชียงฉี,2534)

$$P_{i(\theta)} = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

เมื่อ $P_{i(\theta)}$ คือ โอกาสที่ผู้สอบซึ่งมีระดับความสามารถ (θ) ตอบข้อสอบข้อที่ i ได้ถูกต้อง

θ คือ ระดับความสามารถของผู้สอบ โดยทฤษฎีแล้วจะมีค่าอยู่ระหว่าง -∞ ถึง +∞ ในทางปฏิบัติจะมีค่าอยู่ระหว่าง -3.0 ถึง +3.0

b_i คือ ค่าความยากของข้อสอบข้อที่ iโดยทฤษฎีแล้ว b จะอยู่ระหว่าง -∞ ถึง +∞ แต่ในทางปฏิบัติ b จะมีค่าอยู่ระหว่าง -2.5 ถึ +2.5 โดยข้อสอบที่ b ใกล้ -2.5 จะเป็นข้อสอบที่ง่ายมาก b ใกล้ +2.5 จะเป็นข้อสอบที่ยากมาก

D คือ ค่าคงที่ เพื่อปรับโค้งแสดงลักษณะของข้อสอบ (ICC) ให้มีลักษณะใกล้เคียงกับโค้งแสดงลักษณะข้อสอบของนอร์มัล โอไจฟ์ โมเดลมากที่สุด ซึ่ง D จะมีค่าเท่ากับ 1.7

e คือ ค่าเอ็กซ์โปเนนเชี่ยล (Exponential) เป็นค่าคงที่ตัวหนึ่งซึ่งมีค่าเท่ากับ 2.7182818

2.3 โลจิสติค โมเดล (Logistic Model) พัฒนาขึ้นครั้งแรกโดย เบิร์นบอม (Hambleton and Swaninathan. 1985;citing Birnbaum. 1957, 1958a, 1958b, 1968) โดยพัฒนาแบบ 2 พารามิเตอร์ ซึ่งเป็นโมเดลที่มีโค้งแสดงลักษณะข้อสอบ (Item Characteristic Curve:ICC) เนื่องจากโมเดลนี้สามารถนำไปประยุกต์ใช้ได้ง่ายกว่าแบบนอร์มัล โอไจฟ์ โมเดล ทำให้เป็นที่นิยมมาก ต่อมา เบิร์นบอมได้พัฒนา โลจิสติค โมเดล เป็น 3 พารามิเตอร์โดยเพิ่มค่าการเดา (c) และได้ตัดค่าอำนาจจำแนก (a) ออกจากแบบ 2 พารามิเตอร์ จึงเหลือเพียงพารามิเตอร์เดียว คือค่าความยาก (b) โดยแต่และแบบมีโมเดลทางคณิตศาสตร์ดังนี้ คือ

2.3.1 โมเดล 1 พารามิเตอร์

$$P_{i(\theta)} = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

2.3.2 โมเดล 2 พารามิเตอร์

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

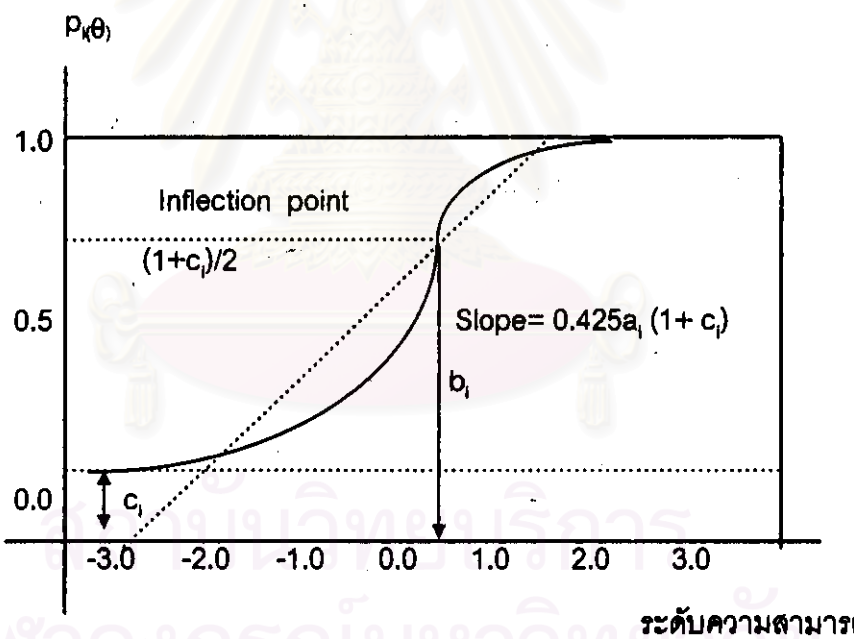
a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i โดยทฤษฎีแล้ว a จะมีค่าอยู่ระหว่าง -∞ ถึง +∞ ข้อสอบที่ใช้ได้ควรมีค่า a เป็นบวก โดยปกติจะไม่เกิน +2.5

2.3.3 โมเดล 3 พารามิเตอร์

$$P_{i(\theta)} = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

c_i คือ ค่าการเดาของข้อสอบข้อที่ i หมายถึง ค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำสุด ทำข้อสอบข้อนี้ได้ถูกต้อง ข้อสอบข้อที่มีค่าการเดาสูงจะไม่ดี ฉะนั้น ข้อสอบควรมีค่าการเดาไม่เกิน 0.3

จากโมเดลทางคณิตศาสตร์ของโมเดล 3 พารามิเตอร์นี้ เมื่อนำมาเขียนเป็นกราฟโค้งแสดงลักษณะข้อสอบ (Item Characteristic Curve : ICC) ได้ดังภาพที่ 1 (Lord, 1980; Hambleton and Swaminathan, 1985)



ภาพที่ 1 โค้งลักษณะข้อสอบ (ICC) แบบ 3 พารามิเตอร์ (Three-Parameters Logistic Model)

ในปัจจุบันทฤษฎีการตอบสนองข้อสอบได้พัฒนาขึ้นไปมากโดยบาร์ตันและลอร์ด (Barton and Lord) ได้เสนอโมเดลโลจิสติค แบบ 4 พารามิเตอร์ โดยเขาเชื่อว่า นักเรียนที่มีความสามารถสูงมากไม่จำเป็นว่าจะต้องตอบข้อสอบถูกต้องเสมอไปแม้ว่าข้อสอบจะง่ายหรือจะระมัดระวังในการตอบข้อสอบอย่างมากแล้วก็ตาม ก็ยังไม่สามารถใช้สารสนเทศที่มีอยู่ตอบข้อสอบได้ถูกต้องเสมอไป และโมเดลโลจิสติคแบบ 4 พารามิเตอร์นี้มีโมเดลทางคณิตศาสตร์ดังนี้

$$P_{i(\theta)} = c_i + (\gamma_i - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

เมื่อ γ_i คือ ค่าพารามิเตอร์ตัวแปรแฝงในข้อกระทง

โมเดลนี้ต่างจากโมเดล 3 พารามิเตอร์ เพราะ γ_i สมมติให้มีค่าต่ำกว่า 1 เล็กน้อย ดังนั้นโค้ง ICC เดิมมีค่าสูงสุดเท่ากับ 1 ในโมเดลนี้โค้ง ICC จะมีค่าต่ำกว่า 1 เล็กน้อย โมเดลนี้ให้ความสนใจศึกษาอยู่ในเชิงทฤษฎี ยังไม่ได้ศึกษาในทางปฏิบัติ หรือการนำไปใช้ (ศิริชัย กาญจนวาสี, 2538)

3. <u>ข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบข้อสอบ</u>

ทฤษฎีการตอบสนองข้อสอบ (IRT) มีข้อตกลงเบื้องต้น ดังนี้

1. แบบสอบวัดคุณลักษณะเดียว (Unidimensionality) สมมติว่าความสามารถของคน (Ability or trait) มีอยู่ทั้งหมด k อย่าง ซึ่งความสามารถแต่ละอย่างนั้น ต่างก็ส่งผลต่อการตอบข้อสอบ (Test Performance) ต่าง ๆ ที่รวมกันเป็นแบบทดสอบ ถ้าผลการตอบข้อสอบหรือคะแนนของผู้สอบนั้นสามารถอธิบายได้เพียงความสามารถเดียวก็เพียงพอแล้ว

2. การตอบข้อสอบของผู้สอบเป็นไปอย่างอิสระ โดยข้อสอบแต่ละข้อต้องมีอิสระต่อกันและผู้สอบแต่ละคนมีอิสระในการตอบข้อสอบแต่ละข้อ หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือการตอบข้อสอบข้อหนึ่งข้อใดถูกหรือผิดจะไม่ส่งผลต่อการตอบข้อสอบข้ออื่น ๆ ด้วย

3. โค้งคุณลักษณะข้อสอบสามารถใช้อธิบายพฤติกรรมการตอบสนองข้อสอบโดยแสดงถึงความสัมพันธ์ระหว่างโอกาสในการตอบข้อสอบข้อนั้นถูกต้องกับระดับความสามารถของผู้สอบที่วัดได้โดยใช้ชุดของข้อสอบ(Hambleton and Swaminathan,1991) จะเห็นได้ว่า โอกาสที่ ผู้สอบตอบข้อสอบถูกต้องจะขึ้นอยู่กับโค้งแสดงลักษณะข้อสอบ (ICC) ของแต่ละข้อ ซึ่งมีคุณสมบัติไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ดังนั้นจึงทำให้โอกาสการตอบข้อสอบถูกต้องในแต่ละข้อไม่แปรเปลี่ยนด้วย (ต่าย เชี่ยงฉี่, 2534)

4. ข้อสอบที่ใช้ต้องไม่เป็นข้อสอบประเภทความเร็ว (ศิริชัย กาญจนวาสี, 2538)

ตอนที่ 3 ระเบียบวิธีวิจัยและพัฒนาโปรแกรมคอมพิวเตอร์ทางการศึกษา

จากการศึกษาเอกสารและตำราต่าง ๆ ดังนี้ คือ "คอมพิวเตอร์และการพัฒนาระบบสารสนเทศ" (วันพร ปั้นเก่า และ ธนาวรรณ จันทรัตนไพบูลย์ , 2537) , "คอมพิวเตอร์พื้นฐาน" (วันชัย ริ้วไพบูลย์ และ มัณฑนา ปราการสมุทร , 2529) และ "การพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการจัดเก็บข้อสอบโดยใช้ไมโครคอมพิวเตอร์" (พลากร กรพิทักษ์ ,2532) ผู้วิจัยสรุปเป็นระเบียบวิธีการวิจัยและพัฒนาโปรแกรมคอมพิวเตอร์ทางการศึกษาได้ดังนี้

การวิจัยและการพัฒนา (RESEARCH & DEVELOPMENT) เป็นระเบียบวิธีการวิจัยหนึ่งที่แตกต่างจากระเบียบวิธีการวิจัยอื่น ๆ เป็นการวิจัยที่มุ่งเน้นการคิดค้นเพื่อประดิษฐ์สิ่งใหม่ ๆ ออกมาเพื่อใช้กับสังคม ในปัจจุบันเทคโนโลยีทางคอมพิวเตอร์เข้ามามีบทบาททางการศึกษามากขึ้น และนโยบายทางการศึกษาก็เน้นการปฏิรูปการศึกษาโดยมีนโยบายว่าโรงเรียนทุกโรงเรียนควรมีการเรียนการสอนคอมพิวเตอร์ให้กับนักเรียน มีการสร้างบทเรียนคอมพิวเตอร์ช่วยสอนออกมาจำหน่ายมาก และยังมีสื่อทางการศึกษาอีกมากที่ใช้คอมพิวเตอร์ ทำให้มีผู้สนใจศึกษาและทำการวิจัยเกี่ยวกับการใช้คอมพิวเตอร์มาช่วยในการเรียนการสอนมากขึ้น ซึ่งจากการศึกษางานวิจัยที่เกี่ยวข้องกับการสร้างโปรแกรมคอมพิวเตอร์เพื่อใช้ในการศึกษา และจากหนังสือการสร้างโปรแกรมที่เกี่ยวข้อง ผู้วิจัยจึงขอเสนอแนวทางในการใช้ระเบียบวิธีการวิจัยและการพัฒนาโปรแกรมคอมพิวเตอร์ทางการศึกษา ตามขั้นตอนดังต่อไปนี้

1. กำหนดปัญหา
2. วิเคราะห์ปัญหา
3. ออกแบบโปรแกรมคอมพิวเตอร์
4. เลือกภาษาคอมพิวเตอร์และเขียนโปรแกรม
5. ทดสอบโปรแกรม
6. จัดทำเอกสารประกอบการใช้โปรแกรม
7. ประเมินคุณภาพและบำรุงรักษาโปรแกรม

รายละเอียดการดำเนินงานของแต่ละขั้นตอนมีดังนี้

1. กำหนดปัญหา

เป็นกระบวนการหรือขั้นตอนที่ใช้ในการระบุให้ชัดเจนว่าผู้จะทำการวิจัย หรือผู้วิจัยมีความประสงค์ที่จะศึกษาอะไร เมื่อได้ปัญหาที่จะทำวิจัยแล้วผู้วิจัยต้องทำการกำหนดประเด็นที่จะศึกษาที่เกี่ยวข้องกับปัญหาวิจัยให้ชัดเจน (พลากร กรพิทักษ์, 2532)

2. วิเคราะห์ปัญหา

เป็นขั้นตอนที่จะทำการศึกษาถึงรายละเอียดต่าง ๆ ของงานที่สั่งให้คอมพิวเตอร์ทำ ตั้งแต่การทำความเข้าใจกับงานว่าอะไรคือสิ่งที่ต้องการให้คอมพิวเตอร์ทำจริง ๆ ผลที่ได้จากคอมพิวเตอร์เป็นอะไร และถ้าต้องการผลที่ได้นั้นจะต้องมีขั้นตอนการคำนวณอย่างไร มีสูตรอะไรบ้าง ซึ่งส่วนที่สำคัญที่จะต้องวิเคราะห์และแจกแจงมีดังนี้

2.1 สิ่งที่ต้องการ เป็นการพิจารณาอย่างกว้าง ๆ ถึงงานที่ต้องการคอมพิวเตอร์ทำ เช่น ต้องการให้คำนวณคะแนนเฉลี่ย ต้องการให้พิมพ์ค่าการคำนวณที่ได้ ต้องการให้แสดงข้อคำถาม ต้องการให้สุ่มตัวเลข ต้องการให้บันทึกค่าที่คำนวณได้ ซึ่งงานแต่ละชนิดที่ต้องการให้คอมพิวเตอร์ทำควรจะเขียนเป็นข้อ ๆ ให้ชัดเจน การพิจารณาสิ่งที่ต้องการอาจดูได้จากคำสั่งหรือปัญหาที่จะทำว่าต้องการให้คอมพิวเตอร์ทำอะไรบ้าง

2.2 ผลลัพธ์ที่ต้องการแสดง เป็นการวิเคราะห์ถึงลักษณะของการรายงานหรือแบบของผลลัพธ์ที่ต้องการให้คอมพิวเตอร์แสดงผลลัพธ์ออกมาว่าควรจะมีลักษณะอย่างไร มีรายละเอียดที่ต้องการในรายงานมากน้อยเพียงใด ปัญหาหรืองานบางอย่างอาจไม่สามารถกำหนดลักษณะของรายงานออกมาได้ชัดเจนว่าต้องการรายงานอย่างไร มีรายละเอียดอย่างไร ผู้วิจัยจะต้องทำการศึกษาว่าระบบเดิมหรือระบบที่สร้างมีรูปแบบใดบ้างที่ต้องการออกรายงาน เพื่อความสะดวกของผู้นำผลลัพธ์ไปใช้ การวิเคราะห์ผลลัพธ์ หรือรายงานนั้นเป็นส่วนสำคัญและจะต้องพิจารณาอย่างละเอียดรอบคอบ เพราะการวิเคราะห์รายงานได้ดีนั้น จะทำให้เราทราบถึงจุดหมายที่ต้องการให้คอมพิวเตอร์ทำ และจะได้หาวิธีที่นำไปสู่จุดมุ่งหมายนั้นได้ ซึ่งเป็นการกำหนดขอบเขตของงานที่จะทำนั้นเอง ในการวิเคราะห์ผลลัพธ์อาจวางรูปแบบออกมาอย่างคร่าว ๆ เหมือนกับที่จะให้คอมพิวเตอร์แสดงออกมา

2.3 ข้อมูลที่ต้องนำเข้า เป็นขั้นตอนที่หลังจากที่ได้ลักษณะของงานที่แน่นอนแล้วก็มาวิเคราะห์ต่อว่าถ้าผลลัพธ์ออกมาอย่างนั้นแล้วข้อมูลที่จะนำเข้าควรมีอะไรบ้าง มีลักษณะหรือรูปแบบอย่างไร เพื่อให้ได้ผลลัพธ์ตามที่ต้องการ ซึ่งการวิเคราะห์ลักษณะข้อมูลนี้นอกจากจะดูจากรายงานแล้ว บางครั้งต้องคำนึงถึงขั้นตอนการประมวลผลเพื่อให้ได้รายงานตามที่ต้องการด้วย

เพราะค่าบางค่าในรายงานได้มาจากการคำนวณไม่ใช่เป็นค่าที่ได้มาจากข้อมูล ดังนั้นการวิเคราะห์ข้อมูลจะวิเคราะห์ว่าข้อมูลที่จะทำให้ได้ผลลัพธ์นั้น ๆ ได้มาจากไหน เป็นค่าคงที่หรือเป็นค่าที่ได้มาจากการคำนวณ

2.4 ตัวแปรที่ใช้เป็นการกำหนดชื่อแทนความหมายของข้อมูลต่าง ๆ เพื่อความสะดวกในการอ้างอิงถึงข้อมูลนั้น และรวมไปถึงการเขียนโปรแกรมด้วย การตั้งชื่อตัวแปรที่ใช้ในงานหรือปัญหาใด ๆ ควรตั้งชื่อให้มีความหมาย และเกี่ยวข้องกับข้อมูลถ้าเป็นไปได้ควรตั้งอยู่ภายใต้กฎเกณฑ์ของภาษาคอมพิวเตอร์ที่ใช้ในการเขียนโปรแกรม โดยทั่ว ๆ ไปการตั้งชื่อตัวแปรจะพิจารณาความหมายของข้อมูล ว่าตรงกับคำใดในภาษาอังกฤษ แล้วนำมาดัดแปลง หรือย่อให้เข้ากับหลักเกณฑ์ของภาษาคอมพิวเตอร์ที่ใช้

2.5 วิธีการประมวลผล เป็นการบอกขั้นตอนของวิธีการหรือการคำนวณเพื่อให้ได้ผลลัพธ์ตามที่ต้องการ โดยเริ่มตั้งแต่การสั่งให้เครื่องรับข้อมูลแล้วนำไปประมวลผลแสดงผลลัพธ์ออกมาขั้นตอนนี้ต้องแสดงการทำงานที่ต่อเนื่องตามลำดับจึงต้องจัดลำดับก่อน - หลังให้ถูกต้อง ในขั้นตอนนี้ถ้ายิ่งกระทำให้ละเอียดก็จะช่วยให้เขียนโปรแกรมได้ง่ายขึ้น

3. การออกแบบโปรแกรมคอมพิวเตอร์

ขั้นตอนนี้เป็นการออกแบบรายละเอียดขั้นตอนสำหรับโปรแกรม การกำหนดลำดับและความครบถ้วนของขั้นตอนของโปรแกรมทั้งโปรแกรมให้ถูกต้อง โดยไม่มีการออกแบบไว้ล่วงหน้านั้นย่อมเป็นการเสี่ยงต่อการผิดพลาดในการเขียนโปรแกรมได้ง่ายขึ้น ดังนั้น งานในส่วนนี้เป็นการออกแบบโปรแกรมว่าควรจะต้องมีขั้นตอนตั้งแต่ต้นจนจบอะไรบ้าง มีลำดับขั้นตอนก่อน - หลังอย่างไร วิธีการในการออกแบบขั้นตอนนี้ได้มีผู้คิดไว้หลายวิธีด้วยกัน เช่น การเขียนเป็นอัลกอริทึม ซึ่งจะอธิบายการทำงานแต่ละขั้นตอนด้วยคำหรือประโยคที่สื่อความหมายให้คนเข้าใจได้ง่ายหรืออาจจะมีรูปประโยคคล้ายกับคำสั่งของภาษาคอมพิวเตอร์ ซึ่งเรียกว่า Pseudo code หรือการเขียนผังงาน (Flow chart) ซึ่งใช้สัญลักษณ์ที่เป็นรูปแทนความหมายต่าง ๆ และเพื่อให้การอ้างถึงข้อมูลและผลลัพธ์ในขั้นตอนต่าง ๆ ที่ออกแบบสะดวกและถูกต้องในการเขียนขั้นตอนของโปรแกรม จึงควรจะกำหนดชื่อหรือรายการข้อมูลและผลลัพธ์ต่าง ๆ ขึ้นด้วย ดังนั้นการออกแบบขั้นตอนสำหรับโปรแกรมขั้นก่อนการเขียนโปรแกรมนี้จะช่วยให้ขั้นตอนในโปรแกรมถูกต้องรัดกุม และที่สำคัญคือ เป็นการแยกความยุ่งยากในส่วนของขั้นตอนภายในโปรแกรมออกจากงานเขียนคำสั่ง ซึ่งแต่ละคำสั่งเหล่านี้จะต้องเขียนให้ถูกต้องตามกฎเกณฑ์ของภาษาคอมพิวเตอร์และนอกจากนี้ยังได้ประโยชน์อีกต่อหนึ่งคือการเก็บผลการออกแบบนี้ไว้จะช่วยในการทำความเข้าใจขั้นตอนของโปรแกรมในภายหลังได้สะดวกกว่าการทำความเข้าใจจากตัวโปรแกรมโดยตรงมาก

4. เลือกภาษาคอมพิวเตอร์และเขียนโปรแกรม

เป็นการเปลี่ยนขั้นตอนการทำงานต่าง ๆ ที่แสดงไว้ในผังงานมาเขียนให้อยู่ในรูปของภาษาคอมพิวเตอร์ภาษาใดภาษาหนึ่งตามต้องการ การจะเลือกภาษาคอมพิวเตอร์ภาษาใดนั้นขึ้นอยู่กับลักษณะและประเภทของงานนั้น ๆ ว่าเป็นงานที่ควรใช้ภาษาใด เช่น ถ้าเป็นงานด้านวิทยาศาสตร์และคณิตศาสตร์ควรใช้ภาษาฟอร์แทน ถ้าเป็นงานด้านธุรกิจ ควรใช้ภาษาโคบอล ถ้าเป็นงานประเภทที่ต้องการพิมพ์รายงานมาก ๆ หรือคล้ายกับการกรอกแบบฟอร์มควรใช้ภาษาอาร์พีจี แต่ไม่ว่าจะเลือกภาษาใดก็ตาม โดยทั่วไปแล้วลักษณะของโปรแกรมที่ดีควรเป็นโปรแกรมที่แบ่งเป็นส่วน ๆ ได้ชัดเจน และในแต่ละส่วนจะเขียนด้วยคำสั่งที่ไม่ยุ่งยากมากนัก นอกจากนี้ยังควรคำนึงถึงขีดจำกัดของเครื่องและตัวแปรภาษาของเครื่องคอมพิวเตอร์ (Compiler) ที่ใช้ และข้อสำคัญในการเลือกใช้โปรแกรมคือ ความถนัดและความชำนาญของผู้เขียนโปรแกรมว่าสามารถใช้ภาษาที่เลือกนั้นได้หรือไม่ และในการเขียนคำสั่งด้วยภาษาคอมพิวเตอร์นั้นยังต้องคำนึงถึงกฎเกณฑ์และหลักของภาษาที่ใช้ให้ถูกต้อง เพราะถ้ามีข้อผิดพลาดเกิดขึ้น ซึ่งเรียกว่า Syntax error โปรแกรมแปลภาษาจะไม่สามารถแปลความหมายของคำสั่งนั้นได้ ขั้นตอนการปฏิบัติการโปรแกรมก็เกิดขึ้นไม่ได้ โปรแกรมแปลภาษาส่วนมากจะรายงาน หรือแสดงข้อความออกมาให้ทราบว่าเกิดการผิดพลาดตรงคำสั่งไหนเพราะอะไร เพื่อจะช่วยให้การแก้ไขข้อผิดพลาดนั้นสะดวกยิ่งขึ้น

5.ทดสอบโปรแกรม

เมื่อโปรแกรมที่เขียนผ่านการแปลภาษาและได้ผลลัพธ์เป็น Object code แล้ว ควรมีการทดสอบโปรแกรมก่อนที่จะทำการ Link ให้เป็น Execute file เพราะการทำงานของเครื่องคอมพิวเตอร์ตามคำสั่งของโปรแกรมนั้นมิได้หมายความว่าจะได้ผลลัพธ์ถูกต้องเสมอไปอาจจะมีขั้นตอนที่ไม่ถูกต้องทำให้ได้ผลลัพธ์ที่ผิดได้ ฉะนั้นเพื่อให้ได้โปรแกรมที่ใช้งานได้ถูกต้องตามจุดประสงค์ที่ตั้งไว้และมีความเชื่อถือได้ก็ควรที่จะต้องทดสอบโปรแกรมเสียก่อน วิธีการทดสอบโปรแกรมนี้กระทำได้โดยการสั่งให้เครื่องทำงานตามคำสั่งในโปรแกรมถ้าโปรแกรมนั้นมีการกำหนดให้เครื่องรับข้อมูลเข้าไปประมวลผลก็จะต้องนำข้อมูลตัวอย่างหรือข้อมูลจริงส่งเข้าไป แล้วนำผลลัพธ์ที่ได้ไปตรวจสอบกับผลลัพธ์ที่มีความถูกต้องที่ได้มาจากวิธีการอื่น เช่น ส่งข้อมูลเข้าไปประมวลหลาย ๆ ชุด ถ้าเปรียบเทียบแล้วได้ผลลัพธ์ตรงกัน จึงจะยอมรับว่าโปรแกรมนั้นใช้งานได้ แต่ถ้าผลลัพธ์ไม่ตรงกันจะต้องพิจารณาว่า ความผิดพลาดนั้นเกิดขึ้นที่ตรงไหนที่ข้อมูลหรือที่ตัวโปรแกรมถ้าผิดที่ข้อมูลก็แก้ไขข้อมูลให้ถูกต้องแล้วนำเข้าไปให้เครื่องคำนวณใหม่ถ้าผิดที่

ตัวโปรแกรมก็ต้องดูว่าผิดที่ขั้นตอนไหนและทำการแก้ไขโปรแกรมแล้วทำการทดสอบโปรแกรมใหม่อีกครั้ง ข้อผิดพลาดที่เกิดจากโปรแกรม สามารถแบ่งได้ 3 แบบ คือ

1. การให้รหัสผิดรูปแบบภาษา (Syntax error) ภาษาคอมพิวเตอร์แต่ละภาษามีกฎเกณฑ์การให้รหัสเฉพาะตัวที่แน่นอนในการเขียนรหัสโปรแกรม ถ้าให้รหัสผิดกฎเกณฑ์ของภาษาเครื่องก็จะให้ข่าวสารเกี่ยวกับข้อมูลผิดพลาดออกมาเพื่อให้ผู้เขียนโปรแกรมรู้ตำแหน่งที่ผิดได้ง่ายขึ้น

2. ข้อผิดพลาดที่เกิดขึ้นกับการใช้งานแต่ละครั้ง (Run - time - error)โปรแกรมที่ใส่รหัสถูกต้องไม่จำเป็นว่าจะทำงานได้สำเร็จเสมอไป ทั้งนี้อาจจะมีข้อผิดพลาดที่เกิดจากการวิ่งโปรแกรม (Running program) ได้ เช่น ในการคำนวณมีการหารค่าใดค่าหนึ่งที่ด้วยตัวแปรที่มีค่าเป็นศูนย์ หรือกรณีที่ตัวแปรที่ให้หารากที่สองมีค่าเป็นลบ ข้อผิดพลาดเช่นนี้โปรแกรมก็จะให้ข่าวสารเกี่ยวกับการผิดพลาดออกมาเพื่อให้ผู้เขียนโปรแกรมรู้และแก้ไขได้ถูกต้อง

3. ข้อผิดพลาดทางตรรก (Logical error) ข้อผิดพลาดทางตรรกเกิดขึ้นเนื่องจากการกำหนดลักษณะของงานหรือการกำหนดวิธีการและขั้นตอนการทำงานผิดหรือไม่ตรงกับจุดประสงค์ของงาน ข้อผิดพลาดนี้ถ้าเกิดขึ้นเครื่องจะไม่บอกข่าวสารมาให้ทำให้ผู้เขียนโปรแกรมแก้ไขข้อผิดพลาดได้ยาก

6. จัดทำเอกสารประกอบโปรแกรม

เนื่องจากการใช้คอมพิวเตอร์ในการประมวลผลข้อมูลต่าง ๆ ส่วนมากจะเป็นการทำงานต่อเนื่องและใช้เวลายาวนาน ฉะนั้นโปรแกรมคอมพิวเตอร์ที่พัฒนาขึ้นก็จะถูกใช้ไปตามกำหนดระยะเวลานั้นด้วย และเมื่อใช้ประยะหนึ่งอาจมีการเปลี่ยนแปลง ทำให้โปรแกรมนั้นอาจจะไม่เหมาะสมกับการใช้งาน ซึ่งจะต้องมีการพัฒนาโปรแกรมขึ้นใหม่ หรืออาจจะแก้ไขโปรแกรมให้มีความเหมาะสมกับการใช้งานมากขึ้น ถ้าได้มีการทำเอกสารประกอบการพัฒนาโปรแกรมไว้ ก็จะเป็นการสะดวกในการใช้เป็นแนวทางในการแก้ไขโปรแกรม ในการทำเอกสารนั้นควรเริ่มทำและรวบรวมตั้งแต่ขั้นตอนแรกของการพัฒนาโปรแกรมมาตามลำดับ มิฉะนั้นอาจทำให้ลืมได้โดยเฉพาะการพัฒนาโปรแกรมที่มีขนาดใหญ่ซึ่งใช้เวลานาน เอกสารสำหรับโปรแกรมเขียนได้ 2 แบบ ดังนี้

1. เอกสารสำหรับผู้ใช้โปรแกรม

ผู้ใช้โปรแกรมอาจไม่มีความรู้เกี่ยวกับโปรแกรมเลย ดังนั้นเนื้อหาของเอกสารควรประกอบด้วย

1.1 แสดงรายละเอียดขอบเขตความสามารถของโปรแกรม เช่น โปรแกรมนี้จะสามารถทำอะไรได้บ้าง และมีข้อกำหนดอะไร

1.2 แสดงรายละเอียดของข้อมูลที่ใช้กับโปรแกรม

1.3 แสดงรายละเอียดของผลลัพธ์ที่ได้จากโปรแกรม

1.4 แสดงรายละเอียดของคำสั่งที่จะใช้ เพื่อให้โปรแกรมเริ่มต้นทำงาน

1.5 แสดงรายละเอียดของการโต้ตอบระหว่างผู้ใช้กับโปรแกรม

1.6 แสดงความหมายต่าง ๆ ของข่าวสารที่โปรแกรมแสดงออกมา

2. เอกสารแสดงเทคนิคที่ใช้ในโปรแกรม

เอกสารแสดงเทคนิคในการใช้โปรแกรมจะมีประโยชน์มากสำหรับผู้เขียนโปรแกรมผู้อื่นที่จะดัดแปลง ปรับปรุงแก้ไขโปรแกรมเพื่อใช้งานในลักษณะที่ใกล้เคียงกัน ฉะนั้นเอกสารฉบับนี้ควรประกอบด้วยเนื้อหาดังนี้

2.1 รายละเอียดของส่วนอธิบายต่าง ๆ ที่ใช้ในโปรแกรม ส่วนนี้จะแสดงให้ทราบว่าโปรแกรมนี้มีขั้นตอนการทำงานอะไรบ้าง

2.2 รายละเอียดเทคนิคต่าง ๆ ที่ใช้ในแต่ละส่วนของโปรแกรม

เอกสารประกอบโปรแกรมอาจอนุโลมให้เขียนเป็นเอกสารฉบับเดียวได้ โดยรวมข้อมูลเกี่ยวกับการใช้งาน และข้อมูลทางด้านเทคนิคโปรแกรมเข้าด้วยกัน ในลักษณะนี้เอกสารประกอบโปรแกรมควรประกอบด้วย เนื้อหาของปัญหา การวิเคราะห์ปัญหา สูตรหรือทฤษฎีที่ใช้ในขั้นตอนต่าง ๆ ของโปรแกรม รูปแบบของข้อมูลและผลลัพธ์ เนื้อหาโปรแกรม รายละเอียดการใช้และข้อจำกัดของโปรแกรม ตลอดจนตัวอย่างของการทำงานของโปรแกรม

7. ประเมินคุณภาพของโปรแกรม

การประเมินคุณภาพของโปรแกรมผู้วิจัยทำการประเมิน 2 แนวทาง คือ

7.1 ประเมินคุณภาพของโปรแกรมโดยผู้พัฒนาโปรแกรม เป็นการประเมินระบบการทำงานภายในโปรแกรม โดยประเมินด้านต่าง ๆ ดังนี้

7.1.1 ความทนทานต่อความผิดพลาดของผู้ใช้โปรแกรม โดยการทดลองทำให้เกิดข้อผิดพลาด แล้วดูว่าโปรแกรมมีวิธีป้องกันข้อผิดพลาดหรือไม่ อย่างไร

7.1.2 ความเชื่อถือได้ของโปรแกรม โดยการตรวจสอบว่าโปรแกรมที่สร้างขึ้นเมื่อใช้ครั้งแรกและครั้งต่อไปนั้นมีความเชื่อถือได้หรือไม่

7.1.3 ความเร็วของโปรแกรม เป็นการประเมินโปรแกรมในเรื่องความเร็วของโปรแกรม

7.1.4 ความสามารถของโปรแกรมที่สามารถบันทึกข้อมูลของผู้ใช้ในการใช้โปรแกรมในแต่ละด้าน เป็นการวิเคราะห์ความสามารถของโปรแกรมที่สามารถจัดเก็บข้อมูลพื้นฐานของผู้ใช้โปรแกรมโดยอัตโนมัติ เช่น ข้อมูล เลขประจำตัว ระดับชั้น วิชาที่เรียน

7.1.5 ความถูกต้องในการสั่งงานตามต้องการ เป็นการประเมินโปรแกรมว่าสามารถทำงานได้ถูกต้องตามต้องการหรือไม่

7.2 ประเมินคุณภาพของโปรแกรมโดยผู้ใช้โปรแกรม เป็นการประเมินโปรแกรมในเรื่องผลย้อนกลับของผู้ใช้โปรแกรม ในด้านต่าง ๆ ดังนี้

7.2.1 ความชัดเจนและความสอดคล้องของคู่มือการใช้โปรแกรม

7.2.2 รูปแบบการใช้โปรแกรม เป็นการประเมินโปรแกรมในด้านการรับข้อมูล การดำเนินงานของโปรแกรม ความรู้พื้นฐานของผู้ใช้โปรแกรม การแสดงผล และขั้นตอนการใช้งาน

7.2.3 ประสิทธิภาพและประโยชน์ของโปรแกรมโดยส่วนรวม เป็นการประเมินโปรแกรมในด้านประสิทธิภาพ ผลที่ได้รับ และประโยชน์ของโปรแกรม

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

จากการศึกษางานวิจัยที่เกี่ยวข้องผู้วิจัยได้แบ่งงานวิจัยที่เกี่ยวข้องเป็น 2 ส่วนคือ

<u>1.งานวิจัยที่เกี่ยวกับการทดสอบแบบปรับเหมาะกับระดับความสามารถ</u>

งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบ มีผู้ศึกษาไว้ดังนี้

นันทิยา พึ่งคำ (2531) ได้ศึกษาเปรียบเทียบคุณภาพของการทดสอบ ซีเอที กับแบบสอบประเพณีนิยม ในการวัดความสามารถทางคำศัพท์ภาษาอังกฤษโดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 120 ผลการศึกษาพบว่า

1. การทดสอบ ซีเอทีกับแบบสอบประเพณีนิยม มีความตรงเชิงเกณฑ์สัมพันธ์ไม่แตกต่างกัน แต่การทดสอบ ซีเอที ใช้จำนวนข้อสอบน้อยลงครึ่งหนึ่ง

2. ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบจากการทดสอบ ซีเอที มีค่าต่ำกว่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของ ผู้สอบจากแบบสอบประเพณีนิยม อย่างมีนัยสำคัญทางสถิติ

เสรี ปรมขวลิตโรจน์ (2531) ศึกษาความตรงในการจำแนกความรอบรู้และการประมาณค่าความสามารถโดยใช้แบบสอบรูปแบบเฟล็กซิเลเวล จำนวน 2 ฉบับ และรูปแบบปิรามิดแบบขนาดขั้นแปรผันจำนวน 2 ฉบับ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 5,987 คน ผลการศึกษาพบว่า ความตรงในการจำแนกความรอบรู้เมื่อใช้ข้อสอบที่สร้างขึ้น

I 17719074O

เป็นเกณฑ์ ปรากฏว่าข้อสอบแบบเฟล็กซิเลเวลและแบบปิรามิดแบบขนาดขั้นแปรผันทุกฉบับ มีความตรงในการจำแนกความรอบรู้ได้อย่างมีนัยสำคัญทางสถิติและเมื่อเปรียบเทียบค่าความตรงแต่ละแบบ พบว่า แบบสอบแบบเฟล็กซิเลเวลทั้งสองฉบับ มีความตรงในการจำแนกความรอบรู้ได้สูงกว่าแบบสอบรูปแบบปิรามิดอยู่หนึ่งฉบับอย่างมีนัยสำคัญ

ตรึงใจ พูลผลอำนวย (2534) ได้พัฒนาแบบสอบเฉพาะบุคคล โดยใช้รูปแบบแยกทางคงที่รูปแบบปิรามิดขนาดขั้นคงที่ 10 ขั้น ในวิชาวิทยาศาสตร์ โดยใช้คอมพิวเตอร์ในการทดสอบ ผลการวิจัยพบว่าแบบสอบมีความตรงตามสภาพ โดยค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.7517 และพบว่าแบบสอบที่ทดสอบโดยใช้คอมพิวเตอร์และผู้สอบแต่ละคนทำ 10 ข้อ มีความสัมพันธ์กับแบบสอบวัดผลสัมฤทธิ์วิชาวิทยาศาสตร์ทั้ง 55 ข้อ ซึ่งใช้การเขียนตอบ โดย มีค่าสัมประสิทธิ์สหสัมพันธ์ เท่ากับ 0.7818 และความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของแบบสอบเฉพาะบุคคลเท่ากับ 0.443

ระพีพรรณ ศรีวิเชียร (2534) ได้พัฒนาแบบสอบตามระดับความสามารถ ในวิชาคณิตศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ผลการวิจัยพบว่า

1. แบบสอบตามระดับความสามารถมีความตรงตามเนื้อหา โดยพิจารณาจากค่าเฉลี่ยของคะแนนการตัดสินความสอดคล้องของผู้ทรงคุณวุฒิ ผ่านเกณฑ์ที่ยอมรับ คือ 0.5

2. แบบสอบตามระดับความสามารถมีความตรงตามเกณฑ์ เมื่อใช้เกรดเฉลี่ยวิชาคณิตศาสตร์เป็นเกณฑ์ ได้ค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.5874 และมีนัยสำคัญทางสถิติที่ระดับ 0.001

รังสรรค์ มณีเล็ก (2539) ได้ศึกษาผลของตัวแปรบางตัวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อสอบที่ใช้ในการทดสอบที่มีความเหมาะสมกับผู้สอบโดยใช้คอมพิวเตอร์ โดยตัวแปรที่ศึกษาคือ ลักษณะของข้อสอบที่ต่างกัน กฎการเลือกข้อสอบต่างกัน วิธีการประมาณค่าความสามารถของผู้สอบต่างกัน เกณฑ์การยุติการทดสอบต่างกัน และความสามารถของผู้ทดสอบต่างกัน ซึ่งผลการวิจัยพบว่า ลักษณะของข้อสอบต่างกันจะมีความตรงเชิงสภาพไม่ต่างกัน การเลือกข้อสอบโดยวิธี $I_{(\theta)max}$ จะมีความตรงเชิงสภาพสูงสุด การประมาณค่าความสามารถของผู้สอบโดยวิธีของเบส์ มีความตรงเชิงสภาพสูงสุด การยุติการทดสอบเมื่อค่าความคลาดเคลื่อนในการประมาณค่าความสามารถต่ำกว่า 0.3 จะมีความตรงเชิงสภาพสูงสุดและผู้ทดสอบที่มีความสามารถสูงจะทำให้ผลการทดสอบมีความตรงเชิงสภาพสูงสุด จากการวิจัยนี้ถ้าไม่พิจารณาปฏิ

สัมพันธ์ระหว่างตัวแปรแล้วพบว่าตัวแปรอิสระทั้ง 5 ตัว ต่างก็ส่งผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบอย่างมีนัยสำคัญ

ลอร์ด (Lord, 1971) ได้ทำการประเมินกระบวนการทดสอบด้วยแบบทดสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบ โดยใช้คอมพิวเตอร์จำลองสถานการณ์ ด้วยวิธีแบบจำลองนอร์มอลโอไจฟ์ในการประมาณค่าความสามารถของผู้สอบ (θ) และคำนวณค่าพารามิเตอร์ของข้อสอบโดยใช้เกณฑ์ในการคัดเลือกข้อสอบ คือ ค่าอำนาจจำแนกเท่ากับ .05 ศึกษาเปรียบเทียบกับแบบสอบประเพณีนิยมที่คัดเลือกข้อสอบด้วยเกณฑ์เดียวกัน โดยใช้แบบสอบประเพณีนิยมจำนวน 60 ข้อ และแบบสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบ จำนวน 119 ข้อ แต่ผู้สอบตอบเพียง 60 ข้อ ผลการศึกษาพบว่า เมื่อขนาดของช่วงห่างระหว่างความยากของข้อสอบใหญ่ขึ้น จะทำให้สารสนเทศของผู้สอบลดลงทุกระดับความสามารถ และจะมีประสิทธิภาพสูงสุดเมื่อขนาดของช่วงห่างระหว่างค่าความยากอยู่ระหว่าง 0.033 และ 0.067 และจะไม่มีประสิทธิภาพเมื่อมีค่าใกล้เคียง 0.1 นอกจากนี้แบบสอบปรับเหมาะกับระดับความสามารถของ ผู้สอบจะให้สารสนเทศสูงกว่าแบบสอบประเพณีนิยม ในกลุ่มผู้สอบที่มีระดับความสามารถทั้งสูงและต่ำ แต่เมื่อผู้สอบมีระดับความสามารถใกล้เคียงกับ 0 หรือมีระดับความสามารถปานกลาง จะให้สารสนเทศต่ำกว่าแบบสอบประเพณีนิยม และเมื่อไม่มีการเดาในการทดสอบ ต้องใช้แบบสอบประเพณีนิยม จำนวน 58,60,69 และ 86 ข้อ ที่ระดับความสามารถเท่ากับ 0,+1,+2 และ +3 และเมื่อการเดาเป็น 0.2 จะต้องใช้แบบสอบประเพณีนิยมจำนวน 58,60,70,83 และ 114 ข้อ ที่ระดับความสามารถ 0,+1,+2 หรือ +2.25,3 และ -3

ผลการศึกษาของลอร์ดพบว่า แบบสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบจะมีประสิทธิภาพสูงเมื่อผู้สอบมีความสามารถสูงหรือต่ำ แต่ที่ระดับความสามารถปานกลาง แบบสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบจะให้สารสนเทศต่ำกว่าแบบสอบประเพณีนิยมและเมื่อใช้ข้อสอบจำนวน 60 ข้อเท่ากัน แบบสอบปรับเหมาะกับระดับความสามารถของผู้สอบจะมีประสิทธิภาพสูงกว่าแบบสอบประเพณีนิยม และใช้เวลาน้อยกว่าแบบสอบประเพณีนิยม

Dodd and Koch (1993)ได้ศึกษาผลกระทบของลักษณะของกลุ่มข้อสอบและความแตกต่างของกฎการยุติการทดสอบ โดยใช้ Partial Credit Model โดยการสมมติข้อมูล พบว่า การยุติโดยการดูค่าความคลาดเคลื่อนมาตรฐาน จะให้ผลดีกว่าการยุติโดยดูค่าสารสนเทศของข้อสอบ และยังพบว่ากลุ่มข้อสอบที่มีข้อสอบน้อยที่สุดที่สามารถใช้ได้คือ 30 ข้อ และควรมีค่าความยากกลาง ๆ

Dodd, Ayala and Koch (1995) ได้ศึกษาการใช้ข้อสอบที่ให้คะแนนแบบ Polytomous ใน CAT โดยการเปรียบเทียบ GRM กับ NRM พบว่า GRM และ NRM มีความเหมาะสมที่จะนำมาใช้ต่างกัน เช่น ถ้าข้อคำถามที่ไม่เรียงลำดับควรใช้ NRM หรือคลังข้อสอบ 30 ก็สามารถประมาณค่าได้เมื่อใช้ GRM โดยสรุปพบว่าการให้คะแนนแบบ Polytomous มีความเหมาะสมที่จะใช้ใน CAT

2.งานวิจัยที่เกี่ยวกับการพัฒนาโปรแกรม

งานวิจัยที่เกี่ยวข้องกับการพัฒนาโปรแกรม มีผู้ศึกษาไว้ดังนี้

พลากร กรพิทักษ์ (2532) ได้พัฒนาโปรแกรมคอมพิวเตอร์สำหรับการจัดเก็บข้อสอบ จากการประเมินสรุปได้ว่าโปรแกรมที่พัฒนาขึ้นมีความสามารถในการบันทึกข้อมูล มีความถูกต้องในการสั่งงานตามต้องการ มีความเชื่อถือได้ และผู้ใช้โปรแกรมมีความเห็นว่าโปรแกรมที่พัฒนาขึ้นนี้ง่ายต่อการใช้งาน และสามารถนำมาใช้แทนระบบการจัดเก็บข้อสอบแบบกระดาษได้เป็นอย่างดี

ศรีไพร ศักดิ์รุ่งพงศากุล (2533) ได้พัฒนาโปรแกรมไมโครคอมพิวเตอร์ เพื่องานประเมินผลการศึกษาของโรงเรียนอัสสัมชัญคอนแวนต์ โปรแกรมพัฒนาขึ้นโดยใช้คำสั่งของโปรแกรม dBASE III PLUS โปรแกรมที่พัฒนาขึ้นประกอบด้วย การสร้างแฟ้มข้อมูล การใส่หรือแก้ไขข้อมูล การประมวลผล และการรายงานผล จากผลการประเมินโปรแกรม พบว่า โปรแกรมมีความถูกต้องในการสั่งงานตามต้องการ มีความเชื่อถือได้ในการนำไปใช้งาน มีความเร็วในการประมวลผลข้อมูลและมีประสิทธิภาพในการทำงาน และยังสามารถทำงานต่อไปได้เมื่อผู้ใช้โปรแกรมสั่งการผิดพลาด

ชัยณรงค์ วิริยะศรีสุวรรณ (2534) ได้พัฒนาโปรแกรมไมโครคอมพิวเตอร์เพื่อวิเคราะห์แบบแผนการตอบข้อสอบโดยใช้สูตรดัชนีชี้เตือนที่ปรับปรุงโดยฮาร์นิสและลินน์ โปรแกรมพัฒนาขึ้นโดยใช้ภาษา FOXPRO โปรแกรมที่ได้สามารถวิเคราะห์คำตอบจากการสอบที่ใช้แบบสอบชนิดเลือกตอบ ตัวเลือกไม่เกิน 5 ตัวเลือก โดยแสดงผลทั้งบนจอภาพและทางเครื่องพิมพ์ได้ และจากการประเมินโปรแกรมพบว่า โปรแกรมสามารถทำงานตามที่สั่งได้อย่างถูกต้องและมีประสิทธิภาพ

ทวีศิลป์ กุลนภาดล (2539) พัฒนาโปรแกรมคอมพิวเตอร์เพื่อกำหนดขนาดตัวอย่างและอำนาจทางสถิติ สำหรับทดสอบความแตกต่างของค่ามัชฌิมเลขคณิต โดยโปรแกรมที่พัฒนาขึ้นใช้ภาษา Foxpro veasion 2.5 โปรแกรมสามารถกำหนดขนาดตัวอย่างและอำนาจทางสถิติ สำหรับสถิติทดสอบ t-test 1 กลุ่ม , 2 กลุ่มอิสระ, 2 กลุ่มสัมพันธ์และสถิติทดสอบ F-test 3 กลุ่มถึง 10 กลุ่ม จากการประเมินโปรแกรมพบว่า คู่มือมีความชัดเจน โปรแกรมมีประสิทธิภาพและเป็นประโยชน์ในการกำหนดขนาดตัวอย่างและอำนาจทางสถิติ

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า การทดสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบให้ค่าสารสนเทศในการทดสอบดีกว่าการทดสอบแบบประเพณีนิยม ดังที่ ลอร์ด(Lord, 1971)และนันทิยา พึ่งคำ ได้ศึกษาไว้ นอกจากนี้ยังพบว่าการศึกษาเกี่ยวกับการทดสอบแบบปรับเหมาะกับระดับความสามารถของผู้สอบในประเทศจะเป็นการพัฒนาแบบทดสอบที่มีวิธีการเลือกข้อสอบโดยใช้รูปแบบทางแยกคงที่ จัดการทดสอบโดยให้ผู้สอบตอบข้อสอบลงในกระดาษคำตอบ การเปรียบเทียบวิธีการเลือกข้อสอบในรูปแบบต่างกัน เพื่อดูว่าวิธีการเลือกข้อสอบรูปแบบใดจะมีประสิทธิภาพในการทดสอบสูงกว่า หรือการศึกษาตัวแปรที่ส่งผลต่อความตรงเชิงสภาพและข้อสอบในการทดสอบแบบปรับเหมาะโดยใช้คอมพิวเตอร์ (CAT) ในส่วนของงานวิจัยที่เกี่ยวกับการพัฒนาโปรแกรมคอมพิวเตอร์ พบว่า มีการพัฒนาโปรแกรมคอมพิวเตอร์สำหรับใช้งานทางด้านสถิติวิจัย การพัฒนาโปรแกรมคอมพิวเตอร์สำหรับการจัดเก็บข้อมูล การพัฒนาโปรแกรมคอมพิวเตอร์สำหรับงานการประเมินผล และการพัฒนาโปรแกรมสำหรับการวิเคราะห์แบบแผนการตอบ

สถาบันวิทยบริการ
จุฬาลงกรณ์มหาวิทยาลัย