บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ เป็นการศึกษาเกี่ยวกับการสร้างแบบทดสอบเทเลอร์ ดำเนินการสอบโดยใช้ไมโครคอมพิวเตอร์ ในวิชาคณิตศาสตร์ (ค 102) ชั้นมัธยมศึกษาปีที่ 1 โดยนำเอาทฤษฎีการตอบสนองข้อสอบ (Item Response Theory : IRT) มาประยุกต์ใช้กับการทดสอบแบบเทเลอร์ และการประมาณค่าความสามารถของผู้สอบ ซึ่งผู้วิจัยขอเสนอเอกสารและงานวิจัยที่เกี่ยวข้องดังนี้

ตอนที่ 1 ทฤษฎีการตอบสนองข้อสอบ (IRT)

1. หลักการของทฤษฎีการตอบข้อสอบ
2. ข้อตกลงเบื้องต้นเกี่ยวกับการใช้ทฤษฎีการตอบสนองข้อสอบ
3. ค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ
4. โมเดลของทฤษฎีการตอบสนองข้อสอบ
5. การนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้

ตอนที่ 2 การทดสอบแบบเทเลอร์ (Tailored Testing)

1. ความหมายของการทดสอบแบบเทเลอร์
2. รูปแบบของการทดสอบแบบเทเลอร์
3. วิธีดำเนินการสอบแบบเทเลอร์
4. วิธีประมาณค่าความสามารถของผู้สอบจากการทดสอบแบบเทเลอร์

ตอนที่ 3 คอมพิวเตอร์และการเขียนโปรแกรมคอมพิวเตอร์

1. ความหมายของคอมพิวเตอร์
2. ภาษาคอมพิวเตอร์
3. การเขียนโปรแกรมคอมพิวเตอร์

ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบเทเลอร์แบบแยกทางแปรผันและการประมาณค่าความสามารถของผู้สอบโดยใช้วิธีการของเบย์

1. งานวิจัยภายในประเทศ
2. งานวิจัยในต่างประเทศ

ตอนที่ 1 ทฤษฎีการตอบสนองข้อสอบ (IRT)

1. หลักการของทฤษฎีการตอบสนองข้อสอบ

ทฤษฎีการตอบสนองข้อสอบ (IRT) หรือทฤษฎีคุณลักษณะแฝง (Latent Trait Theory) มีความเชื่อว่าค่าพารามิเตอร์ (Parameter) ต่าง ๆ ของข้อสอบไม่ว่าจะเป็นค่าความยาก (b) ค่าอำนาจจำแนก (a) หรือค่าการเดา (c) ของข้อสอบแต่ละข้อเป็นคุณลักษณะที่มีอยู่ประจำ และคงที่พอสมควรในตัวข้อสอบนั้นจริง ฉะนั้นค่าพารามิเตอร์เหล่านี้จึงไม่ควรแปรเปลี่ยนไปตามกลุ่มตัวอย่าง (Sample - free) และในทำนองเดียวกันความสามารถของผู้สอบ (Ability) ก็เป็นคุณลักษณะที่มีอยู่ในตัวผู้สอบนั้นจริง จึงไม่ควรจะเปลี่ยนไปตามค่าความยากของข้อสอบ (Test - free) ซึ่งเป็นคุณลักษณะภายนอก แต่เนื่องจากความสามารถของผู้สอบเป็นคุณลักษณะแฝง (Latent Trait) ซึ่งเราไม่สามารถที่จะวัดหรือสังเกตได้โดยตรง (Unobservable) แต่ก็จะเป็นตัวพยากรณ์ (Predict) หรืออธิบาย (Explain) ผลการสอบ (Test Performance) หรือ คะแนน (Score) ซึ่งเป็นสิ่งที่เราสามารถสังเกตหรือวัดได้ (Observable) (Lord & Novick, 1968, p. 358; Hambleton & Cook, 1977, p. 75; Hambleton & Swaminathan, 1985, p. 9 อ้างใน สำเริง บุญเรืองรัตน์, 2527, หน้า 98)

2. ข้อตกลงเบื้องต้นเกี่ยวกับการใช้ทฤษฎีการตอบสนองข้อสอบ (IRT)

ข้อตกลงเบื้องต้นเกี่ยวกับการใช้ทฤษฎีการตอบสนองข้อสอบ ที่สำคัญมีดังนี้

2.1 ความเป็นเอกมิติของแบบทดสอบ (Unidimensionality) กล่าวคือ ข้อสอบแต่ละข้อในแบบทดสอบจะต้องวัดความสามารถหรือคุณลักษณะเดียว หรือมีความเป็นเอกพันธ์กัน สำหรับการที่จะตรวจสอบว่าข้อสอบนั้นมีลักษณะเป็นมิติเดียวหรือไม่นั้น สามารถใช้วิธีการวิเคราะห์ตัวประกอบ (Factor Analysis) ได้ แล้วดูว่ามีองค์ประกอบสำคัญอยู่มากกว่าหนึ่งองค์ประกอบหรือไม่ ถ้าหากมีมากกว่า ก็ถือว่าแบบทดสอบนั้นไม่เป็นแบบทดสอบที่มีมิติเดียว (Warm, 1978, unpaged และ สำเริง บุญเรืองรัตน์. 2527, ไม่มีเลขหน้า อ้างใน ชูเกียรติ ลอองแก้ว, 2537, หน้า 11)

2.2 ความเป็นอิสระต่อกัน (Local Independence) กล่าวคือ การที่ผู้สอบคนหนึ่งมีอิสระจากผู้สอบคนอื่น ๆ ในการทำข้อสอบ และการเป็นอิสระของผู้สอบในการทำข้อสอบข้อหนึ่ง ๆ จากการทำข้อสอบข้ออื่น ๆ (สุวิมล ติรกานันท์, 2539, หน้า 5)

2.3 โค้งคุณลักษณะข้อสอบ (Item Characteristic Curve : ICC) หมายถึง ความน่าจะเป็นที่ผู้สอบผู้หนึ่งจะตอบข้อสอบถูกนั้น ขึ้นอยู่กับโค้งคุณลักษณะข้อสอบของแต่ละโมเดลที่ใช้ ไม่ได้ขึ้นอยู่กับการแจกแจงของกลุ่มประชากร ซึ่งหมายความว่า รูปร่างของเส้นโค้งคุณลักษณะข้อสอบไม่ได้ขึ้นอยู่กับการกระจายความสามารถของกลุ่มผู้สอบ สำหรับโค้งคุณลักษณะข้อสอบนั้นจะเป็นโค้งในลักษณะของโค้งปกติสะสม (Normal Ogive) (Warm, 1978, unpaged และ สำเริง บุญเรืองรัตน์. 2527, ไม่มีเลขหน้า อ้างใน ชูเกียรติ ลอองแก้ว, 2537, หน้า 11)

3. ค่าพารามิเตอร์ตามทฤษฎีการตอบสนองข้อสอบ

ค่าพารามิเตอร์แบ่งออกเป็น 2 ชนิด คือ ค่าพารามิเตอร์ของผู้สอบ และค่าพารามิเตอร์ของข้อสอบ (Hambleton & Cook, 1977, unpaged อ้างใน ชูเกียรติ ลอองแก้ว, 2537, หน้า 13)

3.1 **ค่าพารามิเตอร์ของผู้สอบ** (Examinee Parameter) คือ ระดับความสามารถของผู้สอบ (θ) ที่คำนวณจากคะแนนจริง (True Score) โดยปกติ θ จะมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ ถ้า θ เป็นลบแสดงว่าผู้สอบมีความสามารถต่ำ ถ้า θ มีค่าเป็นบวก แสดงว่า ผู้สอบมีความสามารถสูง

3.2 **ค่าพารามิเตอร์ของข้อสอบ** (Item Parameters) ประกอบด้วย ค่าความยาก ค่าอำนาจจำแนก และค่าการเดา ซึ่งมีรายละเอียดดังนี้

3.2.1 **ค่าความยาก** (b) เป็นค่าที่แสดงระดับความสามารถที่จุดเปลี่ยนโค้งในโค้งคุณลักษณะข้อสอบ ในกรณีที่ไม่มีการเดา ค่าความยากก็คือ θ ณ จุดที่ $P_x(\theta) = 0.5$ ดังนั้น ค่าความยากจึงเป็นมาตรชนิดเดียวกันกับ θ โดยปกติแล้ว b จะมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ ถ้ามีค่าเป็นลบมากแสดงว่าข้อสอบนั้นง่ายมาก ถ้ามีค่าเป็นบวกมากแสดงว่าข้อสอบนั้นยากมาก

3.2.2 **ค่าอำนาจจำแนก** (a) เป็นค่าที่เป็นสัดส่วนโดยตรงกับความชัน (Slope) ของโค้งลักษณะข้อสอบ ณ จุดเปลี่ยนโค้ง หรือ จุดบนโค้งที่ $\theta = b$ โดยปกติค่า a จะมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ ถ้าค่า a เป็นลบ แสดงว่าข้อสอบนั้นไม่ดี ถ้าค่า a เท่ากับ 0 แสดงว่าข้อสอบไม่มีอำนาจจำแนก ถ้าค่า a มีค่าเป็นบวกมาก ๆ แสดงว่าข้อสอบมีอำนาจจำแนกสูง

3.2.3 **ค่าการเดา** (c) เป็นค่าที่แสดงถึงโอกาสในการตอบข้อสอบข้อใด ๆ ได้ถูกโดยที่ผู้สอบไม่มีความรู้ หรือมีความรู้น้อยมาก

4. โมเดลของทฤษฎีการตอบสนองข้อสอบ

รูปแบบของทฤษฎีการตอบสนองข้อสอบได้พัฒนาในรูปแบบต่าง ๆ ทำให้เกิดโมเดลเฉพาะขึ้นหลายโมเดล ซึ่งแต่ละโมเดลจะแตกต่างกันที่ฟังก์ชั่นทางคณิตศาสตร์ และจำนวนพารามิเตอร์ที่ใช้อธิบายโค้งลักษณะข้อสอบ เช่น โมเดลโค้งปกติสะสม (Normal Ogive Model) โมเดลเส้นตรง (The Linear Model) และโมเดลโลจีสติค (Logistic Model) จากโมเดลทั้งสามนี้ จะกล่าวถึงเฉพาะโมเดลโลจีสติค เพราะเป็นโมเดลที่นำมาใช้กันมาก ซึ่งโมเดลโลจีสติคมีโมเดลย่อยอยู่ 3 โมเดล คือ

4.1 **โมเดลที่ใช้พารามิเตอร์ตัวเดียวหรือราส์ชโมเดล** (One - Parameter Logistic Model or Rasch Model)

ปี ค.ศ. 1966 ราส์ช (Rasch) ได้พัฒนาทฤษฎีการตอบสนองข้อสอบ (IRT) และได้เสนอโมเดลนี้ซึ่งเรียกว่า โมเดลของราส์ช (Rasch Model) โดยโมเดลของราส์ช เป็นฟังก์ชั่นที่สามารถอธิบายได้ด้วยพารามิเตอร์เพียงตัวเดียวคือ ค่าความยาก (b) และโมเดลนี้ตรงกับโมเดลที่ใช้พารามิเตอร์ตัวเดียวที่ เบิร์นบอม (Birnbaum) ได้พัฒนาขึ้นในปี ค.ศ. 1968 ข้อตกลงของโมเดลนี้คือ ข้อสอบทุกข้อมีค่าอำนาจจำแนก (a) เท่ากัน จึงใช้ค่าเฉลี่ยของค่าอำนาจจำแนกแทนค่าอำนาจจำแนกของข้อสอบ และข้อสอบทุกข้อมีโอกาสการเดา (c) เท่ากับ 0 ซึ่งมีฟังก์ชั่นทางคณิตศาสตร์ดังนี้ (Hambleton & Swaminathan, 1985, p.47)

$$P_i(\theta) = \frac{e^{D\bar{a}(\theta - b_i)}}{1 + e^{D\bar{a}(\theta - b_i)}} \quad ; \quad i = 1, 2, 3, \ldots, n$$

เมื่อ $P_i(\theta)$ คือ ค่าความน่าจะเป็นในการตอบข้อสอบข้อที่ i ถูก เมื่อผู้สอบมีความสามารถ θ

b_i คือ ค่าความยากของข้อสอบข้อที่ i

$\bar{a}$ คือ ค่าเฉลี่ยของค่าอำนาจจำแนก

D คือ ค่าปรับสเกลมีค่าประมาณ 1.702

e คือ ค่าคงที่ประมาณ 2.718

n คือ จำนวนข้อสอบ

4.2 **โมเดลที่ใช้พารามิเตอร์สองตัว** (Two - Parameters Logistic Model)

ในปี ค.ศ. 1968 เบิร์นบอม (Birnbaum) ได้เสนอโมเดลโค้งลักษณะข้อสอบซึ่งใช้ค่าพารามิเตอร์ 2 ตัว คือ ค่าความยาก (b) และค่าอำนาจจำแนก (a) ข้อตกลงของโมเดลนี้คือ การตอบข้อสอบทุกข้อไม่มีการเดา (c=0) ซึ่งมีฟังก์ชั่นทางคณิตศาสตร์ดังนี้ (Hambleton & Swaminathan, 1985, p.36)

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

เมื่อ a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i

4.3 **โมเดลที่ใช้พารามิเตอร์สามตัว** (Three - Parameters Logistic Model) เป็นโมเดลที่ดัดแปลงมาจากโมเดลที่ใช้พารามิเตอร์สองตัว โดยเพิ่มค่าการเดา (c) เข้าไปเป็นค่าตัวแปรที่สาม ซึ่งมีฟังก์ชั่นทางคณิตศาสตร์ดังนี้ (Hambleton & Swaminathan, 1985, p.37)

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

เมื่อ c_i คือ ค่าการเดาของข้อสอบข้อที่ i

5. การนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้

การนำเอาทฤษฎีการตอบสนองข้อสอบ (IRT) มาประยุกต์ใช้จริงในระยะแรก ๆ ยังมีปัญหาหลายประการ คือ (Hambleton & CooK, 1977, pp. 75-76 อ้างใน ต่าย เชี่ยงฉี, 2534, หน้า 32-35)

(1) ตัวทฤษฎีอธิบายด้วยโมเดลทางคณิตศาสตร์ที่ซับซ้อนยากแก่การที่จะเข้าใจ

(2) ผู้เสนอทฤษฎีส่วนใหญ่เน้นในทางทฤษฎีมากกว่าที่จะนำมาประยุกต์ใช้

(3) ยังขาดโปรแกรมคอมพิวเตอร์ที่มีประสิทธิภาพ ที่ใช้สะดวกและรวดเร็วในการประมาณค่าความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบ

(4) ยังมีนักวิจัยหลายท่าน ตั้งข้อสงสัยถึงประโยชน์ที่จะได้รับจากการวิจัย และพัฒนาในทฤษฎีนี้

(5) เป็นทฤษฎีที่มีข้อตกลงเบื้องต้นที่ค่อนข้างเข้มงวด (Restrictive) ทำให้มีข้อจำกัดในการนำมาใช้

การนำเอาทฤษฎีการตอบสนองข้อสอบ (IRT) มาใช้กันอย่างจริงจังนั้น ได้เริ่มขึ้นหลังจากที่มีการพัฒนาโปรแกรมคอมพิวเตอร์มาใช้ในการประมาณค่าความสามารถของผู้สอบ และค่าพารามิเตอร์ต่าง ๆ ของข้อสอบ ประกอบกับในระยะหลังนี้ ได้มีไมโครคอมพิวเตอร์ (Microcomputer) ซึ่งเป็นคอมพิวเตอร์ (Computer) ขนาดเล็กแต่มีประสิทธิภาพสูงและนำไปใช้งานได้ค่อนข้างง่ายจึงมีการประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) ด้วยไมโครคอมพิวเตอร์ (Microcomputer) กันอย่างแพร่หลาย และดังที่กล่าวไว้แล้วในหลักการของทฤษฎีการตอบสนองข้อสอบ (IRT) เกี่ยวกับคุณลักษณะที่สำคัญในเรื่อง ค่าพารามิเตอร์ต่าง ๆ ของข้อสอบไม่ขึ้นอยู่กับความสามารถของผู้สอบ (Sample - Free) และความสามารถของผู้สอบก็ไม่ได้ขึ้นอยู่กับความยากหรือง่ายของแบบทดสอบ (Test - Free) จึงสามารถที่จะนำเอาคุณสมบัติเหล่านี้ มาประยุกต์ใช้ในทางการวัดผลการศึกษาได้อย่างมีประสิทธิภาพในหลาย ๆ เรื่อง ดังนี้

5.1 **การสร้างคลังข้อสอบ** (Item Bank) โดยการนำเอาทฤษฎีการตอบข้อสอบ (IRT) มาใช้ในการวิเคราะห์และสร้างคลังข้อสอบ ทำให้เกิดผลดีที่เห็นเด่นชัด 2 ประการ คือ ประการแรก ด้านค่าพารามิเตอร์ของข้อสอบ (Item Parameters) อันได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ที่วิเคราะห์โดยใช้ทฤษฎีการตอบข้อสอบ (IRT) มีคุณสมบัติไม่แปรเปลี่ยน (Invariant) ไปตามกลุ่มตัวอย่าง หรือตัวผู้สอบ ประการที่สอง ด้านการรายงานคุณภาพของข้อสอบในทฤษฎีการตอบข้อสอบ (IRT) จะรายงานคุณภาพของข้อสอบในรูปของอินฟอร์เมชั่น (Information) ซึ่งสามารถรายงานได้ทั้งเป็นรายข้อ (Item Information) และทั้งฉบับ (Test Information) ค่าอินฟอร์เมชั่นนี้ จะเป็นตัวบงชี้ถึงความถูกต้องแม่นยำ (Accuracy) ในการประมาณค่า ซึ่งสามารถใช้แทนค่าความเชื่อมั่น (Reliability) และค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) ในทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) ได้ (Hambleton & Swamenathan, 1985, p. 236 อ้างใน ต่าย เชี่ยงฉี, 2534, หน้า 34)

5.2 **การหาความลำเอียงของข้อสอบ** (Item Bias) ข้อสอบที่ใช้กันอยู่ทั่วไปโดยปกติจะมีบางข้อที่เหมาะสมเฉพาะคนบางกลุ่ม แต่จะไม่เหมาะสมกับคนอีกบางกลุ่ม โดยมากจะเกี่ยวข้องกับเพศ ศาสนา วัฒนธรรม ความเชื่อต่าง ๆ เป็นต้น ซึ่งจะเป็นผลทำให้เกิดการได้เปรียบและเสียเปรียบขึ้นในการทดสอบ เราสามารถใช้โค้งแสดงคุณลักษณะข้อสอบ (Item Characteristic Curve : ICC) มาตรวจสอบความลำเอียงของข้อคำถามได้ โดยปกติถ้าข้อคำถามนั้นไม่มีความลำเอียง โค้งลักษณะข้อสอบ (ICC) ไม่ว่าจะได้มาจากผู้สอบกลุ่มใดก็ตาม จะมีรูปร่าง (Shape) เหมือนกัน แต่ถ้าข้อคำถามนั้นมีความลำเอียง โค้งแสดงลักษณะข้อสอบก็จะมีรูปร่างแตกต่างกัน (Hambleton & Swaminathan, 1985, pp. 285-289 อ้างใน ต่าย เซี่ยงฉี, 2534, หน้า 34-35)

5.3 **การเทียบมาตรา** (Test Equating) หมายถึง กระบวนการทางสถิติเพื่อปรับคะแนนที่ได้จากการทดสอบด้วยข้อสอบต่างชุดกัน ที่มีโครงสร้างเดียวกัน ให้สามารถเทียบกันได้ (เรวดี อินทสะระ, 2530, หน้า 7) แต่ถ้าการทดสอบนั้นวัดความสามารถ (Ability or Trait) ที่ต่างกันแล้ว จะนำมาเทียบกันไม่ได้ เนื่องจากความสมารถของผู้สอบที่ประมาณค่ามาจากทฤษฎีการตอบสนองข้อสอบ (IRT) จะมีหน่วยการวัดเดียวกัน และไม่ขึ้นอยู่กับชุดของแบบทดสอบ จึงสามารถนำมาเปรียบเทียบกันได้ (Hambleton & Swaminathan, 1985, p. 202 อ้างใน ต่าย เซี่ยงฉี, 2534, หน้า 35)

5.4 **การทดสอบแบบเทเลอร์** (Tailored Testing) หมายถึง การทดสอบที่มีการจัดข้อสอบให้มีความยากที่เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล นับเป็นการนำเอาทฤษฎีการตอบสนองข้อสอบ (IRT) มาประยุกต์ใช้ได้อย่างมีประสิทธิภาพ (Urry, 1977, p. 181) โดยการนำเอาคุณสมบัติความไม่แปรเปลี่ยน (Invariant) ของค่าพารามิเตอร์ของข้อสอบและความสามารถของผู้สอบมาใช้ การทดสอบแบบเทเลอร์นั้นผู้สอบแต่ละคนจะได้รับชุดของข้อสอบที่แตกต่างกันตามระดับความสามารถของผู้สอบ บางคนอาจได้รับข้อสอบชุดที่ยาก และอีกบางคนอาจจะได้รับข้อสอบชุดที่ง่ายแต่ก็ยังสามารถนำเอาความสามารถที่ประมาณค่าได้จากการทดสอบแบบเทเลอร์มาเปรียบเทียบกันได้ ทั้งนี้เพราะ ความสามารถที่ประมาณค่าได้ยังอยู่ในมาตราวัดเดียวกัน (Common Ability Scale) (Hambleton & Cook, 1977, p. 90-91; Hambleton & Swaminathan, 1985, p. 296 อ้างใน ต่าย เซี่ยงฉี, 2534, หน้า 35)

ตอนที่ 2 การทดสอบแบบเทเลอร์ (Tailored Testing)

1. ความหมายและหลักการในการทดสอบแบบเทเลอร์

การทดสอบแบบเทเลอร์ เป็นการทดสอบที่จัดข้อสอบให้มีความยากให้เหมาะสมกับระดับความสามารถของผู้สอบแต่ละคน (Lord, 1971, p. 43) และผู้สอบแต่ละคนก็ไม่จำเป็นต้องทำข้อสอบเหมือน ๆ กันทุกข้อ จำนวนข้อก็ไม่จำเป็นต้องเท่ากัน ทั้งนี้ขึ้นอยู่กับรูปแบบและผลการตอบข้อสอบในการทดสอบแบบเทเลอร์นั้น ๆ โดยทั่วไปแล้ว จะให้ผู้สอบเริ่มทำข้อสอบข้อที่มีความยากปานกลางก่อน ถ้าผลการตอบข้อสอบถูกข้อต่อไปจะมีความยากมากขึ้น หากตอบข้อสอบข้อนั้นผิดข้อต่อไปก็จะง่ายขึ้น การทำข้อสอบขั้นต่อไปจะดำเนินการเช่นนี้ไปจนสิ้นสุดการทดสอบ (ต่าย เชี่ยงฉี, 2534, หน้า 36)

2. รูปแบบของการทดสอบแบบเทเลอร์

การทดสอบแบบเทเลอร์ แบ่งออกได้เป็นหลายลักษณะทั้งนี้ขึ้นอยู่กับเกณฑ์ที่ใช้ในการจำแนก ในที่นี้ขอเสนอเกณฑ์ที่ใช้ในการจำแนกที่สำคัญได้แก่ การจำแนกตามเครื่องมือที่ใช้ในการทดสอบ และการจำแนกตามยุทธวิธีที่ใช้ในการคัดเลือกข้อสอบ (ศิริชัย กาญจนวาสี, 2538, หน้า 4-5)

ฮูลิน, ดราส์โกว์, และ พาร์สันส์ (Hulin, Drasgow & Parsons, 1983, unpaged อ้างใน ศิริชัย กาญจนวาสี, 2538, หน้า 4) ได้จำแนกประเภทของการทดสอบแบบเทเลอร์ โดยพิจารณาถึงเครื่องมือที่ใช้ในการทดสอบ ซึ่งจำแนกเป็นการทดแบบเทเลอร์ที่ไม่ใช้คอมพิวเตอร์และการทดสอบที่ใช้คอมพิวเตอร์ การทดสอบแบบเทเลอร์ที่ใช้คอมพิวเตอร์ (Computerized Adaptive Testing) เป็นการทดสอบแบบเทเลอร์ที่นำคอมพิวเตอร์มาใช้เพื่อความสะดวกในการคัดเลือกข้อสอบ และประมาณค่าความสามารถของผู้สอบ ซึ่งนิยมใช้การกำหนดทางแบบแปรผัน และใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นพื้นฐานในการประมาณค่า

ไวส์, แฮมเบิลตัน, และสวามินาธาน (Weiss, 1974, pp. 78-110; Hambleton & Swaminathan, 1985, p.297 อ้างใน ต่าย เชี่ยงฉี่, 2534, หน้า 36; ศิริชัย กาญจนวาสี, 2538, หน้า 5) ได้จำแนกประเภทของการทดสอบแบบเทเลอร์ โดยพิจารณาถึงยุทธวิธีที่ใช้ในการคัดเลือกข้อสอบ ซึ่งจำแนกเป็นการทดสอบแบบเทเลอร์ที่ใช้ยุทธวิธีสองขั้นตอนและยุทธวิธีหลายขั้นตอน

2.1 ยุทธวิธีสองขั้นตอน (Two-stage Strategies)

เป็นการทดสอบแบบเทเลอร์ที่แบ่งการทดสอบออกเป็นสองขั้นตอน ประกอบด้วยขั้นตอนแรกเป็นการสอบกำหนดทิศทาง และขั้นตอนที่สองเป็นการสอบวัดผล

2.2 ยุทธวิธีหลายขั้นตอน (Multi-Stage Strategies)

เป็นการทดสอบแบบเทเลอร์ที่แบ่งการทดสอบออกเป็นหลายขั้นตอน โดยมีการจัดโครงสร้างของขั้นตอนจัดเรียงข้อสอบ การเลือกข้อสอบ และการยุติการทดสอบหลายรูปแบบ แบ่งออกเป็น

2.2.1 ยุทธวิธีหลายขั้นตอนแบบแยกทางคงที่ (Fixed-Branching)

2.2.1.1 รูปปิรามิดแบบขนาดขั้นคงที่ (Constant Step Size Pyramid)

2.2.1.2 รูปปิรามิดแบบขนาดขั้นแปรผัน (Variable Step Size Pyramid)

2.2.1.3 รูปปิรามิดแบบข้างตัด (Truncated Pyramid)

2.2.1.4 รูปปิรามิดแบบมีหลายข้อในแต่ละขั้น (Multiple-Item Pyramid)

2.2.1.5 รูปปิรามิดแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง (Differential Response Option Branching)

2.2.1.6 แบบทดสอบเฟล็กซิเลเวล (Flexilevel Test)
แบบทดสอบสเตร็ดเด็ฟตีฟ (Stradaptive Test)

2.2.2 ยุทธวิธีหลายขั้นตอนแบบแยกทางแปรผัน (Variable-Branching)

2.2.2.1 ยุทธวิธีของเบย์ (Bayesian Strategies)

2.2.2.2 ยุทธวิธีความเป็นไปได้สูงสุด (Maximum Liklihood Strategies)

สำหรับการวิจัยครั้งนี้ เป็นการวิจัยที่ศึกษาเกี่ยวกับการทดสอบแบบเทเลอร์ยุทธวิธีหลายขั้นตอนแบบแยกทางแปรผันโดยใช้ยุทธวิธีของเบย์

3. วิธีดำเนินการทดสอบแบบเทเลอร์

วิธีดำเนินการทดสอบแบบเทเลอร์สามารถแบ่งออกได้เป็น 2 วิธี คือ (สวงค์ บุญปลูก, 2534, หน้า 26 อ้างใน ประสาร สุขสอน, 2536, หน้า 21)

3.1 จัดข้อสอบแยกออกเป็นชุด ๆ โดยการใช้ค่าพารามิเตอร์ของข้อสอบเป็นหลัก โดยจัดให้มีความยากหลาย ๆ ระดับ เพื่อจะได้แยกระดับให้ผู้สอบได้ทดสอบตามความสามารถของผู้สอบ ภายหลังจากการทดสอบในแบบทดสอบชุดแรก ดังนั้นผู้สอบจะต้องสอบทีละฉบับ

3.2 ผู้สอบต้องสอบทีละข้อ โดยผลการสอบในแต่ละข้อจะถูกประมาณค่าความสามารถเพื่อจะคัดเลือกข้อสอบที่เหมาะสมกับความสามารถของผู้สอบต่อไป

4. วิธีประมาณค่าความสามารถของผู้สอบจากการทดสอบแบบเทเลอร์

ได้มีผู้เสนอวิธีการประมาณค่าความสามารถของผู้สอบไว้หลายวิธีการ พอสรุปได้ดังนี้ (ต่าย เชี่ยงฉี, 2534, หน้า 49-55)

4.1 ประมาณค่าความสามารถของผู้สอบตามค่าความยากของข้อสอบข้อที่ยากที่สุดที่ทำถูก (The Most Difficulty of Item Answered Correctly) โดยดูจากการตอบข้อสอบในการทดสอบแบบเทเลอร์ที่ผู้สอบตอบมาทั้งหมด ข้อใดมีค่าความยากสูงสุดที่ผู้สอบทำถูก ถือว่าเป็นค่าความสามารถของผู้สอบคนนั้น โดยค่าความยากที่ว่านี้ ควรจะเป็นค่าความยากที่วิเคราะห์มาจากทฤษฎีการตอบสนองข้อสอบ (IRT)

4.2 ประมาณค่าความสามารถของผู้สอบตามค่าเฉลี่ยความยากของข้อสอบข้อที่ทำถูก (Average Difficulty of all Item Answered Correctly) เป็นการประมาณค่าความสามารถของผู้สอบ โดยนำค่าความยากของข้อสอบจากการทดสอบแบบเทเลอร์ทุกข้อที่ผู้สอบทำถูกมาคำนวณหาค่าเฉลี่ย ค่าเฉลี่ยความยากนี้ ถือเป็นค่าความสามารถของผู้สอบคนนั้น

4.3 ประมาณค่าความสามารถของผู้สอบตามค่าความยากของข้อสอบทุกข้อที่ทำ (Average Difficulty of all Item Answered) เป็นการนำเอาค่าความยากของข้อสอบจากการทดสอบแบบเทเลอร์ทุกข้อที่ผู้สอบได้ทำ โดยไม่คำนึงว่าจะตอบถูกหรือไม่มาหาค่าเฉลี่ย ซึ่งค่าเฉลี่ยที่ได้นี้จะถือว่าเป็นค่าความสามารถของผู้สอบคนนั้น

4.4 ประมาณค่าความสามารถของผู้สอบตามค่าความยากของข้อสอบข้อสุดท้ายที่ทำ (Difficulty of the Final Term) นั่นคือ ไม่คำนึงถึงว่าขั้นสุดท้ายจะตอบถูกหรือไม่ก็ตามจะถือว่าค่าความยากจากการทดสอบแบบเทเลอร์ในขั้นสุดท้าย ที่ผู้สอบได้ทำ คือค่าความสามาถของผู้สอบคนนั้น

4.5 ประมาณค่าความสามารถของผู้สอบตามค่าความยากของข้อสอบที่ต่อจากคำถามในขั้นสุดท้ายที่ทำ (Difficulty of $(N+1)^{th}$ Item) ซึ่งมีลักษณะคล้ายกับข้อที่ 4 แต่จะเป็นการตรวจสอบต่อไปอีกขั้นหนึ่ง กล่าวคือ ถ้าข้อสอบในขั้นสุดท้ายทำถูก ค่าความยากในขั้นต่อไปจะเพิ่มขึ้น แต่ถ้าทำผิด ขั้นต่อไปค่าความยากจะลดลง

4.6 ประมาณค่าความสามารถของผู้สอบโดยใช้หลักการของเบย์ (Bayesian Strategies) การใช้หลักการของเบย์นี้ยังมีวิธีการย่อยหลายวิธี แต่วิธีการที่นิยมใช้กันมากคือ วิธีที่นักสถิติชื่อ โรเจอร์ เจ โอเวน (Roger J. Owen) ได้เป็นผู้เสนอขึ้นเพื่อจะนำไปใช้ประมาณค่าความสามารถของผู้สอบจากการทดสอบแบบเทเลอร์โดยเฉพาะโดยให้ชื่อว่า เบย์เซี่ยน อัพเดทติ้ง (Bayesian Updating) ซึ่งสามารถประมาณค่าความสามารถของผู้สอบได้ค่อนข้างคงที่

4.7 ประมาณค่าความสามารถของผู้สอบโดยใช้หลักการความเป็นไปได้สูงสุด (Maximum Likelihood) การใช้หลักการความเป็นไปได้สูงสุดนี้ ยังมีวิธีการย่อยอีกหลายวิธี แต่วิธีที่นิยมใช้กันมากคือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) การประมาณค่าโดยวิธีนี้ มีข้อจำกัดในกรณีที่ผู้สอบตอบข้อสอบถูกหมดหรือผิดหมด จะไม่สามารถประมาณค่าความสามารถของผู้สอบคนนั้นได้

สำหรับการวิจัยในครั้งนี้ การประมาณค่าความสามารถของผู้สอบเลือกใช้วิธีประมาณค่าความสามารถของผู้สอบโดยใช้หลักการของเบย์ ที่ชื่อว่า เบย์เซี่ยน อัพเดทติ้ง (Bayesian Updating)

ตอนที่ 3 คอมพิวเตอร์และการเขียนโปรแกรมคอมพิวเตอร์

1. ความหมายของคอมพิวเตอร์

คอมพิวเตอร์ คือ อุปกรณ์อิกเล็กทรอนิกส์ซึ่งมนุษย์ใช้เพื่อช่วยในการคำนวณ และเก็บ หรือค้นหาข้อมูล โดยคอมพิวเตอร์จะปฏิบัติตามคำสั่งของมนุษย์เท่านั้น มนุษย์สั่งงานคอมพิวเตอร์โดยใช้คำสั่งหลาย ๆ คำสั่งโดยต่อเนื่องกันเรียกว่า ชุดคำสั่ง หรือโปรแกรม (Program) คอมพิวเตอร์มีความสามารถเก็บโปรแกรม รวมทั้งคำสั่งในการตัดสินใจเลือกหนทางปฏิบัติตามเงื่อนไขที่เราได้กำหนดไว้ สามารถทำการคำนวณตามคำสั่ง สามารถเก็บข้อมูลทั้งประเภทตัวเลขและตัวอักษร (เป็นคำ) เพื่อใช้ในการประมวลผล สามารถปฏิบัติตามขั้นตอนซึ่งเรากำหนดไว้ในโปรแกรม และสามารถแสดงผลที่อุปกรณ์แสดงผล (พีรศักดิ์ ศรีกังวาล, 2534, หน้า 16)

2. ภาษาคอมพิวเตอร์

ภาษาคอมพิวเตอร์เป็นสื่อกลางในการติดต่อสั่งงานระหว่างคนกับเครื่อง มีจำนวนมากมายหลายภาษา แต่ละภาษามีกฎเกณฑ์ โครงสร้าง และไวยกรณ์แตกต่างกันออกไป แต่วัตถุประสงค์ของแต่ละภาษามีจุดมุ่งหมายเดียวกัน คือ เป็นภาษาที่มนุษย์สั่งงานเพื่อให้คอมพิวเตอร์ทำงานตามที่ตนต้องการ ภาษาคอมพิวเตอร์แบ่งออกเป็น 3 ประเภท (ประพัฒน์ อุทโยภาส, 2531, ไม่มีเลขหน้า, อ้างใน โกชัญ ทองคำ, 2540, หน้า 7) ดังนี้คือ

2.1 ภาษาเครื่อง (Machine Language) เป็นภาษาที่ใช้สั่งงานเครื่องคอมพิวเตอร์โดยตรง ภาษานี้อยู่ในลักษณะของเลขฐานสอง เป็นภาษาที่เครื่องคอมพิวเตอร์รับรู้ได้โดยตรงไม่ต้องอาศัยตัวแปลภาษา แต่มนุษย์อ่านเข้าใจยากเพราะมีแต่ตัวเลขล้วน ๆ และการเขียนโปรแกรมก็ยุ่งยาก

2.2 ภาษาแอสเซมบลีย์ (Assembly Language) เป็นภาษาที่พัฒนามาจากภาษาเครื่องเพื่อให้เขียนโปรแกรมได้ง่ายขึ้น มีลักษณะการใช้อักษรปนตัวเลขอยู่ และต้องใช้ตัวแปรภาษาด้วยโปรแกรมแอสเซมเบลอร์ (Assembler)

2.3 ภาษาระดับสูง (High Level Language) เป็นภาษาที่พัฒนาขึ้นมามีลักษณะใกล้เคียงภาษาพูด ทำให้ง่ายต่อความเข้าใจ ผู้ใช้ไม่จำเป็นต้องเรียนรู้ภาษาเครื่อง เพราะจะมีตัวแปรภาษาระดับสูงไปเป็นภาษาเครื่อง ตัวแปรนี้เรียกว่า คอมไพเลอร์ (Complier)

ภาษาระดับสูงนี้มีอยู่มากมาย ถ้าแบ่งตามลักษณะการใช้งานแล้ว อาจแบ่งได้เป็น 4 กลุ่ม (อนุชิต, 2526, หน้า 46 อ้างใน โกชัญ ทองคำ, 2540, หน้า 7) คือ

2.3.1 ภาษาฟอร์แทรน (FORTRAN) ภาษาปาสคาล (PASCAL) เป็นภาษาที่เหมาะสำหรับใช้เขียนโปรแกรมสำหรับงานวิจัย วิศวกรรม วิทยาศาสตร์ และคณิตศาสตร์ เนื่องจากลักษณะงานประเภทนี้มีการคำนวณตัวเลข และฟังก์ชั่นทางคณิตศาสตร์มาก จึงเน้นความเร็วในการคำนวณเป็นหลัก

2.3.2 ภาษาโคบอล (COBOL) ภาษาอาร์พีจี (RPG) เป็นภาษาที่เหมาะสำหรับใช้เขียนโปรแกรมสำหรับงานเกี่ยวกับธุรกิจ เป็นภาษาที่เน้นการจัดทำรายงาน เก็บข้อมูล การคำนวณไม่ซับซ้อน

2.3.3 ภาษาไดนาโม (DYNAMO) ภาษาแกส (GASS) เป็นภาษาที่เหมาะสำหรับใช้เขียนโปรแกรมเฉพาะงาน เช่น ใช้เขียนโปรแกรมสำหรับการจำลองสถานการณ์

2.3.4 ภาษาเบสิค (BASIC) และภาษาพีแอลวัน (PL/1) เป็นภาษาที่เหมาะสำหรับใช้เขียนโปรแกรมสำหรับงานทั่วไป หรือเอนกประสงค์ เป็นภาษาที่ใช้กับงานวิทยาศาสตร์ หรือธุรกิจก็ได้ แต่ไม่ดีเท่าภาษาเฉพาะงาน

ปัจจุบันภาษาเบสิคได้ถูกพัฒนาเป็นโปรแกรมสำเร็จรูปสำหรับทำงานบนระบบวินโดว์รุ่นต่างๆ ซึ่งเรียกว่า ไมโครซอฟท์วิชวลเบสิค (Microsoft Visual Basic) หรือเรียกสั้น ๆ ว่า

วิชวลเบสิค (Visual Basic) ซึ่งได้มีการพัฒนาขึ้นมาด้วยกันหลายรุ่น ประสิทธิภาพในการทำงานก็สูงขึ้นเรื่อย ๆ

วิชวลเบสิค (Visual Basic)

การเขียนโปรแกรมเป็นสิ่งสำคัญที่จะสามารถทำให้คอมพิวเตอร์ทำงานได้ตามความต้องการของผู้ใช้ เริ่มจากการรวบรวมแนวความคิดก่อนว่าจะให้โปรแกรมทำงานอะไร มีขั้นตอนเป็นอย่างไร รวมถึงการตรวจสอบผลลัพธ์ด้วยว่าถูกต้องหรือไม่ ซึ่งจะมีรูปแบบการเขียนเป็นลักษณะเฉพาะของแต่ละโปรแกรมแตกต่างกันไป ตามความต้องการและการใช้งาน สำหรับวิชวลเบสิคแล้ว ผู้เขียนโปรแกรมไม่จำเป็นที่จะต้องเคยเขียนโปรแกรมอย่างเชี่ยวชาญมาก่อน เพราะวิชวลเบสิคได้เตรียมเครื่องมือในการพัฒนาโปรแกรมไว้อย่างมากมาย และมีรูปแบบการใช้งานที่คุ้นเคยเช่นเดียวกับแอพพลิเคชั่นต่าง ๆ บนวินโดว์ เช่น มีปุ่มที่อยู่บนทูลบาร์ ปุ่มตัวเลือก กรอบข้อความ กรอบรูปภาพ สกรอลบาร์ เป็นต้น ทำให้ง่ายต่อการเขียนโปรแกรม นอกจากนี้ ยังสามารถหาอุปการณ์เสริมการทำงาน และแหล่งข้อมูลอ้างอิงได้ง่ายกว่า เช่น ค้นหาที่ http:/www.microsoft.com หรือจากที่อื่น ๆ อีกมาก เพียงแต่เรียนรู้ถึงวิธีการเขียนโปรแกรม และกำหนดหน้าที่การทำงานให้กับเครื่องมือต่าง ๆ โดยให้สัมพันธ์กับเหตุการณ์ที่เกิดขึ้น ซึ่งเป็นหลักการเขียนที่เรียกว่า Event-Driven Programming นั่นเอง เช่น จะให้โปรแกรมทำงานอะไรต่อไปเมื่อคลิกเมาส์ที่ปุ่มหรือตัวเลือก หรือเมื่อเลื่อนสกรอลบาร์ เพื่อดูข้อมูลที่ซ่อนอยู่ในกรอบข้อความ การใส่รูปภาพเพิ่มความน่าสนใจให้กับงาน เป็นต้น (สัจจะ จรัสรุ่งรวีวร, และ กฤษณะ สถิตย์, ม.ป.ป., หน้า 4)

สำหรับการวิจัยครั้งนี้ได้ใช้โปรแกรมสำเร็จรูปไมโครซอฟท์วิชวลเบสิค รุ่น 5.0 (Microsoft Visual Basic 5.0) ในการเขียนโปรแกรมสำหรับดำเนินการสอบ

3. การเขียนโปรแกรมคอมพิวเตอร์

การเขียนโปรแกรม เพื่อให้คอมพิวเตอร์ทำงานในด้านต่าง ๆ นั้น ผู้เขียนโปรแกรมควรปฏิบัติตามขั้นตอนต่อไปนี้ (สานนท์ เจริญฉาย, 2533, หน้า 14)

3.1 **วิเคราะห์งาน** (Job Analyzation) โดยศึกษาวิเคราะห์ดังนี้

3.1.1 ข้อมูลหรือรายละเอียดต่าง ๆ ที่จะใช้ดำเนินงาน (Input)

3.1.2 ผลลัพธ์หรือรายงานที่ต้องการ (Output) ต้องศึกษาถึงรูปแบบ เนื้อหาสาระ ลำดับ และหัวข้อ ว่าต้องการแสดงออกมาในรูปแบบใด

3.1.3 กระบวนการทำงานหรือขั้นตอนการปฏิบัติงาน (Process) เป็นขั้นที่มีความสำคัญมาก คือ เป็นขั้นนำข้อมูล (Input) มาเข้ากระบวนการต่าง ๆ เพื่อให้ได้ผลลัพธ์ตามเป้าหมายที่ต้องการ

3.1.4 พิจารณาข้อผิดพลาด (Error) ที่อาจเกิดขึ้น โดยศึกษาข้อผิดพลาดต่าง ๆ ที่อาจเกิดขึ้นในขณะป้อนข้อมูล หรือขณะที่เครื่องกำลังประมวลผลข้อมูลอยู่ พร้อมทั้งข้อมูลต่าง ๆ ในการตรวจสอบข้อผิดพลาดต่าง ๆ เพื่อหาทางป้องกันและแก้ไขต่อไป

3.2 **เขียนผังระบบ** (System Flowchart) และผังโปรแกรม (Programming Flowchart)

เมื่อทราบรายละเอียดต่าง ๆ เกี่ยวกับ Input, Output, และ Error แล้วผู้เขียนโปรแกรมก็นำสิ่งต่าง ๆ ที่วิเคราะห์มาเขียนผังระบบและผังโปรแกรม แสดงขั้นตอนต่าง ๆ ของการทำงาน เพื่อเข้ากระบวนการ (Process) ให้ได้ผลลัพธ์ตามต้องการ

การเขียนผังระบบ (System Flowchart) เป็นการเขียนแผนภูมิเพื่อแสดงการทำงานทั้งหมดอย่างคร่าว ๆ โดยเน้นการส่งข้อมูลในแต่ละขั้นว่าจะให้ปฏิบัติอย่างไร อุปกรณ์ที่ใช้คืออะไร

การเขียนผังโปรแกรม (Programming Flowchart) เป็นการเขียนแผนภูมิเพื่อแสดงรายละเอียดแต่ละขั้นของงาน หรือเป็นการนำผังระบบมาแยกรายละเอียดปลีกย่อย

3.3 **เขียนโปรแกรม** (Programming)

การเขียนโปรแกรม คือ การนำภาษาคอมพิวเตอร์ภาษาใดภาษาหนึ่งมาเขียนเป็นคำสั่งให้เครื่องคอมพิวเตอร์ทำงานตามผังโปรแกรมที่เตรียมเอาไว้แล้ว ในการเขียนโปรแกรมที่ดีนั้นต้องมีการวางรูปของโปรแกรมให้มีโครงสร้างที่ดี ง่ายต่อการตรวจสอบ มีขนาดเนื้อที่ภายในหน่วยความจำไม่มาก และที่สำคัญเครื่องไมโครคอมพิวเตอร์ต้องทำงานได้เร็ว (ยืน ภู่วรรณ, 2526, หน้า 92 อ้างใน โกชัญ ทองคำ, 2540, หน้า 10)

3.4. **การทดสอบและการแก้ไขโปรแกรม** (Program and Debugging)

การทดสอบและการแก้ไขโปรแกรม คือ การตรวจสอบความถูกต้องของโปรแกรมทั้งในด้านหลักภาษาและเงื่อนไขการทำงาน โดยการทดลองใช้โปรแกรมเพื่อตรวจสอบความผิดพลาดที่อาจเกิดขึ้น แล้วปรับปรุงแก้ไขโปรแกรมจนกระทั่งได้โปรแกรมที่สมบูรณ์ หลังจากนั้นจึงเขียนคู่มือการใช้งานและนำโปรแกรมไปใช้งานจริง

วันพร ปั้นเก่า (2530, ไม่มีเลขหน้า อ้างใน โกชัญ ทองคำ, 2540, หน้า 11) กล่าวว่า การออกแบบและพัฒนาโปรแกรมเป็นวิธีการที่จะช่วยให้นักพัฒนาโปรแกรม (Programer) สามารถพัฒนาโปรแกรมได้อย่างรวดเร็วและถูกต้อง ซึ่งความยุ่งยากในการพัฒนาโปรแกรมที่สำคัญมีอยู่ 2 ประการ คือ ขั้นตอนของโปรแกรม และกฎเกณฑ์ของภาษาคอมพิวเตอร์

1. **ขั้นตอนในการออกแบบและพัฒนาโปรแกรม**

โดยทั่วไป ขั้นตอนในการออกแบบและการพัฒนาโปรแกรมมีดังนี้

1.1 การวิเคราะห์ปัญหา

1.2 การออกแบบขั้นตอนสำหรับโปรแกรม

1.3 การเขียนโปรแกรม

1.4 การทดสอบโปรแกรม

1.5 การทำเอกสารประกอบโปรแกรม

1.6 การบำรุงรักษาโปรแกรม

2. **การวิเคราะห์ปัญหา**

การวิเคราะห์ปัญหา คือ การศึกษาปัญหานั้นเสียก่อนว่าต้องทำอะไรบ้าง และทำอย่างไร ข้อมูลและผลลัพธ์มีอะไรบ้าง จะต้องอาศัยสูตรหรือทฤษฎีอะไร ตลอดจนมีเงื่อนไขหรือข้อจำกัดหรือไม่

3. **การออกแบบขั้นตอนสำหรับโปรแกรม**

เมื่อได้ทำการวิเคราะห์ปัญหาให้เข้าใจแล้ว ให้ทำการกำหนดชื่อข้อมูล ผลลัพธ์และผลลัพธ์ชั่วคราว (ถ้ามี) แล้วทำการออกแบบขั้นตอนสำหรับโปรแกรม ซึ่งจะต้องประกอบด้วยขั้นตอนการทำงานอย่างครบถ้วน พร้อมทั้งลำดับการทำงานที่ถูกต้อง การออกแบบขั้นตอนนี้มีหลายวิธี วิธีการที่นิยมคือเขียนเป็นผังงาน เมื่อได้ทบทวนหรือตรวจสอบขั้นตอนเหล่านั้นถูกต้องแล้ว จึงทำการเขียนโปรแกรม

4. **การเขียนโปรแกรม**

ขั้นตอนนี้เป็นการเขียนคำสั่งด้วยภาษาคอมพิวเตอร์ เพื่อให้เครื่องทำงานตามขั้นตอนที่ได้ออกแบบไว้ ในการเขียนคำสั่งต้องคำนึงถึงกฎเกณฑ์และหลักการของภาษาคอมพิวเตอร์ที่ใช้ได้ถูกต้องด้วย เพราะถ้ามีข้อผิดพลาด ซึ่งเรียกว่า Syntax Error เกิดขึ้น โปรแกรมแปลภาษาจะไม่สามารถแปลความหมายของคำสั่งนั้นได้ และส่วนใหญ่จะมีรายงานหรือแสดงข้อความออกมาให้

ทราบว่ามีข้อผิดพลาดตรงไหน เพราะอะไร เพื่อให้ผู้เขียนโปรแกรม แก้ไขโปรแกรมได้อย่างถูกต้องและรวดเร็ว

5. **การทดสอบโปรแกรม**

ขั้นตอนนี้เป็นการทดสอบว่าโปรแกรมที่พัฒนาขึ้นนั้นสั่งให้เครื่องทำงานได้อย่างถูกต้องหรือไม่ วิธีการทดสอบกระทำได้โดยการสั่งให้เครื่องปฏิบัติตามคำสั่งในโปรแกรม (Object Program) ซึ่งถ้ามีการทดสอบให้เครื่องรับข้อมูลเข้าไปทำการประมวลผลก็จะต้องนำผลตัวอย่างหรือข้อมูลจริงส่งเข้าไป แล้วนำผลลัพธ์ที่ได้มาตรวจสอบกับผลลัพธ์ที่ถูกต้อง ถ้าได้ตรงกันก็จะยอมรับว่าโปรแกรมนั้นใช้งานได้ แต่ถ้าไม่ตรงกันจะต้องพิจารณาว่าความผิดพลาดที่เกิดขึ้นอยู่ที่ข้อมูลหรือตัวโปรแกรม ถ้าข้อมูลผิดพลาดจะต้องแก้ไขและส่งเข้าไปประมวลผลใหม่ แล้วตรวจสอบผลลัพธ์ แต่ถ้าโปรแกรมผิด ซึ่งข้อผิดพลาดที่เกิดขึ้นอาจเกิดจากการกำหนดการคำนวณหรือการเปรียบเทียบผิด การอ้างชื่อข้อมูลหรือผลลัพธ์ผิด มีการสลับที่หรือขาดบางตอนไป เป็นต้น ซึ่งข้อผิดนี้เรียกว่า Logic Error การค้นหาข้อผิดพลาดเกี่ยวกับขั้นตอนนั้นจะค้นหาในผังงานที่ออกแบบไว้ ซึ่งสะดวกกว่าการค้นหาในโปรแกรม

6. **การทำเอกสารประกอบโปรแกรม**

ขั้นตอนนี้เป็นการรวบรวมตั้งแต่ขั้นตอนแรกมาไว้เป็นเอกสาร ซึ่งจะประกอบด้วยส่วนสำคัญ ๆ ได้แก่ เนื้อหาของปัญหา (โจทย์) ผลการวิเคราะห์ปัญหานั้น ๆ ขั้นตอนสำหรับโปรแกรม (ผังงาน) รูปแบบของข้อมูลและผลลัพท์ เนื้อหาของ Source Program รายละเอียดการใช้และข้อจำกัดของโปรแกรม ตลอดจนตัวอย่างของผลลัพธ์ที่ได้จากโปรแกรมนั้น ๆ ซึ่งเอกสารนี้จะเป็นประโยชน์อย่างยิ่งในเวลาต่อไป

7. **การบำรุงรักษาโปรแกรม**

ขั้นตอนนี้เป็นขั้นตอนที่จะดูแลโปรแกรมให้มีความเหมาะสมกับงานตลอดเวลาตามที่ต้องการ ทั้งนี้เพราะในงานใด ๆ ก็ตาม เมื่อเวลาผ่านไปสักระยะหนึ่ง อาจมีการเปลี่ยนแปลงในส่วนของข้อมูลผลลัพธ์หรือวิธีการไปจากเดิมได้ ซึ่งโปรแกรมที่มีอยู่เดิมไม่สามารถทำงานได้ถูกต้องทั้งหมด จึงจำเป็นต้องทำการแก้ไขโปรแกรม โดยอาศัยเอกสารประกอบโปรแกรมเป็นแนวทางในการศึกษาถึงจุดเปลี่ยนแปลงแก้ไข

ตอนที่ 4 งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบเทเลอร์แบบแยกทางแปรผันและการประมาณค่าความสามารถของผู้สอบโดยใช้วิธีการของเบย์

1. งานวิจัยภายในประเทศ

นันทิยา พึ่งคำ (2531, หน้า 92) ได้เปรียบเทียบคุณภาพของการทดสอบแบบเทเลอร์ด้วยคอมพิวเตอร์ กับการทดสอบแบบเดิม กลุ่มตัวอย่างได้แก่นักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 120 คน แบบทดสอบแบบเดิมมีข้อสอบ 40 ข้อ ส่วนข้อสอบในคลังข้อสอบที่ใช้สำหรับการทดสอบแบบเทเลอร์ด้วยคอมพิวเตอร์นั้นมี 361 ข้อสอบทั้งสองลักษณะวัดความสามารถด้านคำศัพท์ภาษาอังกฤษ ผลจากการศึกษาพบว่า การทดสอบทั้งสองลักษณะมีความเที่ยงตรงเชิงเกณฑ์สัมพันธ์ไม่ต่างกัน โดยที่การทดสอบแบบเทเลอร์ด้วยคอมพิวเตอร์ใช้ข้อสอบน้อยกว่าครึ่งหนึ่งของการทดสอบแบบเดิม นอกจากนั้นยังมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำกว่าการทดสอบแบบเดิมอีกด้วย

วิชุดา บัวคง (2532) ได้ศึกษาเปรียบเทียบประสิทธิผลของวิธีประมาณค่าพารามิเตอร์ของแบบจำลองโลจิสติก 3 พารามิเตอร์ ระหว่างวิธีแมกซิมัมไลค์ลิฮูด วิธีฮิวริสติกและวิธีของเบย์ในแบบสอบวัดผลสัมฤทธิ์และแบบสอบความถนัด ผลจากการศึกษาพบว่า แบบสอบวัดผลสัมฤทธิ์ที่ประมาณค่าพารามิเตอร์ของข้อสอบด้วยวิธีแมกซิมัมไลค์ลิฮูด มีประสิทธิภาพสูงสุด ในกลุ่มผู้เข้าสอบที่มีความสามารถสูง รองลงมาคือ แบบสอบที่ประมาณค่าด้วยวิธีของเบย์และวิธีฮิวริสติก ตามลำดับ ส่วนในกลุ่มผู้เข้าสอบที่มีความสามารถปานกลางและต่ำนั้น แบบสอบที่ประมาณค่าด้วยวิธีของเบย์ มีประสิทธิภาพสูงสุด รองลงมาคือ แบบสอบที่ประมาณค่าด้วยวิธีแมกซิมัมไลค์ลิฮูดและวิธีฮิวริสติกตามลำดับ สำหรับค่าความตรงร่วมสมัยที่เป็นผลมาจากการประมาณค่าความสามารถของผู้เข้าสอบทั้ง 3 วิธี แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.001 โดยวิธีของเบย์ ให้ค่าความร่วมสมัยสูงที่สุด รองลงมาคือ วิธีฮิวริสติก หรือวิธีแมกซิมัมไลค์ลิฮูด

วินัย วงศ์ฤทัยวัฒนา (2532) ได้ศึกษาเปรียบเทียบผลการประมาณค่าพารามิเตอร์ของโมเดลโลจิสติกแบบสองพารามิเตอร์ระหว่างวิธีของเบย์กับวิธีของแมกซิมัมไลค์ลิฮูด ผลจากการศึกษาพบว่า

1. ค่าอำนาจจำแนก (a) ที่ประมาณค่าด้วยวิธีของเบย์และวิธีแมกซิมัมไลค์ลิฮูดมีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

2. ค่าความยาก (b) ที่ประมาณค่าด้วยวิธีของเบย์และวิธีแมกซิมัมไลค์ลิฮูดมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ($p<0.05$)

3. ค่าอำนาจจำแนก (a) ที่ประมาณค่าด้วยวิธีของเบย์และวิธีแมกซิมัมไลค์ลิฮูดมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ ($p<0.01$)

4. ค่าความยาก (b) ที่ประมาณค่าด้วยวิธีของเบย์และวิธีแมกซิมัมไลค์ลิฮูดมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ ($p<0.01$)

สิทธิชัย หาญสมบัติ (2533, หน้า ซ) ศึกษาประสิทธิผลของแบบทดสอบแบบเทเลอร์บางรูปแบบในการตัดสินความรอบรู้ในการเรียนวิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 ผลจากการศึกษาพบว่า แบบทดสอบแบบเทเลอร์ทั้ง 3 รูปแบบ คือ แบบปิรามิดที่ใช้ขนาดขั้นคงที่ แบบปิรามิดที่มีสองข้อในแต่ละขั้นและแบบเปลี่ยนระดับ ไม่ว่าจะมีจำนวนข้อที่นักเรียนต้องตอบเป็น 6 ข้อ หรือ 10 ข้อ ต่างก็ให้ผลการตัดสินที่มีความเที่ยงตรงอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ผลการตัดสินจากแบบทดสอบแบบเทเลอร์ทั้ง 3 รูปแบบที่แต่ละรูปแบบมีจำนวนข้อที่นักเรียนต้องตอบ 6 ข้อ ให้ผลการตัดสินแตกต่างจากการตัดสินตามปกติของครู แต่ไม่แตกต่างจากการทดสอบความรอบรู้แบบเทเลอร์ตามยุทธวิธีของเบย์อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนผลการตัดสินจากแบบทดสอบแบบเทเลอร์ทั้ง 3 รูปแบบที่แต่ละรูปแบบมีจำนวนข้อที่นักเรียนต้องตอบ 10 ข้อ มีความแตกต่างจากผลการตัดสินตามปกติของเฉพาะในรูปแบบเปลี่ยนระดับ ส่วนระหว่างแบบทดสอบแบบเทเลอร์ที่มีรูปแบบเดียวกันแต่มีจำนวนข้อที่นักเรียนต้องตอบต่างกันให้ผลการตัดสินที่ไม่แตกต่างกัน

รังสรรค์ มณีเล็ก (2540, หน้า172-174) ได้ศึกษาผลของตัวแปรบางตัวต่อความเที่ยงตรงเชิงสภาพและจำนวนข้อที่ใช้ในการทดสอบแบบเทเลอร์ด้วยคอมพิวเตอร์ กับนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 1 ปีการศึกษา 2539 จำนวน 1,620 คน เครื่องมือที่ใช้ในการวิจัยคือ คลังข้อสอบ เรื่องเศษส่วน แบบทดสอบวัดความสามารถทางคณิตศาสตร์ และโปรแกรมคอมพิวเตอร์สำหรับดำเนินการสอบแบบเทเลอร์ด้วยคอมพิวเตอร์ ผลจากการศึกษาพบว่า วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบมีผลต่อความเที่ยงตรงเชิงสภาพของการทดสอบ ส่วนผลต่อตัวแปรตามอีกตัวแปรหนึ่งนั้น อันได้แก่ ค่าความยากของข้อสอบข้อแรก เกณฑ์การคัดเลือกข้อสอบจากคลังข้อสอบ วิธีการประมาณค่าความสามารถของผู้สอบ เกณฑ์ยุติการทดสอบ และความสามารถของผู้สอบต่างก็มีผลต่อจำนวนข้อสอบที่ใช้ในการทดสอบ

2. งานวิจัยในต่างประเทศ

แกรริสันและบอมการ์เตน (Garrison & Baumgarten, 1986, unpaged อ้างใน รังสรรค์ มณีเล็ก, 2540, หน้า 89-90) ได้ทดลองใช้การทดสอบแบบเทเลอร์ด้วยคอมพิวเตอร์ เพื่อวัดทักษะ

คณิตศาสตร์ของนักศึกษา จำนวน 60 คน โดยมีการเปรียบเทียบความคงเส้นคงวาของค่าความสามารถที่ประมาณได้จากการทดสอบแบบเดิมและการทดสอบแบบเทเลอร์ด้วยคอมพิวเตอร์ ผลจากการศึกษาพบว่า การทดสอบแบบเทเลอร์ด้วยคอมพิวเตอร์มีประสิทธิภาพในการวัดสูงกว่าการทดสอบแบบเดิม ความยาวของแบบทดสอบลดลงจากเดิม และผู้สอบมีเจตคติที่ดีต่อการสอบสูงกว่าการทดสอบแบบเดิม

- ลี (Lee, 1986, unpaged อ้างใน รังสรรค์ มณีเล็ก, 2540, หน้า 89) ได้ศึกษาความเป็นไปได้ของการนำการทดสอบแบบเทเลอร์ด้วยคอมพิวเตอร์มาใช้กับการวัดผลสัมฤทธิ์ทางการเรียนโดยสร้างแบบทดสอบ เอ็ม เอ เอ ที (Microcomputerized adaptive achievement test, MAAT) จากคลังข้อสอบที่สร้างโดยแผนกคลังข้อสอบของ พี อี ที เอ (Pitt Educational Testing Aids, PETA) กลุ่มตัวอย่างได้แก่ นิสิตระดับบัณฑิตศึกษาจำนวน 3 ห้องเรียนที่กำลังเรียนวิจัยการศึกษาเบื้องต้น ใช้วิธีการทดสอบแบบเทเลอร์ 2 วิธี ได้แก่ วิธี เอส พี อาร์ ที ของวอลด์ (Walds's Sequential Probability Ratio Test) และแบบเฟล็กซิเลเวลของลอร์ด (Lord's Flexilevel) นอกจากนั้นยังมีการวัดเจคติของผู้สอบต่อการทดสอบและสัมภาษณ์ผู้ดำเนินการสอบด้วย ผลจากการศึกษาพบว่า มีความเป็นไปได้ที่จะนำการทดสอบแบบ เอ็ม เอ เอ ที มาใช้วัดผลสัมฤทธิ์ทางการเรียน คะแนนจากการทดสอบด้วยคอมพิวเตอร์มีความสัมพันธ์ทางบวกกับผลการทดสอบกลางเทอม และผู้สอบมีเจตคิที่ดีต่อการทดสอบแบบ เอ็ม เอ เอ ที

แฮนกิ้น (Hankins, 1987, p. 3031-A อ้างใน ต่าย เซี่ยงฉี, 2534, หน้า 59) ได้ศึกษาการทดสอบแบบเทเลอร์โดยใช้ยุทธวิธีของเบย์ (Bayesian Adaptive Test) 2 วิธี คือ แบบกำหนดความสามารถเริ่มต้นของผู้สอบคงที่ (Fixed Entry) กับแบบกำหนดความสามารถเริ่มต้นของผู้สอบแปรผัน (Variable Entry) ผลจากการศึกษาพบว่า ทั้ง 2 วิธี ไม่มีความลำเอียง (Bias) เมื่อประมาณค่าความสามารถของผู้สอบที่มีความสามารถปานกลาง แต่จะมีความลำเอียง เมื่อผู้สอบมีความสามารถค่อนไปทางเก่ง หรืออ่อนมาก ๆ ทั้ง 2 วิธี ให้ค่าอินฟอร์เมชั่น (Information) มีการแจกแจงเป็นรูปสมมาตร (Symmetry) ที่มีโค้งค่อนข้างโด่ง และการทดสอบแบบกำหนดความสามารถเริ่มต้นของผู้สอบแปรผัน (Variable Entry) จะต้องใช้จำนวนข้อสอบมากกว่าแบบกำหนดความสามารถเริ่มต้นของผู้สอบคงที่ (Fixed Entry)

โฮ (Ho, 1989, p. 421-A อ้างใน ต่าย เซี่ยงฉี, 2534, หน้า 60) ได้ใช้ไมโครคอมพิวเตอร์ศึกษาเปรียบเทียบยุทธวิธีการใช้ อแดบติฟ เทสติ้ง (Adaptive Testing) 3 วิธี คือ ยุทธวิธีของเบย์

(Bayesian) โมดัลเบย์ (Modal Bayesian) และวิธีแมกซิมัมไลค์ลิฮูด (Maximum Likelihood) ยุทธวิธีทั้ง 3 นี้ได้เปรียบเทียบโดยใช้ขนาดของคลังข้อสอบ (Bank Sizes) 3 ขนาด คือ ขนาด 86 ข้อ 71 ข้อ และ 56 ข้อ ชนิดของคลังข้อสอบ (Bank Type) 3 แบบ คือ แบบที่มีข้อสอบแบบสุ่ม แบบที่มีเฉพาะข้อสอบที่มีค่าอำนาจจำแนกสูง และแบบที่มีเฉพาะข้อสอบที่ง่าย ๆ ส่วนระดับความสามารถของผู้สอบ มี 3 ระดับ คือ สูง ปานกลาง และต่ำ การวิเคราะห์ข้อมูลโดยใช้การวิเคราะห์ความแปรปรวน 3 ทาง (Three ways ANOVA with repeated measures) ผลจากการศึกษาพบว่า

1. ยุทธวิธีโมดัลเบย์ (Modal Bayesian) จะมีประสิทธิภาพ (Efficient) สูงสุด ยุทธวิธีของเบย์ (Bayesian) จะให้ค่าความเชื่อมั่น (Reliable) สูงสุด และวิธีแมกซิมัมไลค์ลิฮูด (Maximum Likelihood) จะให้ค่าไม่คงที่ (Inconsistent) ภายใต้การทดสอบในสถานการณ์ต่าง ๆ

2. การเลือกขนาดคลังข้อสอบ ดูเหมือนว่าจะส่งผลน้อยมาก ยกเว้นในกลุ่มผู้สอบที่มีความสามารถต่ำ เมื่อใช้ข้อสอบเริ่มต้นที่มีค่าความยากสูงกว่าความสามารถที่แท้จริงของผู้สอบ

3. ถ้าใช้ข้อสอบเริ่มต้นที่มีความยากเท่ากับ หรือน้อยกว่าความสามารถที่แท้จริงของผู้สอบแล้ว ความถูกต้องในการประมาณค่าจะมีมากขึ้น

4. มีปฏิสัมพันธ์ (Interaction) ระหว่างชนิดของคลังข้อสอบ กับยุทธวิธีของ อแดบติฟเทสติ้ง และระหว่างชนิดของคลังข้อสอบกับระดับความสามารถของผู้สอบ

5. ความแตกต่างระหว่างคุณลักษณะทางสถิติของการเลือกข้อสอบ กับไม่เลือกข้อสอบมีความสัมพันธ์กันต่ำ

อยาลาและคนอื่น ๆ (Ayala & others, 1990, unpaged, อ้างใน รังสรรค์ มณีเล็ก, 2540, หน้า 89) ได้เปรียบเทียบการทดสอบโดยใช้แบบทดสอบเฟล็กซิเลเวลกับการทดสอบแบบเทเลอร์ด้วยคอมพิวเตอร์ที่ประมาณค่าความสามารถของผู้สอบด้วยวิธีของเบส์ ใช้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถเป็นเกณฑ์ในการยุติการสอบ โดยเปรียบเทียบที่จุดต่าง ๆ กันของค่าของความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ ผลจากการศึกษาพบว่า ค่าความสามารถของผู้สอบที่ได้จากการทดสอบสองวิธีมีความถูกต้องแม่นยำไม่ต่างกัน