# บทที่ 2

# เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยเพื่อพัฒนาระบบจำแนกความสามารถผู้เรียนโดยใช้คลังข้อสอบ ผู้วิจัยได้ทำการศึกษารวบรวมเอกสารและงานวิจัยที่เกี่ยวข้อง เสนอเป็นลำดับดังนี้

1. ทฤษฎีการตอบสนองข้อสอบ (IRT)
   - 1.1 หลักการของทฤษฎีการตอบสนองข้อสอบ
   - 1.2 ค่าตัวแปรเสริมการตอบสนองข้อสอบ
   - 1.3 รูปแบบของทฤษฎีการตอบสนองข้อสอบ
   - 1.4 การนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้
2. คลังข้อสอบและการทดสอบแบบปรับเหมาะ
   - 2.1 ความหมายของคลังข้อสอบ
   - 2.2 ประโยชน์ของคลังข้อสอบโดยใช้เครื่องคอมพิวเตอร์
   - 2.3 ความหมายของการทดสอบแบบปรับเหมาะ
   - 2.4 หลักการของการทดสอบแบบปรับเหมาะ
   - 2.5 ประเภทของการทดสอบแบบปรับเหมาะ
   - 2.6 วิธีดำเนินการทดสอบแบบปรับเหมาะ
   - 2.7 วิธีประมาณค่าความสามารถของผู้สอบ
3. การจัดกลุ่มข้อมูล
   - 3.1 ประเภทการแบ่งกลุ่มข้อมูล
   - 3.2 อัลกอริทึมเคมีน (K-Means Algorithm)
   - 3.3 การวัดประสิทธิภาพการแบ่งกลุ่มข้อมูล
4. การสร้างและพัฒนาโปรแกรมคอมพิวเตอร์
   - 4.1 วงจรการพัฒนาระบบ
   - 4.2 ประสิทธิภาพของระบบ
5. งานวิจัยที่เกี่ยวข้อง
   - 5.1 งานวิจัยที่เกี่ยวข้องกับการทดสอบแบบปรับเหมาะ
   - 5.2 งานวิจัยที่เกี่ยวข้องกับการจัดกลุ่มข้อมูล

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY

## ทฤษฎีการตอบสนองข้อสอบ (IRT)

### 1. หลักการของทฤษฎีการตอบสนองข้อสอบ

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

### 2. ค่าตัวแปรเสริมตามทฤษฎีการตอบสนองข้อสอบ

ค่าตัวแปรเสริมแบ่งออกเป็น 2 ชนิด คือ ค่าตัวแปรเสริมของผู้สอบ และค่าตัวแปรเสริมของข้อสอบ (Hambleton & Cook, 1977, unpaged อ้างใน ชูเกียรติ ละอองแก้ว. 2537 : 13)

2.1 ค่าตัวแปรเสริมของผู้สอบ (Examinee Parameter) คือ ระดับความสามารถของผู้สอบ ($\theta$) ที่คำนวณจากคะแนนจริง (True Score) โดยปกติ $\theta$ จะมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ ถ้า $\theta$ เป็นลบแสดงว่าผู้สอบมีความสามารถต่ำ ถ้า $\theta$ มีค่าเป็นบวก แสดงว่า ผู้สอบมีความสามารถสูง

2.2 ค่าตัวแปรเสริมของข้อสอบ (Item Parameters) ประกอบด้วย ค่าความยาก ค่าอำนาจจำแนก และค่าการเดา ซึ่งมีรายละเอียดดังนี้

2.2.1 ค่าความยาก (b) เป็นค่าที่แสดงระดับความสามารถที่จุดเปลี่ยนโค้งในโค้งคุณลักษณะข้อสอบในกรณีที่ไม่มีการเดา ค่าความยากก็คือ $\theta$ ณ จุดที่ $P_x(\theta) = 0.5$ ดังนั้นค่าความยากจึงเป็นมาตรชนิดเดียวกันกับ $\theta$ โดยปกติแล้ว b จะมีค่าตั้งแต่ $-\infty$ ถึง $+\infty$ ถ้ามีค่าเป็นลบมากแสดงว่าข้อสอบนั้นง่ายมาก ถ้ามีค่าเป็นบวกมากแสดงว่าข้อสอบนั้นยากมาก

มหาวิทยาลัยราชภัฏมหาสารคาม RAJABHAT MAHASARAKHAM UNIVERSITY

2.3.2 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

2.3.3 ค่าการเดา(c) เป็นค่าที่แสดงถึงโอกาสในการตอบข้อสอบใด ๆ ได้ถูกโดยที่ผู้สอบไม่มีความรู้ หรือมีความรู้น้อยมาก แต่สามารถตอบข้อสอบข้อนั้นได้ถูก ข้อสอบข้อใดที่มีค่าการเดา (c) มากกว่า0.30 ขึ้นไปแสดงว่าข้อสอบข้อนั้นๆไม่ดีควรจะพิจารณาตัดออก

**3. รูปแบบของทฤษฎีการตอบสนองข้อสอบ** (อ้างในนุจรี สุภีสุทธิ์. 2541 : 9-10)

รูปแบบของทฤษฎีการตอบสนองข้อสอบได้พัฒนาในรูปแบบต่างๆ ทำให้เกิดรูปแบบเฉพาะขึ้นหลายรูปแบบ ซึ่งแต่ละรูปแบบจะแตกต่างกันที่ฟังก์ชันทางคณิตศาสตร์ และจำนวนตัวแปรเสริม ที่ใช้อธิบายโค้งลักษณะข้อสอบ เช่น รูปแบบโค้งปกติสะสม (Normal Ogive Model) รูปแบบเส้นตรง (The Linear Model) และรูปแบบโลจีสติก (Logistic Model) จากรูปแบบทั้งสามนี้ จะกล่าวถึงเฉพาะรูปแบบโลจีสติก เพราะเป็นรูปแบบที่นำมาใช้กันมาก ซึ่งรูปแบบโลจีสติกมีรูปแบบย่อยอยู่ 3 รูปแบบคือ

3.1รูปแบบที่ใช้ตัวแปรเสริมตัวเดียวหรือรูปแบบราส์ช (One Parameter Logistic Model or Rasch Model) ปี ค.ศ. 1966 ราส์ช (Rasch ) ได้พัฒนาทฤษฎีการตอบสนองข้อสอบ (IRT) และได้เสนอรูปแบบนี้ซึ่งเรียกว่า รูปแบบของราส์ช (Rasch Model) โดยรูปแบบราส์ชเป็นฟังก์ชั่นที่สามารถอธิบายได้ด้วยตัวแปรเสริมเพียงตัวเดียวคือ ค่าความยาก (b) และรูปแบบนี้ตรงกับรูปแบบที่ใช้ตัวแปรเสริมที่เบริ์นบอม (Birnbaum)ได้พัฒนาขึ้นในปีค.ศ.1968 ข้อตกลงของรูปแบบนี้คือข้อสอบทุกข้อมีค่าอำนาจจำแนก (a) เท่ากัน จึงใช้ค่าเฉลี่ยของค่าอำนาจจำแนกแทนค่าอำนาจจำแนกของข้อสอบ และข้อสอบทุกข้อมีโอกาสการเดา (c) เท่ากับ 0

3.2 รูปแบบที่ใช้ตัวแปรเสริมสองตัว (Two Parameters Model) ในปี ค.ศ. 1968 เบิร์นบอม (Birnbaum) ได้เสนอรูปแบบโค้งลักษณะข้อสอบซึ่งใช้ค่าตัวแปรเสริม 2 ตัว คือ ค่าความยาก (b) และค่าอำนาจจำแนก (a) ข้อตกลงของรูปแบบนี้คือ การตอบข้อสอบทุกข้อที่ไม่มีการเดา ($c = 0$)

3.3 รูปแบบที่ใช้ตัวแปรเสริมสามตัว (Three Parameters Logistic Model) เป็นรูปแบบที่ดัดแปลงมาจากรูปแบบที่ใช้ตัวแปรเสริมสองตัวโดยเพิ่มค่าการเดา(c) เข้าไปเป็นค่าตัวแปรที่สาม

**4. การนำเอาทฤษฎีการตอบสนองข้อสอบมาประยุกต์ใช้**

4.1การสร้างคลังข้อสอบ(Itembank)โดยการนำเอาทฤษฎีการตอบสนองข้อสอบ (IRT) มาใช้ในการวิเคราะห์และสร้างคลังข้อสอบ ทำให้เกิดผลดีที่เห็นเด่นชัด 2 ประการ คือ ประการแรกด้านค่าตัวแปรเสริมของข้อสอบ (Item Parameters) อันได้แก่ ค่าความยาก (b) ค่าอำนาจจำแนก (a) และค่าการเดา (c) ที่วิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบ (IRT) มีคุณสมบัติไม่แปรเปลี่ยน(Invariant) ไปตามกลุ่มตัวอย่างหรือตัวผู้สอบประการที่สองด้านรายงานคุณภาพของข้อสอบในทฤษฎีการตอบสนองข้อสอบ (IRT) จะรายงานคุณภาพของข้อสอบในรูปของอินฟอร์เมชั่น (Information) ซึ่งสามารถรายงานได้ทั้งเป็นรายข้อ (Item Information) และทั้งฉบับ (Test Information) ค่าอินฟอร์เมชั่นนี้ จะเป็นตัวบ่งชี้ถึงความถูกต้องแม่นยำ (Accuracy) ในการประมาณค่าซึ่งสามารถใช้แทนค่าความเชื่อมั่น (Reliability) และค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) ในทฤษฎีการทดสอบแบบดั้งเดิม หรือแบบประเพณีนิยม (Classical Test Theory : CTT) ได้ (Hambleton & Swamenathan, 1985 , p. 236 อ้างใน ต่าย เชี่ยงฉี. 2534 : 34)

4.2 การทดสอบแบบปรับเหมาะ (Adaptive Testing) หมายถึง การทดสอบที่มีการจัดข้อสอบให้มีความยากที่เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคลนับเป็นการนำเอาทฤษฎีการตอบสนองข้อสอบ(IRT) มาประยุกต์ใช้ได้อย่างมีประสิทธิภาพ (Urry, 1977, p. 181) โดยการนำเอาคุณสมบัติความไม่แปรเปลี่ยน (Invariant) ของค่าตัวแปรเสริมของข้อสอบและความสามารถของผู้สอบมาใช้ การทดสอบแบบปรับเหมาะนั้นผู้สอบแต่ละคนจะได้รับชุดข้อสอบที่แตกต่างกันตามระดับความสามารถของผู้สอบ บางคนอาจได้รับข้อสอบชุดที่ยาก และอีกบางคนอาจจะได้รับข้อสอบชุดที่ง่ายแต่ก็ยังสามารถนำเอาความสามารถที่ประมาณค่าได้จากการทดสอบแบบปรับเหมาะมาเปรียบเทียบกันได้ทั้งนี้เพราะ ความสามารถที่ประมาณค่าได้ยังอยู่ในมาตราวัดเดียวกัน (Common Ability Scale) (Hambleton & Cook, 1977, p. 90-91 ; Hambleton & Swaminathan, 1985, p. 296 อ้างใน ต่าย เชี่ยงฉี. 2534 : 35)

**คลังข้อสอบและการทดสอบแบบปรับเหมาะ**

ความรู้เรื่องคลังข้อสอบได้มีมาช้านานโดยเริ่มแพร่หลายมาจากประเทศตะวันตกและอเมริกาสำหรับประเทศไทยนั้นได้มีการแพร่เข้ามาเมื่อเรารับการศึกษาตามแนวทางใหม่เข้ามาใช้ในการวัดผลประเมินผลทางการศึกษา ทำให้ได้รับความรู้ที่สำคัญในเรื่องการสร้างข้อสอบที่ดีมีความเชื่อถือได้ ความแม่นตรงสูง มีอำนาจจำแนกในการแยกคนเก่งคนไม่เก่งของข้อสอบมีอำนาจจำแนก ความถนัด ไม่ถนัด ความรู้ ความไม่รู้ ด้วยการอาศัยเทคนิคการออกข้อสอบและ

แบบแผนการวัดประเมินผลแบบต่าง ๆ ได้ และข้อสอบเหล่านั้นสามารถนำมาใช้ซ้ำแล้วซ้ำอีกได้ โดยยังคงประสิทธิภาพของข้อสอบที่ดีอยู่เสมอ สำหรับในประเทศไทยนั้นความรู้เรื่องคลังข้อสอบได้แพร่เข้ามาซึ่งปรากฏในประเทศไทยชัดเจนมาก ประมาณหลังปี พ.ศ. 2500 และคำว่า Item Bank ประเทศไทยนิยมเรียกกันว่าธนาคารข้อสอบ หรือคลังข้อสอบ (อุทัย บุญประเสริฐ. 2531 : 4)

**1. ความหมายของคลังข้อสอบ**

อุทัย บุญประเสริฐ (2532 : 5-9) ได้ให้ความหมาย คลังข้อสอบหรือธนาคารข้อสอบว่า หมายถึง ที่รวมของข้อสอบ ซึ่งมีไว้เพื่อประโยชน์ในการใช้ข้อสอบเหล่านั้น ตามวัตถุประสงค์ของการทดสอบในโอกาสต่อไป

บุญชม ศรีสะอาด (2533 : 32 อ้างในอะโนทัย ลุนบุดดา. 2546 : 17) ได้ให้ความหมายของคลังข้อสอบไว้ว่า เป็นที่รวบรวมข้อสอบที่มีคุณภาพได้มาตรฐานเข้าไว้ด้วยกัน

อุทุมพร (ทองอุทัย) จามรมาน (2535 : 87) ได้ให้ความหมายคลังข้อสอบว่า เป็นที่เก็บข้อสอบ ที่ดีเข้าไว้ด้วยกัน เพื่อประโยชน์ในการวัดผลและเรียกใช้ในคราวต่อ ๆ ไป

สุพัฒน์ สุกมลสันต์ (2542 : 290) กล่าวว่า ธนาคารข้อสอบหรือคลังข้อสอบหมายถึงแหล่งเก็บรวบรวมข้อสอบที่มีการจัดเก็บและการใช้อย่างเป็นระบบ ในปัจจุบันนี้นิยมที่จะจัดเก็บข้อทดสอบไว้ในเครื่องคอมพิวเตอร์ เพราะว่าประหยัดเนื้อที่สำหรับการจัดเก็บมากกว่าและการค้นหาข้อทดสอบทำได้รวดเร็วกว่าการจัดเก็บลงในสื่ออย่างอื่น

จากเอกสารดังกล่าวข้างต้นสามารถสรุปได้ว่าคลังข้อสอบหมายถึงแหล่งรวบรวมข้อสอบที่มีคุณภาพ เป็นระบบ สามารถนำมาใช้ได้อย่างสะดวก เพื่อประโยชน์ในการวัดผล

คลังข้อสอบด้วยคอมพิวเตอร์ หมายถึง แหล่งรวบรวมข้อสอบ ที่มีคุณภาพ มีการบริหารจัดการจัดเก็บอย่างเป็นระบบด้วยคอมพิวเตอร์ สามารถสืบค้นได้สะดวกรวดเร็ว

**2. ประโยชน์ของคลังข้อสอบโดยใช้เครื่องคอมพิวเตอร์**

สุพัฒน์ สุกมลสันต์ (2542 : 290 - 291) ได้กล่าวถึงประโยชน์ของการจัดทำคลังข้อสอบโดยใช้คอมพิวเตอร์ดังนี้

2.1 ทำให้เนื้อหาของการทดสอบมีความเป็นไปได้มากขึ้นที่จะสอดคล้องสัมพันธ์กับจุดมุ่งหมายของรายวิชาและเนื้อหาของรายวิชา อันจะทำให้กระบวนของการเรียนการสอนรายวิชาต่าง ๆ มีความสมบูรณ์ยิ่งขึ้น

2.2 เป็นการพัฒนาข้อทดสอบให้มีคุณภาพที่ดีและน่าเชื่อถือมากยิ่งขึ้นอันจะทำให้ผลการทดสอบมีความน่าเชื่อถือมากขึ้น

2.3 เป็นการพัฒนาข้อทดสอบให้มีมาตรฐานสูงขึ้น เช่น มีความแม่นตรง และความเชื่อมั่นมากขึ้น ไม่มีอคติ หรือทำให้มีการเดาน้อยลง เป็นต้น

2.4 สามารถสร้างแบบทดสอบที่เหมาะสมกับกลุ่มผู้สอบเป้าหมาย (target examinees) ได้อย่างรวดเร็วและถูกต้องตามหลักการทดสอบ ทั้งแบบทดสอบสัมฤทธิผล แบบทดสอบสมิทธิผล (proficiency test) แบบทดสอบวินิจฉัย และแบบทดสอบคัดเลือก (screening test) เป็นต้น

2.5 สามารถใช้ข้อทดสอบข้อเดียวกันได้หลายครั้งและเหมาะสมกับจุดมุ่งหมายในการสอบแต่ละครั้งได้อย่างรวดเร็วซึ่งถือได้ว่าเป็นการประหยัดทั้งแรงคน การเงิน และสติปัญญา ได้เป็นอย่างดี

2.6 ทำให้ข้อทดสอบและแบบทดสอบมีความปลอดภัย (security) มากที่สุดจากปัญหาข้อสอบรั่วไหล

2.7 ทำให้กระบวนการทดสอบวิชาต่างๆ (หรือการทดสอบต่างๆ) มีความพร้อมเพรียงอยู่ตลอดเวลา และสามารถสร้างแบบทดสอบเพื่อการสอบได้ทุกเวลาตามที่ผู้บริหารการสอบต้องการ

2.8 ทำให้แบบทดสอบมีลักษณะคู่ขนานทั้งเชิงเนื้อหาและเชิงสถิติ (content and statistical parallel test forms) ได้ง่าย ทั้งนี้ เพื่อประโยชน์ในการเทียบคะแนน (score equating) ของแบบทดสอบต่างชุดกันแต่มีจุดประสงค์เหมือนกัน

2.9 เป็นการประหยัดเนื้อที่ของสถานที่สำหรับเก็บรักษาแบบทดสอบจำนวนมาก เพื่อนำไว้ใช้อีกหรือเพื่อการพัฒนาคุณภาพให้ดีขึ้น

2.10 เป็นประโยชน์ในการสร้างแบบทดสอบสำหรับการทดสอบรายบุคคลจำนวนมาก (tailored test หรือ adaptive test) ในอนาคตได้ด้วย

2.11 เอื้ออำนวยความสะดวกสบายให้แก่ครูอาจารย์ที่ประสงค์จะใช้แบบทดสอบสำหรับการเรียนการสอนรายวิชาต่างๆ ในกรณีฉุกเฉิน

2.12 ช่วยกระตุ้นและเปิดโอกาสให้ครู อาจารย์ได้ใช้ความรู้และความสามารถในการสร้าง ปรับปรุง และพัฒนาข้อทดสอบหรือแบบทดสอบให้มีคุณภาพดียิ่งขึ้นได้อย่างมากมาย ทั้งนี้ เพื่อให้ครู อาจารย์ได้มีประสบการณ์ตรงในเรื่องต่างๆ ดังกล่าวได้เป็นอย่างดีและเป็นระยะเวลานานๆ ซึ่งจะทำให้เกิดมีความรู้และความชำนาญได้เป็นอย่างดีในภายหลัง

มหาวิทยาลัยราชภัฏพระนครศรีอยุธยา
RAJABHAT MAHASARAKHAM UNIVERSITY

**3. ความหมายของการทดสอบแบบปรับเหมาะ**

การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ เป็นการใช้ข้อสอบจากคลังข้อสอบมาสร้างเป็นแบบสอบ แบบสอบแต่ละชุดมีการออกแบบให้เหมาะสมกับความสามารถ ของผู้สอบแต่ละคน ซึ่งเรียกว่า "Tailored Test" คำนี้ใช้เป็นครั้งแรก โดย W.W. Turnbull (1951) (Cited in Lord, 1980) ต่อมา ไวส์ ได้เสนอให้ใช้คำว่า "Adaptive Test" (Weiss, 1974) การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ (Adaptive Test) หมายถึง การทดสอบที่ใช้แบบสอบต่างชุดกันสำหรับผู้สอบต่างกัน โดยมีการคัดเลือกข้อสอบที่มีความยากเหมาะสมกับระดับความสามารถของผู้สอบ ( ศิริชัย กาญจนวาสี. 2545 : 175-175)

**4. หลักการของการทดสอบแบบปรับเหมาะ**

ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ หลักการคัดเลือกข้อสอบสำหรับแต่ละบุคคลอยู่บนพื้นฐานของผลการตอบข้อสอบข้อที่ผ่านมาของผู้สอบ เมื่อผู้สอบข้อสอบข้อเริ่มต้นหรือชุดแรก (ขึ้นอยู่กับการออกแบบ) จากคลังข้อสอบแล้ว จะมีการวิเคราะห์ระดับความสามารถหรือประเมินความสามารถของผู้สอบเบื้องต้น เพื่อคัดเลือกข้อสอบต่อไปที่มีค่าความยากและอำนาจจำแนกเหมาะสมที่จะใช้วัดระดับความสามารถของผู้สอบประมาณระดับความสามารถของผู้สอบใหม่ จากนั้นก็จะเลือกข้อสอบที่เหมาะสมข้อต่อไป โดยอาศัยหลักการที่ว่าถ้าการทำข้อสอบที่ผ่านมาถูก ข้อถัดไปจะยากขึ้น แต่ถ้าทำข้อสอบที่ผ่านมาผิด ข้อถัดไปจะง่ายลง กระบวนการนี้จะดำเนินการต่อไปเรื่อย จนสามารถประมาณระดับความสามารถของผู้สอบได้อย่างน่าเชื่อถือ ตามเกณฑ์ที่กำหนดไว้ การทดสอบก็จะยุติลง (ศิริชัย กาญจนวาสี. 2545 : 177)

**5. ประเภทของการทดสอบแบบปรับเหมาะ**

ไวส์, แฮมเบิลตัน, และสวามินาธาน (Weiss, 1974, pp. 78-110; Hambleton & Swaminathan, 1985, p.297 อ้างใน ต่าย เซี่ยงฉี่.2534 : 36 ; ศิริชัย กาญจนวาสี. 2538 : 5) ได้จำแนกประเภทของการทดสอบแบบปรับเหมาะ โดยพิจารณาถึงยุทธวิธีที่ใช้ในการคัดเลือกข้อสอบ ซึ่งจำแนกเป็นการทดสอบแบบปรับเหมาะที่ใช้ยุทธวิธีสองขั้นตอนและยุทธวิธีหลายขั้นตอน

5.1 ยุทธวิธีสองขั้นตอน (Two stage Strategies)

เป็นการทดสอบแบบปรับเหมาะที่แบ่งการทดสอบออกเป็นสองขั้นตอน ประกอบด้วยขั้นตอนแรกเป็นการสอบกำหนดทิศทาง และขั้นตอนที่สองเป็นการสอบวัดผล

5.2 ยุทธวิธีหลายขั้นตอน (Multi Stage Strategies)

มหาวิทยาลัยราชภัฏมหาสารคาม RAJABHAT MAHASARAKHAM UNIVERSITY

เป็นการทดสอบแบบปรับเหมาะที่แบ่งการทดสอบออกเป็นหลายขั้นตอน โดยมีการจัดโครงสร้างของขั้นตอนจัดเรียงข้อสอบ การเลือกข้อสอบ และการยุติการทดสอบ รูปแบบแบ่งออกเป็น

5.2.1 ยุทธวิธีหลายขั้นตอนแบบแยกทางคงที่ (Fixed Branching)

5.2.2 ยุทธวิธีหลายขั้นตอนแบบแยกทางแปรผัน (Variable Branching)

5.2.2.1 ยุทธวิธีของเบย์ (Bayesian Strategies)

5.2.2.2 ยุทธวิธีความเป็นไปได้สูงสุด (Maximum Likelihood Strategies)

สำหรับการวิจัยครั้งนี้ ผู้วิจัยได้เลือกศึกษาเกี่ยวกับการทดสอบแบบปรับเหมาะยุทธวิธีหลายขั้นตอนแบบแยกทางแปรผันโดยใช้ยุทธวิธีของเบย์

**6. วิธีดำเนินการทดสอบแบบปรับเหมาะ**

วิธีดำเนินการทดสอบแบบปรับเหมาะมีลำดับขั้นตอนดังนี้ (ศิริชัย กาญจนวาสี. 2545 : 181-182) หลังจากที่ผู้สอบลงทะเบียนในระบบการทดสอบและกดรหัสผ่านบนแป้นพิมพ์คอมพิวเตอร์ เครื่องคอมพิวเตอร์จะแนะนำวิธีการตอบข้อสอบ มีการซักซ้อมและทดลองตอบจนเข้าใจดีแล้ว กระบวนการทดสอบก็จะเริ่มขึ้น โดยสุ่มข้อสอบข้อแรก (Initial item) จากคลังข้อสอบ ตามหลักทั่วไปข้อสอบข้อแรกจะมีความยากปานกลาง หรือเป็นข้อสอบที่คัดเลือกให้เหมาะกับผู้สอบตามสารสนเทศที่ได้รับ จากผลการตอบจะมีการประมาณค่าความสามารถเบื้องต้นของผู้สอบ ถ้าผลการทดสอบออกมาถูก ข้อถัดไปจะยากขึ้น แต่ถ้าผลการตอบออกมาผิด ข้อถัดไปจะง่ายลง กระบวนการ การทดสอบจะทำซ้ำเช่นนี้ไปเรื่อยจนบรรลุเกณฑ์ที่กำหนดไว้ คือการประมาณค่าความสามารถที่แท้จริงของผู้สอบนั้นได้อย่างน่าเชื่อถือ มีความคลาดเคลื่อนต่ำ โปรแกรมก็จะยุติการทำงานบันทึกผลการสอบ รายงานผลความสามารถที่แท้จริงของผู้สอบบนหน้าจอคอมพิวเตอร์และพิมพ์รายงานไปยังผู้เกี่ยวข้อง

**7. วิธีประมาณค่าความสามารถของผู้สอบ**

ได้มีผู้เสนอวิธีการประมาณค่าความสามารถของผู้สอบไว้หลายวิธีการ พอสรุปดังนี้ (ต่าย เชี่ยงฉี่. 2534 : 49-55)

7.1 ประมาณค่าความสามารถของผู้สอบตามค่าความยากของข้อสอบข้อที่ยากที่สุดที่ทำถูก โดยดูจากการตอบข้อสอบในการทดสอบแบบปรับเหมาะที่ผู้สอบตอบมาทั้งหมด ข้อใดมีค่าความยากสูงสุดที่ผู้สอบทำถูก ถือว่าเป็นค่าความสามารถของผู้สอบคนนั้น โดยค่าความยากที่ว่านี้ ควรจะเป็นค่าความยากที่วิเคราะห์มาจากทฤษฎีการตอบสนองข้อสอบ (IRT)

7.2 ประมาณค่าความสามารถของผู้สอบตามค่าความยากของข้อสอบทุกข้อที่ทำถูก เป็นการประมาณค่าความสามารถของผู้สอบ โดยนำค่าความยากของข้อสอบจากการทดสอบแบบปรับเหมาะทุกข้อที่ผู้สอบทำถูกมาคำนวณหาค่าเฉลี่ย ค่าเฉลี่ยความยากนี้ ถือเป็นค่าความสามารถของผู้สอบคนนั้น

7.3 ประมาณค่าความสามารถของผู้สอบตามค่าความยากของข้อสอบข้อสุดท้ายที่ทำ ไม่คำนึงถึงว่าขั้นสุดท้ายจะตอบถูกหรือไม่ก็ตามจะถือว่าค่าความยากจากการทดสอบแบบปรับเหมาะในขั้นสุดท้าย ที่ผู้สอบได้ทำ คือค่าความสามารถของผู้สอบคนนั้น

7.4 ประมาณค่าความสามารถของผู้สอบโดยใช้หลักการความเป็นไปได้สูงสุด (Maximum Likelihood) การใช้หลักการความเป็นไปได้สูงสุดนี้ ยังมีวิธีการย่อยมีอีกหลายวิธี แต่วิธีที่นิยมใช้กันมากคือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) การประมาณค่าโดยวิธีนี้ มีข้อจำกัดในกรณีที่ผู้สอบตอบข้อสอบถูกหรือผิดหมดจะไม่สามารถประมาณค่าความสามารถของผู้สอบคนนั้นได้

7.5 การประมาณค่าความสามารถของผู้สอบนี้ใช้วิธีการของเบย์ ซึ่งนักสถิติชื่อโรเจอร์ เจ โอเว่น (Roger J. Owen) เป็นผู้เสนอขึ้น เพื่อใช้ในการประมาณค่าความสามารถของผู้สอบ ที่ปรับใหม่ เบย์เซี่ยน อัพเดทติ้ง (Bayesian Updating) ซึ่งสามารถประมาณค่าความสามารถของผู้สอบได้ค่อนข้างคงที่ (Owen, 1969, 1975, pp. 351-356; Thorndike, 1982, p. 303;citing Owen, 1975, unpaged อ้างในต่ายเซี่ยงฉี่, 2534 : 50-52) โดยมีสูตรคำนวณดังนี้

ในกรณีที่ตอบข้อสอบถูก

$$\theta_{m+1} = \theta_m + (1-c_i)\left[\frac{\sigma_m^2}{\sqrt{\frac{1}{a_i^2}+\sigma_m^2}}\right]\left[\frac{\phi(D)}{A}\right]$$

และ

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

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY

ในกรณีที่ตอบข้อสอบผิด

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

และ

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

เมื่อ $\phi(D)$ คือค่าออร์ดิเนท (Ordinate) ของโค้งปกติ ณ จุด D

$\Phi(D)$ คือพื้นที่ใต้โค้งปกติจากค่า D ต่ำสุดจนถึงจุด D

$$D = \frac{b_i-\theta_m}{\sqrt{\frac{1}{a_g^2}+\sigma_m^2}}$$

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

$$SEE_{m+1} = \sqrt{\sigma_{m+1}^2}$$

$\theta_m$ = ความสามารถของผู้สอบที่ประมาณค่าได้ก่อนตอบข้อสอบข้อที่ $m+1$ โดยปกติถ้าเป็นการประมาณค่าครั้งแรก ที่ยังไม่ทราบค่าความสามารถของผู้สอบ จะให้ $\theta_m = 0.000$

$\sigma_m^2$ = ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบเมื่อตอบข้อสอบข้อที่ m โดยปกติถ้าเป็นการประมาณค่าครั้งแรกที่ยังไม่ทราบค่าความแปรปรวน จะให้ $\sigma_m^2 = 1.000$

$SEE_{m+1}$ = ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ เมื่อตอบข้อสอบข้อที่ m ถ้าเป็นการประมาณค่าครั้งแรก จะให้ $SEE_m = 1.000$

$\theta_{m+1}$ = ความสามารถของผู้สอบที่ประมาณได้เมื่อตอบข้อสอบข้อที่ $m+1$แล้ว

$\sigma_{m+1}^2$ = ความแปรปรวนในการประมาณค่าความสามารถของผู้สอบเมื่อตอบข้อสอบข้อที่ $m+1$ แล้ว

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY

$a_i$ = ค่าอำนาจจำแนกของข้อสอบข้อที่ $m+1$

$b_i$ = ค่าความยากของข้อสอบข้อที่ $m+1$

$c_i$ = ค่าการเดาของข้อสอบข้อที่ $m+1$

สำหรับงานวิจัยในครั้งนี้ การประมาณค่าความสามารถของผู้สอบ ผู้วิจัยเลือกใช้วิธีประมาณค่าความสามารถของผู้สอบโดยใช้หลักการของเบย์ ที่ชื่อว่า เบย์เซี่ยน อัพเดทติ้ง (Bayesian Updating) ซึ่งสามารถประมาณค่าความสามารถของผู้สอบได้ค่อนข้างคงที่

## การจัดกลุ่มข้อมูล

การวิเคราะห์การจัดกลุ่มข้อมูล เป็นกระบวนการในการจำแนกวัตถุหรือข้อมูลจากระดับประชากร ให้อยู่ในรูปแบบของข้อมูลประชากรย่อยหรือข้อมูลระดับตัวอย่างที่มีความหมายเพื่อประสิทธิภาพในการอธิบายรูปแบบปัญหา และลักษณะเฉพาะของแต่ละกลุ่มย่อยของข้อมูลได้ ลักษณะทั่วไปในการจัดกลุ่มข้อมูล (Jain and Dubes. 1998) เป็นการจำแนกลักษณะของวัตถุด้วยความสัมพันธ์ระหว่างวัตถุที่แทนด้วยเมตริกซ์ใกล้เคียง (Proximity Matrix) ตามแถวและคอลัมน์ในรูปแบบมิติของข้อมูล (D-dimensional) หรือคุณลักษณะของข้อมูล (Attribute) ซึ่งเป็นการวัดความใกล้เคียงหรือความเหมือนกัน (Similarity) ของวัตถุ โดยวัดระยะห่างของวัตถุตามความเหมือนหรือความแตกต่าง (Dissimilarity) เช่น การวัดระยะห่างแบบ Euclidean การวัดระยะห่างแบบ Hamming และการวัดระยะห่างแบบ Mahalanobis เป็นต้น

มหาวิทยาลัยราชภัฏพระนครศรีอยุธยา
RAJABHAT MAHASARAKHAM UNIVERSITY

### 1. ประเภทการแบ่งกลุ่มข้อมูล

การจัดประเภทของการแบ่งกลุ่มข้อมูล (Jain and Dubes. 1998 : 56-57) มีดังนี้

1.1 Exclusive versus Nonexclusive เป็นการจำแนกข้อมูลแบบผูกขาดกับการจัดกลุ่มของข้อมูล (Exclusive) แต่ละวัตถุจะอยู่ในกลุ่มย่อยเดียวเท่านั้น สำหรับการจำแนกข้อมูลแบบไม่ผูกขาด (Non-exclusive) มีการซ้อนทับกันระหว่างกลุ่มของวัตถุหรืออาจกล่าวได้ว่าวัตถุสามารถอยู่ได้หลายกลุ่ม เช่น การจัดกลุ่มคนโดยอายุ หรือเพศ เป็นแบบ Exclusive ในขณะที่การจัดกลุ่มผู้ป่วยเป็นแบบ Nonexclusive เพราะคนหนึ่งคนอาจป่วยมากกว่าหนึ่งโรคได้ ตัวอย่าง อัลกอริทึมที่ใช้ในการจัดกลุ่มข้อมูลแบบ Nonexclusive เช่น Fuzzy clustering เป็นต้น

1.2 Intrinsic versus Extrinsic การจำแนกแบบ Intrinsic ใช้เฉพาะเมตริกซ์ความใกล้เคียง (Proximity Matrix) ในการจัดกลุ่มข้อมูล ที่เรียกว่า "Unsupervised Learning" ในการจดจำรูปแบบ เพราะไม่มีการกำหนดค่าของกลุ่มข้อมูลให้กับแต่ละข้อมูลล่วงหน้า แต่

สำหรับการจำแนกข้อมูลแบบ Extrinsic เป็นการให้ค่าจากการจัดกลุ่มข้อมูลแก่วัตถุโดยเมตริกซ์ความใกล้เคียงเมื่อจำแนกตามปัญหาแล้ว จะกำหนดหรือจำแนกวัตถุอื่นตามลักษณะ ของข้อมูลที่ให้ค่าไว้แล้ว การจัดกลุ่มแบบ Extrinsic ยึดหลักข้อมูลที่มีการสอนก่อนที่เรียกว่า "Teacher" หรือ "Supervised Learning" การจำแนกการจัดข้อมูลแบบ Intrinsic นั้น ได้จากรูปแบบการให้ค่าของวัตถุมีการกำหนดค่าของกลุ่มข้อมูลระหว่างการจัดกลุ่มข้อมูลหรือไม่ หรือมีการจัดคู่หรือให้ค่าความสำคัญของข้อมูลหรือไม่ เช่น การวัดสุขภาพโดยการรวบรวมข้อมูลจากผู้ที่สูบบุหรี่ และไม่สูบบุหรี่ การจำแนกแบบ Intrinsic คือการจัดกลุ่มโดยวัดความคล้ายคลึงกันระหว่างเครื่องชี้วัดสุขภาพของข้อมูลที่มี แล้วอธิบายลักษณะของกลุ่มข้อมูลจากปัจจัยเกี่ยวกับการสูบบุหรี่ และแนวโน้มในการเจ็บป่วยอื่นๆ ในกลุ่มข้อมูล ขณะที่การจำแนกแบบ Extrinsic เป็นการศึกษาเพื่อวินิจฉัย ผู้สูบบุหรี่จากเครื่องชี้วัดสุขภาพของบุคคลนั้นๆ

1.3 Hierarchical versus Partitional เป็นการจัดกลุ่มข้อมูลแบบลำดับชั้นกับการจัดกลุ่มข้อมูลแบบแบ่งส่วน ซึ่งเป็นส่วนย่อยของ Intrinsic โดยอัลกอริทึม Hierarchical ทั่วไป ได้แก่

1.3.1 การรวมกลุ่ม (Agglomerative) เป็นการจำแนกแบบแบ่งลำดับชั้น ที่แต่ละวัตถุหรือข้อมูลเริ่มจากการมีกลุ่มของตัวเอง และค่อยๆ รวมกับแต่ละกลุ่มของวัตถุเป็นลำดับชั้นเพื่อให้ได้กลุ่มที่ใหญ่ขึ้น จนกระทั่งทุกข้อมูลอยู่ในกลุ่มเดียวกัน

1.3.2 การแบ่งแยก (Divisive) เป็นการจำแนกแบบแบ่งลำดับชั้นที่มีขั้นตอนการทำแบบย้อนกลับโดยเริ่มจากทุกวัตถุอยู่ในกลุ่มเดียวกัน และมีการแบ่งย่อยจากกลุ่มใหญ่สู่กลุ่มที่เล็กลง

ปัจจุบันมีวิธีการจัดกลุ่มข้อมูลหลากหลายรูปแบบ ได้แก่ การจัดกลุ่มแบบลำดับชั้น (Hierarchical clustering), การจัดกลุ่มแบบแบ่งส่วน (Partitional clustering), การจัดกลุ่มแบบความน่าจะเป็น (Probabilistic clustering), การจัดกลุ่มแบบกราฟ (Graph based clustering), การจัดกลุ่มแบบฟัซซี่ (Fuzzy clustering), กาจัดกลุ่มแบบโครงข่ายประสาทเทียม (Neural Network based clustering) และการจัดกลุ่มในรูปแบบผสม (Hybrid) ซึ่งแต่ละประเภทมีข้อดีและข้อเสียต่างกัน ทั้งด้านความเร็วในการจัดกลุ่มข้อมูล รูปร่างข้อมูล ประเภทของข้อมูล รวมทั้งขนาดของข้อมูล

**2.อัลกอริทึมเคมีน (K-means Algorithm)**

วิธีการของ K-means Algorithm (Han and Kamber. 2001 : 349) เริ่มจากกำหนดจุดที่ต้องการแบ่งข้อมูลแบบสุ่ม กำหนดขอบเขตของการแบ่งข้อมูล จากนั้นทำการเลื่อนจุดให้

มหาวิทยาลัยมหาสารคาม MAHASARAKHAM UNIVERSITY

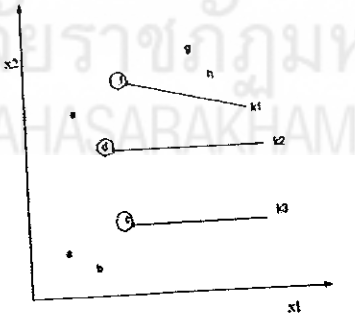
ไปอยู่กึ่งกลางที่สุดของแต่ละ Cluster เมื่อได้จุดใหม่แล้วให้ทำการลากเส้นกำหนดขอบเขตของ Cluster ใหม่ ซึ่งจะได้ขอบเขตของ Cluster ที่ดีที่สุด โดยมีขั้นตอนดังนี้

**ตารางที่ 1** ตัวอย่างข้อมูลสำหรับการแบ่งกลุ่มเคมีน

| Item | Observation | |
|---|---|---|
| | x1 | x2 |
| A | 1 | 2 |
| B | 2 | 1 |
| C | 3 | 3 |
| D | 2.5 | 5 |
| E | 1.5 | 6 |
| F | 3 | 7 |
| G | 5 | 8 |
| H | 6 | 7 |

2.1 ทำการสุ่มส่วนของข้อมูลเป็นกลุ่มเซตย่อย k เซต โดยที่ผู้ทำการแบ่งกลุ่มข้อมูล (Cluster) จะต้องทราบจำนวนกลุ่ม (k) ที่ต้องการแบ่ง ซึ่งสามารถแสดงได้ดังภาพที่ 2

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY



**ภาพที่ 2** การสุ่มข้อมูลเพื่อทำการแบ่งกลุ่มเคมีน

จากภาพที่ 2 มีการแบ่งกลุ่มข้อมูลออกเป็น 3 กลุ่ม (k) โดยทำการสุ่มจุดเริ่มต้นของแต่ละกลุ่ม (k) นำข้อมูลทุกตัวมาคำนวณระยะทางกับ k การคำนวณระยะทางทำได้หลายวิธีเช่น ระยะทางยูคลิเดียน (Euclidean distance) โดยถ้า d คือระยะทางยูคลิเดียน จะได้ระยะทางคือ

$$d(i,j) = \sqrt{\left(\left|X_{i1} - X_{j1}\right|^2 + \left|X_{i2} - X_{j2}\right|^2 + \ldots + \left|X_{ip} + X_{jp}\right|^2\right)}$$

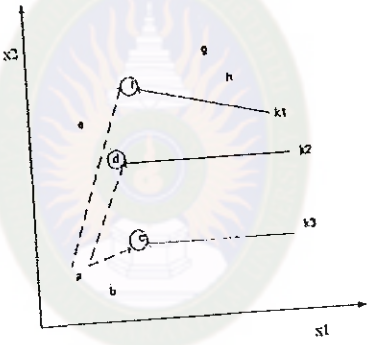
โดยที่ $d(i,j) \geq 0$, $d(i,j) = 0$, $d(i,j) = d(i,j)$, $d(i,j) \leq d(i,k)+d(k,j)$ ซึ่งสามารถคำนวณหาระยะทางยูคลิเดียน โดยใช้ข้อมูลตัวอย่างจากตารางที่ 2-1 และภาพที่ 2-1 ซึ่งได้มีการกำหนดจุดเริ่มต้นของแต่ละกลุ่มข้อมูลคือ จุด c, จุด d, และจุด f ดังนั้นเมื่อคำนวณระยะทางจากจุด a ไปยังจุดเริ่มต้นของแต่ละกลุ่มข้อมูลจะได้ระยะทางดังนี้

ระยะทางจาก a ไป c = $\sqrt{\left(|1-3|^2 + |2-3|^2\right)} = 2.24$

ระยะทางจาก a ไป d = $\sqrt{\left(|1-2.5|^2 + |2-5|^2\right)} = 3.12$

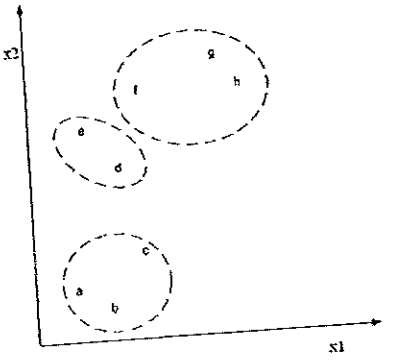
ระยะทางจาก a ไป f = $\sqrt{\left(|1-3|^2 + |2-7|^2\right)} = 5.39$

จากการคำนวณระยะทางระหว่างข้อมูลกับกลุ่มข้อมูล (k) นั้นสามารถแสดงได้ดังภาพที่ 3




**ภาพที่ 3** ระยะทางของข้อมูลกับข้อมูลเริ่มต้นของกลุ่ม

หลังจากนั้นทำการหาระยะทางของข้อมูลทุกจุดกับกลุ่มข้อมูล (k) ทุกกลุ่มและตรวจดูว่าข้อมูลห่างจากกลุ่มข้อมูล (k) ชุดไหนน้อยที่สุดนำข้อมูลไปสังกัดอยู่ในกลุ่มข้อมูล (k) นั้น และได้กลุ่มข้อมูลขึ้นมา โดยสามารถแสดงได้ดังภาพที่ 4




**ภาพที่ 4** ข้อมูลที่ได้รับการแบ่งโดยอาศัย k

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY

209181

2.2 เมื่อได้กลุ่มข้อมูลแต่ละกลุ่มแล้ว แล้วทำการคำนวณหาจุดศูนย์กลางของกลุ่มที่เรียกว่าเซ็นทรอย (centroid/prototype) โดยจุดศูนย์กลางสามารถคำนวณจาก

$$M_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ik}$$

โดยที่ $x = \{x \in k / k$ คือ กลุ่มข้อมูล$\}$

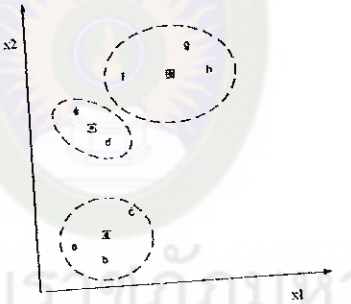
$x = \{x_1, x_2, x_3, ....., x_i\}$

$M_k$ คือ ค่ากลางของข้อมูลในกลุ่มที่ k

$n_k$ คือ จำนวนข้อมูลของกลุ่มที่ k

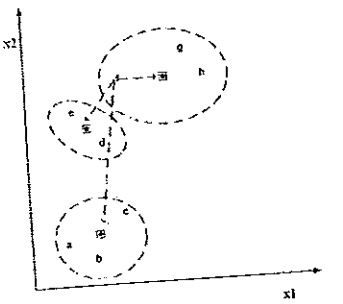
$x_{ik}$ คือ ข้อมูล x ตัวที่ i ของกลุ่มข้อมูลที่ k

เมื่อได้ จุดศูนย์กลางของข้อมูลแล้วนำข้อมูลทุกตัวในแต่ละกลุ่มข้อมูล (cluster) ไปคำนวณกับจุดศูนย์กลาง (centroid) ของทุกกลุ่มข้อมูล (cluster) ดังแสดงในภาพที่ 5



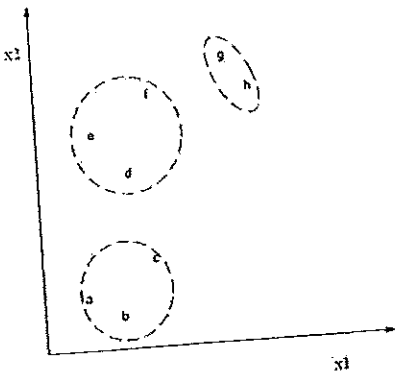
**ภาพที่ 5** จุดศูนย์กลางของข้อมูลในแต่ละกลุ่มข้อมูล

2.3 ตรวจดูระยะทางระหว่างข้อมูลกับจุดศูนย์กลางข้อมูล (centroid) โดยข้อมูลที่ห่างจากจุดศูนย์กลางข้อมูล (centroid) ไหนสั้นที่สุดก็นำข้อมูลไปสังกัดกับกลุ่มดังกล่าว จากจุดนี้จะทำให้เกิดการย้ายกลุ่มข้อมูลซึ่งสามารถแสดงได้ดังภาพที่ 6



**ภาพที่ 6** การวัดระยะทางของข้อมูลกับจุดศูนย์กลางข้อมูล (centroid)

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY



ภาพที่ 7 กลุ่มข้อมูลใหม่

2.4 ทำการคำนวณจุดศูนย์กลางของข้อมูล (centroid) และกำหนดข้อมูลให้กลุ่มข้อมูลใหม่ไปเรื่อยๆ ดังภาพที่ 7 จนกว่าจะถึงเงื่อนไขให้หยุด (stopping criterion) โดยเงื่อนไขให้หยุดคือ 1) ข้อมูลในแต่ละกลุ่มถึงสภาวะเสถียร ไม่สามารถเปลี่ยนกลุ่มได้ดีกว่านี้ได้ 2) ค่าผิดพลาดที่ต่ำสุด (minimum square error) สำหรับทุกกลุ่มข้อมูลโดยหาค่าผิดพลาดที่ต่ำสุดได้จาก

$$E_k^2 = \sum_{k=1}^{k} e_k^2 \quad \text{โดยที่} \quad e_k^2 = \sum_{i=1}^{n_k} \left(x_{ik} - M_k\right)^2$$

โดยที่ $x = \{x \in k / k$ คือ กลุ่มข้อมูล$\}$

$x = \{x_1, x_2, x_3, \ldots\ldots, x_i\}$

$x_{ik}$ คือ ข้อมูลที่ i ของกลุ่มที่ k

$M_k$ คือ ค่ากลางของข้อมูลในกลุ่มที่ k

$n_k$ คือ จำนวนข้อมูลของกลุ่มที่ k

$e_k^2$ คือ ค่าผิดพลาดในกลุ่มที่ k

$E_k^2$ คือ ค่าผิดพลาดรวมของทุกกลุ่ม

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY

**3. การวัดประสิทธิภาพการแบ่งกลุ่มข้อมูล**

3.1 ค่าความแปรปรวน (Variance)

ค่าความแปรปรวน เป็นค่าที่ใช้วัดการกระจายของข้อมูล โดยที่ค่าความแปรปรวนนั้นจะพิจารณาจากผลรวมของค่าความแตกต่างระหว่างค่าของข้อมูลและค่าเฉลี่ยเลข

คณิต ซึ่งถ้าหากค่าแตกต่างสูงมากแสดงว่าข้อมูลมีการกระจายมาก ซึ่งมีสูตรดังต่อไปนี้ (วามิณี นิยากาศ. 2549 : 29-30)

ค่าความแปรปรวนของข้อมูลที่ไม่ได้จัดกลุ่ม

$$\sigma^2 = \sum(x_i-\mu)^2 \Big/ N \quad \text{หรือ} \quad \sigma^2 = \left[\sum x_i^2-(\sum x_i)^2 \Big/ N\right] \Big/ N$$

ค่าความแปรปรวนของข้อมูลที่จัดกลุ่มแล้ว

$$\sigma^2 = \sum f_i(x_i-\mu)^2 \Big/ N \quad \text{หรือ} \quad \sigma^2 = \left[\sum f_i x_i^2-(\sum f_i x_i)^2 \Big/ N\right] \Big/ N$$

3.2 ค่าเบี่ยงเบนมาตรฐาน (Standard Deviation)

ค่าเบี่ยงเบนมาตรฐานคือ รากที่สอบของค่าแปรปรวน ที่เป็นค่าบวกเท่า (มัลลิกา และคณะ, 2540) มีสูตรดังนี้

ค่าเบี่ยงเบนมาตรฐานของข้อมูลที่ไม่ได้จัดกลุ่ม

$$\sigma = \sqrt{\sum(x_i-\mu)^2}/N$$

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY

ค่าเบี่ยงเบนมาตรฐานของข้อมูลที่จัดกลุ่ม

$$\sigma = \sqrt{\sum f_i(x_i-\mu)^2}/N$$

3.3 ค่าความแตกต่างของข้อมูลภายในกลุ่ม (Root Mean Square Standard Deviation : RMSSTD)

การวัดค่าความแตกต่างของข้อมูลภายในกลุ่ม โดยค่าความแตกต่างของข้อมูลภายในกลุ่มนั้นแสดงให้เห็นถึงประสิทธิภาพการจัดกลุ่มว่ามีมากน้อยเพียงใด ซึ่งหากค่าความแตกต่างภายในกลุ่มน้อย นั้นย่อมหมายถึงการแบ่งกลุ่มที่ดี ข้อมูลภายในกลุ่มมีความ

ใกล้เคียงกันมาก สูตรการคำนวณค่าความแตกต่างภายในกลุ่ม มีดังนี้ (Jain et al., 1999 อ้างอิงใน วามิณี นิยากาศ. 2549 : 31)

$$RMSSTD = \sqrt{\sum_{j=1..d}^{i=1..c} \sum_{k=1}^{nij} (x_k - \bar{x}_j)^2}$$

โดยที่

c คือ จำนวนกลุ่มที่แบ่งได้ทั้งหมด

d คือ จำนวนคอลัมน์ทั้งหมดภายในชุดข้อมูล

$\bar{x}_j$ คือ ค่าเฉลี่ยของข้อมูลคอลัมน์ที่ j

$n_{ij}$ คือ จำนวนข้อมูลในกลุ่มที่ i คอลัมน์ที่ j

3.4 ค่าความแตกต่างของข้อมูลระหว่างกลุ่ม (R-Squared : RS)

การวัดค่าความแตกต่างของข้อมูลระหว่างกลุ่ม ค่าความแตกต่างของข้อมูลระหว่างกลุ่ม เป็นค่าซึ่งแสดงให้เห็นถึงประสิทธิภาพการจัดกลุ่มข้อมูล ซึ่งหากค่าความแตกต่างระหว่างกลุ่มมาก แสดงถึงการแบ่งกลุ่มที่ดี กลุ่มแต่ละกลุ่มมีลักษณะที่แตกต่างกัน โดยที่ค่า RS นี้จะอยู่ระหว่าง 0 ถึง 1 สูตรการคำนวณค่าความแตกต่างระหว่างกลุ่ม มีดังนี้ (Jain et al., 1999 อ้างอิงใน วามิณี นิยากาศ. 2549 : 31-32)

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY

$$RS = \frac{SS_t - SS_w}{SS_t}$$

$$SS_t = \sum_{j=1}^{d} \sum_{k=1}^{n_j} (x_k - \bar{x}_j)^2$$

$$SS_w = \sum_{j=1..d}^{i=1..c} \sum_{k=1}^{n_{ij}} (x_k - \bar{x}_j)^2$$

โดยที่

$SS_t$ คือ ผลรวมของผลต่างกำลังสองของทุกข้อมูลแต่ละตัวแปร

$SS_w$ คือ ผลรวมของผลต่างกำลังสองทุกข้อมูลภายในกลุ่ม

c คือ จำนวนกลุ่มที่แบ่งได้ทั้งหมด

d คือ จำนวนตัวแปรทั้งหมด

$\bar{x}_j$ คือ ค่าเฉลี่ยของข้อมูลคอลัมน์ที่ j

$n_{ij}$ คือ จำนวนข้อมูลในกลุ่มที่ i คอลัมน์ที่ j

## การสร้างและพัฒนาโปรแกรมคอมพิวเตอร์

การพัฒนาระบบสารสนเทศเป็นกระบวนการในการสร้างระบบสารสนเทศขึ้นมาเพื่อใช้ สำหรับแก้ปัญหาหรือสร้างมูลค่าเพิ่มให้กับธุรกิจ และด้วยระบบสารสนเทศในยุคปัจจุบันนับวัน จะทวีความซับซ้อนยิ่งขึ้นและมีขนาดใหญ่ ดังนั้น โครงการพัฒนาระบบสารสนเทศจึงจำเป็น ต้องได้รับการวางแผนที่ดีปกติแล้วคำว่า "วงจรชีวิต (Life Cycle)"มักจะใช้กับสิ่งมีชีวิตบนพื้นโลก ไม่ว่าจะเป็นวงจรชีวิตของมนุษย์ สัตว์ หรือพืช ซึ่งข้องเกี่ยวกับการเกิด การดำเนินชีวิตและการตาย ซึ่งจัดเป็นวงจรชีวิตของมนุษย์โดยปกติ ในทำนองเดียวกันเมื่อนำวงจรชีวิตนี้มาใช้กับซอฟต์แวร์ ซึ่งริเริ่มจากการวางแผนเพื่อวิเคราะห์ถึงปัญหาของระบบงานเดิม จากนั้นจึงดำเนินการศึกษาความเป็นไปได้ในแง่มุมต่างๆ จนกระทั่งได้มีโครงการริเริ่มนำซอฟต์แวร์มาใช้งาน และเมื่อมีการนำซอฟต์แวร์มาใช้งานไปตามกาลเวลา สิ่งแวดล้อมต่างๆ รวมถึงเทคโนโลยีก็อาจเปลี่ยนไปตามยุคสมัย ซอฟต์แวร์ดังกล่าวก็อาจไม่สามารถตอบสนองการใช้งานที่ดีได้อีกต่อไป ดังนั้นจึงมีความจำเป็นจะต้องดำเนินการปลดระวางซอฟต์แวร์เหล่านี้ออกไปเมื่อถึงกาลเวลา และดำเนินการวางแผนเพื่อเริ่มต้นศึกษาถึงปัญหาใหม่ ด้วยการพัฒนาระบบใหม่หรือนำซอฟต์แวร์ใหม่ที่เหมาะสมมาใช้งานแทน และด้วยเหตุดังกล่าว ซอฟต์แวร์จึงมีลักษณะเป็นวงจรชีวิตเช่นเดียวกัน ที่เรียกว่าวงจรการพัฒนาระบบ (System Development Life Cycle) หรือมักเรียกสั้นๆ ว่า SDLC (โอภาส เอี่ยมสิริวงศ์. 2548 : 50)

**1. วงจรการพัฒนาระบบ (System Development Life Cycle : SDLC)** ประสงค์ ประณีตพลกรัง และคณะ(2543 : 169) ได้กล่าวถึงขั้นตอนการพัฒนาระบบสารสนเทศ ซึ่งผู้วิจัยนำมาอ้างอิงเพื่อเป็นขั้นตอนในการพัฒนาระบบจำแนกความสามารถผู้เรียนโดยใช้คลังข้อสอบมีขั้นตอนต่างๆ ดังนี้

1.1 การกำหนดปัญหาของระบบงานเดิม (Problem definition)

เป็นขั้นตอนของการกำหนดขอบเขตปัญหา สาเหตุของปัญหา และแนวทางการแก้ปัญหา เพื่อกำหนดวัตถุประสงค์ ของเขตของการพัฒนาโปรแกรม

1.2 การวิเคราะห์ระบบ (System analysis)

เป็นขั้นตอนวิเคราะห์ว่าโปรแกรมที่พัฒนาขึ้นใหม่นี้จะประกอบด้วยอะไรบ้าง มีความเกี่ยวข้องกับสิ่งใด กำหนดรูปแบบการทำงานโปรแกรมเพื่อให้ทำงานได้อย่างมีประสิทธิภาพ

1.3 การออกแบบระบบ(System design)

มหาวิทยาลัยราชภัฏมหาสารคาม RAJABHAT MAHASARAKHAM UNIVERSITY

เป็นขั้นตอนที่นำผลลัพธ์จากการวิเคราะห์ระบบมาออกแบบรูปแบบการทำงานของโปแกรม เช่น การออกแบบฐานข้อมูล รูปแบบหน้าจอ การบันทึกข้อมูล การแสดงผลข้อมูล การรายงานข้อมูล เป็นต้น

1.4 การพัฒนาโปรแกรม (Development)

เป็นขั้นตอนการพัฒนาโปรแกรมเพื่อสร้างระบบงานใหม่ขึ้นมา

1.5 การทดสอบระบบ (Testing)

เป็นขั้นตอนของการทดสอบระบบก่อนนำไปใช้งานจริง ทดสอบโดยผู้วิจัยและให้ผู้เชี่ยวชาญทดลองใช้งานโปรแกรม ถ้าพบข้อผิดพลาดจะย้อนกลับไปพัฒนาโปรแกรมใหม่

1.6 การติดตั้งระบบ (Implementation)

เป็นขั้นตอนหลังจากที่ได้ทำการทดลองโปรแกรมเรียบร้อยแล้วว่าโปรแกรมสามารถทำงานได้ตามต้องการแล้ว จึงทำการติดตั้งโปรแกรมบนระบบเครือข่ายอินเทอร์เน็ต

1.7 การประเมินผล (Evaluation)

เป็นขั้นตอนสุดท้ายของการพัฒนาโปรแกรม ซึ่งต้องมีการประเมินผลเพื่อให้ทราบถึงความพึงพอใจของผู้ใช้งานระบบจำแนกความสามารถผู้เรียนโดยใช้คลังข้อสอบ

**2. ประสิทธิภาพของระบบ (Efficiency)**

สุณี บุญเทวี (2547 : 59 - 60) กล่าวว่าการหาประสิทธิภาพและคุณภาพการทำงานของระบบสารสนเทศ สามารถทำงานได้อย่างถูกต้อง แม่นยำ รวดเร็วและตรงตามความต้องการของผู้ใช้ โดยแบ่งออกเป็น 2 ส่วน ได้แก่ การทดสอบโดยผู้พัฒนาโปรแกรม และการทดสอบโดยผู้เชี่ยวชาญ ผู้ใช้งานและผู้ดูแลระบบ

มนต์ชัย เทียนทอง (2552 : 198 - 200) กล่าวว่าประสิทธิภาพ (Efficiency) ตามความหมายทั่วไป หมายถึง ความสัมพันธ์ระหว่างปริมาณทรัพยากรที่ใช้ไปกับปริมาณผลผลิตที่เกิดจากกระบวนการ ส่วนความหมายอื่น ๆ หมายถึง ความสามารถในการผลิตหรือการใช้งานและความคุ้มค่าของการลงทุน ดังนั้น ประสิทธิภาพจึงเป็นการพิจารณาทางด้านผลลัพธ์ที่ได้รับเมื่อเปรียบเทียบกับสิ่งที่ใช้ไปไม่ว่าจะเป็นทรัพยากร งบประมาณ เวลา หรือปัจจัยอื่น ๆ ในการวิจัยเชิงทดลองทางด้านเทคโนโลยีสารสนเทศ หลังจากที่พัฒนาระบบ เครื่องมือ อุปกรณ์หรือซอฟท์แวร์ขึ้นมาใหม่แล้ว ผู้วิจัยส่วนใหญ่นิยมหาประสิทธิภาพ โดยกำหนดประสิทธิภาพเป็นตัวแปรการทดลองอยู่เสมอ ๆ เนื่องจากเป็นตัวบ่งชี้คุณภาพของระบบที่พัฒนาขึ้นที่เข้าใจได้ง่าย การหาประสิทธิภาพสำหรับการวิจัยเชิงทดลองทางด้านเทคโนโลยีสารสนเทศได้แก่การหา

มหาวิทยาลัยราชภัฏมหาสารคาม RAJABHAT MAHASARAKHAM UNIVERSITY

ประสิทธิภาพสำหรับการวิจัยเชิงทดลองตามแนวทางการวิจัยด้านระบบสารสนเทศโดยวิธี Blackbox รายละเอียดมีดังนี้

การหาประสิทธิภาพสำหรับการวิจัยเชิงทดลองตามแนวทางการวิจัยด้านระบบสารสนเทศโดยวิธี Black box เมื่อแปลความหมายตรงตัวก็คือกล่องดำซึ่ง หมายถึง การประเมิน ที่ไม่พิจารณาภายในของระบบอันได้แก่ ตัวโปรแกรม โครงสร้าง ข้อมูล อัลกอริทึม การจัดการข้อมูล ตัวแปร นิพจน์และอื่น ๆ จึงเปรียบเสมือนภายในระบบเป็นกล่องดำแต่จะพิจารณาเฉพาะ ส่วนนำเข้า (Input) และส่วนแสดงผล (Output) การประเมินด้วยวิธี Blackbox มีวัตถุประสงค์เพื่อการประเมินข้อผิดพลาดต่าง ๆ ได้แก่ 1) การทำงานของส่วนต่าง ๆ 2) การปฏิสัมพันธ์ 3) ข้อมูลและฐานข้อมูล 4) สมรรถนะ และ 5) ผลลัพธ์

การหาประสิทธิภาพด้วยวิธี Blackbox ประเมินได้จากผู้ใช้หรือผู้ที่เกี่ยวข้องกับการใช้ระบบสารสนเทศที่พัฒนาขึ้น ซึ่งไม่จำเป็นต้องมีความรู้ ทักษะ และประสบการณ์ด้านตัวโปรแกรมแต่อย่างใด เนื่องจากเป็นการประเมินภาพรวมของการนำเข้าและการแสดงผลจึงประเมินได้ง่าย ซึ่งนิยมใช้แบบสอบถามแบบมาตราส่วนประเมินค่าเพื่อประเมินระดับความคิดเห็นของผู้ใช้ในแต่ละส่วน ผลการประเมินจะถูกวิเคราะห์โดยใช้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน สำหรับรายการประเมินด้วยวิธี Blackbox จะมีประเด็นหลัก ๆ ที่สำคัญดังนี้

1. ด้านหน้าที่และความถูกต้องในการทำงานของระบบ (Functional testing) เป็นการทดสอบแต่ละส่วนในลักษณะภาพรวม นับตั้งแต่ส่วนนำเข้าส่วนประมวลผล จนถึงส่วนแสดงผล

2. ด้านหน้าที่และความถูกต้องในการทำงานของระบบ (Functional requirement testing) เป็นการทดสอบว่าเป็นไปตามความต้องการหรือไม่ ตั้งแต่ส่วนนำเข้า ส่วนประมวลผล จนถึงส่วนแสดงผล ซึ่งมีลักษณะคล้ายกับการประเมินด้าน Functional test แตกต่างกันที่การประเมินในด้านนี้ จะต้องเปรียบเทียบกับความต้องการหรือข้อกำหนดต่าง ๆ ที่มีอยู่

3. ด้านการใช้งาน (Usability testing) เป็นการทดสอบ เช่น ความง่ายในการติดตั้ง การใช้งานในส่วนต่าง ๆ การปฏิสัมพันธ์การนำเสนอ และการแสดงผลลัพธ์และคู่มือเป็นต้น

4. ด้านความปลอดภัยของระบบ (Security testing) เป็นการทดสอบ เช่น ระบบการพิสูจน์สิทธิ์การรักษาความปลอดภัย และการเข้ารหัส เป็นต้น

5. ด้านความสามารถในการทำงานของระบบ (Performance testing)เป็นการทดสอบ เช่น ความถูกต้อง ความรวดเร็ว สมรรถนะ และประสิทธิภาพโดยรวม

มหาวิทยาลัยราชภัฏมหาสารคาม
RAJABHAT MAHASARAKHAM UNIVERSITY

สุณี บุญเทวี (2547 : 59 - 60) กล่าวว่าการหาประสิทธิภาพและคุณภาพการทำงานของระบบสารสนเทศ สามารถทำงานได้อย่างถูกต้อง แม่นยำ รวดเร็วและตรงตามความต้อง การของผู้ใช้ โดยแบ่งออกเป็น 2 ส่วน ได้แก่ การทดสอบโดยผู้พัฒนาโปรแกรม และการทดสอบโดยผู้เชี่ยวชาญ ผู้ใช้งานและผู้ดูแลระบบ

การใช้วิธีการทดสอบแบบแบล็กบอกซ์ (Blackbox testing)เป็นกระบวนการทดสอบ การทำงานของระบบโดยรวมทั้งหมดว่ามีกระบวนการทำงานถูกต้องตามวัตถุประสงค์ที่ต้องการหรือไม่ โดยทำการทดสอบการทำงานแต่ละฟังก์ชันการทำงานทั้งหมดหาข้อบกพร่องของโปรแกรม หลังจากนั้นทำการปรับปรุงแก้ไขโปรแกรมให้ดีขึ้น โดยแบ่งการประเมินประสิทธิภาพและคุณภาพของโปรแกรมออกเป็น 4 ด้าน ดังนี้

1. ด้านการทำงานตรงตามความต้องการของผู้ใช้ (Functional requirement test)
2. ด้านการทำงานได้อย่างถูกต้องตามขีดความสามารถ (Function test)
3. ด้านการติดต่อระหว่างโปรแกรมกับผู้ใช้ (Usability test)
4. ด้านความปลอดภัยในการทำงานของระบบ (Security test)

จากที่กล่าวมาข้างต้นอาจกล่าวสรุปได้ว่า ประสิทธิภาพ หมายถึง ความสามารถในการผลิตหรือการใช้งานและความคุ้มค่าของการลงทุน การหาประสิทธิภาพสำหรับการวิจัยเชิงทดลองทางด้านเทคโนโลยีสารสนเทศโดยวิธี Blackbox หมายถึง การทดสอบการทำงานของระบบโดยรวมทั้งหมดว่ามีกระบวนการทำงานถูกต้องตามวัตถุประสงค์ที่ต้องการหรือไม่ การประเมินจะไม่พิจารณาภายในของระบบแต่จะพิจารณาเฉพาะส่วนนำเข้า (Input) และส่วนแสดงผล (Output) การประเมินด้วยวิธี Blackbox จะมีประเด็นหลัก ๆ ที่สำคัญประกอบด้วย 5 ด้าน ได้แก่

1. ด้านหน้าที่และความถูกต้องในการทำงานของระบบ
2. ด้านหน้าที่และความถูกต้องในการทำงานของระบบ
3. ด้านการใช้งาน
4. ด้านความปลอดภัยของระบบ
5. ด้านความสามารถในการทำงานของระบบ

มหาวิทยาลัยมหาสารคาม RAJABHAT MAHASARAKHAM UNIVERSITY

## งานวิจัยที่เกี่ยวข้อง

ในการพัฒนาการจำแนกผู้เรียนโดยใช้คลังข้อสอบร่วมกับวิธีการจัดกลุ่มข้อมูลด้วยอัลกอริทึมเคมีน (K-means Algorithm) นั้นผู้วิจัยได้ศึกษางานวิจัยที่เกี่ยวข้องดังนี้

**1. งานวิจัยที่เกี่ยวกับการทดสอบแบบปรับเหมาะ**

นันทิยา พึ่งคำ (2531 : 92) ได้เปรียบเทียบคุณภาพของการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ กับการทดสอบแบบเดิม กลุ่มตัวอย่างได้แก่นักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 120 แบบทดสอบเดิมมีข้อสอบ 40 ข้อ ส่วนข้อสอบในคลังข้อสอบที่ใช้สำหรับการทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์นั้นมี 361 ข้อ ข้อสอบทั้งสองลักษณะวัดความ สามารถด้านคำศัพท์ภาษาอังกฤษ ผลจากการศึกษาพบว่า การทดสอบทั้งสองลักษณะมีความเที่ยงตรงเชิงเกณฑ์สัมพันธ์ไม่ต่างกัน โดยที่การทดสอบแบบปรับเหมาะด้วยคอมพิวเตอร์ใช้ข้อสอบน้อยกว่าครึ่งหนึ่งของการทดสอบแบบเดิม นอกจากนั้นยังมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าต่ำการทดสอบแบบเดิมอีกด้วย

จากการศึกษาของต่าย เชี่ยงฉี่ (2534) พบว่า การประมาณค่าความสามารถตามวิธีการของเบย์ และวิธีความเป็นได้สูงสุดแบบมีเงื่อนไข (conditional maximum likelihood) นั้นให้ผลการประเมินที่ไม่ต่างกัน การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบรูปพีระมิดที่มีรูปแบบ จำนวนขั้นและวิธีการให้คะแนนที่ต่างกันจะส่งผลต่อค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่านอกจากนั้นยังพบว่าค่าความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบพีระมิดแบบขนาดขั้นคงที่ที่มี 16 ขั้น ประมาณค่าความสามารถด้วยวิธีของเบย์มีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบสูงสุด และจากการศึกษาพบว่า ถ้าใช้กลุ่มตัวอย่างขนาดใหญ่ตั้งแต่ 2000 คนขึ้นไปและข้อสอบตั้งแต่ 35 ข้อขึ้นไป การประมาณค่าทั้ง 2 วิธีนี้จะมีความถูกต้องและให้ผลไม่ต่างกัน ถ้ากลุ่มผู้สอบน้อยกว่า 500 คน การประมาณค่าด้วยวิธีของเบย์ จะให้ความคงเส้นคงวา และความถูกต้องแม่นยำมากกว่า แต่ถ้าใช้ข้อสอบน้อยกว่า 15 ข้อแล้วการประมาณค่าด้วยวิธีความเป็นไปได้สูงสุดจะให้ความถูกต้องแม่นยำมากกว่า นอกจากนี้กิ๊ฟฟอร์ดและสวามินาทาน (Gifford and Swaminathan) ก็ได้ศึกษาในทำนองเดียวกัน และยังพบว่าถ้ากลุ่มตัวอย่างมีขนาดเล็กและใช้ข้อสอบน้อยข้อแล้ว การประมาณค่าตัวแปรเสริมต่างๆ ด้วยวิธีของเบย์จะมีความถูกต้องมากกว่า เมื่อเทียบกับวิธีความเป็นไปได้สูงสุด (joint maximum likelihood)

นุจรี สุภีสุทธิ์ (2541 : บทคัดย่อ) ได้ทำการสร้างชุดทดสอบแบบเทเลอร์(ชุดทดสอบแบบปรับเหมาะ) ตามทฤษฎีการตอบสนองข้อสอบ ในวิชาคณิตศาสตร์ (ค102) เรื่องสมการ จุดประสงค์การเรียนรู้ข้อที่ 1 นักเรียนสามารถแก้สมการและตรวจคำตอบได้ ผลจากการวิจัย ได้ข้อสอบที่อยู่ในเกณฑ์ใช้ได้ จำนวน 107 ข้อ สามารถนำไปใช้ได้กับนักเรียนในทุกระดับความสามารถ

มหาวิทยาลัยราชภัฏพระนครศรีอยุธยา RAJABHAT MAHASARAKHAM UNIVERSITY

ธวัฒน์ชัย ศิลังค์ประชา (2544 )พัฒนาการสอบแบบปรับเปลี่ยนได้บนคอมพิวเตอร์ โดยใช้รูปแบบแยกทางคงที่ รูปปีรามิดแบบขนาดคงที่ (Constant Step Size Pyramid) และใช้วิธีการประมาณค่าความสามารถผู้สอบด้วยวิธีการประมาณค่าความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood ) แล้วนำข้อมูลการตอบของผู้สอบมาวิเคราะห์ค่าความสัมพันธ์ระหว่างคะแนนเฉลี่ยสะสม(GPA) ของผู้สอบกับคะแนนความ สามารถที่ได้จากโปรแกรมระบบการสอบแบบปรับเปลี่ยนได้ การศึกษาทดลองสรุปได้ว่า โปรแกรมระบบการสอบแบบปรับเปลี่ยนได้มีประสิทธิภาพตามทฤษฎีการตอบสนองข้อสอบโดยมีความสัมพันธ์ไปในทิศทางเดียวกันกับระดับความสามารถของผู้สอบ โดยมีคะแนนเฉลี่ยสะสม (GPA) ของผู้สอบกับค่าคะแนนความสามารถที่ระดับ 0.374 อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05

โสพส สุขานนท์สวัสดิ์ (2545 : บทคัดย่อ) พัฒนาโปรแกรมสำหรับการทดสอบแบบปรับเหมาะตามระดับความสามารถของผู้สอบโดยใช้คอมพิวเตอร์ ผลการวิจัย ได้พัฒนาโปรแกรมสำหรับทดสอบแบบปรับเหมาะตามระดับความสามารถของผู้สอบ ใช้รูปแบบแยกทางแบบแปรผัน และวิธีการประมาณค่าความสามารถของผู้สอบโดยวิธีของเบย์ จากการประเมินคุณภาพของโปรแกรมสำหรับการทดสอบแบบปรับเหมาะตามระดับความสามารถของผู้สอบ โดยผู้ทดลองใช้โปรแกรม พบว่า โปรแกรมมีประสิทธิภาพและเป็นประโยชน์ต่อการวัดและประเมินผล

ธัญญรัตน์ น้อมพลกรัง (2548 : 93) พัฒนาระบบจัดการและประเมินผลการสอบแบบปรับเปลี่ยนได้โดยใช้คอมพิวเตอร์ โดยใช้รูปแบบแยกทางแปรผัน ด้วยกลวิธีการประมาณค่าความเป็นไปได้สูงสุด พบว่าระบบที่พัฒนาขึ้นมีประสิทธิภาพโดยรวมอยู่ในระดับดี

ภาณุพงษ์ ชัยศรีทิพย์ (2549 : 91)ได้พัฒนาระบบทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบ โดยใช้รูปแบบแยกทางแปรผัน ด้วยวิธีการประมาณค่าความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood ) พบว่าวิธีการนี้มีข้อจำกัดโดยไม่สามารถประมาณค่าความสามารถของผู้สอบที่ทำข้อสอบถูกหรือผิดหมดทุกข้อได้

ยูริ (Urry. 1977 : 183-184 อ้างในรังสรรค์ มณีเล็ก. 2540 : 82) พบว่า การทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบนั้นจะไม่ค่อยมีประสิทธิผลหากเลือกใช้รูปแบบการวิเคราะห์ ข้อมูลที่ไม่ค่อยเหมาะสม เช่น ค่าความเที่ยงตรงเชิงโครงสร้างของแบบทดสอบจะลดลงเมื่อเทียบกับการใช้แบบทดสอบที่มีความยาวเท่ากันแต่เลือกใช้โมเดลการวิเคราะห์ข้อสอบที่ดีกว่า รูปแบบของราส์ช นับว่าไม่ค่อยเหมาะสมในการนำมาใช้กับการวิเคราะห์ข้อสอบแบบเลือกตอบ ทั้งนี้เนื่องจากรูปแบบของราส์ชมีข้อตกลงเบื้องต้นว่า อำนาจจำแกนของข้อสอบแต่ละข้อต้องคงที่

มหาวิทยาลัยราชภัฏรำไพพรรณี
RAJABHAT MAHASARAKHAM UNIVERSITY

และค่าการเดาต้องมีค่าน้อยมากหรือเป็นศูนย์ ซึ่งตามสภาพความเป็นจริงแล้วนับว่าเป็นเรื่องที่ยากมากที่จะทำให้ข้อสอบแต่ละข้อมีลักษณะตามข้อตกลงดังที่กล่าวมารูปแบบโลจิสติก3 ตัวแปรเสริมนับว่ามีความเหมาะสมมากกว่ารูปแบบอื่นๆ

ไวส์และแม็คไบรด์(Weiss and McBride) ได้ศึกษาเกี่ยวกับวิธีการให้คะแนน หรือวิธีการประมาณค่าความสามารถของผู้สอบตามวิธีการของเบย์ที่ปรับใหม่ (bayesian updating) พบว่าเป็นวิธีการประมาณค่าที่ไม่ลำเอียง นอกจากนี้ยูริ(Urry. 1971) เจนเซมา(Jensema. 1972) ยังพบว่า ความสามารถจริงกับความสามารถที่ได้จากการประมาณค่าตามวิธีของเบย์มีความสัมพันธ์กันสูงมาก(ต่าย เชี่ยงฉี. 2534 :13 ; อ้างอิงมาจาก Weiss and McBride. 1984 : 274)

ไวส์(Weiss. 1988 : 374-375 อ้างใน รังสรรค์ มณีเล็ก. 2540 : 82) ได้เสนอแนะว่าคลังข้อสอบที่มีขนาดพอเหมาะนั้นควรมีประมาณ 100-200 ข้อ ซึ่ง 116-150 ข้อจะให้ผลดีที่สุดซึ่งสอดคล้องกับข้อค้นพบของยูริ(Urry. 1970,1977) ที่พบว่า คลังข้อสอบควรมีข้อสอบอย่างน้อย 100 ข้อ และข้อสอบที่จะช่วยยกระดับคุณภาพของการวัดนั้น ควรมีค่าอำนาจจำแนกมากกว่า 0.8 ค่าความยากอยู่ในช่วง -2.0 ถึง 2.0 และค่าการเดาน้อยกว่า 0.3 (Urry. 1977 : 184-187)

โฮ (Ho. 1989 : 421-A) พบว่า ขนาดของคลังข้อสอบส่งผลน้อยมากต่อการประมาณค่าความสามารถของผู้สอบ ยกเว้นกลุ่มผู้สอบที่มีความสามารถต่ำ เมื่อใช้ข้อสอบเริ่มต้นที่มีค่าความยากง่ายสูงกว่าความสามารถที่แท้จริงของผู้สอบ

**2. งานวิจัยที่เกี่ยวกับการจัดกลุ่มข้อมูล**

สุรชัย วิวัฒนเจริญชัย (2546 : บทคัดย่อ) ทำการศึกษาพฤติกรรมของลูกค้าที่ทำธุรกรรมทางอินเทอร์เน็ต เพื่อเป็นข้อมูลขององค์กรในการวางแผนกลยุทธ์ในการแข่งขัน โดยการเปรียบเทียบการแบ่งกลุ่มด้วยอัลกอริทึม 2 ประเภท คือ K-Means Algorithm และ Neural Networks Algorithm และใช้ข้อมูลจากแฟ้ม System Log ของลูกค้าธนาคารพาณิชย์ที่เข้ามาทำธุรกรรมทางอินเทอร์เน็ตในช่วงเวลา 1 ปี ผลที่ได้คือ สามารถแบ่งกลุ่มลูกค้าและวิเคราะห์พฤติกรรมการใช้อินเทอร์เน็ตของข้อมูลแต่ละกลุ่มได้ชัดเจน และสามารถสรุปได้ว่าการแบ่งกลุ่มข้อมูลโดยใช้ K-Means Algorithm และ Neural Networks Algorithm มีความเหมาะสมสำหรับข้อมูลที่มีต่อปริมาณที่แตกต่างกัน

สุพาพร บรรดาศักดิ์ (2548 : 76) ได้ประยุกต์ใช้อัลกอริทึมเคมีนในการจัดกลุ่มผู้เรียน สำหรับระบบการเรียนการสอนที่จัดเนื้อหาตามกลุ่มผู้เรียนแบบออนไลน์ โดยวิธีการเรียนการสอนสามารถปรับเปลี่ยนเนื้อหาและแบบทดสอบตามระดับความรู้และความเข้าใจของผู้เรียนได้เอง จากการจัดกลุ่มผู้เรียนด้วยอัลกอริทึมเคมีน ทำให้ผู้เรียนที่เรียนแบบอีเลิร์นนิ่งมีผล

มหาวิทยาลัยราชภัฏพิบูลสงคราม PIBULSONGKRAM RAJABHAT UNIVERSITY

การเรียนรู้ดีขึ้น เมื่อเปรียบเทียบกับผู้เรียนที่เรียนแบบแบบปกติในชั้นเรียน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสามารถกล่าวได้ว่าระบบการเรียนการสอนตามความรู้ของผู้เรียนแบบอีเลิร์นนิ่งที่มีการจัดกลุ่มที่สร้างขึ้นนั้นมีประสิทธิภาพเชื่อถือได้

ฟอร์จี (Forgy. 1965 อ้างใน สิริชัย ดีเลิศ. 2550 :15-16 ) ได้นำเสนอการเลือกค่าเริ่มต้นอย่างง่าย โดยเลือกค่าจุดเริ่มต้นแบบสุ่มในฐานข้อมูล ถ้าค่าที่เลือกใกล้กับค่ากึ่งกลางความหนาแน่นของข้อมูลในแต่ละกลุ่มแล้ว ผลลัพธ์ที่ได้จากการจัดกลุ่มข้อมูลจะมีประสิทธิภาพ แต่การเลือกค่าเริ่มต้นด้วยวิธีการนี้ไม่สามารถรับประกันได้ว่า ค่าเริ่มต้นที่ได้มีความเหมาะสมหรือใกล้ค่ากึ่งกลางของข้อมูลจริง ถึงแม้ว่าใช้การเลือกค่าซ้ำหลายๆ รอบ

ฟอร์แมน และ แซ้ง (Forman and Zhang. 2001) ศึกษาและเปรียบเทียบประสิทธิภาพการทำงานของการแบ่งกลุ่มข้อมูล 3 วิธีคือ ใช้หลักการ Parallel Code K-Means Algorithm, K-Harmonic Means Algorithm และ Expectation-Maximization โดยการที่ได้จากโปรแกรม MatLab ในการทดสอบพบว่า K-Means Algorithm มีประสิทธิภาพในการจัดกลุ่มข้อมูลที่ดีที่สุด ซึ่งมีประสิทธิภาพที่ดีกว่า K-Harmonic Means Algorithm อยู่ 6 เท่า และดีกว่า Expectation-Maximization 200 เท่า

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องจะเห็นว่าทฤษฎีการตอบสนองข้อสอบ(IRT) เป็นทฤษฎีที่สามารถประมาณค่าความสามารถของผู้สอบได้จริงโดยวิธีการประมาณค่าความสามารถส่วนใหญ่จากการศึกษาจะนิยมใช้อยู่ 2 วิธีการ คือ วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) และวิธีที่นักสถิติชื่อ โรเจอร์ เจ โอ-เวน (Roger J. Owen) ได้เป็นผู้เสนอขึ้นโดยให้ชื่อว่า เบย์เซี่ยน อัพเดทติ้ง (Bayesian Updating) ซึ่งสามารถประมาณค่าความสามารถได้ค่อนข้างคงที่

ดังนั้นผู้วิจัยจึงสนใจที่จะพัฒนาระบบจำแนกความสามารถผู้เรียนโดยใช้คลังข้อสอบด้วยเทคนิคและวิธีการตามยุทธวิธีหลายขั้นตอนแบบแยกทางแปรผัน ยุทธวิธีของเบย์โดยการประมาณค่าความสามารถของผู้สอบด้วยวิธีการของเบย์เซี่ยน อัพเดทติ้ง (Bayesian Updating) และทำการปรับปรุงประสิทธิภาพในการจำแนกผู้เรียนโดยการจำแนกกลุ่มผู้เรียนตามระดับความสามารถด้วยเทคนิควิธีการอัลกอริทึมเคมีน (K-means Algorithm) ซึ่งเป็นวิธีการที่ได้รับความนิยมสูง เนื่องจากความเร็ว ความง่ายต่อความเข้าใจ และการนำไปใช้ในการอธิบายผลลัพธ์ของกลุ่มข้อมูลได้เป็นอย่างดี

มหาวิทยาลัยราชภัฏมหาสารคาม RAJABHAT MAHASARAKHAM UNIVERSITY