

## บทที่ 3

### วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อพัฒนาเกณฑ์ตัดสินคุณภาพการเปรียบเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ ซึ่งเป็นค่าที่แสดงถึงระดับคุณภาพที่ยอมรับได้ของผลการเปรียบเทียบคะแนน ที่คำนวณจากดัชนีความแตกต่าง AMD (Absolute Mean Difference) ดัชนี MAD (Mean Absolute Difference) และดัชนี RMS (Root Mean Squared) หาได้จากรูปแบบการหาเกณฑ์การเปรียบเทียบคะแนนกลับสู่แบบสอบเดิมและการใช้กลุ่มสอบทานผล มีขั้นตอนดำเนินการวิจัย ดังนี้

ตอนที่ 1 ตรวจสอบคุณภาพของการจำลองข้อมูลจากโปรแกรม IRTDATA

ตอนที่ 2 จำลองข้อมูลเพื่อดำเนินการเปรียบเทียบคะแนนตามแบบแผนที่กำหนด จัดกลุ่มดัชนีความแตกต่างและวิเคราะห์หาดัชนีความแตกต่าง

ตอนที่ 3 พัฒนาเกณฑ์จากดัชนีความแตกต่าง

ตอนที่ 4 ตรวจสอบคุณภาพเกณฑ์ที่พัฒนาขึ้น

ตอนที่ 1 ตรวจสอบคุณภาพของการจำลองข้อมูลจากโปรแกรม IRTDATA

ข้อมูลในการวิจัยครั้งนี้ใช้การจำลองข้อมูลด้วยคอมพิวเตอร์โดยใช้โปรแกรม IRTDATA ที่พัฒนาโดย โจแฮนสัน (Johanson, 1992) ลักษณะของโปรแกรมเป็นดังนี้

1) IRTDATA เป็นโปรแกรมในการสร้างข้อมูลของการตอบแบบสอบของกลุ่มตัวอย่าง คำตอบที่ได้เป็นการให้คะแนน 0-1 (ตอบถูกได้ 1 คะแนน ตอบผิดได้ 0 คะแนน) โดยสร้างขึ้นตามแนวทฤษฎีการตอบสนองข้อสอบ ตามโมเดล 1, 2 และ 3 พารามิเตอร์

2) สามารถสร้างคำตอบได้จากสถานการณ์การสอบโดยมีจำนวนผู้สอบได้สูงสุด 3,000 คน และจำนวนข้อสอบสูงสุด 200 ข้อ

3) การสร้างข้อมูลคำตอบสามารถกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน และรูปร่างการแจกแจงเป็นแบบสม่ำเสมอ (Uniform) หรือการแจกแจงเป็นแบบปกติ (Normal) ของพารามิเตอร์ a, b และ c นอกจากนี้ยังสามารถกำหนดค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ผู้สอบได้อีกด้วย

4) สิ่งที่สามารถเปลี่ยนแปลงได้คือเลขสุ่มของการสร้างข้อมูลแต่ละครั้งได้แก่ เลขสุ่มของข้อสอบ เลขสุ่มของผู้สอบ เป็นเลข 4 หลัก เช่น 6245, 8572 เป็นต้น ผู้จำลองสามารถเลือกกำหนดเป็นเลขอะไรก็ได้ ส่วนเลขสุ่มอีกชุดหนึ่งเป็นเลขสุ่มกลไกการสอบประกอบด้วยตัวเลข 5 หลัก เลขตัวแรกถ้ากำหนดเป็น 1 แสดงว่าเป็นการเปรียบเทียบแบบแผนกลุ่มไม่เท่าเทียม

กันใช้แบบสอบร่วมภายใน หลักที่ 2 และ 3 เป็นจำนวนข้อของข้อสอบร่วม ส่วนที่เหลือเป็นตัวเลขที่บอกถึงครั้งที่ของการสอบ เช่น กำหนดเลขสุ่มเป็น 10802 แสดงว่าเป็นการกำหนดให้แบบแผนการปรับเทียบเป็นแบบใช้แบบสอบร่วมภายใน มีจำนวนข้อสอบร่วม 8 ข้อ และเป็นการสอบครั้งที่ 2

5) ค่าพารามิเตอร์ของข้อสอบและพารามิเตอร์ความสามารถ (พารามิเตอร์ผู้สอบ) เป็นการสุ่มมาจากประชากรที่มีค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานตามที่กำหนด

6) ข้อมูลที่ได้จากการใช้โปรแกรม IRTDATA ให้ผลลัพธ์ออกมา 3 ลักษณะ คือ พารามิเตอร์ข้อสอบ พารามิเตอร์ผู้สอบ (พร้อมคะแนนดิบและคะแนนจริง) และเมตริกซ์ของคำตอบซึ่งเป็นคะแนนดิบของการตอบข้อสอบแต่ละข้อจากการตอบของผู้สอบแต่ละคน

7) ข้อมูลที่ได้จากการสร้างขึ้นจากโปรแกรมนี้สามารถนำไปเป็นข้อมูลสำหรับการวิเคราะห์โดยใช้โปรแกรมอื่นต่อไปได้อีก เช่น Excel for Windows หรือ SPSSx

ซึ่งผู้วิจัยคาดว่าโปรแกรม IRTDATA นี้สามารถให้ข้อมูลเพื่อใช้ในการวิจัยในครั้งนี้ได้ แต่เพื่อความมั่นใจว่าข้อมูลที่ได้มีความถูกต้องเชื่อถือได้และเป็นไปตามข้อตกลงเบื้องต้นของทฤษฎีการตอบสนองข้อสอบ ดังนั้นผู้วิจัยจึงดำเนินการตรวจสอบความถูกต้องของข้อมูลและตรวจสอบความเป็นเอกมิติของข้อมูลที่จำลองจากโปรแกรม IRTDATA ดังรายละเอียดต่อไปนี้

## 1.1 ตรวจสอบความถูกต้องของข้อมูลที่ได้จากการจำลองโดยใช้โปรแกรม IRTDATA

1.1.1 ตรวจสอบความถูกต้องของข้อมูลจากการจำลองข้อมูลจากโปรแกรม IRTDATA โมเดล 1 พารามิเตอร์ โดยจำลองข้อมูลในเงื่อนไขใช้แบบสอบที่มีจำนวนข้อสอบ 40 ข้อ ผู้สอบจำนวน 300 คน และใช้แบบสอบที่มีจำนวนข้อสอบ 80 ข้อ ผู้สอบจำนวน 1,000 คน นำคะแนนที่ได้จากการจำลองข้อมูลซึ่งเป็นคะแนนดิบของการตอบข้อสอบแต่ละข้อจากการตอบของผู้สอบแต่ละคนไปวิเคราะห์โดยใช้โปรแกรม BILOG พิจารณาจากพารามิเตอร์ผู้ตอบ ( $\theta$ ) และค่าพารามิเตอร์ข้อสอบ ค่าความยาก ( $b$ ) นำไปเปรียบเทียบกับผลที่ได้จากโปรแกรม IRTDATA แล้วได้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้สอบ ค่าพารามิเตอร์ของข้อสอบที่ได้จากโปรแกรมทั้งสอง โดยใช้ t-test สำหรับทดสอบกลุ่มตัวอย่างสองกลุ่มที่เกี่ยวข้องกัน (Two Dependent Samples Test)

1.1.2 ตรวจสอบความถูกต้องของข้อมูลจากการจำลองข้อมูลจากโปรแกรม IRTDATA โมเดล 3 พารามิเตอร์ โดยจำลองข้อมูลในเงื่อนไขใช้แบบสอบที่มีจำนวนข้อสอบ 40 ข้อ ผู้สอบจำนวน 2,000 คน และใช้แบบสอบที่มีจำนวนข้อสอบ 60 ข้อ ผู้สอบจำนวน 500 คน นำคะแนนที่ได้จากการจำลองข้อมูลซึ่งเป็นคะแนนดิบของการตอบข้อสอบแต่ละข้อ

จากการตอบของผู้สอบแต่ละคนไปวิเคราะห์โดยใช้โปรแกรม BILOG พิจารณาจากพารามิเตอร์ผู้ตอบ ( $\theta$ ) ค่าพารามิเตอร์ข้อสอบ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และความน่าจะเป็นในการเดาถูก (c) นำไปเปรียบเทียบกับผลที่ได้จากโปรแกรม IRTDATA แล้วได้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยพารามิเตอร์ผู้สอบ ค่าพารามิเตอร์ของข้อสอบที่ได้จากโปรแกรมทั้งสอง โดยใช้ t-test สำหรับทดสอบกลุ่มตัวอย่างสองกลุ่มที่เกี่ยวข้องกัน (Two Dependent Samples Test )

## 1.2 ตรวจสอบความเป็นเอกมิติของข้อมูลที่จำลองจากโปรแกรม IRTDATA

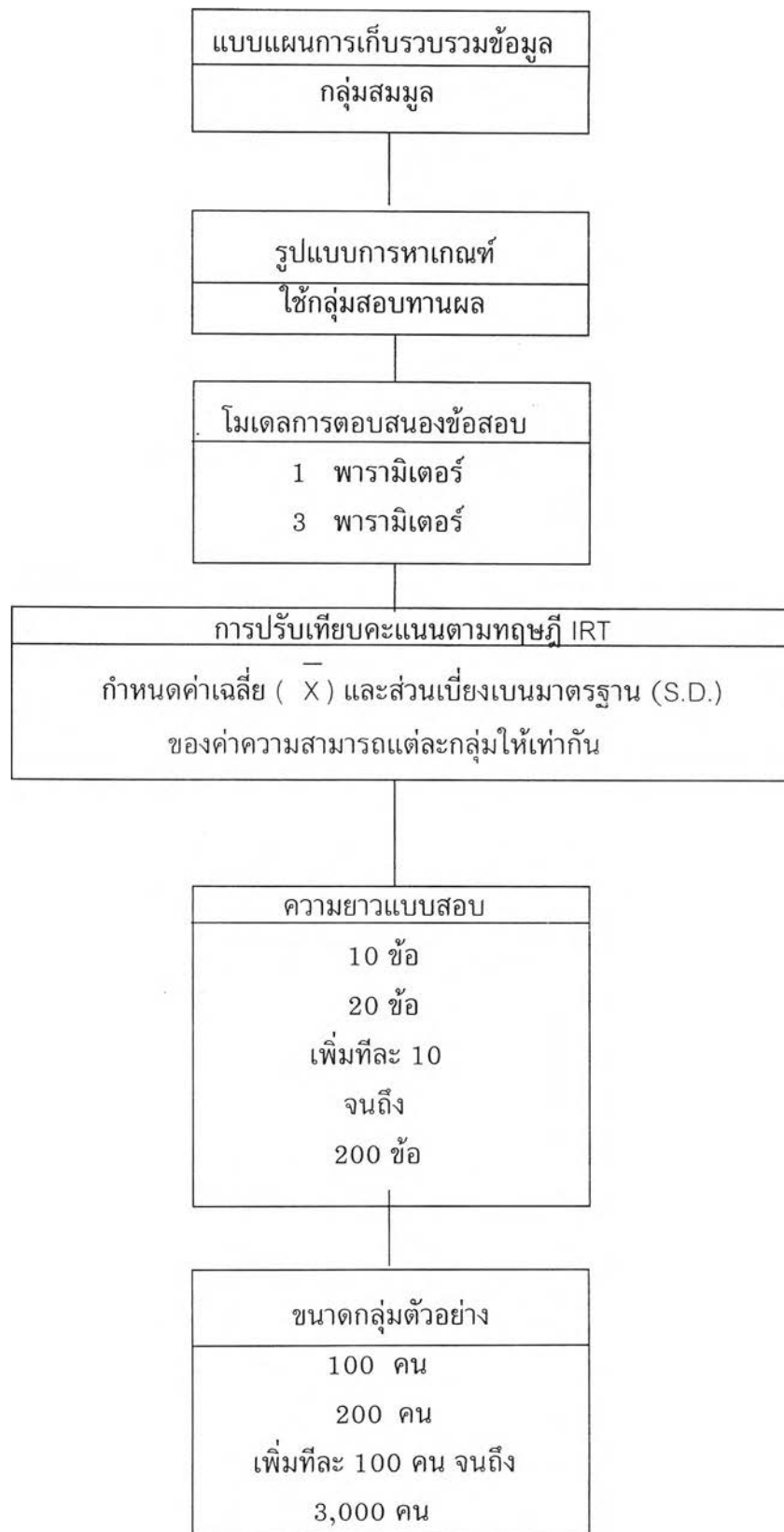
1.2.1 ตรวจสอบความเป็นเอกมิติของข้อมูล ตามทฤษฎีการตอบสนองข้อสอบ โมเดล 1 พารามิเตอร์ โดยการนำข้อมูลผลการตอบข้อสอบรายข้อที่จำลองจากโปรแกรม IRTDATA ข้อสอบ 40 ข้อ ผู้สอบ 300 คน และข้อสอบ 80 ข้อ ผู้สอบ 1,000 คน ไปวิเคราะห์ตัวประกอบ (Factor Analysis) ด้วยวิธีวิเคราะห์ตัวประกอบสำคัญ (Principal Component Analysis) และหมุนแกนด้วยวิธี Varimax ตามที่ Lord (1980) เสนอแนะ ใช้โปรแกรมสำเร็จรูป SPSS

1.2.1 ตรวจสอบความเป็นเอกมิติของข้อมูล ตามทฤษฎีการตอบสนองข้อสอบ โมเดล 3 พารามิเตอร์ โดยการนำข้อมูลผลการตอบข้อสอบรายข้อที่จำลองจากโปรแกรม IRTDATA ข้อสอบ 40 ข้อ ผู้สอบ 2,000 คน และข้อสอบ 60 ข้อ ผู้สอบ 500 คน ไปวิเคราะห์ตัวประกอบ (Factor Analysis) ด้วยวิธีวิเคราะห์ตัวประกอบสำคัญ (Principal Component Analysis) และหมุนแกนด้วยวิธี Varimax ตามที่ Lord (1980) เสนอแนะ ใช้โปรแกรมสำเร็จรูป SPSS

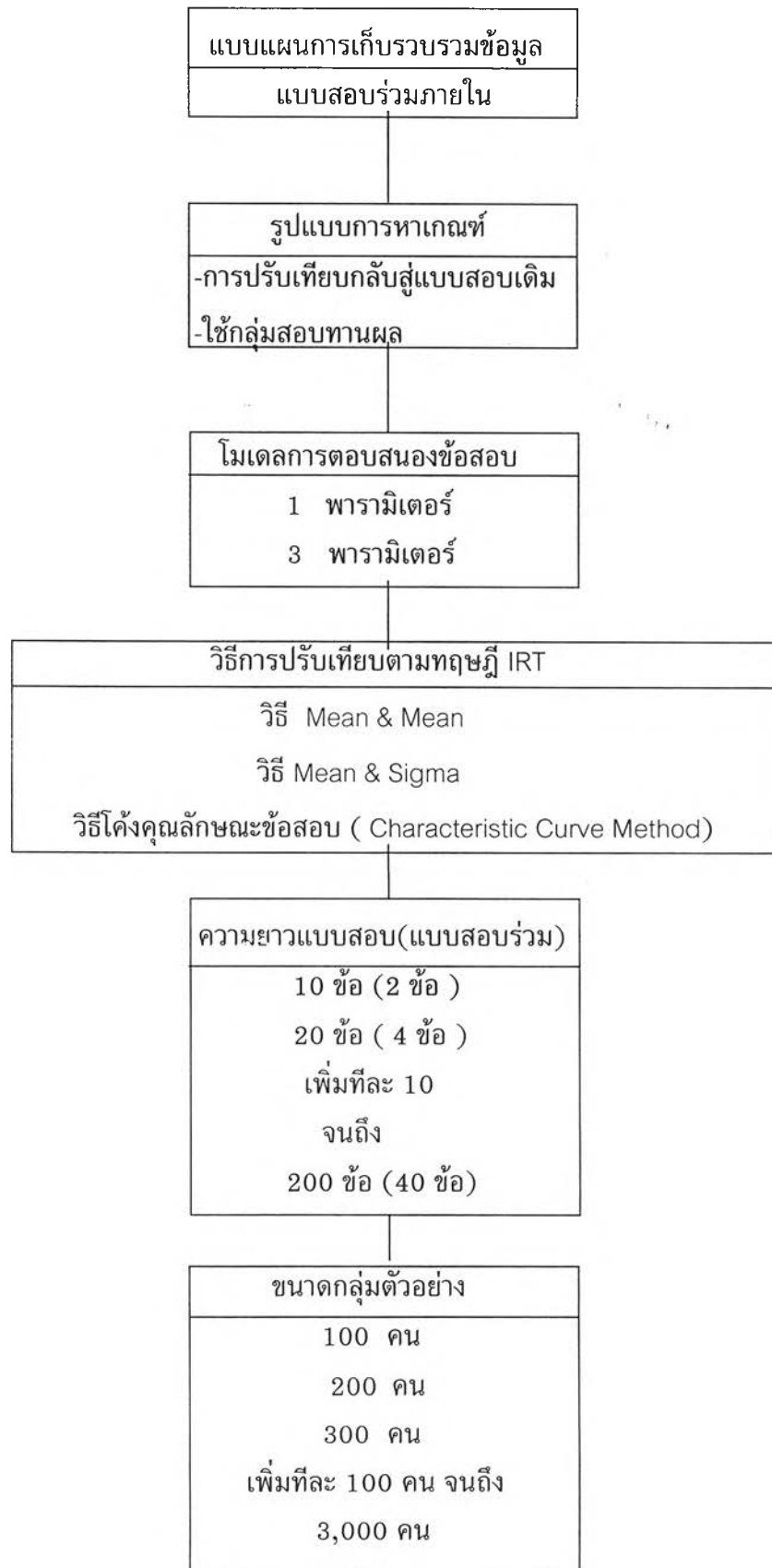
ตอนที่ 2 จำลองข้อมูล ปรับเทียบคะแนนตามแบบแผนที่กำหนด จัดกลุ่มดัชนีความแตกต่างและวิเคราะห์หาดัชนีความแตกต่าง

### 2.1 แบบแผนการจำลองข้อมูล

แบบแผนการจำลองข้อมูลเป็นแบบกลุ่มสมมูล และแบบกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วมภายใน ตัวแปรที่กำหนดไว้สามารถนำมาเสนอด้วยแผนภาพที่ 1 และแผนภาพที่ 2



ภาพที่ 1 แผนผังการจำลองข้อมูลแบบแผนกลุ่มสมมูล



ภาพที่ 2 แผนผังการจำลองข้อมูลตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบรวม

## 2.2 การจำลองข้อมูลตามแบบแผนกลุ่มสมมุติ ปรับเทียบคะแนนตามรูปแบบการใช้กลุ่มสอบทานผล

ในการปรับเทียบตามแบบแผนกลุ่มสมมุติจะต้องมีการจำลองข้อมูลทั้งตามโมเดล 1 พารามิเตอร์ และ 3 พารามิเตอร์ ต้องเปลี่ยนความยาวแบบสอบและขนาดกลุ่มตัวอย่าง ดังนั้นจะต้องมีการจำลองข้อมูลตามเงื่อนไขต่าง ๆ เป็นจำนวนมาก แต่การจำลองแต่ละครั้งใช้วิธีการที่คล้ายคลึงกัน เช่นในกรณีจำลองตามโมเดล 1 พารามิเตอร์ ความยาวข้อสอบ 40 ข้อ จำนวนผู้สอบ 500 คน มีขั้นตอนดำเนินการดังนี้

2.2.1 กำหนดลักษณะการแจกแจง ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ผู้สอบให้คงที่ไว้ค่าหนึ่ง ค่าเฉลี่ยเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และส่วนการแจกแจงให้เป็นโค้งปกติ

2.2.2 กำหนดลักษณะการแจกแจง ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ข้อสอบตามแบบแผนที่กำหนด เช่น ค่าเฉลี่ยค่าความยาก (b) เท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และส่วนการแจกแจงให้เป็นโค้งปกติ

2.2.3 กำหนดจำนวนพารามิเตอร์ให้เป็น 1 พารามิเตอร์

2.2.4 กำหนดจำนวนข้อสอบเป็น 40 ข้อ

2.2.5 กำหนดจำนวนผู้สอบเป็น 500 คน

2.2.6 กำหนดเลขสุ่มของแบบสอบฉบับที่ 1 (X) และฉบับที่ 2 (Y) เป็น 5346 และกำหนดเลขสุ่มของผู้สอบทั้ง กลุ่มที่ 1 เป็น 1543 (A) กลุ่มที่ 2 เป็น 5137(B) ส่วนกลุ่มที่ 3 เป็นกลุ่มสอบทานผล (Validation Samples) ที่ทำแบบสอบทั้งสองชุดเป็น 4124 (C)

2.2.7 จำลองข้อมูล

2.2.8 ปรับสเกลพารามิเตอร์ข้อสอบ

เนื่องจากในการจำลองข้อมูลรูปแบบกลุ่มสมมุติ ได้มีการกำหนดให้พารามิเตอร์ความสามารถหรือพารามิเตอร์ผู้สอบ ( $\theta$ ) ให้มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 ดังนั้นค่า  $\theta$  ที่ได้จากแบบสอบฉบับ X จึงอยู่ในสเกลเดียวกันกับค่า  $\theta$  ที่ได้จากแบบสอบฉบับ Y จึงสามารถนำมาเทียบกันได้ (Kolen and Brennan, 1995) เพราะ

$$\theta_Y = (\sigma_{\theta_Y}/\sigma_{\theta_X}) \theta_X + [\mu_{\theta_Y} - (\sigma_{\theta_Y}/\sigma_{\theta_X}) \mu_{\theta_X}]$$

เมื่อ  $\mu_{\theta_X}$  เป็น ค่าเฉลี่ยของพารามิเตอร์ความสามารถจากแบบสอบฉบับ X

$\mu_{\theta_Y}$  เป็น ค่าเฉลี่ยของพารามิเตอร์ความสามารถจากแบบสอบฉบับ Y

$\sigma_{\theta_X}$  เป็น เป็นส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความสามารถของผู้สอบฉบับ X

$\sigma_{\theta_Y}$  เป็น เป็นส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ความสามารถของผู้สอบฉบับ Y

ดังนั้นค่าคะแนนจริง จากแบบสอบฉบับ X และ Z ที่ระดับความสามารถใด ๆ จะอยู่บนสเกลเดียวกัน

2.2.9 คำนวณคะแนนจริง (True Score) โดยใช้ค่าพารามิเตอร์ผู้สอบจากกลุ่มสอบทานผลกับพารามิเตอร์ข้อสอบจากแบบสอบชุด X ได้คะแนนจริง  $\tau$  กับพารามิเตอร์จากแบบสอบชุด Y ได้คะแนนจริง  $\tau'$

2.2.10 ตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนจริง  $\tau$  กับ  $\tau'$  โดยใช้สถิติ t-test paired sample ผลที่ได้จำแนกได้ 3 กรณี คือ กรณีที่ค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01 จัดเป็นกลุ่มเกณฑ์คุณภาพต่ำ กรณีที่ค่า t คำนวณเท่ากับค่า tวิกฤตจัดเป็นจุดตัด ส่วนกรณีที่ค่าเฉลี่ยไม่แตกต่างจัดเป็นกลุ่มเกณฑ์มีคุณภาพ

2.2.11 นำ  $\tau$  กับ  $\tau'$  ในทั้งสองกลุ่มจากข้อ 2.1.10 ไปคำนวณหาค่าดัชนีความแตกต่าง ดัชนี AMD ดัชนี MAD และ ดัชนี RMS

2.2.12 เปลี่ยนความยาวแบบสอบเป็น 10, 20, 30, 50, 60, 70, 80, ..., 200 แล้วจำลองข้อมูลเช่นเดียวกันกับเมื่อความยาวของแบบสอบ 40 ข้อ

2.2.13 เปลี่ยนขนาดกลุ่มตัวอย่างเป็น 100, 200, 300, 400, 600, ..., 3,000 คน แล้วจำลองข้อมูลเช่นเดียวกันกับขนาดกลุ่มตัวอย่าง 500 คน

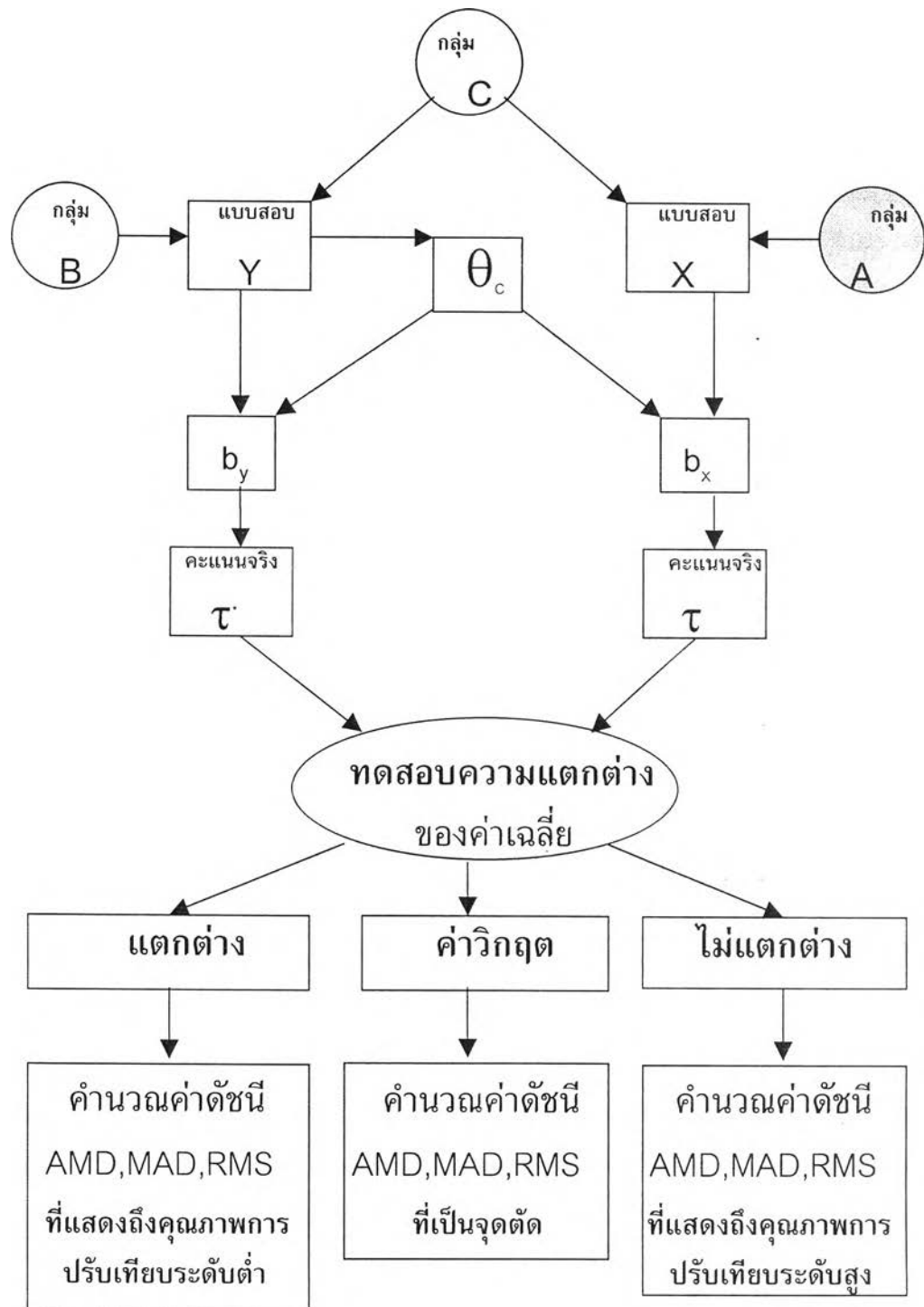
2.2.14 ทำตามขั้นตอนเดียวกันทั้งหมดนี้กับโมเดล 3 พารามิเตอร์

2.2.15 ทำซ้ำในขั้นตอนเดียวกันจากข้อ 2.1.1 ถึง ข้อ 2.1.14

2.2.16 จำลองข้อมูลทั้งหมด  $4 \times 2 \times 20 \times 30 \times 2$  เท่ากับ 9,600 ครั้ง ได้ดัชนีความแตกต่าง AMD ดัชนี MAD และ ดัชนี RMS ดัชนีละ 2,400 ตัว

วิธีการจำลองข้อมูล ปรับเทียบคะแนนและคำนวณค่าดัชนีความแตกต่าง ตามแบบแผนกลุ่มสมมูล รูปแบบการใช้กลุ่มสอบทานผล แสดงได้ดังแผนภาพที่ 3





ภาพที่ 3 แผนผังการจำลองข้อมูล เปรียบเทียบคะแนนและคำนวณค่าดัชนีความแตกต่าง ตามแบบแผนกลุ่มสมมูล รูปแบบการใช้กลุ่มสอบทานผล



## 2.3 การจำลองข้อมูลตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วมภายในกรณีใช้การปรับเทียบกลับสู่แบบสอบเดิม

ในการปรับเทียบตามแบบแผนกลุ่มกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วมภายในใช้การปรับเทียบกลับสู่แบบสอบเดิมต้องมีการจำลองข้อมูลทั้งตามโมเดล 1 พารามิเตอร์ และ 3 พารามิเตอร์ ต้องเปลี่ยนความยาวแบบสอบและขนาดกลุ่มตัวอย่าง ดังนั้นจะต้องมีการจำลองข้อมูลตามเงื่อนไขต่าง ๆ เป็นจำนวนมาก แต่การจำลองแต่ละครั้งใช้วิธีการที่คล้ายคลึงกัน เช่นในกรณีจำลองตามโมเดล 1พารามิเตอร์ ความยาวข้อสอบ 40 ข้อ จำนวนผู้สอบ 500 คน มีขั้นตอนดำเนินการดังนี้

2.3.1 กำหนดลักษณะการแจกแจง ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ผู้สอบให้คงที่ไว้ที่ค่าหนึ่ง คือ ค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานให้เป็น 1 และส่วนการแจกแจงให้เป็นโค้งปกติ

2.3.2 กำหนดจำนวนพารามิเตอร์ให้เป็น 1 พารามิเตอร์

2.3.3 กำหนดจำนวนข้อสอบเป็น 40 ข้อ

2.3.4 กำหนดจำนวนผู้สอบเป็น 500 คน

2.3.5 กำหนดเลขสุ่มของแบบสอบฉบับที่ 1 (X) และกำหนดเลขสุ่มของผู้สอบฉบับนี้ (A) เป็น 1462 และ 4217 ตามลำดับ ส่วนเลขสุ่มกลไกการสอบเป็น 10801 หมายถึงเป็นข้อสอบร่วมภายในมีจำนวนข้อสอบร่วม 8 ข้อ สอบครั้งที่ 1 แล้วจำลองข้อมูล

2.3.6 กำหนดเลขสุ่มของแบบสอบฉบับที่ 2 (Y) เป็น 1462 เช่นเดียวกับแบบสอบฉบับที่ 1 (X) เพราะวัดเนื้อหาเดียวกัน ต่างกันที่กำหนดให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพาราข้อสอบต่างกัน กำหนดเลขสุ่มของผู้สอบฉบับนี้ (B) เป็น 3275 ส่วนเลขสุ่มกลไกการสอบเป็น 10802 หมายถึงเป็นข้อสอบร่วมภายในมีจำนวนข้อสอบร่วม 8 ข้อ สอบครั้งที่ 2 แล้วจำลองข้อมูล

2.3.7 กำหนดเลขสุ่มของแบบสอบฉบับที่ 3 (Z) เป็น 1462 เช่นเดียวกับแบบสอบฉบับที่ 1 (X) เพราะวัดเนื้อหาเดียวกัน ต่างกันที่กำหนดให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพาราข้อสอบต่างกัน กำหนดเลขสุ่มของผู้สอบฉบับนี้ (C) เป็น 6245 ส่วนเลขสุ่มกลไกการสอบเป็น 10801 หมายถึงเป็นข้อสอบร่วมภายในมีจำนวนข้อสอบร่วม 8 ข้อ สอบครั้งที่ 3 แล้วจำลองข้อมูล

2.3.8 ทำสูตรปรับเทียบพารามิเตอร์ข้อสอบจากแบบสอบฉบับ X ไปสู่แบบสอบฉบับ Y ใช้สมการแปลงค่าพารามิเตอร์ดังต่อไปนี้ (Kolen and Brennan, 1995)

$$a^* = \frac{a}{A}$$

$$b^* = Ab + B$$

$$c^* = c$$

การปรับเทียบคะแนนแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม สามารถวิเคราะห์หาค่า A และ B จากค่าพารามิเตอร์ข้อสอบของแบบสอบร่วม ดังต่อไปนี้

1) วิธี Mean and Mean คำนวณหาค่า A และ B จากสมการต่อไปนี้

$$A = \frac{\mu(a_{cx})}{\mu(a_{cy})}$$

$$B = \mu(b_{cy}) - A\mu(b_{cx})$$

2) วิธี Mean and Sigma คำนวณหาค่า A และ B จากสมการต่อไปนี้

$$A = \frac{\sigma(b_{cy})}{\sigma(b_{cx})}$$

$$B = \mu(b_{cy}) - A\mu(b_{cx})$$

3) วิธี Characteristic Curve เป็นวิธีการคำนวณค่าสัมประสิทธิ์ A และ B โดยอาศัยโค้งลักษณะแบบสอบในรูปของฟังก์ชันของผลบวกของความแตกต่างกำลังสองระหว่างโค้งลักษณะข้อสอบ 2 โค้ง กำหนดให้  $P_i(\theta_j)$  แทนความน่าจะเป็นของการตอบข้อสอบร่วม ถูกในแบบสอบที่ไม่ได้ปรับเทียบคะแนน และ  $P_i^*(\theta_j)$  แทน ความน่าจะเป็นของการตอบถูกของข้อสอบร่วมในแบบสอบที่ปรับเทียบคะแนนแล้ว ถ้าหากว่า  $[P_i(\theta_j) - P_i^*(\theta_j)] = 0$  แสดงว่าการปรับเทียบมีความสมบูรณ์ ดังนั้นการปรับเทียบคะแนนคือการหาสัมประสิทธิ์การปรับเทียบ A และ B ได้จากการกำหนดค่าผลต่างจากสมการ

$$F = \frac{1}{N} \sum_{j=1}^N \left\{ \sum_{i=1}^n [P_i(\theta_j) - P_i^*(\theta_j)] \right\}^2$$

การหาค่าสัมประสิทธิ์การปรับเทียบคะแนน A และ B ได้จากการใช้เทคนิค The Iterative Multivariate Search Technique )ทำให้ฟังก์ชัน F ซึ่ง เดวิดออน (Davidon, 1959 cited in Baker, 1997) และ เฟลทเซอร์กับพาวเวลล์ (Fletcher and Powell, 1963 cited in Baker, 1997) พบว่าเมื่อ F ลดลงมากที่สุดจะได้ค่าสัมประสิทธิ์การปรับเทียบ ถ้ามีการปรับเทียบที่สมบูรณ์ ค่า  $F = 0$  ซึ่ง Baker, Al-karni และ Al-Dosory ได้นำมาพัฒนาเป็นโปรแกรม EQUATE โดยใช้ภาษา FORTRAN เพื่อใช้กับคอมพิวเตอร์ขนาดเล็ก ดังที่ได้นำมาใช้ในการวิจัยครั้งนี้

การวิเคราะห์ด้วยโปรแกรม EQUATE ทำการวิเคราะห์โดยใช้เมตริกซ์ความแปรปรวนร่วมที่ได้จากการวิเคราะห์ด้วยโปรแกรม BILOG version 3.04 ครั้งละ 2 เมตริกซ์ จากแบบสอบ 2 ฉบับ ถึงจะได้ค่าสัมประสิทธิ์การปรับเทียบ A และ B จำนวน 1 ชุด

2.3.9 ทำสูตรปรับเทียบพารามิเตอร์ข้อสอบจากแบบสอบฉบับ Y ไปสู่แบบสอบฉบับ Z ใช้กระบวนการเดียวกันกับข้อ 2.3.8

2.3.10 ทำสูตรปรับเทียบพารามิเตอร์ข้อสอบจากแบบสอบฉบับ Z ไปสู่แบบสอบฉบับ X ใช้กระบวนการเดียวกันกับข้อ 2.3.8

2.3.11 ปรับเทียบพารามิเตอร์ข้อสอบ ค่าความยาก (b) ในแบบสอบ X รวมข้อสอบร่วมด้วยทุกข้อ ใช้สูตรการปรับเทียบจากแบบสอบชุด X เข้าสู่สเกลพารามิเตอร์ของแบบสอบชุด Y ได้  $b^*$  ปรับเทียบพารามิเตอร์  $b^*$  ต่อไปโดยใช้ สูตรการปรับเทียบจากแบบสอบชุด Y เข้าสู่สเกลพารามิเตอร์ของแบบสอบชุด Z ได้  $b^{**}$  แล้วปรับเทียบพารามิเตอร์  $b^{**}$  ต่อไปโดยใช้ สูตรการปรับเทียบจากแบบสอบชุด Z เข้าสู่สเกลพารามิเตอร์ของแบบสอบชุด X ได้  $b^{***}$

2.3.12 คำนวณคะแนนจริง (True Score) โดยใช้ค่าพารามิเตอร์ผู้สอบจากผู้สอบกลุ่ม X กับพารามิเตอร์ข้อสอบจากแบบสอบชุด X ที่ไม่ได้ปรับ คือค่าความยาก  $b$  ได้คะแนนจริง  $\tau$  และกับพารามิเตอร์ที่ปรับผ่านแบบสอบฉบับ Y และ Z คือค่าความยาก  $b^{***}$  ได้คะแนนจริง  $\tau'$

2.3.13 ตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนจริง  $\tau$  กับ  $\tau'$  โดยใช้สถิติ t-test paired sample ผลที่ได้จำแนกได้ 3 กรณี คือ กรณีที่ค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01 จัดเป็นกลุ่มเกณฑ์คุณภาพต่ำ กรณีที่ค่า  $t$  จำนวนเท่ากับค่า  $t$  วิฤตจัดเป็นจุดตัด ส่วนกรณีที่ค่าเฉลี่ยไม่แตกต่างจัดเป็นกลุ่มเกณฑ์มีคุณภาพ

2.3.14 นำ  $\tau$  กับ  $\tau'$  ในทั้งสองกลุ่มจากข้อ 2.3.12 ไปคำนวณหาค่าดัชนีความแตกต่าง ดัชนี AMD ดัชนี MAD และ ดัชนี RMS

2.3.15 เปลี่ยนความยาวแบบสอบเป็น 10, 20, 30, 50, 60, 70, 80, ..., 200 แล้วจำลองข้อมูลเช่นเดียวกันกับเมื่อความยาวของแบบสอบ 40 ข้อ

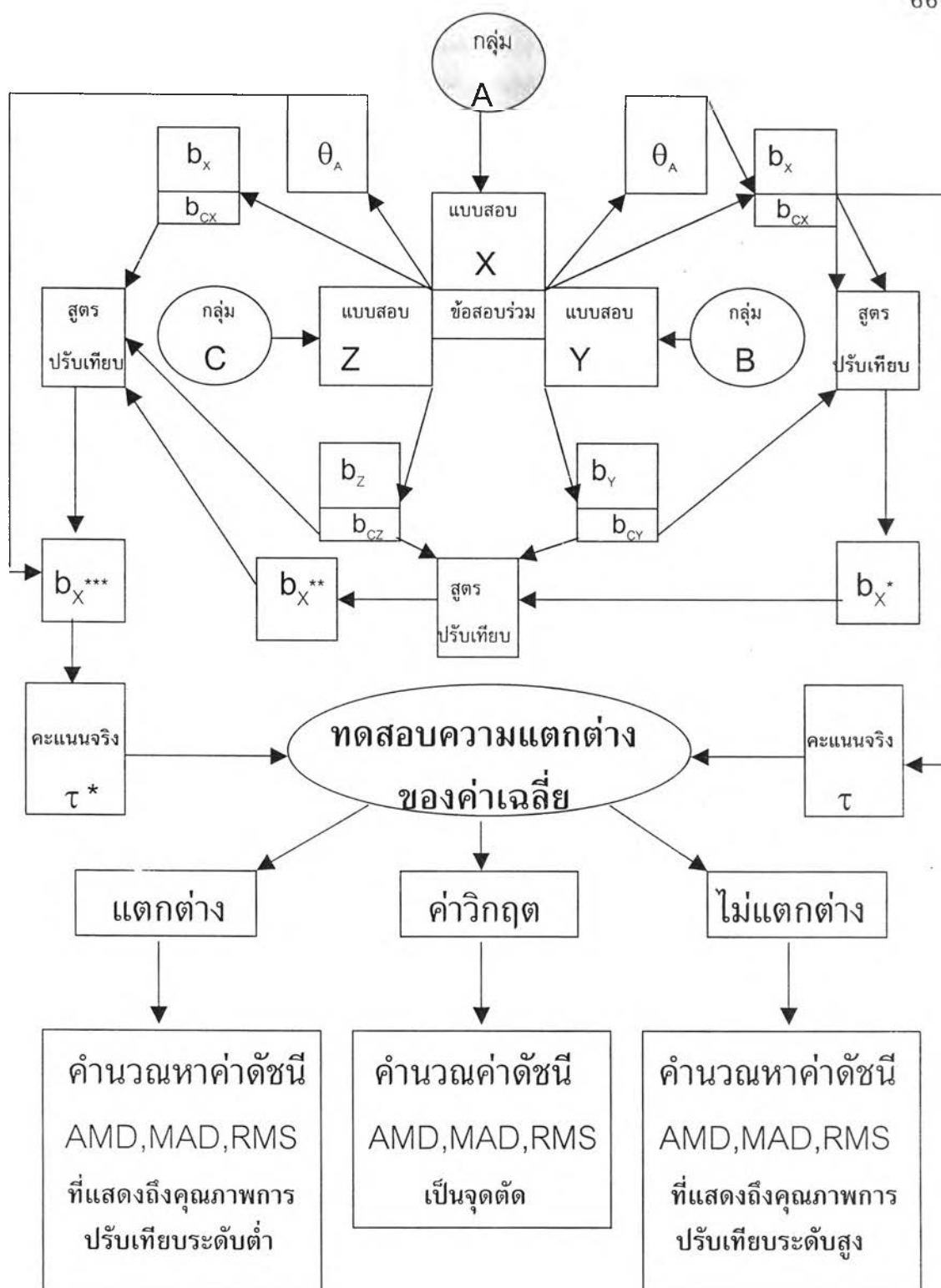
2.3.16 เปลี่ยนขนาดกลุ่มตัวอย่างเป็น 100, 200, 300, 400, 600, ..., 3000 คน แล้วจำลองข้อมูลเช่นเดียวกันกับขนาดกลุ่มตัวอย่าง 500 คน

2.3.17 ทำตามขั้นตอนเดียวกันทั้งหมดนี้กับโมเดล 3 พารามิเตอร์

2.3.18 ซ้ำในขั้นตอนเดียวกันจากข้อ 2.3.1 ถึง ข้อ 2.3.16

2.3.19 จำลองข้อมูลทั้งหมด  $3 \times 2 \times 20 \times 30 \times 2$  เท่ากับ 7,200 ครั้ง ได้ดัชนีความแตกต่าง AMD ดัชนี MAD และ ดัชนี RMS ดัชนีละ 2,400 ตัว

วิธีการจำลองข้อมูล ปรับเทียบคะแนนและหาค่าดัชนีความแตกต่าง ตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม รูปแบบการหาเกณฑ์ใช้การปรับเทียบคะแนนกลับสู่แบบสอบเดิม แสดงได้ดังแผนภาพที่ 4



ภาพที่ 4 แผนผังการจำลองข้อมูล ปรับเทียบคะแนนและหาค่าดัชนีความแตกต่าง ตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบรวมภายใน ในรูปแบบการปรับเทียบคะแนนกลับสู่แบบสอบเดิม

## 2.4 การจำลองข้อมูลตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม ปรับเทียบคะแนนตามรูปแบบการใช้กลุ่มสอบทานผล

ในการปรับเทียบตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม จะต้องมีการจำลองข้อมูลทั้งตามโมเดล 1 พารามิเตอร์ และ 3 พารามิเตอร์ ต้องเปลี่ยนความยาวแบบสอบและขนาดกลุ่มตัวอย่าง ดังนั้นจะต้องมีการจำลองข้อมูลตามเงื่อนไขต่าง ๆ เป็นจำนวนมาก แต่การจำลองแต่ละครั้งใช้วิธีการที่คล้ายคลึงกัน เช่นในการจำลองตามโมเดล 1 พารามิเตอร์ ความยาวข้อสอบ 40 ข้อ จำนวนผู้สอบ 500 คน มีขั้นตอนดำเนินการดังนี้

2.4.1 กำหนดลักษณะการแจกแจง ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของค่าพารามิเตอร์ผู้สอบให้คงที่ไว้ค่าหนึ่ง ค่าเฉลี่ยเท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และส่วนการแจกแจงให้เป็นโค้งปกติ

2.4.2 กำหนดลักษณะการแจกแจง ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ข้อสอบตามแบบแผนที่กำหนด เช่น ค่าเฉลี่ยค่าความยาก ( $b$ ) เท่ากับ 0 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1 และส่วนการแจกแจงให้เป็นโค้งปกติ

2.4.3 กำหนดจำนวนพารามิเตอร์ให้เป็น 1 พารามิเตอร์

2.4.4 กำหนดจำนวนข้อสอบเป็น 40 ข้อ

2.4.5 กำหนดจำนวนผู้สอบเป็น 500 คน

2.4.6 กำหนดเลขสุ่มของแบบสอบฉบับที่ 1 (X) และกำหนดเลขสุ่มของผู้สอบฉบับนี้ (A) เป็น 2462 และ 4217 ตามลำดับ ส่วนเลขสุ่มกลไกการสอบเป็น 10801 หมายถึงเป็นข้อสอบร่วมภายในมีจำนวนข้อสอบร่วม 8 ข้อ สอบครั้งที่ 1 แล้วจำลองข้อมูล

2.4.7 กำหนดเลขสุ่มของแบบสอบฉบับที่ 2 (Y) เป็น 2462 เช่นเดียวกับแบบสอบฉบับที่ 2 (X) เพราะวัดเนื้อหาเดียวกัน ต่างกันที่กำหนดให้ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานของพารามิเตอร์ข้อสอบต่างกัน กำหนดเลขสุ่มของผู้สอบฉบับนี้ (B) เป็น 2475 ส่วนเลขสุ่มกลไกการสอบเป็น 10802 หมายถึงเป็นข้อสอบร่วมภายในมีจำนวนข้อสอบร่วม 8 ข้อ สอบครั้งที่ 2 แล้วจำลองข้อมูล และเลขสุ่มผู้สอบที่เป็นกลุ่มสอบทานผล (C) ที่ทำแบบสอบทั้งสองฉบับเป็น 5126

2.4.8 จำลองข้อมูล

2.4.9 ทำสูตรปรับเทียบพารามิเตอร์ข้อสอบจากแบบสอบฉบับ X ไปสู่แบบสอบฉบับ Y ใช้สมการแปลงค่าพารามิเตอร์ดังในข้อ 2.3.8

2.4.10 ปรับเทียบพารามิเตอร์ข้อสอบ ค่าความยาก ( $b_x$ ) ในแบบสอบ X รวมข้อสอบร่วมด้วยทุกข้อ ที่ได้จากกลุ่มสอบทานผล (C) ใช้สูตรการปรับเทียบจากแบบสอบชุด X เข้าสู่สเกลพารามิเตอร์ของแบบสอบชุด Y ได้  $b_x^*$

2.4.11 คำนวณคะแนนจริง (True Score) โดยใช้ค่าพารามิเตอร์ผู้สอบจากกลุ่มสอบทานผล กับพารามิเตอร์ข้อสอบจากแบบสอบชุด X ที่ไม่ได้ปรับ คือค่าความยาก  $b_x$  ได้คะแนนจริง  $\tau$  และกับค่าความยากที่ปรับสเกลเข้าสู่แบบสอบฉบับ Y คือ  $b_x^*$  ได้คะแนนจริง  $\tau^*$

2.4.12 ตรวจสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนจริง  $\tau$  กับ  $\tau^*$  โดยใช้สถิติ t-test paired sample ผลที่ได้จำแนกได้ 3 กรณี คือ กรณีที่ค่าเฉลี่ยแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ .01 จัดเป็นกลุ่มเกณฑ์คุณภาพต่ำ กรณีที่ค่า  $t$  จำนวนเท่ากับค่า  $t$  วิฤตจัดเป็นจุดตัด ส่วนกรณีที่ค่าเฉลี่ยไม่แตกต่างจัดเป็นกลุ่มเกณฑ์มีคุณภาพ

2.4.13 นำ  $\tau$  กับ  $\tau^*$  ในทั้งสองกลุ่มจากข้อ 2.4.12 ไปคำนวณหาค่าดัชนีความแตกต่าง ดัชนี AMD ดัชนี MAD และ ดัชนี RMS

2.4.14 เปลี่ยนความยาวแบบสอบเป็น 10,20,30, 50, 60, 70, 80,..., 200 แล้วจำลองข้อมูลเช่นเดียวกันกับเมื่อความยาวของแบบสอบ 40 ข้อ

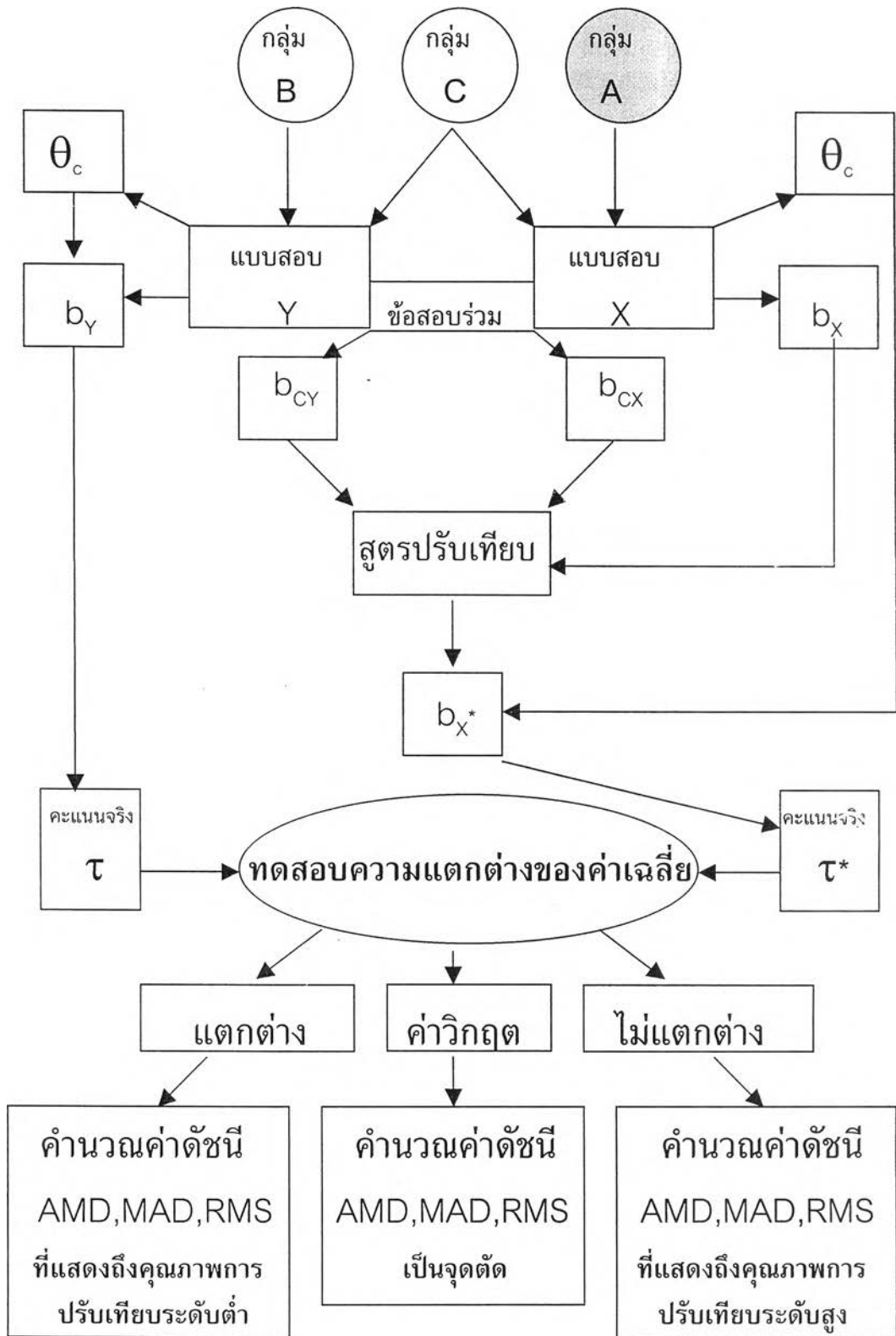
2.4.15 เปลี่ยนขนาดกลุ่มตัวอย่างเป็น 100, 200,300,400, 600,...,3000 คน แล้วจำลองข้อมูลเช่นเดียวกันกับขนาดกลุ่มตัวอย่าง 500 คน

2.4.16 ทำตามขั้นตอนเดียวกันทั้งหมดนี้กับโมเดล 3 พารามิเตอร์

2.4.17 ซ้ำในขั้นตอนเดียวกันจากข้อ 2.4.1 ถึง ข้อ 2.4.16

2.4.18 จำลองข้อมูลทั้งหมด  $4 \times 2 \times 20 \times 30 \times 2$  เท่ากับ 9,600 ครั้ง ได้ดัชนีความแตกต่าง AMD ดัชนี MAD และ ดัชนี RMS ดัชนีละ 2,400 ตัว

วิธีการจำลองข้อมูล ปรับเทียบคะแนนและหาค่าดัชนีความแตกต่าง ตามแบบแผนกลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม รูปแบบการหาเกณฑ์ใช้กลุ่มสอบทานผล แสดงได้ผังแผนภาพที่ 5



ภาพที่ 5 แผนผังการจำลองข้อมูล เปรียบเทียบคะแนนและหาค่าดัชนีความแตกต่างตามแบบแผน กลุ่มไม่เท่าเทียมกันใช้แบบสอบร่วม รูปแบบการหาเกณฑ์ใช้กลุ่มสอบทานผล

### ตอนที่ 3 พัฒนาเกณฑ์จากดัชนีความแตกต่าง

ในการพัฒนาเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนนของดัชนี AMD, MAD และ RMS ผู้วิจัยดำเนินการดังนี้

1. กำหนดจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ โดยพิจารณาจากค่าค่าวิกฤต t-test แบบ Two Dependent Sample Test ที่นัยสำคัญทางสถิติระดับ .01 จากการทดสอบค่าเฉลี่ยระหว่างคะแนนจริงที่ไม่ได้ปรับกับคะแนนจริงที่ปรับเทียบมาจากอีกแบบสอบ แล้วคำนวณค่าดัชนี AMD, ค่าดัชนี MAD และค่าดัชนี RMS ที่กำหนดเป็นจุดตัดจำแนกดัชนีออกเป็น 2 กลุ่ม คือกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ กับกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูง

2. การวิจัยครั้งนี้เป็นการพัฒนาเกณฑ์จากเงื่อนไขที่เป็นไปได้เกือบทั้งหมดของการปรับเทียบคะแนนตามทฤษฎีการตอบสนองข้อสอบ เช่น โมเดลการตอบสนองข้อสอบ การเก็บรวบรวมข้อมูล แบบแผนการหาเกณฑ์ วิธีการปรับเทียบคะแนน จำนวนผู้สอบและจำนวนข้อสอบ ซึ่งทำให้เกิดผลการปรับเทียบที่หลากหลาย แต่เงื่อนไขที่สำคัญและอิทธิพลอย่างมากต่อการปรับเทียบคะแนนและเป็นเงื่อนไขที่นิยมใช้กันมากคือ โมเดลการตอบสนองข้อสอบ แบบแผนการเก็บรวบรวมข้อมูลและแบบแผนการหาเกณฑ์ เพื่อให้ได้เกณฑ์ที่หลากหลายเป็นทางเลือกสำหรับผู้ใช้ผลการปรับเทียบคะแนน ผู้วิจัยจึงจัดดัชนีทั้งในกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ตามเงื่อนไขดังนี้

2.1 จัดรวมเงื่อนไขโมเดลการตอบสนองข้อสอบ การเก็บรวบรวมข้อมูล และแบบแผนการหาเกณฑ์ เป็นเกณฑ์รวม

2.2 จำแนกตามโมเดลการตอบข้อสอบตามทฤษฎีการตอบสนองข้อสอบเป็น 2 เกณฑ์ คือเกณฑ์สำหรับโมเดล 1 พารามิเตอร์ และเกณฑ์สำหรับโมเดล 3 พารามิเตอร์

2.3 จำแนกตามแบบแผนการเก็บรวบรวมข้อมูลเป็น 2 เกณฑ์ คือเกณฑ์สำหรับกลุ่มสมมูล และเกณฑ์สำหรับการใช้ข้อสอบร่วม

2.4 จำแนกตามแบบแผนการหาเกณฑ์ เป็น 2 เกณฑ์ คือ เกณฑ์สำหรับการปรับเทียบกลับสู่แบบสอบเดิม และเกณฑ์สำหรับการใช้กลุ่มสมมูล

ได้กลุ่มดัชนีเพื่อนำไปกำหนดเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนน ทั้ง 7 เกณฑ์ คือ 1) เกณฑ์รวมทุกเงื่อนไข 2) เกณฑ์สำหรับโมเดล 1 พารามิเตอร์ 3) เกณฑ์สำหรับโมเดล 3 พารามิเตอร์ 4) เกณฑ์สำหรับการใช้กลุ่มสมมูล 5) เกณฑ์สำหรับการใช้ข้อสอบร่วม 6) เกณฑ์สำหรับการปรับเทียบกลับสู่แบบสอบเดิม และ 7) เกณฑ์สำหรับการใช้กลุ่มสอบทานผล



3. กำหนดระดับคุณภาพในแต่ละกลุ่มของดัชนี ทั้งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำ สำหรับแต่ละเกณฑ์ทั้ง 7 เกณฑ์

ถ้าผู้วิจัยจัดดัชนีทั้งในกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำให้เป็นเกณฑ์ตัดสินคุณภาพการเปรียบเทียบซึ่งมี 2 ระดับเป็นเกณฑ์ เกณฑ์ที่ได้จะให้สารสนเทศในการตัดสินผลการเปรียบเทียบคะแนนได้น้อย ประกอบกับการพิจารณาจุดตัดที่เป็นค่าดัชนีความแตกต่างที่ได้ทดสอบความแตกต่างระหว่างค่าเฉลี่ยคะแนนจริงที่เปรียบเทียบแล้วกับค่าเฉลี่ยคะแนนจริงที่ยังไม่ได้เปรียบเทียบ ด้วยสถิติ t-test แบบ Two Dependent Sample Test เมื่อค่า t จากการคำนวณมีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติที่ระดับ .01 มีโอกาสที่จะเกิดความคลาดเคลื่อนได้ และโอกาสที่ค่า t คำนวณได้มีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติที่ระดับ .02 จะเกิดขึ้นเมื่อผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของคะแนนทั้งสองไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทำนองเดียวกันค่าที่คำนวณได้มีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติที่ระดับ .001 จะเกิดขึ้นเมื่อผลการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของคะแนนทั้งสองแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และการทดสอบนัยสำคัญทางสถิติที่ระดับ .001 สำหรับการใช้สถิติ t-test ทดสอบความแตกต่างค่าเฉลี่ยสามารถพบเห็นได้ในงานวิจัยทางการศึกษา และนัยสำคัญทางสถิติที่ระดับ .001 ต่างจากนัยสำคัญทางสถิติที่ระดับ .01 เท่ากับ 0.010 หรือคิดเป็นร้อยละ 10 ซึ่งเท่ากับผลต่างระหว่างนัยสำคัญทางสถิติที่ระดับ .01 กับ .02 ดังนั้นผู้วิจัยจึงกำหนดจุดตัดเพื่อแบ่งทั้งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงและกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำดังต่อไปนี้

3.1 จุดตัดที่เป็นค่าดัชนีความแตกต่างที่ได้ทดสอบความแตกต่างระหว่างคะแนนทั้งสองด้วยสถิติ t-test เมื่อค่า t คำนวณมีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติระดับ .02 เพื่อแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย

3.2 จุดตัดที่เป็นค่าดัชนีความแตกต่างที่ได้ทดสอบความแตกต่างระหว่างคะแนนทั้งสองด้วยสถิติ t-test เมื่อค่า t คำนวณมีค่าเท่ากับค่าวิกฤตที่นัยสำคัญทางสถิติระดับ .001 เพื่อแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย

4. กำหนดความหมายช่วงดัชนีความแตกต่าง 4 ช่วง ที่ได้จากจุดตัดแบ่งกลุ่มดัชนีความแตกต่างของแต่ละเกณฑ์ทั้ง 7 เกณฑ์ ดังต่อไปนี้

4.1 ระดับคุณภาพการเปรียบเทียบน่าพอใจอย่างยิ่ง หมายถึง ค่าดัชนีความแตกต่างที่น้อยกว่าดัชนีความแตกต่างที่เป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการเปรียบเทียบระดับสูงออกเป็น 2 กลุ่มย่อย

4.2 ระดับคุณภาพการเปรียบเทียบน่าพอใจ หมายถึง ค่าดัชนีความแตกต่างที่มี

ค่าตั้งแต่ค่าดัชนีความแตกต่างที่เป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูง ออกเป็น 2 กลุ่มย่อย แต่ไม่ถึงค่าดัชนีความแตกต่างที่เป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูงออกจากกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ

4.3 ระดับคุณภาพการปรับเทียบไม่น่าพอใจ หมายถึง ค่าดัชนีความแตกต่างที่มีค่าตั้งแต่ค่าดัชนีความแตกต่างที่เป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับสูง ออกจากกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ แต่ไม่ถึงค่าดัชนีความแตกต่างที่เป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำออกเป็น 2 กลุ่มย่อย

4.4 ระดับคุณภาพการปรับเทียบไม่น่าพอใจอย่างยิ่ง หมายถึง ค่าดัชนีความแตกต่างตั้งแต่ค่าดัชนีความแตกต่างที่เป็นจุดตัดแบ่งกลุ่มดัชนีที่แสดงคุณภาพการปรับเทียบระดับต่ำ ออกเป็น 2 กลุ่มย่อย

#### ตอนที่ 4 ตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้น

1. หาความสัมพันธ์กับเกณฑ์ของ ปีเตอร์เซนและคณะ ที่ได้กำหนดเกณฑ์ตัดสินคุณภาพการปรับเทียบคะแนน ดังนี้

ระดับน่าพอใจอย่างมาก เมื่อ  $MSE < (.05 S_x)^2$

ระดับน่าพอใจ เมื่อ  $(.05 S_x)^2 \leq MSE < (.10 S_x)^2$

ระดับปานกลาง เมื่อ  $(.10 S_x)^2 \leq MSE < (.15 S_x)^2$

ระดับไม่น่าพอใจ เมื่อ  $(.15 S_x)^2 \leq MSE < (.20 S_x)^2$

ระดับไม่น่าพอใจอย่างมาก เมื่อ  $(.20 S_x)^2 \leq MSE$

เมื่อ  $MSE = \frac{\sum fd^2}{nS_x^2}$  และ d เป็นผลต่างของคะแนนที่ปรับแล้วกับไม่ได้ปรับ

โดยการจำลองข้อมูลที่ให้ผลการปรับเทียบคะแนนมีค่าดัชนีที่บอกระดับคุณภาพการปรับเทียบคะแนนในแต่ละช่วงทั้ง 4 ช่วง ทั้ง 7 เกณฑ์ที่พัฒนาขึ้น แล้วเทียบกับเกณฑ์ของปีเตอร์เซนและคณะว่าคุณภาพการปรับเทียบอยู่ในระดับใด เมื่อใช้ข้อมูลชุดเดียวกัน

2. หาความสัมพันธ์กับเกณฑ์ของลอร์ด ที่กล่าวว่าถ้ามีการปรับเทียบคะแนนจากแบบสอบชุดฉบับที่ 1 ไปสู่แบบสอบฉบับที่ 2 อย่างมีคุณภาพดีแล้ว การแจกแจงคะแนนของแบบสอบฉบับที่ 1 ที่ปรับสเกลเข้าสู่แบบสอบชุดที่ 2 ก็กับการแจกแจงคะแนนของแบบสอบชุดที่ 2 จะไม่แตกต่างกัน ในการวิจัยครั้งนี้จึงได้มีการตรวจสอบคุณภาพของเกณฑ์ที่พัฒนาขึ้นดังนี้

2.1 สำหรับการปรับเทียบกลับสู่แบบสอบเดิม ตรวจสอบการแจกแจงของคะแนนจริงของแบบสอบฉบับที่ 1 ที่ปรับผ่านแบบสอบฉบับที่ 2 และฉบับที่ 3 แล้วปรับเข้าสู่แบบสอบฉบับที่ 1 ก็กับการแจกแจงของคะแนนจริงของแบบสอบฉบับที่ 1 ที่ไม่ได้ปรับสเกล มีความแตกต่างกันหรือไม่ โดยใช้สถิติ Wilcoxon sign-rank test วิเคราะห์ด้วย SPSS for

Windows ซึ่งมีการวิเคราะห์จากข้อมูลที่จำลองตามเงื่อนไขการปรับเทียบคะแนนทุกระดับคุณภาพการปรับเทียบจากทั้ง 7 เกณฑ์ที่พัฒนาขึ้น

2.2 สำหรับการปรับเทียบใช้กลุ่มสอบทานผล ตรวจสอบการแจกแจงของคะแนนจริงของกลุ่มสอบทานผลที่สอบแบบสอบฉบับที่ 1 แล้วปรับสเกลเข้าสู่แบบสอบฉบับที่ 2 กับการแจกแจงของคะแนนจริงของแบบสอบฉบับที่ 1 ที่ไม่ได้ปรับสเกล มีความแตกต่างกันหรือไม่ โดยใช้สถิติ Wilcoxon sign-rank test วิเคราะห์ด้วย SPSS for Windows ซึ่งมีการวิเคราะห์จากข้อมูลที่จำลองตามเงื่อนไขการปรับเทียบคะแนนทุกระดับคุณภาพการปรับเทียบจากทั้ง 7 เกณฑ์ที่พัฒนาขึ้น

ค่าดัชนีความแตกต่างที่ใช้ในการวิจัยครั้งนี้คำนวณจากสูตร ดังต่อไปนี้

$$1. \quad \text{AMD} = \frac{\sum_{n=1}^N (\tau - \tau^*)}{N}$$

SD

$$2. \quad \text{MAD} = \frac{\sum_{n=1}^N |\tau - \tau^*|}{N}$$

SD

$$3. \quad \text{RMS} = \sqrt{\frac{\sum_{n=1}^N (\tau - \tau^*)^2}{N}}$$

SD

เมื่อ  $\tau$  คือ คะแนนจริงจากแบบสอบฉบับที่ 1 สำหรับการปรับเทียบกลับสู่แบบสอบเดิม และเป็นคะแนนจริงแบบสอบฉบับที่ 2 สำหรับการปรับเทียบใช้กลุ่มสอบทานผล

$\tau^*$  คือ คะแนนจริงจากแบบสอบฉบับที่ 1 ที่ปรับสเกลผ่านแบบสอบฉบับที่ 2 และฉบับที่ 3 แล้วปรับเข้าสู่แบบสอบฉบับที่ 1 เดิม ในการปรับเทียบกลับสู่แบบสอบเดิม และเป็นคะแนนจริงจากฉบับที่ 1 ที่ปรับสเกลเข้าสู่แบบสอบชุดที่ 2 สำหรับกลุ่มสอบทานผล

$N$  คือ จำนวนผู้สอบ

$SD$  คือ ค่าเฉลี่ยของส่วนเบี่ยงเบนมาตรฐานจากคะแนน  $\tau$  และ คะแนน  $\tau^*$