

### บทที่ 3

#### วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ และเปรียบเทียบค่าความเที่ยงของแบบทดสอบ ที่มีจำนวนข้อสอบทำหน้าที่ต่างกันแตกต่างกัน โดยในการศึกษาครั้งนี้ ผู้วิจัยได้นำข้อมูลทฤษฎีซึ่งเป็นผลการตอบข้อสอบวัดผลสัมฤทธิ์ทางการเรียน ระดับชาติ วิชาภาษาไทย ของกลุ่มตัวอย่างจำนวน 2,000 คน มาจัดกระทำข้อมูลให้มีจำนวนข้อสอบทำหน้าที่ต่างกัน 7 เงื่อนไข ได้แก่ 0%, 5%, 10%, 15%, 20%, 25% และ 30% ในการวิเคราะห์ข้อมูลจะนำข้อมูลผลการตอบข้อสอบมาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี ชิปเทสต์ โดยใช้โปรแกรมชิปเทสต์ จากนั้นนำผลการตรวจสอบมาจัดกระทำข้อมูลให้เป็นไปตามเงื่อนไขจำนวนข้อสอบทำหน้าที่ต่างกันทั้ง 7 เงื่อนไข แล้วนำข้อมูลที่จัดกระทำดังกล่าวมาวิเคราะห์ค่าความเที่ยงของแบบทดสอบ และทำการเปรียบเทียบค่าความเที่ยงของแบบทดสอบ ซึ่งมีวิธีดำเนินการวิจัย ดังนี้

#### ประชากรและกลุ่มตัวอย่าง

##### ประชากร

ประชากร เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2546 ทั้งประเทศ ที่เข้าสอบวิชาภาษาไทย จำนวน 750,978 คน

##### กลุ่มตัวอย่าง

##### 1. การกำหนดขนาดกลุ่มตัวอย่าง

กลุ่มตัวอย่าง เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2546 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาภาษาไทย จำนวน 2,000 คน แบ่งเป็นนักเรียนชาย 1,000 คน และนักเรียนหญิง 1,000 คน ได้มาโดยการสุ่มแบบแบ่งชั้นไม่กำหนดสัดส่วน ใช้ระดับความสามารถ (คะแนนรวม) เป็นชั้นในการแบ่ง

##### 2. การสุ่มตัวอย่าง

ขั้นตอนที่ 1 แบ่งนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เข้าสอบวิชาภาษาไทยเป็นชั้นตามความสามารถ 3 ระดับ ดังนี้

ตารางที่ 1 ช่วงคะแนนที่จำแนกนักเรียนตามกลุ่มความสามารถ และจำนวนนักเรียนในแต่ละระดับความสามารถ

ระดับความสามารถ	ช่วงคะแนน	จำนวนนักเรียน (คน)	คิดเป็นร้อยละ
ดี	26-40	94,189	21.54
พอใช้	16-25	325,635	43.36
ปรับปรุง	0-15	331,154	44.10
รวม		750,978	100.00

ขั้นตอนที่ 2 สุ่มนักเรียนมาระดับละ 666 คน โดยวิธีสุ่มอย่างง่ายใช้คอมพิวเตอร์ในการสุ่มจำแนกเป็นนักเรียนชายและหญิงกลุ่มละเท่า ๆ กัน คือชาย 333 คน และหญิง 333 คน (สุ่มโดยเจ้าหน้าที่สำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ) ได้ผลการสุ่มดังตารางที่ 2

ตารางที่ 2 จำนวนกลุ่มตัวอย่างที่ได้มาโดยการสุ่มแบบแบ่งชั้นไม่กำหนดสัดส่วน จำแนกตามระดับความสามารถ

ระดับความสามารถ	นักเรียน (คน)		รวม
	ชาย	หญิง	
ดี	333	333	666
พอใช้	334	334	668
ปรับปรุง	333	333	666
รวม	1,000	1,000	2,000

### เครื่องมือที่ใช้ในการวิจัย

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 วิชาภาษาไทย ประกอบด้วยข้อสอบแบบ 4 ตัวเลือก จำนวน 40 ข้อ คะแนนเต็ม 40 คะแนน สร้างโดยสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ คณะกรรมการออกข้อสอบ ประกอบด้วย ครูผู้สอนศึกษานิเทศน์ นักวิชาการ และผู้เชี่ยวชาญที่เกี่ยวข้อง มีโครงสร้างของแบบทดสอบ ดังนี้

1. วัดโครงสร้างความรู้ จำนวน 15 ข้อ
2. วัดกระบวนการ จำนวน 25 ข้อ

สำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ ได้ดำเนินการวิเคราะห์คุณภาพของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาภาษาไทย ทั้งรายข้อและรายฉบับ ปรากฏว่าเป็นแบบทดสอบที่มีคุณภาพเหมาะสม คือ

1. มีค่าความยากง่ายเหมาะสม คือ มีข้อสอบค่อนข้างง่าย ประมาณ 25% ยากปานกลาง 50 % และข้อสอบค่อนข้างยาก 25% เฉลี่ยทั้งฉบับมีค่าความยากง่ายประมาณ .50
2. มีค่าอำนาจจำแนกเหมาะสมคือมีค่าอยู่ระหว่าง .20 ขึ้นไป เฉลี่ยทั้งฉบับมีค่าอำนาจจำแนกมากกว่า .40
3. มีค่าความเที่ยงประมาณ .80 ขึ้นไป
4. ครูผู้สอนและผู้เชี่ยวชาญได้ตรวจสอบความตรงเชิงเนื้อหา มีค่าความตรง 100% หรือมีค่า IOC = 1.00

### การเก็บรวบรวมข้อมูล

ผู้วิจัยรวบรวมข้อมูลผลการตอบของนักเรียนที่เป็นกลุ่มตัวอย่าง ซึ่งเป็นข้อมูลทุติยภูมิ จากฐานข้อมูลเก่าของวิทยาลัยการวิจัยและวิทยาการปัญญา ซึ่งได้นำมาจากสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ

### การวิเคราะห์ข้อมูล

#### 1. วิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบ

คำนวณค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ค่าต่ำสุด ค่าสูงสุด ค่าพิสัย (Rank) ค่ามัธยฐาน (Median) ค่าฐานนิยม (Mode) ค่าความเบ้ (Skewness) ค่าความโค้ง (Kurtosis) เพื่อบรรยายลักษณะการแจกแจงของคะแนน โดยใช้โปรแกรม SPSS

#### 2. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีซิปเทสท์ โดยใช้โปรแกรมซิปเทสท์

โปรแกรมซิปเทสท์เป็น โปรแกรมจากสถาบันการวัดผล William Stout ซึ่งเป็น โปรแกรมที่ใช้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ที่มีลักษณะของข้อสอบแบบ Dichotomous (ข้อสอบที่มีการให้คะแนน 0,1) โดยใช้วิธีการทดสอบการทำหน้าที่ต่างกัน โมเดลพหุมิติของ Roussor, Shealy และ Stout

โปรแกรมซิปเทสท์ ถูกออกแบบมาเพื่อวัดและประมาณค่าการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางเดียว โดยใช้หลักเกณฑ์การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป โปรแกรมนี้มีการประมาณค่าทางสถิติและมีการทดสอบสมมติฐาน

โปรแกรมซิปเทสต์ มีจุดมุ่งหมายเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและกลุ่มข้อสอบ ระหว่างกลุ่มผู้สอบสองกลุ่ม (กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ) จากสมมติฐานของการทดสอบและการประมาณค่าจากการวัดการทำหน้าที่ต่างกันของข้อสอบและกลุ่มข้อสอบ โปรแกรมซิปเทสต์ จะประมาณค่าดัชนี  $\beta_{umi}$  ของข้อสอบหรือกลุ่มข้อสอบ ที่มีลักษณะข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว ซึ่งอยู่บนพื้นฐานตามทฤษฎีของ Roussor, Shealy และ Stout โดยใช้โมเดล IRT พหุมิติอธิบาย การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในงานวิจัยนี้มีขั้นตอน ดังนี้

1. หลังจากติดตั้งโปรแกรมแล้ว เลือก SIBTEST
2. นำเข้าข้อมูลทั้งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ โดยกลุ่มอ้างอิง คือ กลุ่มผู้สอบเพศหญิง กลุ่มเปรียบเทียบ คือ กลุ่มผู้สอบเพศชาย เข้าสู่โปรแกรมซิปเทสต์ ซึ่งข้อมูลที่นำเข้าโปรแกรมเป็นข้อมูลผลการตอบข้อสอบแบบ 0.1 บรรจุอยู่ใน File Notepad
3. วิเคราะห์ข้อมูลพร้อมเลือกระดับของผลลัพธ์ที่ต้องการ  
ลักษณะของข้อมูลเมื่อจะทำการ RUN โปรแกรมซิปเทสต์
  1. ข้อมูลจะแยกระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ
  2. ข้อสอบที่แสดงออกมาจะเป็นข้อสอบที่น่าสงสัยและจะทำการจับคู่เพื่อทดสอบ
  3. การวิเคราะห์ประเภทของ DIF/DBF สำหรับกลุ่มข้อสอบที่สงสัย ผู้ใช้สามารถเลือกวิเคราะห์ DIF ทีละข้อใน 1 ครั้ง ได้ทุกข้อที่สงสัยว่า DIF การวิเคราะห์ทุกข้อในกลุ่มข้อที่สงสัยพร้อมกัน เป็นลักษณะของการวิเคราะห์ DBF ซึ่งการวิเคราะห์กลุ่มย่อยของข้อที่สงสัยเป็นการประเมินผลสำหรับ DBF เช่น ในข้อสอบ 25 ข้อ สงสัยข้อที่ 1 – 5 และทำการจับคู่กับข้อ 6 – 25 ผู้ใช้สามารถแยกรันข้อที่สงสัย 1 – 5 ซึ่งผลที่ได้จะบอกถึงค่า DBF หรืออาจรันกลุ่มย่อยทั้งหมดพร้อมกัน
4. การจับคู่ข้อที่สงสัยแบบเฉพาะเจาะจงสำหรับการรันโปรแกรมซิปเทสต์ ผู้ใช้สามารถเลือกเฉพาะข้อที่จับคู่ในข้อ 2 หรือผู้ใช้สามารถเลือกข้อที่เหลือทั้งหมดในแบบทดสอบ เช่น เมื่อต้องการรัน โปรแกรมซิปเทสต์ ข้อ 3 (เฉพาะข้อที่สงสัยข้อ 1 – 5) สามารถประเมินได้เมื่อจับคู่ย่อยกับข้อที่เหลือ (ข้อ 6 – 25) หรือจับทั้งหมด (1,2,3,4 ซึ่งเป็นข้อที่สงสัยกับข้อ 6 – 25)
5. การบรรยายระดับนัยสำคัญโดยค่า p-value ของการทดสอบสมมติฐานโดยวิธีซิปเทสต์ ถ้าค่า p-value มีค่า  $\leq .05$  หรือ ค่า  $\beta_{umi} \neq 0$  แสดงว่าข้อสอบทำหน้าที่ต่างกัน และถ้า  $\beta_{umi} > 0$  แสดงว่าข้อสอบทำหน้าที่ต่างกันโดยเบี่ยงเบนเข้าหากกลุ่มเปรียบเทียบ และถ้าค่า  $\beta_{umi} < 0$  แสดงว่าข้อสอบทำหน้าที่ต่างกันโดยเบี่ยงเบนเข้าหากกลุ่มอ้างอิง

6. จำนวนผู้เข้าสอบที่น้อยที่สุดของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ที่จำเป็นในการจับคู่เพื่อให้โปรแกรมชิปเทสท์วิเคราะห์โดยพิจารณาจากข้อจำกัดของทฤษฎีซึ่งแนะนำว่าจำนวนผู้สอบที่เหมาะสมคือ 20 คน หรือมีผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบกลุ่มละ 10 คน

7. ลักษณะของพารามิเตอร์การเคา การวิเคราะห์ด้วยโปรแกรมนี้อิทธิพลของการเคามีค่าน้อยมาก อาจเป็น 0 สำหรับการตอบข้อสอบอย่างอิสระ หรือไม่นำค่าการเคามาพิจารณาสำหรับข้อสอบแบบเลือกตอบ ค่าการเคาที่ตั้งไว้คือ  $1/(จำนวนตัวเลือก + 1)$  ซึ่งเป็นค่าที่เหมาะสมแล้ว

#### ผลลัพธ์ของการรันโปรแกรมชิปเทสท์

ผลการรันโปรแกรมชิปเทสท์ให้ผลลัพธ์ใน 3 ระดับ จากย่อยสุดจนถึงละเอียดที่สุด ผลลัพธ์ในแต่ละระดับจะให้สารสนเทศทางสถิติจากผลการรันโปรแกรมชิปเทสท์ ดังนี้

#### ผลลัพธ์แบบย่อ

1. เป็นลักษณะเฉพาะของข้อมูลประกอบด้วย จำนวนข้อสอบ ชื่อ File ข้อมูล ขนาด Cell ที่เล็กที่สุด จำนวนการรันโปรแกรมชิปเทสท์ จำนวนผู้เข้าสอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

2. สถิติเกี่ยวกับผู้สอบ

3. สถิติของข้อสอบ (ประกอบด้วย p-value ในทฤษฎีการวัดแบบดั้งเดิม และค่าสหสัมพันธ์)

4. สรุปการรัน โปรแกรมชิปเทสท์ จากการประมาณค่า  $\beta_{um}$  และค่า  $\beta_{cro}$  ค่าคลาดเคลื่อนมาตรฐานและค่า p-value

#### ผลลัพธ์แบบมาตรฐาน

เพิ่มสิ่งต่อไปนี้ในผลลัพธ์แบบย่อ

1. รายการของข้อสอบที่น่าสงสัยและการจับคู่ของข้อสอบที่น่าสงสัย

2. การคำนวณการกระจายโดย  $KR - 20$  สำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

3. ผลสรุปของค่าสถิติในการจับคู่กลุ่มย่อย

#### ผลลัพธ์อย่างละเอียด

เพิ่มสิ่งต่อไปนี้ในผลลัพธ์แบบย่อและแบบมาตรฐาน

1. สำหรับความเป็นไปได้ในการจับคู่กลุ่มย่อย ค่าสถิติที่รายงานคือ

1.1 จำนวนผู้เข้าสอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบกับคะแนนที่จับคู่กัน

1.2 สำหรับการจับคู่คะแนนของกลุ่มผู้เข้าสอบ ค่าเฉลี่ยคะแนนของกลุ่มคะแนน

ที่น่าสงสัยของทั้งสองกลุ่ม

1.3 คะแนนที่แก้ไขได้เหมาะสม โดยการตรวจแก้ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

1.4 ความแตกต่างของคะแนนเฉลี่ยที่ปรับแก้ระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

1.5 นำหนักความแตกต่างของคะแนนปรับแก้ของคะแนนที่น่าสงสัยระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

2. จำนวนคู่ของคะแนนสอบย่อยที่ใช้ได้

3. สัดส่วนคะแนนย่อยที่ใช้สำหรับการจับคู่

### 3. วิเคราะห์เปรียบเทียบค่าเที่ยงของแบบสอบตามเงื่อนไขที่กำหนด

ในการวิจัยนี้วิเคราะห์เปรียบเทียบค่าความเที่ยงของแบบทดสอบ ดังนี้

3.1 นำแบบทดสอบที่วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแล้ว ตัดข้อสอบที่ทำหน้าที่ต่างกันออก หาค่าความเที่ยงของแบบทดสอบตามวิธีของของครอนบัก (Cronbach Alpha Procedure) คือ

$$\alpha = \frac{k}{k-1} \left\{ 1 - \frac{\sum \sigma_i^2}{\sigma^2} \right\}$$

เมื่อ	$\alpha$	คือ	สัมประสิทธิ์ความเที่ยง
	$k$	คือ	จำนวนข้อสอบ
	$\sigma_i^2$	คือ	ค่าความแปรปรวนของคะแนนสอบเป็นรายข้อของแบบทดสอบ
	$\sigma^2$	คือ	ค่าความแปรปรวนของคะแนนรวมของแบบทดสอบทั้งฉบับ

เช่น ข้อสอบ 40 ข้อ มีข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ทำการตัดข้อสอบที่ทำหน้าที่ต่างกันออกทั้ง 10 ข้อ หาค่าความเที่ยงของแบบทดสอบที่เหลือ 30 ข้อ จากนั้นตัดข้อสอบที่ไม่ทำหน้าที่ต่างกันใน 30 ข้อ ออกบางส่วน แล้วนำข้อสอบที่ทำหน้าที่ต่างกันตามเงื่อนไขจำนวนข้อสอบที่ทำหน้าที่ต่างกันเข้ามาแทนที่ข้อสอบที่ตัดออก เช่น ในข้อสอบ 30 ข้อ ต้องการให้มีข้อสอบที่ทำหน้าที่ต่างกัน 10 % หรือ 3 ข้อ ก็ตัดข้อสอบที่ไม่ทำหน้าที่ต่างกันออก 3 ข้อ เหลือข้อสอบ 27 ข้อ นำข้อสอบที่ทำหน้าที่ต่างกัน 3 ข้อ เข้ามาแทน แล้วหาค่าความเชื่อมั่นของแบบทดสอบฉบับที่มีข้อสอบที่ทำหน้าที่ต่างกัน 10%

3.2 เปรียบเทียบค่าความเที่ยงของแบบทดสอบฉบับไม่มีข้อสอบที่ทำหน้าที่ต่างกัน กับฉบับที่มีข้อสอบที่ทำหน้าที่ต่างกันตามเงื่อนไขจำนวนข้อสอบทำหน้าที่ต่างกัน เพื่อตรวจสอบสมมติฐาน โดยการทดสอบสถิติ Z ตามสมการ

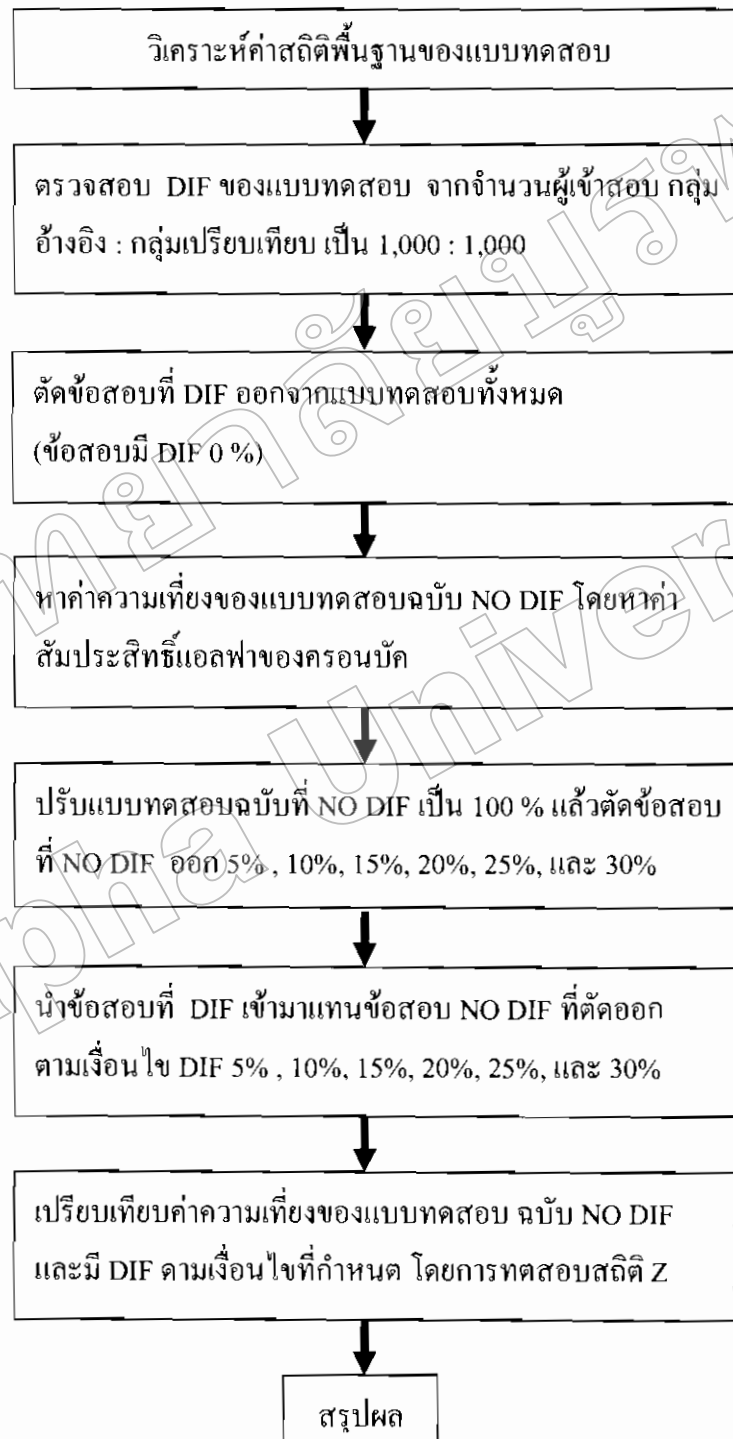
$$Z = \frac{Z_{r1} - Z_{r2}}{\sigma_{Z_{r1}-Z_{r2}}}$$

เมื่อ $Z$	คือ	ค่าแจกแจงแบบ Z (Z-Distribution)
$Z_{r1}$	คือ	ค่าการแปลงการแจกแจงซีของฟิชเชอร์ ของค่าความเที่ยงของแบบทดสอบฉบับที่มีจำนวนข้อสอบทำหน้าที่ต่างกันตามเงื่อนไขที่กำหนด
$Z_{r2}$	คือ	ค่าการแปลงการแจกแจงซีของฟิชเชอร์ ของค่าความเที่ยงของแบบทดสอบฉบับที่ไม่มีข้อสอบทำหน้าที่ต่างกัน
$\sigma_{Z_{r1}-Z_{r2}}$	คือ	ส่วนเบี่ยงเบนมาตรฐานของคะแนนจากการแปลงการแจกแจงซีของฟิชเชอร์ หาได้จาก

$$\begin{aligned}\sigma_{Z_{r1}-Z_{r2}} &= \sqrt{\sigma_{Z_{r1}}^2 + \sigma_{Z_{r2}}^2} \\ &= \sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}\end{aligned}$$

เมื่อ $\sigma_{Z_{r1}}$	คือ	ส่วนเบี่ยงเบนมาตรฐานของคะแนนจากการแปลงการแจกแจงซีของฟิชเชอร์ ของค่าความเที่ยงของแบบทดสอบฉบับที่มีข้อสอบทำหน้าที่ต่างกันตามเงื่อนไขที่กำหนด
$\sigma_{Z_{r2}}$	คือ	ส่วนเบี่ยงเบนมาตรฐานของคะแนนจากการแปลงการแจกแจงซีของฟิชเชอร์ ของค่าความเที่ยงของแบบทดสอบฉบับที่ไม่มีข้อสอบทำหน้าที่ต่างกัน
$n_1$	คือ	จำนวนข้อสอบฉบับที่มีข้อสอบทำหน้าที่ต่างกันตามเงื่อนไขที่กำหนด
$n_2$	คือ	จำนวนข้อสอบฉบับที่ไม่มีข้อสอบทำหน้าที่ต่างกัน

ในการวิจัยนี้สรุปขั้นตอนในการวิจัย ดังภาพที่ 5



ภาพที่ 5 ขั้นตอนการวิจัย