

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการศึกษาอิทธิพลที่เกิดขึ้นจากการตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ โดยศึกษาจากผลการตอบข้อสอบวัดผลฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทย ชั้นประถมศึกษาปีที่ 6 กรณีจำแนกกลุ่มผู้สอบตามตัวแปรเพศ ซึ่งมีวิธีดำเนินการวิจัย ดังนี้

กลุ่มตัวอย่าง

ใช้ข้อมูลทุติยภูมิ ที่เป็นผลการตอบข้อสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 วิชาภาษาไทย ปีการศึกษา 2546 ของนักเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ กลุ่มตัวอย่างเป็นผลการตอบของนักเรียนในเขตพื้นที่การศึกษานครศรีธรรมราช จำนวน 2,000 คน แยกเป็นผลการตอบของนักเรียนเพศชาย จำนวน 1,000 คน ผลการตอบของนักเรียนเพศหญิง จำนวน 1,000 คน ซึ่งกลุ่มตัวอย่างได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) แบบไม่กำหนดสัดส่วน โดยใช้ระดับความสามารถของผู้สอบเป็นชั้น ซึ่งแบ่งระดับดี ระดับพอใช้ และระดับปรับปรุง สุ่มแยกตามเพศ โดยสุ่มมาชั้นละเท่า ๆ กัน ซึ่งการใช้ขนาดกลุ่มตัวอย่าง 2,000 คน เนื่องจากเป็นขนาดที่สามารถตรวจสอบพบข้อสอบทำหน้าที่ต่างกันได้อย่างถูกต้อง แม่นยำ และมีความพอเพียงสำหรับประมาณค่าพารามิเตอร์ได้อย่างถูกต้อง แม่นยำ โดยมีขั้นตอนในการสุ่มกลุ่มตัวอย่าง ดังนี้

ขั้นที่ 1 แบ่งนักเรียนเป็นชั้นตามระดับความสามารถ แยกตามเพศ โดยยึดเกณฑ์ที่สำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการได้แบ่งเอาไว้ เป็น 3 ระดับ คือ ระดับดี (26 - 40 คะแนน) ระดับพอใช้ (16 - 25 คะแนน) และระดับปรับปรุง (0 - 15 คะแนน)

ขั้นที่ 2 สุ่มนักเรียนจากแต่ละชั้นในขั้นที่ 1 มาระดับความสมารถเท่า ๆ กัน โดยสุ่มแยกตามเพศ เพื่อให้ได้กลุ่มตัวอย่างทั้งหมด จำนวน 2,000 คน แยกเป็นเพศชาย 1,000 คน เพศหญิง 1,000 คน โดยใช้คอมพิวเตอร์ในการสุ่ม ได้ผลการสุ่มดังตารางที่ 1

ตารางที่ 1 จำนวนกลุ่มตัวอย่างที่ได้มาโดยการสุ่มแบบแบ่งชั้นแบบไม่กำหนดสัดส่วน จำแนกตามเพศ และระดับความสามารถ

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

เครื่องมือที่ใช้ในการวิจัย

เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ ชั้นประถมศึกษาปีที่ 6 วิชาภาษาไทย ประกอบด้วยข้อสอบแบบ 4 ตัวเลือก จำนวน 40 ข้อ คะแนนเต็ม 40 คะแนน สร้างโดยสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ คณะกรรมการออกข้อสอบประกอบด้วย ครูผู้สอน ศึกษานิเทศก์ นักวิชาการ และผู้เชี่ยวชาญที่เกี่ยวข้อง มีโครงสร้างของแบบทดสอบ ดังนี้

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

สำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ ได้ดำเนินการวิเคราะห์คุณภาพของแบบทดสอบวิชาภาษาไทยทั้งรายข้อและรายฉบับ ปรากฏว่า

1. แบบทดสอบมีค่าความยากเหมาะสม คือ มีข้อสอบค่อนข้างง่าย ประมาณ 25 % ข้อสอบยากปานกลาง 50 % และข้อสอบค่อนข้างยาก 25% (เฉลี่ยทั้งฉบับมีค่าความยากง่าย ประมาณ .50)
2. แบบทดสอบมีค่าอำนาจจำแนกที่ดี ข้อสอบแต่ละข้อจะมีค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป (เฉลี่ยทั้งฉบับมีค่าอำนาจจำแนกมากกว่า .40)
3. แบบทดสอบมีค่าความเที่ยงที่ดี มีค่าค่อนข้างสูง ประมาณ .80 ขึ้นไป
4. ครูผู้สอนและผู้เชี่ยวชาญได้ตรวจสอบความตรงเชิงเนื้อหาของแบบทดสอบ ซึ่งมีความตรงต่อหลักสูตร 100 % (ค่า $IOC = 1.00$)

การเก็บรวบรวมข้อมูล

ผู้วิจัยรวบรวมข้อมูลผลการตอบของนักเรียนที่เป็นกลุ่มตัวอย่าง ซึ่งเป็นข้อมูลทุติยภูมิ จากสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ โดยมีขั้นตอนการเก็บรวบรวมข้อมูล ดังนี้

1. ติดต่อขอหนังสือจากบัณฑิตวิทยาลัย เพื่อขอความร่วมมือในการรวบรวมข้อมูลผลการตอบข้อสอบของนักเรียน จากสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ
2. ขอความร่วมมือจากเจ้าหน้าที่ในการสุ่มข้อมูลผลการตอบของนักเรียนแยกตามระดับความสามารถและตามตัวแปรเพศ เพื่อให้ได้กลุ่มตัวอย่างตามจำนวนที่ต้องการ
3. ตรวจสอบความเรียบร้อยของข้อมูล แล้วนำข้อมูลมาวิเคราะห์ตามวัตถุประสงค์ของการวิจัย

การวิเคราะห์ข้อมูล

1. คำนวณค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (Standard Deviation) ค่าความเบ้ (Skewness) ค่าความโด่ง (Kurtosis) เพื่อบรรยายลักษณะการแจกแจงของคะแนนการสอบวิชาภาษาไทย โดยใช้โปรแกรม SPSS
2. วิเคราะห์ค่าสถิติของข้อสอบรายข้อ และค่าความเที่ยง (Reliability) ของแบบทดสอบวิชาภาษาไทย โดยคำนวณค่าความเที่ยงแบบความสอดคล้องภายใน (Internal Consistency) ใช้สูตรสัมประสิทธิ์แอลฟาของครอนบาค วิเคราะห์โดยใช้โปรแกรม LERTAP 5
3. ตรวจสอบความตรงเชิงโครงสร้างของแบบทดสอบ เนื่องจากเป็นแบบทดสอบมาตรฐานที่เป็นความลับของทางราชการ ผู้วิจัยทราบลักษณะโครงสร้างของแบบทดสอบจากคู่มือดำเนินการสอบ โครงสร้างของแบบทดสอบ วิชาภาษาไทย วัด 2 องค์ประกอบ ดังนี้
 - 3.1 วิชาภาษาไทย คะแนนเต็ม 40 คะแนน (40 ข้อ)
 - วัดโครงสร้างความรู้ (15 ข้อ)
 - วัดกระบวนการ (25 ข้อ)
 ผู้วิจัยดำเนินการวิเคราะห์ความตรงเชิงโครงสร้างตามขั้นตอน ดังต่อไปนี้
 - 3.1.1 คำนวณค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันของข้อสอบแต่ละข้อ โดยใช้โปรแกรมคอมพิวเตอร์ SPSS ในการวิเคราะห์
 - 3.1.2 วิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (Second-Order Confirmatory Factor Analysis) โดยใช้โปรแกรมคอมพิวเตอร์ LISREL 8.50 โดยค่าสถิติที่ใช้ตรวจสอบความสอดคล้องของโมเดลองค์ประกอบกับข้อมูลผลการตอบของผู้สอบ คือ ค่าสถิติไค-สแควร์ (Chi-square: χ^2) ค่าดัชนีวัดระดับความกลมกลืน (Goodness of Fit Index: GFI) ค่าดัชนีวัดระดับ

ความกลมกลืนที่ปรับแก้แล้ว (Adjust Goodness of Fit Index: AGFI) ค่าดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (Comparative Fit Index: CFI) ค่าดัชนีวัดความสอดคล้องปกติ (Normed fit Index: NFI) และค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน โดยประมาณ (Root Mean Squared Error of Approximation: RMSEA)

4. ตรวจสอบความคงที่ของโครงสร้างองค์ประกอบ เมื่อจำแนกกลุ่มประชากร ตามตัวแปรเพศ โดยดำเนินการตามขั้นตอน ดังต่อไปนี้

4.1 กำหนดค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันของข้อสอบแต่ละข้อ กำหนดแยกเป็นกลุ่มผู้สอบเพศชาย ($n = 1000$) และกลุ่มผู้สอบเพศหญิง ($n = 1000$) โดยใช้โปรแกรมคอมพิวเตอร์ SPSS ในการวิเคราะห์

4.2 วิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง โดยใช้โปรแกรมคอมพิวเตอร์ LISREL 8.50 ค่าสถิติที่ใช้ตรวจสอบความสอดคล้องของโมเดลกับข้อมูลผลการตอบของผู้สอบ คือ ค่าสถิติไค-สแควร์ (Chi-square: χ^2) ค่าดัชนีวัดระดับความกลมกลืน (Goodness of Fit Index: GFI) ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (Adjust Goodness of Fit Index: AGFI) ค่าดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (Comparative Fit Index: CFI) ค่าดัชนีวัดความสอดคล้องปกติ (Normed Fit Index: NFI) และค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน โดยประมาณ (Root Mean Squared Error of Approximation: RMSEA)

5. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีชิปเทสท์ที่ปรับใหม่ โดยในการวิเคราะห์จะนำจำนวนผู้สอบกลุ่มอ้างอิง และกลุ่มเปรียบเทียบมาแบ่งกลุ่มผู้สอบออกเป็น 2 กลุ่มตามระดับความสามารถ คือ กลุ่มผู้สอบที่มีความสามารถสูง และกลุ่มผู้สอบที่มีความสามารถต่ำ โดยใช้ค่าเฉลี่ยของคะแนนรวมเป็นเกณฑ์ในการแบ่งกลุ่มผู้สอบ แล้ววิเคราะห์ดัชนีการทำหน้าที่ต่างกันด้วยวิธีชิปเทสท์ในกลุ่มผู้สอบทั้งสอง โดยแยกวิเคราะห์กันคนละกลุ่ม ใช้โปรแกรม SIBTEST ของสเตทท์ และรูสโซ (Stout & Roussos, 1995) ในการวิเคราะห์ ส่วนข้อมูลที่ใช้ป้อนเข้าคือแฟ้มข้อมูลผลการตอบที่ให้คะแนนแบบ 0, 1 ในการตัดสินข้อสอบทำหน้าที่ต่างกัน พิจารณาจากค่านัยสำคัญของดัชนี β_{uni} ถ้าข้อสอบข้อใดมีนัยสำคัญแสดงว่าข้อสอบทำหน้าที่ต่างกัน

6. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีถดถอยโลจิสติก ใช้โปรแกรม SPSS ในการวิเคราะห์ ส่วนข้อมูลที่ใช้ป้อนเข้าเป็นรหัสของกลุ่มผู้สอบ และผลการตอบที่ให้คะแนนแบบ 0, 1 โดยวิธีการตรวจสอบวิธีนี้เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยโมเดลทางสถิติ Logistic Regression เพื่อทำนายโอกาสหรือความน่าจะเป็นในการตอบข้อสอบได้ถูกหรือผิดของผู้สอบ โดยตัวที่ใช้ทำนาย ได้แก่ กลุ่มผู้สอบ (g) ซึ่งแบ่งเป็นกลุ่มอ้างอิง และ

กลุ่มเปรียบเทียบ ความสามารถของผู้สอบ (x) และปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบ และความสามารถของผู้สอบ (g by x) การพิจารณาว่าข้อสอบข้อใดทำหน้าที่ต่างกันพิจารณาจากการมีนัยสำคัญทางสถิติ (Significant) ของค่าปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบ และความสามารถของผู้สอบ (g by x) หรือพิจารณาจากการมีนัยสำคัญทางสถิติของกลุ่มผู้สอบ (g)

7. ตัดข้อสอบที่พบว่าทำหน้าที่ต่างกันออกจากแบบทดสอบ โดยข้อสอบที่ถูกตัดสินว่าทำหน้าที่ต่างกัน และถูกตัดออกจากข้อสอบไม่นำมาคิดคะแนนให้กับผู้สอบ ได้มาจากวิธีตรวจสอบสองวิธีที่พบ DIF ตรงกัน

8. ตรวจสอบคุณภาพของแบบทดสอบ ฉบับหลังจากตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ โดยดำเนินการ ดังนี้

8.1 กำหนดค่าความเที่ยงแบบความสอดคล้องภายในของแบบทดสอบ ฉบับหลังจากตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ โดยใช้สูตรสัมประสิทธิ์แอลฟาของครอนบาค วิเคราะห์โดยใช้โปรแกรม LERTAP 5

8.2 วิเคราะห์ความตรงเชิงโครงสร้างของแบบทดสอบ ฉบับหลังจากตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ ด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง โดยใช้โปรแกรมคอมพิวเตอร์ LISREL 8.50

8.3 ตรวจสอบความคงที่ของโครงสร้างองค์ประกอบ เมื่อจำแนกกลุ่มประชากรตามตัวแปรเพศ แยกเป็นกลุ่มผู้สอบเพศชาย ($n = 1000$) และกลุ่มผู้สอบเพศหญิง ($n = 1000$) ฉบับหลังจากตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ โดยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง โดยใช้โปรแกรมคอมพิวเตอร์ LISREL 8.50

9. เปรียบเทียบค่าความเที่ยง ของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ โดยดำเนินการดังนี้

9.1 ปรับขยายจำนวนข้อสอบของแบบทดสอบฉบับหลัง ให้มีสัดส่วนจำนวนข้อเท่ากับแบบทดสอบฉบับก่อนตัดข้อสอบ แล้วคำนวณค่าความเที่ยงของแบบทดสอบฉบับใหม่ เพื่อให้สามารถเปรียบเทียบค่าความเที่ยงของแบบทดสอบสองฉบับที่มีจำนวนข้อสอบไม่เท่ากันได้ โดยใช้สูตรสเปียร์แมนบราวน์ (Spearman Brown) ดังนี้ (เสรี ชัดเข้ม, 2544, หน้า 135)

$$r'_{11} = \frac{Kr_{11}}{1 + (K - 1)r_{11}}$$

เมื่อ r'_{11} คือ ค่าสัมประสิทธิ์ความเที่ยงของแบบทดสอบฉบับใหม่

r_{11} คือ ค่าสัมประสิทธิ์ความเที่ยงของแบบทดสอบฉบับเดิม

K คือ สัดส่วนระหว่างจำนวนข้อสอบของแบบทดสอบฉบับใหม่
กับจำนวนข้อสอบของแบบทดสอบฉบับเดิม

9.2 ทดสอบนัยสำคัญความแตกต่างของค่าความเที่ยง ของแบบทดสอบฉบับก่อนกับ
แบบทดสอบฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ (ค่าความเที่ยงหลังปรับ
ขยาย) โดยใช้สูตรดังนี้ (Fleht, 1987, p. 99 อ้างถึงใน พิศิษฐ คัณฑวณิช, 2532, หน้า 53)

$$t = \frac{(r_{u1} - r_{u2})\sqrt{N - 2}}{\sqrt{4(1 - r_{u1})(1 - r_{u2})(1 - r_{lu2})}}$$

เมื่อ t คือ ค่าสถิติทดสอบ (t -test)

r_{u1} คือ ค่าความเที่ยงของแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF

r_{u2} คือ ค่าความเที่ยงของแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF

r_{lu2} คือ ค่าสัมประสิทธิ์สหสัมพันธ์ของแบบทดสอบฉบับก่อนกับฉบับหลัง
ตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ

N คือ จำนวนกลุ่มตัวอย่าง

10. เปรียบเทียบคุณภาพด้านความตรงเชิงโครงสร้างของแบบทดสอบ ฉบับก่อนกับ
ฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ โดยพิจารณาเปรียบเทียบจากค่าสถิติ
ไค-สแควร์สอดแทรก (Nested Chi-square), ค่าไค-สแควร์สัมพัทธ์ (Relative Chi-square), ค่าสถิติ
วัดระดับความกลมกลืน (Goodness of Fit Measures) ของแบบทดสอบทั้งสองฉบับ ได้แก่
ค่าสถิติไค-สแควร์ (Chi-square) ค่าดัชนีวัดระดับความกลมกลืน (Goodness of Fit Index: GFI)
ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (Adjust Goodness of Fit Index: AGFI) ค่าดัชนี
วัดระดับความกลมกลืนเปรียบเทียบ (Comparative Fit Index: CFI) ค่าดัชนีวัดความสอดคล้อง
ปกติ (Normed Fit Index: NFI) ค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อน โดยประมาณ
(Root Mean Squared Error of Approximation: RMSEA) และพิจารณาเปรียบเทียบค่าน้ำหนัก
องค์ประกอบของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจาก
แบบทดสอบ

11. เปรียบเทียบความคงที่ของโครงสร้างองค์ประกอบ ของกลุ่มประชากรเพศชาย และ
เพศหญิง ฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ โดยพิจารณา
เปรียบเทียบจากค่าสถิติไค-สแควร์สอดแทรก (Nest Chi-square), ค่าไค-สแควร์สัมพัทธ์ (Relative
Cchi-square), ค่าสถิติวัดระดับความกลมกลืน (Goodness of Fit Measures) ของแบบทดสอบทั้ง

สองฉบับ ได้แก่ ค่าสถิติไค-สแควร์ (Chi-square) ค่าดัชนีวัดระดับความกลมกลืน (Goodness of Fit Index: GFI), ค่าดัชนีวัดระดับความกลมกลืนที่ปรับแก้แล้ว (Adjust Goodness of Fit Index: AGFI) ค่าดัชนีวัดระดับความกลมกลืนเปรียบเทียบ (Comparative Fit Index: CFI) ค่าดัชนีวัดความสอดคล้องปกติ (Normed Fit Index: NFI) ค่ารากของค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนโดยประมาณ (Root Mean Squared Error of Approximation: RMSEA) และเปรียบเทียบค่าน้ำหนักองค์ประกอบของแบบทดสอบฉบับก่อนกับฉบับหลังตัดข้อสอบทำหน้าที่ต่างกันออกจากแบบทดสอบ ซึ่งทำการเปรียบเทียบแบบ Single Group จำนวน 2 กลุ่ม ได้แก่ เพศชาย (ฉบับก่อน) กับ เพศชาย (ฉบับหลัง), เพศหญิง (ฉบับก่อน) กับ เพศหญิง (ฉบับหลัง)

12. จัดลำดับที่ของผู้สอบโดยใช้คะแนนรวมเป็นเกณฑ์ในการตัดสินใจ โดยจัดลำดับทั้งคะแนนจากแบบทดสอบฉบับก่อนตัดข้อสอบที่พบ DIF และคะแนนจากแบบทดสอบฉบับหลังตัดข้อสอบที่พบ DIF จากนั้นก็คำนวณค่าสัมประสิทธิ์สหสัมพันธ์อันดับของผู้สอบ โดยใช้สูตรการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์อันดับของสเปียร์แมน (Spearman Rank Correlation Coefficient) วิเคราะห์โดยใช้โปรแกรม SPSS