

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบพหุมิติ ระหว่างวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก ภายใต้เงื่อนไข 18 เงื่อนไข (3×3×2) คือ ขนาดของกลุ่มตัวอย่าง 3 ขนาด (2,000 คน 1,000 คน และ 300 คน) ความยาวของแบบทดสอบ 3 ขนาด (40 ข้อ 30 ข้อ และ 20 ข้อ) และเกณฑ์การจับคู่ 2 เกณฑ์ (คะแนนรวมทั้งฉบับ และคะแนนแบบทดสอบย่อย) ซึ่งมีขั้นตอนในการดำเนินการวิจัยดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2546 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ จำนวน 29,204 โรงเรียน และมีนักเรียนที่เข้าสอบวิชาภาษาไทย จำนวน 750,978 คน

กลุ่มตัวอย่าง

1. การกำหนดขนาดกลุ่มตัวอย่าง

กลุ่มตัวอย่าง เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ ปีการศึกษา 2546 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทย ได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) แบบไม่กำหนดสัดส่วน โดยแบ่งชั้นตามระดับความสามารถ

เมเซอร์ และคณะ (Mazor et al., 1992, pp. 443-451) พบว่า เมื่อใช้กลุ่มตัวอย่างขนาดใหญ่ทำให้ตรวจพบข้อสอบทำหน้าที่ต่างกันได้มากกว่าการใช้กลุ่มตัวอย่างขนาดเล็ก คือ เมื่อใช้กลุ่มตัวอย่าง 2,000 คน มีความถูกต้องในการตรวจสอบ ร้อยละ 70 ถึง 75 และผลการศึกษาของจิตติมา วรณศรี (2539) พบว่า เมื่อใช้กลุ่มอ้างอิง 1,000 คน และกลุ่มเปรียบเทียบ 1,000 คน สามารถตรวจสอบพบข้อสอบทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 100 และอัตราความคลาดเคลื่อนอยู่ในเกณฑ์ที่เหมาะสม

การวิจัยครั้งนี้ใช้กลุ่มตัวอย่าง 2,000 คน จำแนกเป็นเพศชาย จำนวน 1,000 คน และเพศหญิง จำนวน 1,000 คน ซึ่งการใช้กลุ่มตัวอย่าง กลุ่มละ 1,000 คน เป็นกลุ่มตัวอย่างที่มั่นใจได้ว่า มีขนาดพอเพียงสำหรับการประมาณค่าพารามิเตอร์ ได้อย่างถูกต้องแม่นยำ

2. การสุ่มตัวอย่าง

ขั้นตอนที่ 1 แบ่งนักเรียนชั้นประถมศึกษาปีที่ 6 ที่เข้าสอบวิชาภาษาไทยเป็นชั้นตามระดับความสามารถ (เกณฑ์ที่สำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ ได้แบ่งเอาไว้เป็น 3 ระดับ) ดังตารางที่ 1

ตารางที่ 1 ช่วงคะแนนที่แบ่งนักเรียนตามกลุ่มความสามารถ และจำนวนนักเรียนในแต่ละระดับความสามารถของวิชาภาษาไทย

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

ขั้นตอนที่ 2 สุ่มนักเรียนมาระดับละ 666 คน โดยวิธีสุ่มอย่างง่ายใช้คอมพิวเตอร์ในการสุ่ม จำแนกเป็นนักเรียนชาย และหญิงกลุ่มละเท่า ๆ กัน คือ ชาย 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 วิชาภาษาไทย ชั้นประถมศึกษาปีที่ 6 ประกอบด้วยข้อสอบแบบ 4 ตัวเลือก จำนวน 40 ข้อ คะแนนเต็ม 40 คะแนน สร้างโดยสำนักทดสอบทางการศึกษากระทรวงศึกษาธิการ คณะกรรมการออกข้อสอบประกอบด้วย ครูผู้สอน ศึกษานิเทศก์ นักวิชาการ และผู้เชี่ยวชาญที่เกี่ยวข้อง มีโครงสร้างของแบบทดสอบ ดังนี้

1. วิชาภาษาไทย จำนวน 40 ข้อ
 - 1.1 วัดโครงสร้างความรู้ จำนวน 15 ข้อ
 - 1.2 วัดกระบวนการ จำนวน 25 ข้อ

การสร้างและพัฒนาแบบทดสอบ

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ ชั้นประถมศึกษาปีที่ 6 มีเนื้อหาสาระตามหลักสูตรมัธยมศึกษา พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533) เน้นวัดความเข้าใจ คิดวิเคราะห์ และทักษะด้านกระบวนการ ไม่เน้นวัดความจำ โดยมีครูผู้สอนร่วมเป็นคณะกรรมการออกข้อสอบ

การสร้างและพัฒนาแบบทดสอบ เป็นไปตามกระบวนการสร้างแบบทดสอบมาตรฐาน โดยมีขั้นตอนการสร้างและพัฒนา ดังนี้

1. คณะกรรมการออกข้อสอบประกอบด้วย ครูผู้สอน ศึกษานิเทศก์ นักวิชาการ และผู้เชี่ยวชาญที่เกี่ยวข้อง มีการวิจารณ์ข้อสอบทุกข้อ เพื่อให้ได้ข้อสอบที่มีคุณลักษณะที่ดีในเบื้องต้น

2. นำแบบทดสอบที่ได้ไปทดลองในภาคสนาม (Try-Out) กับนักเรียนที่เป็นกลุ่มตัวอย่าง โดยวิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) เพื่อให้เป็นตัวแทนของสังกัดและขนาดโรงเรียน

3. นำผลที่ได้จากการทดลองมาวิเคราะห์ประมวลผล เพื่อหาค่าสถิติต่างๆ เช่น ค่าความยาก ค่าอำนาจจำแนก และค่าความเที่ยง เป็นต้น โดยนำค่าสถิติที่ได้มาใช้พิจารณาปรับปรุงและพัฒนาข้อสอบจนเป็นแบบทดสอบมาตรฐาน ดังนี้

- 3.1 แบบทดสอบมีค่าความยากเหมาะสม คือ มีข้อสอบค่อนข้างง่าย ประมาณ 25% ข้อสอบยากปานกลาง 50% และข้อสอบค่อนข้างยาก 25% (เฉลี่ยทั้งฉบับมีค่าความยากประมาณ .50)

- 3.2 แบบทดสอบมีค่าอำนาจจำแนกที่ดี ข้อสอบแต่ละข้อมีค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป (เฉลี่ยทั้งฉบับมีค่าอำนาจจำแนกมากกว่า .40)

- 3.3 แบบทดสอบมีค่าความเที่ยงที่ดี มีค่าค่อนข้างสูง ประมาณ .80 ขึ้นไป

4. ครูผู้สอนและผู้เชี่ยวชาญได้ตรวจสอบความตรงเชิงเนื้อหาของแบบทดสอบ ซึ่งมีค่าความตรงเชิงเนื้อหาทั้งหมด 100% (ค่า IOC = 1.00)

คำอธิบายวิธีทำข้อสอบ

1. แบบทดสอบในแต่ละตอนเป็นแบบหลายตัวเลือก มี 4 ตัวเลือก มีข้อคำถาม 40 ข้อ
2. ให้นักเรียนเลือกคำตอบที่ถูกต้องที่สุดหรือเหมาะสมที่สุดเพียงคำตอบเดียว จากตัวเลือก 1, 2, 3 หรือ 4 ที่ให้ไว้ เมื่อนักเรียนเลือกคำตอบได้แล้วให้ใช้ดินสอดำ (2B) ระบายลงใน O ที่มีหมายเลขตรงกับตัวเลือกที่ต้องการลงในกระดาษคำตอบ ดังตัวอย่างการตอบตัวเลือกที่ 3 ดังนี้

0. ① ② ● ④

ถ้านักเรียนต้องการเปลี่ยนคำตอบใหม่เป็นตัวเลือกที่ 4 ให้นักเรียนใช้ยางลบดินสอลบคำตอบเดิมให้สะอาดก่อน แล้วจึงค่อยระบายคำตอบที่ต้องการใหม่ ดังนี้

00. ① ② ③ ●

3. ถ้าพบข้อใดยาก จงเว้นข้ามไปทำข้ออื่น ๆ ต่อไปก่อน เมื่อมีเวลาเหลือจึงค่อยย้อนกลับมาทำข้อที่เว้นไว้ใหม่ เพราะอาจจะมีข้อง่ายอยู่ตอนหลัง ๆ ก็ได้
4. จงระบายคำตอบให้ตรงกับข้อคำถามเสมอ
5. จงพยายามใช้ความคิดและพิจารณาคำตอบจากตัวเลือกให้ดี ไม่ควรเดาเพราะการเดาไม่ทำให้คะแนนดีขึ้น
6. ห้ามขีดเขียนหรือกรอกข้อความอื่นใด นอกเหนือจากที่ระบุไว้ในกระดาษคำตอบ และระวังอย่าให้กระดาษยับ ย่น มีรอยพับ หรือหักงอ
7. ห้ามขีด เขียน ข้อความใด ๆ ในแบบทดสอบ

ตัวอย่างคำถามและวิธีการตอบ (ภาษาไทย)

เป็นแบบทดสอบปรนัยเลือกตอบ 4 ตัวเลือก จำนวน 40 ข้อ ใช้เวลาทำข้อสอบ 50 นาที

ตัวอย่าง (วิชาภาษาไทย)

0. ข้อใดมีตัวสะกดในมาตราแม่กน

- 1) ศศิธร
- 2) สัตตบุษย์
- 3) ม่วงกษัตริย์
- 4) ลาภประเสริฐ

วิธีตอบ ถ้านักเรียนคิดว่าตัวเลือกที่ 1 เป็นคำตอบที่ถูกต้อง หรือเหมาะสมที่สุด
ให้นักเรียนใช้ดินสอระบายลงใน ① ซึ่งตรงกับหมายเลขของตัวเลือกที่ต้องการลงในกระดาษ
คำตอบ ดังนี้

กระดาษคำตอบ

0. ● ② ③ ④

แนวข้อสอบ วิชาภาษาไทย

1. ข้อใดใช้สรรพนามบุรุษที่ 3 ได้ถูกต้อง
- 1) ฉันไปเป็นเพื่อนคุณก็ได้
 - 2) วันเพ็ญเธอทำการบ้านหรือยัง
 - 3) คุณพ่อของผมท่านไม่อยู่บ้านครับ
 - 4) ได้เท่าท่านจะเดินทางวันไหนขอรับ
- (ค่าความยาก = .569 ค่าอำนาจจำแนก = .401)

ความสำเร็จทุกสิ่งจะเป็นผล
ไม่เคยได้ด้วยมนต์การร้องขอ
เพียงหัวใจยังสู้ มือไม่ถอย
ไม่นานวันที่เรารอจะรอเรา

2. จุดประสงค์ของข้อความนี้คือข้อใด
- 1) งานสำเร็จได้ด้วยใจ
 - 2) วันเวลาไม่เคยคอยใคร
 - 3) งานจะลุล่วงได้ต้องลงมือทำ
 - 4) การร้องขอจะทำให้งานได้ผลดี
- (ค่าความยาก = .559 ค่าอำนาจจำแนก = .481)

วิธีเก็บรวบรวมข้อมูล

ผู้วิจัยเก็บข้อมูลผลการตอบข้อสอบของนักเรียนที่เป็นกลุ่มตัวอย่าง ซึ่งเป็นข้อมูลทุติยภูมิ จากสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ โดยมีขั้นตอนการเก็บรวบรวมข้อมูล ดังนี้

1. ขอนหนังสือจากบัณฑิตวิทยาลัย มหาวิทยาลัยบูรพา เพื่อขอความร่วมมือในการรวบรวมข้อมูลผลการตอบข้อสอบของนักเรียนที่เป็นกลุ่มตัวอย่าง จากสำนักทดสอบทางการศึกษา กระทรวงศึกษาธิการ
2. ขอความร่วมมือจากเจ้าหน้าที่ในการส่งข้อมูลผลการตอบข้อสอบของนักเรียน แยกตามเพศ และระดับความสามารถ เพื่อให้ได้กลุ่มตัวอย่างตามจำนวนที่ต้องการ
3. ตรวจสอบความถูกต้อง สมบูรณ์ของข้อมูล แล้วนำข้อมูลที่ได้มาวิเคราะห์ต่อไป

การจัดเตรียมข้อมูลสำหรับวิเคราะห์การหาหน้าที่ต่างกันของข้อสอบ

ขนาดกลุ่มตัวอย่าง

การวิจัยครั้งนี้กำหนดกลุ่มตัวอย่างออกเป็น 3 ขนาด คือ ขนาดใหญ่ (2,000 คน) ขนาดกลาง (1,000 คน) และขนาดเล็ก (300 คน) ซึ่งในการกำหนดขนาดกลุ่มตัวอย่างครั้งนี้ ผู้วิจัยกำหนดจากกลุ่มตัวอย่างขนาดใหญ่ (2,000 คน) โดยวิธีสุ่มอย่างง่ายใช้คอมพิวเตอร์ในการสุ่ม จำแนกเป็น นักเรียนชาย และหญิงกลุ่มละเท่า ๆ กัน

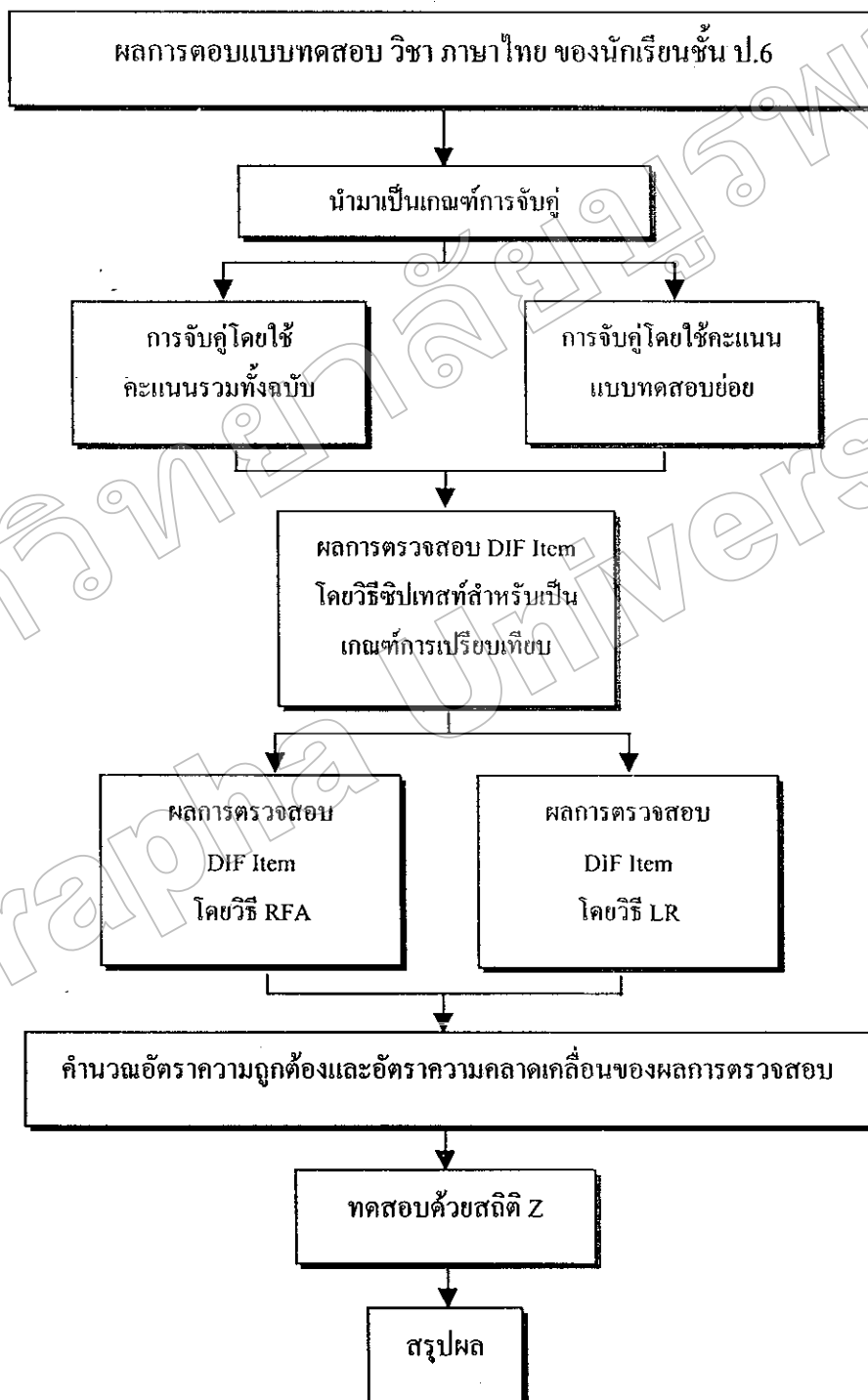
ความยาวของแบบทดสอบ

การวิจัยครั้งนี้กำหนดความยาวของแบบทดสอบออกเป็น 3 ขนาด คือ แบบทดสอบ 40 ข้อ 30 ข้อ และ 20 ข้อ ซึ่งในการกำหนดความยาวของแบบทดสอบครั้งนี้ ผู้วิจัยกำหนดจากแบบทดสอบ 40 ข้อ โดยวิธีสุ่มอย่างง่ายใช้คอมพิวเตอร์ในการสุ่ม ตามโครงสร้างของแบบทดสอบ ดังนี้

1. จำนวน 30 ข้อ
 - 1.1 วัดโครงสร้างความรู้ จำนวน 11 ข้อ
 - 1.2 วัดกระบวนการ จำนวน 19 ข้อ
2. จำนวน 20 ข้อ
 - 2.1 วัดโครงสร้างความรู้ จำนวน 7 ข้อ
 - 2.2 วัดกระบวนการ จำนวน 13 ข้อ

การวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ ผู้วิจัยดำเนินการวิเคราะห์ข้อมูล ตามขั้นตอนดังภาพที่ 4



ภาพที่ 4 ขั้นตอนการวิเคราะห์ข้อมูล ตามเกณฑ์การจับคู่

ตารางที่ 3 เงื่อนไขการตรวจสอบข้อสอบทำหน้าที่ต่างกัน ระหว่างวิธีการวิเคราะห์องค์ประกอบ
จำกัดกับวิธีถดถอยโลจิสติก

ขนาดกลุ่มตัวอย่าง	ความยาวของแบบทดสอบ	เกณฑ์การจับคู่
2,000 คน	40 ข้อ	คะแนนรวม (40 ข้อ)
		คะแนนแบบทดสอบย่อย (15 ข้อ, 25 ข้อ)
	30 ข้อ	คะแนนรวม (30 ข้อ)
		คะแนนแบบทดสอบย่อย (11 ข้อ, 19 ข้อ)
	20 ข้อ	คะแนนรวม (20 ข้อ)
		คะแนนแบบทดสอบย่อย (7 ข้อ, 13 ข้อ)
1,000 คน	40 ข้อ	คะแนนรวม (40 ข้อ)
		คะแนนแบบทดสอบย่อย (15 ข้อ, 25 ข้อ)
	30 ข้อ	คะแนนรวม (30 ข้อ)
		คะแนนแบบทดสอบย่อย (11 ข้อ, 19 ข้อ)
	20 ข้อ	คะแนนรวม (20 ข้อ)
		คะแนนแบบทดสอบย่อย (7 ข้อ, 13 ข้อ)
300 คน	40 ข้อ	คะแนนรวม (40 ข้อ)
		คะแนนแบบทดสอบย่อย (15 ข้อ, 25 ข้อ)
	30 ข้อ	คะแนนรวม (30 ข้อ)
		คะแนนแบบทดสอบย่อย (11 ข้อ, 19 ข้อ)
	20 ข้อ	คะแนนรวม (20 ข้อ)
		คะแนนแบบทดสอบย่อย (7 ข้อ, 13 ข้อ)

ผู้วิจัยดำเนินการวิเคราะห์ข้อมูลดังนี้

1. จัดเตรียมเพิ่มข้อมูลสำหรับการวิเคราะห์ โดยแยกผลการตอบข้อสอบตามกลุ่มผู้เข้าสอบ ได้แก่ กลุ่มเปรียบเทียบ และกลุ่มอ้างอิง
2. วิเคราะห์ค่าสถิติพื้นฐาน เพื่อบรรยายการแจกแจงของคะแนนของแบบทดสอบความสามารถทางวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 6 โดยใช้โปรแกรม SPSS
3. หาความตรงเชิงโครงสร้างของแบบทดสอบความสามารถทางวิชา ภาษาไทย ชั้นประถมศึกษาปีที่ 6 ด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน โดยใช้โปรแกรม LISREL 8.50
4. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีชิปเทสต์ ในแบบทดสอบวิชาภาษาไทย เมื่อกลุ่มตัวอย่างมีจำนวน 2,000 คน 1,000 คน และ 300 คน และความยาวของแบบทดสอบ 40 ข้อ 30 ข้อ และ 20 ข้อ โดยใช้คะแนนรวมทั้งฉบับ และคะแนนแบบทดสอบย่อย โดยใช้โปรแกรม SIBTEST ซึ่งผลที่ได้คือ ค่าดัชนี β_{SIB} และการทดสอบนัยสำคัญทางสถิติ โดยพิจารณาจากค่า p-value ถ้าพบว่ามีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าข้อสอบนั้นทำหน้าที่ต่างกัน (DIF) ซึ่งในการวิจัยครั้งนี้เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธีการวิเคราะห์

องค์ประกอบจำกัดกับวิธีดัดแปลงโลจิสติก การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีชิปเทสต์ ซึ่งใช้เป็นเกณฑ์สำหรับการเปรียบเทียบประสิทธิภาพของวิธีตรวจสอบ DIF

5. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์องค์ประกอบจำกัด ในแบบทดสอบวิชาภาษาไทย เมื่อกลุ่มตัวอย่างมีจำนวน 2,000 คน 1,000 คน และ 300 คน และความยาวของแบบทดสอบ 40 ข้อ 30 ข้อ และ 20 ข้อ โดยใช้คะแนนรวมทั้งฉบับ และคะแนนแบบทดสอบย่อย แบ่งออกเป็น 2 ขั้นตอน ดังนี้

5.1 เตรียมข้อมูลให้อยู่ในรูปเมทริกซ์สหสัมพันธ์ โดยใช้โปรแกรม SPSS

5.2 ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้โปรแกรม LISREL 8.50 โดยพิจารณาค่าดัชนีดัดแปรโมเดล (Modification Indices: MI) และค่าพารามิเตอร์ที่คาดหวัง (Expected Parameter Change: EPC) ว่าแตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 หรือไม่ โดย ออร์ต (Oort, 1998) เสนอแนะว่าให้ตัดข้อสอบที่มีค่า MI และดัชนี EPC ที่แตกต่างจาก 0 ออกจากแบบทดสอบชุดนั้นทีละข้อ โดยตัดข้อสอบที่มีค่า MI สูงที่สุดออกก่อน นำข้อสอบที่เหลือมาวิเคราะห์ซ้ำไปเรื่อย ๆ จนกว่าไม่มีข้อสอบที่มีค่า MI และดัชนี EPC ที่แตกต่างจาก 0 เหลืออยู่ในแบบทดสอบชุดนั้น

6. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีดัดแปลงโลจิสติก ในแบบทดสอบวิชาภาษาไทย เมื่อกลุ่มตัวอย่างมีจำนวน 2,000 คน 1,000 คน และ 300 คน และความยาวของแบบทดสอบ 40 ข้อ 30 ข้อ และ 20 ข้อ โดยใช้คะแนนรวมทั้งฉบับ และคะแนนแบบทดสอบย่อย โดยนำแฟ้มข้อมูลผลการตอบข้อสอบของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบไปวิเคราะห์โดยใช้โปรแกรม SPSS ซึ่งวิธีดัดแปลงโลจิสติกเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ทดสอบด้วยโมเดลดัดแปลงโลจิสติก เพื่อทำนายโอกาสในการตอบข้อสอบได้ถูกต้อง โดยพิจารณาจากการมีนัยสำคัญทางสถิติของค่าปฏิสัมพันธ์ระหว่างกลุ่มผู้เข้าสอบและความสามารถของผู้เข้าสอบ (g by x) หรือพิจารณาจากการมีนัยสำคัญทางสถิติของกลุ่มผู้สอบ

7. คำนวณอัตราความถูกต้องและอัตราความคลาดเคลื่อนในการตรวจสอบ โดยนำผลการวิเคราะห์ด้วยวิธีการวิเคราะห์องค์ประกอบจำกัด กับวิธีดัดแปลงโลจิสติก มาเปรียบเทียบกับผลการตรวจสอบด้วยวิธีชิปเทสต์ โดยมีเกณฑ์ว่า ข้อสอบที่ถูกตรวจพบว่าทำหน้าที่ต่างกันด้วยวิธีชิปเทสต์ถือว่าเป็นข้อสอบทำหน้าที่ต่างกัน หากวิธีการวิเคราะห์องค์ประกอบจำกัด และวิธีดัดแปลงโลจิสติกตรวจพบว่าข้อสอบทำหน้าที่ต่างกันตรงกับข้อที่ตรวจพบด้วยวิธีชิปเทสต์ ถือว่าวิธีนั้นมีความถูกต้องในการตรวจสอบ แต่ถ้าตรวจพบว่าข้อสอบข้อใดทำหน้าที่ต่างกันไม่ตรงกับข้อที่ตรวจพบด้วยวิธีชิปเทสต์ ถือว่ามีความคลาดเคลื่อนในการตรวจสอบ

7.1 การคำนวณอัตราความถูกต้อง

นำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการวิเคราะห์องค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก เมื่อใช้เกณฑ์การจับคู่ต่างกันมาเปรียบเทียบกับผลการตรวจสอบด้วยวิธีชิปเทสต์ ซึ่งใช้เป็นเกณฑ์สำหรับการเปรียบเทียบประสิทธิภาพของวิธีตรวจสอบ DIF ดังนั้นถ้าข้อสอบข้อใดถูกตรวจพบว่าทำหน้าที่ไม่ต่างกันด้วยวิธีชิปเทสต์จะถือว่าเป็นข้อสอบทำหน้าที่ไม่ต่างกัน ถ้าวิธีการวิเคราะห์องค์ประกอบจำกัดหรือวิธีถดถอยโลจิสติกตรวจพบข้อสอบทำหน้าที่ไม่ต่างกันตรงกับที่ตรวจพบด้วยวิธีชิปเทสต์ ถือว่าวิธีนั้นมีความถูกต้องประเภทที่ 1 ในการตรวจสอบ และถ้าตรวจพบข้อสอบข้อใดทำหน้าที่ต่างกันตรงกับที่ตรวจพบด้วยวิธีชิปเทสต์ ถือว่าวิธีนั้นมีความถูกต้องประเภทที่ 2 ในการตรวจสอบ การคำนวณอัตราความถูกต้องในการตรวจสอบประเภทที่ 1 และ 2 ดังนี้

อัตราความถูกต้องประเภทที่ 1

$$= \frac{\text{จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ไม่ต่างกันตรงกับวิธีชิปเทสต์}}{\text{จำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบทดสอบ}}$$

อัตราความถูกต้องประเภทที่ 2

$$= \frac{\text{จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันตรงกับวิธีชิปเทสต์}}{\text{จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ}}$$

7.2 การคำนวณอัตราความคลาดเคลื่อน

นำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีการวิเคราะห์องค์ประกอบจำกัด (RFA) กับวิธีถดถอยโลจิสติก (LR) เมื่อใช้เกณฑ์การจับคู่ต่างกันมาเปรียบเทียบกับผลการตรวจสอบด้วยวิธีชิปเทสต์ ถ้าข้อสอบข้อใดวิเคราะห์ด้วยวิธี RFA หรือวิธี LR ตรวจพบว่าทำหน้าที่ต่างกัน แต่วิธีชิปเทสต์ตรวจพบว่าเป็นข้อสอบที่ทำหน้าที่ไม่ต่างกัน ถือว่าวิธีนั้นมีการระบุผิดพลาดทำให้เกิดความคลาดเคลื่อนประเภทที่ 1 และถ้าวิธี RFA หรือวิธี LR ตรวจพบข้อสอบข้อใดทำหน้าที่ไม่ต่างกัน แต่วิธีชิปเทสต์ตรวจพบว่าเป็นข้อสอบทำหน้าที่ต่างกัน ถือว่าวิธีนั้นมีการระบุผิดพลาดซึ่งทำให้เกิดความคลาดเคลื่อนประเภทที่ 2 การคำนวณอัตราความคลาดเคลื่อนในการตรวจสอบประเภทที่ 1 และ 2 ดังนี้

อัตราความคลาดเคลื่อนประเภทที่ 1

$$= \frac{\text{จำนวนข้อสอบที่วิธี RFA หรือวิธี LR ระบุผิดพลาดว่าทำหน้าที่ต่างกัน}}{\text{จำนวนข้อสอบที่วิธี SIBTEST ระบุว่าทำหน้าที่ไม่ต่างกันทั้งหมดในแบบทดสอบ}}$$

อัตราความคลาดเคลื่อนประเภทที่ 2

$$= \frac{\text{จำนวนข้อสอบที่วิธี RFA หรือวิธี LR ระบุผิดพลาดว่าทำหน้าที่ไม่ต่างกัน}}{\text{จำนวนข้อสอบที่วิธี SIBTEST ระบุว่าทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ}}$$

8. การทดสอบอัตราความถูกต้องและอัตราความคลาดเคลื่อนในการตรวจสอบด้วยสถิติ Z จึงเป็นการเปรียบเทียบความแตกต่างของอัตราความถูกต้องและอัตราความคลาดเคลื่อนในการตรวจสอบระหว่างวิธีการวิเคราะห์หองค์ประกอบจำกัดกับวิธีถดถอยโลจิสติก โดยการทดสอบด้วยค่าสถิติ Z ที่ระดับนัยสำคัญ .05 โดยใช้สูตรดังนี้

$$Z_p = \frac{|p_1 - p_2|}{\sqrt{PQ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

เมื่อ p_1, p_2 เป็นสัดส่วนที่ได้จากการตรวจสอบด้วยวิธี RFA และวิธี LR ตามลำดับ

$$P \quad \text{เท่ากับ} \quad \frac{(P_1)(n_1) + (P_2)(n_2)}{n_1 + n_2}$$

$$Q \quad \text{เป็นค่าที่ได้จาก} \quad 1 - P$$

$$n_1, n_2 \quad \text{จำนวนของข้อสอบที่ตรวจด้วยวิธี RFA และวิธี LR}$$