

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อหาประสิทธิภาพของโมเดลในการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายใน ที่พัฒนาตามทฤษฎีเขาวัวปัญญาของชาร์ล สเปียร์แมน และหาปัจจัยที่มีอิทธิพลต่อค่าดัชนีวัดความสอดคล้องซึ่งได้แก่ จำนวนกลุ่มตัวอย่าง จำนวนข้อสอบต่อองค์ประกอบ เพื่อให้ได้โมเดลที่สอดคล้องกับทฤษฎีและเกณฑ์ที่เหมาะสมในการตัดสินใจสำหรับการตรวจสอบความสอดคล้องของโมเดลกับข้อมูลเชิงประจักษ์ในการวิเคราะห์องค์ประกอบเชิงยืนยันสำหรับข้อมูลพหุมิติโดยการเปรียบเทียบกับโมเดลที่นิยมใช้อยู่ในปัจจุบันและตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยผู้วิจัยได้ออกแบบการวิจัยดังนี้

ประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 สังกัดสำนักงานเขตพื้นที่การศึกษานุริรัมย์ เขต 2 ปีการศึกษา 2548 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ จำนวน 114 โรงเรียน และมีนักเรียนที่เข้าสอบวิชาภาษาไทย คณิตศาสตร์ สังคมศึกษา ศาสนาและวัฒนธรรม วิทยาศาสตร์ และภาษาอังกฤษ จำนวน 5,011 คน

กลุ่มตัวอย่าง

1. การกำหนดขนาดกลุ่มตัวอย่างในการหาประสิทธิภาพของโมเดลการวิเคราะห์องค์ประกอบเชิงยืนยันเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 สังกัดสำนักงานเขตพื้นที่การศึกษานุริรัมย์ เขต 2 ปีการศึกษา 2548 ที่เข้าสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติโดยมีขั้นตอนดังนี้

ขั้นที่ 1 แบ่งนักเรียนออกเป็นกลุ่มตามระดับความสามารถ (เกณฑ์สำนักทดสอบกระทรวงศึกษาธิการ) โดยแบ่งชั้นตามระดับความสามารถดังนี้

ตารางที่ 3-1 ช่วงคะแนนที่ใช้แบ่งนักเรียนตามกลุ่มความสามารถของแต่ละรายวิชา

ระดับความสามารถ	ช่วงคะแนน
ปรับปรุง	0-60
พอใช้	61-100
ดี	100-200

ขั้นที่ 2 สุ่มนักเรียนระดับความสามารถดี คิดเป็นร้อยละ 33 ความสามารถพอใช้คิดเป็นร้อยละ 34 และความสามารถปรับปรุง คิดเป็นร้อยละ 33 คัดเลือกกลุ่มตัวอย่างด้วยวิธีการสุ่มอย่างง่าย ให้มีกลุ่มตัวอย่างขนาดใหญ่พอสำหรับการวิเคราะห์ ชูมาร์คเกอร์ และ โลแม็กซ์ (Schumacher & Lomax, 1960) ได้สรุปผลงานวิจัยการศึกษาขนาดกลุ่มตัวอย่างที่เหมาะสมกับการวิเคราะห์องค์ประกอบว่าส่วนใหญ่ใช้กลุ่มตัวอย่างขนาด 100 – 200 และให้ผลการวิจัยเป็นที่น่าพอใจ บูมส์มา (Boomsma, 1983) ได้เสนอขนาดกลุ่มตัวอย่างที่เหมาะสม ควรเป็น 400 และนอกจากนี้ ชูมาร์คเกอร์ และ โลแม็กซ์ (Schumacker & Lomax, 1996) แฮร์ และคณะ (Hair et al., 1998) ได้เสนอกฎแห่งความชัดเจน (Rule of Thumb) ที่นักวิเคราะห์ทางสถิติใช้กันมากในการวิเคราะห์โมเดลสมการ โครงสร้าง คือขนาดกลุ่มตัวอย่าง 10 – 20 คนต่อตัวแปรในการวิจัยหนึ่งตัวแปรในการวิจัยครั้งนี้จึงใช้กลุ่มตัวอย่างจำนวนเท่ากับ 800 คน

ขั้นที่ 3 สุ่มตัวอย่างตามเงื่อนไขของการทดสอบจำนวน 100, 200, 400, 800, 1,200, 1,600 และ 2,000 คน โดยใช้วิธีตามขั้นตอนที่ 2

2. การกำหนดขนาดกลุ่มตัวอย่างในการทดสอบการทำหน้าที่ต่างกันของข้อสอบ

2.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีระหว่างการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายในใช้กลุ่มตัวอย่างตามแนวคิดเมเซอร์ 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 สามารถตรวจสอบข้อสอบทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 100 และอัตราความคลาดเคลื่อนอยู่ในเกณฑ์ที่เหมาะสม การวิจัยครั้งนี้ใช้กลุ่มตัวอย่าง 2,000 คน จำแนกเป็นเพศชาย จำนวน 1,052 คน และเพศหญิง จำนวน 948 คน ได้มาโดยการสุ่มอย่างง่ายเป็นกลุ่มตัวอย่างที่มั่นใจได้ว่ามีขนาดพอเพียงสำหรับการประมาณค่าพารามิเตอร์ ได้อย่างถูกต้องแม่นยำ

2.2 การศึกษาประสิทธิภาพการทำหน้าที่ต่างกันของข้อสอบโดยใช้เงื่อนไขโดยใช้กลุ่มตัวอย่างขนาด 1,600, 2,000, 2,400, 2,800, 3,200, 3,600 และ 4,000 คนได้มาโดยวิธีการสุ่มอย่างง่ายจากนักเรียนที่มีความสามารถในระดับร้อยละ 33 ระดับพอใช้ร้อยละ 34 และระดับปรับปรุงร้อยละ 33

เครื่องมือที่ใช้ในการวิจัย

1. เครื่องมือที่ใช้ในการวิจัย

สำหรับเครื่องมือที่ใช้ในการวิจัยครั้งนี้ได้แก่แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับชาติ วิชาภาษาไทย ชั้นมัธยมศึกษาปีที่ 3 ประกอบด้วยข้อสอบแบบ 4 ตัวเลือก จำนวน 200 ข้อ คะแนนเต็ม 200 คะแนน สร้างโดยสำนักทดสอบ คณะกรรมการออกข้อสอบประกอบด้วย ครูผู้สอนศึกษานิเทศก์ นักวิชาการ และผู้เชี่ยวชาญที่เกี่ยวข้อง มีรายละเอียดของของแบบทดสอบในการสร้างแบบทดสอบมุ่งวัดความสามารถด้านภาษาไทย คณิตศาสตร์ สังคมศึกษา ศาสนาและวัฒนธรรม วิทยาศาสตร์และภาษาอังกฤษ ตามวัตถุประสงค์ที่กำหนดในหลักสูตรชั้นมัธยมศึกษาปีที่ 3 เป็นความสามารถหลัก (θ) ส่วนคุณลักษณะอื่น ๆ เช่น ความสามารถในการอ่านและแก้ปัญหา โจทย์ที่มีความซับซ้อน ไหวพริบในการทำข้อสอบ ความคุ้นเคยหรือประสบการณ์ในสถานการณ์ของโจทย์ที่ผู้ตอบข้อสอบนำมาใช้ในการตอบข้อสอบ ได้ถูกต้อง จัดเป็นความสามารถรอง (η) ซึ่งประกอบด้วยข้อสอบดังนี้

1. วิชาภาษาไทย	จำนวน 40 ข้อ
2. คณิตศาสตร์	จำนวน 40 ข้อ
3. สังคมศึกษา ศาสนาและวัฒนธรรม	จำนวน 40 ข้อ
4. วิทยาศาสตร์	จำนวน 40 ข้อ
5. ภาษาอังกฤษ	จำนวน 40 ข้อ

2. ขั้นตอนการสร้างและพัฒนาแบบทดสอบ

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนเพื่อประเมินคุณภาพการศึกษาระดับชาติ ชั้นมัธยมศึกษาปีที่ 3 มีเนื้อหาสาระตามหลักสูตรมัธยมศึกษา พุทธศักราช 2542 (ฉบับปรับปรุง พ.ศ. 2544) เน้นความเข้าใจ คิดวิเคราะห์ และทักษะด้านกระบวนการไม่เน้นวัดความจำ โดยมีครูผู้สอนร่วมเป็นคณะกรรมการออกข้อสอบ การสร้างและพัฒนาแบบทดสอบ เป็นไปตามกระบวนการสร้างแบบทดสอบมาตรฐาน โดยมีขั้นตอนการสร้างและการพัฒนา ดังนี้

2.1 คณะกรรมการออกข้อสอบประกอบด้วย ครูผู้สอน ศึกษานิเทศก์ นักวิชาการและผู้เชี่ยวชาญที่เกี่ยวข้อง มีการวิจารณ์ข้อสอบทุกข้อ เพื่อให้ได้ข้อสอบที่มีคุณลักษณะที่ดีในเบื้องต้น

2.2 นำแบบทดสอบที่ได้ไปทดสอบในภาคสนาม (Try-Out) กับนักเรียนที่เป็นกลุ่มตัวอย่างโดยวิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) เพื่อให้เป็นตัวแทนของสังกัดและขนาดโรงเรียน

2.3 นำผลที่ได้จากการทดลองมาวิเคราะห์ประมวลผล เพื่อหาค่าสถิติต่าง ๆ เช่น ค่าความยาก ค่าอำนาจจำแนก และค่าความเที่ยง เป็นต้น โดยนำค่าสถิติที่ได้มาใช้พิจารณาปรับปรุงและพัฒนาข้อสอบจนเป็นแบบทดสอบมาตรฐาน ดังนี้

2.3.1 แบบทดสอบที่ค่าความยากเหมาะสม คือ มีข้อสอบก่อนข้างง่าย ประมาณ 25% ข้อสอบยากปานกลาง 50% และข้อสอบก่อนข้างยาก ประมาณ 25% (เฉลี่ยทั้งฉบับมีค่าความยากประมาณ 50%)

2.3.2 แบบทดสอบมีค่าอำนาจจำแนกที่ดี ข้อสอบแต่ละข้อมีค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป (เฉลี่ยทั้งฉบับมีค่าอำนาจจำแนกมากกว่า .40)

2.3.3 แบบทดสอบมีค่าความเที่ยงที่ดี มีค่าก่อนข้างสูง ประมาณ .80 ขึ้นไป

2.4 ครูผู้สอนและผู้เชี่ยวชาญได้ตรวจสอบความตรงเชิงเนื้อหาของแบบทดสอบ ซึ่งมีค่าความตรงเชิงเนื้อหาทั้งหมด 100 % (ค่า IOC = 1.00)

การเก็บรวบรวมข้อมูล

ผู้วิจัยเก็บข้อมูลผลการตอบข้อสอบของนักเรียนจากกลุ่มประชากร ซึ่งเป็นข้อมูลทุติยภูมิจากนักสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 2 โดยมีขั้นตอนการเก็บรวบรวมข้อมูล ดังนี้

1. ขออนุญาตจากภาควิชาวิจัยและจิตวิทยาประยุกต์ คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา เพื่อขอข้อมูลผลการตอบข้อสอบของนักเรียนที่เป็นกลุ่มตัวอย่าง จากนักสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 2
2. ขอความร่วมมือจากเจ้าหน้าที่ในการคัดเลือกข้อสอบที่ผู้สอบที่ตอบข้อสอบครบทุกข้อ ได้จำนวน 5,011 ฉบับ นำมาวิเคราะห์ด้วยโปรแกรมสุ่มข้อมูลผลการตอบข้อสอบของนักเรียนแยกตามเพศ และระดับความสามารถ เพื่อให้ได้กลุ่มตัวอย่างตามจำนวนที่ต้องการ
3. นำข้อมูลที่ได้มาตรวจสอบคุณภาพโดยการวิเคราะห์ค่าอำนาจจำแนกและค่าความยากง่าย โดยใช้โปรแกรม BILOG MG 3.0 เลือกข้อสอบที่มีค่าอำนาจจำแนกที่มีค่ามากกว่า 0.5 ตามแนวคิดทฤษฎีตอบสนองข้อสอบวิชาละ 15 ข้อเพื่อนำมาใช้ในการวิจัย และการแจกแจงตามข้อตกลงเบื้องต้นของการวิเคราะห์องค์ประกอบและนำไปวิเคราะห์ต่อไป

การจัดกระทำข้อมูลสำหรับการวิเคราะห์

1. การคัดเลือกข้อสอบที่นำมาใช้ในวิจัย ผู้วิจัยได้นำข้อสอบที่เก็บรวบรวมได้มาวิเคราะห์คุณภาพข้อสอบอีกครั้งเพื่อคัดเลือกข้อสอบแบ่งออกเป็น 5 องค์ประกอบ โดยใช้เกณฑ์การตรวจสอบคุณภาพในการคัดเลือกข้อสอบที่สร้างขึ้นจากคำแนะนำของแฮมเบิลตันที่ว่าข้อสอบที่ดีควรมีค่าอำนาจจำแนกสูงคือมีค่ามากกว่า 0.5 โดยใช้โปรแกรม BILOGMG 3.0 มีรายละเอียดดังนี้

องค์ประกอบที่ 1 วัดความรู้ทางภาษาไทย จำนวน 15 ข้อ ประกอบด้วยข้อสอบข้อที่ TH1, TH8, TH9, TH10, TH16, TH21, TH26, TH27, TH29, TH32, TH33, TH35, TH36 และ TH39

องค์ประกอบที่ 2 วัดความสามารถทางคณิตศาสตร์ จำนวน 15 ข้อ ประกอบด้วยข้อสอบข้อที่ MA4, MA7, MA9, MA13, MA14, MA15, MA16, MA19, MA23, MA25, MA28, MA30, MA31, MA38 และ MA39

องค์ประกอบที่ 3 วัดความรู้ทางวิทยาศาสตร์ จำนวน 15 ข้อ ประกอบด้วยข้อสอบข้อที่ SC2, SC4, SC6, SC12, SC15, SC21, SC23, SC24, SC26, SC29, SC31, SC35, SC37, SC39 และ SC 40

องค์ประกอบที่ 4 วัดความรู้ทางสังคมศึกษา ศาสนาและวัฒนธรรม จำนวน 15 ข้อ ประกอบด้วยข้อสอบข้อที่ SO2, SO4, SO5, SO9, SO11, SO13, SO17, SO18, SO22, SO24, SO25, SO27, SO28, SO31, และ SO32

องค์ประกอบที่ 5 วัดความรู้ทางภาษาอังกฤษ จำนวน 15 ข้อ ประกอบด้วยข้อสอบข้อที่ EN2, EN8, EN10, EN12, EN15, EN16, EN19, EN20, EN20, EN21, EN25, EN28, EN29, EN30, EN31 และ EN32

2. ตรวจสอบความเที่ยงตรงเชิงโครงสร้างของข้อสอบทุกวิชาที่คัดเลือกโดยการวิเคราะห์ยืนยันองค์ประกอบโดยใช้โปรแกรม AMOS Version 16.0

การวิเคราะห์ข้อมูล

ตอนที่ 1 การหาประสิทธิภาพของวิเคราะห์องค์ประกอบเชิงยืนยันพหุ

1. การวิเคราะห์ประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยัน

ขั้นที่ 1 การวิเคราะห์ค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด ค่าต่ำสุด ค่าความเบ้และค่าความโด่ง เพื่อให้ทราบลักษณะการแจกแจงของตัวแปรโดยใช้โปรแกรมสำเร็จรูป SPSS Version 16.0

2. วิเคราะห์ประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง อันดับสองและโมเดลวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง อันดับสองแฝงภายใน เพื่อหาโมเดลวัดที่เหมาะสมกับข้อมูลเชิงประจักษ์ การวิเคราะห์ข้อมูลเพื่อหาประสิทธิภาพของโมเดลวิเคราะห์องค์ประกอบเชิงยืนยัน โดยใช้โปรแกรม AMOS 16.0 และตรวจสอบดัชนีวัดความสอดคล้อง (Fit Indexes) ในการตรวจสอบประสิทธิภาพในการวิเคราะห์องค์ประกอบเชิงยืนยันซึ่งเลือกใช้ดัชนีวัดความสอดคล้อง (Brown, 2006) โดยไม่มีการปรับโมเดล ได้แก่

2.1 ดัชนีวัดความสอดคล้องสมบูรณ์ (Absolute Fit) ซึ่งเป็นดัชนีในการประเมินค่าสุดท้ายของเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมของตัวแปรสังเกตได้กับเมทริกซ์ในการประมาณค่า ซึ่งประกอบไปด้วยดัชนีไค-สแควร์ (χ^2) โดยต้องการให้ผลลัพธ์ไม่มีนัยสำคัญทางสถิติ ($p > .05$) โดยเมทริกซ์ความแปรปรวนและความแปรปรวนร่วมของตัวแปรสังเกตได้กับเมทริกซ์ในการประมาณค่ามีค่าเท่ากันซึ่งดัชนีดังกล่าวมีความไวต่อขนาดกลุ่มตัวอย่าง ถ้ากลุ่มตัวอย่างขนาดใหญ่มีแนวโน้มจะมีนัยสำคัญทางสถิติ มีความไวต่อการแจกแจงปกติของกลุ่มตัวอย่าง และเป็นที่ยอมรับในการนำเสนอสำหรับโปรแกรมในการวิเคราะห์โมเดลสมการโครงสร้างในทุกโปรแกรม

2.2 ดัชนีวัดการปรับโมเดล (Parsimony Correction Fit) มีลักษณะคล้ายกับดัชนีวัดความสอดคล้องสมบูรณ์แต่ดัชนีวัดความสอดคล้องสมบูรณ์มีค่าดัชนีที่ต่ำกว่าดัชนีวัดปรับโมเดล (Kalinowski, 2006, p. 13) ค่าเฉลี่ยกำลังสองความคลาดเคลื่อนที่พอเหมาะ (Root Mean Square Error of Approximation: RMSEA) ซึ่งจะมีค่าจาก 0 - $+\infty$ แต่ซึ่งควรค่า $\leq .06$ (Hu & Bentler, 1999; Thompson, 2004)

2.3 ดัชนีวัดความกลมกลืนเปรียบเทียบ (Comparative Fit)

2.3.1 ดัชนีวัดความสอดคล้องเปรียบเทียบ (Comparative Fit Index: CFI) มีค่าอยู่ระหว่าง 0 - 1 ซึ่งควรค่า $\geq .95$ (Hu & Bentler, 1999; Thompson, 2004)

2.3.2 ดัชนี Tucker-Lewis Index (TLI) โดยปกติมีค่าอยู่ระหว่าง 0 - 1.0 ค่าดัชนีควร $\geq .95$ (Hu & Bentler, 1999)

2.3.3 ดัชนีวัดความสอดคล้องปกติ (Normed Fit Index: NFI) มีค่าอยู่ระหว่าง 0 - 1.0 และควรมีค่า $\geq .95$ (Thompson, 2004)

3. การศึกษาเปรียบเทียบผลของความแตกต่างของจำนวนของกลุ่มตัวอย่างต่อค่าดัชนีวัดความสอดคล้องในการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง การวิเคราะห์เชิงยืนยันอันดับสอง และการวิเคราะห์องค์ประกอบเชิงยืนยันแฝงภายใน และการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน โดยกำหนดขนาดกลุ่มตัวอย่างเท่ากับ 100, 200, 400, 800, 1,200, 1,600 และ

2,000 คน โดยเปรียบเทียบจากดัชนีไค-สแควร์ (χ^2) ค่าเฉลี่ยกำลังสองความคลาดเคลื่อนที่พอเหมาะ (Root Mean Square Error of Approximation: RMSEA) ดัชนีวัดความสอดคล้องเปรียบเทียบ (Comparative Fit Index: CFI) ดัชนี Tucker-Lewis Index (TLI) และดัชนีวัดความสอดคล้องปกติ (Normed Fit Index: NFI)

4. การศึกษาอัตราส่วนของตัวดัชนีการวัดต้องประกอบที่ส่งผลต่อค่าดัชนีวัดความสอดคล้องในการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่ง การวิเคราะห์เชิงยืนยันอันดับสอง การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับหนึ่งแฝงภายในและการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองแฝงภายใน โดยการกำหนดดัชนีการวัดต้องประกอบเท่ากับ 3, 5, 8, 10 และ 15 โดยการ เปรียบเทียบจากดัชนีไค-สแควร์ (χ^2) ค่าเฉลี่ยกำลังสองความคลาดเคลื่อนที่พอเหมาะ (Root Mean Square Error of Approximation: RMSEA) ดัชนีวัดความสอดคล้องเปรียบเทียบ (Comparative Fit Index: CFI) ดัชนี Tucker-Lewis Index (TLI) และดัชนีวัดความสอดคล้องปกติ (Normed Fit Index: NFI)

ตอนที่ 2 หาประสิทธิภาพของการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์องค์ประกอบแฝงภายใน

การวิเคราะห์องค์ประกอบ (Jensen, 1980, pp. 446-450) เป็นวิธีหนึ่งในการนำมาใช้ในการวิเคราะห์ความลำเอียงของข้อสอบ โดยถือว่าข้อสอบที่มีน้ำหนักองค์ประกอบ (Factor loading) ซึ่งมีขนาดและจำนวนไม่เท่ากันจากผู้สอบต่างกลุ่มกันเป็นข้อสอบที่มีความลำเอียงซึ่งหมายความว่าข้อสอบข้อนั้นไม่ได้วัดความสามารถเดียวกันในผู้สอบแต่ละกลุ่ม วิธีการวิเคราะห์นี้ เป็นวิธีที่เอื้อคะแนนการตอบของผู้ตอบแต่ละคน จากแต่ละกลุ่มย่อยมาทำการวิเคราะห์ข้อสอบทีละข้อ ในลักษณะของการยืนยันทฤษฎี (Confirmatory) และทำการสกัดองค์ประกอบด้วยวิธีการวิเคราะห์แกนสำคัญ (Principal Component Analysis: PCA) โดยการใช้การหมุนแกนแบบออร์โธโกนอล (Orthogonal) ตามวิธีความแปรปรวนสูงสุด (Varimax) เมื่อได้ค่าน้ำหนักแต่ละองค์ประกอบย่อยจากข้อสอบข้อเดียวกัน แล้วนำน้ำหนักองค์ประกอบแต่ละกลุ่มย่อยมาเปรียบเทียบกัน หากพบว่าน้ำหนักองค์ประกอบแต่ละกลุ่มย่อยไม่เท่ากัน และแตกต่างกันอย่างมีนัยสำคัญ แสดงว่าข้อสอบข้อนั้นมีความลำเอียง ผู้วิจัยดำเนินการวิเคราะห์ข้อมูลดังนี้

1. จัดเตรียมเพิ่มข้อมูลสำหรับการวิเคราะห์ โดยแยกผลการตอบข้อสอบตามกลุ่มผู้เข้าสอบ ได้แก่ กลุ่มเปรียบเทียบ และกลุ่มอ้างอิง
2. วิเคราะห์ค่าสถิติพื้นฐาน เพื่อบรรยายการแจกแจงของคะแนนของแบบทดสอบความสามารถทางวิชาภาษาไทย คณิตศาสตร์ วิทยาศาสตร์ สังคมศึกษาและภาษาอังกฤษ ชั้นมัธยมศึกษาปีที่ 6 โดยใช้โปรแกรม SPSS for Windows Version 16.0

3. วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์องค์ประกอบเชิงยืนยัน อันดับสองแฝงภายในซึ่งเป็น โมเดลที่ผู้วิจัยทำการทดสอบแล้วว่ามีความสอดคล้องกับข้อมูลเชิงประจักษ์มากที่สุด โดยใช้โปรแกรม AMOS Version 16.0

4. เปรียบเทียบผลการทดสอบการทำหน้าที่ต่างกันของข้อสอบ แบบทดสอบด้วยจำนวนกลุ่มตัวอย่างที่มาทดสอบต่างกัน

5. เปรียบเทียบระดับความเชื่อมั่นของแบบทดสอบก่อนตัดข้อสอบที่ทำหน้าที่ต่างกัน และหลังการจัดข้อสอบที่ทำหน้าที่ต่างการ โดยการทดสอบ G – Coefficient โดยใช้โปรแกรม GNOVA

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. การวิเคราะห์องค์ประกอบ โดยใช้โปรแกรม AMOS Version16.0 โดยทำการสกัดองค์ประกอบด้วยวิธีวิเคราะห์แกนสำคัญ (Principal Component Analysis: PCA) แล้วหมุนแกนแบบออร์ทोगอนอล (Orthogonal) ตามวิธีความแปรปรวนสูงสุด (Varimax) แล้วนำค่าน้ำหนักองค์ประกอบ (Factor loading) ของแต่ละกลุ่มข้อสอบแต่ละข้อ มาเปรียบเทียบกับน้ำหนักองค์ประกอบหาความแตกต่าง ด้วยสูตรไค-สแควร์ (Chi-square) โดยเปลี่ยนน้ำหนักองค์ประกอบเป็นคะแนนมาตรฐานโดยใช้ตาราง Fisher's Z ในแต่ละกลุ่มหากพบค่าไค-สแควร์ที่คำนวณได้มีนัยสำคัญ แสดงว่าข้อสอบข้อนั้นมีค่าเฉลี่ย ทดสอบความแตกต่างของน้ำหนักองค์ประกอบระหว่างกลุ่มต่าง ๆ ในข้อสอบแต่ละข้อ โดยใช้สูตรไค-สแควร์ (Chi-square) (Jensen,1980, p. 449)

$$\chi^2 = \frac{\sum_{i=1}^n (z_A - z_B)^2}{\frac{n}{N_A - 3} + \frac{n}{N_B - 3}} \quad (32)$$

เมื่อ	χ^2	แทน ค่าไค-สแควร์
	Z_A	แทน ค่าน้ำหนักองค์ประกอบของข้อสอบที่แปลงเป็นคะแนนมาตรฐาน โดยใช้ตาราง Fisher's Z ของกลุ่ม A
	Z_B	แทน ค่าน้ำหนักองค์ประกอบของข้อสอบที่แปลงเป็นคะแนนมาตรฐาน โดยใช้ตาราง Fisher's Z ของกลุ่ม B
	n	แทน จำนวนข้อสอบในแบบสอบ

N_A แทน จำนวนคนในกลุ่ม A

N_B แทน จำนวนคนในกลุ่ม B

2. เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยการวิเคราะห์

G – Coefficient โดยใช้โปรแกรม GNOVA

มหาวิทยาลัยบูรพา
Burapha University