# การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลระหว่างนักเรียนชายและนักเรียนหญิง: การวิเคราะห์จากผลการทดสอบระดับชาติ

**นางสาวนิตยา ดวงจันทร์ทิพย์**

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาศึกษาศาสตรมหาบัณฑิต
แขนงวิชาการวัดและประเมินผลการศึกษา สาขาวิชาศึกษาศาสตร์ มหาวิทยาลัยสุโขทัยธรรมาธิราช
พ.ศ. 2562

# Differential Item Functioning in Reasoning Abilities between Males and females: An analysis of the National Test Result

**Miss Nitaya Dungchantip**

A Thesis Submitted in Partial Fulfillment of the Requirements for

the Degree of Master of Education in Educational Evaluation

School of Educational Studies

Sukhothai Thammathirat Open University

2019

# สารบัญ

หน้า


บทคัดย่อภาษาไทย ..... ง
บทคัดย่อภาษาอังกฤษ ..... จ
กิตติกรรมประกาศ ..... ฉ
สารบัญตาราง ..... ฌ
สารบัญภาพ ..... ญ
บทที่ 1 บทนำ ..... 1
ความเป็นมาและความสำคัญของปัญหา ..... 1
วัตถุประสงค์การวิจัย ..... 3
ขอบเขตของการวิจัย ..... 3
นิยามคำศัพท์เฉพาะ ..... 4
ประโยชน์ที่คาดว่าจะได้รับ ..... 5
บทที่ 2 วรรณกรรมที่เกี่ยวข้อง ..... 6
การทดสอบความสามารถพื้นฐานของผู้เรียนระดับชาติ ..... 6
การทำหน้าที่ต่างกันของข้อสอบ ..... 9
งานวิจัยที่เกี่ยวข้อง ..... 30
บทที่ 3 วิธีดำเนินการวิจัย ..... 38
ประชากร ..... 38
เครื่องมือที่ใช้ในการวิจัย ..... 38
การเก็บรวบรวมข้อมูล ..... 42
สถิติที่ใช้ในการวิเคราะห์ข้อมูล ..... 43
บทที่ 4 ผลการวิเคราะห์ข้อมูล ..... 51
ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติพื้นฐาน ..... 52
ตอนที่ 2 ค่าความยากง่ายและค่าอำนาจจำแนกของข้อสอบด้านเหตุผล ของแบบทดสอบระดับชาติ ..... 53
ตอนที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านเหตุผลของ แบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง ..... 57

# สารบัญ (ต่อ)

หน้า

ตอนที่ 4 การแสดงกราฟของการทำหน้าที่ต่างกันของข้อสอบด้านเหตุผล
ของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง ..... 65
บทที่ 5 สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ ..... 70
สรุปการวิจัย ..... 71
อภิปรายผล ..... 72
ข้อเสนอแนะ ..... 73
บรรณานุกรม ..... 72
ภาคผนวก ..... 80
ก ค่าสถิติพื้นฐาน ค่าความยาก ค่าอำนาจจำแนกของข้อสอบด้านเหตุผล
ของแบบทดสอบระดับชาติ ..... 81
ข หนังสือเชิญผู้เชี่ยวชาญ ..... 86
ค ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ..... 90
ง แบบทดสอบวัดความสามารถด้านเหตุผล ปีการศึกษา 2557 ..... 97
ประวัติผู้วิจัย ..... 111

# สารบัญตาราง

หน้า

ตารางที่ 2.1 กรอบโครงสร้างการประเมินความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ..... 8
ตารางที่ 2.2 ค่า $R^2$ ของการวัดขนาดการทำหน้าที่ต่างกันของข้อสอบ ..... 26
ตารางที่ 4.1 สถิติพื้นฐานของคะแนนจากการตอบข้อสอบของนักเรียนจำแนกตามเพศผู้สอบ ..... 53
ตารางที่ 4.2 ค่าความยากง่ายและค่าอำนาจจำแนกของข้อสอบด้านเหตุผลของแบบทดสอบระดับชาติ ..... 53
ตารางที่ 4.3 การแสดงความสัมพันธ์ระหว่างค่าความยากง่ายกับค่าอำนาจจำแนกของข้อสอบด้านเหตุผลของแบบทดสอบระดับชาติ ..... 55
ตารางที่ 4.4 ค่าความยากง่ายและค่าอำนาจจำแนกของข้อสอบด้านเหตุผลของแบบทดสอบระดับชาติแยกเป็นนักเรียนชายและนักเรียนหญิง ..... 56
ตารางที่ 4.5 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง ..... 58

# สารบัญภาพ

หน้า

ภาพที่ 2.1 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป ..... 12
ภาพที่ 2.2 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป ..... 13
ภาพที่ 2.3 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง ..... 14
ภาพที่ 2.4 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว ..... 15
ภาพที่ 3.1 ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป ..... 50
ภาพที่ 3.2 ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ..... 50

# บทที่ 1
# บทนำ

## 1. ความเป็นมาและความสำคัญของปัญหา

การทดสอบวัดความสามารถพื้นฐานของผู้เรียนระดับชาติ เป็นการประเมินผู้เรียนเพื่อสะท้อนคุณภาพในการจัดการศึกษาของสถานศึกษา และสำนักงานเขตพื้นที่การศึกษา อีกทั้งยังเป็นตัวบ่งชี้คุณภาพการศึกษาขั้นพื้นฐานในภาพรวม ใช้ประกอบการตัดสินใจในการกำหนดนโยบายในระดับต่าง ๆ ในระดับสถานศึกษาเป็นการประเมิน เพื่อปรับปรุงและพัฒนาการเรียนการสอนของครู สร้างความมั่นใจกับสถานศึกษาและหน่วยงานต้นสังกัดจัดการศึกษาได้อย่างมีประสิทธิภาพตามมาตรฐานที่กำหนด การวัดและประเมินผลการเรียนรู้ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ก็เป็นหลักประกันการเรียนรู้ (accountability) อย่างหนึ่ง ในการประเมินคุณภาพของนักเรียนว่ามีความรู้หรือคุณลักษณะที่ต้องการวัดอยู่ในระดับใด ซึ่งผลที่ได้จากการวัดมีความสำคัญต่อการพัฒนาคุณภาพของการศึกษา การทดสอบเป็นวิธีการวัดผลการศึกษาวิธีหนึ่งที่นิยมใช้มาก

การทดสอบเพื่อวัดความสามารถของผู้เรียนนั้นต้องมีการกำหนดจุดมุ่งหมายของการวัดที่ชัดเจนว่าจุดประสงค์หลักที่ต้องการวัด เครื่องมือวัดที่มีคุณภาพ ปัจจุบันเครื่องมือวัดทางการศึกษาและจิตวิทยา ส่วนใหญ่ใช้วัดพัฒนาการของผู้เรียน เช่น แบบสอบ แบบสัมภาษณ์ แบบสังเกตและแบบสอบถาม แบบประเมินการเขียน เป็นต้น ผลของการวัดถูกกำหนดค่าให้เป็นคะแนนซึ่งถือว่าเป็นเพียงตัวอย่างของพฤติกรรมที่แสดงออกมา แต่ไม่ใช่พฤติกรรมทั้งหมดของบุคคล เมื่อแปลผลของคะแนนแล้วต้องมีความหมายที่สื่อถึงคุณลักษณะภายในที่ต้องการจะวัดได้ถูกต้องตามความเป็นจริง คะแนนที่ได้เป็นคะแนนดิบไม่มีความสมบูรณ์มีคุณลักษณะเชิงสัมพัทธ์ต้องเปรียบเทียบความหมายกับเกณฑ์มาตรฐานอื่นแล้วจึงแปลความหมายต่อไปได้ (ศิริชัย กาญจนวาสี, 2550, น. 10) ดังนั้นในการสร้างเครื่องมือต้องคำนึงถึงคุณภาพเครื่องมือ

การวิเคราะห์ข้อสอบเป็นเทคนิคของการตรวจสอบคุณภาพของข้อสอบเป็นรายข้อ ตัวบ่งชี้คุณภาพของข้อสอบที่สำคัญได้แก่ ค่าความยาก ค่าอำนาจจำแนกของข้อสอบ ส่วนคุณภาพของข้อสอบทั้งฉบับ ประกอบด้วยหลักฐานแสดงความเที่ยง และความตรง ที่ได้มาตรฐาน โดยเฉพาะประเด็นด้านความตรง เป็นหัวใจสำคัญของคุณภาพของแบบทดสอบ ทั้งนี้เพราะความ

ตรงเป็นคุณสมบัติของแบบทดสอบที่แสดงถึงความสามารถในการวัดได้ถูกต้องแม่นยำ และการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ (item and test unfairness) ก็เป็นอีกลักษณะหนึ่งของการตรวจสอบคุณภาพด้านความตรง ในประเด็นของความยุติธรรม (ศิริชัย กาญจนวาสี, 2555, น. 115) การตรวจสอบในประเด็นของความยุติธรรมของข้อสอบและแบบสอบ เป็นการขจัดข้อสอบที่ทำให้เกิดปัญหาความยุติธรรมระหว่างข้อสอบกับกลุ่มผู้สอบต่างๆ ที่มีลักษณะบางอย่างแตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบมักเกิดจากการปฏิสัมพันธ์ (interaction) ระหว่างลักษณะของข้อสอบกับลักษณะของกลุ่มผู้สอบ ที่มีผลต่อคะแนน และลักษณะข้อสอบเช่น การใช้ภาษา ความเฉพาะของเนื้อหา โครงสร้างของการวัด วิธีการที่เลือกใช้ เป็นลักษณะของกลุ่มผู้สอบ เช่น เชื้อชาติ ศาสนา วัฒนธรรม ภูมิลำเนา สังคม เพศ ภาษา อายุ ประสบการณ์ เป็นต้น เมื่อทั้งสองปัจจัยมีปฏิสัมพันธ์ต่อกันจะทำให้ข้อสอบข้อนั้นเกิดการทำหน้าที่แตกต่างกัน (differential item functioning; DIF) (ศิริชัย กาญจนวาสี, 2555, น. 116) จึงมีผลต่อคะแนนของข้อสอบข้อนั้น อันอาจนำไปสู่ความลำเอียงของข้อสอบด้านต่างๆ เช่น ความลำเอียงด้านเพศ ความลำเอียงด้านเชื้อชาติ เป็นต้น เพื่อพัฒนาแบบสอบให้มีคุณภาพเหมาะสมสำหรับนำไปใช้ทดสอบต่อไป วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่นิยมใช้ตั้งแต่เริ่มศึกษาการทำหน้าที่ต่างกันของข้อสอบ จนถึงปัจจุบันก็ยังได้รับความนิยมอยู่ คือ วิธีแมนเทล-แฮนส์เซล (Mantel – Haenszel ; MH) วิธีซิปเทสท์ (SIBTEST) และวิธีถดถอยโลจิสติก (logistic regression ; LR) ทุกวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ แต่จะมีข้อดี ข้อเสีย ข้อจำกัด ขนาดของกลุ่มตัวอย่าง ความยาวของข้อสอบ และลักษณะของเครื่องมือ ของแต่ละวิธีแตกต่างกันออกไป ดังนั้นผู้วิจัยควรศึกษาให้เข้าใจและเลือกวิธีที่เหมาะสม เพื่อความถูกต้องและประสิทธิภาพของการศึกษา ในการวิจัยครั้งนี้ผู้วิจัยเลือกวิธีถดถอยโลจิสติก เพราะเหมาะกับการสอบที่มีผู้สอบจำนวนมากๆ เป็นวิธีที่สอดคล้องกับธรรมชาติของการทำหน้าที่ต่างกันของข้อสอบได้ดี สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและอเนกรูปได้ดี

การวัดความสามารถพื้นฐานของนักเรียนชั้นประถมศึกษาปีที่ 3 ได้ดำเนินการจัดทำแบบทดสอบวัดความสามารถพื้นฐานที่สำคัญ 3 ด้าน คือ ด้านภาษา (literacy) ด้านคำนวณ (numeracy) และด้านเหตุผล (reasoning abilities) ซึ่งถือเป็นความสามารถพื้นฐานเบื้องต้นสำคัญที่ใช้ในการเรียนรู้ในระดับที่สูงขึ้น สำหรับการประเมินความสามารถด้านเหตุผลของผู้เรียนนั้น มีผลการประเมินหลายแหล่งและกระแสวิพากษ์วิจารณ์ว่า เด็กไทยควรได้รับการพัฒนาทักษะด้านการคิดวิเคราะห์ สังเคราะห์และการให้เหตุผลดังนั้นจึงมีการประเมินความสามารถด้านเหตุผลของผู้เรียนด้วย โดยใช้บริบทหรือสถานการณ์เกี่ยวกับวิทยาศาสตร์และสิ่งแวดล้อม สังคมศาสตร์และเศรษฐศาสตร์รวมถึงการดำรงชีวิต (สำนักทดสอบทางการศึกษา, 2556, น. 3)

สำหรับการทดสอบวัดความสามารถพื้นฐานของผู้เรียนระดับชาติเพื่อการประเมินคุณภาพการศึกษาในระดับประเทศ เป็นการประเมินเพื่อกำกับ ติดตาม และควบคุม คุณภาพการจัดการศึกษาขั้นพื้นฐานของประเทศ ซึ่งผลการประเมินยังสะท้อนสู่คุณภาพการดำเนินการจัดการศึกษาของสถานศึกษา และเขตพื้นที่การศึกษา ในการหาคุณภาพของข้อสอบ ด้านความตรงความเที่ยงและค่าอำนาจจำแนกของข้อสอบนั้นมีผู้วิจัยหลายท่านได้ศึกษาไปมากแล้ว แต่ในด้านของความยุติธรรมของข้อสอบ ยังมีการศึกษาน้อย ผู้วิจัยจึงสนใจศึกษาการทำหน้าที่ต่างกันของข้อสอบ ในความแตกต่างกันทางด้านเพศ ซึ่งอาจจะส่งผลให้มีการได้เปรียบเสียเปรียบกันเกิดขึ้นในการทำข้อสอบหรือแบบทดสอบ สำหรับแบบทดสอบที่ใช้มีลักษณะเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก ทดสอบกับนักเรียนชั้นประถมศึกษาปีที่ 3 โดยสำนักทดสอบทางการศึกษาของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ด้วยเหตุนี้ทำให้ผู้วิจัยสนใจที่จะศึกษาการทำหน้าที่ต่างกันของข้อสอบด้านเหตุผลระหว่างนักเรียนชายกับนักเรียนหญิง โดยใช้การวิเคราะห์จากผลการทดสอบระดับชาติ ระดับชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 ด้วยวิธีของโลจิสติก โดยจำแนกตามตัวแปรเพศ เพื่อเป็นแนวทางในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่อไป

## 2. วัตถุประสงค์การวิจัย

2.1 เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

2.2 เพื่อวิเคราะห์สาเหตุการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของการทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

## 3. ขอบเขตของการวิจัย

3.1 ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลทุติยภูมิจากฐานข้อมูลของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน และกลุ่มตัวอย่างคือนักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 ข้อมูลที่ใช้ในการศึกษาเป็นคะแนนที่ได้จากการทดสอบวัดความสามารถด้านเหตุผลของการทดสอบระดับชาติ ที่สร้างโดยสำนักทดสอบทางการศึกษาของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานโดยผ่านกระบวนการตรวจสอบของแบบสอบทั้งด้านความตรงและความเที่ยงจากผู้เชี่ยวชาญเป็นที่เรียบร้อยแล้ว

3.2 การวิจัยครั้งนี้มุ่งตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้วิธีการวิเคราะห์ถดถอยโลจิสติก (logistic regression) ทั้งนี้การตัดข้อสอบที่ทำหน้าที่ต่างกันออกนั้นโดยพิจารณาข้อสอบที่มีขนาดการทำหน้าที่ต่างกันของข้อสอบ ที่มีขนาดกลางและขนาดใหญ่ ถือว่าข้อสอบทำหน้าที่ต่างกัน

## 4. นิยามศัพท์เฉพาะ

**4.1 การทำหน้าที่ต่างกันของข้อสอบ** หมายถึง ข้อสอบที่ทำให้ผู้สอบต่างกลุ่มกันและมีความสามารถเท่ากันหรือคุณลักษณะที่ต้องการวัดเท่ากัน แต่มีโอกาสในการตอบข้อสอบได้ถูกต้องต่างกัน ในการวิจัยครั้งนี้ คือ กลุ่มผู้สอบนักเรียนชายและนักเรียนหญิง การทำหน้าที่ต่างกันของข้อสอบคำนวณจาก วิธีการวิเคราะห์ถดถอยโลจิสติก หมายถึง วิธีการพิจารณาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปหรืออเนกรูป

**4.2 ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป** หมายถึง ข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มหนึ่งอย่างสม่ำเสมอกันในทุกระดับความสามารถ เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างกลุ่มในทุกระดับความสามารถ

**4.3 ข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป** หมายถึง ข้อสอบที่ทำให้โอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มไม่เสมอกันในทุกระดับความสามารถเมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่ามีปฏิสัมพันธ์ระหว่างกลุ่มในบางระดับความสามารถ

**4.4 แบบวัดผลสัมฤทธิ์ในด้านความสามารถด้านเหตุผล** หมายถึง แบบทดสอบความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นฐาน ปีการศึกษา 2557

## 5. ประโยชน์ที่คาดว่าจะได้รับ

5.1 เพื่อเป็นแนวทางในการปรับปรุงข้อสอบหรือแบบสอบให้มีความยุติธรรมต่อกลุ่มผู้สอบ ซึ่งเป็นทางเลือกหนึ่งในการพัฒนาแบบทดสอบให้มีคุณภาพ

5.2 เป็นการขยายขอบเขตในการศึกษาค้นคว้าเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและเป็นการขยายองค์ความรู้ด้านวัดผลการศึกษา

# บทที่ 2
# วรรณกรรมที่เกี่ยวข้อง

การศึกษาการทำหน้าที่ต่างกันของข้อสอบด้านเหตุผลระหว่างนักเรียนชายและนักเรียนหญิง : การวิเคราะห์จากผลการทดสอบระดับชาติ ผู้วิจัยได้ทำการศึกษาค้นคว้าเอกสารและงานวิจัยที่เกี่ยวข้องโดยนำเสนอผลการศึกษาตามลำดับดังนี้

1. การประเมินคุณภาพการศึกษาขั้นพื้นฐาน
   1.1 การทดสอบวัดความสามารถของผู้เรียนด้านเหตุผล
   1.2 ความสามารถและตัวชี้วัดด้านเหตุผล
2. การทำหน้าที่ต่างกันของข้อสอบ
   2.1 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ
   2.2 ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ
   2.3 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ
   2.4 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
   2.5 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
   2.6 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก
   2.7 สาเหตุการทำหน้าที่ต่างกันของข้อสอบ
3. งานวิจัยที่เกี่ยวข้องกับการหน้าที่ต่างกันของข้อสอบ

## 1. การประเมินคุณภาพการศึกษาขั้นพื้นฐาน

การประเมินคุณภาพการศึกษาขั้นพื้นฐาน เป็นการประเมินภาพรวมของการจัดการศึกษาของประเทศและสะท้อนคุณภาพนักเรียน เพื่อพัฒนาและปรับปรุงการเรียนการสอนของครู และยังเป็นกระบวนการตรวจสอบการเพิ่มประสิทธิภาพของสถานศึกษา ให้ได้มาตรฐานตามที่กำหนด นักเรียนมีความรู้ความสามารถ และคุณลักษณะอันพึงประสงค์ตามที่หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กำหนด โดยสำนักทดสอบทางการศึกษาได้วัดความสามารถพื้นฐานของนักเรียนชั้นประถมศึกษาปีที่ 3 ที่สำคัญ 3 ด้าน คือ ด้านภาษา (literacy) ด้านคำนวณ (numeracy) และด้านเหตุผล (reasoning abilities) ในที่นี้จะขอกล่าวเฉพาะด้านเหตุผล ดังต่อไปนี้

**1.1 การทดสอบวัดความสามารถของผู้เรียนด้านเหตุผล**

ความสามารถด้านเหตุผล หมายถึง ความสามารถในการเชื่อมโยงความรู้ และประสบการณ์ด้านวิทยาศาสตร์และสิ่งแวดล้อม ด้านสังคมศาสตร์และเศรษฐศาสตร์ และด้านการดำเนินชีวิต โดยการวิเคราะห์ สังเคราะห์ ประเมินค่า แก้ปัญหา หรือตัดสินใจอย่างมีหลักการและเหตุผล บนพื้นฐานของข้อมูล สถานการณ์ หรือสารสนเทศที่เพียงพอ โดยยึดหลักคุณธรรมและจริยธรรม โดยมีคำสำคัญดังนี้

***1.1.1 ความรู้*** หมายถึง ข้อเท็จจริง ทฤษฎี หลักการ กระบวนการที่ศึกษา รวมทั้งคุณธรรมและจริยธรรม

***1.1.2 ประสบการณ์*** หมายถึง ความรู้เดิมที่เกิดจากการเรียนรู้ ปฏิบัติหรือได้พบเห็น เรื่องต่างๆ ในระดับบุคคล สังคม และสังคมโลก

***1.1.3 วิเคราะห์*** หมายถึง ความสามารถในการเปรียบเทียบ บอกความต่าง ความเหมือน สรุปหลักการ บอกความสัมพันธ์เชื่อมโยงอย่างมีเหตุผล บนพื้นฐานของหลักการทางวิทยาศาสตร์ สังคมศาสตร์ และการดำเนินชีวิตอย่างมีคุณธรรมและจริยธรรม

***1.1.4 สังเคราะห์*** หมายถึง ความสามารถในการสร้างข้อสอบสรุปใหม่ ออกแบบคิดสร้างสรรค์ บนพื้นฐานของข้อมูลที่ผ่านการวิเคราะห์ ประเมินค่าแล้วอย่างสมเหตุสมผล

***1.1.5 ประเมินค่า*** หมายถึง ความสามารถในการตัดสินใจเลือกทางเลือกอย่างสมเหตุสมผล มีประโยชน์ และสร้างสรรค์

***1.1.6 เหตุผลทางวิทยาศาสตร์*** หมายถึง การนำความรู้ ประสบการณ์ ที่เกิดจากการเรียนรู้มาประกอบการตัดสินใจในสถานการณ์ที่เกิดขึ้นในสังคมให้สมเหตุสมผลตามหลักเกณฑ์ทางวิทยาศาสตร์

***1.1.7 เหตุผลทางสังคมศาสตร์*** หมายถึง การนำความรู้ ประสบการณ์จากกฎเกณฑ์ ความเชื่อ วัฒนธรรม ค่านิยมทางสังคมศาสตร์ มาประกอบการตัดสอนใจในสถานการณ์ที่เกิดขึ้นในสังคมได้อย่างสมเหตุสมผล

***1.1.8 เหตุผลทางการดำเนินชีวิต*** หมายถึง การนำความรู้ หลักการ กฎเกณฑ์มาใช้ในการดำเนินชีวิตหรือประกอบการตัดสินใจในสถานการณ์ที่เกิดขึ้นในสังคมอย่างมีคุณธรรมและจริยธรรม

**1.2 ความสามารถและตัวชี้วัดด้านเหตุผล**

ความสามารถในการเชื่อมโยงความรู้ประสบการณ์ ด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิต โดยการวิเคราะห์ สังเคราะห์ ประเมิน

ค่า แก้ปัญหาหรือตัดสินใจอย่างมีหลักฐานและเหตุผล บนพื้นฐานของข้อมูล สถานการณ์ หรือสารสนเทศที่เพียงพอโดยยึดหลักคุณธรรมและจริยธรรม

***1.2.1 ตัวชี้วัด***

*1) มีความเข้าใจในข้อมูล* สถานการณ์หรือสารสนเทศทางด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิตอย่างมีเหตุผล

*2) วิเคราะห์ข้อมูล* สถานการณ์ หรือสารสนเทศ โดยใช้องค์ความรู้ ด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิตอย่างมีเหตุผล

*3) สามารถสร้างข้อสอบสรุปใหม่ออกแบบ* วางแผนบนพื้นฐานของข้อมูลสถานการณ์หรือสารสนเทศที่ผ่านการวิเคราะห์ โดยใช้องค์ความรู้ ด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์และด้านการดำเนินชีวิตอย่างมีเหตุผล

*4) สามารถตัดสินใจและแก้ปัญหาอย่างมีหลักการและเหตุผลหรือให้ข้อสนับสนุนข้อโต้แย้งที่สมเหตุสมผล* โดยคำนึงถึง คุณธรรมและจริยธรรม ค่านิยม ความเชื่อ ในกรณีที่มีสถานการณ์ที่ต้องการตัดสินใจหรือมีปัญหา

***1.2.2 กรอบโครงสร้างการประเมินความสามารถด้านเหตุผลชั้นประถมศึกษาปีที่ 3*** ดังตารางที่ 2.1

ตารางที่ 2.1 กรอบโครงสร้างในการประเมินความสามารถด้านเหตุผลชั้นประถมศึกษาปีที่ 3

| ตัวชี้วัด | เลือกตอบ | เขียนตอบ |
|---|---|---|
| 1. มีความเข้าใจในข้อมูล สถานการณ์ หรือสารสนเทศทางด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิตอย่างมีเหตุผล | 3 | 1 |
| 2. วิเคราะห์ข้อมูล สถานการณ์ หรือสารสนเทศ โดยใช้องค์ความรู้ด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิตอย่างมีเหตุผล | 7 | 1 |
| 3. สามารถสร้างข้อสรุปใหม่ ออกแบบ วางแผน บนพื้นฐานของข้อมูล สถานการณ์ หรือสารสนเทศที่ผ่านการวิเคราะห์ โดยใช้องค์ความรู้ด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิตอย่างมีเหตุผล | 8 | |

ตารางที่ 2.1 (ต่อ)

| ตัวชี้วัด | เลือกตอบ | เขียนตอบ |
|---|---|---|
| 4. สามารถตัดสินใจและแก้ปัญหาอย่างมีหลักการและเหตุผล หรือให้ข้อสนับสนุนข้อโต้แย้งที่สมเหตุสมผล โดยคำนึงถึงคุณธรรมและจริยธรรม ค่านิยม ความเชื่อ ในกรณีที่มีสถานการณ์ที่ต้องการตัดสินใจหรือมีปัญหา | 9 | 1 |

## 2. การทำหน้าที่ต่างกันของข้อสอบ

### 2.1 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ เดิมเรียกว่าความลำเอียงของข้อสอบ (item bias) ในระยะหลังเปลี่ยนมาใช้คำว่าการทำหน้าที่ต่างกันของข้อสอบ เพราะเป็นคำที่เป็นกลางและมีความเหมาะสมมากกว่า (Holland and Thayer, 1988 ; Holland and Wainer, 1993. อ้างถึงใน ศิริชัย กาญจนวาสี, 2555, น. 116) มีนักวิจัยทางการวัดผลหลายท่านได้ให้ความหมายของการลำเอียงของข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ ดังนี้

Scheuneman (1979) กล่าวว่า ความลำเอียงของข้อสอบ หมายถึง สัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องไม่เท่ากันในแต่ละกลุ่มประชากรที่ใช้ในการศึกษา เมื่อกลุ่มผู้สอบมีคะแนนเท่ากันและข้อสอบมีความเป็นเอกพันธ์

Holland and Wainer (1993) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง สารสนเทศทางสถิติของข้อสอบที่ได้จากผลการตอบของผู้สอบต่างกลุ่มกัน และมีความสามารถเท่ากันแต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน

Mazor and Clauser (1992) กล่าวว่า ข้อสอบทำหน้าที่ต่างกันเมื่อผู้สอบที่มีความสามารถระดับเดียวกัน แต่เป็นสมาชิกของกลุ่มย่อยต่างกัน มีโอกาสในการตอบข้อสอบถูกต้องแตกต่างกัน

Shealy and Stout (1993) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่เข้าข้าง (favor) ผู้สอบกลุ่มหนึ่งมากกว่าผู้สอบอีกกลุ่มหนึ่งที่นำมาจับคู่เปรียบเทียบกัน ซึ่งทำให้ผู้สอบกลุ่มหนึ่งได้ประโยชน์แต่ผู้สอบอีกกลุ่มหนึ่งเสียประโยชน์

Narayanan and Swaminathan (1996) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ฟังก์ชันการตอบสนองข้อสอบซึ่งคำนวณจากกลุ่มผู้สอบกลุ่มย่อยที่แตกต่างกัน มีค่าไม่เท่ากัน

ศิริชัย กาญจนวาสี (2555) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง การที่ข้อสอบทำให้ผู้สอบจากต่างกลุ่มกันที่มีความสามารถหรือคุณลักษณะที่มุ่งวัดเท่ากัน มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน หรือมีฟังก์ชันการตอบสนองข้อสอบแตกต่างกัน การทำหน้าที่ต่างกันของข้อสอบเกิดขึ้นเมื่อนำข้อสอบไปทดสอบกับผู้สอบกลุ่มย่อยต่างกัน ที่มีความสามารถหลัก (primary ability) ระดับเดี่ยวกันหรือมีลักษณะแฝง (latent trait) ที่ต้องการวัดเท่ากันแต่มีความสามารถรอง (secondary ability) แตกต่างกัน ทำให้ผู้สอบต่างกลุ่มที่นำมาเปรียบเทียบมีโอกาสตอบข้อสอบถูกต่างกัน

จากนิยามการทำหน้าที่ต่างกันของข้อสอบหรือแบบทดสอบดังกล่าวสามารถสรุปรวมได้ว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่ทำให้ผู้สอบต่างกลุ่มกันและมีความสามารถเท่ากันหรือคุณลักษณะที่ต้องการวัดเท่ากัน แต่มีโอกาสในการตอบข้อสอบได้ถูกต้องต่างกัน

**2.2 ความเป็นมาของการทำหน้าที่ต่างกันของข้อสอบ**

การศึกษาถึงคุณภาพของข้อสอบจากผลการตรวจข้อสอบของผู้สอบกลุ่มต่างๆ ในประชากรมีมานานแล้ว แต่การศึกษาคุณภาพด้านความยุติธรรมของข้อสอบหรือแบบสอบระหว่างผู้สอบกลุ่มต่างๆ เริ่มศึกษากันอย่างจริงจังมากขึ้นในช่วงปลายทศวรรษของปี ค.ศ. 1960 มีการเสนอวิธีการต่างๆ เพื่อตรวจสอบความลำเอียงของแบบสอบ (test bias) ความพยายามของการตรวจสอบความลำเอียงดังกล่าวในยุกต์ต้นๆ เพื่อจำแนกข้อสอบที่ทำหน้าที่ไม่เหมาะสมหรือไม่ยุติธรรมสำหรับปรับปรุง หรือตัดข้อสอบข้อนั้นออกจากแบบสอบ เป็นการขจัดข้อสอบที่ทำให้เกิดปัญหาความยุติธรรมระหว่างกลุ่มข้อสอบกลุ่มต่าง ๆ ซึ่งมีลักษณะบางอย่างแตกต่างกัน เช่น เชื้อชาติ วัฒนธรรม ภูมิลำเนา สังคม เพศ อายุ ประสบการณ์ เป็นต้น เพื่อพัฒนาแบบสอบให้มีคุณภาพเหมาะสมสำหรับนำไปใช้ทดสอบต่อไป ต่อมาได้มีนักวัดผลการศึกษาได้ทำการศึกษากันอย่างกว้างขวางและเกิดความสับสนของการใช้คำและความหมาย ส่งผลต่อการบรรลุจุดมุ่งหมายของการใช้แบบสอบ และเกณฑ์ที่ใช้สำหรับตัดสินความลำเอียงยังมีความคลุมเครือและค่อนข้างสับสนจึงไม่เหมาะสมที่จะใช้คำว่าข้อสอบลำเอียง ดังนั้นจึงควรเปลี่ยนมาใช้คำว่าการทำหน้าที่ต่างกันของข้อสอบการทำหน้าที่ต่างกันของข้อสอบกับความลำเอียงของข้อสอบ มีแนวความคิดที่แตกต่างกัน สำหรับการทำหน้าที่ต่างกันของข้อสอบ เป็นกระบวนการที่เน้นการใช้วิธีการทางสถิติสำหรับตรวจสอบ เพื่อให้ได้สารสนเทศเกี่ยวกับการทำหน้าที่ของข้อสอบสำหรับกลุ่มผู้สอบกลุ่ม

ย่อยที่มีลักษณะเฉพาะบางอย่างแตกต่างกัน ส่วนความลำเอียงของข้อสอบ เป็นกระบวนการตัดสินความยุติธรรมของข้อสอบ โดยนำสารสนเทศการทำหน้าที่ต่างกันของข้อสอบมาวิเคราะห์เชิงตรรกะ (logical analysis) โดยผู้เชี่ยวชาญพิจารณาถึงการเขียนข้อสอบ เนื้อหาสาระของข้อสอบ และจุดมุ่งหมายของการวัด เพื่อระบุว่าข้อสอบข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มใดหรือไม่ เพราะเหตุใดจึงเป็นการตัดสินความลำเอียงของข้อสอบ (Camilli and Shapard, 1994 อ้างถึงในศิริชัย กาญจนวาสี, 2555, น. 116)

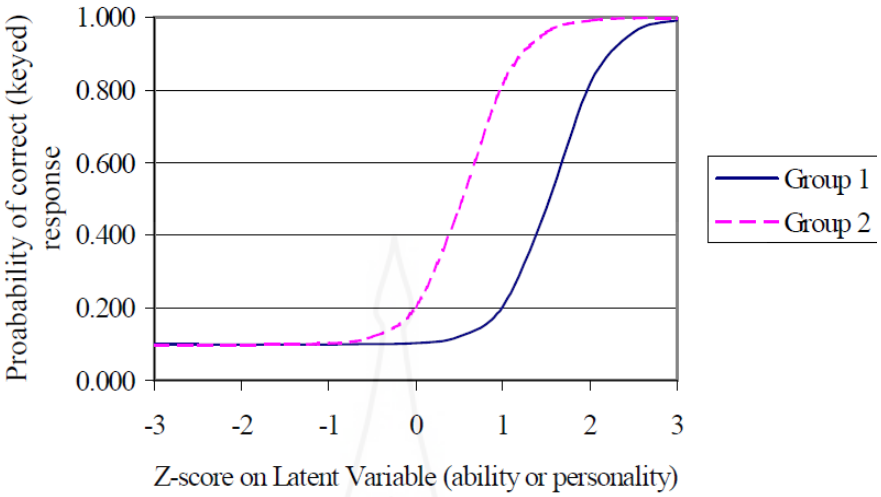
จากการศึกษาเอกสารและงานวิจัยข้างต้นผู้วิจัยสามารถสรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ แต่เดิมใช้คำว่า ความลำเอียงของข้อสอบ เป็นการขจัดข้อสอบที่ทำให้เกิดปัญหาความยุติธรรมระหว่างกลุ่มข้อสอบกลุ่มต่างๆ และเป็นภาษาที่มีความหมายในเชิงลบ และเกณฑ์ที่ใช้ในการตัดสินความลำเอียงยังมีความคลุมเครือ ในระยะหลังนักวิจัยได้เปลี่ยนมาใช้คำว่าการทำหน้าที่ต่างกันของข้อสอบ

**2.3 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ**

การทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่มขึ้นไป ในการเปรียบเทียบนิยมทำระหว่างกลุ่มที่มีความสามารถในระดับเดียวกัน และได้กำหนดผู้สอบ 2 กลุ่มคือ กลุ่มอ้างอิง (reference group หรือกลุ่ม R) เป็นกลุ่มที่คาดว่าจะได้เปรียบในการตอบข้อสอบได้ถูกต้องหรือได้ประโยชน์ และกลุ่มเปรียบเทียบ (focal group หรือ กลุ่ม F) เป็นกลุ่มที่สนใจศึกษาและคาดว่าจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบหรือเสียประโยชน์ และในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ Mellenbergh (1982 อ้างถึงในศิริชัย กาญจนวาสี, 2555, น. 117) ได้ระบุลักษณะของข้อสอบที่ทำหน้าที่ต่างกันใน 2 ประเภทดังนี้

***2.3.1 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป***

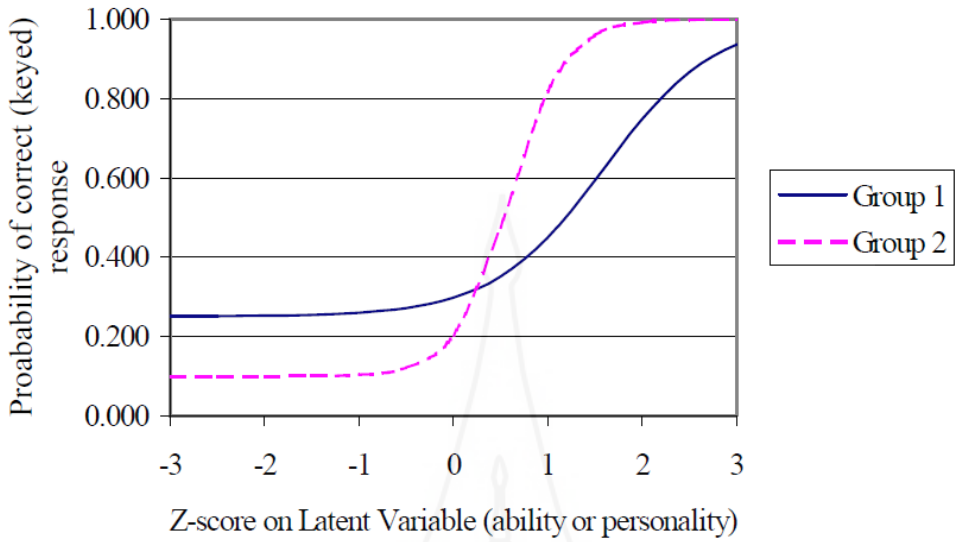
ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (uniform DIF) หมายถึง ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งในทุกระดับความสามารถหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่งอย่างคงที่ทุกระดับความสามารถ เมื่อพิจารณาจากโค้งคุณลักษณะของข้อสอบ (item characteristic curves ; ICCs) ระหว่างกลุ่มผู้สอบย่อย 2 กลุ่ม จะพบว่ากราฟจะขนานกันและไม่ตัดกันในลักษณะตรงข้ามหรือไม่เกิดปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม (group membership) ดังภาพที่ 2.1




ภาพที่ 2.1 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป มาจาก (Zumbo, 1999, p. 20)

### *2.3.2 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป*

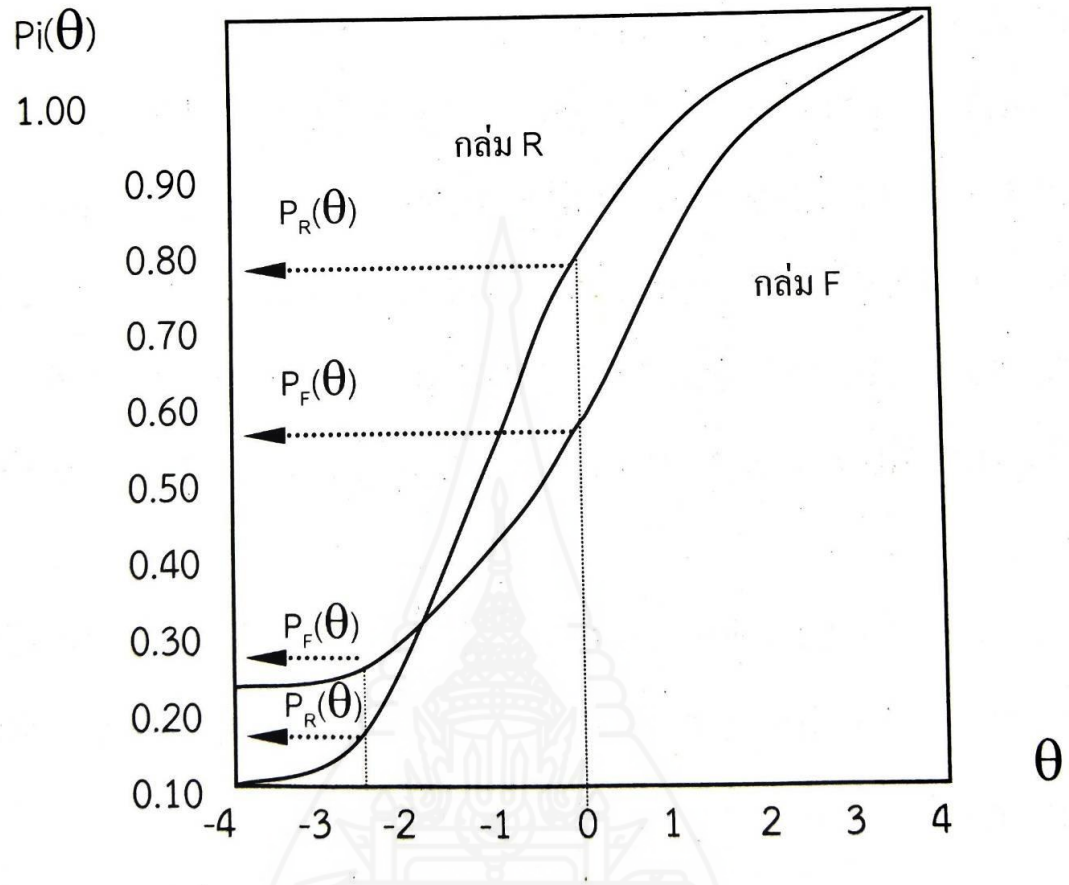
ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (nonuniform DIF) หมายถึง ข้อสอบที่ทำให้โอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอกันในทุกระดับความสามารถ นั้นคือ ความแตกต่างของความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบไม่ เหมือนกันทุกระดับความสามารถ เช่น ที่ระดับความสามารถหนึ่งผู้สอบที่กลุ่มอ้างอิงมีความน่าจะเป็นในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบกลุ่มเปรียบเทียบแต่ที่ระดับความสามารถอีกระดับหนึ่ง กลุ่มผู้สอบกลุ่มเปรียบเทียบหรือกลุ่มสนใจมีความน่าจะเป็นโอกาสตอบถูกน้อยกับมีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบกลุ่มอ้างอิง ซึ่งมีโอกาสตอบถูกมากกว่า ข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบประเภทนี้เมื่อมีปฏิสัมพันธ์ระหว่างความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม เมื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่าไม่ขนานกันและมีปฏิสัมพันธ์ร่วมกันระหว่างระดับความสามารถของผู้สอบกับสมาชิกของกลุ่มผู้สอบหรือมีฟังก์ชันการตอบข้อสอบ (IRFs) ต่างกันนั่นเองดังภาพที่ 2.2



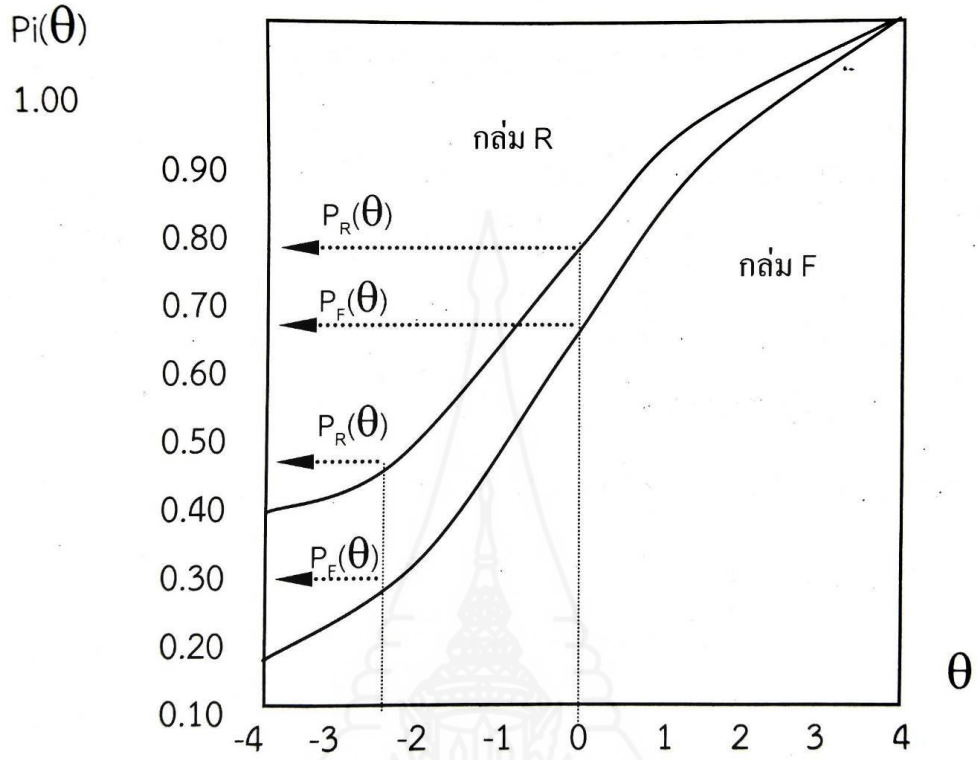

ภาพที่ 2.2 ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป มาจาก (Zumbo,1999 : 21)

ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory ; IRT) สามารถพิจารณา “ปฏิสัมพันธ์” ได้จากความแตกต่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบ ระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม คือ ถ้าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แล้วโค้งลักษณะข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะขนานกัน หรือมีฟังก์ชันการตอบสนองข้อสอบ (item response functions: IRFs) เหมือนกัน แต่ถ้าข้อสอบทำหน้าที่ต่างกัน แบบอเนกรูปแล้วโค้งลักษณะข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะไม่ขนานกัน หรือมีฟังก์ชันการตอบสนองข้อสอบต่างกัน

นอกจากนี้ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป ยังจำแนกได้ 2 ลักษณะภายใต้ทฤษฎีการตอบสนองข้อสอบ พบว่าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปโดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (disordinal interaction) การทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะตัดกันเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (non - directional DIF) ดังภาพที่ 2.3 และข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูปแบบมีปฏิสัมพันธ์เป็นลำดับ (ordinal Interaction) เป็นการทำหน้าที่ต่างกันสำหรับผู้สอบซึ่งเกิดขึ้นเมื่อโค้งลักษณะข้อสอบตัดกันไม่สม่ำเสมอ แต่ไม่ตัดกันหรืออาจตัดกันนอกช่วงเรียกลักษณะข้อสอบนี้ว่า ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว ดังภาพที่ 2.4




ภาพที่ 2.3 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (nondirectional DIF)




ภาพที่ 2.4 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว (unidirectional DIF)

**2.4 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ**

กระบวนการในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมี 2 ขั้นตอน คือ การตรวจสอบในระหว่างการสร้างข้อสอบโดยให้ผู้เชี่ยวชาญพิจารณาจากการวิเคราะห์หลักสูตรว่าข้อสอบสามารถวัดได้ตรงกับจุดมุ่งหมาย พฤติกรรม หรือเนื้อหาที่ต้องการวัดหรือไม่และรวมถึงภาษาที่ใช้ในการสื่อความหมายมีความชัดเจนหรือคลุมเครือเพื่อป้องกันการเกิดปัญหาในการตีความในการทำข้อสอบไปใช้ การตรวจสอบโดยใช้ผู้เชี่ยวชาญเป็นการตรวจสอบความตรงเชิงเนื้อหา ขั้นตอนที่ 2 การตรวจสอบโดยใช้ค่าสถิติ เมื่อมีการนำแบบทดสอบไปใช้ กระบวนการตรวจการทำหน้าที่ต่างกันของข้อสอบเป็นการเปรียบเทียบผลการตอบข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบอย่างน้อยสองกลุ่ม ที่มีความสามารถระดับเดี่ยวกัน โดยผู้สอบกลุ่มหนึ่งเป็นกลุ่มอ้างอิงซึ่งคาดว่าน่าจะได้เปรียบในการตอบข้อสอบข้อนั้น หรือมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่าอีกกลุ่มคือ กลุ่มเปรียบเทียบ ซึ่งเป็นกลุ่มที่สนใจศึกษาและคาดว่าน่าจะเป็นกลุ่มที่เสียเปรียบ ในการเปรียบเทียบนั้นจำเป็นต้องจับคู่ (matching) ผู้สอบตามความสามารถ ซึ่งเป็นเงื่อนไขสำคัญของการ

ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การจับคู่ (matching criteria) ที่นิยมใช้กันมี 2 วิธี คือเกณฑ์ภายนอก(external criterion) และเกณฑ์ภายใน (internal criterion) ดังนี้

***2.4.1 การตรวจสอบโดยใช้เกณฑ์ภายนอก***

การวิเคราะห์การทำหน้าที่ต่างกันโดยใช้เกณฑ์ภายนอกนี้ สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อและแบบสอบทั้งฉบับ โดยการใช้คะแนนจากแบบสอบอื่นเป็นเกณฑ์ภายนอก แล้วใช้เทคนิคการวิเคราะห์การถดถอย (regression analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์ กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ หลักการนี้มีจุดมุ่งหมาย เพื่อสร้างสมการทำนายตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบสอบอื่นจากตัวแปรทำนายซึ่งเป็นคะแนนรายข้อ หรือคะแนนแบบสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะใช้คะแนนรายข้อเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์ข้อสอบการทำหน้าที่ต่างกันของแบบสอบ จะใช้คะแนนรวมของแบบสอบทั้งฉบับเป็นตัวแปรทำนาย (Cronbach, 1970) สมการทำนายสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงได้ดังนี้ (ศิริชัย กาญจนวาสี, 2555, น. 122)

กลุ่มอ้างอิง (R) $Y_i = A_R + B_R X_i$

กลุ่มเปรียบเทียบ (F) $Y_i = A_F + B_F X_i$

เมื่อ $Y_i$ = คะแนนของตัวแปรเกณฑ์ภายนอก

$X_i$ = คะแนนของตัวแปรทำนาย

A = ค่าคงที่หรือค่าตัดแกน (intercept)

B = ค่าความชัน (slope)

จากฟังก์ชันการทำนายทั้ง 2 สมการ สามารถเปรียบเทียบค่าตัดแกน (A) และค่าความชัน(B) ของเส้นกราฟระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบได้ ถ้าเส้นกราฟดังกล่าวมีความชันหรือค่าตัดแกนแตกต่างกัน แสดงว่าข้อสอบหรือแบบสอบนั้นมีการทำหน้าที่ต่างกัน โดยลำเอียงเข้าข้างกลุ่มผู้สอบที่มีค่าตัดแกนที่สูงกว่าหรือค่าความชันที่สูงกว่า การใช้เกณฑ์ภายนอกมีข้อดี คือเกณฑ์ที่ใช้มีความเป็นอิสระจากข้อสอบ และแบบสอบที่ต้องการตรวจสอบ แต่มีจุดอ่อนตรงที่ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ ในทางปฏิบัติเป็นการยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบสอบฉบับอื่นที่มีความตรงเชิงทำนาย และมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าว จะทำให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบขาดความแม่นยำและความสมบูรณ์

***2.4.2 การตรวจสอบโดยใช้เกณฑ์ภายใน***

การวิเคราะห์การทำหน้าที่ต่างกันโดยใช้เกณฑ์ภายในเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบโดยเน้นการพิจารณาจากโครงสร้างของแบบสอบเป็นหลัก ด้วยการวิเคราะห์ผลจากการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบที่ได้จากแบบสอบฉบับนั้น การนำผลการตอบเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถหรือคะแนนจริงเท่ากัน ว่าจะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ การวิเคราะห์ลักษณะนี้นิยมใช้ค่าสถิติต่างๆ เป็นตัวบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ ซึ่งเป็นวิธีการสนับสนุนการตรวจสอบความตรงเชิงโครงสร้างเป็นวิธีที่ได้รับความนิยมเป็นจำนวนมาก เนื่องจากตัดปัญหาเรื่องเกณฑ์ที่ไม่มีมาตรฐาน การตรวจสอบสามารถดำเนินการได้จากการศึกษาโครงสร้างภายในของข้อสอบ โดยใช้ค่าสถิติทดสอบที่นิยมนำมาใช้พอสรุปได้ดังนี้

*1) การทดสอบปฏิสัมพันธ์ (interaction)* โดยการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อทดสอบปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับข้อสอบ ถ้าข้อสอบมีนัยสำคัญต้องวิเคราะห์ต่อด้วยวิธี Post Hoc เพื่อระบุข้อสอบที่มีผลต่อการเกิดปฏิสัมพันธ์ ซึ่งเป็นข้อที่ทำหน้าที่ต่างกัน วิธีการนี้มีข้อดีที่สามารถศึกษาผู้สอบหลายๆ กลุ่มได้สะดวก แต่มีจุดอ่อนในเรื่องการควบคุมกลุ่มต่างๆให้มีความสามารถที่ทัดเทียมกัน ขนาดกลุ่มตัวอย่างของกลุ่มต่างๆ และอัตราความคลาดเคลื่อนจะสูงขึ้น ถ้าจำนวนข้อสอบเพิ่มมากขึ้น

*2) การวัดความเบี่ยงเบนสัมพันธ์ (relative deviation)* เป็นการคำนวณค่าความยากของข้อสอบ เช่น p, b เป็นต้น เมื่อคำนวณแยกระหว่างกลุ่ม และแปลงให้เป็นค่าความยากมาตรฐาน ($\Delta$) สามารถนำมาพล็อตเปรียบเทียบเป็นรายข้อ รวมทั้งสามารถคำนวณค่าสหสัมพันธ์ระหว่างค่าความยากรายข้อระหว่างกลุ่ม เพื่อแสดงถึงการทำหน้าที่ต่างกันของแบบสอบ ถ้าสหสัมพันธ์เข้าใกล้ 1.00 แสดงว่าค่าความยากสัมพันธ์ของข้อสอบมีค่าใกล้เคียงกันระหว่างกลุ่ม ดังนั้นแบบสอบวัดคุณลักษณะคล้ายกันระหว่างกลุ่ม วิธีการนี้มีข้อดีและข้อเสียคล้ายการดสอบปฏิสัมพันธ์ นอกจากนี้ค่าความยากของข้อสอบ (p) มิใช่ตัวแทนของค่าวามยากจริงของข้อสอบ และได้รับอิทธิพลจากค่าแทรกซ้อนอื่นได้แก่ ค่าอำนาจจำแนก และความสามารถของผู้สอบ

***2.4.3 การเปรียบเทียบน้ำหนักตัวประกอบ*** เป็นการวิเคราะห์ตัวประกอบ (factor analysis) โดยใช้เทคนิคทางสถิติที่นิยมใช้ในการตรวจสอบความตรงเชิงทฤษฎีหรือโครงสร้าง (construct validity) เมื่อทำการวิเคราะห์ตัวประกอบมาใช้ในการวิเคราะห์โครงสร้างของแบบสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะสำคัญที่มุ่ง

วัด หรือความแตกต่างของค่าเฉลี่ยคะแนน ตัวประกอบ (factor scores) ระหว่างกลุ่มผู้สอบ ย่อยสะท้อนการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ การใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงสำรวจ (exporatory factor analysis ; EFA) จะมีจุดอ่อนในเรื่องความไม่สอดคล้องระหว่างน้ำหนักตัวประกอบอาจเกิดจากความแตกต่างของความสามารถระหว่างกลุ่มก็ได้ แนวทางที่เหมาะสมจึงควรใช้เทคนิคการวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis ; CFA) นอกจากนี้ยังสามารถตรวจสอบความแตกต่างระหว่างกลุ่ม ในด้านคุณลักษณะหรือความสามารถหลักและความสามารถรองได้อีกด้วย (Camilli and Shepard, 1994 อ้างใน ศิริชัย กาญจนวาสี, 2555, น. 123)

***2.4.4 การเปรียบเทียบโอกาสตอบข้อสอบถูก*** เป็นการวิเคราะห์โอกาสตอบข้อสอบถูกของผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากัน เป็นแนวทางสำคัญที่นิยมใช้กันและเป็นที่ยอมรับในปัจจุบัน สำหรับบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ มีการคำนวณค่าสถิติ 2 แนวทาง คือ เปรียบเทียบค่าสัดส่วนหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มที่มีความสามารถเท่ากัน เช่น วิธีแมนเทล-แฮนส์เซล เป็นต้น และเปรียบเทียบค่าฟังก์ชันการตอบสนองข้อสอบหรือโค้งลักษณะข้อสอบระหว่างกลุ่มที่มีระดับความสามารถเท่ากัน เป็นวิธีที่อยู่บนพื้นฐานของทฤษฎี IRT เช่น วิธีวัดความแตกต่างของพื้นที่ วิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก วิธีการทดสอบไค-สแควร์ของ ลอร์ด (Lord's $\chi^2$ – test) เป็นต้น วิธีการนี้มีข้อดีที่สำคัญได้แก่ การคำนวณค่าสถิติของข้อสอบมีความน่าเชื่อถือมีกลไกลควบคุมความสามารถของผู้สอบโดยการจับคู่กลุ่มความสามารถ เพื่อทำการเปรียบเทียบ ณ ตำแหน่งต่างๆ ที่มีความสามารถเท่ากัน จึงเป็นวิธีการที่ยอมรับกันทั่วไป แต่มีข้อจำกัดในด้านความสลับซับซ้อนของแนวคิดพื้นฐานและการวิเคราะห์มีความจำเป็นต้องใช้โปรแกรมคอมพิวเตอร์โดยเฉพาะ (ศิริชัย กาญจนวาสี, 2555, น. 123)

**2.5 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ**

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จำแนกตามลักษณะการตรวจให้คะแนนได้เป็น 2 ประเภทคือ ข้อสอบที่มีการให้คะแนนแบบทวิวิภาคหรือ 2 ค่า (dichotomous scoring) คือให้คะแนนแบบ 0, 1 และข้อสอบที่มีการให้คะแนนแบบพหุวิภาค หรือหลายค่า polytomous scoring) คือมีการให้คะแนนมากกว่า 2 ค่าในข้อสอบ ในวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแต่ละประเภท ยังสามารถจำแนกได้อีก 2 มิติ ได้แก่ มิติลักษณะของตัวแปรเกณฑ์ และมิติลักษณะของสถิติวิเคราะห์ มีรายละเอียดดังนี้

***2.5.1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิวิภาค***

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิวิภาค สามารถจำแนกได้ 2 กลุ่ม คือ กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้ และกลุ่มวิธีการที่ใช้คุณลักษณะแฝง เป็นต้น

*1) กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้* ซึ่งมักวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม หรือกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (non-IRT approach) ใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่ การวิเคราะห์ความแปรปรวน (Cleary and Hilton, 1968) วิธีการวิเคราะห์การถดถอยโลจิสติก วิธีการแปลค่าความยากของข้อสอบ (transformed item difficulty ; TID) วิธีแมนเทล-แฮนส์เซล วิธีดัชนีมาตรฐาน (Standardization ; STND) การปรับให้เป็นมาตรฐานด้วยน้ำหนักประกอบ (Dorans and Kulick , 1986)

*2) กลุ่มวิธีการที่ใช้คุณลักษณะแฝง* คือ วิธีในกลุ่มนี้ใช้คุณลักษณะหรือตัวแปรแฝง ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) สำหรับใช้เป็นเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ ได้แก่ วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนองข้อสอบ (IRT – $D^2$) วิธีไค-สแควร์ของลอร์ด วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (general IRT likelihood ratio) วิธีอัตราส่วนไลค์ลิฮูด ลอกลิเนียร์ (loglinear IRT likelihood ratio) วิธีซิปเทสท์

***2.5.2 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบพหุวิภาค*** ได้แก่ กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้ และกลุ่มวิธีที่ใช้คุณลักษณะแฝง ดังนี้

*1) กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้* ได้แก่ วิธีการวิเคราะห์ความแปรปรวน วิธีการวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (polytomous logistic regression) วิธีดัชนีมาตรฐานพหุวิภาค (polytomous standardization) วิธีแมนเทล – แฮนส์เซลทั่วไป (General Mantel – Haenzel ; GMH)

*2) กลุ่มวิธีที่ใช้คุณลักษณะแฝง* ได้แก่ วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (general IRT likelihood ratio) วิธีการให้คะแนนบางส่วน (Partial Credit Model ; PCM) วิธีซิปเทสท์พหุวิภาค (polytomous SIBTEST) วิธีการให้คะแนนบางส่วนทั่วไป (Generalized Partial Credit Model; GPCM) (ศิริชัย กาญจนวาสี, 2555, น. 126)

**2.6 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก**

Swaminathan and Rogers (1990, pp. 362 - 363) ได้พัฒนาโมเดลการถดถอยโลจิสติก จากวิธีลอกลิเนียร์ ของ Mellenberg (1982) วิธีแมนเทล – แฮนส์เซล ของ Holland &

Thayer (1988) พร้อมทั้งเชื่อมแนวความคิดของวิธีที่ใช้คุณลักษณะแฝงของทฤษฎีการตอบข้อสอบ วิธีนี้ตั้งอยู่บนฐานของโมเดล (model-based methods) ที่ให้ความสำคัญกับธรรมชาติของความต่อเนื่องบนสเกลความสามารถ เพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิภาค หรือสองค่าหลักการตรวจสอบจะใช้โมเดลการถดถอยโลจิสติกทำนายความน่าจะเป็นในการตอบสนองของข้อสอบถูก โดยปกติแล้วการใช้คะแนนแบบทดสอบรวมแทนระดับความสามารถซึ่งสมมุติว่าเป็นตัวแปรความสามารถแบบต่อเนื่อง โมเดลดังกล่าวมีพจน์สำหรับทดสอบปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มผู้สอบกับระดับความสามารถ ดังนั้นจึงสามารถตรวจสอบได้ทั้งข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปและข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป โมเดลการถดถอยโลจิสติกมีความยืดหยุ่นสามารถนำไปปรับขยายเพื่อใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบพหุวิภาค หรือหลายค่า

Swaminathan and Rogers (1990) กล่าวว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีคะแนนแบบสองค่าจะใช้โมเดลการถดถอยโลจิสติก ทำนายความน่าจะเป็นในการตอบข้อสอบถูก ดังนี้

$$P(u = 1 \mid \theta) = \frac{e^{(\beta_0+\beta_1\theta)}}{1+e^{(\beta_0+\beta_1\theta)}}$$

เมื่อ $u$ แทน ผลการตอบข้อสอบ

$\theta$ แทน ความสามารถที่สังเกตได้ของผู้สอบแต่ละคน

$\beta_0$ แทน พารามิเตอร์จุดตัด

$\beta_1$ แทน พารามิเตอร์ความชัน

สมการดังกล่าวเป็นโมเดลการถดถอยโลจิสติกแบบมาตรฐานสำหรับใช้ทำนายตัวแปรตาม แบบสองค่าจากตัวแปรอิสระที่กำหนดให้ โมเดลการถดถอยโลจิสติกสามารถนำไปประยุกต์ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยระบุสมการแยกกลุ่มผู้สอบสองกลุ่มที่สนใจดังนี้

$$P(u_{ij} = 1/\theta_{ij}) = \frac{e^{(\beta_{0j}+\beta_{1j}\theta_{1j}}}{\left[1+e^{(\beta_{0j}+\beta_{1j}\theta_{1j}}\right]} \quad , \quad i = 1,2,\ldots,n_j : j = 1,2$$

เมื่อ $U_{ij}$ แทน โอกาสของการตอบข้อสอบถูกของบุคคลที่ i กลุ่มที่ j

$\theta_{ij}$ แทน ค่าความสามารถของผู้สอบที่ i กลุ่มที่ j

$\beta_{0j}$ แทน พารามิเตอร์จุดตัดในกลุ่ม j

$\beta_{1j}$ แทน ค่าพารามิเตอร์ความชันในกลุ่ม j

โดยที่ $\beta_{01} = \beta_{02}$ และ $\beta_{11} = \beta_{12}$ เป็นลักษณะโค้งถดถอยโลจิสสำหรับผู้สอบสองกลุ่มเหมือนกันแสดงว่าข้อสอบทำหน้าที่ไม่แตกต่างกัน

$\beta_{01} \neq \beta_{02}$ แต่ $\beta_{11} = \beta_{12}$ เป็นลักษณะโค้งถดถอยโลจิสสำหรับผู้สอบสองกลุ่มเท่าเทียมกัน แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป

$\beta_{01} = \beta_{02}$ แต่ $\beta_{11} \neq \beta_{12}$ เป็นลักษณะโค้งถดถอยโลจิสสำหรับผู้สอบสองกลุ่มไม่เท่าเทียมกัน แสดงว่าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป

จากสมการการถดถอยโลจิสติกข้างต้น สามารถเขียนเป็นสมการใหม่ในรูปแบบการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้โมเดลที่สมมูลกับโมเดลของสมการ ได้ดังนี้

$$P(u_{ij}= 1) = \frac{e^{Z_{ij}}}{1+e^{Z_{ij}}}$$

โดยที่ $Z_{ij} = \tau_0 + \tau_1\theta + \tau_2 g + \tau_3(\theta g)$

เมื่อ $P(u_{ij} = 1)$ = ความน่าจะเป็นการตอบข้อสอบถูกของผู้สอบที่มีความสามารถ $\theta$ คนที่ i ในกลุ่ม j

$\theta$ แทน ความสามารถของผู้สอบ

g แทน กลุ่มผู้สอบ( g = 1 สมาชิกกลุ่ม 1 หรือกลุ่มอ้างอิง, g = 0 สมาชิกกลุ่มที่ 2 หรือกลุ่มเปรียบเทียบ)

$\theta g$ แทน ปฏิสัมพันธ์ของตัวแปรความสามารถและกลุ่ม

$\tau_0$ แทน ค่าพารามิเตอร์จุดตัด

$\tau_1$ แทน ผลของความสามารถของผู้สอบ

$\tau_2$ แทน ความแตกต่างระหว่างกลุ่มในในการตอบข้อสอบถูก

โดยที่ $\tau_2 = \beta_{01} - \beta_{02}$

$\tau_3$ แทน ปฏิสัมพันธ์ระหว่างกลุ่มกับความสามารถผู้สอบ

โดยที่ $\tau_3 = \beta_{11} - \beta_{12}$

ในการตัดสินว่าข้อสอบทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปหรืออเนกรูป พิจารณาดังนี้

ถ้า $\tau_2 \neq 0$ และ $\tau_3 = 0$ แสดงว่าข้อสอบข้อนั้นเป็นเอกรูป

แต่ถ้า $\tau_2 = 0$ และ $\tau_3 \neq 0$ หรือไม่ก็ได้ แสดงว่าข้อสอบข้อนั้นเป็นอเนกรูป

Zumbo (1999) ได้ศึกษาโมเดลการถดถอยโลจิสติกที่มีการให้คะแนนแบบสองค่าของ Swaminathan and Rogers. (1990) และปรับปรุงโมเดลการถดถอยโลจิสติกที่ใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า ซึ่งได้แก่โมเดลโลจิทของอัตราส่วนแบบต่อเนื่อง (continuation ratio logits) โมเดลโลจิทของข้อสอบที่อยู่ติดกัน( Adjacentcategory logits) และโมเดลโลจิทสะสม (cumulative logits) ของ French and Miller (1996) โดยที่แต่ละโมเดลค่าการคิดคำนวณจะแยกกันแต่ละโลจิท ทำให้ยากต่อการระบุการทำหน้าที่ต่างกันของข้อสอบ Zumbo (1999) ได้นำโมเดลโลจิทสะสม มาพัฒนาใช้กับข้อสอบแบบจัดอันดับ โดยทำให้โลจิททั้งหมดสามารถทดสอบได้พร้อมกัน และโมเดลการถดถอยโลจิสติกมีพื้นฐานทางสถิติของความน่าจะเป็นในการตอบข้อสอบหนึ่งข้อที่ให้คะแนนแบบสองค่าได้ถูกต้อง วิธีนี้จะใช้ในการตอบข้อสอบ (0 หรือ 1) เป็นตัวแปรตาม (dependent variable ) ที่มีตัวแปรกลุ่ม (ซึ่งกำหนดรหัสดัมมีเป็น 1 สำหรับกลุ่มอ้างอิงและ 0 สำหรับกลุ่มสนใจหรือเปรียบเทียบ ในที่นี้เป็นตัวแปร Group คะแนนรวมสำหรับผู้เข้าสอบแต่ละคน อาจเป็นความสามารถ หรือคุณลักษณะ ในที่นี้เป็นตัวแปร TOT (total test score ) และปฏิสัมพันธ์ระหว่างคะแนนรวมกับกลุ่มผู้สอบ(TOT * GROUP) เป็นตัวแปรอิสระ( independent variable) วิธีนี้ได้ให้การทดสอบการทำหน้าที่ต่างกันของข้อสอบโดยมีเงื่อนไขในความสัมพันธ์ระหว่างการตอบข้อสอบและคะแนนรวม การทดสอบผลของกลุ่มผู้สอบเพื่อตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเอกรูป และการทดสอบปฏิสัมพันธ์ของคะแนนรวมและกลุ่มผู้สอบ เพื่อตรวจสอบการทำหน้าที่ต่างกันของ ข้อสอบที่เป็นรูปแบบอเนกรูป

สมการการถดถอยโลจิสติกดังกล่าวเป็นดังนี้

$$Y = b_0 + b_1TOT + b_2GROUP + b_3TOT*GROUP \qquad (1)$$

เมื่อ Y เป็นลอการิทึมธรรมชาติของอัตราส่วนแต้มต่อ (odds ratio)สมการ (1) เขียนใหม่ได้ดังนี้

$$\ln\left[\frac{P_i}{1-p_i}\right] = b_0 + b_1tot + b_2group + b_3(tot*group) \qquad (2)$$

เมื่อ P แทนสัดส่วนของผู้สอบที่ตอบข้อสอบถูก

ขณะเดียวกันสามารถที่จะแสดงการถดถอยโลจิสติกในรูปการถดถอยเชิงเส้น (linear regression) ของตัวแปรที่ทำนาย (predictor variable) บนตัวแปรสุ่มที่มีการแจกแจงแบบต่อเนื่องที่ไม่สามารถสังเกตได้ ดังนั้นสามารถเขียนสมการ (1) ได้ดังนี้

$$y^* = b_0 + b_1TOT + b_2GROUP + b_3TOT*GROUP + \varepsilon_i \qquad (3)$$

y* แทน ตัวแปรตามแบบต่อเนื่องที่ไม่สามารถสังเกตได้

$\varepsilon_i$ แทน ความคลาดเคลื่อน ซึ่งการแจกแจงของ $\varepsilon_i$ ที่มีค่าเฉลี่ยเป็น 0 และความแปรปรวนเป็น $\pi^2/3$

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่า โดยใช้โมเดลการถดถอยโลจิสติกแบบจัดอันดับตามแนวคิดของ Zumbo (1999) จะประมาณค่าความน่าจะเป็นสะสม (cumulative probability) ในข้อสอบข้อที่ 1 ถึง J-1 เมื่อ J เป็นจำนวนข้อสอบในมาตรแบบจัดอันดับ โดยที่ความน่าจะเป็นสะสมในข้อสอบของคำตอบสุดท้ายที่มีค่าเท่ากับ 1 ตัวอย่างเช่นข้อสอบมีรายการคำตอบ 3 รายการ โมเดลการถดถอยจะประมาณค่าความน่าจะเป็นสะสมในรายการที่ 1 ถึง 2 คือความน่าจะเป็นสะสมสำหรับการตอบในรายการที่ 1หรือต่ำกว่า และความน่าจะเป็นสะสมสำหรับการตอบในรายการที่ 2 หรือต่ำกว่า โดยที่ความน่าจะเป็นสะสมในรายการที่ 3 ซึ่งเป็นรายการสุดท้ายที่มีค่าเท่ากับ 1 ดังนั้นสามารถเขียนสมการแสดงการถดถอยโลจิสติกแบบจัดอันดับในรูปลอการิทึมธรรมชาติของอัตราส่วนแต้มต่อ

$$\ln\left[\frac{P(Y\leq j)}{P(Y>j)}\right] = \alpha_j + bX$$

โดยที่

$$\ln\left[\frac{p\,(Y\leq j)}{P\,(Y>j)}\right] = \text{logit}[P(Y \leq j)]$$

จะได้ว่า $\text{logit}[P(Y \leq j)] = \alpha_j + bX \qquad (4)$

เมื่อ $P(Y \leq j)$ แทน ความน่าจะเป็นสะสมสำหรับผู้ตอบในข้อสอบที่ j หรือต่ำกว่า

$\alpha_j$ แทน พารามิเตอร์ส่วนตัดของความน่าจะเป็นสะสมในแต่ละรายการ

$b$ แทน พารามิเตอร์ที่เกี่ยวข้องกับระดับความสามารถของผู้สอบ

$X$ แทน ระดับความสามารถของผู้สอบ

โมเดลสมการ (4) เป็นสมการที่ไม่ปรากฏการที่หน้าที่ต่างกันของข้อสอบโมเดลนี้ทำนายคะแนนข้อสอบจากระดับความสามารถ (คะแนนรวม) ผู้สอบเท่านั้น เมื่อพิจารณาโมเดลในสมการ (4) ทำให้สามารถเขียนสมการที่เหมาะสมสำหรับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบสำหรับสมการ (3) ดังนี้

$$\text{Logit}\,[P(P(Y \leq j)\} = \alpha_j + b_1\text{tot} + b_2\text{group} + b_3\,(\text{tot*group}) \qquad (5)$$

เมื่อ $P(Y \leq j)$ แทน ความน่าจะเป็นสะสมสำหรับผู้ตอบในข้อสอบที่ jหรือต่ำกว่า

$\alpha_j$ แทน พารามิเตอร์ส่วนตัดของความน่าจะเป็นสะสมในข้อสอบที่ j

tot แทน ระดับความสามารถของผู้สอบ

group แทน กลุ่มผู้สอบ

tot*group แทน ปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับกลุ่มผู้สอบ

$b_1$ แทน พารามิเตอร์ที่เกี่ยวข้องกับระดับความสามารถของผู้สอบ (X)

$b_2$ แทน พารามิเตอร์ที่เกี่ยวข้องกับกลุ่มผู้สอบ (G)

$b_3$ แทน พารามิเตอร์ที่เกี่ยวข้องกับปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับกลุ่มผู้สอบ

จากโมเดลสมการ (5) เป็นโมเดลเต็ม (full model) ซึ่งประกอบด้วยการทำหน้าที่ต่างกันของข้อสอบที่เป็นแบบเอกรูปและการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป ในโมเดลเต็ม คะแนนข้อสอบถูกทำนายจากระดับความสามารถของผู้สอบ (คะแนนรวม) การเป็นสมาชิกกลุ่มและปฏิสัมพันธ์ของระดับความสามารถกับการเป็นสมาชิกกลุ่มผู้สอบ

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีการถดถอยโลจิสติกจะใช้กระบวนการตรวจสอบความเหมาะสมของข้อมูลกับโมเดลโดยจะทดสอบโมเดล 3 ขั้นดังนี้

ขั้นที่ 1 ทดสอบตัวแปรที่ใช้ในการจับคู่ความสามารถ (matching variable) หรือตัวแปรเงื่อนไข (conditioning variable) ในที่นี้ คือ คะแนนเป็นระดับความสามารถของผู้สอบ ในขั้นตอนนี้ถือว่าขนาดของตัวแปรคะแนนรวมมีค่าเท่ากันในทุกข้อ และสมมติให้ค่าประมาณของพารามิเตอร์ความชันมีเพียงค่าเดียว การนำตัวแปรดังกล่าวเข้าไปในสมการถดถอยมีลักษณะดังนี้

$$\text{Logit}[P(Y \leq j)] = \alpha_j + b_1\text{tot} \qquad (6)$$

ขั้นที่ 2 ทดสอบตัวแปรสมาชิกกลุ่มของผู้สอบในที่นี้จะกำหนดรหัสเป็นตัวแปรดัมมี (dummy variable) โดยให้กลุ่มอ้างอิงแทนด้วย 1 และกลุ่มสนใจแทนด้วย 0 และเพิ่มตัวแปรดังกล่าวในสมการ (6) ของขั้นตอนหนึ่งดังนี้

$$\text{Logit}\,[P(Y \leq j)] = \alpha_j + b_1\text{tot} + b_2\text{group} \qquad (7)$$

ขั้นที่ 3 ทดสอบตัวแปรปฏิสัมพันธ์ระหว่างระดับความสามารถ กับสมาชิกของกลุ่มผู้สอบ โดยเพิ่มพจน์ทั้งสองดังกล่าวในโมเดลสมการ (7) ของขั้นตอนที่หนึ่งดังนี้

$$\text{logit}\,[P(Y \leq j)] = \alpha_j + b_1\,\text{tot} + b_2\text{group} + b_3(\text{tot*group}) \qquad (8)$$

การทดสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเอกรูปและอเนกรูปสามารถทดสอบได้พร้อมกัน โดยกำหนดสมมติฐานศูนย์ ($H_0$) และสมมติฐานทางเลือก ($H_1$) ดังนี้

$H_0 : b_2 = 0$ และ $b_3 = 0$

$H_1 : b_2 \neq 0$ และ $b_3 \neq 0$

การทดสอบนัยสำคัญทางสถิติของการทำหน้าที่ต่างกันของข้อสอบจะดำเนินการตามกระบวนการทดสอบความเหมาะสมของข้อมูลกับโมเดล 3 ขั้นตอนในการนำตัวแปรเข้าทดสอบกับโมเดลตามลำดับขั้นดังได้กล่าวมาแล้ว ขั้นที่ 1 ใส่ตัวแปรควบคุม(คะแนนรวม เข้าไปในสมการ ขั้นที่ 2 ใส่ตัวแปรกลุ่ม เข้าไปในสมการของขั้นที่ 1 และ ขั้นที่ 3 ใส่ตัวแปรปฏิสัมพันธ์ระหว่างระดับความสามารถ (คะแนนรวม) กับสมาชิกกลุ่มผู้สอบ เข้าไปในสมการของขั้นที่ 2 ดังกล่าวและการทดสอบไค-สแควร์สำหรับการถดถอยโลจิสติกหนึ่งการถดถอยนั้นสามารถคำนวณการทดสอบทางสถิติของการทำหน้าที่ต่างกันของข้อสอบซึ่งในการตัดสินใจว่าข้อสอบใดทำหน้าที่ต่างกันหรือไม่พิจารณาได้จากความแตกต่างระหว่างค่าไค-สแควร์ที่ได้รับมานั้นจากขั้นตอนที่ 3 และขั้นตอนที่ 1 ผลของค่าไค-สแควร์ที่ได้สามารถเปรียบเทียบกับฟังก์ชั่นการแจกแจงแบบไค-สแควร์ที่มีระดับชั้นความเป็นอิสระเท่ากับ 2 (df=2) โดยมีค่า $\chi^2$ ซึ่งค่าของ df ดังกล่าวเกิดจากความแตกต่างของ df ของขั้นตอนทั้งสองเป็นเช่นนี้เพราะสถิติไค-สแควร์ของโมเดล ในขั้นตอนที่ 3 มีระดับชั้นความเป็นอิสระเท่ากับ 3 (df=3 )และสถิติไค-สแควร์ของโมเดลขั้นตอนที่ 1 มีระดับชั้นความเป็นอิสระเท่ากับ 1(df=1 )ผลจากการการทดสอบค่าไค-สแควร์ที่ระดับ df=2 เป็นการทดสอบพร้อมกันของการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเอกรูป และการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบอเนกรูป ถ้าผลการทดสอบมีค่านัยสำคัญทางสถิติ โดยปฏิเสธสมมติฐานศูนย์ นั่นแสดงว่าข้อสอบทำหน้าที่ต่างกันซึ่งเป็นผลมาจากอิทธิพลของการเป็นสมาชิกของกลุ่มผู้สอบ และ/หรืออิทธิพลของปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มกับระดับความสามารถ ( Zumbo, 1999, p. 26)

Swaminathan and Rogers (1990) มุ่งเน้นที่ความสนใจบนการทดสอบไค-สแควร์ที่มีระดับ df=2 ดังนั้นขนาดของผล( effect size)ที่ได้จะเป็น $R^2$ (มาจาก $R^2$ ของขั้นตอนที่ 3 ลบด้วย $R^2$ ของขั้นตอนที่ 1) ซึ่งเป็นผลมาจากทั้งอิทธิพลของการเป็นสมาชิกของกลุ่ม หรืออิทธิพลของปฏิสัมพันธ์ระหว่างการเป็นสมาชิกของกลุ่มกับระดับความสามารถของผู้สอบพร้อมกัน ถือเป็นการ

วัดการทำหน้าที่ต่างกันของข้อสอบทั้งที่เป็นรูปแบบเอกรูปและรูปแบบอเนกรูป นอกจากนี้วิธีดังกล่าวทำให้เราสามารถเปรียบเทียบค่า $R^2$ ขั้นตอนที่ 1 กับค่า $R^2$ ขั้นตอนที่ 2 ซึ่งเป็นผลมาจากอิทธิพลของกลุ่ม (ความแตกต่างกลุ่ม)และเปรียบเทียบค่า $R^2$ ขั้นตอนที่ 2 กับค่า $R^2$ ขั้นตอนที่ 3 ซึ่งเป็นผลมาจากอิทธิพลของปฏิสัมพันธ์กลุ่มและความสามารถซึ่งวิธีการดังกล่าวสามารถนำไปใช้ตัดสินใจว่าข้อสอบข้อใดทำหน้าที่ต่างกันแบบเอกรูป ข้อใดทำหน้าที่อเนกรูป โดยพิจารณาจากค่าของ $R^2$ ที่คิดคำนวณได้ระหว่างช่วงแรก (ขั้นตอนที่ 1 กับ 2) และช่วงหลัง (ขั้นตอนที่ 2 กับ 3 ) ถ้าค่าความแตกต่าง $R^2$ ในช่วงแรกมากกว่าความแตกต่าง $R^2$ ในช่วงหลังสามารถสรุปได้ว่าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แต่ถ้าค่าความแตกต่าง $R^2$ ในช่วงแรกน้อยกว่าค่าความแตกต่าง $R^2$ ในช่วงหลัง สรุปได้ว่าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (Zumbo, 1999, p. 26)

### *2.6.1 การวัดขนาดของการทำหน้าที่ต่างกันของข้อสอบ*

การวัดขนาดของการทำหน้าที่ต่างกันของข้อสอบ พิจารณาจากตารางที่ 2.2

ตารางที่ 2.2 ค่า $R^2$ ของการวัดขนาดของการทำหน้าที่ต่างกันของข้อสอบ(Zumbo, 1999. p. 28)

| Item Scoring | Measure | Notes |
|---|---|---|
| Ordinal | R-squared for ordinal | McKelvey & Zavoina(1975) |
| Binary (nominal) | Nagelkerke R – squared | Nagelkerke Zc.f.,Thomas & Zumbo, 1998) |
| Binary (nominal) | Weighted-least-squares R-squared | Thomas & Zumbo (1998) |
| Binary (nominal) | R-squared for ordinal (i.e.,same as above) | McKelvey & Zavoina(1975) |

จากตารางที่ 2.2 พบว่า การคำนวณหาค่า $R^2$ ในการวัดขนาดของการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีรูปแบบการให้คะแนนแบบพหุวิภาคมีเพียงวิธีเดียว คือ $R^2$ ที่เป็นลำดับขั้น( R-squared for ordinal) ข้อสอบที่มีรูปแบบการให้คะแนนแบบทวิวิภาคมีการคำนวณ $R^2$ สามวิธีประกอบด้วย

1) วิธีการคำนวณ $R^2$ ของ Nagelkerke (Nagelkerke R – squared) เป็นกลุ่มนามบัญญัติเป็นวิธีการคำนวณที่ง่ายที่สุดสามารถใช้โปรแกรมสำเร็จรูป SPSS ในการคำนวณได้

2) วิธีการคำนวณ $R^2$ ของ WLS (Weighted-least-squares Squared) เป็นกลุ่มนามบัญญัติ วิธีการนี้เหมาะสำหรับใช้ประโยชน์กับกลุ่มตัวอย่างที่มีตัวแปรเป็นลำดับขั้นภายในกลุ่ม

3) วิธีการคำนวณ $R^2$ สำหรับการจัดลำดับข้อมูลหรือ $R^2$ ที่เป็นลำดับขั้น(R-squared for ordinal) เป็นกลุ่มจัดลำดับ เหมาะกับกลุ่มตัวแปรที่อยู่ในงานวิจัยประเภทพฤติกรรมศาสตร์และสังคมศาสตร์ เป็นตัวแปรที่มีความเป็นคุณลักษณะแฝงที่มีความต่อเนื่อง (Zumbo, 1999. p. 28)

Swaminathan and Rogers (1990) ได้ให้ความสำคัญกับความเป็นอิสระของทั้งสองระดับในการทดสอบแบบไคสแควร์ ดังนั้นขนาดของการทำหน้าที่ต่างกันของข้อสอบที่เกี่ยวข้องจะเป็น R-squared ที่เกี่ยวข้องกับกลุ่มและเงื่อนไข(เช่น R-squared ที่ขั้นตอนที่ 3 ลบ R-squared ในขั้นตอนที่ 1 ) ในขณะเดี่ยวกันวิธีการถดถอยโลจิสติกในการวัดการทำหน้าที่ต่างกันของข้อสอบจะวัดจากการทดสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและอเนกรูป ตามแบบจำลองลำดับขั้นตอนข้างต้นช่วยให้สามารถเปรียบเทียบค่า R – squared ในขั้นตอนที่ 2 กับค่า R - squared ในขั้นตอนที่ 1 เพื่อวัดรูปแบบที่ไม่ซ้ำกันซึ่งเป็นผลมาจากความแตกต่างของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่สามารถคำนวณได้จากคะแนนรวมของการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป นอกจากนี้การเปรียบเทียบค่า R-squared สำหรับขั้นตอนที่ 3 และ 2 จะวัดรูปแบบข้อสอบที่เป็นแบบอเนกรูปซึ่งเป็นผลมาจากการปฏิสัมพันธ์ดังนั้นข้อได้เปรียบในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบของวิธีการถดถอยโลจิสติกคือสามารถทดสอบทั้งการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปและแบบอเนกรูปได้ในเวลาเดียวกัน (Zumbo, 1999, p. 27)

**2.7 สาเหตุการทำหน้าที่ต่างกันของข้อสอบ**

สาเหตุของการทำหน้าที่ต่างกันของข้อสอบนั้นมีการศึกษาทำวิจัยอย่างต่อเนื่องและค้นพบว่ามีงานวิจัยหลายชิ้นที่มุ่งประเด็นไปที่สาเหตุการทำหน้าที่ต่างกันของข้อสอบ ซึ่งในระยะหลังมีการกล่าวถึงสาเหตุการทำหน้าต่างกันของข้อสอบอย่างมากมายโดยนักการศึกษาต่างๆ มีความคิดเห็นเกี่ยวกับสาเหตุการทำหน้าที่ต่างกันของข้อสอบที่ แตกต่างกัน ดังนี้

Scheuneman (1982) อ้างถึงใน สุมาลี แกล้วทนงค์, 2547. น. 27) กล่าวว่า สาเหตุของการทำหน้าที่ต่างกันของแบบสอบน่าจะเกิดจากสาเหตุดังนี้

1. การเดา (guessing) ซึ่งอาจเกิดจากข้อสอบมีความยากเกินไปหรือผู้เข้าสอบไม่เคยผ่านประสบการณ์เกี่ยวกับเรื่องที่ถามสำหรับกลุ่มผู้เข้าสอบบางกลุ่ม จึงทำให้ผู้เข้าสอบตอบคำตอบโดยการเดาข้อสอบ ทำให้เกิดความไม่เท่าเทียมกันเกี่ยวกับโอกาสในการการตอบข้อนั้นถูก

2. ความเร็ว (speed) หรือเวลาในการตอบ ซึ่งเป็นผลให้ทำข้อสอบไม่ทันในกรณีที่เวลาน้อย โดยจะมีผลกับการทำข้อสอบข้อหลังๆ

3. ความกำกวมหรือความไม่ชัดเจนของข้อคำถาม (unclear) คือ ข้อคำถามขาดความเป็นปรนัยหรือการใช้สำนวนภาษาที่เป็นภาษาเฉพาะซึ่งเป็นสิ่งที่ก่อให้เกิดความแตกต่างระหว่างกลุ่ม

4. ลำดับขั้นของคำถาม (series) เช่นการตั้งคำถามโดยการลำดับขั้นของคำถามี่มีความสับสน

5. สถานภาพทางสังคมหรือความเกี่ยวข้องทางสังคม (social implication)

6. ประสบการณ์หรือการฝึกของแต่ละกลุ่มย่อย

7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ ประเพณี เชื้อชาติ

Taylor and Orlando (1987 อ้างถึงใน สุมาลี แกล้วทนงค์, 2547, น. 27) จำแนกสาเหตุของความลำเอียงจากสิ่งที่เกิดขึ้นกับผู้สอบว่าเนื่องมาจากการสื่อสารที่ไม่ตรงกันจากแบบสอบ ประกอบด้วย

1. ความลำเอียงจากสถานการณ์ในข้อสอบ (situation bias) เกิดจากความเข้าใจไม่ตรงกัน ระหว่างผู้ออกข้อสอบและผู้เข้าสอบเนื่องจากมีการใช้สถานการณ์ในการออกข้อสอบและใช้ภาษาที่ไม่เหมาะของผู้ออกข้อสอบ

2. ความลำเอียงจากคำแนะนำหรือคำสั่งในแบบสอบ (directions bias) เกิดจากคำแนะนำ หรือคำสั่งจากแบบสอบรวมไปถึงความซับซ้อน ความกำกวมของคำสั่งและความไม่คุ้นเคยจากคำแนะนำ หรือคำสั่งในแบบสอบของผู้สอบ

3. ความลำเอียงจากการตีค่า (value bias) เกิดจากการตีค่าที่แตกต่างกันของผู้เข้าสอบ โดยผู้สอบจะเลือกคำตอบจากการตีค่าหรือความรู้สึกของตนเองเป็นหลัก

4. ความลำเอียงที่เกิดจากภาษา (linguistics bias) ได้แก่ ความแตกต่างจากการใช้ภาษาหรืออิทธิพลของภาในข้อสอบซึ่งเกิดการพูดหรือใช้ภาษาอย่างไม่เป็นมาตรฐาน ดังนั้น การเลือกตัวเลือกของ ข้อสอบจะใช้มาตรฐานที่ต่างกัน

5. ความลำเอียงจากรูปแบบ (format bias) เกิดจากการปฏิบัติของข้อสอบหรือข้อกำหนดที่ไม่สอดคล้องกับความเข้าใจหรือรูปแบบการเรียนรู้ของผู้สอน

6. ความลำเอียงทางด้านวัฒนธรรม (cultural bias) เกิดจากผู้เข้าสอบเข้าใจเนื้อหาในแบบสอบไขว้เขว หรือแปลเรื่องไปทางที่เกี่ยวกับวัฒนธรรมของตนจนทำให้เกิดการเข้าใจผิด

7. ความลำเอียงจากการแปลความหมายผิด (misinterpretation bias) เกิดจากแปลความหมายไปตามประสบการณ์และการรับรู้เดิมของตนที่มีมาก่อน

8. ความลำเอียงจากสิ่งเร้า (stimulus bias) เกิดจากผู้ออกข้อสอบแสดงสิ่งเร้าในวัตถุประสงค์ และเนื้อหาสาระที่นำมาวิเคราะห์ไปยังผู้เข้าสอบ

Hambleton and Rogers (1995 อ้างอิงในสุมาลี แกล้วทนงค์, 2547, น. 28) เชื่อว่าสาเหตุของการทำหน้าที่ต่างกันของข้อสอบมาจากแบบสอบ จำแนกสาเหตุของความลำเอียงและวิธีการสังเกต ดังนี้

1. ความลำเอียงที่เกิดจากเนื้อหาของแบบสอบ (content bias) สังเกตได้จาก

1.1 ข้อสอบจะประกอบไปด้วยเนื้อหาของที่มีความแตกต่างและไม่คุ้นเคยความหมายระหว่างกลุ่ม

1.2 สมาชิกในกลุ่มสามารถตอบข้อสอบได้ถูกหรือผิดโดยมีเหตุผลการตอบข้อสอบที่ต่างกัน

1.3 เนื้อหาในข้อสอบและการตอบถูกสะท้อนให้เห็นข้อมูลบอกใบ้ และมีความเกี่ยวข้องกับพื้นฐานทางการศึกษาของผู้สอบบางคนเท่านั้น

1.4 เนื้อหาของข้อสอบประกอบด้วยข้อมูลที่ได้เปรียบแก่ผู้สอบบางกลุ่ม

2. ความลำเอียงทางด้านภาษา (language bias) สังเกตได้จาก

2.1 ข้อสอบจะประกอบด้วยคำซึ่งมีความแตกต่างกันหรือไม่คุ้นเคยเกี่ยวกับความหมายของคำระหว่างกลุ่มผู้สอบ

2.2 ข้อสอบจะประกอบไปด้วยคำศัพท์ยาก

2.3 ข้อสอบจะประกอบด้วยภาษาเฉพาะกลุ่ม คำศัพท์เฉพาะ และสรรพนามอ้างอิงเฉพาะ

3. ความลำเอียงจากโครงสร้างและรูปแบบของข้อสอบ (item structure and format bias) สังเกตจาก

3.1 มีข้อสอบบางข้อซึ่งทำให้เกิดความไขว้เขวแก่ผู้เข้าสอบ เช่นมีคำที่มีความแตกต่างกันเมื่อเปรียบเทียบกับคำในภาษาแม่ของผู้สอบบางคน

3.2 มีความบกพร่องเกี่ยวกับข้อสอบซึ่งผู้สอบสอบสามารถรู้สึกได้เมื่อทำข้อสอบ

3.3 ข้อสอบประกอบด้วยตัวเลือกที่คลาดเคลื่อนหรือความไม่คงที่ในการตอบหรือสามรถตอบได้หลายข้อ

3.4 มีการอธิบายในเรื่องที่ต้องการให้สำเร็จในข้อสอบซึ่งเอนเอียงและสร้างความยุ่งยากในการสอบสำหรับผู้ที่เข้าสอบบางกลุ่ม

3.5 เนื้อหาที่อยู่ในแบบสอบอำนวยความสะดวกให้กับกลุ่มใดกลุ่มหนึ่งเท่านั้น

3.6 มีคำว่า “ฉันไม่ทราบ” ประกอบอยู่ในตัวเลือกในข้อสอบเพื่อตัดโอกาสในการเดาของผู้เข้าสอบ

3.7 มีรูปแบบหรือโครงสร้างของข้อสอบที่สร้างปัญหาให้แก่ผู้สอบที่มีพื้นฐานที่ต่างกัน

3.8 มีการใช้ข้อสอบที่เป็นลบ (negative) ในแบบสอบซึ่งเป็นสาเหตุให้เกิดความแตกต่างในการปฏิบัติ

3.9 มีความกำกวมในโครงสร้างของแบบสอบ ข้อคำถามและการเฉลย

นอกจากนี้ Taylor และ Orlando พบว่า ยังมีความลำเอียงจากสาเหตุอื่น ๆ อีก เช่น ความลำเอียงจากเพศ (sex bias) ความลำเอียงจากวัฒนธรรม (cultural bias) ความลำเอียงจากเชื้อชาติ (ethnic bias) ความลำเอียงจากศาสนา (religious bias) และความลำเอียงจากระดับชั้นเรียน (class bias) เป็นต้น

ดังนั้นผู้วิจัยขอสรุปสาเหตุการทำหน้าที่ต่างกันของข้อสอบหรือความลำเอียงดังต่อไปนี้

1. ลำเอียงทางด้านภาษา เกิดจากความเข้าใจไม่ตรงกันหรือการใช้ภาษาที่ไม่เหมาะสมของผู้ออกข้อสอบ ภาษาที่ออกข้อสอบเข้าใจเฉพาะกลุ่ม หรือคำศัพท์ยาก

2. ลำเอียงจากโครสร้างและรูปแบบข้อสอบ มีความบกพร่องเกี่ยวกับข้อสอบ เนื้อหาในข้อสอบอำนวยความสะดวกให้กับกลุ่มผู้สอบกลุ่มใดกลุ่มหนึ่งเท่านั้น

3. ลำเอียงจากเนื้อหาของแบบสอบ เนื้อหาที่มีความแตกต่างกันหรือไม่คุ้นเคย มีความคิดเห็นในการตอบถูกหรือผิดที่ต่างกัน เนื้อหาของข้อสอบมีการเอื้อให้กับผู้สอบบางกลุ่ม

4. ความแตกต่างด้านวัฒนธรรม มีความเข้าใจหรือแปลเรื่องไปทางเกี่ยวกับวัฒนธรรมของตนจนทำให้เกิดการเข้าใจผิด

## 3. งานวิจัยที่เกี่ยวข้อง

วริษยา ชะม้อย (2550) ได้ทำการศึกษาการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาคระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2550 โรงเรียนมัธยมศึกษา สังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กรุงเทพมหานคร เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนรู้วิชาคณิตศาสตร์ 5 ตัวเลือกที่มีคำตอบถูกหลายคำตอบและให้ตัดสินทุกคำตอบ (Multiple True-false Answer) เงื่อนไขที่ศึกษาได้แก่ (1) ความยาวของแบบสอบ มี 2 ระดับ คือ แบบสอบที่ตรวจให้คะแนนหลายค่าที่มีจำนวน 30 ข้อ และ 40 ข้อ และ (2) ขนาดกลุ่มตัวอย่างมี 3 ระดับ คือ 200 คน 500 คน และ 1,000 คน ผลการวิจัยพบว่า 1) เมื่อเปรียบเทียบอัตราความคลาดเคลื่อนประเภท 1 พบว่าทั้ง 6 เงื่อนไขของการทดสอบ วิธี

LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 0 ถึง ร้อยละ 15.6 และวิธี MACS มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 8.3 ถึงร้อยละ 21.9 โดยที่วิธี LDFA เกือบทุกเงื่อนไขมีค่าอยู่ในเกณฑ์ที่กำหนด ส่วนวิธี MACS พบว่าส่วนใหญ่มีค่าเกินกว่าเกณฑ์ที่กำหนด และพบว่าทั้งวิธี LDFA ทั้งวิธี LDFA และวิธี MACS จะมีอัตราความคลาดเคลื่อนประเภทที่ 1 ในเงื่อนไขความยาวแบบสอบ 30 ข้อต่ำกว่าเงื่อนไขความยาวแบบสอบ 40 ข้อต่ำกว่าเงื่อนไข 2) เมื่อเปรียบเทียบอำนาจการทดสอบ พบว่า ทั้ง 6 เงื่อนไขของแบบทดสอบ วิธี LDFA มีอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 62.5 ถึงร้อยละ 100 และวิธี MACS จะมีอำนาจการทดสอบในเงื่อนไขความยาวแบบสอบ 30 ข้อสูงกว่าเงื่อนไขความยาวแบบ 40 ข้อทุกเงื่อนไข

อิทธิฤทธิ์ พงษ์ปิยะรัตน์ (2551) ศึกษาการวิเคราะห์ข้อสอบและการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยการวิเคราะห์พหุระดับ โดยประยุกต์ใช้โปรแกรมโมเดลเชิงเส้นตรงระดับลดหลั่น(HLM) และโปรแกรม BILOG-MG ซึ่งกลุ่มตัวอย่างที่ใช้ นักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 1588 คน ผู้บริหาร จำนวน 32 คน โดยใช้เครื่องมือ แบบทดสอบวัดความรู้วิชาคณิตศาสตร์ แบบสอบถามคุณลักษณะของผู้เรียน แบบวัดเจตคติต่อการเรียนวิชาคณิตศาสตร์ แบบวัดความเครียดของนักเรียน และแบบประเมินความเป็นผู้นำทางวิชาการของผู้บริหาร ผลการศึกษาพบว่า ระดับนักเรียน(Level 2) ตัวแปรผลการเรียนวิชาคณิตศาสตร์ในภาคเรียนที่ผ่านมาส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบได้ถูกต้องอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และระดับโรงเรียน (Level 3) พบว่าตัวแปรขนาดของโรงเรียนและตัวแปรความเป็นนำทางด้านวิชาการของผู้บริหารส่งผลต่อค่าเฉลี่ยของโอกาสในการตอบข้อสอบได้ถูกต้องอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างโมเดล HGLM กับการวิเคราะห์จากโปรแกรม BILOG – MG มีความสอดคล้องตรงกัน

จันทนา เปรมฤดีปรีชาชาญ (2551) การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบระหว่างวิธีการของแมนเทล-แฮนส์เซล วิธีการทดสอบไค-สแควร์ของ ลอร์ด และวิธี Closed Interval Area ซึ่งมี 2 วิธีย่อย คือ วิธี Closed – Interval Signed Area และวิธี Closed – Interval Unsigned Area กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2549 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) จำนวน 1,533 คน เครื่องมือที่ใช้ในการศึกษา เป็นแบบทดสอบโจทย์ปัญหากลุ่มสาระการเรียนรู้วิชาคณิตศาสตร์ ตามหลักสูตรขั้นพื้นฐาน พุทธศักราช 2544 ผลการวิจัยพบว่า 1) จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ได้จากการตรวจสอบ ระหว่าง แมนเทล-แฮนส์เซล

วิธีการทดสอบไค-สแควร์ของ ลอร์ด และ วิธี Closed Interval Area แตกต่างกันเมื่อขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน ได้แก่ กรณีแบบทดสอบ 40 ข้อ ที่ใช้กลุ่มตัวอย่างขนาด 600คน และ 750 คนและแบบทดสอบ 60 ข้อ ที่ใช้กลุ่มตัวอย่างขนาด 600 คน และ 750 คน 2) ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อพิจารณาจากกลุ่มตัวอย่างขนาด 600 คน กับ 750 คน วิธีการตรวจสอบเหมือนกัน แบบทดสอบจำนวน 40 ข้อพบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด คือ วิธี แมนเทล-แฮนส์เซล แบบทดสอบจำนวน 60 ข้อ พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด คือ วิธี Closed – Interval Unsigned Area

อมลณัฐ อุบลรัตน์ (2553) ได้ศึกษาการทำหน้าที่ต่างกันด้วยวิธีถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล สำหรับข้อสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นมัธยมศึกษาปีที่ 2 สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2550 ในโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 จำนวน 1,846 คน ได้มาโดยการสุ่มแบบแบ่งชั้นภูมิ แบบไม่กำหนดสัดส่วน เครื่องมือที่ใช้ในการศึกษาครั้งนี้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนระดับเขตพื้นที่การศึกษา สำนักงานเขตพื้นที่การศึกษาปัตตานี เขต 2 วิชาภาษาไทยและวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เป็นแบบข้อสอบ 4 ตัวเลือก จำนวน 39 ข้อ ผลการวิจัยพบว่าการทำหน้าที่ต่างกันของข้อสอบวิชาคณิตศาสตร์และภาษาไทย ด้วยวิธีถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์เซล พบว่า 1) เพศเป็นตัวแปรที่มีผลต่อการทำหน้าที่ต่างกันของข้อสอบ ในวิชาคณิตศาสตร์ข้อสอบที่ทำหน้าที่ต่างกันส่วนใหญ่จะลำเอียงเข้าข้างกลุ่มเพศชายมากกว่ากลุ่มเพศหญิง ในกลุ่มภาษาที่ใช้ในชีวิตประจำวันจะพบว่าข้อสอบที่ทำหน้าที่ต่างกันทั้งในวิชาคณิตศาสตร์และวิชาภาษาไทย ส่วนใหญ่จะเข้าข้างกลุ่มผู้สอบที่ใช้ภาษาไทยในชีวิตประจำวันมากกว่ากลุ่มผู้สอบที่ใช้ภาษามลายูถิ่นในชีวิตประจำวัน 2) การเปรียบเทียบจำนวนข้อสอบของแบบทดสอบที่ทำหน้าที่ต่างกันในวิชาคณิตศาสตร์และวิชาภาษาไทย เมื่อจำแนกตามกลุ่มเพศและกลุ่มภาษาที่ใช้ในชีวิตประจำวัน พบว่าวิธีแมนเทล-แฮนส์เซล กับวิธีถดถอยโลจิสติก ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน แตกต่างกันโดยที่วิธีแมนเทล-แฮนส์เซล จะให้ผลจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากกว่าวิธีถดถอยโลจิสติก ในเกือบทุกกลุ่ม ยกเว้นวิชาคณิตศาสตร์กลุ่มภาษาที่ใช้ในชีวิตประจำวันที่วิธีถดถอยโลจิสติกพบมากกว่า 3 ข้อ นั้นแสดงว่าการตรวจสอบการทำหน้าที่ต่างกันวิธีแมนเทล-แฮนส์เซลเป็นวิธีที่มีความไวในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันมากกว่าวิธีถดถอยโลจิสติก 3) ค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีการทำหน้าที่ต่างกันของข้อสอบจำแนกตามกลุ่มเพศและกลุ่มภาษาที่ใช้ในชีวิตประจำวันระหว่างวิธีถดถอยโลจิสติกกับวิธีแมนเทล-แฮนส์เซล ในวิชาคณิตศาสตร์มีความสัมพันธ์ทางบวก

และในวิชาภาษาไทยไม่มีความสัมพันธ์กัน ทั้งนี้อาจเป็นเพราะวิชาภาษาไทยที่ใช้ข้อสอบในครั้งนี้มีค่าความยากและอำนาจจำแนกสูง ทำให้การวิเคราะห์ด้วยวิธีถดถอยโลจิสติก มีประสิทธิภาพต่ำ

เรืองเดช ศิริกิจ (2554) ได้ศึกษาการวิเคราะห์เปรียบเทียบโมเดลการประเมินคุณภาพการจัดการศึกษาวิชาคณิตศาสตร์โดยการประยุกต์ใช้โมเดลมูลค่าเพิ่มที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและการวิเคราะห์การทำหน้าที่ต่างกันของตัวลวง ข้อมูลที่ใช้ในการศึกษาเป็นโครงการการศึกษาแนวโน้มการจัดการศึกษาคณิตศาสตร์หรือ TIMSS ปี 2007 การวิเคราะห์ข้อมูลใช้วิธีการวิเคราะห์แมนเทล – แฮนเซล ร่วมกับลอกของอัตราส่วนแต้มต่อและการวิเคราะห์มูลค่าเพิ่ม โดยใช้โปรแกรมโมเดลเชิงเส้นระดับลดหลั่น ผลการศึกษาพบว่า ข้อสอบที่ทำหน้าที่ต่างกันจะพบในตัวแปรเพศมากที่สุด รองลงมาคือการศึกษาของผู้ปกครอง ความมั่นคงของครอบครัว ตามลำดับ ส่วนการทำหน้าที่ต่างกันของตัวลวงพบว่ามีความสอดคล้องกับการทำหน้าที่ต่างกันของข้อสอบคือส่วนใหญ่กลุ่มเปรียบเทียบ มีโอกาสที่จะเลือกตอบตัวเลือกที่เป็นตัวลวงมากกว่ากลุ่มอ้างอิง และผลของข้อสอบที่ทำหน้าที่ต่างกันในแบบสอบประเมินความรู้วิชาคณิตศาสตร์ เมื่อใช้โมเดลการประเมินคุณภาพการจัดการศึกษาที่ต่างกัน พบว่ามีความสอดคล้องกันระดับคุณภาพการจัดการศึกษาของสถานศึกษาและจัดอันดับคุณภาพการศึกษาของสถานศึกษาไม่แตกต่างกัน แสดงให้เห็นว่าการทำหน้าที่ต่างกันของข้อสอบ และการทำหน้าที่ต่างกันของตัวลวง ในแบบสอบวิชาคณิตศาสตร์ไม่มีผลต่อการประเมินคุณภาพการจัดการศึกษา

ธเกียรติกมล ทองงอก (2554) ได้ศึกษาประสิทธิภาพการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในวิธีถดถอยโลจิสติกโดยใช้เกณฑ์ขนาดอิทธิพล 2 วิธี สำหรับข้อสอบที่มีรูปแบบการตรวจให้คะแนนแบบทวิภาค : ข้อมูลจำลองและข้อมูลเชิงประจักษ์ ในการศึกษาครั้งนี้ได้จำลองข้อมูลภายใต้ทฤษฎีการตอบสนองข้อสอบแบบสองพารามิเตอร์ จำลองผลการตอบภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย รวมข้อมูลที่ศึกษาทั้งหมด 24 เงื่อนไข ($2 \times 3 \times 2 \times 2$) คือ รูปแบบของข้อสอบที่ทำหน้าที่ต่างกัน(อเนกรูปและเอกรูป) ขนาดของการทำหน้าที่ต่างกัน ( 0.1,0.2 และ 0.4) จำนวนข้อสอบที่ทำหน้าที่ต่างกัน(ทั้งฉบับคิดเป็นร้อยละ 10 และ 20 )และความยาวของแบบสอบทั้งฉบับ (40 และ 50 ข้อ) ในทุกเงื่อนไขจำลองข้อมูลซ้ำ 25 ครั้ง วิเคราะห์ข้อมูลในแต่ละเงื่อนไข ด้วยวิธีถดถอยโลจิสติระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และเกณฑ์ Zumbo and Thomas ผลการวิจัยพบว่า ผลการเปรียบเทียบอัตราความคลาดเคลื่อนประเภทที่ 1 ในวิธีถดถอยโลจิสติกระหว่างการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl และการวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas พบว่า การวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้อัตราความถูกต้องสูงกว่า การวัดขนาดอิทธิพลตามเกณฑ์ Zumbo and Thomas ในเกือบทุกเงื่อนไขที่เป็นเช่นนี้อาจเนื่องมาจากเกณฑ์ของ Zumbo and Thomas ไม่ได้มีความระเอียดอ่อน

มากนัก ช่วงการแบ่งงขนาดมีความกว้างกว่าเกณฑ์ของ Jodoin and Gierl (2001) ผลของอัตราความคลาดเคลื่อนประเภทที่ 1 ระหว่างผลการตรวจสอบจากการทดสอบระดับนัยสำคัญกับผลการตรวจสอบจากการวัดขนาดอิทธิพลตามเกณฑ์ทั้ง 2 เกณฑ์พบว่า รูปแบบการทำหน้าที่ต่างกันแบบอเนกรูปตามเกณฑ์ Jodoin and Gieel ให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าเกณฑ์ Zumbo and Thomas แต่รูปแบบการทำหน้าที่ต่างกันแบบเอกรูปตามเกณฑ์ Jodoin and Gierl ให้ค่าอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าเกณฑ์ Zumbo and Thomas และการตรวจสอบประสิทธิภาพด้านอัตราความถูกต้องโดยการวัดขนาดอิทธิพลตามเกณฑ์ Jodoin and Gierl ให้ผลประสิทธิภาพที่ดีกว่าการวัดขนาดอิทธิพล ตามเกณฑ์ Zumbo and Thomas โดยมีความถูกต้องสูงเมื่อแบบสอบมีจำนวนข้อสอบมากขึ้น อัตราความคลาดเคลื่อนประเภทที่ 1 ก็จะลดน้อยลง

พนัส จันทร์เปล่ง (2557) ได้ศึกษาการวิเคราะห์เปรียบเทียบโมเดลประเมินคุณภาพการจัดการศึกษาวิทยาศาสตร์ : การประยุกต์ใช้โมเดลมูลค่าเพิ่มพหุระดับที่มีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบ ข้อมูลที่ใช้ในการวิจัยครั้งนี้เป็นข้อมูลทุติยภูมิจากโครงการประเมินผลนักเรียนนานาชาติเกี่ยวกับการรู้เรื่องวิทยาศาสตร์ (PISA) ปี ค.ศ. 2009 จำนวน 9 ฉบับ ของกลุ่มตัวอย่างในประเทศไทย ซึ่งประกอบด้วยนักเรียนจำนวน 4,292 คน จากสถานศึกษา 230 แห่ง ผลการวิจัยพบว่า มีข้อสอบเอนเอียงเข้าข้างนักเรียนหญิงมากกว่านักเรียนชาย 3 ข้อ เอนเอียงเข้าข้างนักเรียนที่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษามากกว่านักเรียนที่ไม่เรียนพิเศษวิทยาศาสตร์นอกสถานศึกษา 1 ข้อ เอนเอียงเข้าข้างนักเรียนนักเรียนที่มีเศรษฐานะของครอบครัวต่ำมากกว่านักเรียนนักเรียนที่มีเศรษฐานะของครอบครัวสูง 2 ข้อ เอนเอียงเข้าข้างนักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูงมากกว่านักเรียนที่มีแหล่งทรัพยากรการศึกษาที่บ้านสูงต่ำ 5 ข้อ และเอนเอียงเข้าข้างนักเรียนที่มีความมั่นคงทางครอบครัวสูงมากกว่านักเรียนที่มีความมั่นคงทางครอบครัวต่ำ 2 ข้อ ค่าความเที่ยงของแบบทดสอบในโมเดลที่ 2 ลดลงเมื่อข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันถูกตัดออก ส่วนผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบพบว่า ขนาดอิทธิพลการทำหน้าที่ต่างกันของแบบสอบ 2 ฉบับ ลดลงหลังจากตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันออกไป ส่วนแบบสอบอีก 7 ฉบับมีขนาดอิทธิพลการทำหน้าที่ต่างกันของแบบสอบไม่เปลี่ยนแปลง และหลังจากตัดข้อสอบบางข้อที่ตรวจพบการทำหน้าที่ต่างกันออกไปในโมเดลที่ 2 พบว่า ผลการวิเคราะห์องค์ประกอบเชิงยืนยันทั้งแบบสอบรวมและแยกรายฉบับมีความตรงเชิงโครงสร้างที่ยอมรับได้ ผลการศึกษาสัมประสิทธิ์การทำนาย ($R^2$) พบว่า ค่าเฉลี่ยรวมคะแนนการประเมินผลการรู้เรื่องวิทยาศาสตร์ที่มีการควบคุมตัวแปรระดับนักเรียนและสถานศึกษาหลังตัดข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันสามารถอธิบายสัดส่วนความแปรปรวนได้มากกว่าข้อสอบก่อนข้อสอบที่ตรวจพบการทำหน้าที่ต่างกัน และการเปรียบเทียบความสอดคล้องการจัด

อันดับคุณภาพของโรงเรียนจากคะแนนมูลค่าเพิ่มโดยสถิติทดสอบ wilcoxon signed ranks พบว่า โมเดลที่ 1และโมเดลที่ 2 มีการจัดอันดับคุณภาพคะแนนมูลค่าเพิ่มไม่สอดคล้องกัน ส่วนผลการเปรียบเทียบความสอดคล้องของการจัดกลุ่มคุณภาพโรงเรียน 5 กลุ่มตามคะแนนมูลค่าเพิ่ม พบว่า โมเดลที่ 1 กับโมเดลที่ 2 มีการจัดกลุ่มคุณภาพโรงเรียนที่สอดคล้องกันอย่างมีนัยสำคัญทางสถิติ

สุชาติ สิริมีนนันท์ (2554) ได้ศึกษาการเปรียบเทียบวิธีโพลีโตมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า ในการศึกษาครั้งนี้ใช้ข้อมูลจำลองโดยจำลองภายใต้โมเดลพาร์เชียลเครดิตทั่วไปแบบมิติเดี่ยวของมูรากิ ซึ่งแต่ละข้อจะมีรายรายการตอบ 5 รายการ ให้คะแนนเป็น 1,2,3,4 หรือ 5 คะแนน โดยศึกษาภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน 2 รูปแบบ ความยาวของแบบสอบ 3 ขนาด สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน 3 ขนาด และตัวอย่าง 3 ขนาด รวมข้อมูลที่ใช้ในการศึกษาทั้งหมด 54 เงื่อนไข ในแต่ละเงื่อนไขจำลองซ้ำ 50 ครั้ง ผลการศึกษาพบว่า อัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าของวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก สูงกว่าวิธีโพลีโตมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับในทุกเงื่อนไขและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกต่ำกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับในทุกเงื่อนไข และโดยภาพรวมวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสูงกว่าวิธีโพลีโตมัสซิบเทสท์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ

Swaminathan and Rogers (1990) ได้เปรียบเทียบวิธีถดถอยโลจิสติกกับวิธีแมลเทลขแฮนส์เซล กลุ่มตัวอย่างขนาด 250 และ 500 คน ความยาวของข้อสอบ 40, 60, 80 ข้อ ผลการวิจัยพบว่าวิธีถดถอยโลจิสติกและวิธีแมนเทล-แฮนส์ดเซล ให้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบเอกรูปได้ถูกต้องร้อยละ 70 กรณีกลุ่มตัวอย่าง 250 คน ตรวจสอบได้ร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน ในทุกความยาวของแบบสอบ สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอเนกรูป พบว่าวิธีแมนเทล-แฮนส์เซลสามารถตรวจสอบได้เล็กน้อย ส่วนวิธีถดถอยโลจิสติกสามารถตรวจสอบได้ถูกต้องร้อยละ 50 กรณีกลุ่มตัวอย่างน้อย ข้อสอบสั้นและถูกต้องร้อยละ 75 กรณีแบบสอบยาวและกลุ่มตัวอย่างที่มีขนาดใหญ่

Lei et al. (2006) ได้ศึกษาการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบแบบปรับเหมาะโดยใช้ระบบการทดสอบด้วยคอมพิวเตอร์ในการวิจัยครั้งนี้ได้จำลองข้อมูลจำลองภายใต้

เงื่อนไขอัตราส่วนขนาดตัวอย่างที่แตกต่างกัน ซึ่งเงื่อนไขกลุ่มตัวอย่างที่ 1 จำนวนกลุ่มตัวอย่าง 1,000 คน แบ่งเป็นกลุ่มอ้างอิง 500 คน และกลุ่มเปรียบเทียบ 500 คน และเงื่อนไขกลุ่มตัวอย่างที่ 2 จำนวนกลุ่มตัวอย่าง 1,000 คน เท่ากันแบ่งเป็นกลุ่มอ้างอิง 900 คน และกลุ่มเปรียบเทียบ 100 คน มีการแจกแจงความสามารถที่แตกต่างกัน การแจกแจงความสามารถแบบปกติที่มีค่าเฉลี่ยเท่ากับ 0 และส่วนเบี่ยงเบนมาตรฐานเป็น 1 ประเภทของข้อสอบที่ศึกษามี 3 แบบ คือ ไม่เกิดการทำหน้าที่ต่างกันของข้อสอบ เกิดการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทาง และ เกิดการทำหน้าที่ต่างกันของข้อสอบแบบไม่มีทิศทาง การวิเคราะห์ข้อมูลด้วยวิธีถดถอยโลจิสติก วิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT (IRT likelihood ratio test) และวิธีแคทซิบ ผลการวิจัยพบว่า การวิเคราะห์ข้อมูลด้วยวิธีการถดถอยโลจิสติก และวิธีการทดสอบอัตราส่วนไลด์ลิฮู้ดแบบ IRT สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบมีทิศทางและแบบไม่มีทิศทางได้ดีเท่ากันและทั้งสองวิธีสามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่าวิธีแคทซิบ การควบคุมอัตราความคลาดเคลื่อนของข้อมูลที่ดีที่สุดคือ IRT- LRT)

Marie (2009) ศึกษาการทำหน้าที่ต่างกันของข้อสอบของแบบวัดความสามารถระดับสูง mastery tests เปรียบเทียบของสามวิธีโดยใช้ข้อมูลจริง เพื่อเปรียบเทียบวิธีลอกลิเนียร์โมเดลโลจิสติกรีเกรสชันและวิธีแมนเทลแฮนส์เซลเซลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบวัดความสามารถระดับสูง ข้อมูลที่ใช้ในการวิเคราะห์เป็นผลการสอบจากเครื่องมือ swedish theory driving license test (SDLT) และแบบวัดความสามารถระดับสูง ประกอบด้วยข้อสอบจำนวน 65 ข้อ ในระดับยากผู้เข้าร่วมต้องทำข้อสอบได้อย่างน้อย 52 ข้อขึ้นไปถึงจะผ่านการทดสอบและจากผู้เข้าสอบ 5,404 คนและสุ่มคัดเลือกข้อสอบมา 15 ข้อที่ครอบคลุมหลักสูตรเพื่อนำมาตรวจสอบ DIF สถิติที่ใช้ในการวิเคราะห์ข้อมูล คือ ลอกลิเนียร์โมเดล โลจิสติกรีเกรสชันและวิธีแมนเทลแฮนส์เซล โดยใช้โปรแกรม R package (R-development-core-team, 2007) ผลการวิจัยพบว่า มีความสัมพันธ์กันค่อนข้างสูงเกี่ยวกับขนาดของการทำหน้าที่ต่างกันของข้อสอบ การใช้วิธีแมนเทลแฮนส์เซลให้ผลที่เหมือนกันกับทั้งสองวิธี วิธีโลจิสติกกรีเกรสชันและลอกลิเนียร์โมเดลให้ผลที่สอดคล้องกัน ส่วนวิธีลอกลิเนียร์โมเดลจะมีประโยชน์ในการให้ค่าช่วงคะแนนในการสอบที่แน่นอนกว่า ซึ่งถือเป็นสิ่งที่น่าสนใจเป็นพิเศษในแบบวัดความสามารถระดับสูงนี้ ในการทดสอบคะแนนส่วนนี้วิธีการโลจิสติกกรีเกรสชันและวิธีแมนเทลแฮนส์เซลให้ผลลัพธ์ที่แตกต่างกัน

Jodoin and Gierl (2001) ได้ศึกษาประสิทธิภาพของการวัดขนาดอิทธิพลในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ขนาดของกลุ่มตัวอย่าง 3 ขนาดทั้งกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ คือ 250, 500 และ 1,000 คน วิธีที่ใช้แมนเทลแฮนส์เซล วิธี simultaneous item

bias) และ วิธีถดถอยโรจิสติก พบว่า การวัดขนาดอิทธิพลโดยสถิติ $R^2$ ร่วมกับการทดสอบนัยสำคัญทางสถิติจะได้ค่าที่ลดลงจนเกือบจะเป็นศูนย์ของการสรุปผิดว่าข้อสอบทำหน้าที่ต่างกันของข้อสอบ ทั้งที่ความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน (False Positive; FP) โดยเมื่อข้อสอบยิ่งมีความยากมากขึ้น FP ยิ่งใกล้ศูนย์และในทางกลับกันการทดสอบนัยสำคัญของสถิติเพียงอย่างเดียวจะทำให้ได้ FP สูงกว่าเล็กน้อยหรือใกล้เคียงจากค่าปกติทั่วไป อย่างไรก็ตามการวัดขนาดอิทธิพลโดยสถิติ $R^2$ ให้ผลของอำนาจการทดสอบ (1-$\beta$) ที่ต่ำลงจากการทดสอบนัยสำคัญทางสถิติ ซึ่งผลการวิจัยสนับสนุนให้ศึกษาการวัดขนาดอิทธิพลโดยสถิติ $R^2$ ร่วมกับการทดสอบนัยสำคัญทางสถิติจะทำให้ได้สารสนเทศมากยิ่งขึ้น

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่า การทำหน้าที่ต่างกันของข้อสอบหมายถึง ข้อสอบที่ทำให้ผู้สอบต่างกลุ่มกันและมีความสามารถเท่ากันหรือคุณลักษณะที่ต้องการวัดเท่ากัน แต่มีโอกาสในการตอบข้อสอบได้ถูกต้องต่างกัน และการทำหน้าที่ต่างกันของข้อสอบ มีลักษณะข้อสอบที่ทำหน้าที่ต่างกัน 2 ประเภท คือ ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูปและข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีหลายวิธี สามารถจำแนกได้ดังนี้ คือข้อสอบที่มีการให้คะแนนแบบทวิวิภาคหรือ 2 ค่า คือให้คะแนนแบบ 0 , 1 และข้อสอบที่มีการให้คะแนนแบบพหุวิภาค หรือหลายค่า คือมีการให้คะแนนมากกว่า 2 ค่าในข้อสอบ ในวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแต่ละประเภท ยังสามารถจำแนกได้อีก 2 มิติ ได้แก่ มิติลักษณะของตัวแปรเกณฑ์ซึ่งแบ่งเป็นกลุ่มวิธีที่ใช้คะแนนสังเกตได้และกลุ่มวิธีที่ใช้คะแนนสังเกตไม่ได้หรือคะแนนของตัวแปรแฝง และมิติลักษณะของสถิติวิเคราะห์ ซึ่งแบ่งเป็นกลุ่มวิธีที่ใช้สถิติพาราเมตริก และกลุ่มวิธีที่ใช้สถิตินันพาราเมตริก วิธีการวิธีการตรวจสอบการทำหน้าที่ต่างกันข้อสอบส่วนใหญ่ที่ใช้ในปัจจุบันคือ วิธีแมนเทล-แฮนส์เซล วิธี SIBTEST และวิธีถดถอยโลจิสติก วิธีโพลีโตมัสซิบเทสท์ เป็นต้น

# บทที่ 3
# วิธีดำเนินการวิจัย

การวิจัยเรื่อง การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลระหว่างนักเรียนชายและนักเรียนหญิงจากการวิเคราะห์ผลการทดสอบระดับชาติ มีวัตถุประสงค์ของการวิจัยเพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง และเพื่อวิเคราะห์สาเหตุการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของการทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง โดยใช้ผู้เชี่ยวชาญในการวิเคราะห์ โดยมีขั้นตอนดำเนินการวิจัยดังนี้

1. ประชากร
2. เครื่องมือที่ใช้ในการวิจัย
3. การเก็บรวบรวมข้อมูล
4. การวิเคราะห์ข้อมูล

## 1. ประชากร

ประชากรในการวิจัยครั้งนี้เป็นนักเรียนชั้นประถมศึกษาปีที่ 3 ที่ศึกษาในปีการศึกษา 2557 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานมีนักเรียนทั้งหมด 464,532 คน

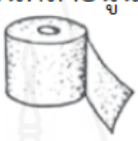
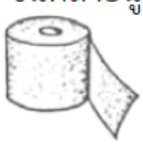
## 2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 27 ข้อ ลักษณะของแบบทดสอบที่สร้างจากสำนักทางการศึกษา สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน มีลักษณะดังนี้

**2.1 ตัวชี้วัด**

ตัวชี้วัดที่ 1 มีความเข้าใจในข้อมูล สถานการณ์หรือสารสนเทศทางวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์และด้านการดำรงชีวิตอย่างมีเหตุผลมีตัวอย่างข้อสอบตัวอย่างเช่น

ผลการสำรวจร้านค้าเกี่ยวกับการใช้กระดาษชำระหรือกระดาษทิชชู

| ร้านค้า | ร้านขายข้าวสาร | ร้านอาหาร | ร้านก๋วยเตี๋ยว | ร้านขายยา |
|---|---|---|---|---|
| ลักษณะกระดาษชำระ (ทิชชู) | ชนิดแผ่นเรียบ | ชนิดลายนูน  | ชนิดลายนูน  | ชนิดแผ่นเรียบ |

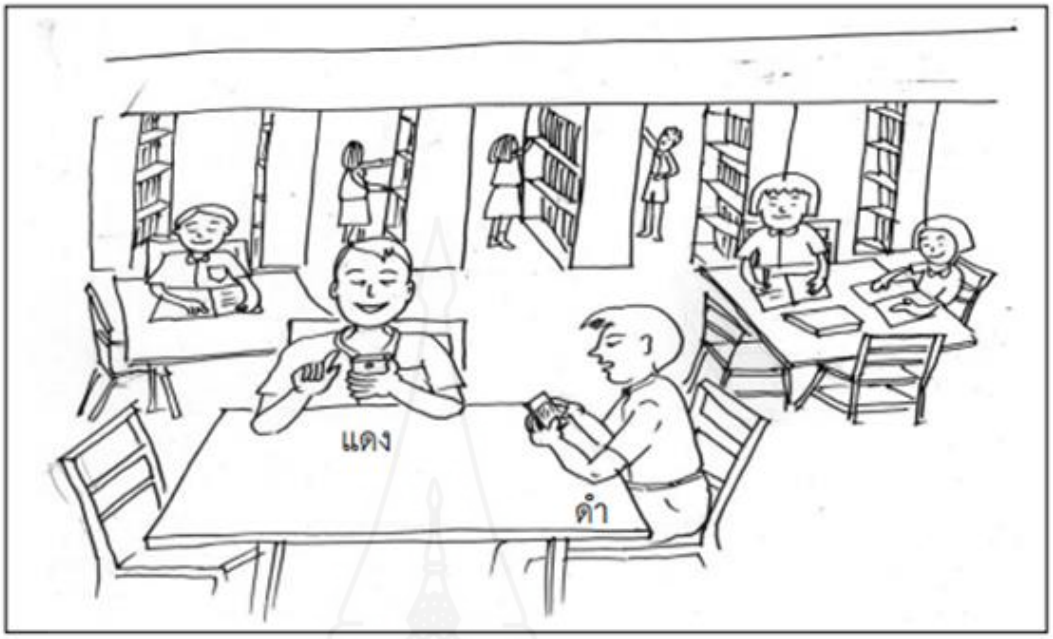
แอ๋วขายส้มตำควรเลือกใช้กระดาษชำระ (ทิชชู) แบบใด จึงจะเหมาะสมเพราะเหตุใด

1) ชนิดลายนูน เพราะมีความยืดหยุ่นและน้ำหนักเบา
2) ชนิดลายนูน เพราะมีความยืดหยุ่นและซึมซับได้ดี
3) ชนิดแผ่นเรียบ เพราะมีความยืดหยุ่นย่อยและสลายได้ดี
4) ชนิดแผ่นเรียบ เพราะมีความยืดหยุ่นและได้ปริมาณมากกว่า

ที่มา : สำนักทดสอบทางการศึกษา (2556, น.59)

ตัวชี้วัดที่ 2 วิเคราะห์ข้อมูลสถานการณ์หรือสารสนเทศ โดยใช้องค์ความรู้ด้านวิทยาศาสตร์ สิ่งแวดล้อมด้านสังคมศาสตร์ เศรษฐศาสตร์ และด้านการดำเนินชีวิตอย่างมีเหตุผล ตัวอย่างเช่น

พิจารณาภาพแล้วตอบคำถาม




ถ้านักเรียนเป็นเพื่อนกับ แดงและดำ คำพูดใดควรใช้มากที่สุด

1) น่าจะเอามาแบ่งให้เพื่อนๆ เล่นบ้าง
2) มีความสุขจริงนะ ขอเล่นบ้างได้ไหม
3) เดี๋ยวฉันจะเก็บเงินเพื่อมาซื้อเล่นบ้าง
4) ที่บ้านของฉันก็มี แต่ฉันจะเล่นในเวลาว่าง

ที่มา : สำนักทดสอบทางการศึกษา (2556, น. 2)

ตัวชี้วัดที่ 3 สามารถสร้างข้อสรุปใหม่ออกแบบ วางแผน บนพื้นฐานของข้อมูลสถานการณ์หรือสารสนเทศที่ผ่านการวิเคราะห์โดยใช้องค์ความรู้ ด้านวิทยาศาสตร์ สิ่งแวดล้อม ด้านสังคมศาสตร์ เศรษฐศาสตร์และด้านการดำเนินชีวิตอย่างมีเหตุผล ตัวอย่างเช่น

พิจารณาภาพ แล้วตอบคำถาม

จากภาพให้ข้อคิดในการดำเนินชีวิตอย่างไร

1) กินดีอยู่ดีชีวีมีสุข
2) กินอยู่อย่างพอดี
3) กินเป็นเรื่องสำคัญ
4) กินวันละนิดชีวิตยืนยาว

ที่มา : แบบทดสอบความสามารถด้านเหตุผล สำนักทดสอบทางการศึกษา( 2556, น. 3)

ตัวชี้วัดที่ 4 สามารถตัดสินใจและแก้ปัญหาอย่างมีหลักการและเหตุผลหรือให้ข้อสนับสนุนข้อโต้แย้งที่สมเหตุสมผล โดยคำนึงถึงคุณธรรมและจริยธรรม ค่านิยม ความเชื่อ ในกรณีที่มีสถานการณ์ที่ต้องการตัดสินใจหรือมีปัญหา ตัวอย่างเช่น

อ่านข้อความแล้วตอบคำถาม

บ้านของกิ่งและแก้วเป็นห้องแถวเล็กๆ ทั้งสองคนนอนและทำการบ้านในห้องเดียวกันกิ่งทำการบ้านเสร็จแล้วปิดไฟนอนก่อนเสมอ ทำให้แก้วซึ่งทำงานเสร็จช้ากว่าต้องนอนและทำให้ไม่มีโอกาสได้ทบทวนบทเรียน

แก้วควรทำตามข้อเสนอใดจึงจะได้ทบทวนบทเรียน

1) พยายามทำงานให้เร็วขึ้น
2) บอกพ่อแม่บังคับให้กิ่งเปิดไฟนอน
3) ให้พ่อสร้างห้องนอนใหม่คนละห้อง
4) ให้กิ่งช่วยทำงานเพื่อปิดไฟนอนพร้อมกัน

ที่มา : แบบทดสอบความสามารถด้านเหตุผล สำนักทดสอบทางการศึกษา( 2556, น. 3)

**2.2 เกณฑ์การตรวจให้คะแนน** นักเรียนตอบถูกให้ 1 คะแนน นักเรียนตอบผิดให้ 0 คะนน

**2.3 การตรวจสอบคุณภาพของแบบสอบ** แบบทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 ผ่านการตรวจสอบคุณภาพด้านความตรงเชิงเนื้อหา ค่าความยาก ค่าอำนาจจำแนกและความเที่ยง จากสำนักทดสอบทางการศึกษาของสำนักงานคณะกรรมกการศึกษาขั้นพื้นฐานเป็นที่เรียบร้อยแล้ว

## 3. การเก็บรวมรวมข้อมูล

การวิจัยครั้งนี้ผู้วิจัยใช้ข้อมูลทุติยภูมิ ซึ่งเป็นผลการตอบข้อสอบความสามารถด้านเหตุผล ของนักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 ของสำนักทดสอบทางการศึกษาของสำนักงานการศึกษาขั้นพื้นฐาน ชนิดเลือกตอบ 4 ตัวเลือก มี 27 ข้อ โดยมีขั้นตอนในการรวบรวมข้อมูล ดังนี้

3.1 ขอหนังสือจากสาขาศึกษาศาสตร์ แขนงวิชาการวัดและประเมินผลการศึกษาของมหาวิทยาลัยสุโขทัยธรรมาธิราช เพื่อขอข้อมูลแบบบันทึกผลการทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จากผู้อำนวยการสำนักทดสอบทางการศึกษาของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

3.2 ผู้วิจัยได้ตรวจสอบความถูกต้องสมบูรณ์ของข้อมูลเพื่อนำไปใช้ในการวิเคราะห์ผลต่อไป

3.3 ข้อมูลที่ใช้ในการวิจัย ได้แก่ คะแนนการสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 464,532 คน

## 4. การจัดเตรียมข้อมูล

**4.1 กำหนดกลุ่มผู้เข้าสอบ** เป็น 2 กลุ่ม

***4.1.1 กลุ่มอ้างอิง*** กำหนดให้เป็นเพศชาย เป็นกลุ่มที่คาดว่าจะได้ประโยชน์จากการตอบข้อสอบได้ถูกต้อง

***4.1.2 กลุ่มเปรียบเทียบ*** กำหนดให้เป็นเพศหญิง เป็นกลุ่มที่เสียประโยชน์ในการตอบข้อสอบ

**4.2 กำหนดค่าของตัวแปรที่ใช้ในการวิเคราะห์** โดยแยกข้อมูลออกเป็น 2 ส่วน

***4.2.1 กลุ่มอ้างอิง*** แทนด้วย 1 จำนวน 224,615 คน

***4.2.2 กลุ่มเปรียบเทียบ*** แทนด้วย 0 จำนวน 239,917 คน

## 5. สถิติที่ใช้ในการวิเคราะห์ข้อมูล

การวิเคราะห์หาคุณภาพของข้อสอบแบบวัดความสามารถด้านเหตุผล ประกอบด้วย

**5.1 สถิติพื้นฐาน** ได้แก่ ค่าเฉลี่ย (mean) ส่วนเบี่ยงเบนมาตรฐาน (standard deviation) ค่าความเบ้ (skewness) ค่าความโด่ง (kurtosis) ค่าความยาก และค่าอำนาจจำแนก (discrimination) เป็นต้น เพื่อให้ทราบการลักษณะของการแจกแจงและการกระจายข้อมูล

***5.1.1 ค่าเฉลี่ย*** โดยใช้สูตร (อนุวัติ คูณแก้ว, 2558, น. 233)

$$\bar{X} = \frac{\sum X}{N}$$

เมื่อ $\bar{X}$ แทน คะแนนเฉลี่ย

$\sum X$ แทน ผลรวมของคะแนนทั้งหมด

$N$ แทน จำนวนนักเรียนทั้งหมด

***5.1.2 ค่าความเบี่ยงเบนมาตรฐาน*** โดยใช้สูตร(อนุวัติ คูณแก้ว, 2558, น. 241)

$$\text{S.D.} = \sqrt{\frac{n\sum X^2-(\sum X)^2}{n(n-1)}}$$

เมื่อ S.D. แทน ความเบี่ยงเบนมาตรฐาน

$\sum X$ แทน ผลรวมของคะแนน

$\sum X^2$ แทน ผลรวมของคะแนนแต่ละตัวกำลังสอง

$n$ แทน จำนวนนักเรียน

***5.1.3 การหาค่าความยาก(p)***

1) สูตรการหาค่าความยาก $p = \frac{R}{N}$

เมื่อ p แทน ค่าความยาก

R แทน จำนวนผู้สอบที่ตอบแต่ละข้อคำถามถูก

N แทน จำนวนผู้สอบทั้งหมด

2) การแปลความหมายค่าความยากมีเกณฑ์ในการพิจารณา (วรรณ์ดี แสงประทีปทอง, 2555, น. 54) ดังนี้

| ค่าความยาก | ความหมาย |
|---|---|
| 0.81-1.00 | ข้อสอบง่ายมาก |
| 0.61-0.80 | ข้อสอบค่อนข้างง่าย |
| 0.41-0.60 | ข้อสอบยากง่ายพอเหมาะ |
| 0.20-0.40 | ข้อสอบค่อนข้างยาก |
| 0.00-0.19 | ข้อสอบยากมาก |

ในการพิจารณาค่าความยากที่เหมาะสมจึงกำหนดว่าควรมีค่าระหว่าง 0.2-0.8

***5.1.4 การหาค่าอำนาจจำแนก***

1) สูตรการหาค่าอำนาจจำแนก ( r ) โดยใช้วิธีของพอยท์ไบซีเรียล($P_{pbis}$) (Ebel and Frisbie 1986, p. 234) ดังนี้

$$r_{pbis} = \frac{M_p - M_q}{S_y} \times \sqrt{pq}$$

เมื่อ $M_p$ แทน ค่าเฉลี่ยของคะแนนรวมของนักเรียนกลุ่มที่ตอบข้อสอบข้อนั้นถูก

$M_q$ แทน ค่าเฉลี่ยของคะแนนรวมของนักเรียนกลุ่มที่ตอบข้อสอบข้อนั้นผิด

$S_y$ แทน ความเบี่ยงเบนมาตรฐานของคะแนนรวมทั้งหมด

p แทน สัดส่วนของนักเรียนที่ตอบข้อสอบข้อนั้นถูก
(ค่าความยากของข้อสอบ)

q แทน สัดส่วนของนักเรียนที่ตอบข้อสอบนั้นผิดซึ่งมีค่าเท่ากับ 1 – p

2) การแปลความหมายค่าอำนาจจำแนกมีเกณฑ์ในการพิจารณา (Ebel and Frisbie 1986, p. 234) ดังนี้

| ค่าอำนาจจำแนก | ความหมาย |
|---|---|
| 0.40 ขึ้นไป | ข้อสอบที่จำแนกได้ดีมาก |
| 0.30-0.39 | ข้อสอบที่จำแนกได้ค่อนข้างดีแต่ถ้าเป็นไปได้ควรปรับปรุงแก้ไข |
| 0.20-0.29 | ข้อสอบที่จำแนกได้บ้าง แต่ควรปรับปรุงแก้ไขก่อนนำไปใช้ |
| ต่ำกว่า 0.19 | ข้อสอบที่จำแนกได้น้อย ควรตัดออกหรือปรับปรุงใหม่ |

**5.2 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ**

***5.2.1 การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ***

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีของถดถอยโลจิสติก เป็นการทดสอบค่านัยสำคัญทางสถิติ ด้วยสถิติไค – สแควร์ การตรวจสอบความแตกต่างใช้สถิติ $\chi^2$ ในงานวิจัยนี้ ผู้วิจัยใช้หลักการของ Bruno and Zumbo (1999, p. 26) โดยใช้โปรแกรม SPSS เป็นการเปรียบเทียบผลการตอบข้อสอบ 2 กลุ่ม กำหนดให้กลุ่มอ้างอิง (R) คือกลุ่มเพศชาย แทนด้วย 1 และกลุ่มเปรียบเทียบ (F) คือกลุ่มเพศหญิง แทนด้วย 0 เป็นตัวแปรตาม และมีการกำหนดวิธีการศึกษาเป็นรูปแบบโมเดลที่เป็นขั้น(step) โดยมีขั้นสำหรับการทดสอบ 3 ขั้น คือ ขั้นที่ 1 เป็นขั้นการกำหนดตัวแปรควบคุมหรือเลือกตัวแปรควบคุม ขั้นที่ 2 เป็นการนำกลุ่มตัวแปรเพศเข้าสู่สมการโลจิสติก และขั้นที่ 3 เป็นการศึกษาปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับเพศชายในสมการ การวิเคราะห์ผลการทำหน้าที่ต่างกันของข้อสอบมีการดำเนินการเป็น 3 ขั้น ดังนี้

1) ขั้นที่ 1 การกำหนดตัวแปรควบคุมหรือเลือกตัวแปรควบคุม (คะแนนรวมของผู้สอบ)เป็นการทดสอบตัวแปรที่ใช้ในการจับคู่ความสามารถ หรือตัวแปรเงื่อนไข โดยนำคะแนนซึ่งเป็นระดับความสามารถของผู้สอบเพศชายและเพศหญิง ในขั้นตอนนี้ให้ถือว่าอิทธิพลของตัวแปรคะแนนรวมมีค่าเท่ากันในทุกรายการตอบ การนำตัวแปรดังกล่าวเข้าไปในสมการการวิเคราะห์ถดถอยโลจิสติก ดังนี้

$$Y = a + b_1 TOT \quad \text{.....(step 1)}$$

โดยที่ Y แทน คะแนนรายข้อ

a แทน ค่าพารามิเตอร์จุดตัด

$b_1$ แทน สัมประสิทธิ์การถดถอยแสดงอิทธิพลต่อคะแนนรวมที่มีผลต่อคะแนนรายข้อ

TOT แทน ระดับความสามารถของผู้เข้าสอบ(คะแนนรวม)

2) ขั้นที่ 2 การนำตัวแปรเพศเข้าสู่สมการโลจิสติก โดยทดสอบกลุ่มผู้สอบเพศชายและเพศหญิง กำหนดรหัสตัวแปรดัมมี(dummy variable) คือ มีผลต่อการตอบถูกของแต่ละข้อ ผู้วิจัยได้กำหนดตัวแปรเพศชาย (1) เป็นกลุ่มอ้างอิง และเพศหญิง (0) เป็นกลุ่มเปรียบเทียบ เพื่อเข้าสู่สมการการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบดังนี้

$$Y = a + b_1TOT + b_2GENDER \quad \text{..... (step 2)}$$

โดยที่ Y แทน คะแนนรายข้อ

a แทน ค่าพารามิเตอร์จุดตัด

TOT แทน ระดับความสามารถของผู้เข้าสอบ(คะแนนรวม)

GENDER แทน เพศ

$b_1$ แทน สัมประสิทธิ์การถดถอยแสดงอิทธิพลต่อคะแนนรวมที่มีผลต่อคะแนนรายข้อ

$b_2$ แทน สัมประสิทธิ์การถดถอยแสดงอิทธิพลต่อเพศ

จากสมการ ข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบหรือลำเอียง เมื่อควบคุมคะแนนรวมมีหลักการพิจารณาดังนี้

(1) กรณีที่ค่าสัมประสิทธิ์ของ $b_2$ มีค่าเป็นบวกแสดงว่าของสอบข้อนั้นผู้ชายมีโอกาสตอบถูกมากกว่าผู้หญิงเมื่อควบคุมคะแนนรวม

(2) กรณีที่ค่าสัมประสิทธิ์ของ $b_2$ มีค่าเป็นลบแสดงว่าข้อสอบข้อนั้นผู้หญิงมีโอกาสตอบถูกมากกว่าผู้ชายเมื่อควบคุมคะแนนรวม

(3) ถ้า $b_2$ ไม่ sig แสดงว่าผู้หญิงกับผู้ชายมีโอกาสตอบถูกไม่แตกต่างกันนั้นหมายถึงข้อสอบข้อนั้นมีความยุติธรรมกับผู้สอบ

3) ขั้นที่ 3 เป็นการศึกษาปฏิสัมพันธ์ โดยการทดสอบตัวแปรปฏิสัมพันธ์ระหว่างคะแนนรวมของผู้สอบกับเพศ ในขั้นนี้เพิ่มพจน์ทั้งสองในขั้นที่ 2 เพื่อตรวจสอบการทำหน้าที่ต่างกันข้อสอบว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันแบบเอกรูปหรืออเนกรูป มีสมการการวิเคราะห์ถดถอยดังนี้

$Y = a + b_1TOT + b_2GENDER + b_3(GENDER*TOT)$ .....(step 3)

โดยที่ Y แทน คะแนนรายข้อ

a แทน ค่าพารามิเตอร์จุดตัด

TOT แทน ความสามารถของผู้สอบ(คะแนนรวม)

GENDER แทน เพศ

GENDER*TOT แทน ปฏิสัมพันธ์ระหว่างความสามารถของผู้สอบกับเพศ

$b_1$ แทน สัมประสิทธิ์การถดถอยแสดงอิทธิพลต่อคะแนนรวมที่มีผลต่อคะแนนรายข้อ

$b_2$ แทน สัมประสิทธิ์การถดถอยแสดงอิทธิพลต่อเพศ

$b_3$ แทน สัมประสิทธิ์การถดถอยแสดงอิทธิพลต่อปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบ(คะแนนรวม)กับเพศ

### *5.2.2 การตรวจสอบข้อสอบทำหน้าที่ต่างกัน*

โดยการนำค่า $\chi^2$ ที่ได้จากการคำนวณ เป็นรายข้อขั้นที่ 2 ลบด้วยค่า $\chi^2$ ขั้นที่ 1 ดังนี้

$$\chi^2_2 - \chi^2_1$$

นำผลที่ได้มาแปลผล โดยเปรียบเทียบกับค่า $\chi^2$ ในตาราง

ถ้า $\Delta\chi^2_{\text{คำนวณ}} < \Delta\chi^2_{\text{ตาราง}}$ ยอมรับ ขั้นที่ 1 หมายความว่า ค่าไค-สแควร์ ที่คำนวณได้เมื่อนำไปเทียบกับค่าไค-สแควร์ในตาราง ที่ระดับ .05 มีค่าน้อยกว่า 3.84 แสดงว่าข้อสอบข้อนั้นไม่ทำหน้าที่ต่างกัน

ถ้า $\Delta\chi^2_{\text{คำนวณ}} > \Delta\chi^2_{\text{ตาราง}}$ ยอมรับ ขั้นที่ 2 หมายความว่า ค่าไค-สแควร์ ที่คำนวณได้เมื่อนำไปเทียบกับค่าไค-สแควร์ในตาราง ที่ระดับ .05 มีค่ามากกว่า 3.84 แสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกัน

### *5.2.3 หลักการจำแนกประเภทของการทำหน้าที่ต่างกันของข้อสอบ*

พิจารณาตามหลักการ(Bruno & Zumbo, 1999, p. 27) ดังต่อไปนี้

1) การทำหน้าที่ต่างกันแบบเอกรูป พิจารณาจากการเปรียบเทียบค่า $\chi^2$ ของขั้นที่ 1 กับขั้นที่ 2 ถ้า $\Delta\chi^2_{\text{คำนวณ}} > \Delta\chi^2_{\text{ตาราง}}$ ยอมรับ ขั้นที่ 2 แสดงว่าข้อสอบข้อนั้นเป็นเอกรูป

แต่ถ้า $\Delta\chi^2_{\text{คำนวณ}} < \Delta\chi^2_{\text{ตาราง}}$ ยอมรับ ขั้นที่ 1 แสดงว่าข้อสอบข้อนั้นไม่ทำหน้าที่ต่างกัน คือ ผู้หญิงและผู้ชายมีโอกาสตอบถูกไม่ต่างกัน

2) การทำหน้าที่ต่างกันแบบอเนกรูป พิจารณาจากการเปรียบเทียบค่า $\chi^2$ ของขั้นที่ 2 กับขั้นที่ 3 ถ้า $\Delta\chi^2_{\text{คำนวณ}} > \Delta\chi^2_{\text{ตาราง}}$ ยอมรับ ขั้นที่ 3 แสดงว่าข้อสอบข้อนั้นเป็นอเนกรูป แต่ถ้า $\Delta\chi^2_{\text{คำนวณ}} < \Delta\chi^2_{\text{ตาราง}}$ ยอมรับ ขั้นที่ 2 แสดงว่าข้อสอบข้อนั้นเป็นเอกรูป

***5.2.4 การหาขนาดของการทำหน้าที่ต่างกันของข้อสอบ (effect size)***

เมื่อทราบประเภทของการทำหน้าที่ต่างกันของข้อสอบแล้ว ยังสรุปไม่ได้ว่าข้อสอบข้อนั้นไม่ได้มาตราฐานหรือไม่มีความยุติธรรม จึงต้องทำการหาขนาดขออิทธิพลเป็นรายข้อ โดยดูจากค่า $R^2$ ดังนี้

ขั้นตอนที่ 1 นำค่า $R^2$ ของขั้นที่ 2 ลบด้วยขั้นที่ 1 ซึ่งเป็นผลมาจากอิทธิพลของกลุ่ม(ความแตกต่างของเพศ) ตรวจสอบการทำหน้าที่ต่างกันแบบเอกรูป

$$R^2_{\text{เอกรูป}} = R^2_2 - R^2_1$$

ขั้นตอนที่ 2 นำค่า $R^2$ ของขั้นที่ 3 ลบด้วยขั้นที่ 2 ซึ่งเป็นผลมาจากอิทธิพลของปฏิสัมพันธ์ของเพศและความสามารถของนักเรียน ตรวจสอบการทำหน้าที่ต่างกันแบบอเนกรูป ดังนี้

$$R^2_{\text{อเนกรูป}} = R^2_3 - R^2_2$$

นำผลลัพธ์ค่า $R^2$ ทั้งสองขั้นตอนมาพิจารณาขนาดของการทำหน้าที่ต่างกันของข้อสอบตามหลักเกณฑ์ของซัมโบ

***5.2.5 การพิจารณาขนาดของการทำหน้าที่ต่างกันของข้อสอบ(effect size)***

การหาขนาดของการทำหน้าที่ต่างกันของข้อสอบโดยใช้ค่าสถิติ $R^2$ ซึ่งจะตัดสินข้อสอบที่ทำหน้าที่ต่างกันเมื่อขนาดของการทำหน้าที่ต่างกันของข้อสอบที่มีขนาดปานกลางและขนาดใหญ่ ใช้เกณฑ์ของ Jodoin and Gierl (2001, p. 335) เพื่อตัดสินขนาดของการทำหน้าที่ต่างกัน 3 ระดับ ได้แก่

DIF ขนาดเล็กน้อย เมื่อมีค่าความแตกต่าง $\Delta R^2 < .035$

DIF ขนาดปานกลาง เมื่อมีค่าความแตกต่าง $.035 \leq \Delta R^2 \leq .07$

DIF ขนาดใหญ่ เมื่อมีค่าความแตกต่าง $\Delta R^2 > .26$

***5.2.6 การทำตัวแปรควบคุมให้มีความบริสุทธิ์***

การทำให้ตัวแปรควบคุมให้มีความบริสุทธิ์(purifying the matching variable) คือการควบคุมคะแนนรวมที่ส่งผลต่อตัวแปรที่ศึกษา เพื่อให้คะแนนรวมของผู้สอบเป็นเกณฑ์ที่ดีในการจับคู่สำหรับการวิเคราะห์การทำหน้าที่ต่างกันของผู้สอบ ซึ่งในการศึกษาครั้งนี้

อาจจะมีปัจจัยอื่นที่ไม่ได้สนใจศึกษาส่งผลต่อคะแนนสอบของนักเรียนได้ ดังนั้นการศึกษาครั้งนี้พบข้อสอบที่มีการเอนเอียงไปทางเพศหญิงหรือเพศชาย เพื่อทำให้คะแนนรวมมีเฉพาะข้อสอบที่ดีหรือทำหน้าที่ไม่แตกต่างกัน มีขั้นตอนดังนี้

1) เลือกข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบที่มีขนาดของการทำหน้าที่ต่างกันของข้อสอบ (ขนาด DIF) ขนาดปานกลางและขนาดใหญ่

2) ให้ดำเนินการตัดข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบที่มีขนาดกลางและขนาดใหญ่ออกหรือทิ้ง

3) ดำเนินการคำนวณหาผลรวมของคะแนนสอบใหม่ในข้อที่ไม่ได้ทำหน้าที่ต่างกันหรือไม่ได้ตัดทิ้งใหม่ คะแนนรวมใหม่นี้จะถูกจับคู่สำหรับการวิเคราะห์ในสมการโลจิสติก

4) คำนวณตั้งแต่ขั้นที่ 1 ถึงขั้นที่ 3 อีกครั้ง เพื่อดูว่ายังมีข้อสอบที่ทำหน้าที่ต่างกันอีกหรือไม่ ถ้าเป็นขนาดเล็กถือว่าข้อสอบที่คำนวณใหม่นี้เป็นข้อสอบที่ดีแล้ว

***5.2.7 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจากกราฟ*** มีขั้นตอนดังนี้

1) แบ่งคะแนนของนักเรียนออกเป็น 3 กลุ่ม โดยใช้เกณฑ์ 33 % ในการแบ่งกลุ่มผู้สอบ คือ

(1) กลุ่มสูงคือ ผู้ที่ได้คะแนน 19 – 27

(2) กลุ่มปานกลาง คือผู้ที่สอบได้คะแนน 10 – 18

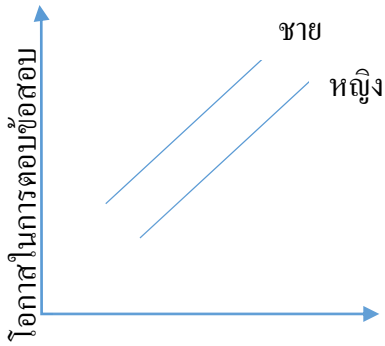
(3) กลุ่มต่ำ คือผู้ที่สอบได้คะแนน 1 – 9

2) ใช้โปรแกรม SPSS ในการวาดกราฟ

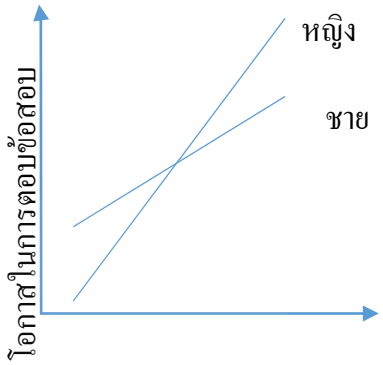
***5.2.8 การพิจารณากราฟในการตัดสินการทำหน้าที่ต่างกันของข้อสอบ***

1) ถ้าพิจารณาโค้งคุณลักษณะของข้อสอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างกลุ่ม เป็นการทำหน้าที่ต่างกันแบบเอกรูป แสดงว่า ข้อสอบข้อนั้นทำให้ผู้สอบกลุ่มหนึ่งมีโอกาสในการตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอกัน ในทุกระดับความสามารถ (ภาพที่ 3.1)

2) ถ้าพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่ามีปฏิสัมพันธ์ร่วมกันระหว่างระดับความสามารถของผู้สอบกับกลุ่มเพศ เช่น ที่ระดับความสามารถหนึ่ง นักเรียนชายมีโอกาสในการตอบข้อสอบถูกมากกว่านักเรียนหญิง แต่ที่ระดับความสามารถอีกระดับหนึ่งนักเรียนหญิงมีโอกาสในการตอบข้อสอบถูกมากกว่านักเรียนชาย เป็นการทำหน้าที่ต่างกันแบบอเนกรูป แสดงว่า ข้อสอบที่ทำให้โอกาสในการตอบข้อสอบถูกของผู้สอบระหว่างกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอกันในทุกระดับความสามารถ (ภาพที่ 3.2)




ภาพที่ 3.1 กลุ่มผู้สอบ




ภาพที่ 3.2 กลุ่มผู้สอบ

**4.5 การวิเคราะห์สาเหตุการทำหน้าที่ต่างกันของข้อสอบ** โดยให้ผู้เชี่ยวชาญด้านเหตุผลวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบว่าเกิดจากสาเหตุใด มีขั้นตอนดังนี้

4.5.1 คัดเลือกผู้เชี่ยวชาญที่จบปริญญาโททางด้านการวัดและประเมินผลและผู้เชี่ยวชาญที่มีประสบการณ์ทางด้านการออกข้อสอบด้านเหตุผลที่ศึกษาจำนวน 1 ท่าน เพื่อทำการวิเคราะห์ข้อสอบ

4.5.2 แต่งตั้งผู้เชี่ยวชาญโดยขอหนังสือจากทางมหาวิทยาลัยสุโขทัยธรรมาธิราช

4.5.3 นำผลการวิเคราะห์ข้อข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบ (DIF) และข้อสอบที่เป็นฉบับจริงที่นักเรียนใช้สอบไปให้กับผู้เชี่ยวชาญ วิเคราะห์สาเหตุหรือเหตุที่ทำให้ข้อสอบเกิดการทำหน้าที่ต่างกันของข้อสอบ

4.5.4 นำผลที่ได้จากผู้เชี่ยวชาญทั้ง 1 ท่านมาสรุปและสังเคราะห์ผลของสาเหตุของการทำหน้าที่ต่างกันของข้อสอบ

# บทที่ 4
# ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีวัตถุประสงค์ (1) วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง (2) วิเคราะห์สาเหตุการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง ในการนำเสนอผลการวิเคราะห์ข้อมูล ผู้วิจัยแบ่งการวิเคราะห์ข้อมูลเป็น 3 ตอน ดังนี้

ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติพื้นฐาน

ตอนที่ 2 ค่าความยากง่ายและค่าอำนาจจำแนกของข้อสอบวัดความสามารถด้านเหตุผลจากการทดสอบระดับชาติ

ตอนที่ 3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลจากการทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

ตอนที่ 4 การแสดงกราฟของการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลจากการทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

เพื่อให้การนำเสนอผลการวิเคราะห์ข้อมูลมีความเข้าใจตรงกันผู้วิจัยได้กำหนดสัญลักษณ์ สมการ และความหมายแทนตัวแปรที่ศึกษาต่าง ๆ ในการวิเคราะห์ข้อมูลและการแปลผลการวิเคราะห์ข้อมูลดังนี้

สมการ ขั้นที่ 1 คือ $Y = a + b_1TOT$

สมการ ขั้นที่ 2 คือ $Y = a + b_1TOT + b_2GENDER$

สมการ ขั้นที่ 3 คือ

$Y = a_1 + b_1TOT + b_2GENDER + b_3(GENDER*TOT)$

| | | |
|---|---|---|
| Y | แทน | คะแนนรายข้อ |
| a | แทน | ค่าพารามิเตอร์จุดตัด |
| TOT | แทน | ความสามารถของผู้สอบ(คะแนนรวม) |
| GENDER | แทน | เพศ |
| GENDER*TOT | แทน | ปฏิสัมพันธ์ระหว่างความสามารถของผู้สอบกับเพศชาย |

| | | |
|---|---|---|
| $b_1$ | แทน | สัมประสิทธิ์การถดถอยแสดงอิทธิพลต่อคะแนนรวมที่มีผลต่อคะแนนรายข้อ |
| $b_2$ | แทน | สัมประสิทธิ์การถดถอยแสดงอิทธิพลต่อเพศ |
| $b_3$ | แทน | สัมประสิทธิ์การถดถอยแสดงอิทธิพลต่อปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบ(คะแนนรวม)กับเพศ |
| $\chi^2$ | แทน | ค่าไค-สแควร์ |
| $\chi^2_1$ | แทน | ค่าไค-สแควร์ ของ ขั้นที่ 1 |
| $\chi^2_2$ | แทน | ค่าไค-สแควร์ ของ ขั้นที่ 2 |
| $\chi^2_3$ | แทน | ค่าไค-สแควร์ ของ ขั้นที่ 3 |
| $R^2$ | แทน | ค่า อา-สแควร์ |
| $R^2_1$ | แทน | ค่า อา-สแควร์ ของ ขั้นที่ 1 |
| $R^2_2$ | แทน | ค่า อา-สแควร์ ของ ขั้นที่ 2 |
| $R^2_3$ | แทน | ค่า อา-สแควร์ ของ ขั้นที่ 3 |

## ตอนที่ 1 ผลการวิเคราะห์ค่าสถิติพื้นฐาน

ผลการวิเคราะห์ข้อมูลจากการทดสอบการวัดความสามารถด้านเหตุผลของนักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 จำนวน 464,532 คน พบว่าผู้ได้สอบ เป็นนักเรียนชาย 224,615 คน และเป็นนักเรียนหญิง 239,917 คน ได้คะแนนมากที่สุด 27 คะแนน คะแนนต่ำสุด 0 คะแนน และมีค่าเฉลี่ยเท่ากับ 12.92 คะแนน ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.16 การแจกแจงของคะแนนมีลักษณะเบ้ขวาและมีความโด่งน้อย (ค่าความเบ้เท่ากับ .159 และ ค่าความโด่งเท่ากับ -.816 )

เมื่อพิจารณาคะแนนและการแจกแจงข้อมูลตามเพศ พบว่าเป็นนักเรียนชาย มีคะแนนต่ำสุด 0 คะแนน และมากที่สุด 27 คะแนน ค่าเฉลี่ยเท่ากับ 11.78 ซึ่งต่ำกว่าค่าเฉลี่ยของนักเรียนทั้งหมด และมีส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.00 การแจกแจงของคะแนนมีลักษณะเบ้ขวาและมีความโด่งน้อย (ค่าความเบ้เท่ากับ .388 และ ค่าความโด่งเท่ากับ -.601) ส่วนนักเรียนหญิงมีคะแนนต่ำสุด 0 คะแนน และมากที่สุด 27 คะแนน ค่าเฉลี่ยเท่ากับ 13.98 และสูงกว่าค่าเฉลี่ยของคะแนนการสอบของนักเรียนทั้งหมด ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 5.09 การแจกแจงของคะแนนมีลักษณะเบ้ซ้าย และความโด่งน้อย (ค่าความเบ้เท่ากับ -.042 และค่าความโด่งเท่ากับ -

.802) ทั้งนี้ลักษณะการแจกแจงคะแนนของนักเรียนชั้นประถมศึกษาปีที่ 3 มีลักษณะเบ้ขวา แสดงดังตารางที่ 4.1

ตารางที่ 4.1 สถิติพื้นฐานของคะแนนจากการตอบข้อสอบของนักเรียนชั้นประถมศึกษาปีที่ 3 จำแนกตามเพศผู้สอบ

| ตัวแปร | จำนวน | ร้อยละ | คะแนนต่ำสุด | คะแนนสูงสุด | ค่าเฉลี่ย | ส่วนเบี่ยงเบนมาตรฐาน | ลักษณะเบ้ | ลักษณะโด่ง |
|---|---|---|---|---|---|---|---|---|
| นักเรียนชาย | 224615 | 48.35 | .00 | 27 | 11.78 | 5.00 | .388 | -.601 |
| นักเรียนหญิง | 239917 | 51.65 | .00 | 27 | 13.98 | 5.09 | -.042 | -.802 |
| รวม | 464532 | 100 | .00 | 27 | 12.92 | 5.16 | .159 | -.816 |

## ตอนที่ 2 ค่าความยากง่ายและค่าอำนาจจำแนกของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติ

ข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติ ของนักเรียนผู้ที่เข้าสอบ 464,532 คน พบว่า ข้อสอบมีคุณภาพรายข้อที่เหมาะสมมีค่าความยากง่ายอยู่ระหว่าง .28 ถึง .65 มีค่าอำนาจจำแนกอยู่ระหว่าง .25 ถึง .54 การหาค่าสัมประสิทธิ์สหสัมพันธ์แบบพอยท์ไบซีเรียล (point biserial) ใช้หลักการพิจารณาของ Ebel and Frisbie (1986, p. 234) ดังตารางที่ 4.2 ดังนี้

ตารางที่ 4.2 ค่าความยากง่ายและค่าอำนาจจำแนกของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติ

| ข้อสอบ | คุณภาพ | | การแปลความหมาย |
|---|---|---|---|
| | ค่าความยากง่าย | ค่าอำนาจจำแนก | |
| 1 | .46 | .16 | ยากง่ายพอเหมาะ,จำแนกได้น้อย |
| 2 | .35 | .18 | ค่อนข้างยาก,จำแนกได้น้อย |

ตารางที่ 4.2 (ต่อ)

| ข้อสอบ | คุณภาพ | | การแปลความหมาย |
|---|---|---|---|
| | ค่าความยากง่าย | ค่าอำนาจจำแนก | |
| 3 | .57 | .36 | ยากง่ายพอเหมาะ,จำแนกได้ค่อนข้างดี |
| 4 | .46 | .39 | ยากง่ายพอเหมาะ, จำแนกได้ค่อนข้างดี |
| 5 | .52 | .33 | ยากง่ายพอเหมาะ, จำแนกได้ค่อนข้างดี |
| 6 | .48 | .27 | ยากง่ายพอเหมาะ, จำแนกได้บ้าง |
| 7 | .61 | .23 | ค่อนข้างง่าย, จำแนกได้บ้าง |
| 8 | .59 | .31 | ยากง่ายพอเหมาะ, จำแนกได้ค่อนข้างดี |
| 9 | .50 | .29 | ยากง่ายพอเหมาะ, จำแนกได้บ้าง |
| 10 | .44 | .29 | ยากง่ายพอเหมาะ, จำแนกได้บ้าง |
| 11 | .55 | .38 | ยากง่ายพอเหมาะ, จำแนกได้ค่อนข้างดี |
| 12 | .65 | .34 | ค่อนข้างง่าย, จำแนกได้ค่อนข้างดี |
| 13 | .43 | .32 | ยากง่ายพอเหมาะ, จำแนกได้ค่อนข้างดี |
| 14 | .38 | .18 | ค่อนข้างยาก, จำแนกได้น้อย |
| 15 | .48 | .32 | ยากง่ายพอเหมาะ,จำแนกได้ค่อนข้างดี |
| 16 | .49 | .24 | ยากง่ายพอเหมาะ,จำแนกได้บ้าง |
| 17 | .38 | .22 | ค่อนข้างยาก,จำแนกได้บ้าง |
| 18 | .45 | .34 | ยากง่ายพอเหมาะ,จำแนกได้ค่อนข้างดี |
| 19 | .46 | .24 | ยากง่ายพอเหมาะ,จำแนกได้บ้าง |
| 20 | .58 | .47 | ยากง่ายพอเหมาะ,จำแนกได้ดีมาก |
| 21 | .38 | .27 | ค่อนข้างยาก, จำแนกค่อนได้บ้าง |
| 22 | .31 | .16 | ค่อนข้างยาก, จำแนกได้น้อย |
| 23 | .49 | .47 | ยากง่ายพอเหมาะ, จำแนกได้ดีมาก |
| 24 | .28 | .19 | ค่อนข้างยาก, จำแนกได้น้อย |
| 25 | .60 | .43 | ยากง่ายพอเหมาะ, จำแนกได้ดีมาก |
| 26 | .56 | .43 | ยากง่ายพอเหมาะ, จำแนกได้ดีมาก |
| 27 | .50 | .42 | ยากง่ายพอเหมาะ, จำแนกได้ดีมาก |

ตารางที่ 4.3 แสดงความสัมพันธ์ระหว่างค่าความยากง่ายกับค่าอำนาจจำแนกของข้อสอบด้านเหตุผลของแบบทดสอบระดับชาติ

| ค่าความยากง่ายของข้อสอบ | ค่าอำนาจจำแนกของข้อสอบ | | | | |
|---|---|---|---|---|---|
| | ความหมาย | ต่ำกว่า 0.19 (จำแนกได้น้อย) | 0.20 – 0.29 (จำแนกได้บ้าง) | 0.30 – 0.39 (จำแนกค่อนข้างดี) | 0.40 ขึ้นไป (จำแนกได้ดีมาก) |
| | 0.00 – 0.19 (ยากมาก) | | | | |
| | 0.20 – 0.40 (ค่อนข้างยาก) | 2,14,22,24 | 17,21 | | |
| | 0.41 – 0.60 ยากง่ายพอเหมาะ | 1 | 6,9,10,16,19 | 3,4,5,8,11,13, 15,18 | 20,23,25,26,27 |
| | 0.61 – 0.80 (ข้อสอบค่อนข้างง่าย) | | 7 | 12 | |
| | 0.81 – 1.00 (ข้อสอบง่ายมาก) | | | | |

เมื่อพิจารณาผลการวิเคราะห์ข้อสอบรายข้อของข้อสอบวัดความสามารถด้านเหตุผลของการทดสอบระดับชาติ ระหว่างนักเรียนชายและนักเรียนหญิง จำนวน 27 ข้อ พบว่านักเรียนชายมีค่าความยากง่ายอยู่ระหว่าง 0.28 ถึง 0.58 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.22 ถึง 0.51 และนักเรียนหญิงมีค่าความยากง่ายอยู่ระหว่าง 0.28 ถึง 0.71 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.26 ถึง 0.55 มีค่าผลต่างความยากง่ายระหว่างนักเรียนชายและนักเรียนหญิงอยู่ระหว่าง .02 ถึง .16 มีค่าผลต่างค่าอำนาจจำแนกอยู่ระหว่าง 0.01 ถึง 0.07 ดังตารางที่ 4.4

ตารางที่ 4.4 ค่าความยากง่ายและค่าอำนาจจำแนกของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติแยกเป็นนักเรียนชายและนักเรียนหญิง

| ข้อที่ | นักเรียนชาย | | นักเรียนหญิง | | ผลต่าง (นักเรียนชาย-นักเรียนหญิง) | |
|---|---|---|---|---|---|---|
| | คุณภาพ | | คุณภาพ | | คุณภาพ | |
| | ค่าความยาก | ค่าอำนาจจำแนก | ค่าความยาก | ค่าอำนาจจำแนก | ค่าความยาก | ค่าอำนาจจำแนก |
| 1 | .48 | .19 | .45 | .15 | .03 | .04 |
| 2 | .33 | .18 | .37 | .18 | .04 | 0 |
| 3 | .52 | .33 | .61 | .36 | .09 | .03 |
| 4 | .40 | .37 | .51 | .39 | .11 | .02 |
| 5 | .49 | .32 | .55 | .32 | .06 | 0 |
| 6 | .44 | .26 | .51 | .26 | .07 | 0 |
| 7 | .57 | .25 | .66 | .19 | .09 | .06 |
| 8 | .58 | .30 | .60 | .31 | .02 | .01 |
| 9 | .46 | .29 | .55 | .27 | .09 | .02 |
| 10 | .43 | .29 | .45 | .29 | .02 | 0 |
| 11 | .50 | .38 | .59 | .36 | .09 | .02 |
| 12 | .58 | .33 | .71 | .31 | .13 | .02 |
| 13 | .35 | .27 | .51 | .32 | .16 | .05 |
| 14 | .34 | .16 | .41 | .18 | .07 | .02 |
| 15 | .44 | .32 | .51 | .30 | .07 | .02 |
| 16 | .45 | .25 | .52 | .21 | .07 | .04 |
| 17 | .36 | .21 | .40 | .23 | .04 | .02 |
| 18 | .40 | .34 | .49 | .32 | .09 | .02 |
| 19 | .42 | .23 | .50 | .24 | .08 | .01 |
| 20 | .50 | .43 | .66 | .48 | .16 | .05 |
| 21 | .35 | .23 | .40 | .30 | .05 | .07 |
| 22 | .28 | .13 | .33 | .17 | .05 | .04 |

ตารางที่ 4.4 (ต่อ)

| ข้อที่ | นักเรียนชาย | | นักเรียนหญิง | | ผลต่าง (นักเรียนชาย-นักเรียนหญิง) | |
|---|---|---|---|---|---|---|
| | คุณภาพ | | คุณภาพ | | คุณภาพ | |
| | ค่าความยาก | ค่าอำนาจจำแนก | ค่าความยาก | ค่าอำนาจจำแนก | ค่าความยาก | ค่าอำนาจจำแนก |
| 23 | .41 | .42 | .57 | .47 | .16 | .05 |
| 24 | .28 | .17 | .28 | .21 | 0 | .04 |
| 25 | .52 | .40 | .67 | .43 | .15 | .03 |
| 26 | .48 | .38 | .63 | .45 | .15 | .07 |
| 27 | .43 | .38 | .56 | .43 | .13 | .05 |

## ตอนที่ 3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติก กระบวนการตรวจสอบความเหมาะสมของข้อมูลกับโมเดลการทดสอบมี 3 ขั้น โดยมีผลการวิเคราะห์การการทดสอบระดับนัยสำคัญของแบบทดสอบวัดความสามารถด้านเหตุผล ของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิงจำนวน 27 ข้อ พบข้อสอบ เมื่อวิเคราะห์ประเภทของการทำหน้าที่ต่างกันของข้อสอบพบข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 2 ข้อ คิดเป็นร้อยละ 7.41 ได้แก่ ข้อ 2 และ ข้อ 10 เมื่อศึกษาค่า $b_2$ .ในขั้นที่ 2 มีค่าเป็นบวก ทั้งสองข้อ ซึ่งหมายความว่าข้อสอบที่มีค่าเป็นบวกมีความลำเอียงเข้าหากลุ่มอ้างอิง กล่าวคือนักเรียนชายมีโอกาสในการตอบข้อสอบถูกมากกว่านักเรียนหญิงอย่างสม่ำเสมอกันในทุกระดับความสามารถ และพบข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป จำนวน 25 ข้อ คิดเป็นร้อยละ 92.59 ได้แก่ 1 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 โดยที่โอกาสในการตอบข้อสอบถูกของนักเรียนทั้งสองกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอในทุกระดับความสามารถ และเมื่อพิจารณาขนาดการทำหน้าที่ต่างกันของข้อสอบพบว่า ทุกข้อมีขนาดของการทำหน้าที่ต่างกันเพียงขนาดเล็ก (small effect size : < 0.13) ดังนั้นเมื่อพิจารณาโดยรวมแล้วข้อสอบทุกข้อทำหน้าที่ไม่แตกต่างกัน ดังปรากฏในตารางที่ 4.5

ตารางที่ 4.5 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

| ข้อ | Step (ขั้น) | a | $b_1$ | $b_2$ | $b_3$ | $\chi^2$ | | | | | $R^2$ | | | | | ขนาดของ DIF | รูปแบบของ DIF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\chi^2_1$ | $\chi^2_2$ | $\chi^2_3$ | $\chi^2_2 - \chi^2_1$ | $\chi^2_3 - \chi^2_2$ | $R^2_1$ | $R^2_2$ | $R^2_3$ | $R^2_2-R^2_1$ | $R^2_3-R^2_2$ | | |
| 1 | 1 | -1.48 | 0.10 | | | 30103.54 | 33193.97 | 33384.64 | 3090.43* | 190.67* | 0.08 | 0.09 | 0.09 | 0.01 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -1.75 | 0.11 | 0.35 | | | | | | | | | | | | | |
| | 3 | -1.63 | 0.10 | 0.12 | 0.02 | | | | | | | | | | | | |
| 2 | 1 | -2.14 | 0.11 | | | 34099.81 | 34243.63 | 34243.67 | 143.82* | 0.04 | 0.10 | 0.10 | 0.10 | 0.00 | 0.00 | เล็ก | เอกรูป |
| | 2 | -2.20 | 0.12 | 0.08 | | | | | | | | | | | | | |
| | 3 | -2.20 | 0.12 | 0.08 | 0.00 | | | | | | | | | | | | |
| 3 | 1 | -2.31 | 0.21 | | | 97945.48 | 97953.84 | 98002.31 | 8.36* | 48.47* | 0.26 | 0.26 | 0.26 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.32 | 0.21 | 0.02 | | | | | | | | | | | | | |
| | 3 | -2.39 | 0.21 | 0.15 | -0.01 | | | | | | | | | | | | |
| 4 | 1 | -3.07 | 0.22 | | | 114054.40 | 114062.03 | 114087.68 | 7.63* | 25.65* | 0.29 | 0.29 | 0.29 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -3.06 | 0.22 | -0.02 | | | | | | | | | | | | | |
| | 3 | -3.11 | 0.23 | 0.08 | -0.01 | | | | | | | | | | | | |

ตารางที่ 4.5 (ต่อ)

| ข้อ | Step (ขั้น) | a | $b_1$ | $b_2$ | $b_3$ | $\chi^2$ | | | | | $R^2$ | | | | | ขนาดของ DIF | รูปแบบของ DIF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\chi^2_1$ | $\chi^2_2$ | $\chi^2_3$ | $\chi^2_2 - \chi^2_1$ | $\chi^2_3 - \chi^2_2$ | $R^2_1$ | $R^2_2$ | $R^2_3$ | $R^2_2-R^2_1$ | $R^2_3-R^2_2$ | | |
| 5 | 1 | -2.27 | 0.18 | | | 84317.22 | 84675.10 | 84686.03 | 357.88* | 10.93* | 0.22 | 0.22 | 0.22 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.37 | 0.19 | 0.13 | | | | | | | | | | | | | |
| | 3 | -2.34 | 0.18 | 0.07 | 0.01 | | | | | | | | | | | | |
| 6 | 1 | -2.07 | 0.15 | | | 62750.84 | 62793.49 | 62802.83 | 42.65* | 9.34* | 0.17 | 0.17 | 0.17 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.11 | 0.15 | 0.04 | | | | | | | | | | | | | |
| | 3 | -2.08 | 0.15 | -0.01 | 0.00 | | | | | | | | | | | | |
| 7 | 1 | -1.30 | 0.14 | | | 50570.66 | 50881.32 | 51520.55 | 310.66* | 639.23* | 0.14 | 0.14 | 0.14 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -1.22 | 0.14 | -0.11 | | | | | | | | | | | | | |
| | 3 | -1.00 | 0.12 | -0.53 | 0.04 | | | | | | | | | | | | |
| 8 | 1 | -1.85 | 0.18 | | | 76070.29 | 78553.93 | 78569.22 | 2483.64* | 15.29* | 0.20 | 0.21 | 0.21 | 0.01 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.11 | 0.19 | 0.33 | | | | | | | | | | | | | |
| | 3 | -2.08 | 0.18 | 0.26 | 0.01 | | | | | | | | | | | | |

ตารางที่ 4.5 (ต่อ)

| ข้อ | Step (ขั้น) | a | $b_1$ | $b_2$ | $b_3$ | $\chi^2$ | | | | | $R^2$ | | | | | ขนาดของ DIF | รูปแบบของ DIF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\chi^2_1$ | $\chi^2_2$ | $\chi^2_3$ | $\chi^2_2 - \chi^2_1$ | $\chi^2_3 - \chi^2_2$ | $R^2_1$ | $R^2_2$ | $R^2_3$ | $R^2_2-R^2_1$ | $R^2_3-R^2_2$ | | |
| 9 | 1 | -2.11 | 0.17 | | | 70675.41 | 70746.79 | 70825.61 | 71.38* | 78.82* | 0.19 | 0.19 | 0.19 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.07 | 0.16 | -0.06 | | | | | | | | | | | | | |
| | 3 | -1.99 | 0.16 | -0.21 | 0.01 | | | | | | | | | | | | |
| 10 | 1 | -2.38 | 0.16 | | | 68394.42 | 70610.27 | 70612.69 | 2215.85* | 2.43 | 0.18 | 0.19 | 0.19 | 0.01 | 0.00 | เล็ก | เอกรูป |
| | 2 | -2.62 | 0.17 | 0.31 | | | | | | | | | | | | | |
| | 3 | -2.64 | 0.17 | 0.34 | 0.00 | | | | | | | | | | | | |
| 11 | 1 | -2.57 | 0.22 | | | 109025.88 | 109061.61 | 109240.23 | 35.73* | 178.62* | 0.28 | 0.28 | 0.28 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.60 | 0.22 | 0.04 | | | | | | | | | | | | | |
| | 3 | -2.47 | 0.21 | -0.21 | 0.02 | | | | | | | | | | | | |
| 12 | 1 | -1.85 | 0.20 | | | 88484.66 | 89675.73 | 89770.77 | 1191.07* | 95.04* | 0.24 | 0.24 | 0.24 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -1.67 | 0.20 | -0.24 | | | | | | | | | | | | | |
| | 3 | -1.58 | 0.19 | -0.41 | 0.02 | | | | | | | | | | | | |

ตารางที่ 4.5 (ต่อ)

| ข้อ | Step (ขั้น) | a | $b_1$ | $b_2$ | $b_3$ | $\chi^2$ | | | | | $R^2$ | | | | | ขนาดของ DIF | รูปแบบของ DIF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\chi^2_1$ | $\chi^2_2$ | $\chi^2_3$ | $\chi^2_2 - \chi^2_1$ | $\chi^2_3 - \chi^2_2$ | $R^2_1$ | $R^2_2$ | $R^2_3$ | $R^2_2-R^2_1$ | $R^2_3-R^2_2$ | | |
| 13 | 1 | -2.63 | 0.18 | | | 80055.20 | 83235.56 | 83507.93 | 3180.36* | 272.37* | 0.21 | 0.22 | 0.22 | 0.01 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.37 | 0.17 | -0.37 | | | | | | | | | | | | | |
| | 3 | -2.53 | 0.18 | -0.06 | -0.02 | | | | | | | | | | | | |
| 14 | 1 | -2.01 | 0.11 | | | 34924.59 | 34954.33 | 34965.19 | 29.74* | 10.86* | 0.10 | 0.10 | 0.10 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -1.99 | 0.11 | -0.04 | | | | | | | | | | | | | |
| | 3 | -2.02 | 0.12 | 0.02 | 0.00 | | | | | | | | | | | | |
| 15 | 1 | -2.40 | 0.18 | | | 79911.45 | 79986.10 | 80122.27 | 74.65* | 136.17* | 0.21 | 0.21 | 0.21 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.44 | 0.18 | 0.06 | | | | | | | | | | | | | |
| | 3 | -2.34 | 0.17 | -0.16 | 0.02 | | | | | | | | | | | | |
| 16 | 1 | -1.84 | 0.14 | | | 52056.72 | 52095.92 | 52426.01 | 39.20* | 330.09* | 0.14 | 0.14 | 0.14 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -1.87 | 0.14 | 0.04 | | | | | | | | | | | | | |
| | 3 | -1.72 | 0.13 | -2.67 | 0.02 | | | | | | | | | | | | |

ตารางที่ 4.5 (ต่อ)

| ข้อ | Step (ขั้น) | a | $b_1$ | $b_2$ | $b_3$ | $\chi^2$ | | | | | $R^2$ | | | | | ขนาดของ DIF | รูปแบบของ DIF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\chi^2_1$ | $\chi^2_2$ | $\chi^2_3$ | $\chi^2_2 - \chi^2_1$ | $\chi^2_3 - \chi^2_2$ | $R^2_1$ | $R^2_2$ | $R^2_3$ | $R^2_2-R^2_1$ | $R^2_3-R^2_2$ | | |
| 17 | 1 | -2.26 | 0.13 | | | 46474.51 | 46866.83 | 46929.37 | 392.32* | 62.54* | 0.13 | 0.13 | 0.13 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.36 | 0.14 | 0.13 | | | | | | | | | | | | | |
| | 3 | -2.43 | 0.14 | 0.27 | -0.01 | | | | | | | | | | | | |
| 18 | 1 | -2.71 | 0.19 | | | 89403.13 | 89408.66 | 89480.32 | 5.53* | 71.66* | 0.23 | 0.23 | 0.23 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.72 | 0.19 | 0.02 | | | | | | | | | | | | | |
| | 3 | -2.64 | 0.19 | -0.15 | 0.01 | | | | | | | | | | | | |
| 19 | 1 | -1.99 | 0.14 | | | 53853.66 | 53863.68 | 53875.08 | 10.02* | 11.40* | 0.15 | 0.15 | 0.15 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -1.97 | 0.14 | -0.02 | | | | | | | | | | | | | |
| | 3 | -2.00 | 0.14 | 0.04 | 0.00 | | | | | | | | | | | | |
| 20 | 1 | -3.24 | 0.29 | | | 159097.04 | 160217.85 | 160678.15 | 1120.81* | 460.30* | 0.39 | 0.39 | 0.39 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -3.06 | 0.28 | -0.24 | | | | | | | | | | | | | |
| | 3 | -3.30 | 0.30 | 0.22 | -0.04 | | | | | | | | | | | | |

ตารางที่ 4.5 (ต่อ)

| ข้อ | Step (ขั้น) | a | $b_1$ | $b_2$ | $b_3$ | $\chi^2$ | | | | | $R^2$ | | | | | ขนาดของ DIF | รูปแบบของ DIF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\chi^2_1$ | $\chi^2_2$ | $\chi^2_3$ | $\chi^2_2 - \chi^2_1$ | $\chi^2_3 - \chi^2_2$ | $R^2_1$ | $R^2_2$ | $R^2_3$ | $R^2_2-R^2_1$ | $R^2_3-R^2_2$ | | |
| 21 | 1 | -2.56 | 0.15 | | | 60769.70 | 61006.68 | 61517.14 | 236.98* | 510.46* | 0.17 | 0.17 | 0.17 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.64 | 0.16 | 0.10 | | | | | | | | | | | | | |
| | 3 | -2.86 | 0.17 | 0.53 | -0.03 | | | | | | | | | | | | |
| 22 | 1 | -2.23 | 0.11 | | | 28328.74 | 28331.60 | 28468.32 | 2.86 | 136.72* | 0.08 | 0.08 | 0.08 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.23 | 0.11 | -0.01 | | | | | | | | | | | | | |
| | 3 | -2.34 | 0.11 | 0.20 | -0.02 | | | | | | | | | | | | |
| 23 | 1 | -3.56 | 0.27 | | | 154813.62 | 156052.22 | 156384.91 | 1238.60* | 332.69* | 0.38 | 0.38 | 0.38 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -3.38 | 0.27 | -0.25 | | | | | | | | | | | | | |
| | 3 | -3.59 | 0.29 | 0.15 | -0.03 | | | | | | | | | | | | |
| 24 | 1 | -2.59 | 0.12 | | | 33966.84 | 35831.51 | 36143.08 | 1864.67* | 311.57* | 0.10 | 0.11 | 0.11 | 0.01 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.83 | 0.13 | 0.30 | | | | | | | | | | | | | |
| | 3 | -3.02 | 0.14 | 0.66 | -0.03 | | | | | | | | | | | | |

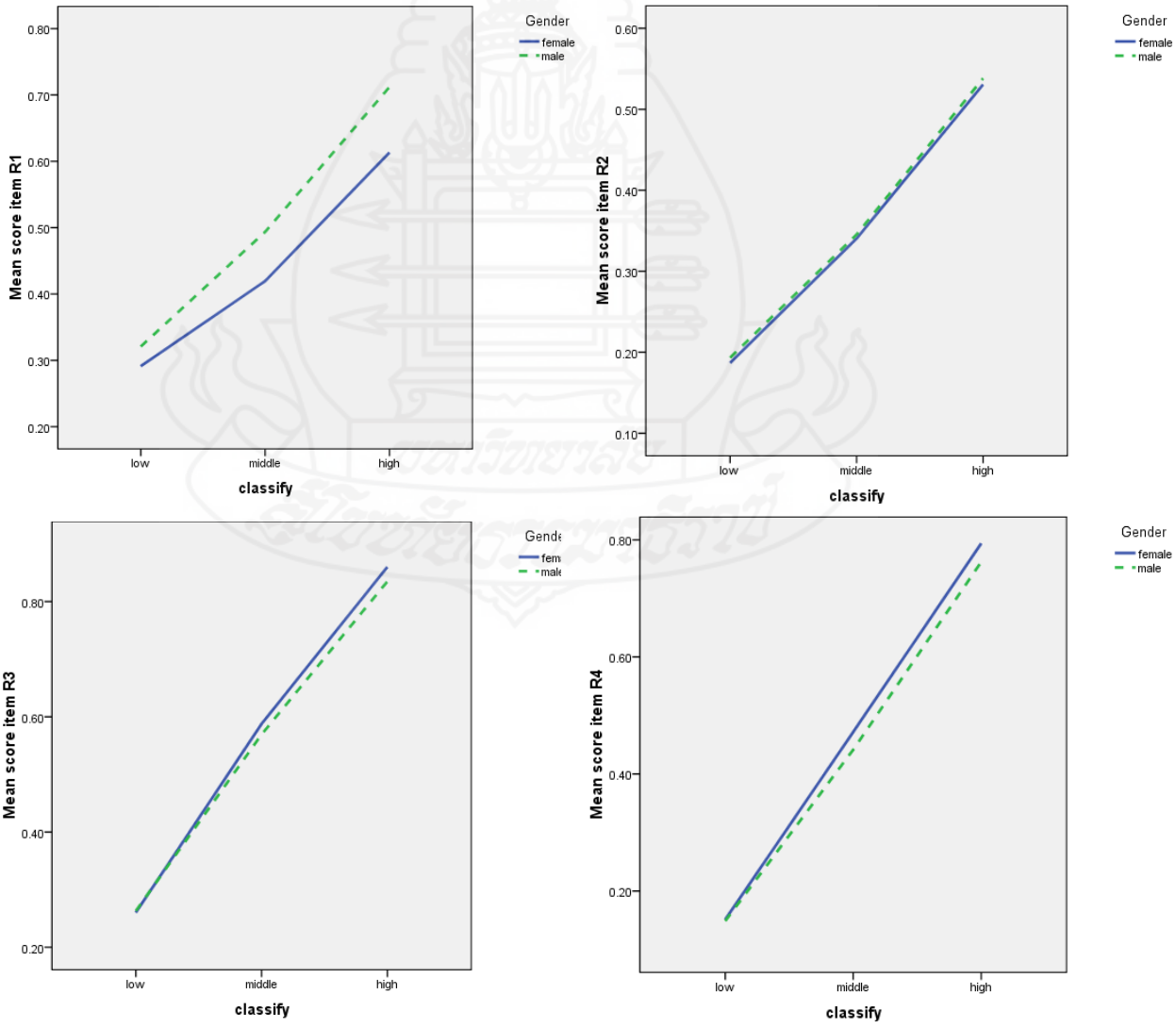
ตารางที่ 4.5 (ต่อ)

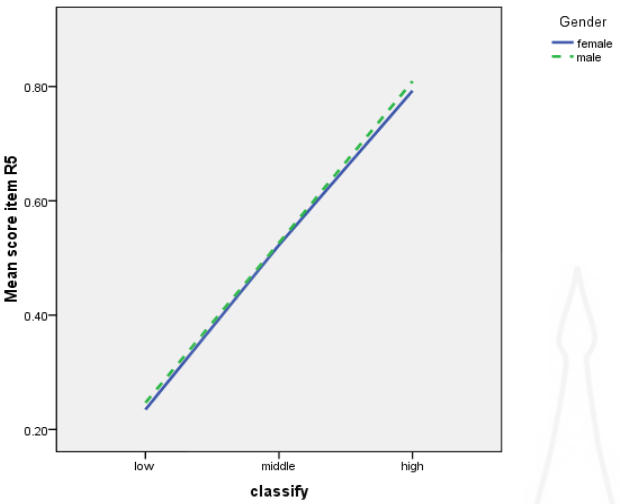
| ข้อ | Step (ขั้น) | a | $b_1$ | $b_2$ | $b_3$ | $\chi^2$ | | | | | $R^2$ | | | | | ขนาดของ DIF | รูปแบบของ DIF |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | $\chi^2_1$ | $\chi^2_2$ | $\chi^2_3$ | $\chi^2_2 - \chi^2_1$ | $\chi^2_3 - \chi^2_2$ | $R^2_1$ | $R^2_2$ | $R^2_3$ | $R^2_2-R^2_1$ | $R^2_3-R^2_2$ | | |
| 25 | 1 | -2.78 | 0.26 | | | 134729.66 | 136258.33 | 136457.61 | 1528.67* | 199.28* | 0.34 | 0.34 | 0.34 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -2.58 | 0.25 | -0.28 | | | | | | | | | | | | | |
| | 3 | -2.73 | 0.27 | 0.01 | -0.02 | | | | | | | | | | | | |
| 26 | 1 | -2.96 | 0.26 | | | 136577.57 | 137770.82 | 138858.47 | 1193.25* | 1087.66* | 0.34 | 0.34 | 0.35 | 0.00 | 0.01 | เล็ก | อเนกรูป |
| | 2 | -2.78 | 0.25 | -0.24 | | | | | | | | | | | | | |
| | 3 | -3.14 | 0.28 | 0.42 | -0.06 | | | | | | | | | | | | |
| 27 | 1 | -3.11 | 0.24 | | | 128516.25 | 128739.62 | 129325.99 | 223.38* | 586.37* | 0.32 | 0.32 | 0.32 | 0.00 | 0.00 | เล็ก | อเนกรูป |
| | 2 | -3.03 | 0.24 | -0.10 | | | | | | | | | | | | | |
| | 3 | -3.29 | 0.26 | 0.39 | -0.04 | | | | | | | | | | | | |

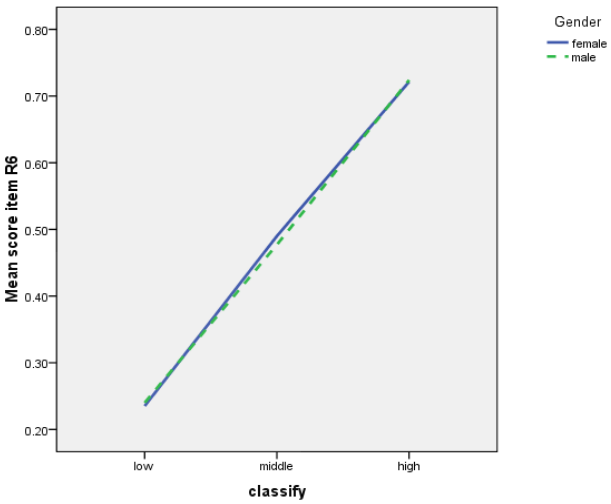
***$\chi^2_{ตาราง}$ = 3.84**

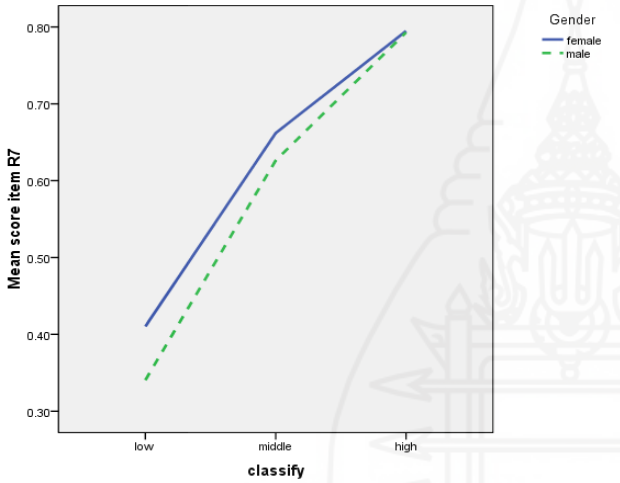
**ตอนที่ 4 การแสดงกราฟของการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง**

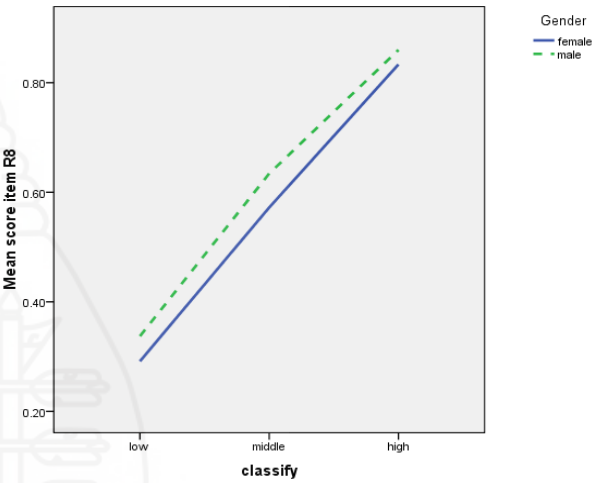
เมื่อพิจารณากราฟแล้วพบว่าข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป คือ 2 และ 10 ทั้งสองข้อ ซึ่งกราฟที่แสดงแต่ละข้อไม่มีปฏิสัมพันธ์ระหว่างกลุ่ม กล่าวคือนักเรียนชายมีโอกาสในการตอบข้อสอบถูกมากกว่านักเรียนหญิงอย่างสม่ำเสมอกันในทุกระดับความสามารถ และเป็นข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป คือข้อ 1 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 กราฟที่แสดงแต่ละข้อพบว่ามีปฏิสัมพันธ์ร่วมกันระหว่างระดับความสามารถของผู้สอบกับกลุ่มเพศ เช่น ที่ระดับความสามารถหนึ่ง นักเรียนชายมีโอกาสในการตอบข้อสอบถูกมากกว่านักเรียนหญิง แต่ที่ระดับความสามารถอีกระดับหนึ่งนักเรียนหญิงมีโอกาสในการตอบข้อสอบถูกมากกว่านักเรียนชายดังแสดงในกราฟการทำหน้าที่ต่างกัน ดังนี้

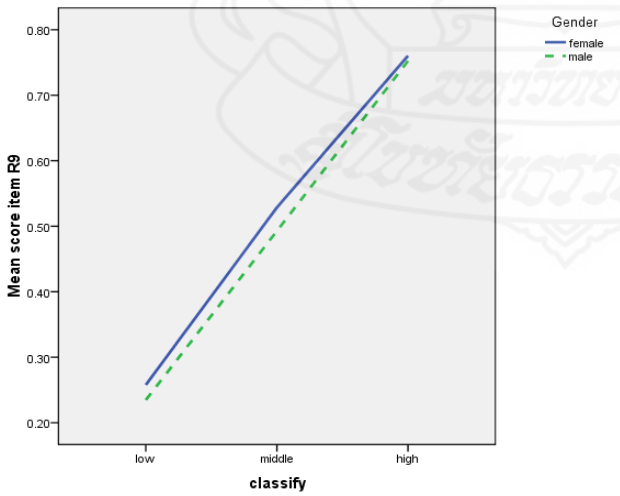


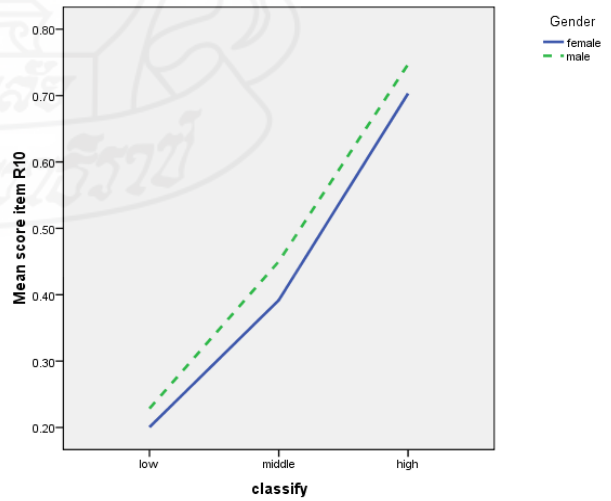

Gender
female
male
Mean score item R5
0.80
0.60
0.40
0.20
low
middle
high
classify

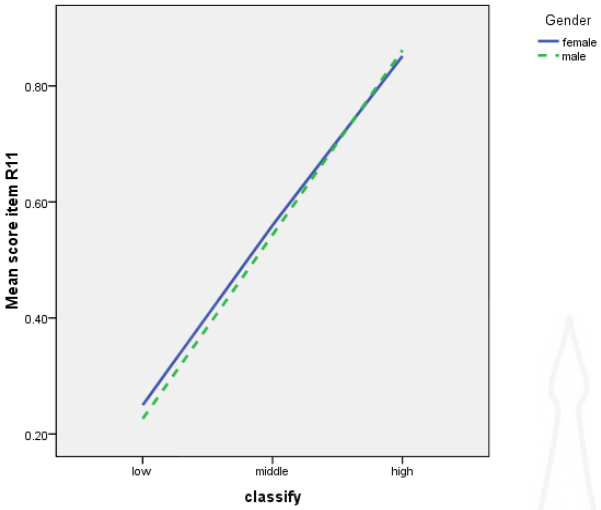

Gender
female
male
Mean score item R6
0.80
0.70
0.60
0.50
0.40
0.30
0.20
low
middle
high
classify

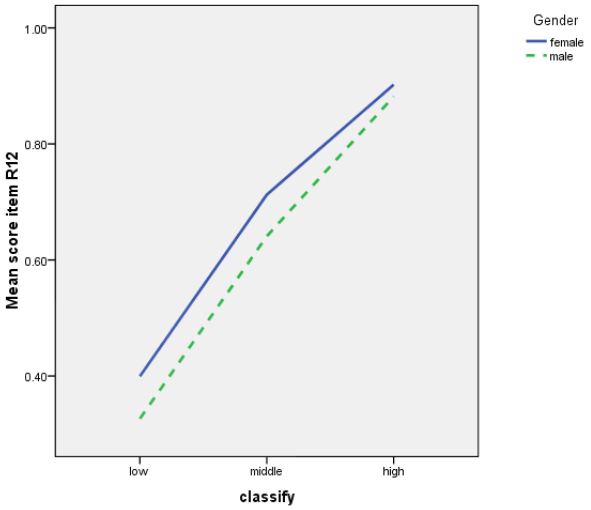

Gender
female
male
Mean score item R7
0.80
0.70
0.60
0.50
0.40
0.30
low
middle
high
classify

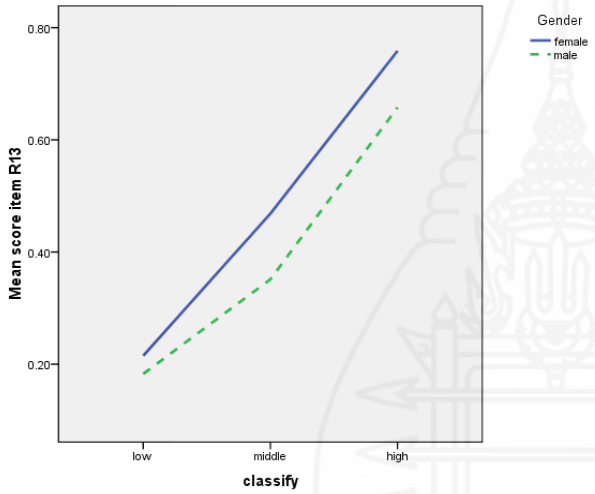

Gender
female
male
Mean score item R8
0.80
0.60
0.40
0.20
low
middle
high
classify

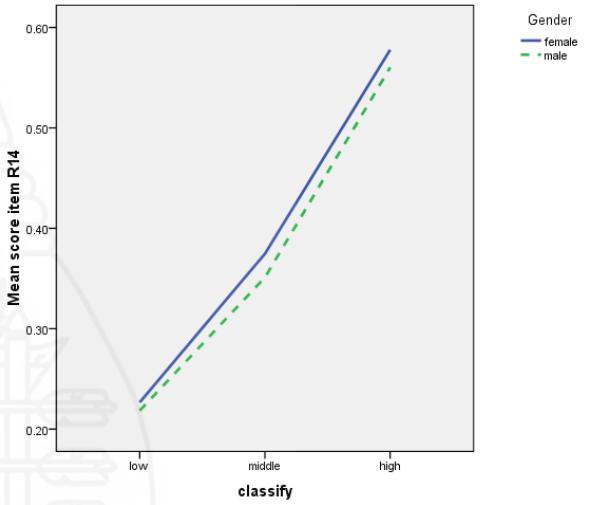

Gender
female
male
Mean score item R9
0.80
0.70
0.60
0.50
0.40
0.30
0.20
low
middle
high
classify

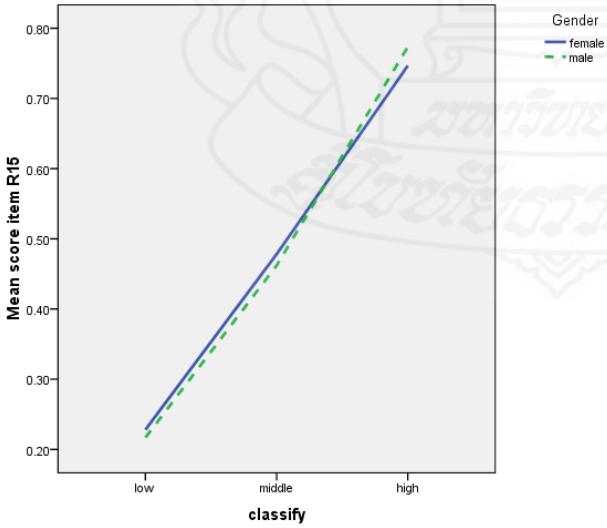

Gender
female
male
Mean score item R10
0.80
0.70
0.60
0.50
0.40
0.30
0.20
low
middle
high
classify

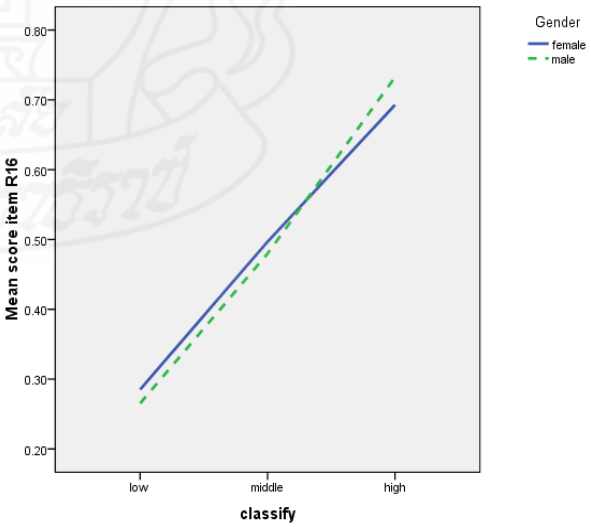

Gender
female
male
Mean score item R11
0.80
0.60
0.40
0.20
low
middle
high
classify

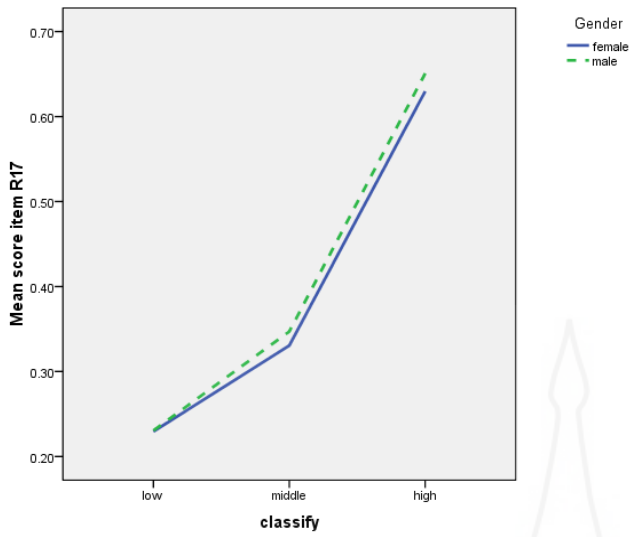

Gender
female
male
Mean score item R12
1.00
0.80
0.60
0.40
low
middle
high
classify

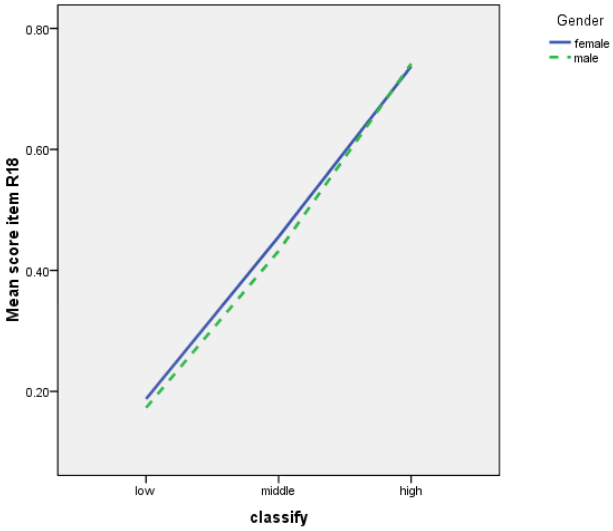

Gender
female
male
Mean score item R13
0.80
0.60
0.40
0.20
low
middle
high
classify

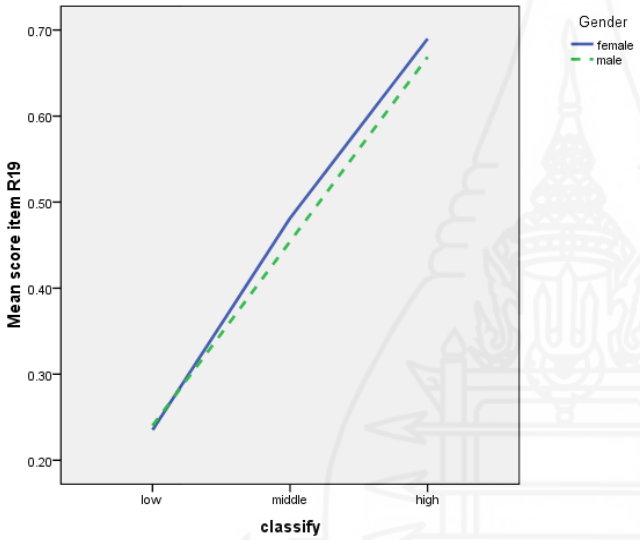

Gender
female
male
Mean score item R14
0.60
0.50
0.40
0.30
0.20
low
middle
high
classify

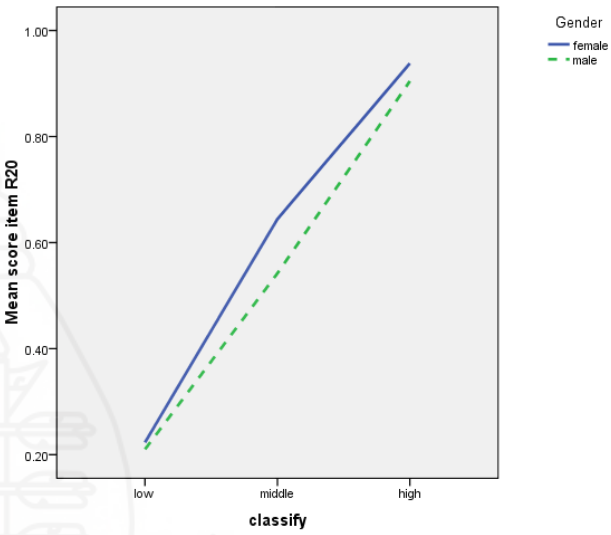

Gender
female
male
Mean score item R15
0.80
0.70
0.60
0.50
0.40
0.30
0.20
low
middle
high
classify

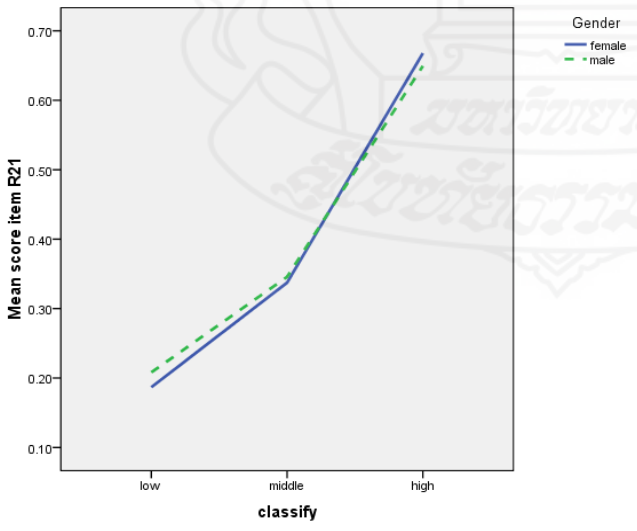

Gender
female
male
Mean score item R16
0.80
0.70
0.60
0.50
0.40
0.30
0.20
low
middle
high
classify

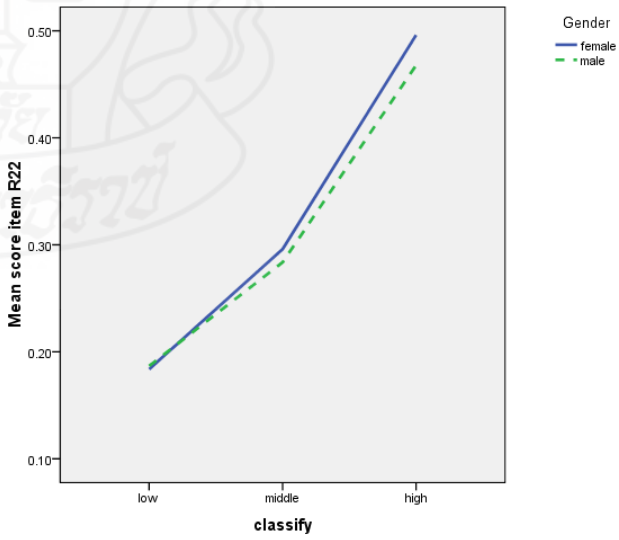

Mean score item R17
0.70
0.60
0.50
0.40
0.30
0.20
low
middle
high
classify
Gender
female
male

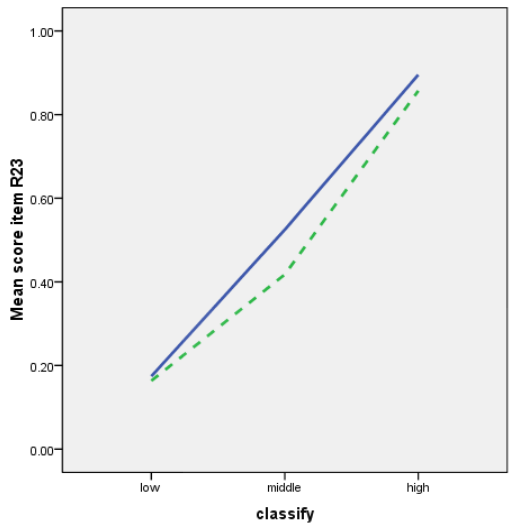

Mean score item R18
0.80
0.60
0.40
0.20
low
middle
high
classify
Gender
female
male

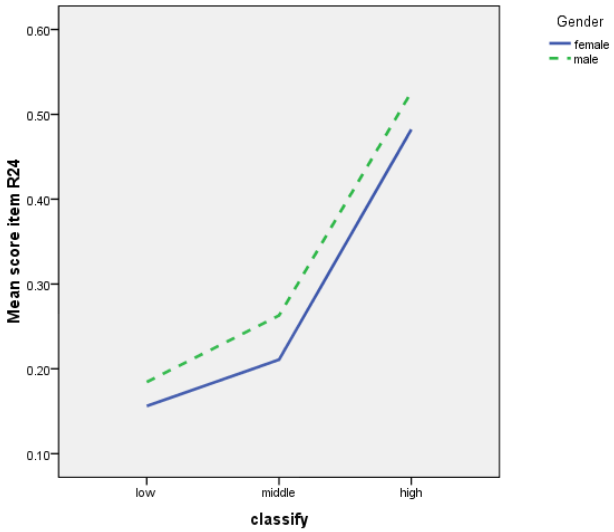

Mean score item R19
0.70
0.60
0.50
0.40
0.30
0.20
low
middle
high
classify
Gender
female
male

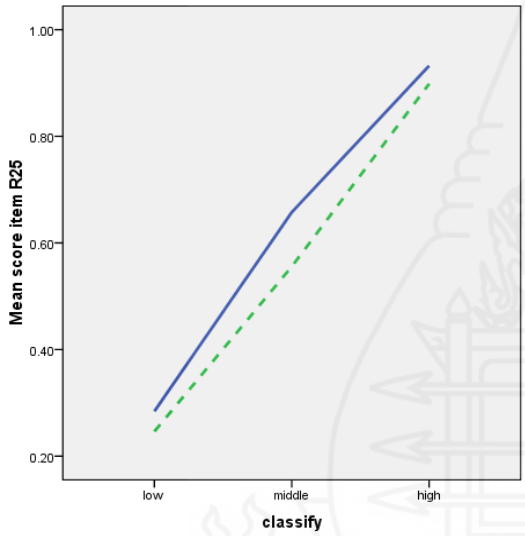

Mean score item R20
1.00
0.80
0.60
0.40
0.20
low
middle
high
classify
Gender
female
male

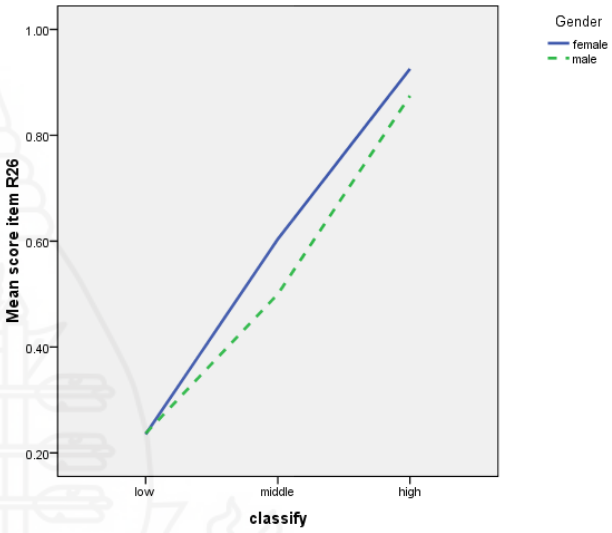

Mean score item R21
0.70
0.60
0.50
0.40
0.30
0.20
0.10
low
middle
high
classify
Gender
female
male

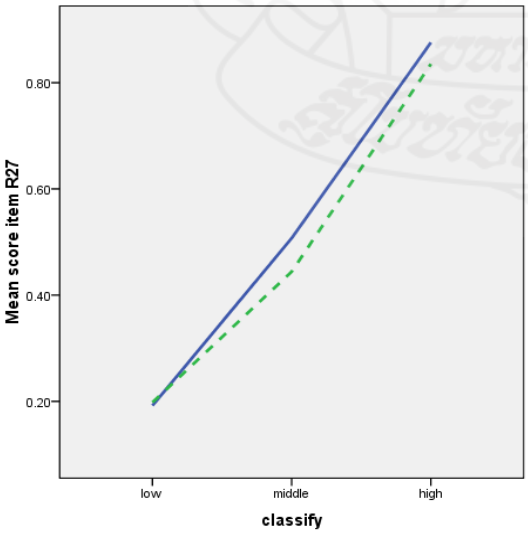

Mean score item R22
0.50
0.40
0.30
0.20
0.10
low
middle
high
classify
Gender
female
male


Mean score item R23
1.00
0.80
0.60
0.40
0.20
0.00
low
middle
high
classify
Gender
female
male


Mean score item R24
0.60
0.50
0.40
0.30
0.20
0.10
low
middle
high
classify
Gender
female
male


Mean score item R25
1.00
0.80
0.60
0.40
0.20
low
middle
high
classify
Gender
female
male


Mean score item R26
1.00
0.80
0.60
0.40
0.20
low
middle
high
classify
Gender
female
male


Mean score item R27
0.80
0.60
0.40
0.20
low
middle
high
classify
Gender
female
male

# บทที่ 5
# สรุปการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีจุดประสงค์ เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง และเพื่อวิเคราะห์สาเหตุการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของการทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง โดยข้อมูลที่ใช้เป็นข้อมูลทุติยภูมิ ซึ่งเป็นผลการตอบข้อสอบความสามารถด้านเหตุผล ของนักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2557 ของสำนักทดสอบทางการศึกษาของสำนักงานการศึกษาขั้นพื้นฐาน มีนักเรียนทั้งข้อมูลได้ทั้งหมด 464,532 คน แบ่งเป็นนักเรียนชาย 224,615 คน และนักเรียนหญิง 239,917 คน

เครื่องมือที่ใช้ในการศึกษาครั้งนี้ คือ แบบทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 27 ข้อ สร้างโดยสำนักทดสอบทางการศึกษาสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

ข้อมูลที่ใช้ในการวิจัย ได้แก่ คะแนนผลการสอบวัดความสามารถด้านเหตุผลชั้นประถมศึกษาปีที่ 3 ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ปีการศึกษา 2557 จำนวน 464,532 คน

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีของถดถอยโลจิสติก เป็นการเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบของกลุ่มผู้สอบ 2 กลุ่ม ที่เรียกว่า กลุ่มอ้างอิง และกลุ่มเปรียบเทียบ การวิเคราะห์ผลการทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อของ Bruno and Zumbo (1999) ในการวิจัยครั้งนี้ได้ดำเนินการวิเคราะห์ข้อมูล ดังนี้

1. คำนวณค่าสถิติพื้นฐานเพื่อบรรยายลักษณะการแจกแจงของคะแนนจากแบบทดสอบวัดความสามารถด้านเหตุผล ด้วยโปรแกรมคอมพิวเตอร์ SPSS

2. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการศึกษาครั้งนี้ใช้วิธีการถดถอยโลจิสติก เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบของแบบทดสอบระดับชาติวัดความสามารถด้านเหตุผลระหว่างนักเรียนชายและนักเรียนหญิง โดยกำหนดให้นักเรียนชายเป็นกลุ่มอ้างอิง และนักเรียนหญิงเป็นกลุ่มเปรียบเทียบ และขั้นตอนในการวิเคราะห์เป็นลำดับขั้นคือ ขั้นที่ 1 ขั้นเลือกควบคุมของตัวแปร ขั้นที่ 2 เป็นการนำตัวแปรเพศเข้าสู่สมการโลจิสติก และขั้นที่ 3 เป็นการศึกษาปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับกลุ่มผู้สอบ สำหรับการ

ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีถดถอยโลจิสติกทดสอบค่านัยสำคัญทางสถิติ .05 ค่าไคล์-สแควร์ และวัดขนาดการทำหน้าที่ต่างกันของข้อสอบ โดยใช้เกณฑ์ Jodoin and Gierl ในการพิจารณาร่วมกันในการวิเคราะห์ผล

3. การวิเคราะห์สาเหตุของการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิงนั้นวิเคราะห์จากค่า R-square เพื่อหาการวัดขนาดของการทำหน้าที่ต่างกันของข้อสอบโดยใช้เกณฑ์ Jodoin and Gierl ในตัดสินการทำหน้าที่ต่างกันของข้อสอบและนำผลการวิเคราะห์ให้ผู้เชี่ยวชาญพิจารณาถึงสาเหตุของการทำหน้าที่ต่างกันของข้อสอบ

## 1. สรุปการวิจัย

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลระหว่างนักเรียนชายและนักเรียนหญิงจากผลการทดสอบระดับชาติ โดยใช้วิธีถดถอยโลจิสติก ได้ผลดังนี้

1.1 การวิเคราะห์การทำหน้าที่ต่างกันของนักเรียนชายและนักเรียนหญิงโดยพิจารณาควบคู่กับขนาดของการทำหน้าที่ต่างกันของข้อสอบ ตามหลักการของ Jodoin and Gierl ดังนี้

1.1.1 ข้อสอบที่มีการทำหน้าที่ต่างกันและมีการเอนเอียงทางนักเรียนชายหรือนักเรียนหญิงเมื่อพิจารณาจากตารางได้แก่

1) ข้อสอบที่มีการเอนเอียงไปทางนักเรียนชาย คือ 1 2 5 6 8 10 11 15 16 17 18 21 และ 24

2) ข้อสอบที่มีการเอนเอียงไปทางนักเรียนหญิง คือ 3 4 7 9 12 13 14 19 20 22 23 25 26 และ 27

1.1.2 ขนาดของการทำหน้าที่ต่างกันของข้อสอบ พบว่า ทุกข้อมีขนาดของการทำหน้าที่ต่างกันเพียงขนาดเล็ก ดังนั้นเมื่อพิจารณาโดยรวมแล้วข้อสอบทุกข้อทำหน้าที่ไม่แตกต่างกัน

1.1.3 การทำหน้าที่ต่างกันของข้อสอบมี 2 แบบ คือ

1.3.1 ข้อสอบที่ทำหน้าที่ต่างกันแบบเอกรูป จำนวน 2 ข้อ คือ ข้อ 2 และ 10 คือทุกระดับความสามารถนักเรียนชายมีโอกาสตอบถูกดีกว่านักเรียนหญิง

1.3.2 ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป จำนวน 25 ข้อ1 3 4 5 6 7 8 9 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 พบว่าโอกาสในการตอบข้อสอบถูกระหว่าง นักเรียนชายและนักเรียนหญิงไม่สม่ำเสมอกันเช่นนักเรียนชายบางระดับ

ความสามารถมีโอกาสตอบถูกได้ดีกว่านักเรียนหญิงและนักเรียนหญิงบางระดับความสามารถมีโอกาสตอบถูกได้ดีกว่านักเรียนชาย

ดังนั้นจึงสรุปได้ว่า ข้อสอบทุกข้อทำหน้าที่ไม่แตกต่างกัน และมีขนาดของการทำหน้าที่ต่างกันของข้อสอบมีขนาดเล็ก ซึ่งถือได้ว่ายังอยู่ในเกณฑ์ดีและข้อสอบวัดความสามารถด้านเหตุผลของการทดสอบระดับชาติมีความยุติธรรมทั้งนักเรียนชายและนักเรียนหญิง

1.2 การวิเคราะห์สาเหตุการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของการทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง

เมื่อวิเคราะห์สาเหตุของการทำหน้าที่ต่างกันของข้อสอบ โดยวิเคราะห์ข้อสอบที่มีขนาดอิทธิพลขนาดกลางและขนาดใหญ่ ข้อสอบด้านเหตุผลของแบบทดสอบระดับชาติมีขนาดของการทำหน้าที่ต่างกันเล็กมาก เนื่องจากนักเรียนชายและนักเรียนหญิงมีความสนใจที่แตกต่างกันเป็นข้อสอบเกี่ยวกับความรู้ความเข้าใจในข้อมูล วิเคราะห์ข้อมูล สถานการณ์ หรือสารสนเทศโดยใช้ความรู้ด้านวิทยาศาสตร์และสิ่งแวดล้อม ด้านสังคมและเศรษฐกิจ และด้านการดำเนินชีวิตอย่างมีเหตุผล ข้อสอบส่วนใหญ่ทำหน้าที่ต่างกันแบบอเนกรูป ลำเอียงเข้าหากลุ่มอ้างอิง เมื่อวิเคราะห์สาเหตุ พบว่า นักเรียนชายและนักเรียนหญิงมีความสนใจที่แตกต่างกัน ข้อสอบมีลักษณะของกิจกรรมที่เหมาะกับนักเรียนหญิงและมีโอกาสพบเจอหรือสัมผัสกับกิจกรรมนั้นอยู่เสมอก็จะตอบข้อสอบข้อนั้นได้ดีกว่านักเรียนชาย ข้อสอบมีลักษณะของกิจกรรมที่เหมาะกับเพศชายหรือนักเรียนชายมีความสนใจก็จะตอบข้อสอบข้อนั้นได้ดีกว่า เช่น กิจกรรมงานช่าง ความรู้วิทยาศาสตร์ ผจญภัยต่างๆ นักเรียนชายชอบแข่งขัน แก้ไขปัญหาจากสถานการณ์ต่าง ๆ ได้อย่างมีเหตุผล จากข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง พบว่านักเรียนชายสามารถทำข้อสอบได้มากกว่านักเรียนหญิง 1 ข้อ สรุปได้ว่า ข้อสอบที่เกี่ยวกับการคิดวิเคราะห์ นักเรียนชายจะทำข้อสอบได้ดีกว่านักเรียนหญิง สอดคล้องกับ อุบลวรรณ เสือเดช (2550) กล่าวว่า เพศส่งผลต่อความสามารถในการคิดวิเคราะห์ของนักเรียนและกลุ่มเพศชายมีความสามารถในการวิเคราะห์แตกต่างจากกลุ่มเพศหญิง

## 2. อภิปรายผล

ผู้วิจัยขออภิปรายในประเด็นการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบในการตรวจสอบผลการทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบ

ระดับชาติ พบว่าเมื่อคำนวณด้วยวิธีถดถอยโลจิสติกพิจารณาค่า $\chi^2$ ของ Bruno and Zumbo (1999, p. 26) แล้วพบข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบเป็นจำนวนมาก แต่ถ้าพิจารณาร่วมกับขนาดของการทำหน้าที่ต่างกันของข้อสอบ Jodoin and Gierl พบว่าขนาดของการทำหน้าที่ต่างกัน มีขนาดเล็ก ซึ่งไม่มีอิทธิพลหรือส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ ดังนั้นจึงสรุปได้ว่า ไม่พบข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบหรือทำหน้าที่ไม่แตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 จากผลการวิจัยการทำหน้าที่ต่างกันของข้อสอบของแบบทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 พบว่านักเรียนชายและนักเรียนหญิงมีโอกาสตอบข้อสอบได้ไม่แตกต่างกันทุกระดับความสามารถ หมายความว่านักเรียนหญิงและนักเรียนชายมีโอกาสตอบข้อสอบถูกเท่ากัน ซึ่งแสดงให้เห็นว่าแบบทดสอบวัดความสามารถด้านเหตุผล ชั้นประถมศึกษาปีที่ 3 มีความยุติธรรมทุกคน สามารถนำผลการทดสอบไปใช้อ้างอิงพัฒนาการของผู้เรียน ได้และถือว่าแบบทดสอบมีคุณภาพสามารถนำไปเก็บในคลังข้อสอบเพื่อลดค่าใช้จ่ายในการออกข้อสอบ

ส่วนสาเหตุของการทำหน้าที่ต่างกันของข้อสอบนั้นผู้เชี่ยวชาญได้พบคือ 1) ข้อคำถามของข้อสอบมีลักษณะของกิจกรรมที่เหมาะกับเพศใดเพศหนึ่ง 2) ภาษาที่ใช้ การใช้สำนวน นักเรียนหญิงมีโอกาสตอบข้อสอบข้อนั้นได้ดีกว่านักเรียนชาย เป็นต้นและยังพบว่าข้อสอบบางข้อมีขนาดอิทธิพลที่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบนั้นมีเพียงเล็กจึงไม่มีอิทธิพลที่เพศมีผลต่อการตอบข้อสอบถูก

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยโลจิสติก ของ Bruno and Zumbo (1999) นี้ถือว่าเป็นวิธีการที่ถูกออกแบบมาสำหรับตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบเอกรูปและอเนกรูป และการวัดขนาดการทำหน้าที่ต่างกันของข้อสอบได้ถูกพัฒนาขึ้นจากวิธีถดถอยโลจิสติก การตัดสินขนาดของการทำหน้าที่ต่างกันของข้อสอบของ Jodoin และ Gierl (2001) ถือว่าเป็นวิธีที่ดี เข้าใจง่ายและสะดวกในการวิเคราะห์ข้อสอบ การวิเคราะห์ข้อสอบสามารถวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและขนาดอิทธิพลของข้อสอบไปพร้อมกันได้ ซึ่งถือว่าลดขั้นตอนของการวิเคราะห์ข้อสอบได้

## 3. ข้อเสนอแนะ

การวิจัยครั้งนี้เป็นการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบวัดความสามารถด้านเหตุผลของแบบทดสอบระดับชาติระหว่างนักเรียนชายและนักเรียนหญิง ด้วยวิธีการถดถอยโลจิสติก (Logistic Regression ) เป็นลำดับขั้นในการวิเคราะห์ข้อสอบด้วยพารามิเตอร์ของข้อสอบ ซึ่งผลการวิจัยพบว่า ข้อสอบวัดความสามรถด้านเหตุผลของแบบทดสอบระดับชาติ ทำหน้าที่ไม่

ต่างกัน แสดงว่านักเรียนชายและนักเรียนหญิงมีโอกาสตอบข้อสอบข้อนั้นถูกเท่ากัน ดังนั้นจากผลการวิเคราะห์ข้อสอบทำให้ได้ข้อมูลสารสนเทศเกี่ยวกับข้อสอบที่ไม่ทำหน้าที่ต่างกันของข้อสอบ ซึ่งเป็นประโยชน์ต่อครูผู้สอนในการพัฒนาข้อสอบและแบบสอบเพื่อยกระดับคุณภาพของสถานศึกษา

การวิจัยครั้งต่อไปควรศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบลักษณะทำนองเดี่ยวกันในตัวแปรอื่น เช่น เชื้อชาติ ศาสนา ภูมิลำเนา ภูมิภาค เป็นต้น เป็นเกณฑ์ในการแบ่งกลุ่มเพื่อให้ได้ข้อมูลสารสนเทศมากยิ่งขึ้นและยืนยันข้อสอบอีกครั้ง โดยเฉพาะการศึกษาด้านภูมิภาคของนักเรียนนั้นจะเห็นชัดเจนถึงการทำหน้าที่ต่างกันของข้อสอบ

# ประวัติผู้วิจัย

| | |
|---|---|
| **ชื่อ** | นางสาวนิตยา ดวงจันทร์ทิพย์ |
| **วัน เดือน ปี เกิด** | 6 กันยายน 2528 |
| **สถานที่เกิด** | อำเภอเมืองบึงกาฬ จังหวัดบึงกาฬ |
| **ประวัติการศึกษา** | ครุศาสตรบัณฑิต (คณิตศาสตร์) มหาวิทยาลัยราชภัฏอุดรธานี ปี 2552 |
| **สถานที่ทำงาน** | โรงเรียนวัดด่านทับตะโก อำเภอจอมบึง จังหวัดราชบุรี |
| **ตำแหน่ง** | ครู |